Biol. Chem., Vol. 379, pp. 633 – 646, June 1998 · Copyright © by Walter de Gruyter & Co · Berlin · New York

Review

The AP2/EREBP Family of Plant Transcription Factors

José Luis Riechmann and Elliot M. Meyerowitz*

Division of Biology 156-29, California Institute of Technology, Pasadena, California 91125, USA

* Corresponding author

AP2 (APETALA2) and EREBPs (ethylene-responsive element binding proteins) are the prototypic members of a family of transcription factors unique to plants, whose distinguishing characteristic is that they contain the so-called AP2 DNA-binding domain. *AP2/EREBP* genes form a large multigene family, and they play a variety of roles throughout the plant life cycle: from being key regulators of several developmental processes, like floral organ identity determination or control of leaf epidermal cell identity, to forming part of the mechanisms used by plants to respond to various types of biotic and environmental stress. The molecular and biochemical characteristics of the AP2/EREBP transcription factors and their diverse functions are reviewed here, and this multigene family is analyzed within the context of the *Arabidopsis thaliana* genome sequence project.

Key words: *Arabidopsis thaliana* / Disease resistance / Flower development / Genome sequence / Homeotic gene / Plant stress.

Introduction

Transcription factors often comprise (super)families of related proteins that share a homologous DNA-binding domain. If the DNA-binding motifs are ancient and have proliferated throughout evolution by means of gene duplication and rearrangement events, the resulting gene families are large and may comprise numerous genes in a single species as well as genes from organisms that belong to different branches of the tree of life. Well-known examples are homeobox-containing genes, which are found in yeast (*Saccharomyces cerevisiae*; 8 genes), plants (25 genes already identified in *Arabidopsis thaliana*), and metazoans [for example, *Drosophila melanogaster* (25) and *Homo sapiens* (101)] (Henikoff *et al.*, 1997), among other organisms, and MADS-box genes, also found in the same species (Shore and Sharrocks, 1995; Riechmann and Meyerowitz, 1997). Other gene families, however, appear to be restricted to certain taxa. For example, GAL4-related transcriptional regulators, one of the largest families in yeast (52 proteins), have been found only in fungi (Henikoff *et al.*, 1997), and POU domain factors, abundant among metazoans, have yet to be identified in fungi or plants (Ryan and Rosenfeld, 1997). The AP2/EREBP family of transcription factors shares with these two groups a limited distribution: so far, *AP2/EREBP*-related genes have been isolated only from plants.

The distinguishing characteristic of proteins of the AP2/EREBP family is that they contain either one or two APETALA2 (AP2) domains. The AP2 domain was first recognized as a repeated motif within the *Arabidopsis thaliana* AP2 protein (Jofuku *et al.*, 1994). Shortly afterwards, four DNA-binding proteins from tobacco were identified that interact with a sequence that is essential for the responsiveness of some promoters to the plant hormone ethylene, and were designated as 'ethylene-responsive element binding proteins' (EREBPs) (Ohme-Takagi and Shinshi, 1995). The DNA-binding domain of EREBP-2 was mapped to a region that was common to all four proteins (Ohme-Takagi and Shinshi, 1995), and that was found to be closely related to the AP2 domain (Weigel, 1995) but that did not bear sequence similarity to previously known DNA-binding motifs.

AP2/EREBP genes form a large family, with many members known in several plant species (Okamuro *et al.*, 1997a). The multigene family can be divided into two subfamilies based on whether the proteins contain one or two AP2 domains, the EREBP subfamily and the AP2 subfamily, respectively. The diversity of functions carried out by different AP2/EREBP proteins throughout the plant life cycle is illustrated by *AP2* and the EREBPs themselves. *AP2* participates in the control of several steps or processes of flower development, among them specification of organ and meristem identity and ovule and seed development (Komaki *et al.*, 1988; Bowman *et al.*, 1989; 1991; 1993; Kunst *et al.*, 1989; Irish and Sussex, 1990; Schultz and Haughn, 1993; Shannon and Meeks-Wagner, 1993; Jofuku *et al.*, 1994; Léon-Kloosterziel *et al.*, 1994; Modrusan *et al.*, 1994; Okamuro *et al.*, 1997b). In contrast, EREBP proteins bind to a *cis*-regulatory sequence widely conserved among ethylene-responsive pathogenesis-related (PR) genes, which are part of the battery of defense genes activated upon plant pathogen attack (Ohme-Takagi and Shinshi, 1995; Zhou *et al.*, 1997). EREBP proteins may therefore form part of the mechanism used by plants to respond to biotic stress.

In this review, we begin by summarizing the organization and features of the AP2/EREBP transcription factors. The *AP2/EREBP* multigene family is then analyzed within the context of the *Arabidopsis thaliana* genome sequence project, which provides insight into the complexity of this

extremely large group of mostly uncharacterized genes and suggestions for their future analyses. Finally, we review the current knowledge of the diverse functional roles of different AP2/EREBP proteins throughout the plant life cycle.

AP2/EREBP Transcription Factors

Sequence-specific DNA binding and the capability to activate (or repress) transcription are two features of transcription factors that are frequently separable into distinct functional domains, and the AP2/EREBP proteins adhere to this rule of modular organization. As described in detail below, sequence similarity among the AP2/EREBP proteins is mostly limited to the AP2 DNA-binding domain itself (approximately 70 aa), which is repeated (domains R1 and R2) within proteins of the AP2 subfamily (herein collectively called AP2-like proteins), such as AP2 (Jofuku *et al.*, 1994), *Arabidopsis* AINTEGUMENTA (ANT; Elliott *et al.*, 1996; Klucher *et al.*, 1996), and maize Glossy15 (Gl15; Moose and Sisco, 1996) (Figure 1). The region between domains R1 and R2, called 'L' for linker (approximately 25 aa), is also conserved among AP2-like proteins (Klucher *et al.*, 1996; Figure 1). Members of the EREBP subfamily (herein collectively called EREBP-like proteins), such as tobacco EREBP-2 (Ohme-Takagi and Shinshi, 1995), and *Arabidopsis* AtEBP (***A**rabidopsis **t**haliana* **e**thylene-responsive element **b**inding **p**rotein; Büttner and Singh, 1997), TINY (Wilson *et al.*, 1996), and CBF1 (**C**-repeat/dehydration responsive element **b**inding **f**actor 1; Stockinger *et al.*, 1997) have a single AP2 domain (Figure 2). In addition to the conserved AP2 DNA-binding domain, AP2/EREBP proteins share two other features that are characteristic of transcription factors: region(s) of biased amino acid composition typical of transcription activation domains, and possible nuclear localization signals.

Eukaryotic transcription activation domains show little conservation in amino acid sequence, but can be broadly divided into several groups according to their amino acid content: acidic-rich, glutamine-rich, proline-rich, and serine-/threonine-rich (Mitchell and Tjian, 1989; Seipel *et al.*, 1992; Gerber *et al.*, 1994). Activation domains consisting of sequences in which more than one of these residue types are clustered (for example, S/T and P; Q and S/T; or S/T and D/E) are also common (Bennicelli *et al.*, 1995; Coustry *et al.*, 1995; Matsuzaki *et al.*, 1995; Dörfler and Busslinger, 1996), as are transcription factors that rely on two or more interdependent domains to activate transcription (Tanaka and Herr, 1990; Winter *et al.*, 1992; Hwang *et al.*, 1993). Sequence analysis reveals significant variability among different AP2/EREBP proteins in the domains that they might use to activate transcription (Figures 1 and 2). For example, AP2 contains an acidic and serine-

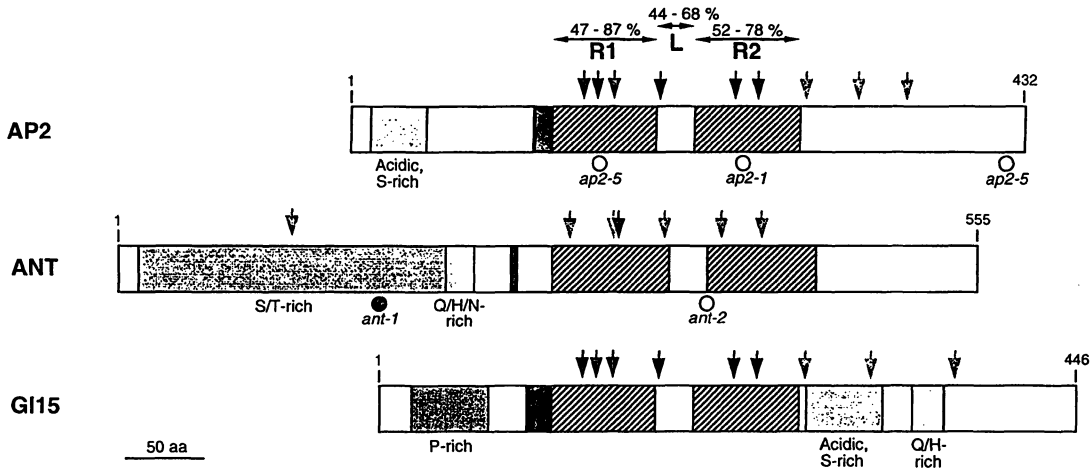

Fig. 1 Organization of AP2/EREBP Plant Proteins: The AP2 Subfamily.
Schematic representation of proteins with two AP2 domains, after the sequences of *Arabidopsis* APETALA2 (AP2; Jofuku *et al.*, 1994), AINTEGUMENTA (ANT; Elliott *et al.*, 1996; Klucher *et al.*, 1996), and maize Glossy15 (Gl15; Moose and Sisco, 1996). The two AP2 domains (R1 and R2, hatched rectangles), and the linker segment (L) that connects them, are highly conserved among AP2, ANT, and Gl15 (the percentage of amino acid identity is indicated). Outside of these three conserved regions, the proteins bear little similarity, but all of them exhibit several hallmarks of transcription factors. A basic region that may function as a nuclear localization sequence is indicated by black rectangles (aa 119-129 of AP2, 252-255 of ANT, and 96-111 of Gl15). Several segments of these proteins exhibit a biased amino acid composition typical of different types of transcription activation domains (Mitchell and Tjian, 1989; Seipel *et al.*, 1992; Bennicelli *et al.*, 1995; Matsuzaki *et al.*, 1995; Dörfler and Busslinger, 1996): acidic and serine-rich (aa 14 to 50 of AP2, and 278 to 326 of Gl15), serine- and threonine-rich (aa 13 to 214 of ANT), and proline-rich (aa 21 to 69 of Gl15). A segment rich in glutamine and histidine residues is present in ANT (aa 214 to 231) and in Gl15 (345-364), and the carboxyl-terminus of Gl15 (aa 364 to 446) is very rich in alanine residues. Positions that correspond to intron/exon boundaries are indicated by arrowheads; black arrowheads indicate those that are conserved in two (or three) of the genes (six positions are conserved between *AP2* and *Gl15*). Positions of amino acids that are changed as a result of the mutations present in different alleles are indicated with circles: the *ap2-1* mutation changes residue Gly-251 to Ser, *ap2-5* changes Gly-159 and Gln-420 to Glu, *ant-2* converts Gly-382 (the C-terminal residue of the L region) to Asp, and *ant-1* consists of a 22 nucleotide deletion that generates a frameshift immediately followed (after one amino acid change) by a stop codon.

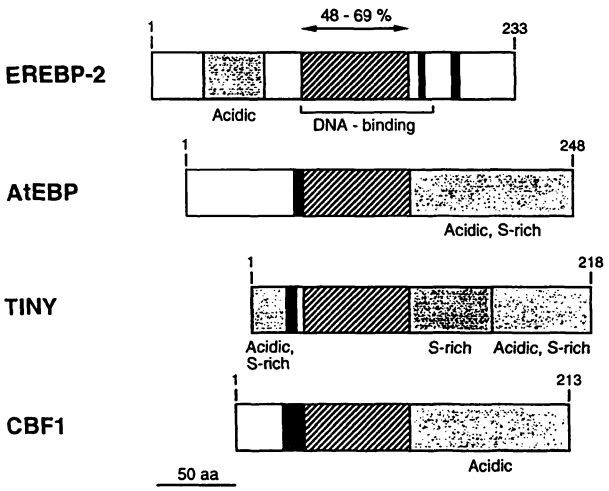

Fig. 2 Organization of AP2/EREBP Plant Proteins: The EREBP Subfamily.

Schematic representation of proteins with one AP2 domain, after the sequences of tobacco EREBP-2 (Ohme-Takagi and Shinshi, 1995), and *Arabidopsis* AtEBP (Büttner and Singh, 1997), TINY (Wilson *et al.*, 1996), and CBF1 (Stockinger *et al.*, 1997). The intron/exon structure of the corresponding genes has not been determined, except for TINY, which contains no introns (Wilson *et al.*, 1996). The AP2 domain (hatched rectangle) is highly conserved among EREBP-2, AtEBP, TINY, and CBF1 (the percentage of amino acid identity is indicated), and all other members of the family. The DNA-binding region of EREBP-2 (Ohme-Takagi and Shinshi, 1995) is indicated; because only a coarse deletion analysis of EREBP-2 was performed, the minimal DNA-binding domain may be smaller than indicated. Outside of the AP2 domain, the proteins bear little similarity, but all of them exhibit several characteristics of transcription factors: basic sequences that may serve as nuclear localization signals (black rectangles; aa 172 to 174 and 193 to 197 of EREBP-2, 71 to 75 of AtEBP, 23 to 28 of TINY, and 32 to 44 of CBF1) and putative transcription activation domains, either acidic (aa 33 to 71 of EREBP-2, and 112 to 213 of CBF1), serine-rich (aa 100 to 153 of TINY), or acidic and serine-rich (aa 143 to 248 of AtEBP, and 1 to 22 and 154 to 218 of TINY).

rich sequence; in ANT, an extended region rich in serine and threonine, but not acidic, is present; Gl15 contains both proline-rich and acidic and serine-rich segments; and other proteins, like EREBP-2, bear acidic domains (Figures 1 and 2). The capability to activate transcription has been demonstrated in heterologous yeast systems for the EREBP-like proteins CBF1 (Stockinger *et al.*, 1997) and Pti5 and Pti6 (from tomato; Zhou *et al.*, 1997), all of which contain acidic domains, and for the AP2-like protein ANT (Vergani *et al.*, 1997). The particular regions that mediate activation in each case remain to be defined.

AP2/EREBP proteins contain short stretches of basic amino acid residues that could function as nuclear localization signals (NLS), by analogy to similar sequences in other plant transcription factors that direct nuclear targeting (van der Krol and Chua, 1991; Raikhel, 1992; Varagona *et al.*, 1992). Those clusters of basic residues are usually (but not always) localized in proximity to the amino terminus of the AP2 domain (Figures 1 and 2).

The AP2 Domain

A large number of sequences encoding AP2/EREBP proteins are already present in the databases and can be easily identified through BLAST searches due to the conservation of the AP2 domain, which, as mentioned above, is the only region conserved among all proteins of the family. Figure 3 shows an alignment of the AP2 domain sequences of 40 proteins, some of which correspond to the products of characterized genes from several plant species whereas others are deduced from *Arabidopsis thaliana* genomic sequences (see legend of Figure 3 for details).

Several features of the AP2 domain, which is unrelated to other known DNA-binding motifs, stand out from the sequence alignment. They are summarized here and were previously noted by Okamuro *et al.* (1997a) in an analysis of a subset of these sequences. The AP2 domains of AP2-like proteins are more related among the members of this subfamily than to those of proteins of the EREBP subfamily, and *vice versa*, and therefore sequences from each subfamily were aligned separately (Figure 3), but many characteristics are common to the AP2 domains from both subgroups. Two conserved segments are found within each AP2 domain, which have been referred to as the YRG element and the RAYD element – both are named for amino acids conserved in most AP2 domains (Okamuro *et al.*, 1997a; Figure 3). The amino terminal part of the AP2 domain (the YRG element) is basic and hydrophilic. The carboxyl RAYD element contains a central region that, in almost all the AP2 domains, is predicted to adopt the configuration of an α-helix of amphipathic character (Jofuku *et al.*, 1994; Okamuro *et al.*, 1997a). That the YRG and RAYD segments may correspond to different structural elements is supported by the fact that the gaps required to align all AP2 domain sequences are introduced between them: gaps (i. e., insertions and deletions) appear to occur far more often in turns and coils between the main secondary structure elements of α helices and β strands of homologous proteins than within them (Pascarella and Argos, 1992). Five amino acid residues are absolutely conserved among all the AP2 domain sequences of both subfamilies, and that number is increased if the two subfamilies are considered separately (Figure 3). The conservation of those residues obviously suggests their importance for the structure/function of the AP2 domain, of which very little is known. It has been suggested that the YRG element could be directly involved in DNA binding due to its basic character (Okamuro *et al.*, 1997a). The RAYD element, because of the presumed amphipathic α-helical structure of its central region, could be involved in protein-protein interactions, but the possibility that it is involved in contacts with the DNA has also been considered (Okamuro *et al.*, 1997a).

EREBP-2, AtEBP, Pti5, and Pti6, all of which are members of the EREBP subfamily, have been shown to bind DNA fragments containing the sequence TAAGAGCCGCC (Ohme-Takagi and Shinshi, 1995; Büttner and

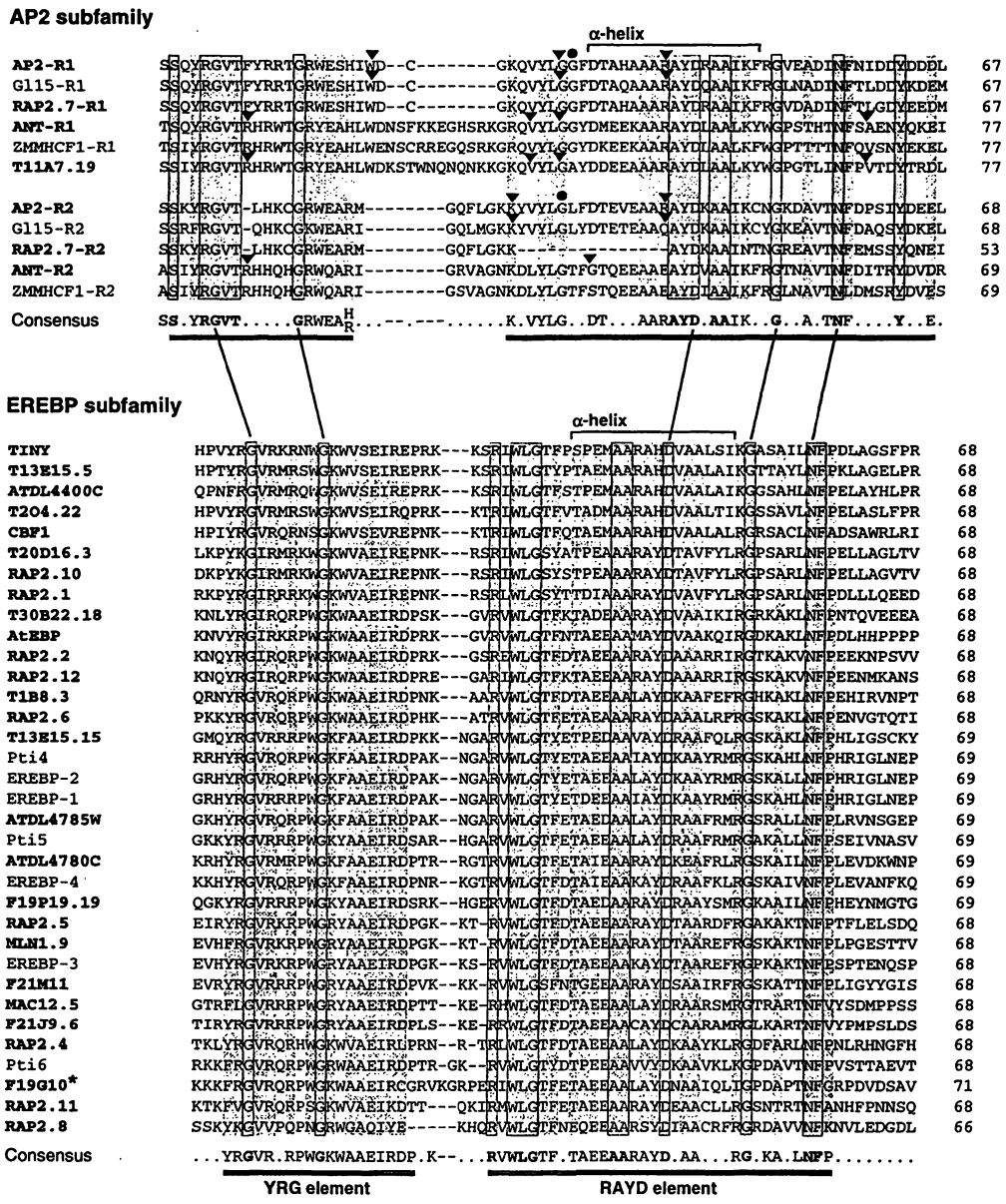

Fig. 3 The AP2 Domain.

Alignment of the amino acid sequences of the AP2 domains of proteins that belong to the AP2 subfamily (both AP2 domains, R1 and R2, are included) and to the EREBP subfamily. Because sequences within one subfamily are more related to one another than to sequences of the other subfamily, proteins from each subfamily were aligned separately. Sequences were identified through BLAST searches (Altschul *et al.*, 1990) and aligned using the Clustal option of the GeneWorks program (IntelliGenetics). Protein names are shown to the left, and those from *Arabidopsis* proteins are in bold. Within each alignment, regions of highest similarity are shaded, and amino acid identities among all the sequences included in the alignment are boxed. The corresponding consensus sequence is shown below each alignment (invariant residues shown in bold). The five amino acid residues that are conserved among all of the AP2/EREBP proteins are indicated with lines connecting the two alignments. Brackets above the alignments indicate the region that is predicted to form an amphipathic α-helix (Jofuku *et al.*, 1994; Okamuro *et al.*, 1997a). The YRG and RAYD elements (see main text) are indicated with solid lines below the consensus sequences. For proteins of the AP2 subfamily, both AP2 domains (R1 and R2) were included in the analysis. Positions that correspond to boundaries between exons in AP2, Gl15, ANT, and T11A7.19 are indicated by inverted triangles; other genes either have no introns (like *TINY*, Wilson *et al.*, 1996, and *RAP2.10*) or their genomic structure has not yet been determined. Amino acids that are changed as a result of the mutations present in *ap2-5* (a G to E change in AP2-R1) and *ap2-1* (a G to S change in AP2-R2) alleles are indicated with filled circles (Jofuku *et al.*, 1994). For some of the sequences that are shown in the alignments the corresponding genes have been characterized; other protein sequences are deduced from genomic sequences made available through the *Arabidopsis* Genome Initiative by the different sequencing groups (for more information on the *Arabidopsis* Genome Project, see the *Arabidopsis thaliana* database -AtDB- at http://genome-www.stanford.edu/Arabidopsis/; Flanders *et al.*, 1998). Some of those genomic sequences are already present in the EST (expressed sequence tag) collections, indicating that they are expressed genes (see below). In order to avoid inaccuracies, no EST-only derived sequences were included in the analysis. Details are as follows. *AP2*, GenBank accession number U12546 (Jofuku *et al.*, 1994; the genomic sequence of *AP2* is available under accession number Z99707); *Gl15*, from maize, U41466 (Moose and Sisco, 1996); *RAP2.1*, *RAP2.2*, *RAP2.4*, *RAP2.5*, *RAP2.6*, *RAP2.7*, *RAP2.8*, *RAP2.9*, *RAP2.10*, *RAP2.11*, and *RAP2.12*, accession numbers AF003094, AF003095, AF003097, AF003098, AF003099, AF003100, AF003101, AF003102, AF003103, AF003104, and AF003105, respectively

Singh, 1997; Zhou *et al.*, 1997), referred to as the GCC box or the PR box. The GCCGCC core of this sequence appears essential for the binding, because binding is lost if the element is mutated to TCCTCC (Ohme-Takagi and Shinshi, 1995; Büttner and Singh, 1997; Zhou *et al.*, 1997). Furthermore, the GCC box is protected in DNase I footprinting analysis performed with AtEBP (Büttner and Singh, 1997). CBF1 binds the C-repeat/dehydration responsive element (DRE), TACCGACAT (Stockinger *et al.*, 1997). The similarity between the GCC box and the C-repeat/DRE sequences prompted the suggestion that those proteins (all of which belong to the EREBP subfamily) may recognize related *cis*-regulatory sequences having CCG as a common core, with the different flanking sequences perhaps dictating binding to distinct AP2 domain proteins (Stockinger *et al.*, 1997). DNA-binding by AP2-like proteins has not been described yet, but it has been advanced that the EREBPs and AP2 may recognize distinct binding sites, perhaps owing to the differences in sequence between the YRG elements of the AP2 domains of the EREBP-like proteins and of the AP2-like proteins (Okamuro *et al.*, 1997a).

The functional and structural significance of having two AP2 domains brought together as a single entity in the AP2-like proteins is unknown, but analysis of *ap2* mutant alleles indicates that both domains (R1 and R2) are required for proper AP2 function (Jofuku *et al.*, 1994; Figure 1). The linker segment between R1 and R2 is highly conserved in sequence among AP2-like proteins, suggesting functional/structural constraints; its importance is suggested by the strong *ant-2* mutant allele, in which the carboxyl-terminal residue of the linker region is changed (Klucher *et al.*, 1996; Figure 1). The configuration of two (or repeated) DNA-binding domains in a single protein is unusual among transcription factors, but not unique to the AP2-like proteins. A bipartite DNA-binding domain is present in the POU domain proteins. The POU domain consists of two structurally independent domains that cooperate functionally as a DNA-binding unit: an amino terminal POU-specific domain (POU_S) and a carboxyl POU-homeodomain (POU_H), both of which make sequence-specific contacts with DNA through helix-turn-helix structures (reviewed in Herr and Cleary, 1995, and Ryan and Rosenfeld, 1997). An example of a repeated DNA-binding motif is provided by the MYB proteins. MYB proteins contain a conserved sequence that is present as imperfect tandem repeats, which constitute the MYB DNA-binding domain (usually three repeats – R1, R2, and R3 – are present in the animal MYB proteins, whereas most MYB proteins from plants contain two repeats – R2 and R3 – but there are proteins that contain only one of these sequences; Lipsick, 1996; Martin and Paz-Ares, 1997). *MYB* genes in plants also form an extremely large multigene family (Martin and Paz-Ares, 1997). Another family of plant transcription factors that, like the AP2/EREBP proteins, contain either a repeated or a single DNA-binding domain is the WRKY family (Ishiguro and Nakamura, 1994; Rushton *et al.*, 1995; 1996; Pater *et al.*, 1996).

The *AP2/EREBP* Multigene Family

AP2/EREBP genes are abundant in plant genomes. The approximate number of *Arabidopsis thaliana AP2/EREBP* genes can be estimated from the genomic sequences made available through the *Arabidopsis* Genome Initiative (see Figure 3 for details). Eighteen different *AP2/EREBP* genes were found in the 14.1 Mb of genomic sequence from completely sequenced BAC clones deposited in the databases as of December 1997. If the size of the *Arabidopsis* genome is 100 Mb (Meyerowitz, 1994; Goodman *et al.*, 1995), then the expected number of *AP2/EREBP* genes is approximately 125, which represents 0.6% of the total complement of *Arabidopsis* genes (estimated as 21 000; The EU Arabidopsis genome project, 1998). In addition, the EREBP subfamily likely contains many more members than the *AP2* subfamily, given that significantly more genes of the *EREBP* subfamily are found in the *Arabidopsis* genomic fragments already sequenced (16 versus 2; Figure 3). This calculation of the total number of

(Okamuro *et al.*, 1997a) (the genomic sequence of *RAP2.10* is available under accession number Z99707); *ANT*, U40256 and U41339 (Elliott *et al.*, 1996; Klucher *et al.*, 1996); *ZMMHCF1*, from maize, Z47554; *TINY*, X94698 (Wilson *et al.*, 1996); *CBF1*, U77378 (Stockinger *et al.*, 1997); *AtEBP*, Y09942 (Büttner and Singh, 1997); *Pti4*, *Pti5*, and *Pti6*, from tomato, accession numbers U89255, U89256, and U89257, respectively (Zhou *et al.*, 1997); *EREBP-1*, *EREBP-2*, *EREBP-3*, and *EREBP-4*, from tobacco, D38123, D38126, D38124, and D38125, respectively (Ohme-Takagi and Shinshi, 1995); *T11A7.19*, *T13E15.5*, *T13E15.15*, *T2O4.22*, *T20D16.3*, *T30B22.18*, and *T1B8.3* (deduced from BAC clones T11A7, T13E15, T2O4, T20D16, T30B22, and T1B8; accession numbers AC002339, AC002388, AC001645, AC002391, AC002535, and U78721, respectively; http://www.tigr.org/tdb/at/at.html); *ATDL4400C* (deduced from nt 186 161 to 185 625 of ESSAI contig fragment number 6, accession number Z97341; http://muntjac.mips.biochem.mpg.de/arabi/index.html); *ATDL4785W* and *ATDL4780C* (deduced from nt 124 417 to 124 980, and 118 592 to 117 747, respectively, of ESSAI contig fragment number 8, accession number Z97343; http://muntjac.mips.biochem.mpg.de/arabi/index.html); *F19P19.19* (deduced from BAC clone F19P19; AC000104; http://pgec-genome.pw.usda.gov); *MLN1.9* and *MAC12.5* (deduced from P1 clones MLN1 and MAC12; accession numbers AB005239 and AB005230, respectively; http://www.kazusa.or.jp/arabi/); *F21M11* (deduced from BAC clone F21M11; AC003027; http://sequence-www.stanford.edu/ara/ArabidopsisSeqStanford.html); *F21J9.6* (deduced from BAC clone F21J9; AC000103; http://cbil.humgen.upenn.edu/~atgc/ATGCUP.html); and *F19G10** (deduced from nt 35 216 to 35 692 of BAC clone F19G10; http://sequence-www.stanford.edu/ara/ArabidopsisSeqStanford.html). *ATDL4785W*, *ATDL4780C*, *T20D16.3*, *T13E15.5*, and *T13E15.15* match the sequences of the following ESTs, respectively: AA394962, T43246, N96896, Z37651, and N97133. Close inspection of the cDNA sequences of *AtEBP*, *RAP2.3* (AF003096; Okamuro *et al.*, 1997a), and *ATCADINP* (Z37504) suggests that they may correspond to the same gene, and therefore only *AtEBP* was included in the analysis.

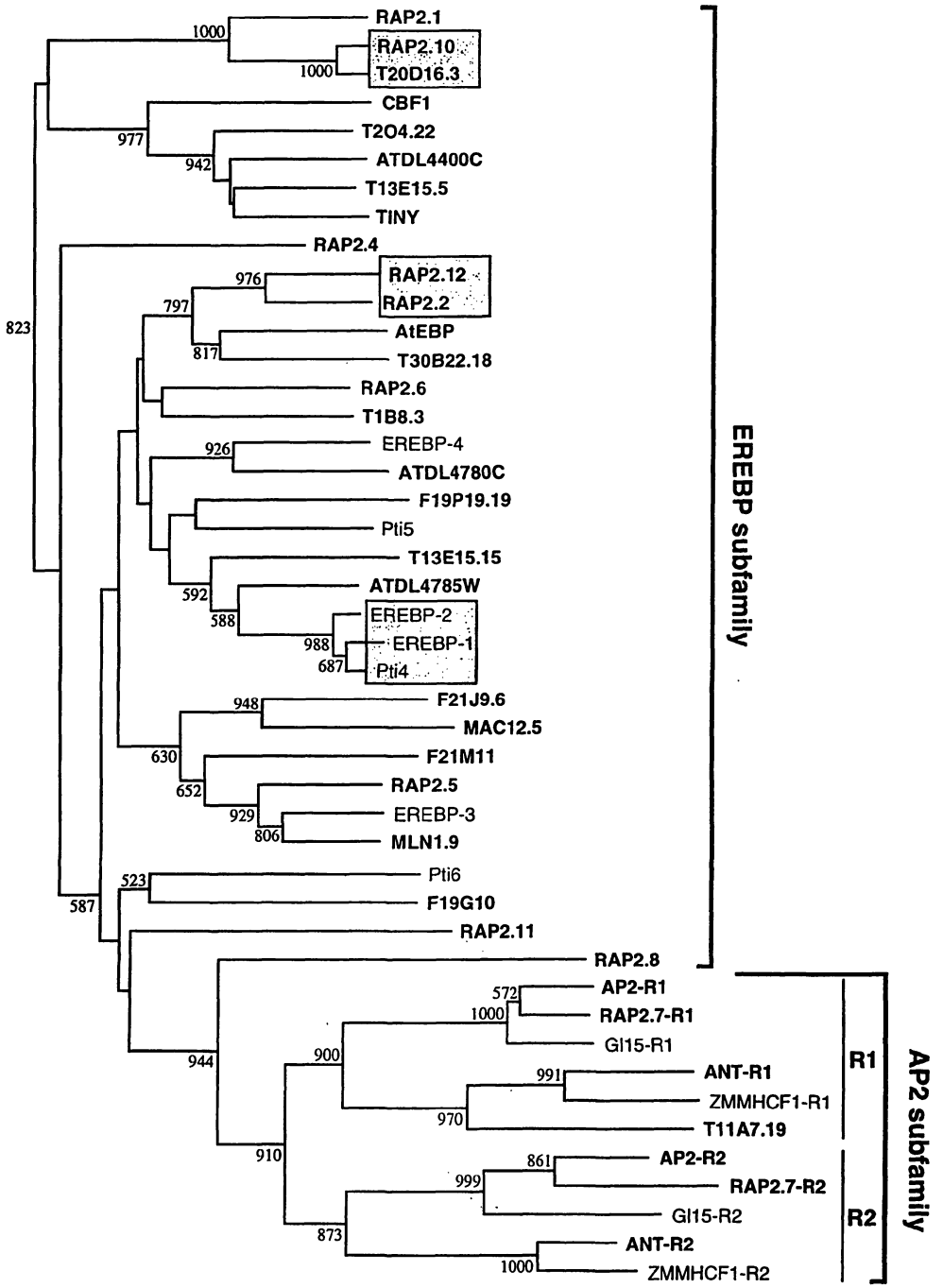

Fig. 4 The *AP2/EREBP* Multigene Family.
Neighbor-Joining tree showing the relationship among the sequences used in Figure 3. All the AP2 domain sequences were aligned together using the Clustal X program, in which the pairwise similarity scores are used to build a phylogenetic tree by the Neighbor-Joining (NJ) method. Only AP2 domain sequences were used in the analysis, because most AP2/EREBP proteins do not share any sequence similarity outside of that region. A few proteins, however, bear sequence similarity in the regions amino- or carboxyl-terminal to the AP2 domain, and this is indicated by enclosing their names in shaded boxes. EREBP-1 and EREBP-2, from tobacco, and Pti4, from tomato, are similar in their N- and C-terminal regions. RAP2.2 and RAP2.12, and RAP2.10 and T20D16.3, have related C-terminal regions. Numbers next to the nodes give bootstrap values from 1000 replicates (only those with more than 50% bootstrap support are indicated). The general topology of the tree is consistent with other features of the AP2/EREBP proteins (see main text). *T11A7.19* is unique in that it apparently codes for a protein with only one AP2 domain, but it clearly belongs to the *AP2* subfamily because of the sequence of that domain, which is of the R1 class and is encoded by several exons (Figure 3). In addition, the AP2 domain is followed by the characteristic L region. *T11A7.19* may therefore be an *AP2*-like gene in which the second (R2) domain has been lost during evolution.

Arabidopsis AP2/EREBP genes assumes that they are not significantly clustered but are instead scattered throughout the genome, as has been found for other *Arabidopsis* multigene families like, for example, those of MADS-box genes (Rounsley *et al.*, 1995) or of 20S proteasome genes (Parmentier *et al.*, 1997). In fact, most of the 18 *AP2/*

EREBP genes were identified in different BAC or YAC clones that, together, sample the five *Arabidopsis* chromosomes.

Furthermore, preliminary mapping experiments of several *RAP2* genes also reveal a widespread distribution (Okamuro *et al.*, 1997a; RAP2 stands for '**r**elated to **AP2**'; the term does not imply belonging to the *AP2* subfamily but, rather, to the *AP2/EREBP* family; see Figures 3 and 4). Some genes, however, are closely linked in the genome, like *AP2*, *ANT*, and *RAP2.10* (Elliott *et al.*, 1996; Klucher *et al.*, 1996; Okamuro *et al.*, 1997a), or T13E15.15 and T13E15.5 (see legend of Figure 3 for details), but such linkage may not be due to simple gene duplication events (see below).

A preliminary phylogenetic analysis of the *AP2/EREBP* family can be carried out using the AP2 domain sequences shown on Figure 3 (which may include 25% of the *Arabidopsis* family if the estimated number of 125 *AP2/EREBP* genes is correct). A Neighbor-Joining tree (Figure 4) shows that the AP2-like proteins are grouped together as a monophyletic group, or subfamily, apart from the EREBP-like proteins. Therefore, it appears that two AP2 domains were present in the product of the gene that was the common ancestor of all *AP2*-like genes. Because *AP2*-like genes are found in both *Arabidopsis* and maize, the origin of such a common ancestor (i. e., the origin of the subfamily) predates the monocot/dicot divergence (200–135 mya, according to molecular estimates or fossil evidence; Wolfe *et al.*, 1989; Crane *et al.*, 1995). The subdivision of the multigene family into two subfamilies, as outlined from the introduction, is therefore further supported and, in summary, is based on:

(i) The presence of either one or two AP2 domains in the proteins, EREBP-like or AP2-like, respectively;
(ii) the conservation among AP2-like proteins of the linker region that joins the two AP2 domains (Klucher *et al.*, 1996);
(iii) the fact that the AP2 domain sequences of each subfamily are more related among subfamily members than to those of the other subfamily (Okamuro *et al.*, 1997a; Figure 3);
(iv) the estimated phylogenetic history of the family (Figure 4); and
(v) the different intron/exon structure of *EREBP*-like and *AP2*-like genes.

All *AP2*-like genes whose structure is known (*AP2*, *ANT*, and maize GI15) contain multiple exons and, in particular, both AP2 domains (R1 and R2) of each protein are encoded by several exons (Figures 1 and 3). In contrast, the sequences coding for the AP2 domain in *EREBP*-like genes are not interrupted by introns. This has been shown for *TINY* and *RAP2.10*, which are intronless genes (Wilson *et al.*, 1996; Figures 2 and 3), and is also apparent for all those subfamily members that are deduced from genomic sequences (see legend of Figure 3). The phylogenetic tree also shows that any R1 domain is more closely related to the R1 domains of the other AP2 subfamily members than to the R2 domain to which it is actually linked (or than to any other R2 domain), and *vice versa*, which may indicate distinct functional constraints for R1 and R2.

The analysis of 1.9 Mb of contiguous sequence from *Arabidopsis* chromosome 4 has revealed a frequent occurrence of close similarities among members of gene families that were located next to each other on the same DNA strand, on the basis of which it has been suggested that simple gene duplication and subsequent divergence may be a common mechanism for expanding gene families in *Arabidopsis* (The EU *Arabidopsis* genome project, 1998). Although some *Arabidopsis* gene families may well have expanded and evolved in part by such mechanism (for example, disease resistance-related genes; Botella *et al.*, 1997), this does not appear to be the case for the *AP2/EREBP* family. *AP2/EREBP* genes located on different chromosomes can be the most highly related, whereas genes that are clustered in the genome may be quite divergent and have different evolutionary histories (Figure 4). The following are several examples: *RAP2.10* (an *EREBP*-like gene) is located at a distance of approximately 10 kb from *AP2* on chromosome 4 (see legend of Figure 3); both genes are also linked to *ANT* (Elliott *et al.*, 1996; Klucher *et al.*, 1996), but *AP2* is more closely related to RAP2.7 than to *ANT* (Okamuro *et al.*, 1997a) (Figure 4); RAP2.10 and T20D16.3 are not only highly similar in the AP2 domain, but also share sequence similarity in the carboxyl region (in contrast to most other proteins of the family) (Figure 4), yet the respective genes are localized on chromosomes 4 (Okamuro *et al.*, 1997a) and 2 (legend of Figure 3); and *T13E15.5* and *T13E15.15*, which lie approximately 42 kb from each other on chromosome 2, are distantly related *EREBP*-like genes (Figure 4).

Functional Roles of the AP2/EREBP Proteins

The subdivision of the *AP2/EREBP* multigene family into *AP2*-like genes and *EREBP*-like genes on the basis of molecular criteria may reflect a functional dichotomy. The *AP2* subfamily genes whose functions have been determined by mutant analyses (*Arabidopsis AP2* and *ANT*, and maize *GI15*) act as key regulators in developmental processes, whereas the *EREBP* subfamily members so far characterized appear to be involved in responses to biotic and environmental stress, although their precise functions are largely unknown because no mutants for the corresponding genes have yet been isolated.

Plant Development

AP2 is involved in the specification of sepal and petal identity through its activity as a homeotic gene that forms part of the combinatorial genetic mechanism of floral organ identity determination (Komaki *et al.*, 1988; Bowman *et al.*, 1989; 1991; Kunst *et al.*, 1989), and in the negative regulation of the expression of the MADS-box floral homeotic

gene *AGAMOUS* (*AG*; Drews *et al.*, 1991). *AP2* also takes part in the specification of floral meristem identity (Irish and Sussex, 1990; Huala and Sussex, 1992; Bowman *et al.*, 1993; Schultz and Haughn, 1993; Shannon and Meeks-Wagner, 1993; Okamuro *et al.*, 1997b); and it is also required for normal ovule and seed development (Jofuku *et al.*, 1994; Léon-Kloosterziel *et al.*, 1994; Modrusan *et al.*, 1994). The *Arabidopsis* flower consists of four concentric whorls or rings of organs occupied by (from the outermost to the innermost whorl) four sepals, four petals, six stamens, and two fused carpels (Figure 5A). *ap2* mutations cause homeotic conversions in the outer two whorls of the flower: in weak *ap2* mutants, such as *ap2-1*, sepals are transformed into leaves and petals are staminoid (Bowman *et al.*, 1989; Figure 5B); in extreme mutants, like *ap2-2*, the medial sepals are converted into carpels, development of the lateral first whorl organs is most often suppressed, the second whorl organs do not arise, and the number of stamens in the third whorl is greatly reduced (Bowman *et al.*, 1991; Figure 5C). The phenotypic differences between weak and strong *ap2* mutants are explained by the different degree to which two of the *AP2* functions – controlling the pathways of sepal and petal development and repressing *AG* expression in whorls one and two – may be affected in each case (Bowman *et al.*, 1991; Drews *et al.*, 1991). *AG* directs stamen and carpel development in whorls three and four of wild type flowers, respectively, and its ectopic expression in whorls one and two of *ap2-2* mutant flowers results in the observed organ identity changes; in contrast, *ap2-1* appears to be a partial loss-of-function allele, in which the pathways of sepal (and petal) development are not properly activated but *AG* expression is still substantially repressed in the first two whorls, which therefore do not exhibit the homeotic changes characteristic of *ap2-2* mutant flowers (Bowman *et al.*, 1991; Drews *et al.*, 1991). Other phenotypes associated with *ap2* mutations are defective seed coat development (Jofuku *et al.*, 1994; Léon-Kloosterziel *et al.*, 1994) and the occasional replacement of ovules by filaments or carpelloid organ-like structures (Modrusan *et al.*, 1994). In addition, the combination of *ap2* mutations with mutations at two other loci, *apetala1* and *leafy*, uncovers the early function that *AP2* has in reinforcing the action of these so-called floral meristem identity genes, i. e., in providing floral identity to the primordia that arise on the flanks of the inflorescence apex (Irish and Sussex, 1990; Huala and Sussex, 1992; Bowman *et al.*, 1993; Schultz and Haughn, 1993; Shannon and Meeks-Wagner, 1993; Okamuro *et al.*, 1997b).

Although all of these *AP2* functions are related to flower development, the *AP2* expression domain includes both floral and vegetative tissues. *AP2* is expressed in the inflorescence meristem and throughout the floral primordia during the early stages of flower development; late in flower development, *AP2* transcripts are detectable in all organs but appear concentrated in petals and specific tissues of stamens and carpels, including ovules; in addition, *AP2* is expressed in vegetative leaves and in the stem (Jofuku *et al.*, 1994). This broad expression domain, that includes tissues and organs that do not appear affected in *ap2* mutant plants, like stems and leaves, contrasts with that of other floral homeotic genes belonging to the MADS-box family of transcription factors, for which the domains of expression and the realms of function (as inferred from the phenotypic analysis of the corresponding mutants) are largely coincident (for review: Riechmann and Meyerowitz, 1997). Partial genetic redundancy for *AP2* function and/or postranscriptional mechanisms of control of *AP2* activity may underlie these differences (see below).

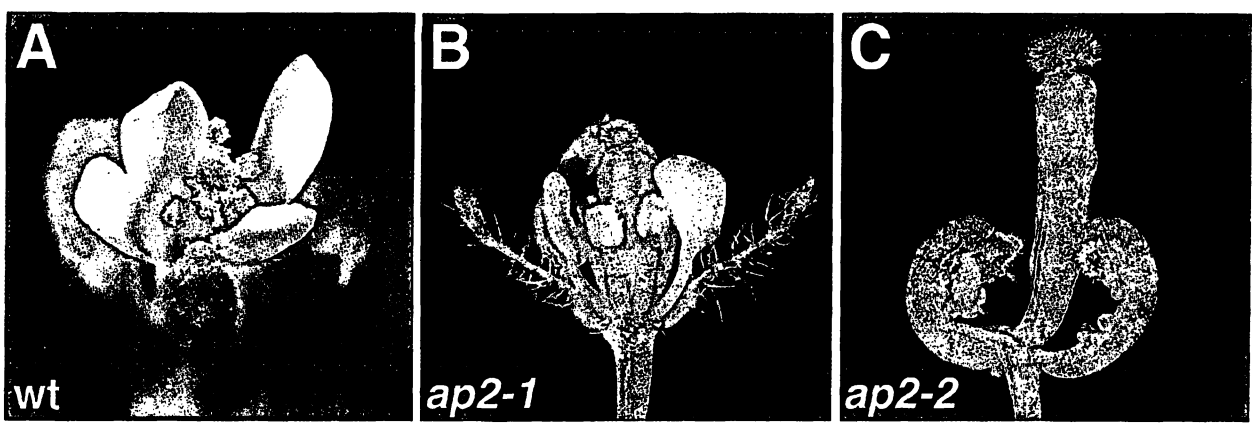

Fig. 5 *AP2* Participates in the Regulation of *Arabidopsis* Flower Development.
(A) Wild-type *Arabidopsis* flower (Landsberg *erecta* ecotype). Organs are: four sepals (first whorl), four petals (second whorl), six stamens (third whorl), and two fused carpels that occupy the center of the flower.
(B) Flower homozygous for the weak *ap2-1* allele (Bowman *et al.*, 1989). First whorl organs are leaf-like, with stellate (branched) trichomes, which are characteristic of leaves. Second whorl organs are slightly staminoid petals. One first-whorl organ has been removed to reveal the organs of the inner whorls.
(C) Flower homozygous for the extreme *ap2-2* allele (Bowman *et al.*, 1991). Lateral first whorl organs are missing, and the two medial first whorl organs are solitary carpels topped with stigmatic tissue and with ovules along their margins. No organs have developed in the second and third whorls, and two fused carpels occupy the fourth whorl.

ANT is required for ovule development and it also plays a role in floral organ growth, although it does not appear to have homeotic or organ identity functions (Elliott *et al.*, 1996; Klucher *et al.*, 1996). Developed angiosperm ovules consist of

(i) an embryo sac (or female gametophyte) formed by seven cells and eight haploid nuclei (after the double fertilization, one of the six haploid cells, the egg cell, generates the plant embryo, whereas the diploid central cell forms the triploid endosperm),
(ii) one or two integuments that enclose the embryo sac (formed by maternally-derived diploid cells and that will develop into the seed coat), and
(iii) a funiculus (a supporting stalk that connects the ovule to the ovary wall) (Gasser and Robinson-Beers, 1993; Reiser and Fischer, 1993).

In ovules of *ant* mutant plants, which are female sterile, integuments do not form and, although the megaspore mother cell arises (which in wild type ovules will eventually result in the seven-celled embryo sac through meiosis and megagametogenesis), the female gametophyte does not develop (Elliott *et al.*, 1996; Klucher *et al.*, 1996; Baker *et al.*, 1997; Schneitz *et al.*, 1997). In addition, *ant* flowers show a pleiotropic phenotype consisting of a reduction in the number of organs, alterations in their shape, and organ fusion as well as the formation of mosaic organs (Elliott *et al.*, 1996; Klucher *et al.*, 1996; Baker *et al.*, 1997).

The *ANT* expression pattern is complex and dynamic, and although it encompasses the organs showing defects in *ant* mutant plants (with the exception of the nucellus, where the megaspore arises; see Elliott *et al.*, 1996), *ANT* (similarly to *AP2*; Jofuku *et al.*, 1994), is also expressed in regions and tissues that are not affected in *ant* mutants (Elliott *et al.*, 1996). Consistent with the mutant phenotype, *ANT* is expressed in the region from which the ovule integuments arise as well as in the primordia of all floral organs; but, in addition, it is expressed in the funiculus, in the primordia of cotyledons and leaves, and in the procambium cells of the developing stem (Elliott *et al.*, 1996). It is possible that *ANT* is involved in controlling the development of these organs (a unifying theme for *ANT* function would be that it could have a general role in promoting primordium initiation and growth), but that its functions are redundant with those of other genes (Elliott *et al.*, 1996). Interestingly, floral organ development is severely disrupted in *ant ap2* double mutants, and this synergistic effect indicates that *ANT* and *AP2* functions partially overlap (Elliott *et al.*, 1996).

Maize *Gl15* regulates leaf epidermal cell identity. Vegetative development in maize (and in higher plants in general) is divided into a juvenile and an adult phase that have distinct morphological and physiological traits (Poethig, 1990). In particular, the epidermis of the basal 5 or 6 juvenile maize leaves consists largely of a single cell type and is characterized by the presence of epicuticular wax and the absence of epidermal hairs, whereas that of adult leaves lacks epicuticular wax, and differentiated bulliform cells with leaf hairs are present (Poethig, 1990; Evans *et al.*, 1994). *Gl15* is specifically required for expression of the juvenile and suppression of the adult epidermal characteristics in leaves 3–4 through 6, and no other juvenile traits nor the general morphology of the plant are affected by its mutation (Evans *et al.*, 1994; Moose and Sisco, 1994). In contrast to *AP2* and *ANT*, the expression pattern of *Gl15* mRNA strictly correlates with the mutant phenotype: it is detected as a low abundance transcript in juvenile leaves 4 to 6 and it is not detected in adult leaves (Moose and Sisco, 1996). In *Arabidopsis* there may be more than one gene closely related to *Gl15* within the *AP2*-subfamily, because sequences specifically related to *Gl15* (i. e., outside the conserved *AP2* domain) are present in low copy number, whereas in maize and several grasses they are present in a single copy (Moose and Sisco, 1996). The maize genome contains more *AP2*-like genes, besides *Gl15*, that are important in development. For example, *Indeterminate spikelets* (Ids) is an *AP2*-like gene that apparently functions to maintain the spikelets as determinate structures. Spikelets are second order branches that arise on the maize inflorescence and that in wild type plants generate a pair of third-order branches that develop into the florets (for a review on maize floral development, see: Veit *et al.*, 1993; Schmidt and Ambrose, 1998). Spikelets from loss-of-function *ids* mutants make more than the normal two florets (George Chuck and Sarah Hake, unpublished results). The sequence of another *AP2*-like gene from maize has been reported, but no data are available on its function (Daniell *et al.*, 1996).

Coping with Stress

The attack of a plant by a pathogen may induce defense responses that lead to resistance to the invasion, and these responses are associated with transcriptional activation of defense-related genes, among them those encoding pathogenesis-related (PR) proteins (reviewed in Hammond-Kosack and Jones, 1996; Ryals *et al.*, 1996). Although the functions of many PR proteins remain to be determined, some have known catalytic activities (β-1,3-glucanases and chitinases, lytic enzymes that can degrade fungal cell wall polysaccharides), and evidence has accumulated causally linking PR proteins to disease resistance (Hammond-Kosack and Jones, 1996; Ryals *et al.*, 1996). The signaling pathways controlling the activation of defense responses are complex, and signal molecules that appear to be involved are salicylic acid, jasmonic acid, and ethylene (Hammond-Kosack and Jones, 1996; Ryals *et al.*, 1996). The role of ethylene, in particular, is far from understood, because different genetic studies with *Arabidopsis* **e**thylene **in**sensitive mutants (like *ein2*) have indicated both independence and dependence of disease resistance responses from the ethylene signal transduction pathway (Lawton *et al.*, 1994; Penninckx *et al.*, 1996). A possible explanation for these conflicting results may be that the role of ethylene could vary depending on the plant-pathogen interaction studied and on the response

analyzed (Penninckx *et al.*, 1996; Knoester *et al.*, 1998). In fact, the rate of ethylene biosynthesis increases during pathogen infection, and ethylene can induce the expression of several PR genes (for review: Ecker, 1995). Many PR gene promoters contain a short *cis*-acting element that mediates their responsiveness to ethylene, which has been referred to as the GCC-box, the PR-box, or the **e**thylene **r**esponsive **e**lement (ERE) (Eyal *et al.*, 1993; Hart *et al.*, 1993; Ohme-Takagi and Shinshi, 1995; Zhou *et al.*, 1997). Tobacco EREBP-1, -2, -3, and -4 proteins were identified owing to their specific binding to the GCC-box (Ohme-Takagi and Shinshi, 1995). All four *EREBP* genes are ubiquitously expressed, and their mRNA levels are substantially increased by ethylene (Ohme-Takagi and Shinshi, 1995), which may explain the requirement of protein synthesis for the transcriptional induction of ERE-containing plant defense response genes (Ecker, 1995).

Studies of a tomato resistance (*R*) gene, *Pto*, have provided more direct evidence linking the *EREBP*s and *EREBP*-like genes to the defense response. *R* genes control race-specific resistance, which depends on specific recognition of the invading pathogen by the plant as a result of the interaction (direct or indirect) between a plant *R* gene and the corresponding pathogen avirulence (*Avr*) gene product (reviewed in Bent, 1996). Pto is a protein kinase that confers resistance against *Pseudomonas syringae* pv *tomato* bacteria expressing the avirulence gene *avrPto* (Bent, 1996). Pto has been shown to directly interact in two-hybrid assays with several proteins: bacterial AvrPto, a protein kinase (Pti1), and the tomato EREBP-like proteins Pti4, Pti5, and Pti6 (Zhou *et al.*, 1997, and references therein). The significance of the Pto-Pti4/5/6 interaction is demonstrated by the lack of interaction of these EREBP-like proteins with pto, the product of the recessive *pto* allele from a tomato cultivar that is susceptible to *P. s. tomato* carrying *avrPto* (Zhou *et al.*, 1997). Furthermore, tobacco EREBP-2 (which is highly related to Pti4; Figure 4) also interacts with Pto and, using transgenic tobacco plants expressing Pto that were challenged with the appropriate avirulent bacteria, *EREBP-1* was shown to be induced early upon *Pto-avrPto* recognition, preceding the accumulation of PR mRNAs (Zhou *et al.*, 1997). All of these observations indicate that Pti4/5/6 and the EREBPs are probable components linking an *R* gene product (Pto) to the expression of defense-related genes (Zhou *et al.*, 1997).

Plants are challenged not only by pathogens, but also by adverse environmental conditions like cold or drought, and EREBP-like proteins appear to be involved in the responses to these abiotic stresses as well. The ability of plants to tolerate adverse conditions involves many biochemical and physiological changes, among them alterations in gene expression that are mediated in part by a cis-element (the C-repeat/dehydration responsive element; DRE) found in the promoters of many **co**ld-**r**egulated (COR) genes (Stockinger *et al.*, 1997). COR gene expression is induced during cold acclimation, the process by which plants increase their resistance to freezing in response to low unfreezing temperatures. The C-repeat/DRE is specifically recognized by the *Arabidopsis* EREBP-like protein CBF1 (Stockinger *et al.*, 1997). Ectopic expression of *CBF1* in *Arabidopsis* transgenic plants induced *COR* gene expression in the absence of a cold stimulus, and the plant freezing tolerance was increased, indicating that *CBF1* is a regulator of the cold acclimation response (Jaglo-Ottosen *et al.*, 1998).

The functions of other *EREBP*-like genes, for example *TINY*, remain more elusive. *TINY* was isolated as a result of a transposon-mutagenesis experiment designed to recover dominant gain-of-function alleles in *Arabidopsis* (Wilson *et al.*, 1996). *tiny* mutant plants show pleiotropic effects (reductions in height and hypocotyl elongation, and reduced fertility) that may be the consequence of a general reduction in cell expansion and result from increased or ectopic expression of *TINY* (Wilson *et al.*, 1996). Interestingly, some of the defects observed in *tiny* mutant plants are similar to those caused by mutations that result in a constitutive activation of the ethylene signal transduction pathway (Kieber *et al.*, 1993). The function of *TINY* in wild type plants, however, cannot be ascertained at present because loss-of-function alleles are not available.

Concluding Remarks and Prospects

The analysis of any multigene family with more than a hundred members is a tremendous challenge, and some of the features of the *AP2/EREBP* family will make the endeavor even more demanding. The characterization of all family members in a single plant species will be facilitated by, and ultimately accomplished owing to, the *Arabidopsis* genome sequence project. Mutants for all the *Arabidopsis AP2/EREBP* genes may eventually be obtained by screening pools of T-DNA or transposon mutagenized lines (McKinney *et al.*, 1995; Krysan *et al.*, 1996) or perhaps by targeted gene disruption (Kempin *et al.*, 1997). When analyzing the effects that such mutations in the *AP2/EREBP* genes have on the plant, at least two potential problems can be expected. Because of the large size of the gene family, it is possible that some of its members are (partially) redundant in function, in which case plants that are singly mutant for those genes would be phenotypically wild-type. Indeed, *AP2* and *ANT* functions have been shown to partially overlap, and functional redundancy is one (but not the only) explanation for why some of the regions or organs in which *AP2* or *ANT* are expressed may not be affected in the corresponding mutants. Another potential problem in mutant analysis is derived from the roles that at least some *EREBP* genes appear to play: if they are involved in controlling the plant responses to particular stresses, a mutant phenotype may not be uncovered unless such particular stress conditions are present during the analysis.

Even if some of the AP2-like and EREBP-like proteins are partially redundant in function within each subfamily,

biological specificity must exist as well. How is that obtained? A primary source of specificity might be differential DNA recognition, and the anticipated differences in binding sites between AP2 and the EREBPs, and between the EREBPs and CBF1, point in this direction. It will be important to characterize the DNA-binding properties of the AP2-like and EREBP-like proteins and, ultimately, determine their crystal structures, which will be interesting on their own because of the novelty of the AP2 DNA-binding domain with its added peculiarity of being either single or repeated in these proteins. However, based on the large size of the *AP2/EREBP* family it is most likely that many proteins within each subfamily will be found to have very related, if not identical, DNA-binding specificities, similar to what has been observed for other large families of transcription factors, like those of the homeodomain proteins (Biggin and McGinnis, 1997) or the MADS domain proteins (Riechmann and Meyerowitz, 1997). Another potential source of functional specificity is specific protein-protein interactions with additional transcription factors and the concomitant formation of ternary or multiprotein complexes with novel properties, which is a common theme in eukaryotic transcriptional regulation. Indeed, the EREBP-like protein AtEBP was cloned because of its specific interactions with *Arabidopsis* basic-region leucine zipper protein OBF4 (***o**cs* element **b**inding **f**actor 4; Büttner and Singh, 1997). The significance of the AtEBP-OBF4 interaction remains to be determined, but the promoters of several PR genes contain both GCC-box and G-box sequences (the G-box resembles an *ocs* element half site and is recognized by a subfamily of bZIP proteins, the G-box-binding factors; Büttner and Singh, 1997, and references therein). Yet additional functional specificity might arise from the diversity of activation domains that the AP2/EREBP proteins exhibit, for example, if different domains have different activation properties (like stimulating transcription from remote or proximal positions; Seipel *et al.*, 1992), or if their activity is regulated by mechanisms like phosphorylation.

That the activity of the AP2/EREBP proteins might be postranscriptionally regulated is suggested from the observation that *CBF1* transcript levels do not change significantly in response to low temperature or water deficit, whereas these conditions induce the expression of genes that are putatively regulated by CBF1 (i. e., expression of CBF1 does not cause activation of promoters containing C-repeat/DRE sequences under normal growth condition; Stockinger *et al.*, 1997). Activation of CBF1 could occur, for example, by phosphorylation. Phosphorylation can control transcription by regulating localization (nuclear translocation), DNA-binding activity, or transactivation potential of transcription factors (Hunter and Karin, 1992; Karin and Hunter, 1995). It should be noted, however, that *E. coli*-produced EREBP-2, Pti5, Pti6, and CBF1 proteins bind to DNA (Ohme-Takagi and Shinshi, 1995; Stockinger *et al.*, 1997; Zhou *et al.*, 1997). Regulation of protein activity by postranslational modifications or by protein stability/degradation is another possible explanation, besides functional redundancy, for the occasional lack of correlation between domains of expression and domains of function (as defined by mutant analysis) of genes like *AP2* or *ANT*.

Because of the involvement of AP2-like proteins in development, the study of the family might be relevant for an understanding of the evolution of plants and of plant form, and the multigene family itself poses interesting evolutionary questions regarding its origin and molecular and functional diversification. Since the family had already diverged into two subfamilies before the monocot/dicot split, it must be ancient in plant evolution. How ancient? Are there *EREBP*-like genes in non vascular plants, for example? When and where did *AP2*-like genes arise? Finally, to conclude this overview of the AP2/EREBP proteins, it should be noted that the interest in this family of plant transcription factors goes beyond all those aspects and extends into their possible biotechnological applications. For example, overexpression of *CBF1* has been shown to enhance plant freezing tolerance (Jaglo-Ottosen *et al.*, 1998). Similarly, the involvement of Pti4/5/6 in PR gene expression upon pathogen attack suggests that an approach to generate crop plants with enhanced disease resistance may make use of these transcription factors to coordinately express a large combination of PR proteins in the plant (Zhou *et al.*, 1997). In addition, plant transcription factors that are not found in animals, like the AP2/EREBP proteins, may provide for novel systems of engineered gene regulation in animals that may not interfere with the endogenous transcriptional regulatory circuitry.

Acknowledgements

We are grateful to George Chuck and Sarah Hake for communicating and allowing us to cite their unpublished results on the maize *Ids* gene, and to Xuemei Chen, Jennifer Fletcher, Toshiro Ito, Steve Jacobsen, Prakash Kumar, Carolyn Ohno, Kevin Roberg, Robert Sablowski, Doris Wagner, and Eva Ziegelhoffer for valuable comments on the manuscript. Work in the laboratory of E. M. M. is supported by grants from the National Science Foundation, the National Institutes of Health, the Department of Energy, the Department of Agriculture, and Zeneca Seeds.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Baker, S.C., Robinson-Beers, K., Villanueva, J.M., Gaiser, J.C., and Gasser, C.S. (1997). Interactions among genes regulating ovule development in *Arabidopsis thaliana*. Genetics *145*, 1109–1124.

Bennicelli, J.L., Fredericks, W.J., Wilson, R.B., Rauscher III, F.J., Barr, F.G. (1995). Wild type PAX3 protein and the PAX3-FKHR fusion protein of alveolar rhabdomyosarcoma contain potent, structurally distinct transcriptional activation domains. Oncogene *11*, 119–130.

Bent, A.F. (1996). Plant disease resistance genes: function meets structure. Plant Cell *8*, 1757–1771.

Biggin, M.D., and McGinnis, W. (1997). Regulation of segmentation and segmental identity by *Drosophila* homeoproteins: the role of DNA binding in functional activity and specificity. Development *124*, 4425–4433.

Botella, M.A., Coleman, M.J., Hughes, D.E., Nishimura, M.T., Jones, J.D.G., and Somerville, S.C. (1997). Map positions of 47 *Arabidopsis* sequences with sequence similarity to disease resistance genes. Plant J. *12*, 1197–1211.

Bowman, J.L., Smyth, D.R., and Meyerowitz, E.M. (1989). Genes directing flower development in *Arabidopsis*. Plant Cell *1*, 37–52.

Bowman, J.L., Smyth, D.R., and Meyerowitz, E.M. (1991). Genetic interactions among floral homeotic genes of *Arabidopsis*. Development *112*, 1–20.

Bowman, J.L., Alvarez, J., Weigel, D., Meyerowitz, E.M., and Smyth, D.R. (1993). Control of flower development in *Arabidopsis thaliana* by *APETALA1* and interacting genes. Development *119*, 721–743.

Büttner, M., and Singh, K.B. (1997). *Arabidopsis thaliana* ethylene-responsive element binding protein (AtEBP), an ethylene-inducible, GCC box DNA-binding protein interacts with an ocs element binding protein. Proc. Natl. Acad. Sci. USA *94*, 5961–5966.

Coustry, F., Maity, S.N., de Crombrugghe, B. (1995). Studies on transcription activation by the multimeric CCAAT-binding factor CBF*. J. Biol. Chem. *270*, 468–475.

Crane, P.R., Friis, E.M., and Pedersen, K.R. (1995). The origin and early diversification of angiosperms. Nature *374*, 27–33.

Daniell, T.J., Fordham-Skelton, A.P., Vergani, P., and Edwards, R. (1996). Isolation of a maize cDNA encoding APETALA2-like binding domains by complementation cloning of an L-isoaspartyl methyltransferase-deficient mutant of *Escherichia coli* (accession No. Z47554) (PGR96-013). Plant Physiol. *110*, 1435.

Dörfler, P., and Busslinger, M. (1996). C-terminal activating and inhibitory domains determine the transactivation potential of BSAP (Pax-5), Pax-2 and Pax-8. EMBO J. *15*, 1971–1982.

Drews, G. N., Bowman, J. L., and Meyerowitz, E. M. (1991). Negative regulation of the *Arabidopsis* homeotic gene *AGAMOUS* by the *APETALA2* product. Cell *65*, 991–1002.

Ecker, J.R. (1995). The ethylene signal transduction pathway in plants. Science *268*, 667–675.

Elliott, R.C., Betzner, A.S., Huttner, E., Oakes, M.P., Tucker, W.Q. J., Gerentes, D., Perez, P., Smyth, D.R. (1996). *AINTEGUMENTA*, an *APETALA2*-like gene of *Arabidopsis* with pleiotropic roles in ovule development and floral organ growth. Plant Cell *8*, 155–168.

Evans, M.M.S., Passas, H.J., and Poethig, R.S. (1994). Heterochronic effects of *glossy15* mutations on epidermal cell identity in maize. Development *120*, 1971–1981.

Eyal, Y., Meller, Y., Lev-Yadun, S., and Fluhr, R. (1993). A basic-type PR-1 promoter directs ethylene responsiveness, vascular and abscission zone-specific expression. Plant J. *4*, 225–234.

Flanders, D.J., Weng, S., Petel, F.X., and Cherry, J.M. (1998). AtDB, the *Arabidopsis thaliana* database, and graphical-web-display of progress by the *Arabidopsis* Genome Initiative. Nucleic Acids Res. *26*, 80–84.

Gasser, C. S., and Robinson-Beers, K. (1993). Pistil development. Plant Cell *5*, 1231–1239.

Gerber, H.-P., Seipel, K., Georgiev, O., Hofferer, M., Hug, M., Rusconi, S., and Schaffner, W. (1994). Transcriptional activation modulated by homopolymeric glutamine and proline stretches. Science *263*, 808–811.

Goodman, H. M., Ecker, J. R., and Dean, C. (1995). The genome of *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA. *92*, 10831–10835.

Hammond-Kosack, K.E., and Jones, J.D.G. (1996). Resistance gene-dependent plant defense responses. Plant Cell *8*, 1773–1791.

Hart, C.M., Nagy, F., and Meins, F., Jr. (1993). A 61 bp enhancer element of the tobacco β-1,3-glucanase B gene interacts with one or more regulated nuclear proteins. Plant Mol. Biol. *21*, 121–131.

Henikoff, S., Greene, E.A., Pietrokovski, S., Bork, P., Attwood, T.K., and Hood, L. (1997). Gene families: the taxonomy of protein paralogs and chimeras. Science *278*, 609–614.

Herr, W., and Cleary, M.A. (1995). The POU domain: versatility in transcriptional regulation by a flexible two-in-one DNA-binding domain. Genes Dev. *9*, 1679-1693.

Huala, E., and Sussex, I.M. (1992). *LEAFY* interacts with floral homeotic genes to regulate *Arabidopsis* floral development. Plant Cell *4*, 901–913.

Hunter, T., and Karin, M. (1992). The regulation of transcription by phosphorylation. Cell *70*, 375–387.

Hwang, J.-J., Chambon, P., and Davidson, I. (1993). Characterization of the transcription activation function and the DNA binding domain of transcriptional enhancer factor-1. EMBO J. *12*, 2337–2348.

Irish, V.F., and Sussex, I.M. (1990). Function of the *apetala1-1* gene during *Arabidopsis* floral development. Plant Cell *2*, 741–753.

Ishiguro, S., and Nakamura, K. (1994). Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5′ upstream regions of genes coding for sporamin and β-amylase from sweet potato. Mol. Gen. Genet. *244*, 563–571.

Jaglo-Ottosen, K.R., Gilmour, S.J., Zarka, D.G., Schabenberger, O., and Thomashow, M.F. (1998). *Arabidopsis CBF1* overexpression induces COR genes and enhances freezing tolerance. Science *280*, 104–106.

Jofuku, K.D., den Boer, B.G.W., van Montagu, M., and Okamuro, J.K. (1994). Control of Arabidopsis flower and seed development by the homeotic gene *APETALA2*. Plant Cell *6*, 1211–1225.

Karin, M., and Hunter, T. (1995). Transcriptional control by protein phosphorylation: signal transmission from the cell surface to the nucleus. Current Biol. *5*, 747–757.

Kempin, S.A., Lijegren, S.J., Block, L.M., Rounsley, S.D., Yanofsky, M.F., and Lam, E. (1997). Targeted disruption in *Arabidopsis*. Nature *389*, 802–803.

Kieber, J.J., Rothenberg, M., Roman, G., Feldmann, K.A., and Ecker, J.R. (1993). *CTR1*, a negative regulator of the ethylene response pathway in *Arabidopsis*, encodes a member of the Raf family of protein kinases. Cell *72*, 427–441.

Klucher, K.M., Chow, H., Reiser, L., and Fischer, R.L. (1996). The *AINTEGUMENTA* gene of *Arabidopsis* required for ovule and female gametophyte development is related to the floral homeotic gene *APETALA2*. Plant Cell *8*, 137–153.

Knoester, M., van Loon, L.C., van den Heuvel, J., Hennig, J., Bol, J.F., and Linthorst, H.J.M. (1998). Ethylene-insensitive tobacco lacks nonhost resistance against soil-borne fungi. Proc. Natl. Acad. Sci. USA *95*, 1933–1937.

Komaki, M.K., Okada, K., Nishino, E., and Shimura, Y. (1988). Isolation and characterization of novel mutants of *Arabidopsis thaliana* defective in flower development. Development *104*, 195–203.

Krysan, P.J., Young, J.C., Tax, F., and Sussman, M.R. (1996). Identification of transferred DNA insertions within *Arabidopsis* genes involved in signal transduction and ion transport. Proc. Natl. Acad. Sci. USA *93*, 8145–8150.

Kunst, L., Klenz, J.E., Martinez-Zapater, J., and Haughn, G.W. (1989). *AP2* gene determines the identity of perianth organs in flowers of *Arabidopsis thaliana*. Plant Cell *1*, 1195–1208.

Lawton, K.A., Potter, S.L., Uknes, S., and Ryals, J. (1994). Acquired resistance signal transduction in *Arabidopsis* is ethylene independent. Plant Cell *6*, 581–588.

Léon-Kloosterziel, K.M., Keijzer, C.J., and Koornneef, M. (1994). A seed shape mutant of *Arabidopsis* that is affected in integument development. Plant Cell *6*, 385–392.

Lipsick, J.S. (1996). One billion years of Myb. Oncogene *13*, 223–235.

Martin, C., and Paz-Ares, J. (1997). MYB transcription factors in plants. Trends Genet. *13*, 67–73.

Matsuzaki, Y., Fujisawa, J-i, and Yoshida, M. (1995). Identification of transcriptional activation domain of TREB5, a CREB/ATF family protein that binds to HTLV-1 enhancer. J. Biochem. *117*, 303–308.

McKinney, E.C., Ali, N., Traut, A., Feldmann, K. A., Belostotsky, D. A., McDowell, J. M., and Meagher, R. B. (1995). Sequence-based identification of T-DNA insertion mutations in *Arabidopsis*: actin mutants *act2-1* and *act4-1*. Plant J. *8*, 613–622.

Meyerowitz, E. M. (1994). Structure and organization of the *Arabidopsis thaliana* nuclear genome. In: *Arabidopsis*. E. M. Meyerowitz and C. R. Somerville, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), pp. 21–36.

Mitchell, P.J., and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science *245*, 371–378.

Modrusan, Z., Reiser, L., Feldmann, K.A., Fischer, R.L., and Haughn, G.W. (1994). Homeotic transformation of ovules into carpel-like structures in *Arabidopsis*. Plant Cell *6*, 333–349.

Moose, S.P., and Sisco, P.H. (1994). *Glossy15* controls the epidermal juvenile-to-adult phase transition in maize. Plant Cell *6*, 1343–1355.

Moose, S.P., and Sisco, P.H. (1996). *Glossy15*, an *APETAL2*-like gene from maize that regulates leaf epidermal cell identity. Genes Dev. *10*, 3018–3027.

Ohme-Takagi, M., and Shinshi, H. (1995). Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. Plant Cell *7*, 173–182.

Okamuro, J.K., Caster, B., Villarroel, R., van Montagu, M., and Jofuku, K.D. (1997a). The AP2 domain of *APETALA2* defines a large new family of DNA binding proteins in *Arabidopsis*. Proc. Natl. Acad. Sci. USA *94*, 7076–7081.

Okamuro, J.K., Szeto, W., Lotys-Prass, C., and Jofuku, K.D. (1997b). Photo and hormonal control of meristem identity in the *Arabidopsis* flower mutants *apetala2* and *apetala1*. Plant Cell *9*, 37–47.

Parmentier, Y., Bouchez, D., Fleck, J., and Genschik, P. (1997). The 20S proteasome gene family in *Arabidopsis thaliana*. FEBS Lett. *416*, 281–285.

Pascarella, S., and Argos, P. (1992). Analysis of insertions/deletions in protein structures. J. Mol. Biol. *224*, 461–471.

Pater, S.d., Greco, V., Pham, K., Memelink, J., and Kijne, J. (1996). Characterization of a zinc-dependent transcriptional activator from *Arabidopsis*. Nucleic Acids Res. *24*, 4624–4631.

Penninckx, I.A.M.A., Eggermont, K., Terras, F.R.G., Thomma, B.P. H.J., Desamblanx, G.W., Buchala, A., Metraux, J.P., Manners, J.M., and Broekaert, W.F. (1996). Pathogen-induced systemic activation of a plant defensin gene in *Arabidopsis* follows a salicylic acid-independent pathway. Plant Cell *6*, 2309–2323.

Poethig, R.S. (1990). Phase change and the regulation of shoot morphogenesis in plants. Science *250*, 923–930.

Raikhel, N. (1992). Nuclear targeting in plants. Plant Physiol. *100*, 1627–1632.

Reiser, L., and Fischer, R.L. (1993). The ovule and the embryo sac. Plant Cell *5*, 1291–1301.

Riechmann, J.L., and Meyerowitz, E.M. (1997). MADS domain proteins in plant development. Biol. Chem. *378*, 1079-1101.

Rounsley, S.D., Ditta, G.S., and Yanofsky, M.F. (1995). Diverse roles for MADS box genes in *Arabidopsis* development. Plant Cell *7*, 1259–1269.

Rushton, P.J., Macdonald, H., Huttly, A.K., Lazarus, C.M., and Hooley, R. (1995). Members of a new family of DNA-binding proteins bind to a conserved *cis*-element in the promoters of α-*Amy2* genes. Plant Mol. Biol. *29*, 691–702.

Rushton, P.J., Tovar Torres, J., Parniske, M., Wernert, P., Hahlbrock, K., and Somssich, I.E. (1996). Interaction of elicitor-induced DNA-binding proteins with elicitor response elements in the promoters of parsley PR1 genes. EMBO J. *15*, 5690–5700.

Ryals, J.A., Neuenschwander, U.H., Willits, M.G., Molina, A., Steiner, H.-Y., and Hunt, M.D. (1996). Systemic acquired resistance. Plant Cell *8*, 1809–1819.

Ryan, A.K., and Rosenfeld, M.G. (1997). POU domain family values: flexibility, partnerships, and developmental codes. Genes Dev. *11*, 1207–1225.

Schmidt, R.J., and Ambrose, B.A. (1998). The blooming of grass flower development. Curr. Opin. Plant Biol. *1*, 60–67.

Schneitz, K., Hülskamp, M., Kopczak, S.D., and Pruitt, R.E. (1997). Dissection of sexual organ ontogenesis: a genetic analysis of ovule development in *Arabidopsis thaliana*. Development *124*, 1367–1376.

Schultz, E.A., and Haughn, G.W. (1993). Genetic analyses of the floral initiation process (FLIP) in *Arabidopsis*. Development *119*, 745–765.

Seipel, K., Georgiev, O., and Schaffner, W. (1992). Different activation domains stimulate transcription from remote ('enhancer') and proximal ('promoter') positions. EMBO J. *11*, 4961–4968.

Shannon, S., and Meeks-Wagner, D.R. (1993). Genetic interactions that regulate inflorescence development in *Arabidopsis*. Plant Cell *5*, 639–655.

Shore, P., and Sharrocks, A.D. (1995). The MADS-box family of transcription factors. Eur. J. Biochem. *229*, 1–13.

Stockinger, E.J., Gilmour, S.J., and Thomashow, M.F. (1997). *Arabidopsis thaliana CBF1* encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. Proc. Natl. Acad. Sci. USA *94*, 1035–1040.

Tanaka, M., and Herr, W. (1990). Differential transcriptional activation by Oct-1 and Oct-2: interdependent activation domains induce Oct-2 phosphorylation. Cell *60*, 375–386.

The EU Arabidopsis genome project (1998). Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. Nature *391*, 485–488.

van der Krol, A. R., and Chua, N.-H. (1991). The basic domain of plant B-ZIP proteins facilitates import of a reporter protein into plant nuclei. Plant Cell *3*, 667–675.

Varagona, M.J., Schmidt, R.J., and Raikhel, N.V. (1992). Nuclear localization signal(s) required for nuclear targeting of the maize regulatory protein Opaque-2. Plant Cell *4*, 1213–1227.

Veit, B., Schmidt, R.J., Hake, S., and Yanofsky, M.F. (1993). Maize floral development: new genes and old mutants. Plant Cell *5*, 1205–1215.

Vergani, P., Morandini, P., and Soave, C. (1997). Complementation of a yeast Δ*pkc1* mutant by the *Arabidopsis* protein ANT. FEBS Lett. *400*, 243–246.

Weigel, D. (1995). The APETALA2 domain is related to a novel type of DNA binding domain. Plant Cell *7*, 388–389.

Wilson, K., Long, D., Swinburne, J., and Coupland, G. (1996). A *Dissociation* insertion causes a semidominant mutation that increases expression of *TINY*, an *Arabidopsis* gene related to *APETALA2*. Plant Cell *8*, 659–671.

Winter, B., Braun, T., and Arnold, H.-H. (1992). Co-operativity of functional domains in the muscle-specific transcription factor Myf-5. EMBO J. *11*, 1843–1855.

Wolfe, K., Gouy, M., Yang, Y., Sharp, P., and Li, W. H. (1989). Date of the monocot dicot divergence estimated from chloroplast DNA sequence data. Proc. Natl. Acad. Sci. USA. *86*, 5201–5205.

Zhou, J., Tang, X., and Martin, G.B. (1997). The Pto kinase conferring resistance to tomato bacterial speck disease interacts with proteins that bind a *cis*-element of pathogenesis-related genes. EMBO J. *16*, 3207–3218.

Biol. Chem., Vol. 379, pp. 647 – 654, June 1998 · Copyright © by Walter de Gruyter & Co · Berlin · New York

Review

H1 Histone as a *trans*-Acting Factor Involved in Protecting Genomic DNA from Full Methylation

Giuseppe Zardo[1], Stefania Marenzi[2] and Paola Caiafa[1,3,*]

[1] Department of Biomedical Sciences and Technologies, University of L'Aquila, Via Vetoio, Loc. Coppito, I-67100 L'Aquila, Italy
[2] Department of Biological Sciences "A. Rossi Fanelli" University of Roma, Piazzale A. Moro 5, I-00185 Rome, Italy
[3] C.N.R. Centre for Molecular Biology, Rome, Italy

* Corresponding author

This review aims to explain why H1 histone can be considered as a protein involved in protecting genomic DNA from full methylation. Some of our results indicated that, to explain the multiple roles in which H1 histone seems to be involved, it is important to consider that it is not a unique protein but a family of genetic somatic variants and that every one of them can be dynamically modified by different post-synthetic enzymatic modifications. Our data show that H1 histone plays an inhibitory effect on DNA methylation through its H1e variant and that poly(ADP-ribosyl)ation is a post-synthetic modification involved in this regulatory role. The idea that the poly(ADP-ribosyl)ated isoform of H1e could be present in decondensed chromatin structure, where the housekeeping genes are located, will be discussed.

Key words: DNA methylation / H1 histone variants / Poly(ADP-ribosyl)ation.

Introduction

Our aim is to explain some reasons pointing to H1 histone as a *trans*-acting factor involved in protecting genomic DNA from full methylation and may be also in maintaining the unmethylated state of 'CpG islands'. DNA methylation is a specific post-synthetic modification of DNA that, by transferring methyl groups from S-adenosyl methionine (S-AdoMet) to cytosine (C), converts these residues into 5-methylcytosine (5mC) (Bestor and Ingram,1983), the best substrate being the cytosine located in the CpG dinucleotide (Gruenbaum *et al.*, 1981).

Three processes are involved in defining the specific DNA methylation pattern, namely maintenance methylation, *de novo* methylation and demethylation (Adams, 1995).

A combination of *de novo* methylation and demethylation reactions is actively involved in defining, during the early stages of embryonic development, the final 'correct' methylation pattern in somatic cells (Brandeis *et al.*, 1993). Demethylation of individual genes also occurs during tissue-specific differentiation, probably as a requirement for gene activation (Razin *et al.*, 1986; Brandeis *et al.*, 1993; Jost and Jost, 1994).

Maintenance methylation preserves the tissue-specific methylation pattern by recognizing, in the first minutes after replication (Leonhardt *et al.*, 1992), hemi-methylated sites generated during DNA replication (Razin and Riggs, 1980), and modifying the newly replicated strand. The methylation pattern is distributed in a non-random fashion in genomic DNA (Yisraeli and Szyf, 1984), methylated cytosines being present in bulk DNA (Bloch and Cedar, 1976) while the unmethylated ones are essentially located in portions of the promoter regions of the housekeeping genes which, because of their enrichment in CpG dinucleotides, are termed 'CpG islands' (Bird *et al.*, 1985).

The important role played by DNA methylation in the modulation of gene expression is mainly correlated to the methylation pattern of 'CpG-islands'. The genes associated with these DNA regions are indeed expressed only when the 'CpG-islands' are in the unmethylated state (Bird, 1986, 1987). That CpG dinucleotides are present in the unmethylated state in 'CpG islands' is surprising, since their frequency in these islands is five times higher than in bulk DNA and they exhibit only a moderate intrinsic resistance to *in vitro* methylation (Carotti *et al.*, 1989; Bestor *et al.*, 1992), but remain totally untouched by the action of DNA methyltransferase in spite of the presence of active enzyme in chromatin (Ysraeli and Szyf, 1984).

The mechanisms defining the bimodal pattern of methylation involved both in cell differentiation and gene expression remain unclear. It has been suggested that the density of CpG dinucleotide inside 'CpG-islands' could be *per se* a signal involved in protecting the unmethylated state of these DNA regions (Frank *et al.*, 1991). Further experiments have suggested that there are some sequence motifs that are intrinsically protected against *de novo* methylation (Szyf *et al.*, 1990; Christman *et al.*, 1995) and/or that there are some *cis*-acting 'centers of methylation' capable of preventing methylation of flanking DNA sequences (Szyf *et al.*, 1990; Szyf, 1991; Mummaneni *et al.*, 1993; Brandeis *et al.*, 1994; Hasse and Schultz, 1994; Macleod *et al.*, 1994; Magewu and Jones, 1994; Mummaneni *et al.*, 1995). Some *trans*-acting proteins able to bind methylated DNA have been identified (Huang *et al.*,

1984; Zhang *et al.*, 1986; Khan *et al.*, 1988; Meehan *et al.*, 1989; Boyes and Bird, 1991; Pawlak *et al.*, 1991; Jost and Hofsteenge, 1992; Lewis *et al.*, 1992; Meehan *et al.*, 1992; Asiedu *et al.*, 1994; Bruhat and Jost, 1995; Zhang *et al.*, 1995; Nan *et al.*, 1997). Although these proteins may play an important role in mediating the methylation-dependent repression of genes, their involvement in creating or maintaining the under-methylated state of CpG islands has not been demonstrated so far.

Our aim is to describe a protein that, in view of its characteristics, could be a *trans*-acting factor directly involved in protecting genomic DNA from full methylation. This protein could be H1 histone, i.e. a protein particularly abundant in chromatin where it seems to be involved in many important structural and functional roles (van Holde, 1988; Zlatanova and van Holde, 1992). To explain how a single protein could play multiple roles, it is important to remember that H1 histone is composed of a family of somatic variants and that every one of them can be dynamically modified by a number of different post-synthetic enzymatic reactions.

This microheterogeneity of H1 histone therefore allows different possible interactions with DNA or with other proteins.

Inhibitory Effect of H1 Histone on *In Vitro* DNA Methylation

Our previous experiments (Caiafa *et al.*, 1991) have shown that the ability of total histones to affect *in vitro* enzymatic DNA methylation, catalyzed by human placenta DNA methyltransferase, was essentially due to H1 histone. This histone was the only component able to exert a severe (90%) inhibition of dsDNA methylation when present at the 'physiological' range, 0.3:1.0 (w/w) histone:DNA ratio (Santoro *et al.*, 1993).

Experiments were carried out in order to assess whether the observed hypomethylation of linker DNA sequences (Razin and Cedar, 1977; Solage and Cedar, 1978; Adams *et al.*, 1984; Caiafa *et al.*, 1986) reflected an intrinsic deficiency in CpG dinucleotides in the linker, or whether the well-documented association between linker DNA and H1 histone could be responsible for a local inhibition of the enzymatic DNA methylation process (D'Erme *et al.*, 1993). The results obtained have shown that the lower level of DNA methylation in linker regions compared to that in 'core' particles was not due to an intrinsic CpG deficiency in linker DNA, which was, in H1-depleted oligonucleosomes, susceptible to extensive *in vitro* methylation. The hypomethylation of linker DNA could rather be ascribed to the inhibition exerted by H1 histone on the process of enzymatic DNA methylation (Caiafa *et al.*, 1991), which selectively occurs in these DNA regions because of their preferential association with this protein.

The ability and specificity of H1 histone to inhibit CpG methylation in linker DNA were assayed by re-adding purified H1 to H1-depleted oligonucleosomes or to DNA purified from them. H1-depletion doubled the methyl-accepting ability of oligonucleosomes, with a further 50% increase when the remaining proteins were also removed. Re-addition of H1 to the H1-depleted oligonucleosomes or to the purified oligonucleosomal DNA, in a protein-to-DNA (w/w) ratio of 0.3, reduced the incorporation of labeled methyl groups to the same level as in native oligonucleosomal particles (Table 1). The inhibition was paralleled by condensation caused by addition of H1 to H1-depleted oligonucleosomes. Both phenomena were apparently specific to H1 histone, since they could not be obtained by addition of other histones or serum albumin up to a 1:1 protein/DNA (w/w) ratio.

The hypothesis of competition between the enzyme and H1 histone for common DNA sites was investigated by adding increasing amounts of purified DNA methyltransferase to *Micrococcus luteus* ds-DNA in the presence of a constant amount of H1 histone, the H1/DNA ratio being fixed to its 'physiological' value of 0.3. DNA methylation was found to be independent of the H1-to-enzyme ratio, disproving the competition hypothesis (Santoro *et al.*, 1993).

It is also worth recalling that in the same set of experiments, the methyl-accepting ability of native oligonucleo-

Table 1 Inhibition by H1 Histone of the Methyl-Accepting Ability of Native Oligonucleosomes, of H1-Depleted Oligonucleosomes and of Oligonucleosomal Purified DNA.

	Histone proteins added			
	None	H1 (0.3 mg/mg DNA)	'core' histones (1 mg/mg DNA)	H2a (1 mg/mg DNA)
Number of experiments:	n = 6	n = 6	n = 3	n = 3
Native oligonucleosomes	48.4 ± 0.7	–	–	–
H1-depleted oligonucleosomes	100.0	58.0 ± 1.3	101.4 ± 3.5	102.6 ± 2.7
Purified DNA from oligonucleosomes	155.0 ± 2.1	41.6 ± 0.8	153.8 ± 5.8	–

The incorporation of labeled methyl groups in the DNA of H1-depleted oligonucleosomes is made equal to 100 and all the other results obtained in a similar set of experiments are referred to this value. Reprinted from Biochim. Biophys. Acta *1173*, 209–216, (1993) D'Erme *et al.*, Inhibition of CpG methylation in linker DNA by H1 histone, with kind permission of Elsevier Science Publishers – NL Sara Burgerhartstraat 25, 1055 KV Amsterdam, The Netherlands.

somes was far from negligible, although it underwent a twofold increase upon H1 histone depletion, with a further 50% increment if all other proteins were also removed.

These results led to the formulation of two hypotheses: the presence of some particular histone variant(s) more or less capable of inhibiting enzymatic DNA methylation and/or the presence of DNA regions escaping the negative control of H1 histone.

H1e Variant: Its Specificity in Inhibiting the *In Vitro* DNA Methylation

As mentioned above, the H1 histone family includes variants that are present in all somatic cells. In addition, other variants are present only in some species and cell types (e.g. H5, H1°, H1t, H1s). The hypothesis that some particular H1 somatic variant(s) could be specifically involved in the in vitro inhibition of enzymatic DNA methylation stems from the fact that this histone is composed of a family of five variants that differ in their primary structures (Kinkade and Cole, 1966; Roll and Cole, 1971). Although a new nomenclature system, based on amino acid sequence data, has recently been proposed by Parseghian *et al.* (1994), we shall still refer to the 'old' but more familiar nomenclature, where these variants are termed, according to Cole (1987), H1a, H1b, H1c, H1d and H1e. The overall variation in molecular mass is approx. 1.0–1.4 kDa and in this range H1b, H1d and H1e are larger than H1a and H1c. All variants have a three domain structure, with a highly conserved central globular domain (98% identity in 80 aa sequence). The differences between the variants are located in the N-terminal and C-terminal tails, which consist of about 40 and 100 amino acids respectively (Cole, 1987). Since the number and relative amounts of these variants differ in various tissues and species throughout the development stages of the organism and in neoplastic systems (Liao and Cole, 1981a,b; Pehrson and Cole, 1982; Lennox and Cohen, 1983; Huang and Cole, 1984; Lennox, 1984; Cole, 1987; Davie and Delcuve, 1991; Baubichon-Cortay *et al.*, 1992; Giancotti *et al.*, 1993; Schulze *et al.*, 1993; De Lucia *et al.*, 1994), it is reasonable to assume that they may play different roles in chromatin organization, and be distributed in a non-random manner in functionally distinct chromatin regions.

Experiments were therefore performed in order to verify whether or not these variants differ in their ability to exert negative control on DNA methylation.

To this end H1 histone somatic variants were purified by using reverse phase HPLC (Santoro *et al.*, 1995; D'Erme *et al.*, 1996; Zardo *et al.*, 1996) and used in an *in vitro* DNA methyltransferase activity assay.

Of all H1 variants tested, only H1e was effective in causing a marked inhibition of DNA methylation, at the H1:DNA 'physiological' ratio (Figure 1A).

Moreover H1e is also the only H1 variant able to bind CpG island-like DNA sequences, as evidenced by gel retardation assays performed on various synthetic oligonucleotides that varied in terms of sequence and of relative abundance in unmethylated CpGs with respect to NpGs (i.e. to all dinucleotide sequences having G as second base). As representative of genomic DNA we used 145 bp DNA fragments prepared by digestion of human placenta chromatin with *Staphylococcus aureus* nuclease and deproteinization.

Among the H1 somatic variants, H1a was able to bind the 145 bp genomic DNA fragments but was unable to bind 44 bp ds-oligonucleotides containing two or more CpG dinucleotides; the other variants were capable of binding sequences containing up to three CpGs (Santoro *et al.*, 1995). The only variant able to bind CpG rich oligonucleotides was H1e (Zardo *et al.*, 1996; Figure 1B).

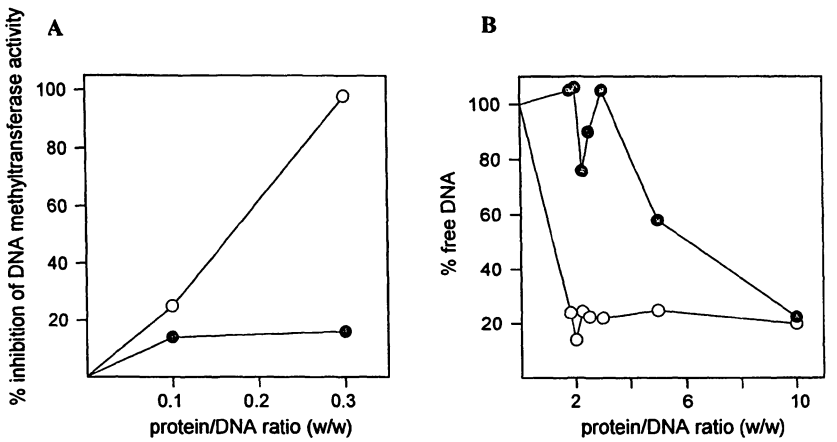

Fig. 1 Dependence of DNA Methylation Inhibition on H1e Variant and Its Affinity for a 44 bp Unmethylated ds-Oligonucleotides with CpG Rich Sequence.

(A) Inhibition of DNA methyltransferase activity by H1e (open circles) or H1c (closed circles), at different protein-to-DNA ratios. (B) Binding of H1e (open circles) and H1c (closed circles) to a 44 bp synthetic 6-CpG duplex oligonucleotide with the cytosines in the CpG moieties in unmethylated form. The binding was evaluated by gel retardation after incubation of the H1e or the H1c variants with the appropriate oligonucleotide, the relative amount of free DNA being measured by densitometric scanning of the autoradiograms. Reprinted from Biochem. Biophys. Res. Comm. *20*, 102 – 107 (1996) Zardo *et al.*, Inhibitory effect of H1e histone somatic variant on *in vitro* DNA methylation process, with kind permission of Academic Press, Inc.

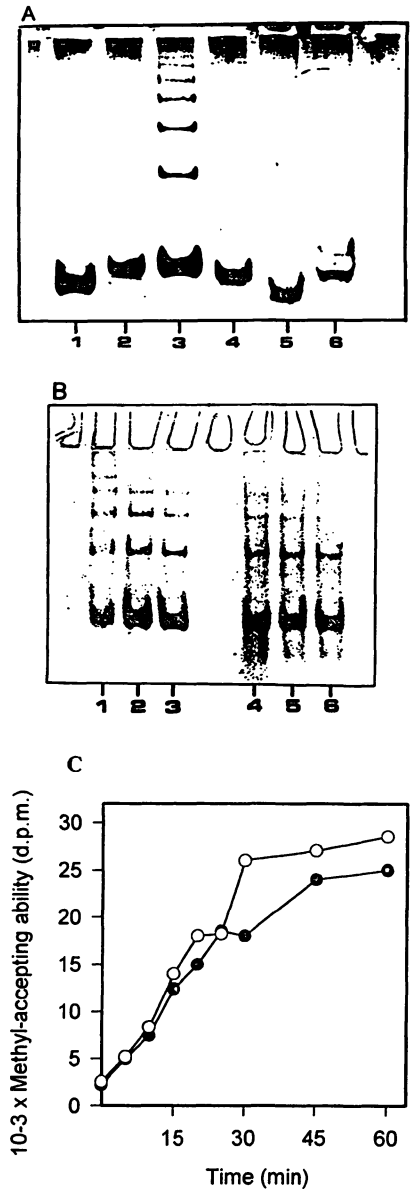

Fig. 2 Analysis of H1-H1 Polymer Formation and Association with linker DNA.

(A) Use of cross-linking analysis to investigate the ability of each H1 histone variant to form H1-H1 polymers when bound to DNA. SDS-PAGE patterns of H1 histone variants, at 30% (w/w) H1:DNA ratio, incubated with 1.2 kb oligonucleosomal DNA in 40 mM NaCl for 1 hour at room temperature and then treated with dithiobis-(succinimidyl)propionate (DSP 0.2 mg/ml) for 20 min: H1a, H1b, H1e and H1c (lanes 1–4). In lanes 5 and 6, untreated histone H1 and histone H1 treated with DSP were run as controls in the absence of DNA. (B) Use of cross-linking analysis to investigate the effect of the 'partial' poly(ADP-ribosyl)ation of H1e variant, on the formation of H1-H1 polymers. SDS-PAGE patterns of the products of cross-linking of H1e at 30%, 20% and 10% of H1e/DNA (lanes 1–3); lanes 4–6, 30%, 20% and 10% (w/w) of partially poly(ADP-ribosyl)ated H1e/DNA. (C) Methyl-accepting ability as an assay to study the association of H1 histone with linker DNA in native (closed circles) vs poly(ADP-ribosyl)ated nuclei (open circles). For experimental details, see Caiafa *et al.* (1991). Reprinted from Biochem. J. *316*, 475–480, (1996) D'Erme *et al.*, Co-operative interactions of oligonucleosomal DNA with the H1e histone variant and its poly(ADP-ribosyl)ated isoform, with kind permission of Portland Press.

Thus, two important characteristics of H1e emerge: it is the only variant that suppresses enzymatic DNA methylation and that can bind to CpG-rich sequences. Moreover, it is the only variant that interacts with DNA in such a way as to allow the formation of H1-H1 polymers upon chemical cross-linking (Figure 2A).

Although the structural interpretation of H1 polymer formation on DNA and/or chromatin is still unclear, the fact that H1e is capable of such interaction while the other H1 variants are not, singles this variant out for further in-depth investigations.

H1 Histone Post-Synthetic Modifications

If histone H1e is involved in inhibition of DNA methylation, the question arises as to whether and how its post-synthetic modifications affect this property.

The post-synthetic modifications in which H1 histone seems to be most involved are phosphorylation and poly(ADP-ribosyl)ation (Wu *et al.*, 1986; van Holde, 1988; Davie, 1995). In nuclei, the modifications are catalyzed by specific enzymes, and are reversed by other enzymatic reactions that dynamically remove the covalently bound groups.

There is a very strong tendency for the modification sites to be concentrated in the tails of H1 histone molecules; the modifications are quite specific not only in terms of variant(s) to be modified but also with respect to the amino acid residue(s) involved. As far as phosphorylation is concerned, the amino acids involved are serine and threonine located in both tails. For poly(ADP-ribosyl)ation, the amino acid residue involved is glutamic acid or the C-terminal lysine residue. The position of modified residue(s) (Parseghian *et al.*, 1994) and/or the amount of polymer (D'Erme *et al.*, 1996) can vary with respect to the variants.

Our attention was focused on poly(ADP-ribosyl)ation, since we noted that DNA-bound H1e, enriched in its poly(ADP-ribosyl)ated form, loses its ability to form H1-H1 polymers when compared to native H1e (Figure 2B). It is also clear that the ribosylated form of H1e remains associated with linker DNA in chromatin (D'Erme *et al.*, 1996; Figure 2C).

These data suggest that even if poly(ADP-ribosyl)ation decreases the H1e-H1e interactions, that may be essential for the formation of the higher levels of chromatin structure, the poly(ADP-ribosyl)ated isoform of H1e could be present in the decondensed chromatin structure where the housekeeping genes are located.

Our attention was also focused on poly(ADP-ribosyl)-ation, since it is known that the demethylation process utilizes an excision-repair mechanism to remove 5-methylcytosine and that the poly(ADP-ribosyl)ation process plays a role in the repair mechanism (de Murcia *et al.*, 1994, 1995) through the poly(ADP-ribosyl)ation of H1 histone (Boulikas, 1989; Realini and Althaus, 1992; Malanga and Althaus, 1994). The H1 histone in its poly(ADP-ribosyl)-ated isoform could therefore, following the demethylation process, remain bound to demethylated regions and regu-

late the *de novo* re-methylation process, keeping the CpG islands in an unmethylated state.

In vivo experiments carried out on L929 mouse fibroblasts preincubated for 24 h with or without 3-aminobenzamide, a well-known inhibitor of poly(ADP-ribose) polymerase, (Griffin *et al.*, 1995) suggested the existence of a negative correlation between poly(ADP-ribosyl)ation and DNA methylation processes. In nuclei obtained upon incubation of cells with 3-aminobenzamide we observed a consistent increase in DNA methylation by endogenous DNA methyltransferase. Likewise, when the methyl-accepting ability assay was performed on DNA purified from the pre-treated nuclei, the subsequent methylation by exogenous enzymes was consequently severely reduced, indicating saturation of possible methylatable sites *in vivo* in conditions of decreased poly(ADP-ribosyl)ation (Zardo *et al.*, 1997). Poly(ADP-ribosyl)ation appears, therefore, to be a mechanism involved in protecting CpG dinucleotides from full methylation.

Further *in vitro* experiments were performed with the aim of examining the possible connection between DNA methylation and poly(ADP-ribosyl)ation of H1 histone. In order to do this, the poly(ADP-ribosyl)ated and poly(ADP-ribose)-free H1 histone isoforms were purified and their effect on *in vitro* DNA methylation, catalyzed by placental DNA methyltransferase, was investigated (Zardo *et al.*, 1997).

These experiments showed that the poly(ADP-ribose)-free form of H1 histone, obtained from mouse fibroblasts preincubated with 3-aminobenzamide, failed to inhibit DNA methylation when added up to a protein:DNA ratio of 0.25 (w/w). The poly(ADP-ribosyl)ated isoform of H1 histone, purified by affinity chromatography on a phenylboronate column, was, in contrast, highly inhibitory under the same conditions (Figure 3A, B). Another interesting result was obtained by studying the direct effect of ADP-ribose polymers on *in vitro* DNA methylation. These protein-free polymers caused an appreciable inhibition of *in vitro* methylation of dsDNA but not of ssDNA. The extent of this inhibition was directly dependent on the size of the

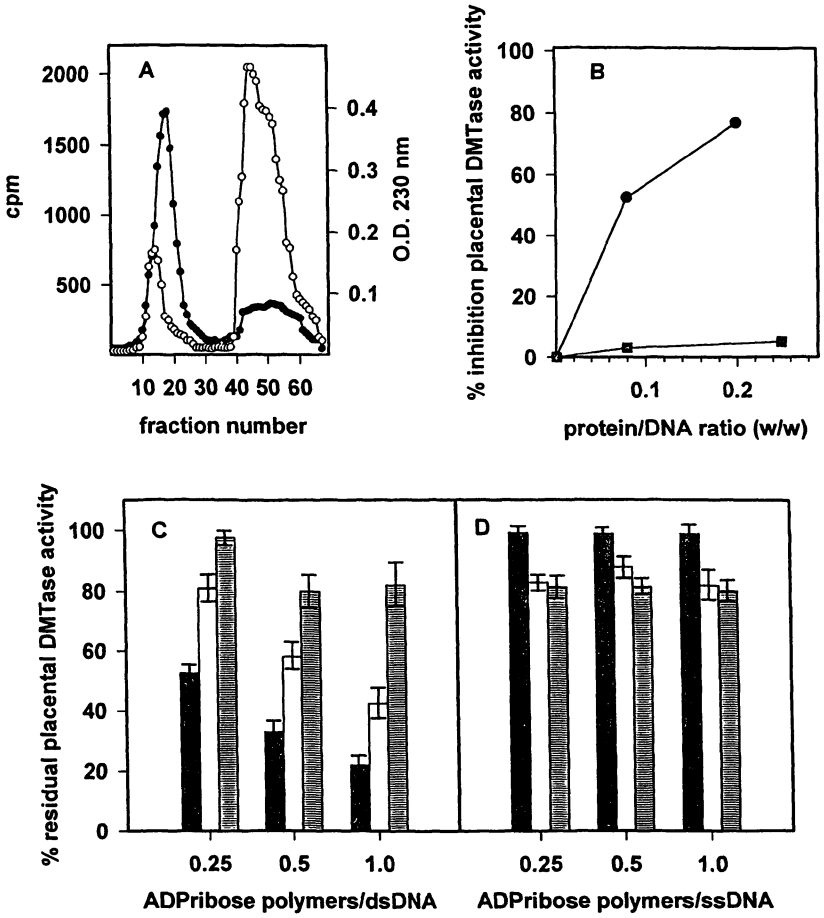

Fig. 3 Effect of Poly(ADP-ribosyl)ation and ADP-Ribose Polymers on DNA Methylation.
(A) Purification of poly(ADP-ribosyl)ated H1 histone isoform on an aminophenylboronate column chromatography, monitoring either the absorbance at 230 nm (closed circles), or the radioactivity (open circles). (B) Comparison between poly(ADP-ribose)-free H1 histone (closed squares) and the purified poly(ADP-ribosyl)ated isoform (closed circles) for their inhibitory effect on *in vitro* DNA methylation. (C) Effect of ADP-ribose polymers of different length (black bars, $n > 40$; white bars; $6 < n < 40$; horizontally striped bars $n < 20$) on *in vitro* DNA methylation. Control assay, taken as 100%, was performed in the absence of polymers. Different polymer/DNA ratios, ranging from 0.25 to 1.00, are indicated in the abscissa. Reprinted from Biochemistry *36*, 7937–7943, (1997) Zardo *et al.*, Does poly(ADP-ribosyl)ation regulate the DNA methylation pattern?, with kind permission of the American Chemical Society.

polymers (Figure 3C). It is worth noting that, since a high ADP-ribose polymer/DNA ratio did not affect methylation of ssDNA (Figure 3D), the polymers can hardly be considered as directly interacting with DNA methyltransferase.

In the close relationship existing between poly(ADP-ribosyl)ation and DNA methylation processes, the poly-(ADP-ribosyl)ation of H1 histone appears therefore to play a key role. Since the association of H1 histone with ADP-ribose polymers can be either covalent (Naegeli and Althaus, 1991) or non-covalent (Panzeter *et al.*, 1992), further investigations are needed to ascertain whether the latter interactions are also effective in maintaining CpG dinucleotides in their unmethylated state. By gel retardation assay we had also shown that poly(ADP-ribosyl)ated H1 histone has a high capacity of linking CpG-rich ds-oligonucleotide, so that it is possible to suppose that it has a preferential location on genomic DNA in regions rich in these nucleotides.

In conclusion, it is our hypothesis that, after DNA packaging into nucleosomes, the access to DNA of a moving methyltransferase can then be limited by the presence of poly(ADP-ribosyl)ated H1 histone. Since, on the other hand, only relatively short poly-ADPribose chain(s) are bound to H1 histone (D'Erme *et al.*, 1996), it is unlikely that they can be responsible by themselves for the intense inhibitory effect exerted on the methylation of ds DNA.

Our hypothesis is that *in vivo* long and branched polymers, linked in a non-covalent way to the histone, could take part in affording protection of the unmethylated state of the CpG-rich DNA regions (Zardo *et al.*, 1997).

Acknowledgement

This work was supported by the Italian Ministry of University and Scientific and Technological Research (40% Progetti di Interesse Nazionale, 60% Ricerca Scientifica dell' Università dell'Aquila) and by Fondazione "Istituto Pasteur-Fondazione Cenci Bolognetti".

References

Adams, R.L.P. (1995). Eukaryotic DNA methyltransferases – structure and function. BioEssays *17*, 139–145.

Adams, R.P.L., David, T., Fulton, J., Kirk, D., Qureshi, M., and Burdon, R.H. (1984). Eukaryotic DNA methylase properties and action on native DNA and chromatin. Curr. Top. Microbiol. Immunol. *108*, 143–156.

Asiedu, C.K., Scotto, L., Assoian, R.K., and Ehrlich, M. (1994). Binding of AP-1/CREB proteins and of MDBP to contiguous sites downstream of the human TGF-1 gene. Biochim. Biophys. Acta *1219*, 55–63.

Baubichon-Cortay, H., Mallet, L., Denoroy, L., and Roux, B. (1992). Histone H1a subtype presents structural differences compared to other histone H1 subtypes. Evidence for a specific motif in the C-terminal domain. Biochim. Biophys. Acta *1122*, 167–171.

Bestor, T.H., and Ingram, V.M. (1983). Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity and mode of interaction with DNA. Proc. Natl. Acad. Sci. USA *80*, 5559–5563.

Bestor, T.H., Gundersen, G., Kolsto, A.B., and Pryd, H. (1992). CpG islands in mammals gene promoter inherently resistant to *de novo* methylation. Genet. Anal. Tech. Appl. *9*, 48–53.

Bird, A.P., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated CpG-rich DNA. Cell *40*, 91–99.

Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. Nature *321*, 209–213.

Bird, A.P. (1987). CpG islands as gene markers in the vertebrate nucleus. Trends Genet. *3*, 342–347.

Bloch, S., and Cedar, H. (1976). Methylation of chromatin DNA. Nucl. Acids Res. *3*, 1507–1519.

Boulikas, T. (1989). DNA strand breaks alter histone ADP-ribosylation. Proc. Natl. Acad. Sci. USA *86*, 3499–3503.

Boyes, J., and Bird, A.P. (1991). DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell *64*, 1123–1134.

Brandeis, M., Ariel, M., and Cedar, H. (1993). Dynamics of DNA methylation during development. Bioessays *15*, 709–713.

Brandeis, M., Frank, D., Keshet, I, Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A., and Cedar, H. (1994). Sp1 elements protect a CpG island from *de novo* methylation. Nature *371*, 435–438.

Bruhat, A., and Jost, J.P. (1995). In vivo estradiol-dependent dephosphorilation of the repressor MDBP-2-H1 correlates with the loss of in vitro preferential binding to methylated DNA. Proc. Natl. Acad. Sci. USA *92*, 3678–3682.

Caiafa, P., Attinà, M., Cacace, F., Tomassetti, A., and Strom, R. (1986). 5-methylcitosine levels in nucleosome subpopulations differently involved in gene expression. Biochim. Biophys. Acta *867*, 195–200.

Caiafa, P., Reale, A., Allegra, P., Rispoli, M., D'Erme, M., and Strom, R. (1991). Histones and DNA methylation in mammalian chromatin. I. Differential inhibition of in vitro methylation by histone H1. Biochim. Biophys. Acta *1090*, 38–42.

Carotti, D., Palitti, F., Lavia, P., and Strom, R. (1989). In vitro methylation of CpG islands. Nucl. Acids Res. *17*, 9219–9229.

Christman, J.K., Sheikhnejad, J., Marasco, C.J., and Sufrin, J.R. (1995). 5-methyl-2′-deoxycytidine in single-stranded DNA can act in cis to signal *de novo* DNA methylation. Proc. Natl. Acad. Sci. USA *92*, 7347–7351.

Cole, R.D. (1987). Microheterogeneity in H1 histones and its consequences. Int. Pept. Protein Res. *30*, 433–449.

Davie, J.R. (1995). The nuclear matrix and the regulation of chromatin organization and function. In: Structural and Functional Organization of the Nuclear Matrix, R. Berenzey and K.W. Jean, eds. (New York: Academic Press, Inc.), pp. 191–249.

Davie, J.R., and Delcuve, J.P. (1991). Characterization and chromatin distribution of H1 histones and high-mobility-group non-histone chromosomal proteins of trout liver and hepato cellular carcinoma. Biochem. J. *280*, 491–497.

De Lucia, F., Faraone-Mennella, M.R., D'Erme, M., Quesada, P., Caiafa, P., and Farina, B. (1994). Histone-induced condensation of rat testis chromatin: testis-specific H1t versus somatic H1 variants. Biochem. Biophys. Res. Commun. *198*, 32–39.

de Murcia, G., and Ménissier de Murcia, J. (1994). Poly(ADP-ribose) polymerase: a molecular nick sensor. Trends Biochem. Sci. *19*, 172–176.

de Murcia, G., Jacobson, M., and Shall, S. (1995). Regulation by ADP-ribosylation. Trends Cell Biol. *5*, 78–81.

D'Erme, M., Santoro, R., Allegra, P., Reale, A., Marenzi, S., Strom, R., and Caiafa, P. (1993). Inhibition of CpG methylation in linker DNA by H1 histone. Biochim. Biophys. Acta *1173*, 209–216.

D'Erme, M., Zardo, G., Reale, A., and Caiafa, P. (1996). Co-operative interactions of oligonucleosomal DNA with the H1e histone variant and its poly(ADP-ribosyl)ated isoform. Biochem. J. *316*, 475–480.

Frank, D., Keshet, I., Shani, M., Levine, A., Razin, A., and Cedar, H. (1991). Demethylation of CpG islands in embryonic cells. Nature *351*, 239–241.

Giancotti, V., Bandiera, A., Ciani, L., Santoro, D., Crane Robinson, C., Goodwin, G.H., Biocchi, M., Dolcetti, R., and Casetta, B. (1993). High-mobility-group (HMG) proteins and histone H1 subtypes expression in normal and tumor tissues of mouse. Eur. J. Biochem. *213*, 825–832.

Griffin, R.J., Curtin, N.J., Golding, B.T., Durkacz, B.W., and Calvert, A.H. (1995). The role of inhibitors of poly(ADP-ribose) polymerase as resistance-modifying agents in cancer therapy. Biochimie *77*, 408–422.

Gruenbaum, Y., Stein, R., Cedar, H., and Razin, A. (1981). Methylation of CpG sequences in eukaryotic DNA. FEBS Lett. *124*, 67–71.

Hasse, A., and Schulz, W.A. (1994). Enhancement of reporter gene *de novo* methylation by DNA fragments from the alpha-fetoprotein control region. J. Biol. Chem. *269*, 1821–1826.

Huang, H.C., and Cole, R.D. (1984). The distribution of H1 histone is non uniform in chromatin and correlates with different degrees of condensation. J. Biol. Chem. *259*, 14237–14242.

Huang, L., Wang, R., Gama-Sosa, M.A., Shenoy, S., and Ehrlich, M. (1984). A protein from human placental nuclei binds preferentially to 5-methylcytosine-rich DNA. Nature *308*, 293–295.

Jost, J.P., and Hofsteenge, J. (1992). The repressor MDBP-2 is a member of the histone H1 family that binds preferentially *in vitro* and *in vivo* to methylated nonspecific DNA sequences. Proc. Natl. Acad. Sci. USA *89*, 9499–9503.

Jost, J.P., and Jost, Y.C. (1994). Transient DNA demethylation in differentiating mouse myoblasts correlates with higher activity of 5-methyldeoxycytidine excision repair. J. Biol. Chem. *269*, 10040–10043.

Khan, R., Zhang, X.Y., Supakar, P.C., Ehrlich, K.C., and Ehrlich, M. (1988). Human methylated DNA-binding protein. J. Biol. Chem. *263*, 14374–14383.

Kinkade, J.M., Jr., and Cole, R.D. (1966). The resolution of four lysine-rich histones derived from calf thymus. J. Biol. Chem. *241*, 5790–5797.

Lennox, R.W. (1984). H1 differences in evolutionary stability among mammalian subtypes. J. Biol. Chem. *259*, 669–672.

Lennox, R.W., and Cohen, L.H. (1983). The histone H1 complements of dividing and nondividing cells of the mouse. J. Biol. Chem. *258*, 262–268.

Leonhardt, H., Page, A.W., Weier, H.U., and Bestor, T.H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell *71*, 865–873.

Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jepsen, P., Klein, F., and Bird, A.P. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell *69*, 905–914.

Liao, L.W.L., and Cole, R.D. (1981a). Condensation of dinucleosomes by individual subfractions of H1 histone. J. Biol. Chem. *256*, 10124–10128.

Liao, L.W.L., and Cole, R.D. (1981b). Differences among H1 histone subfractions in binding to linear and superhelical DNA. J. Biol. Chem. *256*, 11145–11150.

Macleod, D., Charlton, J., Mullins, J., and Bird, A.P. (1994). Sp1 sites in the mouse *aprt* gene promoter are required to prevent methylation of the CpG islands. Genes and Dev. *8*, 2282–2292.

Magewu, A.N., and Jones, P.A. (1994). Ubiquitous and tenacious methylation of the CpG sites in codon 248 of the p53 gene may explain its frequent appearance as a mutational hot spot in human cancer. Mol. Cell. Biol. *14*, 4225–4232.

Malanga, M., and Althaus, F.R. (1994). Poly(ADP-ribose) molecules formed during DNA repair *in vivo*. J. Biol. Chem. *269*, 17691–17969.

Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L., and Bird, A.P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell *58*, 499–507.

Meehan, R.R., Lewis, J.D., and Bird, A.P. (1992). Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. Nucl. Acids Res. *20*, 5085–5092.

Mummaneni, P., Bishop, P.L., and Turker, M.S. (1993). A cis-acting element accounts for a conserved methylation pattern upstream of the mouse adenine phosphoribosyltransferase gene. J. Biol. Chem. *268*, 552–558.

Mummaneni, P., Walker, K.A., Bishop, P.L., and Turker, M.S. (1995). Epigenetic gene inactivation induced by a *cis*-acting methylation centre. J. Biol. Chem. *270*, 788–792.

Naegeli, H., and Althaus, F.R. (1991). Regulation of poly(ADP-ribose) polymerase. J. Biol. Chem. *266*, 10596–10601.

Nan, X., Campoy, F.J., and Bird, A.P. (1997). MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. Cell *88*, 471–481.

Panzeter, P.L., Realini, C.A., and Althaus, F.R. (1992). Noncovalent interactions of poly(adenosine diphosphate ribose) with histones. Biochemistry *31*, 1379–1385.

Parseghian, M.H., Henschen, A.H., Krieglstein, K.G., and Hamkalo, B.A. (1994). A proposal for a coherent mammalian histone H1 nomenclature correlated with amino acid sequences. Protein Science *3*, 575–587.

Pawlak, A., Bryans, M., and Jost, J.P. (1991). An avian 40 kDa nucleoprotein binds preferentially to a promoter sequence containing one single pair of methylated CpG. Nucl. Acids Res. *19*, 1029–1034.

Pehrson, J.R., and Cole, R.D. (1982). Histone H1 subfractions and H1° turnover at different rates in nondividing cells. Biochemistry *21*, 456–460.

Rall, S.C., and Cole, R.D. (1971). Amino acid sequence variability of the amino-terminal regions of Lysine-rich histones. J. Biol. Chem. *246*, 7175–7190.

Razin, A., and Cedar, H. (1977). Distribution of 5-methylcitosine in chromatin. Proc. Natl. Acad. Sci. USA *74*, 2725–2728.

Razin, A., and Riggs, A.D. (1980). DNA methylation and gene function. Science *210*, 604–610.

Razin, A., Szyf, M., Kafri, T., Roll, M., Giloh, H., Scarpa, S., Carotti, D., and Cantoni, G.L. (1986). Replacement of 5-methylcytosine by citosine: a possible mechanism for transient DNA demethylation during differentiation. Proc. Natl. Acad. Sci. USA *83*, 2827–2831.

Realini, C.A., and Althaus, F.R. (1992). Histone shuttling by poly(ADP-ribosyl)ation. J. Biol. Chem. *267*, 18858–18865.

Santoro, R., D'Erme, M., Reale, A., Strom, R., and Caiafa, P. (1993). Effect of H1 histone isoforms on the methylation of single- or double stranded DNA. Biochem. Biophys. Res. Comm. *190*, 86–91.

Santoro, R., D'Erme, M., Mastrantonio, S., Reale, A., Marenzi, S., Saluz, H.P., Strom, R., and Caiafa, P. (1995). Binding of histone H1e-c variants to CpG-rich DNA correlates with the inhibitory effect on enzymic DNA methylation. Biochem. J. *305*, 739–744.

Schulze, E., Trieschemann, L., Schulze, B., Schmidt, E.R., Pitzel, S., Zechel, K., and Grossbach, U. (1993). Structural and functional differences between histone H1 sequence variants with differential intranuclear distribution. Proc. Natl. Acad. Sci. USA *90*, 2481–2485.

Solage, A., and Cedar, H. (1978). Organization of 5-methylcytosine in chromosomal DNA. Biochemistry *17*, 2934–2938.

Szyf, M. (1991). DNA methylation patterns: an additional level of information? Biochem. Cell Biol. *69*, 764–767.

Szyf, M., Tanigawa, G., and McCarthy, P.L., Jr. (1990). A DNA signal from the Thy-1 gene defines *de novo* methylation patterns in embryonic stem cells. Mol. Cell. Biol. *10*, 4396–4400.

van Holde, K. (1988). Chromatin. (New York: Springer-Verlag).

Wu, R.S., Panusz, H.T., Hatch, C.L., and Bonner, W.M. (1986). Histones and their modification. CRC Crit. Rev. Biochem. *20*, 201–263.

Yisraeli, J., and Szyf, M. (1984). Gene methylation patterns and expression. In: DNA Methylation and Its Biological Significance, A. Razin *et al.*, eds. (New York: Springer-Verlag), pp. 353–378.

Zardo, G., Santoro, R., D'Erme, M., Reale, A., Guidobaldi, L., Caiafa, P., and Strom, R. (1996). Specific inhibitory effect of H1e histone somatic variant on *in vitro* DNA-methylation process. Biochem. Biophys. Res. Comm. *220*, 102–107.

Zardo, G., D'Erme, M., Reale, A., Strom, R., Perilli, M., and Caiafa, P. (1997). Does poly(ADP-ribosyl)ation regulate the DNA methylation pattern? Biochemistry *36*, 7937–7943.

Zhang, X.Y., Ehrlich, K.C., Wang, R.Y.H., and Ehrlich, M. (1986). Effect of site-specific DNA methylation and mutagenesis on recognition by methylated DNA-binding protein from human placenta. Nucl. Acids Res. *14*, 8387–8397.

Zhang, X.Y., Ni, Y.-S., Saifudeen, Z., Asiedu, C.K., Supakar, P.C., and Ehrlich, M. (1995). Increasing binding of a transcriptor factor immediately downstream of the cap site of a cytomegalovirus gene represses expression. Nucl. Acids Res. *23*, 3026–3033.

Zlatanova, J., and van Holde, K. (1992). Histone H1 and transcription: still an enigma? J. Cell. Sci. *103*, 889–895.