A plant dialect of the histone language

Peter Loidl

Department of Molecular Biology, Innsbruck Medical University, Peter-Mayr-Strasse 4b, A-6020 Innsbruck, Austria

The genome contains all the information needed to build an organism. However, during differentiation and development, additional epigenetic information determines the functional state of cells and tissues. This epigenetic information can be introduced by cytosine methylation and by marking nucleosomal histones. The code written on histones consists of post-translational modifications, including acetylation and methylation. In contrast to the universal nature of the DNA code, the histone language and its decoding machinery differ among animals, plants and fungi. Plant cells have retained totipotency to generate the entire plant and maintained the ability to dedifferentiate, which suggests that the establishment and maintenance of epigenetic information differs from animals. Here, I aim to summarize the histone code and plant-specific aspects of setting and translating the code.

Nuclear DNA is compacted into chromatin, which is a complex structure built from repeating units, the nucleosomes [1]. These consist of ~145 bp of DNA wrapped around an octamer of basic proteins, the core histones. The octamer is formed by two molecules each of histones H2A, H2B, H3 and H4. At least two different domains can be distinguished in core histones: a globular domain involved in histone–histone interactions (containing the ‘histone fold’ motif); and the flexible N-terminal tails of H3 and H4, and N- and C-terminal tails of H2A and H2B. A series of consecutive nucleosomes produces a ‘beads on a string’ structure or 10 nm fiber. A further level of compaction is the 30 nm fiber with six nucleosomes per turn in a solenoid arrangement.

Our traditional picture of eukaryotic chromatin as a static and largely repressive functional state has, over recent years, changed to a more complex view of chromatin as a highly dynamic state that is essential for regulating cellular functions. The dynamic properties of chromatin are mediated by multiprotein complexes with different functions that set marks overlying the stable information of the DNA. The most prominent factors that influence chromatin structure and function are enzymes that modify the histones and chromatin remodeling machines that utilize ATP. Plant chromatin remodeling complexes have recently been reviewed in detail [2–4].

Histones have been conserved during evolution. However, they are dynamically changed by post-translational modifications. These modifications include acetylation, methylation, phosphorylation, ubiquitination, glycosylation, ADP ribosylation, carboxylation, sumoylation and biotinylation, which can all cause structural and functional rearrangements in chromatin and are therefore essential elements of the complex ‘epigenetic histone code’ [5,6]. To decipher this code, which is recognized and interpreted by transcriptional regulators and chromatin remodeling machines, is one of the central challenges of chromatin research. In this article, I focus on the methylation and acetylation of core histones (Figure 1) because little is known about the other types of modification.

DNA methylation and histone modifications

There is ample evidence for an interrelation between DNA methylation and histone modifications [7]. Cytosine methylation is important for regulating gene activity and genome stability in eukaryotes and is correlated with gene silencing. Silencing governs processes such as X chromosome inactivation and imprinting in vertebrates, as well as suppression of transposon movement in higher plants [8]. DNA methylation also silences protein-coding genes in plant polyploids [9] and RNA polymerase I genes [10]. In contrast to the vertebrate genome, in which methylation is mostly restricted to CG islands, plants and fungi carry methylation on CG dinucleotides and also on non-CG cytosines. In Arabidopsis, the cytosine methyltransferase CMT3 is involved in CNG methylation [11]. CMT3 is a member of the plant-specific DNA methyltransferases with a chromodomain [8,12]. A related enzyme from maize, also carrying a chromodomain, is involved in the maintenance of CNG methylation [13]. The link between DNA methylation and histone modifications was initially supported by two basic observations. Genetic experiments in Arabidopsis showed that the DDM1 gene, encoding a protein related to SWI2/SNF2 chromatin remodeling factors, is implicated in DNA methylation [14]. A genetic search in Neurospora then revealed the dim-5 gene, encoding a SET domain protein, to be necessary for DNA methylation [15]. SET domain [Su(var)3-9, Enhancer-of-zeste, Trithorax] proteins form a protein family that can influence gene expression through histone methylation [16]. The Neurospora protein DIM-5 has histone methyltransferase (HMT) activity and catalyzes the methylation of lysine-9 in H3; loss of DIM-5 function blocks DNA methylation and histone H3 methylation [17]. A related protein has recently been identified in Arabidopsis that has HMT activity at lysine-9 in H3 and a function for DNA methylation [18,19].

HMTs set specific marks on nucleosomes

HMTs catalyze the transfer of methyl groups to either arginine or lysine residues, preferentially within the N-terminal extensions of histones H3 and H4 (Figure 1). In
Histone methylation is correlated with transcriptional repression as well as transcriptional activation, even when it occurs at the same site; H3 lysine-4 methylation in yeast is important for rDNA silencing [24] but is also involved in activating gene transcription [25]. This might depend on the particular gene, the type of the functionally redundant HMTs involved and the level of methylation (di- or trimethylated). Only trimethylated lysine-9 in H3 marks chromatin regions for DNA methylation in Neurospora; the dimethylated form does not [17]. It is interesting in this context that trimethylation at lysine-20 of H4 is correlated with the aging of mammalian cells [26]. An Arabidopsis homolog of Trithorax, ATX-1, has recently been shown to function as an activator of homeotic genes in plants; it specifically methylates H3 lysine-4 and is therefore involved in chromatin remodeling [27].

In Arabidopsis, euchromatin is characterized by high levels of dimethylated lysine-4 in H3, whereas high levels of dimethylated lysine-9 in H3 were observed for heterochromatic chromocenters [28]. It has also been reported for the Arabidopsis gene DDM1 that DNA methylation inheritance is dependent on H3 lysine-9 methylation [29]. However, this might only be true for parts of the plant genome, because a comparative study over the whole genome of different plants revealed that, in small genome plants, high levels of H3 dimethylated lysine-9 were confined to constitutive heterochromatin, whereas, in large genome plants, dimethylated lysine-9 was uniformly distributed. This suggests that large genome plants have to silence dispersed repetitive sequences also within euchromatic regions with a high level of methylated H3 lysine-4 [30]. A recent report shows that the methylation level of H3 lysine-9 is drastically reduced after removal of CpG methylation in an Arabidopsis mutant lacking the DNA maintenance methyltransferase MET1 [31]. However, in this particular case, the loss of both CpG methylation and lysine-9 methylation in H3 at heterochromatic chromocenters had no effect on the structure of this tightly packed chromatin. An earlier study [32] showed that chromocenters in a hypomorphic allele of mel1 were reduced in number and size compared with the wild type. These data emphasize the complex and manifold mechanisms leading to a particular chromatin structure.

The relationships and hierarchy between DNA methylation, H3 methylation, H4 acetylation and heterochromatin assembly have been investigated in Arabidopsis. As a result, a tentative model was proposed in which four key players act in a coordinated manner. Following DNA replication, maintenance DNA methyltransferase acts on chromatin with acetylated H4 lysine-16. DNA methylation precedes and governs H3 lysine-9 methylation. The chromatin remodeling factor DDM1 could finally trigger the deacetylation of H4 lysine-16 [32]. However, this model might not apply to all silenced loci in the Arabidopsis
Acetylation signal on histones is reversible

Of the different types of histone modification, acetylation is the most extensively characterized. e-Amino groups of lysines in the N-terminal extensions of core histones can be post-synthetically acetylated by histone acetyltransferases (HATs) using acetyl-CoA as a cosubstrate; the modification can be reverted by HDACs, a reaction for which a large panel of specific inhibitors is available [40]. HATs and HDACs exist as multiple proteins that belong to distinct protein families [40,41]; an excellent detailed sequence and phylogenetic analysis of HATs and HDACs has recently been published [42].

Although many research reports have been published on histone acetylation in yeast and vertebrates, little is still known about the specific functions of histone acetylation in plants. Studies in animal cells have more and more revealed that HATs and HDACs not only modify histones but also many non-histone regulatory proteins as well as structural proteins: proto-oncogene products like c-Myc, tumor suppressors like p53 and the retinoblastoma tumor suppressor protein (pRb) and tubulin are examples of a rapidly growing list of proteins [43–46]. Therefore, it more and more becomes unclear whether HATs and HDACs are predominantly histone modifying enzymes \textit{in vivo}; they might rather be protein acetyltransferases and deacetylases, and would probably be better named accordingly.

HATs and HDACs basically fulfill similar functions in all eukaryotes. The various HATs are well conserved in their catalytic domains and recruit proteins with homologous functions in different species [42]. By sequence analysis, four distinct families of HATs can be distinguished: (i) the GNAT-MYST family; (ii) the p300/CBP coactivator family; (iii) the TAF\(_1\)250-related family; and (iv) the nuclear receptor coactivator family, which is present in vertebrates but not in plants or fungi [42].

There are few reports on plant HATs. In \textit{Arabidopsis}, the HAT GCN5 is associated with the transcriptional adaptor protein ADA2; mutations disrupting ADA2 or GCN5 had pronounced effects on growth, development and gene expression patterns [47]. In maize, experiments using both a transgenic line expressing the GCN5 antisense transcript or a cell line treated with trichostatin A (a potent HDAC inhibitor) revealed a connection between the acetylation level of histones and the equilibrium between histone synthesis and degradation [48]. Histone acetylation also has an essential role in cell differentiation. Increased acetylation of H3 and H4 at the promoter of a light-regulated plastocyanin gene of pea was correlated with transcriptional activity in response to light [49]. The transcriptional enhancer of the pea plastocyanin gene activates transcription by associating with the nuclear matrix, thereby triggering the acetylation of histones on the promoter and altering chromatin structure [50].

Three families of HDAC have been identified in plants [42]: (i) members of the RPD3/HDA gene family; (ii) members of the sirtuin family related to yeast SIR2; and (iii) members of the HD2 class of enzymes. In contrast to other eukaryotes, plants contain the HD2-type deacetylases, a plant specific class that is unrelated to the other HDAC classes [51,52].

In general, HDACs were often correlated with transcriptional repression and gene silencing. HDACs cooperate with HMTs and DNA methyltransferases [53]. Antisense inhibition of an Rpd3-type HDAC led to various growth and developmental defects [54–56]. Overexpression of rice HDAC1 resulted in increased growth rate and dramatic changes in phenotype [57]. Mutations in the gene encoding the Rpd3-type HDAC HDA6 of \textit{Arabidopsis} revealed it to be associated with transgene silencing [58]. Similarly, antisense inhibition of an HD2-type HDAC in \textit{Arabidopsis} resulted in aborted seed development [59]. Rpd3-, HD2- and HDA1-type deacetylases were able to repress transcription efficiently in reporter gene assays [59–61]. In \textit{Brassica}, a low temperature responsive transcription factor interacts specifically with an Rpd3-type HDAC [62]. Together, these data suggest a global role for HDACs in gene regulation of plants.

Plant-specific peculiarities of histone modifications

The patterns of growth and development differ dramatically between plants and animals. When animals have developed into adult organisms, growth and morphogenesis cease and cell division mainly replaces dead cells or specialized cells that undergo continuous turnover. In plants, morphogenesis and growth continue throughout the lifetime of the organism. In the absence of migration of cells, plant morphogenesis is determined by cell division and expansion. Because cell division occurs preferentially in meristematic regions, the identity of a cell that leaves the meristematic region is determined mostly by the position of the cell relative to the position of its neighbors. In animals, the identity of cells is mostly determined by cell lineage. Totipotency of plant cells can refer to single plant cells that can be isolated and give rise to an entire
plant under appropriate culture conditions, or to developmental stem cells in the shoot apical meristem.

With respect to DNA methylation, plants differ from animals in that DNA methylation patterns can persist throughout development and can be inherited between generations, whereas, in vertebrates, DNA methylation is reprogrammed during early embryogenesis and altered methylation patterns are not usually transmitted to the progeny. In addition, plant cells are much more prone to environmental stress due to their immobility. Altogether, these basic differences indicate that plants might have developed mechanisms of gene regulation that are distinct from those of animals. Indeed, plants differ in the histone code they use and in the enzymes involved. The differences discovered so far concern the sites of modification, a whole plant-specific HDAC family as well as distinct ways of regulating histone modifying enzyme activity.

Additional lysine residues are acetylated or methylated in plants

In plants, more lysine residues in the N-terminal histone regions are subject to modification then in other eukaryotes. Plant H4 can be occasionally acetylated at lysine-20 (Figure 1), eventually leading to penta-acetylated H4 [63]. In animals and fungi, this site is only subject to methylation and it could therefore act as a functional switch in plant chromatin. A similar situation is present at H3 lysine-9 in all species investigated so far; lysine-9 can be either acetylated or methylated and it has been shown that methylation of lysine-9 influences the phosphorylation of serine 10, which in turn influences the acetylation of lysine-14 [64]. This interdependence of different modifications might represent the key element of a complex histone code [65,66]. Plants might have even more combinatorial possibilities, because another three lysines of H3 (lysines 14, 18 and 23) can be methylated, apart from being acetylated (http://research.nhgri.nih.gov/histones/posttrans.shtml). Taking into account that the nucleosome consists of two molecules of each of the core histones, the modification pattern of two H4 molecules, for example, can be identical or different, giving rise to a great number of modification states.

Plants are distinguished by a unique HDAC family

Unlike animals and fungi, plants possess the HD2-type HDAC family [42,51]. This HDAC family has no sequence relations to other known HDAC classes. HD2-type HDACs are distantly related to cis–trans isomerases with respect to DNA sequence but are functionally unrelated [42]. The enzymatic activity of maize HD2 depends on phosphorylation [51,67]. HD2 might have a function in plant-specific pathways or have taken over functions corresponding to those carried out by other HDACs in non-plant organisms. Possible candidates include members of the SIR2 family, because this HDAC family is under-represented in plants compared with animals and fungi. A detailed phylogenetic analysis indicated that a gene duplication might have occurred during dicot evolution, suggesting functional diversification of the HD2 family [42]. It is interesting in this context that the Arabidopsis genome is predicted to encode a total of 16 HDAC proteins (http://www.chromdb.org/), which is more than the number of corresponding genes in other eukaryotic organisms outside the plant kingdom. Most of these HDACs have not been characterized so far and it is not clear whether all of them display histone specificity in vivo. It might also turn out that the different subtypes have distinct functions for particular genes or chromatin areas or only during certain developmental transitions in individual tissues.

Histone deacetylase ZmHDA1 is post-translationally regulated by limited proteolysis

It has been recently reported that maize regulates the activity of an Hda1 homolog by limited proteolysis [60]. ZmHDA1 is synthesized as an enzymatically inactive protein with an apparent molecular weight of 84 kDa that is converted to an enzymatically active 48 kDa HDAC by proteolytic removal of the C-terminal part, presumably by the aid of a 65 kDa intermediate protein. Interestingly, the 84 kDa precursor HDAC is part of a 300 kDa complex of unknown function (Figure 2). Only the processed 48 kDa deacetylase could repress transcription efficiently in a reporter gene assay. Arabidopsis contains not only an HDA1-homologous gene (AtHda5; NP_200914) but also a
distinct gene encoding a small protein of 252 amino acids (Atg61050; NP_200913); this short protein is highly homologous to the C-terminal proteolytic cleavage product of ZmHDA1 and also contains a 17 amino acid stretch homologous to the N-terminus of ZmHDA1. The 84 kDa precursor, a 65 kDa intermediate form and the enzymatically active 48 kDa protein exist as phosphorylated forms; dephosphorylation of p48 stimulates the enzyme activity and changes the substrate specificity [67]. Mammalian HDAC1 is also phosphorylated and it has been reported that phosphorylation alters enzymatic activity and complex formation [68]. It might be that phosphorylation of ZmHDA1 also determines recruitment of complex components, since the enzymatically active 48 kDa form, which is presumably dephosphorylated, acts as a protein monomer. The regulation of maize HDA1 by limited proteolysis is probably a unique, plant-specific level of HDAC regulation.

Future prospects and perspectives
Four decades ago, Vincent Allfrey and co-workers [69] discovered the existence of acetylation and methylation on histones. With time, it became clear that acetylation does not function by simple charge neutralization on nucleosomes but rather sets specific regulatory signals in concert with other modifications [65,70]. We now know that histone modifications represent additional epigenetic information on chromatin that alters the functional

Figure 3. Model of interdependence of histone modifications. As outlined earlier [66], different modifications on adjacent sites influence each other. It has been shown that methylation of H3 lysine-9 (K9) inhibits phosphorylation of serine-10 (S10), and vice versa. The situation is complicated because lysine-9 can alternatively be acetylated. Similarly, serine-10 phosphorylation and lysine-14 acetylation cross-talk to each other. These modifications occur on one of the two nucleosomal H3 molecules, but the modification pattern on the second H3 molecule can be different and has a distinct influence on its second counterpart. There might be a direct influence of modifications of H4 on the modification pattern of H3, and vice versa. Many more factors will change the ‘letters’ of the histone code, leading to a subtle tuning of the functional meaning. The green arrow at lysine-20 of H4 indicates that this residue can be acetylated or methylated in plants. Question marks indicate the expected yet still unclear interdependent influences of modifications.
properties of the underlying genetic information; together with DNA methylation, chromatin remodeling and a range of non-histone factors, this histone code forms a complex epigenetic code that formats the genetic text and would better be named ‘the chromatin code’. At present, we can translate some words of this chromatin language but it will probably take decades to be able to translate and read the entire code.

Figure 3 aims to give an idea of the many open questions and the different levels of complexity of setting and translating the chromatin code. At present, we can analyze the interdependence of modifications exemplified by methylation of lysine-9, phosphorylation of serine-10 and acetylation of H3 lysine-14. However, what about the second H3 molecule in a nucleosome? Does it make a difference if the modification pattern is ‘asymmetric’ within a nucleosome? What are the influences of modifications in one or both H4 molecules on the modification pattern of H3? And there are still two molecules each of H2A and H2B in the nucleosome. What is the effect of arginine-3 methylation in H4, when the methylation is introduced to a mono-acetylated rather than a tetra-acetylated H4? The number of combinations of different post-translational marks is enormous and their functional effects might be widespread. We have little information about post-translational modifications other than acetylation and methylation, or about the influence of histone H1 and non-histone proteins on the modification patterns of nucleosomes (Figure 3). Because many non-histone regulatory proteins get modified by HATs and HDACs, the question arises of whether the histone code is somehow mirrored on these regulatory proteins for interaction with histones, chromatin or DNA. It is therefore possible that the histone code is part of a much more universal protein code that governs interactions of proteins with other molecules.

Although many elements of setting and translating the chromatin code are fairly well conserved throughout different kingdoms and species, we have to await subtle or even pronounced differences in the details of this code. Investigations in plant systems have already revealed such differences and the special properties of plant cells and plant genomes, and their unique exposure to environmental influences suggest a plant dialect of the chromatin language.

Acknowledgements
I apologize for not citing all relevant literature owing to the enormous number of papers published on this subject during the past few years, and the reference limitation. Experimental work in our laboratory was continuously funded by the Austrian Science Foundation (FWF, P-14528) and the Austrian National Bank (Jubilaumsfonds P-9000).

References
26 Sarg, B. et al. (2002) Postsynthetic trimethylation of histone H4 at lysine 20 in mammalian tissues is associated with aging. J. Biol. Chem. 277, 39195–39201
30 Houben, A. et al. (2003) Methyltransferase of histone H3 in euchromatin of plant chromosomes depends on basic nuclear DNA content. Plant J. 33, 967–973
32 Soppe, W.J. et al. (2002) DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. EMBO J. 21, 6549–6559
38 Aufsatz, W. et al. (2002) HDAC6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. EMBO J. 21, 6832–6841
46 Zhao, Y. et al. (2003) HDAC-6 interacts with and deacetylates tubulin and microtubules in vivo. EMBO J. 22, 1168–1179
67 Körle, D. et al. (1999) Different types of maize histone deacetylases are distinguished by a highly complex substrate and site specificity. Biochemistry 38, 6769–6773
73 Rho, J. et al. (2001) Prmt5, which forms distinct homo-oligomers, is a member of the protein–arginine methyltransferase family. J. Biol. Chem. 276, 11393–11401
83 Rayasam, G.V. et al. (2003) NSD1 is essential for early post-implantation development and has a catalytically active SET domain. EMBO J. 22, 3153–3163