Hypothesis

The origin and utility of histone deacetylases

Saadi Khochbin[a,*], Alan P. Wolffe[b]

[a] *Laboratoire de Biologie Moléculaire du Cycle Cellulaire - INSERM U309, Institut Albert Bonniot, Faculté de Médecine, Domaine de la Merci, 38706 La Tronche Cedex, France*

[b] *Laboratory of Molecular Embryology, National Institute of Child Health and Human Development, NIH, Building 18T, Room 106, Bethesda, MD 20892-5430, USA*

Received 3 November 1997

Abstract A large region of two distinct yeast histone deacetylases, RPD3 and HDA1, is highly homologous to several prokaryotic enzymes that catalyze reactions involving various acetylated substrates. Proteins sharing this homology domain are found also in many higher eukaryotes and they all appear to be related to the RPD3 family of histone deacetylases. In each member of the family, the 'prokaryotic homology' domain covers almost two thirds of the protein, with the remaining portion containing the most divergent sequences. These sequences are located at the C-terminal region allowing for a clear definition of variants. Since the involvement of deacetylase members in different distinct regulatory complexes is now well established, the above observation suggests that the C-terminal domain may confer specificity to different members of the family. The RPD3 histone deacetylases thus appear as members of a family with a large conserved domain involved in enzymatic activity targeted to a short C-terminal domain, which probably confers functional specificity. The potential for deacetylases to be involved in multiple regulatory pathways provides an attractive counterpoint to the role of multiple histone acetyltransferases as coactivators.

Key words: Histone deacetylase; Functional specificity; RPD3 family

1. Introduction

The evolutionary appearance of chromatin must have paralleled that of the chromatin structure/function-modulating machinery [1]. It is now clear that the reversible acetylation of core histone tails is involved in the modulation of chromatin structure and function [2,3]. Indeed, lysines present in the relatively short unstructured tail of core histones are the sites of reversible acetylation which involves two types of enzymes, histone acetyltransferases and histone deacetylases. Interestingly, within the members of histone acetyltransferases already identified, several are proteins that are known to be involved in the regulation of transcription: GCN5, P/CAF, P300/CBP, $TAF_{II}250$ [4]. However, the acetyltransferase activity of these factors may not exclusively concern histones since P300 and PCAF are also able to acetylate non-histone proteins such as P53, TFIIF and TFIIE [5,6]. The first deacetylase cloned was of human origin and showed striking homology with a yeast general regulator of transcription, RPD3 [7]. RPD3 homologues were identified in many eukaryotes and showed considerable sequence conservation during evolution [8]. The study of the function of RPD3 members in yeast, *Drosophila* and in mammalian cells suggests an active involvement in transcriptional repression and activation [9,10]. Thus, the involvement of acetylation of core histones in the chromatin-mediated control of transcription is now undoubtedly established. A second important function of acetylation occurs during replication coupled chromatin assembly, when the synthesis, nuclear import and incorporation of histones are also tightly linked to the core histone acetylation/deacetylation machinery [4]. Both specific and general histone acetyltransferases/deacetylases are therefore in action throughout the cell cycle adapting the indispensable dynamic features of chromatin to the requirement of DNA packaging. The evolutionary use of acetylation to achieve control of chromatin dynamics must have been selected concomitantly with the appearance of chromatin in eukaryotes.

Histone deacetylases probably evolved from prokaryotic enzymes which catalyze chemical reactions involving various acetylated substrates. Indeed, extensive sequence homology was found between all members of RPD3, yeast histone deacetylase HDA1, and two prokaryotic proteins, acuC and acetylpolyamine amidohydrolase [8]. Our sequence analysis showed that almost all deacetylases found in higher eukaryotes are related to the RPD3 family. Within these RPD3 members, the domain of sequence homology with the prokaryotic protein covers two thirds of the protein from the N-terminal end. Interestingly, the remaining one third sequence is highly divergent. Detailed analysis of this sequence clearly defines variants and moreover provides clues to the way a histone deacetylase activity may become specific.

2. The highly conserved acuC/APH homology domain is present in distinct families of histone deacetylases

In yeast, at least two distinct histone deacetylases have been identified, RPD3 and HDA1, both of which are present in different multiprotein complexes [11]. These proteins show some sequence homology with each other and are also homologous to two prokaryotic proteins, acetylpolyamine amidohydrolase (APH) and acuC [8]. Interestingly, the only homologous sequence between RPD3 and HDA1 lies within the region that is also homologous to the prokaryotic enzymes. We call this region of histone deacetylases the acuC/APH homology domain. These observations strongly suggest that the deacetylase activity is associated with this portion of the protein. This domain covers two thirds of the sequence from the NH2 terminal (Fig. 1) and contains stretches of absolutely conserved amino acids [8]. In yeast, other open reading frames (ORF) containing the acuC/APH homology domain have

*Corresponding author. Fax: (33) 4 76 54 95 95.
E-mail: khochbin@ujf-grenoble.fr

PII S0014-5793(97)01423-3

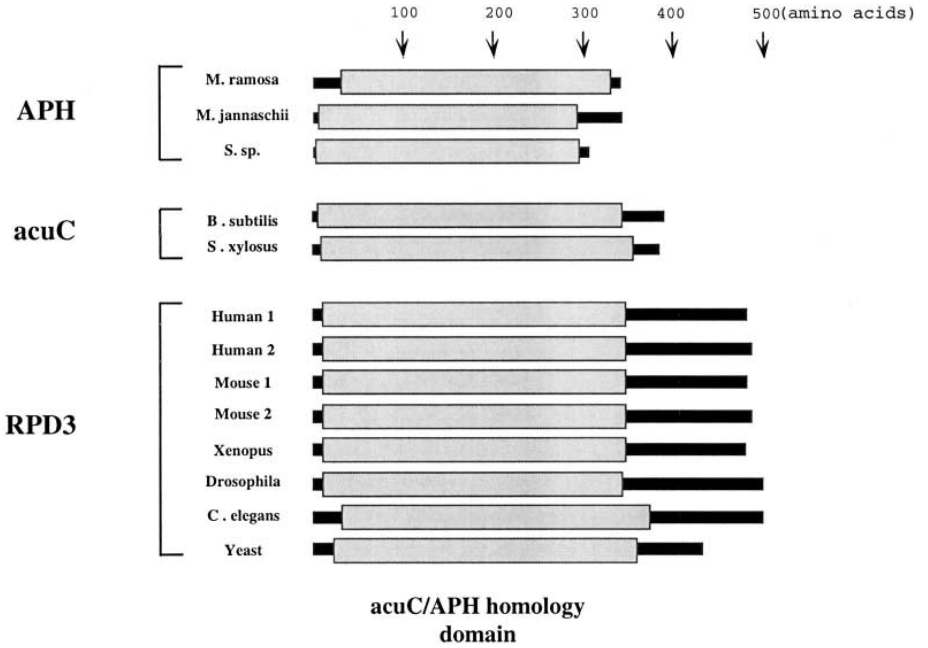

Fig. 1. Schematic representation of members of RPD3, acuC and acetylpolyamine amidohyrolases (APH). The sequence of the acuC/APH homology domain from yeast RPD3 was used to perform a search against a non-redundant protein sequence database (SwissProt current release; Non-redundant SwissProt+GenPept+GenPept updates; gpu cumulative updates of GenPept) with the yeast RPD3 (*S. cerevisiae*, GenBank accession number (gb): S66438) using the BLASTP program [19]. Identified homologous sequences were aligned pairwise using the facility made available by the GeneStream align home page (http://www.tigem.it/TIGEM/HTML/fasta.html). The multiple alignment was performed manually based on the pairwise alignment data. The region of homology found in all proteins called the acuC/APH homology domain is represented as a gray box. Black boxes represent the non-conserved regions. Numbers refer to the positions of the amino acids. Sequences used for this analysis are as follows: RPD3 members; Human 1, gb: U50079; Human 2, gb: U31814; Mouse 1, gb: X98207; Mouse 2, gb: U31758; *Xenopus*, gb: X78454; *Drosophila*, gb: Y09258; *C. elegans*, gb: Z46676; Yeast (*S. cerevisiae*), gb: S66438; acuC members; *B. subtilis*, gb: L17309; *S. xylosus*, gb: X95439; APH members; *Mycoplana ramosa*, D10463; *Methanococcus jannaschii*, U67502; *Synechocystis* sp., D90900.

been also found (HOS1, HOS2 and HOS3), but there is no direct evidence for histone deacetylase activity associated with these proteins [11]. Databank searching with the acuC/APH homology domain of the yeast RPD3 reveals the existence of ORFs in almost all higher eukaryotes. In invertebrates (*Caenorhabditis elegans* and *Drosophila*) and in vertebrates (amphibians, birds and mammals), ORFs with high sequence homology to the acuC/APH domain have been found [8]. These proteins appear to be members of the RPD3 family. Indeed, their acuC/APH domain shows better homology to that of the yeast RPD3 than to that of HDA1 (63% identity between yeast RPD3 and human RPD3.1 over 397 residues versus 29% identity between yeast HDA1 and the same human protein over 283 residues, not shown). No known higher eukaryotic ORFs with convincing homology to yeast HDA1 have been found. However, among mouse and human ESTs, ORFs with significant homology to yeast HDA1 exist (Verdel and Khochbin, unpublished).

3. Definition of species-specific domain in the RPD3 family of histone deacetylases

The definition of the acuC/APH homology domain within RPD3 members allowed for its positioning along the sequence of all known RPD3 proteins and subsequent analysis of associated sequences. The schematic presentation of RPD3 shows that the C-terminal part of the protein flanking the homology domain is variable in length (Fig. 1) while sequence alignment showed little evolutionary conservation (not shown). However, as two different RPD3 sequences from mouse and human have been published [7,12], we investigated sequence homology and divergence between these two species. This analysis clearly revealed the existence of two independent variants, named here RPD3.1 and RPD3.2. The sequence of each variant is highly conserved and is almost identical between human and mouse (Fig. 2A). The sequence divergence between variants 1 and 2 is observed essentially in the 50 most C-terminal amino acids (Fig. 2A). We also analyzed the C-terminal portion of the *Xenopus* RPD3 to investigate putative homology to one of the two mammalian species. Although the 54 C-terminal residues of the *Xenopus* RPD3 share sequence features of both mammalian RPD3 variants, it appears to correlate more closely with the class 1 variant than the class 2 (Fig. 2B). Therefore, the existence of other RPD3 members in *Xenopus* is likely. In *C. elegans*, two ORFs containing an acuC/APH homology domain have been found. One containing 507 residues is clearly a RPD3 member exhibiting high sequence homology to other members of the RPD3 family (not shown), but the other, containing 798 residues, presents unique features. Indeed, this protein would have two adjacent acuC/APH domains, each showing about 30% amino acid identity to the acuC/APH domain of the *C. elegans* RPD3 (not shown).

4. Concluding remarks

The appearance of chromatin during evolution was certainly associated with considerable pressure for the selection of enzymatic activities capable of controlling the dynamics of this structure. Histone-like proteins have been discovered in

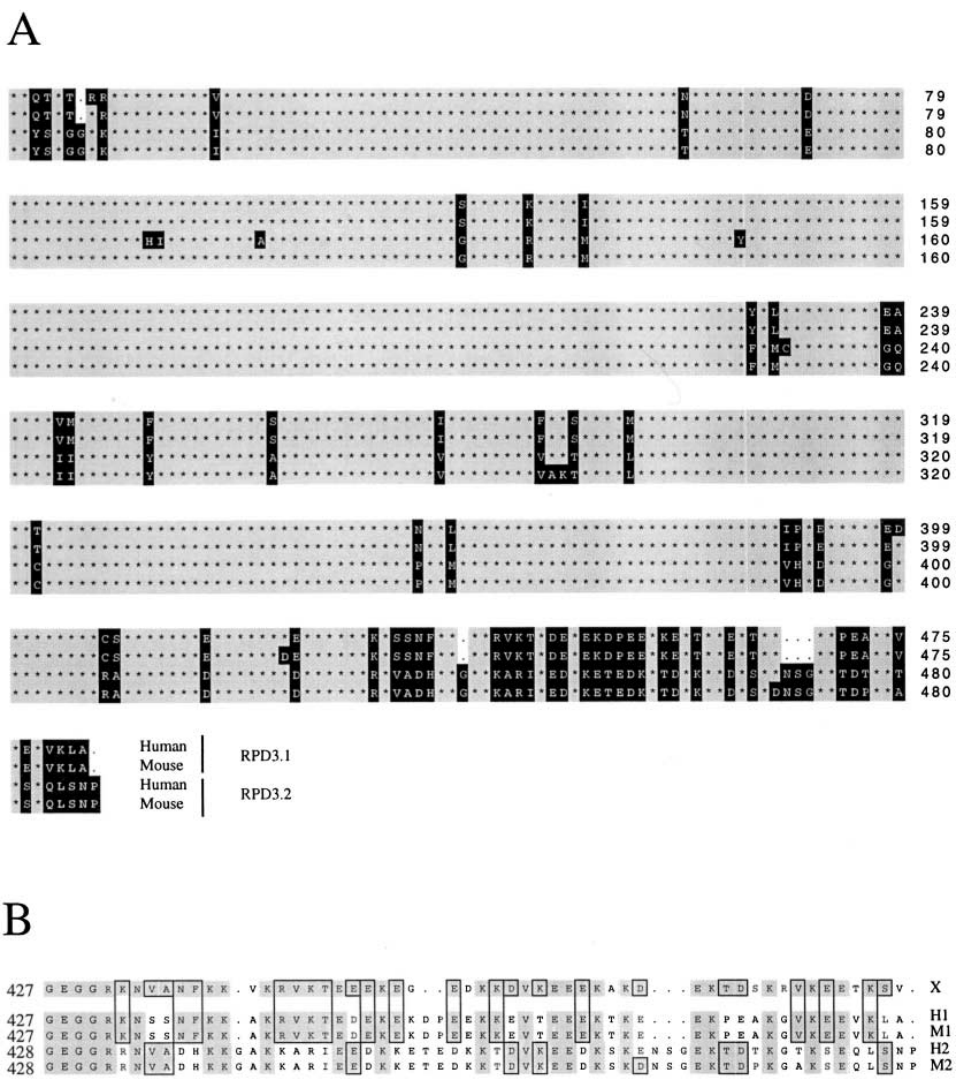

Fig. 2. Definition of RPD3 variants using C-terminal sequence divergence. A: In mammals, two distinct RPD3 variants have been observed. The sequences of the two variants from human and mouse were aligned. Stars in gray boxes represent identical amino acids, not shown. Sequences in black boxes show divergent amino acids. Dots represent gaps introduced for alignment purposes. B: The sequence of the C-terminal region of the *Xenopus* RPD3 (X) was aligned with that of the human and mouse variant 1 (H1 and M1 respectively) as well as with human and mouse variant 2 (H2 and M2 respectively). The identical amino acids present in the *Xenopus* sequence and one or both mammalian variants are boxed (gray). Open boxes show the identical amino acids present in *Xenopus* and only one of the mammalian variants. This alignment shows 68.5% identity in amino acid content between the *Xenopus* and the human 1 variant while the percentage of identical amino acids between *Xenopus* and the human 2 variant is 59.3.

bacteria (i.e. Hmf and Hmt, [13]), and there is increasing evidence for the presence of chromatin-like structure in these organisms [14]. Although Hmf and Hmt possess the histone fold structure and are believed to be true histones, they have very short or no labile termini [15]. These proteins would therefore lack potential sites for post-translational modifications. As the architectural foundation of chromatin was established during evolution, eukaryotic histones may have acquired additional regulatory functions [16]. Database searching with the sequence of the acuC/APH homology domain does not reveal any other prokaryotic protein with significant sequence homology. Bacteria therefore probably lack histone deacetylase-like enzymatic activity and we propose that the appearance of chromatin led to the evolutionary pressure to transform the pre-existing enzymes involved in the metabolism of acetylated substrates to histone deacetylases.

Within all eukaryotes where histone deacetylases were identified, we can distinguish two families, the RPD3 family, found from yeast to human, and HDA1, found only in yeast. HDA1 and RPD3 show significant sequence homology only within the acuC/AHP homology domain. This observation suggests that the acuC/APH domain is the only motif that could be involved in histone deacetylation. Recently, however, a plant histone deacetylase (HD2) has been cloned which has a nucleolar localization and does not show any significant sequence homology to the mentioned histone deacetylases [17]. Therefore the acuC/APH homology domain is not the only structure capable of deacetylating histones. The important question is whether this enzyme is also present in other eukaryotes and if it also corresponds to an evolutionarily conserved family of enzymes.

The finding of an evolutionarily conserved region present in almost all known histone deacetylases is indicative of a functional role for this region in the enzymatic activity. Moreover,

the fact that this same region is found in prokaryotic enzymes using various acetylated substrates strongly suggests the involvement of this region in catalytic activity. The acuC/APH homology domain covers a relatively large portion of the protein suggesting the involvement of a large number of amino acids (more that 200) in the maintenance of the correct structure necessary for substrate interaction and catalytic activity. Interestingly, in the case of APH from *Mycoplana ramosa*, it has been shown that the enzyme is composed of two identical subunits each containing a zinc atom [18]. Consequently, the participation of metals such as zinc in the maintenance of an active structure [8] and a need for the multimerization of the protein may account for the evolutionary conservation of such a large number of amino acids.

After the demonstration of the existence of a large, highly conserved portion of histone deacetylases potentially involved in catalytic activity, an important question concerning the specificity of histone deacetylation arises. Indeed, within RPD3 members, the acuC/APH homology domain shows remarkable sequence identity from yeast to human. Moreover, recent investigations have established the presence of RPD3 in distinct regulatory complexes [4,9,10]. It will therefore be of interest to know how a complex can recruit a specific deacetylase. In solving this problem, one may consider the sequence of the C-terminal part of RPD3s. This portion of the protein shows important sequence divergence and moreover, based only on the sequence of this region, variants could be defined in mouse and human. Considering the high sequence identity between the two variants within the rest of the protein, the only divergent region, the C-terminal, is therefore a good candidate for conferring variant specificity. This potential for targeting to particular corepressor complexes would provide a sophisticated regulatory response for transcriptional repression. Such a response is likely to be necessary in light of the diverse and combinatorial activity of targeted histone acetyltransferases [3,20].

References

[1] Kingston, R.E., Bunker, C.A. and Imbalzano, A.N. (1996) Gene Dev. 10, 905–920.
[2] Roth, S.Y. and Allis, C.D. (1996) Cell 87, 5–8.
[3] Wade, P.A. and Wolffe, A.P. (1997) Curr. Biol. 7, R82–R84.
[4] Wade, P.A., Pruss, D. and Wolffe, A.P. (1997) Trends Biochem. Sci. 22, 128–132.
[5] Gu, W. and Roeder, R.G. (1997) Cell 90, 595–606.
[6] Imhof, A., Yang, X.J., Ogryzko, V.V., Nakatani, Y., Wolffe, A.P. and Ge, H. (1997) Curr. Biol. 7, 689–692.
[7] Taunton, J., Hassing, C.A. and Schreiber, S.L. (1996) Science 272, 408–411.
[8] Leipe, D. and Landsman, D. (1997) Nucleic Acids Res. 18, 3693–3697.
[9] Wolffe, A.P. (1997) Nature 387, 16–17.
[10] Pazin, M.J. and Kadonaga, J. (1997) Cell 89, 325–328.
[11] Rundlett, S.E., Carmen, A.A., Kobayashi, R., Bavkin, S., Turner, B. and Grunstein, M. (1996) Proc. Natl. Acad. Sci. USA 93, 14503–14508.
[12] Yang, W.M., Inouye, C., Zeng, Y., Bearss, D. and Seto, E. (1996) Proc. Natl. Acad. Sci. USA 93, 12845–12850.
[13] Tabassum, R., Sandman, K.M. and Reeves, J.N. (1992) J. Bacteriol. 174, 7890–7895.
[14] Reeves, J.N., Sandman, K. and Daniels, C.J. (1997) Cell 89, 999–1002.
[15] Arents, G. and Moudrianakis, E.N. (1995) Proc. Natl. Acad. Sci. USA 92, 11170–11174.
[16] Wollfe, A.P. and Pruss, D. (1996) Trends Genet. 12, 58–62.
[17] Lussar, A., Borsch, G., Loidl, A., Haas, H. and Loidl, P. (1997) Science 277, 88–91.
[18] Sakurada, K., Ohta, T., Fujishiro, K., Hasegawa, M. and Aisaka, K. (1996) J. Bacteriol. 178, 5781–5786.
[19] Altschul, S.F., Boguski, M.S., Gish, W. and Wootton, J.C. (1994) Nature Genet. 6, 119–129.
[20] Brownell, J.E. and Allis, C.D. (1996) Curr. Opin. Genet. Dev. 6, 176–184.