the r.m.s. difference in B-factors of bonded atoms is 4.1 Å$^2$.

Received 21 January; accepted 5 July 1999.


1. Snider, W. D. Functions of the neurotrophins during nervous system development: what the knockouts are teaching us. *Cell* **77,** 627–638 (1994).
2. Chao, M. *et al.* Neurotrophin receptors: mediators of life and death. *Brain Res. Rev.* **26,** 295–301 (1998).
3. Kaplan, D. R., Hempstead, B. L. Martin-Zanca, D., Chao, M. V. & Parada, L. F. The trk proto-oncogene product: a signal transducing receptor for nerve growth factor. *Science* **252,** 554–558 (1991).
4. Urfer, R. *et al.* An immunoglobulin-like domain determines the specificity of neurotrophin receptors. *EMBO J.* **14,** 2795–2805 (1995).
5. Chao, M. V. Neurotrophin receptors: a window into neuronal differentiation. *Neuron* **9,** 583–593 (1992).
6. Götz, R. *et al.* Neurotrophin-6 is a new member of the nerve growth factor family. *Nature* **372,** 366–269 (1994).
7. Schneider, R. & Schweiger, M. A novel modular mosaic of cell adhesion motifs in the extracellular domains of the neurogenic trk and trkB tyrosine kinase receptors. *Oncogene* **6,** 1807–1811 (1991).
8. Peréz, P., Coll, P. M., Hempstead, B. L., Martin-Zanca, D. & Chao, M. V. NGF binding to the trk tyrosine kinase receptor requires the extracellular immunoglobulin-like domains. *Mol. Cell. Neurosci.* **6,** 97–105 (1995).
9. Ultsch, M. *et al.* Crystal structures of the neurotrophin-binding domain of TrkA, TrkB and TrkC. *J. Mol. Biol.* **290,** 149–159 (1999).
10. McDonald, N. Q. *et al.* New protein fold revealed by a 2.3-Å resolution crystal structure of nerve growth factor. *Nature* **354,** 411–414 (1991).
11. McInnes, C. & Sykes, B. D. Growth factor receptors: structure, mechanism, and drug discovery. *Biopolymers* **43,** 339–366 (1997).
12. Urfer, R. *et al.* The binding epitopes of neurotrophin-3 to its receptors trkC and gp75 and the design of a multifunctional human neurotrophin. *EMBO J.* **13,** 5896–5909 (1994).
13. Urfer, R. *et al.* High resolution mapping of the binding site of TrkA for nerve growth factor and TrkC for neurotrophin-3 on the second immunoglobulin-like domain of the Trk receptors. *J. Biol. Chem.* **273,** 5829–5840 (1998).
14. Ibáñez, C. F., Ebendal, T. & Persson, H. Chimeric molecules with multiple neurotrophic activities reveal structural elements determining the specificities of NGF and BDNF. *EMBO J.* **10,** 2105–2110 (1991).
15. Kullander, K. & Ebendal, T. Neurotrophin-3 acquires NGF-like activity after exchange to five NGF amino acid residues: molecular analysis of the sites in NGF mediating the specific interaction with the NGF high affinity receptor. *J. Neurosci. Res.* **39,** 195–210 (1994).
16. Ibáñez, C. F., Ilag, L. L., Murray-Rust, J. & Persson, H. An extended surface of binding to Trk tyrosine kinase receptors in NGF and BDNF allows the engineering of a multifunctional pan-neurotrophin. *EMBO J.* **12,** 2281–2293 (1993).
17. O'Leary, P. D. & Hughes, R. A. Structure-activity relationships of conformationally constrained peptide analogues of loop 2 of brain-derived neurotrophic factor. *J. Neurochem.* **70,** 1712–1721 (1998).
18. Urfer, R., Tsoulfas, P., O'Connell, L. & Presta, L. G. Specificity determinants in neurotrophin-3 and design of nerve growth factor-based trkC agonists by changing central beta-strand bundle residues to their neurotrophin-3 analogs. *Biochemistry* **36,** 4775–4781 (1997).
19. Windisch, J. M., Marksteiner, R. & Schneider, R. Nerve growth factor binding site on TrkA mapped to a single 24-amino acid leucine-rich motif. *J. Biol. Chem.* **270,** 28133–28138 (1995).
20. Windisch, J. M., Marksteiner, R., Lang, M. E., Auer, B. & Schneider, R. Brain-derived neurotrophic factor, neurotrophin-3, and neurotrophin-4 bind to a single leucine-rich motif of TrkB. *Biochemistry* **34,** 11256–11263 (1995).
21. Dechant, G., Tsoulfas, P., Parada, L. F. & Barde, Y. A. The neurotrophin receptor p75 binds neurotrophin-3 on sympathetic neurons with high affinity and specificity. *J. Neurosci.* **17,** 5271–5287 (1997).
22. Ibáñez, C. F. *et al.* Disruption of the low affinity receptor-binding site in NGF allows neuronal survival and differentiation by binding to the trk gene product. *Cell* **69,** 329–341 (1992).
23. Ryden, M. & Ibanez, C. F. A second determinant of binding to the p75 neurotrophin receptor revealed by alanine-scanning mutagenesis of a conserved loop in nerve growth factor. *J. Biol. Chem.* **272,** 33085–33091 (1997).
24. Wiesmann, C. *et al.* Crystal structure at 1.7 Å resolution of VEGF in complex with domain 2 of the Flt-1 receptor. *Cell* **91,** 695–704 (1997).
25. Sun, P. D. & Davies, D. R. The cystine-knot growth-factor superfamily. *Annu. Rev. Biophys. Biomol. Struct.* **24,** 269–291 (1995).
26. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
27. CCP4. Programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
28. Jones, T. A., Zhou, J.-Y., Cowan, S. W. & Kjelgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
29. Brünger, A. T., Kuriyan, J. & Karplus, M. Crystallographic R factor refinement by molecular dynamics. *Science* **235,** 458–460 (1987).
30. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. Procheck: a program to check to the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26,** 283–291 (1993).



### Acknowledgements

We thank the following colleagues at Genentech: E. Martin and G. Burton for purified NGF; D. Reilly for fermentation runs; W. Henzel for N-terminal sequencing; J. Bourell for mass spectrometry analysis; H. Christinger for help with data collection; and L. Presta and D. Shelton for helpful discussions. We also thank the staff at SSRL, beam line 9-1, and ALS, beam line 5.2; and D. Dawbarn for sharing data before publication.


Correspondence and requests for materials should be directed to A.M.D.V. (e-mail: devos@gene.com). The coordinates have been deposited with the Protein Data Bank for immediate release (accession number 1www).

# Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors

**Michael S. Finnin*, Jill R. Donigian†, Alona Cohen†, Victoria M. Richon‡, Richard A. Rifkind‡, Paul A. Marks‡, Ronald Breslow§ & Nikola P. Pavletich***

† *Cellular Biochemistry and Biophsyics Program and * Howard Hughes Medical Institute,* ‡ *Cell Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA*

§ *Department of Chemistry, Columbia University, New York, New York 10027, USA*

**Histone deacetylases (HDACs) mediate changes in nucleosome conformation and are important in the regulation of gene expression[1]. HDACs are involved in cell-cycle progression and differentiation, and their deregulation is associated with several cancers[2,3]. HDAC inhibitors, such as trichostatin A (TSA) and suberoylanilide hydroxamic acid (SAHA), have anti-tumour effects, as they can inhibit cell growth[4–6], induce terminal differentiation[4,5] and prevent the formation of tumours in mice models[7,8], and they are effective in the treatment of promyelocytic leukemia[3]. Here we describe the structure of the histone deacetylase catalytic core, as revealed by the crystal structure of a homologue from the hyperthermophilic bacterium *Aquifex aeolicus*, that shares 35.2% identity with human HDAC1 over 375 residues, deacetylates histones *in vitro* and is inhibited by TSA and SAHA. The deacetylase, deacetylase–TSA and deacetylase–SAHA structures reveal an active site consisting of a tubular pocket, a zinc-binding site and two Asp–His charge-relay systems, and establish the mechanism of HDAC inhibition. The residues that make up the active site and contact the inhibitors are conserved across the HDAC family. These structures also suggest a mechanism for the deacetylation reaction and provide a framework for the further development of HDAC inhibitors as anti-tumour agents.**

HDACs catalyse the removal of acetyl groups for the ϵ-amino groups of lysine residues clustered near the amino terminus of nucleosomal histones, and this process is associated with transcriptional repression[9]. The deregulation of HDAC recruitment to promoters appears to be one of the mechanisms by which these enzymes contribute to tumorigenesis. For example, in acute promyelocytic leukemia (APL), chromosomal translocations fuse the retinoic acid receptor-α (RARα) to either the promyelocytic leukaemia zinc finger (PLZF) or the promyelocytic leukaemia protein (PML), which recruit a co-repressor and, in turn, HDACs[3]. There are two related classes of eukaryotic HDACs (Class I and II; refs 10–12) and they share a ~390-amino-acid region of homology—the deacetylase core. The deacetylase core belongs to a superfamily[13] of genes that includes the *A. aeolicus* HDAC homologue (35.2% identity with HDAC1), the prokaryotic acetoin utilization proteins (AcuC; 28.1% identity with HDAC1) and the prokaryotic acetylpolyamine amidohydrolases (APAH; 15.0% identity with HDAC1).

The function of the *A. aeolicus* HDAC homologue is unknown, but it can deacetylate histones *in vitro* with a specific activity about 7.5% of that of insect cell-expressed and affinity-purified FLAG-tagged HDAC1 (typically, 7,500 and 100,000 cpm of [$^3$H]acetate released per nmol of enzyme by the HDAC homologue and FLAG-tagged HDAC1, respectively). The *A. aeolicus* HDAC homologue is inhibited by TSA (half-maximal inhibition at 0.4 μM) and SAHA and binds these inhibitors in a gel-filtration chromatography assay (data not shown). The *in vitro* deacetylase activity of the purified HDAC homologue was observed only after incubation with zinc

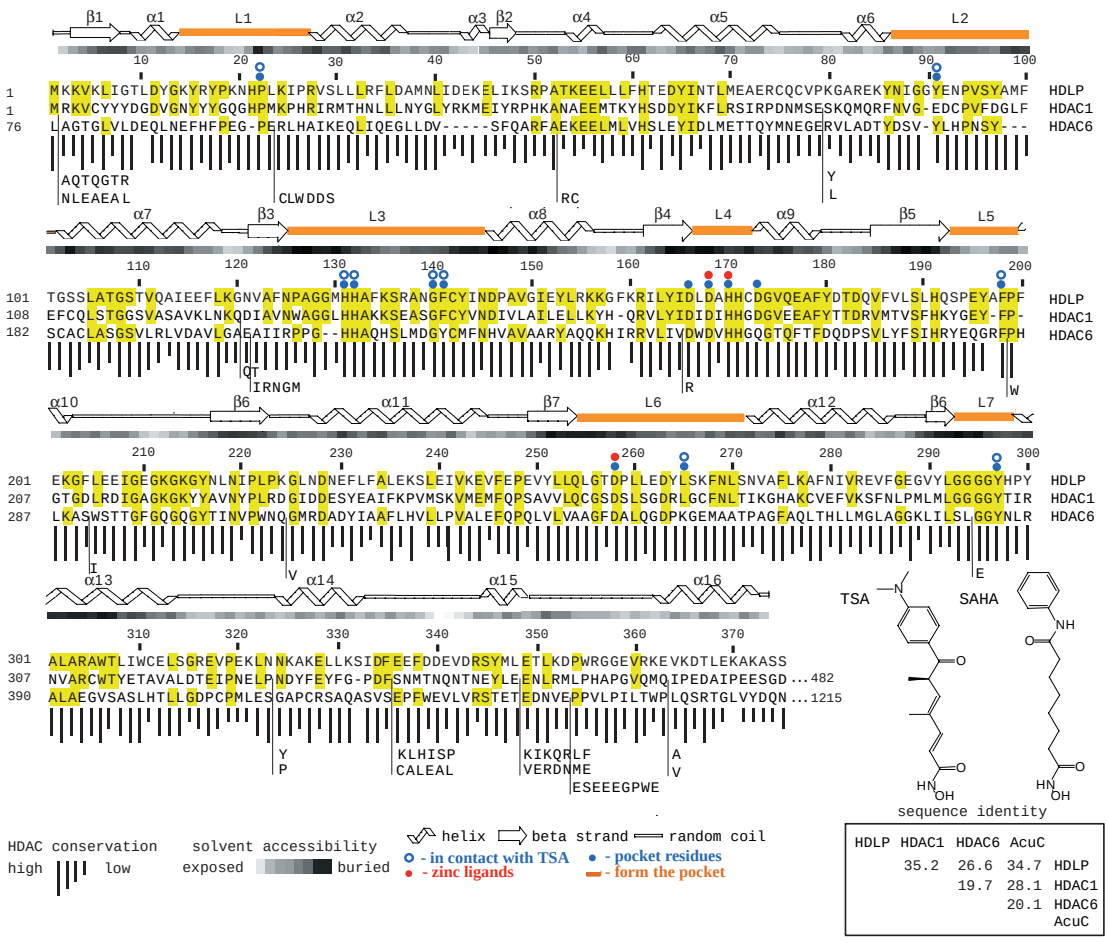


**Figure 1** The *A. aeolicus* HDLP has 35.2% identity to human HDAC1. Sequence alignment of HDLP, human class I HDAC1 and class II HDAC6. Residues identical between HDLP and HDAC1 or HDLP and HDAC6 are highlighted in yellow. The bar graph below the sequence alignment indicates the degree of identity among five class I HDAC genes: HDAC1, and its homologues from thale cress, fruit fly, frog and yeast (RPD3). Insertions in HDAC1 and HDAC6 sequences are dropped below the alignment with a vertical line for clarity. Loops that form the active-site pocket are highlighted orange, and residues that make up the pocket walls and active site and those that contact TSA are indicated. The chemical structures of TSA and SAHA are shown. Class I is characterized by human HDAC1, 2, 3 and yeast RPD3, and class II by human HDAC4, 5, 6 and yeast HDA1 (refs 10–12). The function of HDLP is not known. It has been annotated as *AcuC1* in the genome sequence according to its homology to the *B. subtilis AcuC* acetoin utilization protein, whose precise function is also not known. However, the *A. aeolicus* genome appears to lack the *acuA* and *acuB* genes that are part of the *acuABC* operon of *B. subtilis*[27], and HDLP is as similar to human HDAC1 (35.2% identity) as it is to *B. subtilis* AcuC (34.7% identity).

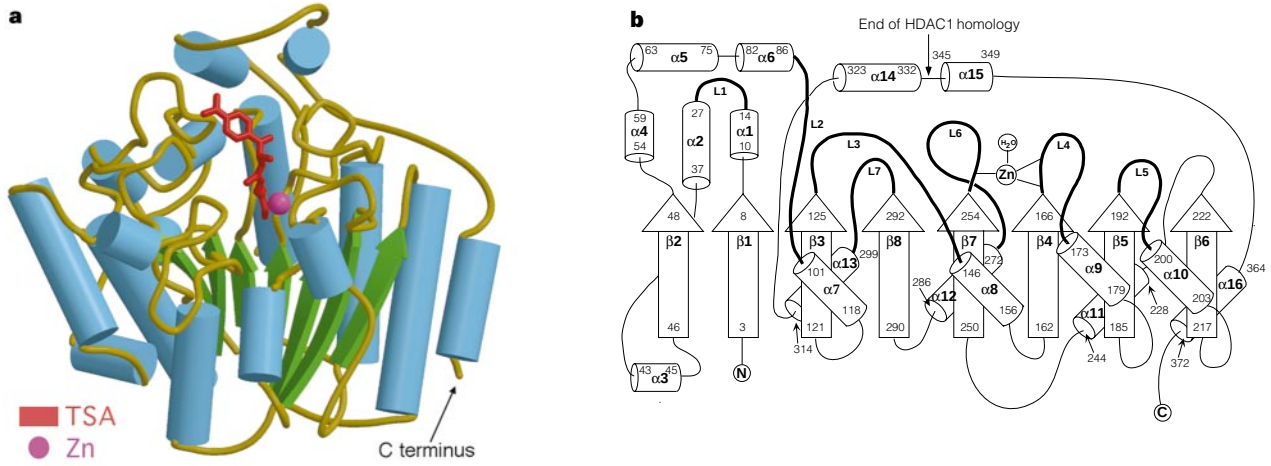


**Figure 2** The histone deacetylase catalytic core belongs to the open α/β-family of folds. **a**, HDLP–$Zn^{2+}$–TSA complex. Figure prepared with the programs MOLSCRIPT[28] and RASTER3D[29]. **b**, Topology diagram for HDLP. Loops forming the active-site pocket are in bold. The end of the deacetylase motif is indicated with an arrow. HDLP has the same topology as arginase[30], which catalyses the hydrolysis of arginine to ornithine, and the two structures can be superimposed with a Cα r.m.s.d. of 2.5 Å for about 35% of the residues, mapping primarily to the β-sheet. The arginase structure differs in that it has only nine helices instead of sixteen, a two-manganese metal cluster and a wider active site that lacks the internal cavity.

chloride, consistent with previous suggestions that HDAC activity requires a metal cofactor[14], and with our observation that the zinc chelator phenanthroline inhibits HDAC1 (data not shown). No significant activity was observed with HDAC homologue preparations reconstituted with $Ca^{2+}$, $Cu^{2+}$, $Fe^{2+}$, $Mg^{2+}$ or $Mn^{2+}$, whereas $Co^{2+}$, which can often substitute for zinc *in vitro*, produced a comparable level of activity (data not shown). Here we refer to the *A. aeolicus* HDAC homologue as HDLP (histone deacetylase-like protein) to reflect its sequence homology to the HDAC family (Fig. 1), its *in vitro* histone deacetylase activity and its inhibition by the HDAC inhibitors.

The crystal structure reveals that HDLP has a single-domain structure belonging to the open α/β class of folds. It consists of a central eight-stranded parallel β-sheet and sixteen α-helices (Fig. 2). Four helices pack on either face of the β-sheet, forming the core α/β structure characteristic of this class of folds. Most of the remaining eight helices are clustered near one side of the β-sheet (Fig. 2). Large, well-defined loops originate from the carboxy-terminal ends of the β-strands (loops L1–L7; Fig. 2b). The extra helices and the large L1–L7 loops are associated with a significant extension of the structure beyond the core α/β motif. This gives rise to two prominent architectural features: a deep, narrow pocket and an internal cavity adjacent to the pocket (Fig. 3a).

The pocket has a tube-like shape with a depth of ~11 Å. The pocket opening constricts halfway down to ~4.5 by 5.5 Å, but becomes wider at the bottom (Fig. 3a). The pocket and its immediate surroundings are made up of seven loops. These originate from strands near the centre of the sheet (L3–L7) and from the additional helices that pack near the ends of the strands (L1 and L2; Figs 1, 2b). The walls of the pocket are covered with hydrophobic and aromatic residues that are identical in HDAC1 (Pro 22, Gly 140, Phe 141, Phe 198, Leu 265 and Tyr 297; Fig. 3b). Of particular significance are Phe 141 and Phe 198, whose phenyl groups face each other in parallel at a distance of 7.5 Å, marking the most slender portion of the pocket (Fig. 3b, c).

In the crystal structure of the zinc-reconstituted enzyme, the zinc ion is positioned near the bottom of the pocket, just beyond its narrowest point. It is coordinated by Asp 168 (Oδ1, 2.1 Å), His 170 (Nδ1, 2.1 Å), Asp 258 (Oδ1, 1.9 Å) and a water molecule (2.5 Å). The protein ligands are arranged in a tetrahedral geometry, but the position of the water molecule, which also hydrogen-bonds to His 131, deviates from this geometry by ~25°.

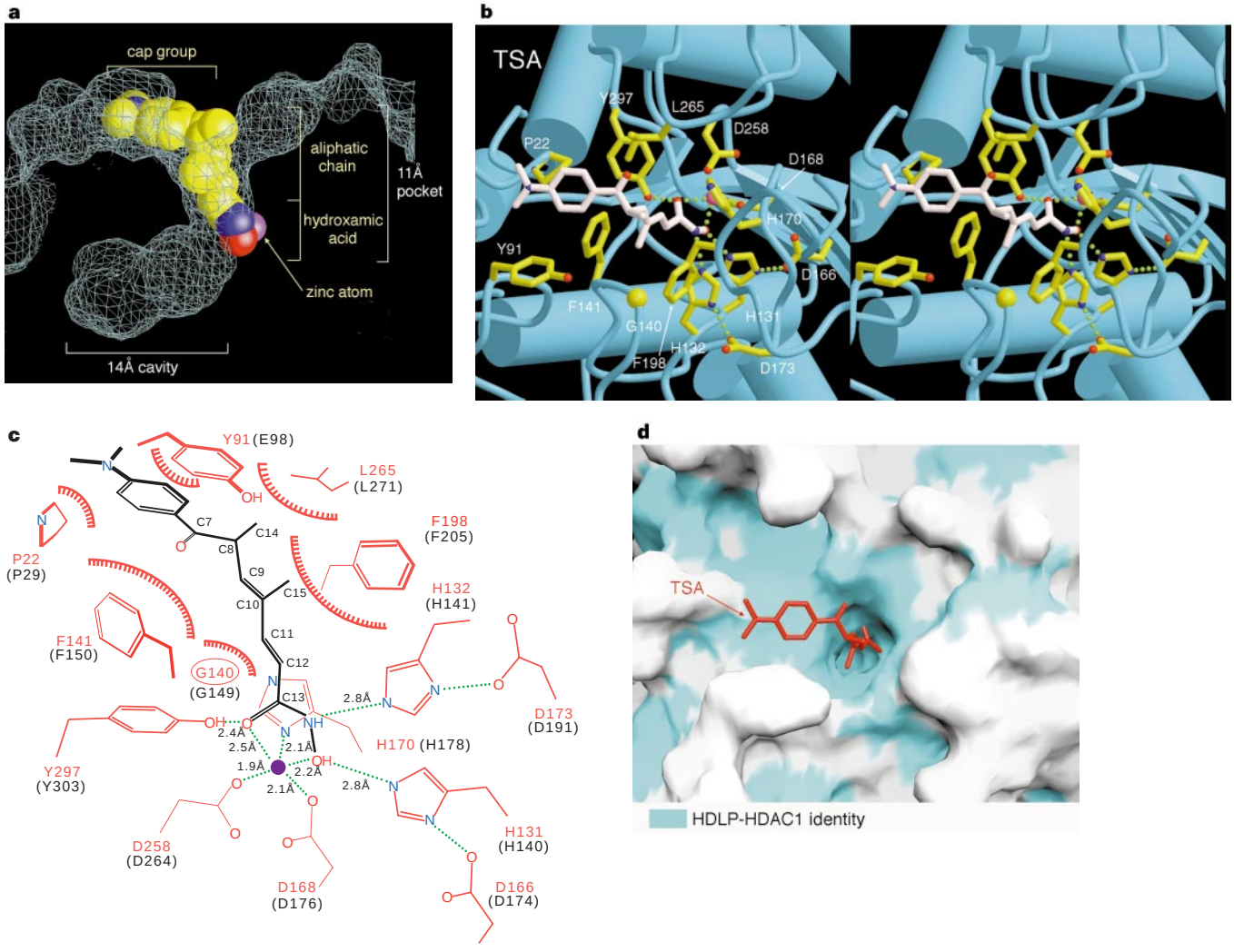


**Figure 3** TSA binds inside the pocket making contacts to residues at the rim, walls and bottom of the pocket. **a**, Space-filling representation of TSA in the active-site pocket. The hydroxamic acid group, most of the aliphatic chain and part of the dimethylamino-phenyl group of TSA are buried (60% of TSA's surface area). The internal cavity has a volume of 144 $Å^3$. **b**, Closeup stereo view of the structure of the HDLP–$Zn^{2+}$–TSA complex in an orientation similar to Fig. 2a except for a 90° rotation about the vertical axis. TSA is in white; active-site residues and residues that contact TSA, which are all identical in HDAC1 except Tyr 91, are in yellow. **c**, Schematic representation of HDLP–TSA interactions. TSA is in black and the protein is in red. HDLP residues are labelled in red with their counterparts in HDAC1 indicated in black. Thatched semi-circles indicate van der Waals contacts between hydrophobic protein residues and TSA. Hydrogen bonds are shown as green dashed lines. **d**, Surface representation of the HDLP–TSA interface in a similar orientation to **b**. The protein surface is coloured according to residue identity with HDAC1. Cyan indicates residues that are identical in HDAC1.

In addition to the zinc ligands, the bottom of the pocket contains two histidines (His 131 and His 132), two aspartic acids (Asp 166 and Asp 173) and a tyrosine (Tyr 297), all of which are identical in HDAC1. Each of the histidines makes a hydrogen bond through its Nδ1 to an aspartic acid carboxylate oxygen, with the oxygen located in the plane of the imidazole ring (Fig. 3b, c). This His–Asp arrangement is characteristic of the charge-relay system found in the active sites of serine proteases, where it polarizes the imidazole Nϵ and increases its basicity[15]. The Asp 166–His 131 charge-relay pair is positioned deeper inside the pocket, and is more buried than the Asp 173–His 132 charge relay, which is partially solvent exposed. The buried charge relay makes a hydrogen bond (2.6 Å) to the zinc-bound water molecule, which could contribute to the deviation of the water–zinc coordination from ideal geometry. Tyr 297 is positioned next to the zinc, opposite to the two charge-relay systems. Its hydroxyl group lies 4.4 Å from the zinc atom and has no interactions with the rest of the protein (Fig. 3b, c). Next to Tyr 297 there is an opening in the pocket wall, which leads to the adjacent internal cavity.

The internal cavity is made up of portions of the L3 and L7 loops as they emerge from the β-strands and the α1–L1–α2 segment. The L1 loop, which demarcates the cavity from the solvent, has higher temperature factors than the rest of the structure, indicating some inherent flexibility that may allow the transient exchange of the cavity contents with the bulk solvent. The cavity is lined primarily with hydrophobic residues and is particularly rich in glycine residues (Fig. 1). There are only two charged residues in the cavity and these are contributed by the L1 loop (Arg 27 and His 21). The function of the cavity is not clear. It is possible that it provides space for the diffusion of the acetate product away from the catalytic centre, which may otherwise be crowded and shielded from the solvent when the substrate is bound.

The HDLP–HDAC1 homology maps mainly to the hydrophobic core and to the L1–L7 loops, with the portions of the loops that make up the pocket and adjacent cavity having the highest level of conservation (Fig. 1). Specifically, all of the polar residues in the active site and the hydrophobic residues that make up the walls of the pocket are identical. Among the residues that make up the internal cavity, the ones closest to the active site are either identical or conservatively substituted (Fig. 1). The HDLP active-site residues that are identical in HDAC1 are also conserved in the class II histone deacetylases. One exception is Asp 173 of the exposed charge-relay system, which is an asparagine in some class II HDACs (Fig. 1). The overall 35.2% HDLP–HDAC1 sequence identity predicts structural similarity with r.m.s. deviation in Cα positions of ~1.5 Å (ref. 16), and indicates that the 375-amino-acid HDLP protein corresponds to the histone deacetylase catalytic core that is conserved across the HDAC family. The one region likely to have a less similar structure is the 40-residue C terminus of HDLP (335–375), which has significantly lower homology to HDAC1.

In the crystal structure of the HDLP–$Zn^{2+}$–TSA complex, TSA binds by inserting its long aliphatic chain into the HDLP pocket, making multiple contacts to the tube-like hydrophobic portion of the pocket (Fig. 3a). The hydroxamic acid group at one end of the aliphatic chain reaches the polar bottom of the pocket, where it coordinates the zinc in a bidentate fashion and also contacts active-site residues. The aromatic dimethylamino-phenyl group at the other end of the TSA chain makes contacts at the pocket entrance and in an adjacent surface groove, capping the pocket. The length of the aliphatic chain appears to be optimal for spanning the length of the pocket and allowing contacts both at the bottom and at the entrance of the pocket.

The hydroxamic acid coordinates the zinc through its carbonyl (2.4 Å) and hydroxyl groups (2.2 Å), resulting in a penta-coordinated $Zn^{2+}$ (Fig. 3b, c). The hydroxamic acid also hydrogen-bonds with both charge-relay histidines and the Tyr 297 hydroxyl group (Fig. 3b, c) and replaces the zinc-bound water molecule of the active structure with its hydroxyl group. The binding of the TSA hydroxamic acid to HDLP shares several features with the binding of hydroxamic-acid-containing inhibitors to zinc metalloproteases, such as neutrophil collagenase[17], collagenase-3 (ref. 18) and thermolysin[19]. Like TSA, these inhibitors also (1) coordinate the active-site zinc in a bidentate fashion using their hydroxamate hydroxyl and carbonyl oxygens; (2) replace the nucleophilic water molecule with their hydroxamate hydroxyl groups; and (3) form hydrogen bonds with a general base in the active site.

The five-carbon-long branched aliphatic chain of TSA fits snugly in the narrow portion of the pocket, making multiple van der Waals contacts with all of the hydrophobic groups lining the pocket. Near its centre, the chain contains a methyl-substituted vinyl group that is sandwiched between the phenyl groups of Phe 141 and Phe 98, at the tightest point of the pocket (Fig. 3a–c).

The dimethylamino-phenyl and adjacent carbonyl groups of TSA form a planar structure and contact residues at the rim of the pocket and in an adjacent surface groove (Pro 22, Tyr 91, Phe 141; Fig. 3b, c). This packing is facilitated by the ~110° angle in the overall structure of TSA at the junction of the aliphatic chain and the dimethylamino-phenyl group (at C8). Upon TSA binding, the side chain of Tyr 91, which is at the periphery of the pocket and mostly solvent exposed, changes conformation to make space for the dimethylamino-phenyl group. This is the only change near the active site observed upon TSA binding. All but one (Tyr 91) of the HDLP residues that TSA contacts are identical in HDAC1 (Figs 1, 3b, c), indicating that TSA may bind and inhibit HDAC1 in a similar fashion.

SAHA, which has ~30-fold weaker inhibitory activity than TSA[6], binds HDLP similarly (Fig. 4). The SAHA hydroxamic acid group makes the same contacts to the zinc and active-site residues, and the importance of these interactions is underscored by the loss of activity of SAHA derivatives lacking the hydroxamic acid group[6]. The six-carbon-long aliphatic chain of SAHA packs in the tube-like hydrophobic portion of the pocket. Compared to TSA, however, SAHA's aliphatic chain packs less snugly and makes fewer van der Waals contacts, in part because SAHA lacks TSA's C15 methyl-group branch (Figs 3b, 4). SAHA also lacks TSA's double bonds in this region, and this may lead to increased flexibility of the aliphatic chain. The cap group of SAHA consists of a phenyl-amino ketone group. In the crystal structure, the phenyl group has weak electron density, suggesting that it may not pack as well as the cap group of TSA. This could be due, in part, to the larger separation between the hydroxamic and cap groups of SAHA than between those of TSA (Fig. 1).

The natural products trapoxin and HC-toxin, which represent a chemically distinct family of HDAC inhibitors[4,20], contain groups that may be analogous to the cap, aliphatic chain and active-site/zinc-binding groups of TSA. These inhibitors do not have a

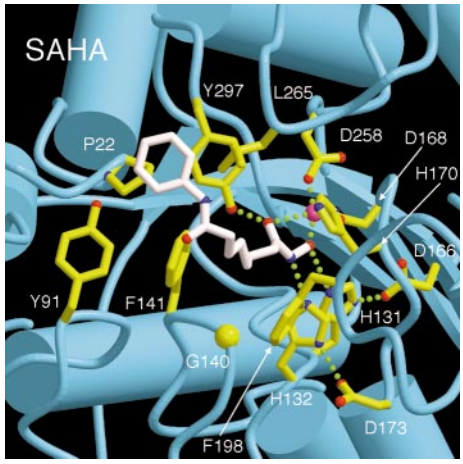


**Figure 4** SAHA binds HDLP like TSA, but its aliphatic chain and cap groups make fewer contacts. Close-up view of the HDLP–$Zn^{2+}$–SAHA complex in the same orientation and colour scheme as Fig. 3b. Note that Tyr 91 is not in the TSA-bound conformation.

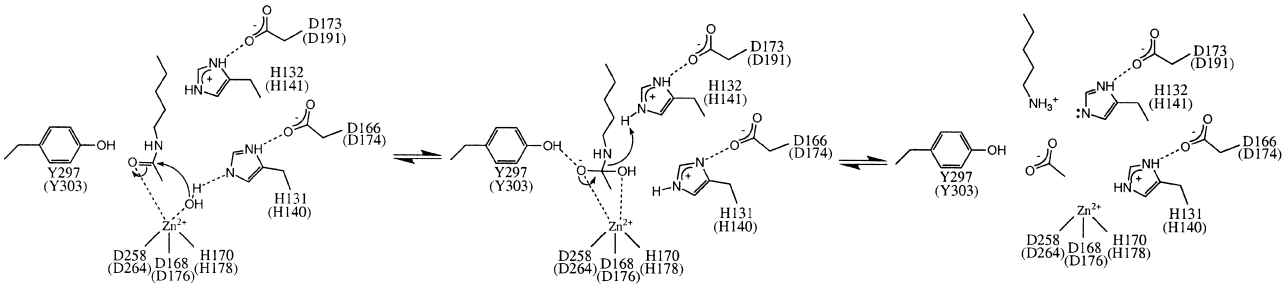


**Figure 5** The proposed catalytic mechanism for the deacetylation of acetylated lysine. HDLP active-site residues and their proposed HDAC1 counterparts (in parenthesis) are labelled.

hydroxamic acid, but their epoxy group may crosslink to an active site nucleophile[4]. Also, their ketone group may interact with polar residues, and possibly the zinc, at the bottom of the active-site pocket. This is supported by the observation that the reduction of the carbonyl to a hydroxyl, or its elimination, causes a large decrease in the activity of HC-toxin[20]. Trapoxin and HC-toxin both have a cyclic tetrapeptide with hydrophobic groups such as L-phenylalanine, D-proline, and L- or D-alanine that could, in principle, serve as a cap. The larger size of their cap group compared to that of TSA might allow them to make more extensive contacts at the rim of the pocket and in the shallow grooves surrounding the pocket entrance (Fig. 3d). The epoxy ketone and cyclic tetrapeptide groups of these inhibitors are linked by a five-carbon aliphatic chain, which probably binds in the hydrophobic tube-like portion of the pocket.

The structures of the HDLP–$Zn^{2+}$–TSA and HDLP–$Zn^{2+}$–SAHA complexes support the overall analogy of the aliphatic chain and hydroxamic acid groups of these inhibitors to the lysine side chain and the acetyl group of the HDAC substrate[10]. The branched C8 carbon, where the TSA structure bends as it enters the pocket, may approximate the Cα of the lysine amino acid. This positioning of the Cα would allow the aliphatic portion of the lysine side chain to span the tube-like portion of the pocket and place the acetylated amino group into the polar bottom of the pocket. The polypeptide backbone could interact with surface residues and in the shallow grooves around the pocket, including the groove where the TSA cap binds (Fig. 3d).

The deacetylase active site has features of both metallo- and serine proteases. We propose a mechanism that has aspects from both families of proteases and that is also consistent with the contacts TSA and SAHA make in the active site (Fig. 5). By analogy to the zinc proteases[21], the carbonyl oxygen of the N-acetyl amide bond could bind the zinc, and the carbonyl carbon could be positioned in close proximity to the water molecule. The zinc ion could polarize the carbonyl group so that the carbon is a better electrophile, and could also help orient the water molecule. The nucleophilicity of the water molecule would be increased by the negative charge of the buried Asp 166–His 131 charge-relay system to which the water is hydrogen bonded. This is analogous to the activation of a serine hydroxyl by a buried charge-relay system in the serine proteases[15], and to the activation of a water molecule by a glutamic acid in the zinc proteases[21]. The nucleophilic attack of the water on the carbonyl carbon would result in a tetrahedral carbon. This oxyanion intermediate could be stabilized by two zinc-oxygen interactions, in a manner analogous to the zinc proteases[21], and possibly by a hydrogen bond from the Tyr 297 hydroxyl group (Fig. 5). In the final step, the carbon–nitrogen bond of the intermediate would break, and the nitrogen of the scissile bond would accept a proton from the exposed Asp 173–His 132 charge relay, yielding the acetate and lysine products (Fig. 5).

The general aspects of this mechanism are supported by mutagenesis studies of *Drosophila* and human HDACs and of HDLP. Mutation of the histidine and aspartic-acid residues of the buried charge-relay system abolished activity[14,22], consistent with Asp 166–His 131 serving as a general base. Mutation of the histidine of the exposed charge relay reduced but did not abolish the activity, consistent with Asp 173–His 132 serving to protonate the epsilon nitrogen atom, and with some class II HDACs having an asparagine instead of aspartic acid in this charge relay. Mutation of Tyr 297 to phenylalanine eliminated the activity of HDLP in our *in vitro* assay (data not shown), supporting a role in catalysis.

The crystal structures of HDLP and the HDLP–$Zn^{2+}$–TSA and HDLP–$Zn^{2+}$–SAHA complexes provide a framework for understanding the catalytic activity and inhibition of the histone deacetylase family, and form a basis for the further development of HDAC inhibitors as antitumour agents. □

## Methods

### Protein expression and purification

The gene for HDLP (genbank accession no. AE000719) was subcloned from an *A. aeolicus* chromosomal DNA preparation (provided by R. Huber) into the pGEX4T3 vector (Pharmacia). The wild-type HDLP, Cys75Ser/Cys77Ser and active-site mutants were expressed in *Escherichia coli* as glutathione *S*-transferase (GST) fusion proteins, purified by glutathione-sepharose affinity chromatography, cleaved with thrombin and further purified by anion-exchange chromatography. HDLP was reconstituted with $Zn^{2+}$ (we presume the lack of metal in the purified HDLP is due, in part, to the use of DTT during purification) by mixing the Cys75Ser/Cys77Ser double mutant, at 10 mg $ml^{-1}$, with a 5-fold molar excess of $ZnCl_2$, followed by fractionation through a G25 desalting column. The use of the Cys75Ser/Cys77Ser mutant largely overcame the problem of protein aggregation during reconstitution, which we presume was due to the binding of free zinc to solvent-exposed cysteines. The HDLP–$Zn^{2+}$–TSA and HDLP–$Zn^{2+}$–SAHA complexes were prepared by mixing the $Zn^{2+}$-reconstituted HDLP Cys75Ser/Cys77Ser double mutant with 1 mM ligand, and were isolated by gel-filtration chromatography. Compared to the apo– or $Zn^{2+}$–HDLP, the concentrated HDLP–$Zn^{2+}$–TSA complex has a lower retention volume in gel-filtration chromatography, and this could correspond to the dimer observed in the crystals of the HDLP–$Zn^{2+}$–TSA complex. Proteins were concentrated to typically 25 mg $ml^{-1}$ in a buffer of 25 mM bis-tris propane, 500 mM NaCl, 1% isopropanol, pH 7.0 by ultrafiltration. FLAG-epitope-tagged human HDAC1 was overexpressed using a baculovirus expression system in Hi5 (Invitrogen) insect cells grown in suspension in serum-free medium (Sf900, Gibco). The fusion protein was purified by affinity chromatography using Anti-FLAG M2 affinity resin (Sigma) and FLAG Peptide (Sigma).

### Crystallization and data collection

We grew crystals of apo–HDLP at room temperature by the hanging-drop vapour-diffusion method, from 7.5% isopropanol, 28% polyethylene glycol (PEG) 1500, 425 mM NaCl, 100 mM Tris-Cl, pH 7.0; they contain one HDLP molecule in the asymmetric unit. Crystals of the HDLP–$Zn^{2+}$ complex were grown from 23% tert-butanol, 27% PEG 1500, 400 mM KCl, 100 mM bis-tris propane-Cl, pH 6.8. Crystals of the HDLP–$Zn^{2+}$–TSA and HDLP–$Zn^{2+}$–SAHA complexes were grown from 23% tert-butanol, 27% PEG 1500, 600 mM KCl, 100 mM bis-tris propane-Cl, pH 6.8, by microseeding. The HDLP–$Zn^{2+}$–TSA crystals contain two complexes in the asymmetric unit. All data were collected on flash-frozen crystals.

### MIR analysis, model building and refinement

Heavy-atom soaks were performed with the apo–HDLP crystals in a buffer of 7.5% isopropanol, 30% PEG 1500, 75 mM NaCl, 100 mM Tris-Cl, pH 8.0, supplemented with 1.0 mM thimerosal for 2 h, 5 mM $KAu(CN)_2$ for 1 h and 1 mM $Pb(Me)_3OAc$ for 2 h. MIR phases, calculated with the program MLPHARE[23] at 2.5 Å, included the anomalous diffraction signal from the thimerosal derivative and had a mean figure of merit of 0.55 (Table 1). The phases were improved by solvent flattening with the program DM[23], and were used to build the initial model with the program O[24]. The model was refined with the program CNS[25].

We determined the structures of the HDLP–$Zn^{2+}$–TSA and HDLP–$Zn^{2+}$–SAHA complexes by molecular replacement with the program AMORE[23] using the apo–HDLP

**Table 1 Statistics from the crystallographic analysis**

| Data set | Native | Hg | Pb | Au | Zn | TSA | SAHA |
|---|---|---|---|---|---|---|---|
| Space group | *C*2 | *C*2 | *C*2 | *C*2 | *C*2 | $P2_12_12_1$ | *C*2 |
| Resolution (Å) | 1.8 | 2.3 | 3.5 | 2.8 | 2.0 | 2.1 | 2.5 |
| Observations | 134,952 | 79,023 | 11,454 | 27,722 | 125,769 | 180,427 | 27,568 |
| Unique reflections | 32,143 | 15,958 | 4,040 | 8,753 | 23,643 | 50,796 | 11,748 |
| Data coverage (%) | 92.3 | 95.7 | 86.4 | 94.3 | 90.6 | 93.8 | 90.6 |
| $R_{sym}$ (%) | 2.9 | 8.4 | 9.6 | 8.9 | 7.2 | 7.1 | 8.6 |
| MIR analysis (20.0–2.5 Å): | | | | | | | |
| Phasing power | – | 1.47 | 1.24 | 1.10 | – | – | – |
| $R_{cullis}$ | – | 0.72 | 0.78 | 0.85 | – | – | – |
| $R_{cullis(ano)}$ | | 0.92 | | | | | |

Refinement statistics:

| Data set | Resolution (Å) | Reflections ($\|F\| > 1\sigma$) | Total atoms | Water atoms | *R*-factor (%) | $R_{free}$ (%) | Bonds (Å) | r.m.s.d. Angles (°) | r.m.s.d. *B*-factor (Å$^2$) |
|---|---|---|---|---|---|---|---|---|---|
| HDLP | 1.8 | 31,550 | 3,214 | 228 | 19.8 | 24.0 | 0.010 | 1.63 | 3.55 |
| HDLP–Zn | 2.0 | 23,582 | 3,424 | 434 | 22.0 | 25.8 | 0.009 | 1.48 | 1.04 |
| HDLP–Zn–TSA | 2.1 | 44,122 | 6,475 | 456 | 22.4 | 25.8 | 0.008 | 1.78 | 3.83 |
| HDLP–Zn–SAHA | 2.5 | 11,113 | 3,152 | 144 | 20.0 | 27.7 | 0.009 | 1.58 | 1.12 |

$R_{sym} = \Sigma_h \Sigma_i |I_{h,i} - \langle I_h \rangle| / \Sigma_h \Sigma_i I_{h,i}$ for the intensity ($I$) of $i$ observations of reflection $h$. Phasing power is $\langle F_h \rangle / E$, where $\langle F_h \rangle$ is the root-mean-square heavy atom structure factor and $E$ is the residual lack of closure error. $R_{cullis}$ is the mean residual lack of closure error divided by the dispersive difference. $R$-factor $= \Sigma||F_{obs} - F_{calc}|| / \Sigma|F_{obs}|$, where $F_{obs}$ and $F_{calc}$ are the observed and calculated structure factors, respectively. Figure of merit $= |F(hkl)_{best}| / |F(hkl)|$. $R_{free}$ is $R$-factor calculated using 5% of the reflection data chosen randomly and omitted from the start of refinement. r.m.s.d.: root mean square deviations from ideal geometry and root mean square variation in the $B$-factor of bonded atoms.

structure as a search model. The structure of TSA from the Cambridge Structural Database (Refcode TRCHST) was used to define stereochemical restraints used in the refinement with the program CNS. The restraints for SAHA were constructed based on stereochemical parameters from TSA and amino acids. The dimer interface in the HDLP–$Zn^{2+}$–TSA crystals primarily involves Phe 200 and Tyr 91 on the protein surface. These two residues are not conserved in HDAC1.

### Histone deacetylase assays.

We assayed purified proteins by incubating them with 10 μg of [$^3$H]acetyl-labelled murine erythroleukaemia cell-derived histone substrate for 30–60 min at 37 °C in a total volume of 30 μl as described[26]. The final concentrations of HDLP–$Zn^{2+}$ and FLAG-tagged HDAC1 were 3.6 μM and 0.24 μM, respectively, and assays were performed in duplicate. Inhibitors were added in the absence of substrate and incubated on ice for 20 min, substrate was added, and the assay performed as described above. Dialysis of FLAG-tagged HDAC1 against 20 μM $ZnCl_2$ had no significant effect on its activity, suggesting that FLAG-tagged HDAC1 retains a metal as purified.

Received 14 June; accepted 30 July 1999.


1. Davie, J. R. & Chadee, D. N. Regulation and regulatory parameters of histone modifications. *J. Cell Biochem.* (Suppl.) **30–31,** 203–213 (1998).
2. Kouzarides, T. Histone acetylases and deacetylases in cell proliferation. *Curr. Opin. Genet. Dev.* **9,** 40–84 (1999).
3. Fenrick, R. & Hiebert, S. W. Role of histone deacetylases in acute leukemia. *J. Cell. Biochem.* (Suppl.) **30–31,** 194–202 (1998).
4. Yoshida, M., Horinouchi, S. & Beppu, T. Trichostatin A and trapoxin: novel chemical probes for the role of histone acetylation in chromatin structure and function. *Bioessays* **17,** 423–430 (1995).
5. Richon, V. M. *et al.* Second generation hybrid polar compounds are potent inducers of transformed cell differentiation. *Proc. Natl Acad. Sci. USA* **93,** 5705–5708 (1996).
6. Richon, V. M. *et al.* A class of hybrid polar inducers of transformed cell differentiation inhibits histone deacetylases. *Proc. Natl Acad. Sci. USA* **95,** 3003–3007 (1998).
7. Cohen, L. A. *et al.* Inhibition of N-methylnitrosourea-induced tumours by the cytodifferentiating agent, suberanilohydroxamic acid (SAHA). *Proc. AACR* **39,** 108, abstr. 736 (1998).
8. Desai, D., El-Bayoumy, K. & Amin, S. Chemopreventive efficacy of suberanilohydroxamic acid (SAHA), a cytodifferentiating agent, against tobacco-specific nitrosamine 4-(methylinitros-amino)-1-(3-pyridyl)-1-butanone (NNK)-induced lung tumorigenesis in female A/J mice. *Proc. AACR* **40,** 2396, abstr. 362 (1999).
9. Struhl, K. Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev.* **12,** 599–606 (1998).
10. Taunton, J. Hassig, C. A. & Schreiber, S. L. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* **272,** 408–411 (1996).
11. Grozinger, C. M., Hassig, C. A. & Schreiber, S. L. Three proteins define a class of human histone deacetylases related to yeast Hda1p. *Proc. Natl Acad. Sci. USA* **96,** 4868–4873 (1999).
12. Fischle, W. *et al.* A new family of human histone deacetylases related to *Saccharomyces cerevisiae* HDA1p. *J. Biol. Chem.* **274,** 11713–11720 (1999).
13. Leipe, D. D. & Landsman, D. Histone deacetylases, acetoin utilization proteins and acetylpolyamine amidohydrolases are members of an ancient protein superfamily. *Nucleic Acids Res.* **25,** 3693–3697 (1997).
14. Hassig, C. A. *et al.* A role for histone deacetylase activity in HDAC1-mediated transcriptional repression. *Proc. Natl Acad. Sci. USA* **95,** 3519–3524 (1998).
15. Fersht, A. R. & Sperling, J. The charge relay system in chymotrypsin and chymotrypsinogen. *J. Mol. Biol.* **74,** 137–149 (1973).
16. Chothia, C. & Lesk, A. M. The relation between the divergence of sequence and structure in proteins. *EMBO J.* **5,** 823–826 (1986).
17. Grams, F. *et al.* Structure determination and analysis of human neutrophil collagenase complexed with a hydroxamate inhibitor. *Biochemistry* **34,** 14012–14020 (1995).
18. Lovejoy, B. *et al.* Crystal structures of MMP-1 and -13 reveal the structural basis for selectivity of collagenase inhibitors. *Nature Struct. Biol.* **6,** 217–221 (1999).
19. Holmes, M. A. & Matthews, B. W. Binding of hydroxamic acid inhibitors to crystalline thermolysin suggests a pentacoordinate zinc intermediate in catalysis. *Biochemistry* **20,** 6912–6920 (1981).
20. Shute, R. E., Dunlap, B. & Rich, D. H. Analogues of the cytostatic and antimitogenic agents chlamydocin and HC-toxin: synthesis and biological activity of chloromethyl ketone and diazomethyl ketone functionalized cyclic tetrapeptides. *J. Med. Chem.* **30,** 71–78 (1987).
21. Christianson, D. W. & Lipscomb, W. N. Carboxypeptidase A. *Acc. Chem. Res.* **22,** 62–69 (1989).
22. Kadosh, D. & Struhl, K. Histone deacetylase activity of Rpd3 is important for transcriptional repression *in vivo*. *Genes Dev.* **12,** 797–805 (1998).
23. The CCP4 suite: Programs for computational crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
24. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
25. Brunger, A. T. *et al.* Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallgr. D* **54,** 905–921 (1998).
26. Hendzel, M. J., Delcuve, G. P. & Davie, J. R. Histone deacetylase is a component of the internal nuclear matrix. *J. Biol. Chem.* **266,** 21936–21942 (1991).
27. Deckert, G. *et al.* The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* **392,** 353–358 (1998).
28. Kraulis, P. Molscript: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24,** 946–950 (1991).
29. Merritt, E. A. & Bacon, D. J. Raster3D-Photorealistic molecular graphics. *Methods Enzymol.* **277,** 505–524 (1997).
30. Kanyo, Z. F., Scholnick, L. R., Ash, D. E. & Christianson, D. W. Structure of a unique binuclear manganes cluster in arginase. *Nature* **383,** 554–557 (1996).



### Acknowledgements

We thank R. Huber for the *A. aeolicus* chromosomal DNA preparation; C. A. Hassig and S. L. Schreiber for the HDAC1-FLAG baculovirus and for helpful discussions; A. S. Mildvan and J. Thornson for advice on the reaction mechanism; and C. Murray for administrative help. Supported by the Howard Hughes Medical Institute, the NIH, the Dewitt Wallace Foundation and the Samuel and May Rudin Foundation.


Correspondence and requests for materials should be addressed to N.P.P. (e-mail: nikola@xray2.mskcc. org). Coordinates have been deposited with the Brookhaven Protein Data Bank (accession numbers 1C3P, 1C3R and 1C3S for HDLP, HDLP-Zn-TSA and HDLP-Zn-SAHA, respectively).