Chapter One

General Introduction

1.1 The transition to flowering

Timing of the transition to flowering is a critical factor in determining the reproductive success of plants, which use both external and internal signals to determine the time of flowering. Environmental signals such as photoperiod, temperature and light quality have a significant influence on flowering time, allowing plants to flower in a season when reproductive success is most likely. For example, many plants which grow in temperate climates, especially those which grow naturally at high latitudes, will flower earlier in long days and after a prolonged period of cold, termed vernalization. This response ensures that flowering will occur in spring and summer, seasons favourable for flower and seed development. Some species, or ecotypes within a species, have an absolute requirement for certain environmental signals; for most, while certain conditions will promote flowering, the plants will flower eventually even in "non-promotive" conditions. Plants also respond to internal signals; many species will not flower, even in "promotive" conditions, until a certain size or age has been attained.

Arabidopsis thaliana is an annual cruciferous plant which forms a vegetative rosette. The transition to flowering is marked by elongation of the internodes (bolting) producing a tall stem, which bears cauline leaves and flowers (Napp-Zinn, 1985; Koornneef *et al*., 1998a). Arabidopsis is a facultative long day plant: almost all ecotypes flower earlier

under long days than under short days, but will flower eventually in short days (Napp-Zinn, 1985; Karlsson *et al*., 1993). Arabidopsis ecotypes can be roughly divided into late flowering winter annuals, which account for the majority of ecotypes, and earlier flowering summer annuals (Napp-Zinn, 1985). Late flowering winter annual ecotypes, such as Pitztal and Innsbruck, show the strongest responses to vernalization, while in the earliest flowering summer annuals, such as Landsberg, the promotion of flowering by vernalization is negligible (Napp-Zinn, 1985; Karlsson *et al*., 1993; Bagnall, 1993). Flowering of Arabidopsis is promoted by high ratios of far red to red light (Martinez-Zapater and Somerville, 1990; Bagnall, 1992, 1993) and by blue light (Eskins, 1992).

Floral meristem identity genes, which convert the meristem from vegetative to floral, and floral organ identity genes, which specify the type and position of floral organs, are required for normal floral development (reviewed recently by Irish, 1999; Pineiro and Coupland, 1998). The genes controlling floral meristem identity can influence flowering time: mutations in the *TERMINAL FLOWER* (*TFL*) gene allow earlier flowering, suggesting that *TFL* is antagonistic to the floral transition (Shannon and Meeks-Wagner, 1991) while *LEAFY* (*LFY*) and *APETALA1* (*AP1*) promote flowering, as overexpression of these genes results in earlier flowering (Mandel and Yanofsky, 1995; Weigel and Nilsson, 1995). Interactions between these floral meristem identity genes and genes which primarily control flowering time must occur in the final stages of floral induction. Many flowering time genes affect expression of *LFY*, which is expressed in vegetative tissues and up-regulated as the plant approaches floral transition (Blazquez *et al*., 1997).

The effect of the flowering time genes is likely to be channeled through regulation of *LFY* expression.

1.2 The vernalization response

The term vernalization refers to the promotion of flower initiation by prolonged cold treatment of imbibed seed or a vegetative plant. The vernalization treatment does not directly cause initiation of flowering; floral buds differentiate some time after the plant is returned to normal growth temperatures, and, in some species, only if specific photoperiodic requirements are met (reviewed in Vince-Prue, 1975). Most, but not all, species which respond to vernalization are long day plants (reviewed in Vince-Prue, 1975). Some species of plants must attain a minimum size or age before they are competent to respond to a cold treatment by flowering early, but others, including Arabidopsis, can be vernalized as imbibed seed (reviewed in Vince-Prue, 1975; Napp-Zinn, 1985). Variation is seen within and among species for the nature and degree of the vernalization response. Some species will flower only after a cold period, while for others such as Arabidopsis, vernalization is merely promotive of flowering (reviewed in Vince-Prue, 1975). The inductive temperature range and duration of treatment required to promote flowering vary, but for most species, temperatures of 1-7°C over a period of one to three months result in earlier flowering (reviewed in Vince-Prue, 1975). The extent of the early flowering response increases with the duration of the vernalization period (reviewed in Vince-Prue, 1975; Lee and Amasino, 1995).

Exposure of the shoot apex to vernalizing temperatures is sufficient to induce early flowering; exposure of other parts of the plant, but not the shoot apex, to the cold temperature does not result in early flowering, demonstrating that the vernalization signal is not transmitted to the shoot apex from other tissues (Purvis, 1940; Schwabe, 1954; Metzger, 1988a). The presence of dividing cells is a prerequisite for vernalization (Wellensiek, 1964). In whole plants, the mitotically active cells in the shoot apex will form the inflorescence, so it is here that the vernalization response is expressed. *In vitro* experiments have shown that early flowering plants can be regenerated from other tissues, such as leaf petioles or root tips, which contained dividing cells at the time of the cold treatment (Wellensiek, 1964; Metzger, 1988a; Burn *et al*., 1993a), indicating that the vernalization signal is transmitted mitotically and is stable through many rounds of cell division, and through de- and re-differentiation. However, the "thermoinduced" state is not transmissible through sexual reproduction, as progeny of vernalized plants will not flower early unless cold treated (reviewed in Lang, 1965).

It should be noted that vernalization responsive perennial plants require re-vernalization each season in order for flowering to occur (reviewed in Vince-Prue, 1975). This appears to contradict the transmission of the vernalization signal through mitosis. It is possible, however, that shoot apices are de-vernalized during the summer, as temperatures in the range of 25 to 40 °C have been shown to reverse the effects of cold treatment in a number of species (Napp-Zinn, 1985; Vince-Prue, 1975; Lang, 1965). High light intensity prevents this de-vernalizing process in Chrysanthemum and Arabidopsis (Schwabe, 1955;

Napp-Zinn, 1960). De-vernalization by exposure to short day conditions has also been observed for several species (reviewed in Vince-Prue, 1975).

1.3 Natural variation in flowering time and vernalization response

Genetic analysis of the variation in flowering time and responsiveness to vernalization among Arabidopsis ecotypes has identified a number of genes controlling these traits, of which two, *FRIGIDA* (*FRI*) and *FLOWERING LOCUS C* (*FLC*) are the major determinants of flowering time. Crosses between a number of late- and early-flowering ecotypes revealed that a dominant gene, *FRI*, which maps to the top of chromosome 4, confers late flowering (Napp-Zinn, 1979; Lee *et al*., 1993; Burn *et al*., 1993b; Clarke and Dean, 1994). The early flowering ecotype Columbia is thought to contain a non-functional *FRI* allele, as mutants of the San Feliu-2 *FRI* allele, which confers late flowering as a wildtype allele, are almost identical in flowering time to Columbia when introgressed into the Columbia background (Michaels and Amasino, 1999). The presence of dominant late flowering alleles at a second locus, the semi-dominant *FLC* locus on chromosome 5, is necessary for expression of the extreme late flowering phenotype seen in ecotypes such as Pitztal and Innsbruck (Koornneef *et al*., 1994; Lee *et al*., 1994a). The extreme late flowering phenotype conferred by the presence of late flowering alleles of both *FRI* and *FLC* can be reversed by vernalization (Napp-Zinn, 1979; Lee *et al*., 1993; Clarke and Dean, 1994; Lee and Amasino, 1995). While most ecotypes carry dominant *FLC* alleles, the laboratory strains Landsberg *erecta* (L*er*) and C24 carry recessive alleles which suppress the late flowering phenotype conferred by *FRI* (Lee *et al*., 1994a; Sanda

and Amasino, 1995). Analysis of null mutants in the late *FLC* allele from ecotype San Feliu-2 revealed that they eliminated late flowering due to *FRI* in the same way as the *flc*-L*er* allele, suggesting that the early *FLC* alleles are null mutations (Michaels and Amasino, 1999).

A number of genes and quantitative trait loci (QTL) have been identified which have minor effects on Arabidopsis flowering time. Analysing the same cross in which *FRI* was identified, Napp-Zinn (1957, 1969) found a further two loci, *KRYOPHILA* (*KRY*) and *JUVENALIS* (*JUV*), for which recessive alleles caused a small delay in flowering time. Grbic and Bleeker (1996) identified the *AERIAL ROSETTE* (*ART*) locus in the very late ecotype Skye. In combination with dominant alleles at a locus on chromosome 4, probably *FRI*, dominant alleles of *ART* delay the floral transition in axillary meristems, leading to the formation of aerial rosettes. *ART* maps close to the *FLC* locus, and therefore the phenotype may be partly due to the presence of late *FLC* alleles (Grbic and Bleeker, 1996). Crosses between late and early ecotypes (Clarke *et al*., 1995) and between early ecotypes (Kowalski *et al*., 1994; Jansen *et al*., 1995; Alonso-Blanco *et al*., 1998) have been analysed for QTLs, and in all cases, multiple QTLs affecting flowering time were identified. These QTLs are located throughout the genome; some may correspond to genes for which mutant alleles are known.

1.4 The "classical" late flowering mutants

Further genes controlling flowering time have been identified by mutagenesis. Late-flowering mutants in early flowering backgrounds reveal genes which promote flowering. The set of 12 "classical" late flowering mutants was identified by Redei (1962), Koornneef *et al*. (1991) and Lee *et al*. (1994a). All the late flowering mutants isolated in the L*er* background by Koornneef *et al.* (1991) flower later than L*er* in both long and short days (Koornneef *et al*., 1991), and all but two mutations are recessive, these two being semi-dominant (*co*) and dominant (*fwa*).

Mutants in the L*er* background were classified based on their responses to photoperiod and vernalization, and on their epistatic interactions (Koornneef *et al*., 1991, 1998b). L*er* plants show only a small response to vernalization in either long or short day conditions, and a moderate promotion of flowering by long days (Koornneef *et al*., 1991). Class 1 mutants *fca*, *fve, fpa, fy,* and *ld* show increased responsiveness to environmental stimuli: they have an increased photoperiod response, and flower much earlier after vernalization in both long and short days (Koornneef *et al.*, 1991). Mutants in Class 2 (*fe, ft, fd, fwa*) have responses to photoperiod and vernalization that are slightly reduced compared to wildtype, while mutants in Class 3 (*co, gi, fha*) are almost entirely insensitive to photoperiod and vernalization (Koornneef *et al.*, 1991).

Genetic interactions among ten of the mutants described above were examined by analysis of the flowering time of double mutants (Koornneef *et al*., 1998b). The mutants

less sensitive to environmental stimuli, in Classes 2 and 3, fall into one epistatic group (group A; Koornneef *et al*., 1998b), while the Class 1 mutants fall into a second epistatic group (group B; Koornneef *et al*., 1998b). In general an additive effect was seen for interactions between mutants from groups A and B, indicating that the mutations in these two groups disrupt different pathways to flowering (Koornneef *et al*., 1998b).

1.5 A model for the control of flowering time

Koornneef *et al*. (1998a) proposed a model for the control of flowering time in Arabidopsis, based on the physiological and genetic classification of the "classical" late flowering mutants (Koornneef *et al*., 1991, 1998b), and on observations by other groups, discussed later. This model postulates the existence of three pathways to flowering: a pathway that constitutively promotes flowering; a pathway that promotes flowering in response to long photoperiods; and a pathway that promotes flowering after vernalization (Koornneef *et al*., 1998a). More than one of these pathways could be active simultaneously; for example, the photoperiod and vernalization pathways can act simultaneously, as, for many ecotypes, long days and vernalization promote flowering more than either treatment singly (Karlsson *et al*., 1993). A description of the model is given below; after this, the interactions and likely roles of genes influencing flowering time will be discussed with reference to the model.

The Class 1 mutants have an increased dependence on environmental signals to promote flowering, suggesting that the wildtype genes promote flowering via an autonomous or

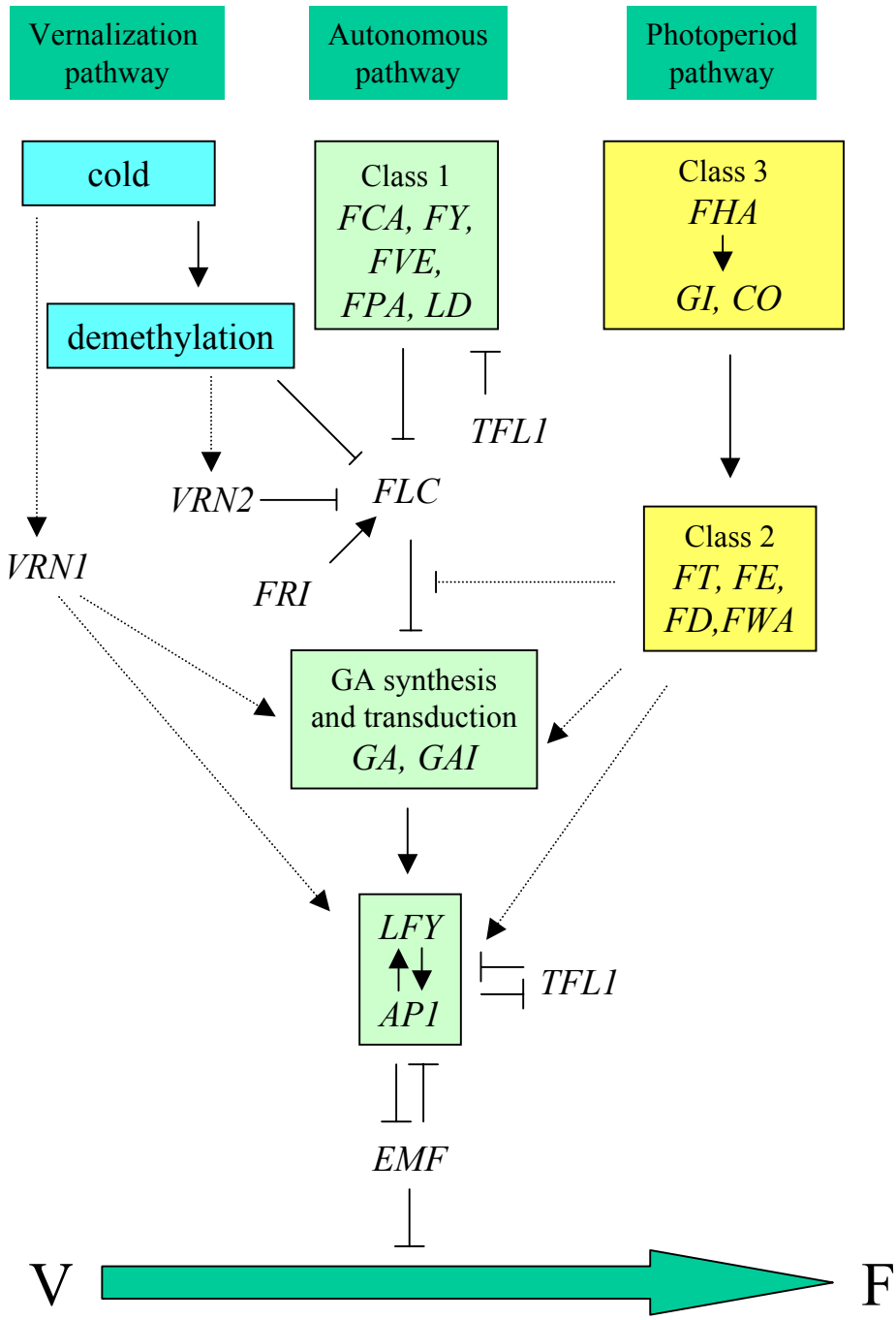
constitutive pathway to flowering (Koornneef *et al*., 1998a, b; Figure 1.1). The Class 2 and 3 mutants show different degrees of reduced sensitivity to environmental signals. However, long vernalization treatments promote flowering of Class 2 and 3 mutants (Koornneef *et al*., 1991; Chandler and Dean, 1994), and double mutants between Class 1 and either Class 2 or 3 are responsive to vernalization (Koornneef *et al*., 1998b). This indicates that the wildtype genes identified in Class 2 and 3 mutants are not required for the vernalization response, but act in a pathway that promotes flowering in response to long photoperiods (Koornneef *et al*., 1998a, b; Figure 1.1).

1.5.1 Floral repressors

The *FRI* and *FLC* genes described above are two major determinants of flowering time and vernalization responsiveness. Introgression of alleles causing late flowering into the early flowering L*er* background revealed that the presence of late alleles of either *FRI* or *FLC* delays flowering slightly, but that late alleles at both loci are required to confer the extreme lateness seen in ecotypes such as Pitztal or San Feliu-2 (Lee *et al*., 1994a). This synergistic effect on flowering time implies an interaction between these two genes whereby one up-regulates the other, perhaps reciprocally.

The recent cloning of the *FLC* gene, which encodes a MADS box protein (Sheldon *et al*, 1999; Michaels and Amasino, 1999), allowed further analysis of the role of *FLC* and the interaction between *FLC* and *FRI*. The presence of the late flowering *FLC* allele delays flowering independently of the presence of *FRI* (Michaels and Amasino, 1999; Sheldon *et*

Figure 1.1: A model of pathways to flowering in Arabidopsis, showing interactions between flowering time genes; genes which are believed to act in the same pathway, based on genetic and physiological characterisation, are grouped in boxes. Promotive effects are shown by "—▶", while repressive effects are shown by "—⊣". Alternative pathways are shown by dotted lines. The order of genes in the photoperiod pathway is adapted from Levy and Dean (1998).


Vernalization pathway
Autonomous pathway
Photoperiod pathway
cold
demethylation
Class 1
FCA, FY, FVE, FPA, LD
Class 3
FHA
GI, CO
TFL1
VRN2
FLC
VRN1
FRI
Class 2
FT, FE, FD,FWA
GA synthesis and transduction
GA, GAI
LFY
AP1
TFL1
EMF
V
F

al., 1999); as observed previously, the presence of both *FLC* and *FRI* has a synergistic effect, causing extreme lateness (Koornneef *et al*., 1994; Lee *et al*., 1994a; Michaels and Amasino, 1999). *FLC* is up-regulated in plants homozygous for the dominant *FRI* allele (Michaels and Amasino, 1999). This provides a mechanism to explain the synergistic action of these two genes in delaying flowering. The *FRI* gene has been cloned (Gendall *et al*., 1999), so it will now be possible to ascertain whether *FLC* up-regulates *FRI*.

Other genes which repress flowering have been identified through studies of early flowering mutants. The *embryonic flower 1* and *2* (*emf1* and *emf2*) mutants are the most severe early flowering mutants identified so far, and therefore the *EMF* genes are thought to play a major role in repression of flowering. *emf1* and *emf2* mutants produce no rosette, but do produce a few cauline leaves followed by an inflorescence of abnormal, incomplete floral buds (Sung *et al.*, 1992; Yang *et al*., 1995). Studies of double mutants between the *emf* mutants and other flowering time mutants have indicated some possible interactions. Late flowering mutants *co* and *gi*, which lack photoperiod and vernalization responsiveness, had no effect on the phenotype of *emf* mutants, indicating that *CO* and *GI* operate upstream of the *EMF* genes and may down-regulate *EMF* activity (Yang *et al*., 1995; Haung and Yang, 1998). Double mutants between *emf* and *fwa* had enhanced vegetative properties, with increased numbers of cauline leaves and either no flowers or leaf-like flowers (Haung and Yang, 1998). As the mutant *fwa* allele is expressed, while the wildtype allele is not (W. Soppe, personal communication), this phenotype indicates that fwa may promote vegetative growth independently of interactions with *EMF* (Haung and Yang, 1998). *FCA* is required for early flowering in *emf* mutants, as the formation of

flowers was delayed in *fca emf* double mutants; thus *FCA* may promote flowering independently of interactions with the *EMF* genes (Page *et al*., 1999).

The floral meristem identity genes *APETALA1* (*AP1*) and *AGAMOUS* (*AG*) are expressed early in *emf* mutants, suggesting that these genes are down-regulated by *EMF* activity (Chen *et al*., 1997). *emf1* is epistatic to the floral meristem identity mutants *ap1*, *ap2*, *leafy* (*lfy*), and *ap1 cauliflower* (Ratcliffe *et al*., 1999), suggesting that the products of these genes may down-regulate *EMF* activity. A reciprocal negative regulation between the *EMF* genes and the floral meristem identity genes may form part of the mechanism regulating the timing of flowering (Chen *et al*., 1997).

Mutants at *TERMINAL FLOWER 1* and *2* (*tfl1* and *tfl2*) flower early, and the wildtype indeterminate meristem is converted to a terminal flower (Shannon and Meeks-Wagner, 1991; Alvarez *et al*., 1992; Larsson *et al*. 1998). *tfl2* has reduced photoperiod sensitivity, which may indicate an interaction of *TFL2* with the photoperiod pathway to flowering (Larsson *et al*., 1998). Double mutants between *tfl1* and *tfl2* show a more severe phenotype, suggesting that the two genes act in distinct pathways (Larsson *et al*., 1998).

The *TFL1* gene is expressed at high levels in the inflorescence apex (Bradley *et al*., 1997). Constitutive over-expression of *TFL1* extends the vegetative and the reproductive growth phases (Ratcliffe *et al*., 1998). The general increase in duration of growth phases counters the contraction of growth phases seen in *tfl1* mutants (Weigel *et al*., 1992), and led Ratcliffe *et al*. (1998) to suggest that *TFL1* interacts with a mechanism determining

phase transitions, rather than having separate effects, firstly on the timing of the transition to flowering and secondly on meristem identity. In plants over-expressing *TFL1*, the up-regulation of the meristem identity genes *LEAFY* (*LFY*) and *APETALA 1* (*AP1*) is significantly delayed (Ratcliffe *et al*., 1998). Plants constitutively expressing either *LFY* or *AP1* have reduced expression of *TFL1*, indicating that *LFY* and *AP1*, which up-regulate each other (Liljegren *et al*., 1999), down-regulate *TFL1* expression directly (Ratcliffe *et al*., 1999; Liljegren *et al*., 1999). The mutual down-regulation of activity between *TFL1* and the meristem identity genes limits expression of *TFL1* to the centre of the shoot apex, causing it to remain indeterminate in wildtype infloresences, and limits expression of the floral meristem identity genes to the periphery, where they cause initiation of floral meristems (Ratcliffe *et al*., 1999).

emf1 was seen to be epistatic to *tfl1* in double mutant analysis, suggesting that *EMF1* is required for *TFL1* function, and that *TFL1* may affect inflorescence development through up-regulation of *EMF1* (Chen *et al*., 1997). Both *TFL1* and *EMF1* have been proposed to have a reciprocal negative regulation relationship with meristem identity genes such as *AP1*, *LFY* and *AG* (Chen *et al*., 1997; Liljegren *et al*., 1999; Ratcliffe *et al*., 1999). This suggests that *TFL1* and *EMF1* may up-regulate each other by down-regulating these meristem identity genes (refer to Figure 1.1).

The *curly leaf* (*clf*) mutant shows a pleiotropic phenotype including upwardly curled leaves, reduced inflorescence height and abnormal flowers. *clf* also flowers early, indicating a possible role for *CLF* in delaying flowering (Goodrich *et al*., 1997). The

floral meristem identity gene *AG* was ectopically expressed in leaves, stems and flowers of *clf* mutants, indicating that *CLF* has a role in negative regulation of *AG* (Goodrich *et al*., 1997). The *CLF* gene has been cloned, and is similar to the Drosophila Polycomb group genes that are involved in transcriptional repression of homeotic genes through chromatin remodelling (Goodrich *et al*., 1997). The *clf* mutant is phenotypically similar to the early flowering *wavy leaves and cotyledons* mutant (*wlc*), which expresses *AG* and *AP3* ectopically in leaves (Bancroft *et al*., 1993). *wlc* shows hypomethylation in centromeric repeats (C. Hutchinson and C. Dean, cited in Levy and Dean, 1998); it is possible that the ectopic expression of *AG* and *AP3* is due to reduced methylation. Plants expressing an antisense construct for the Arabidopsis methyltransferase gene *METI* showed hypomethylation in repeat and low copy sequences, and expressed *AG* and *AP3* ectopically in leaves (Finnegan *et al*., 1996). The observations made on these plant lines suggest that chromatin structure and cytosine methylation have a role in regulating flowering time, possibly through regulating the expression of floral homeotic genes including *AG* and *AP3*.

1.5.2 The constitutive pathway

As described above, the late flowering mutants *fca*, *fpa*, *ld*, *fve* and *fy* have an increased dependence on environmental stimuli to promote flowering, indicating that the wildtype genes *FCA*, *FPA*, *LD*, *FVE* and *FY* promote flowering constitutively. Two of these genes, *FCA* and *LD*, have been cloned, and both encode proteins which may regulate the expression of other genes.

LD encodes a nuclear protein containing a glutamine-rich domain that is found in some transcription factors (Lee *et al*., 1994b; Aukerman *et al*., 1999). *LD* is expressed primarily in shoot and root apices, including the shoot apical meristem and leaf primordia (Aukerman *et al*., 1999). *FLC* expression is increased in the *ld-3* mutant, demonstrating that LD negatively regulates *FLC* (Michaels and Amasino, 1999; Sheldon *et al*., 1999). This suggests that the reduction in flowering time of *ld-3* mutants after vernalization (Lee *et al*., 1994a) is due to down-regulation of *FLC* by vernalization. It also raises the question of whether *LD* is up-regulated by vernalization, since vernalization of *ld-3* does not reduce flowering time to that of the Ws wildtype (Lee *et al*., 1994a).

FCA encodes a protein containing two RNA-binding domains and a WW protein interaction domain, suggesting that it functions in post-transcriptional regulation of other flowering gene transcripts (Macknight *et al*., 1997). Four alternative transcripts of *FCA* have been detected; only the full length transcript contains the WW protein interaction domain. Constitutive over-expression of *FCA* using the 35S promoter led to early flowering (Macknight *et al*., 1997). Analysis of double mutants between *fca* and *tfl1*, *ap1*, and *lfy* indicated that *FCA* promotes flowering both through activation of *LFY* and *AP1* and also through a second pathway not involving these genes (Page *et al*., 1999). *LFY* promoter activity is reduced in *fca* mutants, and the late flowering phenotype can be rescued by constitutive *LFY* expression (Nilsson *et al*., 1998), confirming the relationship of these two genes. Similar observations were made for the *fve* mutant, indicating that *FVE* also up-regulates *LFY* (Nilsson *et al*, 1998).

The late flowering *fld* mutant is responsive to photoperiod and vernalization (Sanda and Amasino, 1996), and therefore has been classified with the Class 1 late flowering mutants thought to be impaired in the constitutive pathway to flowering (Koornneef *et al*., 1991, 1998a, 1998b). Like *ld* and *FRI*, *fld* requires the dominant *FLC* allele for expression of the late flowering phenotype (Sanda and Amasino, 1996), indicating that *FLD* may, like *LD* (Michaels and Amasino, 1999), down-regulate *FLC* expression. Expression of *FLC* is increased in *ld* and *fld*, as well as in the *fca*, *fve*, and *fpa* mutants (Sheldon *et al*., 1999, 2000). This indicates that the *FCA*, *FVE*, and *FPA* genes, like *LD* and possibly *FLD*, down-regulate *FLC* expression, and that *FLC* retards flowering via the constitutive pathway to flowering.

1.5.3 The photoperiod pathway

Mutants in the pathway to flowering promoted by long photoperiods may be impaired in perception or transduction of daylength signals. The product of the *FHA* gene is the blue light receptor cryptochrome 2 (CRY2; Lin *et al*., 1996; Guo *et al*., 1998). The level of the CRY2 protein drops rapidly in the light (Lin *et al*., 1998), indicating that it could provide a mechanism for measuring daylength. CRY2 up-regulates *CO* (Guo *et al*., 1998); thus the light-regulation of CRY2 levels could also be the mechanism regulating *CO* expression, which, in wildtype plants, is very low in short days, but much higher in long days (Putterill *et al*., 1995). Constitutive expression of *CO* leads to early flowering in short days (Putterill *et al*., 1995), confirming that this gene promotes flowering. CO

contains a region homologous to the zinc finger domain of GATA-1 type transcription factors (Putterill *et al*., 1995), and is likely to up-regulate the floral meristem identity genes *LFY*, *TFL1* and *AP1* (Simon *et al*., 1996; Nilsson *et al*. 1998).

GI also appears to up-regulate *LFY* (Nilsson *et al*., 1998), but is likely to have its effect through *CO*; the *gi* mutation has no effect on the flowering time of *CO*-overexpressing plants, and therefore is likely to act upstream of *CO* (Igeno and Coupland cited in Onouchi and Coupland, 1998). *GI* encodes a novel protein likely to be membrane localized; its expression is regulated by the circadian clock, and peaks during the light period. In long days *GI* expression is lower and peaks later, and the reduction in expression from peak levels is slower (Fowler *et al*., 1999; Park *et al*., 1999). *GI* interacts with *ELF3*, which is predicted to input light signals to the circadian clock and to act in blue light-regulated photomorphogenesis (Hicks *et al*., 1996; Zagotta *et al*., 1996), and with *LHY* and *CCA1*, which function in the circadian clock (Park *et al*., 1999; Green and Tobin, 1999; Schaffer *et al*., 1998).

The *FT* and *FWA* genes have been grouped together on the basis of genetic analyses and physiology (Ruíz-Garcia *et al*.,1997; Madueno *et al*., 1996). Analyses of double mutants between *ft* or *fwa* and *lfy* or *ap1* suggest that, when *LFY* is impaired, *FT* and *FWA* are both required to induce the expression of *AP1* and thus flowering (Ruíz-Garcia *et al*., 1997). *FT* and *FWA* are likely to control the competence of plants to respond to *LFY* induction; Nilsson *et al*. (1998) found that the *ft* and *fwa* mutants were epistatic to the effects of constitutively expressed *LFY*, and that the activity of the *LFY* promoter was

only slightly decreased in these mutants. Simon *et al*. (1996) suggested that genes other than *CO* were required for *AP1* up-regulation; *FT* and *FWA* are likely to at least partially fill this role. *CO* upregulates *FT* expression in the early vegetative phase, indicating that *CO* acts upstream of *FT* (Kardailsky *et al.*, 1999; Kobayashi *et al*., 1999). *FWA* and *FT* could both be required, with *LFY*, for the function of other floral meristem identity genes (Ruíz-Garcia *et al*., 1997).

The *FT* gene has strong sequence homology to *TFL1*; these genes show similarity to mammalian membrane-associated proteins, suggesting possible roles in a signaling cascade (Araki *et al*., 1998; Kardailsky *et al.*, 1999; Kobayashi *et al*., 1999). As mutations in *FT* lead to late flowering, while mutations in *TFL1* lead to early flowering, these genes are likely to have different roles; their similarity may indicate interactions with the same targets, such as *LFY*. Like *FT* and *FWA*, *FE* and *FD* may also have a role in enabling the shoot apex to respond to *LFY* expression (Nilsson *et al*., 1998). However, *LFY* promoter activity was decreased in *fe* and *fd*, indicating that these genes also function in up-regulating *LFY* expression (Nilsson *et al*., 1998).

1.5.4 The vernalization pathway

Mutants showing a reduced response to vernalization have been isolated (Chandler *et al*., 1996); these identify genes acting in the vernalization pathway to flowering, in either perception or transduction of the cold stimulus (Koornneef *et al*, 1998a; Figure 1.1). These *vrn* mutants were isolated in the vernalization responsive late flowering mutant *fca*,

as, after vernalization, they flowered later than vernalized *fca* plants (Chandler *et al*., 1996). Five independent recessive mutations were isolated, which fell into three complementation groups; two of these mutants were characterized further. *vrn1 fca* plants flower later than *fca* plants when vernalized, but no later when unvernalized. The *vrn1* mutation was segregated away from *fca*; *vrn1* mutants flower late and show no response to vernalization. *vrn1* also reduces the vernalization response of the late flowering mutants *fve*, *ld*, *fwa*, *fe*, *fpa* and *ft* (Chandler *et al*., 1996). *vrn2*, which could not be separated from *fca*, reduces the vernalization response of *fca* plants, but also delays flowering in unvernalized *fca* plants (Chandler *et al*., 1996), indicating that *vrn2* may act in more than one pathway or downstream of the convergence of two or more pathways.

Expression of the floral repressor *FLC* is decreased after vernalization (Sheldon *et al*., 1999; Michaels and Amasino, 1999), indicating that *FLC* acts as a block to the vernalization pathway. As mentioned above, *FLC* is down-regulated by wildtype genes acting in the constitutive pathway, indicating that *FLC* also retards flowering via the constitutive pathway (Sheldon *et al*., 1999; Michaels and Amasino, 1999). Expression of *FLC* is higher in *fca vrn2* plants than in *fca* plants, suggesting that *FLC* is down-regulated by *VRN2*. However, *FLC* expression is not increased by the presence of the *vrn1* mutation (Sheldon *et al*., 1999). As the double mutant *fca vrn1* retains a vernalization response, albeit reduced, the *vrn1* mutation does not completely block the vernalization pathway. This may indicate that the *vrn1* mutation is leaky, or that a parallel vernalization pathway involving *VRN2* and *FLC* makes this gene partly redundant in mediating the vernalization response (refer to Figure 1.1). As *vrn1* plants do not show a

vernalization response, it is unlikely that the *vrn1* mutation is leaky (Chandler and Dean, 1996). If the second explanation is correct, no vernalization response would be expected in the *vrn1* mutant in the L*er* background, as *FLC* expression in *vrn1* is very low, equivalent to that seen in wildtype L*er* (Sheldon *et al*., 1999). Thus, the effect on flowering time of down-regulation of *FLC* would be negligible, and the *VRN1* component of the vernalization pathway would be blocked by mutation. In the *fca vrn1* mutant, *FLC* is over-expressed (Sheldon *et al*., 1999); thus, vernalization leads to down-regulation of *FLC* and earlier flowering, despite the mutation in *vrn1*.

The *vrn1* mutation delays flowering in unvernalized as well as vernalized plants (Chandler *et al*., 1996), suggesting that it acts in more than one pathway to flowering. *fca* is epistatic to *vrn1*, as *fca vrn1* flowers no later than *fca* (Chandler *et al*., 1996); this suggests that *fca* may act downstream of *vrn1* in the autonomous pathway to flowering. However, any interaction between *FCA* and *VRN1* does not appear to affect the down-regulation of *FLC* by *FCA*, as *vrn1* mutants show no increase in *FLC* expression (Sheldon *et al*., 1999). Cloning of *VRN1* and *VRN2* will allow more informative studies of the interactions between, and roles of, *VRN1*, *VRN2*, *FCA* and *FLC*. As mentioned above, late flowering alleles of *FRI* up-regulate *FLC* (Michaels and Amasino, 1999); now that *FRI* has been cloned (Gendall *et al*., 1999), its role in the pathways discussed above may become clearer.

1.6 The role of gibberellic acid in the response to vernalization

Exogenous gibberellic acid (GA) promotes flowering in a range of plants, including Arabidopsis, although in other plants, it can have inhibitory or negligible effects on flowering (Lang, 1965; Zeevaart, 1983; Pharis and King, 1985). In Arabidopsis, application of GA substitutes for vernalization or long days in promoting flowering, suggesting that cold treatment and increased daylength may induce flowering through increasing levels of active gibberellins (Napp-Zinn, 1969, 1985).

Mutants in gibberellin biosynthesis and responses show altered flowering responses. The *ga1-3* mutant, which is extremely deficient in GA (Zeevaart and Talon, 1991), flowers late in continuous light, and not at all in short days unless treated with exogenous gibberellins, indicating the absolute requirement of GA in short days to allow flowering (Wilson *et al*., 1992). *ga1-3* shows no response to a 45 day cold treatment, suggesting that GA is required for the vernalization response. However, as *ga1-3* is in the L*er* background, it is unlikely to have elevated *FLC* levels, and this may account for its lack of vernalization responsiveness. Also, a double mutant between *fca* and *ga1-3*, which would be expected to have elevated *FLC*, responds to vernalization when grown in continuous light, suggesting that, though GA may have a role in the response to vernalization, other response pathways can substitute for it (Chandler and Dean, unpublished data in Chandler *et al*., 1996). The *gai* mutant, which is blocked in GA responses (Peng *et al*., 1997), flowers late in short days but does respond to a 45 day vernalization treatment, though still flowering later than wildtype. This mutant also

responds to continuous light conditions, flowering no later than wildtype (Wilson *et al*., 1992). It should be noted that while *ga1-3* requires exogenous GA for germination, *gai* does not (Wilson *et al*., 1992); evidently some GA responses are still functioning in the *gai* mutant, and these may have allowed its response to vernalization.

A possible mechanism for the failure of *ga1-3* to flower in short days was suggested by Blázquez *et al*. (1998), who found that in short days, *ga1-3* plants lacked *LFY* promoter activity. Constitutive expression of *LFY* in *ga1-3* plants restored flowering. These results indicate that GA causes activation of the *LFY* promoter (Blázquez *et al*., 1998). Nilsson *et al*. (1998) found that the flowering time of the *gai* mutant is rescued by constitutive expression of *LFY*, again indicating that induction of *LFY* expression is, in part at least, a response to GA.

Chandler and Dean (1994) found that exogenous GA promoted the flowering of all late flowering mutants tested: *fb*, *fca*, *fd*, *fe*, *co*, *fpa*, *ft*, *fve*, *fwa*, *fha* and *fy*; although the last two named had very small responses. The wildtype L*er* response was non-significant. As mutants from both the vernalization responsive group and the less responsive group flowered early in response to GA, the hormone appears to have a general role in promotion of flowering rather than a specific role in the vernalization response. The *vrn1* mutant, which is impaired in vernalization responsiveness, also flowers early when treated with GA (Chandler *et al*., 1996).

While GA may not be essential for the response to vernalization, or act specifically in the vernalization response, studies in the close relative of Arabidopsis, *Thlaspi arvense*, have shown changes in GA metabolism after vernalization. The conversion of kaurenoic acid (KA) to 7β-hydroxy KA was accelerated in shoot apices, though not in leaves, after prolonged exposure to low temperatures (Hazebroek and Metzger, 1990; Hazebroek *et al*., 1993). This early step in GA biosynthesis is catalyzed by kaurenoic acid hydroxylase (KAH); KAH activity was observed to increase in the shoot apex, and KA levels declined (Hazebroek and Metzger, 1990; Hazebroek *et al*., 1993). These changes in GA metabolism are likely to lead to increased GA levels in the shoot apex, which in turn should induce higher expression of *LFY* and thus promote flowering.

As all late flowering mutants tested responded to GA application by flowering early, GA can be presumed to act downstream of all these mutants in the pathways to flowering, possibly after convergence of the various pathways. The *flf-1* mutant, which flowers late due to over-expression of *FLF* (=*FLC*), requires repeated applications of exogenous gibberellin to promote flowering, indicating that the *FLC* gene product may directly antagonize the promotion of flowering by GA (Sheldon *et al*., 1999). *FLC* levels are reduced by vernalization, lessening the effects of this antagonist and allowing the flowering response to GA to proceed.

1.7 Cytosine methylation

Methylation of cytosine, by addition of a methyl residue to the carbon 5 position, is common in prokaryotes and many eukaryotes, including plants. In plants, the proportion of cytosine residues methylated ranges from 6% in Arabidopsis (Kakutani *et al*., 1999) up to 37% in *Helianthus* (Ergle and Katterman, 1961). Methylation levels in mammals are much lower, ranging from 2-7% (Razin and Riggs, 1980). A number of eukaryotic species, including some dipterans such as *Drosophila melanogaster*, all nematodes surveyed, and a number of ascomycetes including the yeast *Saccharomyces cerevisiae*, lack detectable cytosine methylation (Urieli-Shoval *et al*., 1982; Proffitt *et al*., 1984; Regev *et al*., 1998).

Cytosine methylation is most common in repeated sequences (Guseinov *et al*., 1975; Bennetzen *et al*., 1988; Bennetzen, 1996), although it also occurs in single copy DNA (Meyer *et al*., 1994; Jacobsen and Meyerowitz, 1997). The most usual sequence contexts for methyl-cytosine in plant DNA are the symmetrical sequences CpG or CpNpG (where N is any nucleotide), and in these contexts methylation is also symmetrical, so that the cytosines on both strands are methylated (Gruenbaum *et al*., 1981). However, methylation also occurs in other sequence contexts which are not strand symmetrical (Meyer *et al*., 1994; Zhou *et al*., 1998; Goubley *et al*., 1999). In mammals, methyl-cytosine is seen in the sequence CpG (Stein *et al*., 1982; Araujo *et al*., 1998). Vertebrate and plant genomes contain “CpG islands”, sequences of approximately 1-2 kb, which are

G+C-rich, commonly unmethylated at CpG dinucleotides, and often associated with promoters of genes (Bird, 1987; Antequera and Bird, 1988).

1.7.1 Cytosine methylation as genome defense

In prokaryotes, cytosine methylation functions in the restriction-modification system which operates to degrade non-self DNA by restriction at sites which are methylated in self DNA (Noyer-Weidner and Trautner, 1993). Eukaryotes also use cytosine methylation in a defence role, although in different ways. Eukaryote genomes contain large numbers of transposons, retrotransposons and retroviruses (referred to collectively here as transposable elements or TEs), which often occur as repetitive sequence families (reviewed in Federoff, 1989; Finnegan, 1989). TEs pose a threat to genome structure and to normal gene expression, as they may insert into coding sequences, interrupting gene expression, and may induce transpositions and other structural rearrangements of host DNA (reviewed in Federoff, 1989; Finnegan, 1989). Methylation of cytosine residues within TE sequences is associated with loss of transcription and transposition (reviewed in Singer and Selker, 1995; Matzke *et al*., 1996; Fedoroff, 1999). The activity of three maize transposons, Supressor-mutator (Spm), Mutator (Mu), and Activator (Ac) is negatively correlated with methylation of promoter and other sequences within the transposons (Fedoroff *et al*., 1995; Lisch *et al*., 1995; Kunze *et al*., 1996; Fedoroff, 1999). In mice mutated in the Dnmt1 methyltransferase gene, the transcription of endogenous intracisternal A particle (IAP) retroviruses was increased 50-100-fold throughout the tissues of the embryo, indicating down-regulation by methylation (Walsh *et al*., 1998b).

Methylation was seen in the sequence of the Moloney murine leukemia virus which was integrated into the germline of the Mov-3 mouse strain; this methylation correlated with reduced infectivity, compared with the unmethylated cloned sequence (Harbers *et al*., 1981). Similarly, in the fungi *Neurospora crassa* and *Ascobolus immersus*, repeat sequences including TEs are silenced, either by repeat induced point mutation (RIP), which may involve cytosine methylation (Selker *et al*., 1987; Cambareri *et al*., 1989; Singer *et al*., 1995) or by methylation induced premeiotically (MIP, Goyon and Faugeron, 1989; Rhounim *et al*., 1992; Rossignol and Faugeron, 1994) (reviewed in Selker, 1997). Similar processes are likely to occur in other fungi (Freedman and Pukkila, 1993; Nakayashiki *et al*., 1999). *Drosophila*, which lacks cytosine methylation, is likely to use other methods such as chromatin remodelling to control TEs, though Engels (1997) noted the rapid spread of the P transposon through the *Drosophila* population, indicating that *Drosophila* may not control TEs very effectively.

1.7.2 Cytosine methylation and eukaryotic development

Disruption of normal methylation levels and patterns has a profound effect on development in all eukaryote species so far studied, indicating that cytosine methylation has a role in regulation of eukaryote development. Mouse embryos homozygous for a loss-of-function mutation of the major murine methyltransferase, *Dnmt1*, showed a reduction in methylation levels to approximately 30% of wildtype levels (Li *et al*., 1992). The homozygous mutation was embryo lethal, causing spontaneous abortion in mid-gestation, and these embryos showed stunting and delayed development (Li *et al*., 1992).

Ascobolus strains carrying a knockout mutation of the *Masc1* methyltransferase gene were viable but lacked the *de novo* methylation which occurs between fertilization and karyogamy in sexual reproduction (MIP). Crosses between mutant strains were sterile, with development disrupted after fertilization (Malagnac *et al*., 1997). *Neurospora* mutants defective in methylation (*dim-1*) showed reductions of 40-50% in methylation levels throughout the genome, altered patterns of methylation and reduced female fertility (Foss *et al*., 1998). Arabidopsis plants transformed with an antisense construct to the major methyltransferase *METI* (Finnegan and Dennis, 1993) had methylation levels as low as 10% of wildtype (Finnegan *et al*., 1996). These plants showed a pleiotropic phenotype including floral abnormalities; over several generations the phenotype increased in severity leading to sterility (Finnegan *et al*., 1996). A similarly pleiotropic phenotype was seen for the *decreased DNA methylation 1* mutants (*ddm1*), in which methylation levels were reduced to approximately 30% of wildtype (Vongs *et al*., 1993).

The means by which methylation plays a role in development is still under debate. Possible explanations include regulation of expression of tissue specific genes, alteration of chromatin states, and genomic imprinting. These possibilities will be discussed below.

1.7.3 Cytosine methylation and expression of tissue-specific genes

For many genes, including genes expressed only in specific tissues or at specific developmental stages, a correlation has been seen between cytosine methylation, especially in the promoter region, and suppression of transcription, suggesting that

expression patterns of tissue-specific genes could be regulated by programmed methylation and demethylation of specific sequences (reviewed in Eden and Cedar; 1994, Adams, 1995). However, no definitive experiments have shown a causal link between methylation and expression patterns of tissue-specific genes, while Walsh and Bestor (1998) have shown that Dnmt1-deficient mouse embryos did not show ectopic or precocious expression of three genes previously reported to be regulated by promoter methylation (Loo and Cauchi, 1992; Rhodes *et al*., 1994; Weiss *et al*., 1996). Also, methylation patterns of a number of tissue-specific genes could not be correlated with their expression patterns in tissues of fetal and newborn mice (Walsh and Bestor, 1998), calling into question the role of methylation in regulation of tissue specific genes.

1.7.4 Cytosine methylation and transgene silencing

Studies of epigenetic events often seen in transgenic plants have shed more light on the possible role of methylation in regulating gene expression. Two major types of transgene silencing occur. The first relates to the position of integration of the transgene, due to negative effects of flanking plant DNA on transgene expression. Several silenced or unstably expressed transgenes have been shown to be localized in regions of heterochromatic and repetitive DNA (Pröls and Meyer, 1992; Iglesias *et al*., 1997; Jakowitsch *et al*., 1999). In two transgenic petunia lines, the characteristic methylation pattern of the flanking DNA was imposed on the border regions of the transgenes; the line showing higher expression was hypomethylated relative to the line with lower expression. The transgene of a third line was integrated into a highly repetitive, highly methylated

region and was silenced (Pröls and Meyer, 1992). The correlation between methylation and silencing, and the differential methylation of the two expressing lines, suggests that methylation patterns of flanking DNA may "migrate" into transgenes and affect expression. Thus, the high methylation levels typical of heterochromatic regions may be responsible for silencing adjacent transgenes.

A second type of transgene silencing commonly occurs when multiple copies of the transgene are present, or when there is significant homology between the transgene and an endogenous sequence. This is referred to as homology dependent gene silencing (HDGS; reviewed in Meyer and Saedler, 1996; Stam *et al*., 1997; Kooter *et al*., 1999). Two forms of HDGS have been observed: transcriptional gene silencing (TGS) and post-transcriptional gene silencing (PTGS). TGS normally involves sequences with homology in promoter sequences, and is characterised by increased promoter methylation, which may be meiotically heritable (Meyer *et al*., 1993; Park *et al*., 1996). Dieguez *et al*. (1998) showed that methylation at symmetrical sites was not required to induce transcriptional inactivation, but was essential for maintenance of silencing, suggesting that other changes such as chromatin remodelling may initiate silencing. Chromatin condensation is associated with TGS (van Blokland *et al*., 1997). Transcriptionally silenced transgenes also show methylation at non-symmetrical sites (Meyer *et al*., 1994), but methylation at these sites is not sufficient to maintain silencing (Dieguez *et al*., 1998). Methylation of non-symmetrical sites can be stable through mitosis (Meyer *et al*., 1994). No mechanism for the perpetuation of non-symmetrical methylation patterns is known. Meyer (1999) suggests that the methylation of non-symmetrical sites is directed by chromatin structure,

and work by Pelissier *et al*. (1999) suggests that interactions of RNA with homologous DNA sequences can result in dense methylation at both symmetrical and non-symmetrical sites.

PTGS, also known as co-suppression, involves transgenes, or transgenes and endogenous sequences, which have homology in the transcribed regions. Transcription occurs normally, but no RNA message accumulates in the cytoplasm, suggesting that a highly specific RNA degradation mechanism is operating. Several models for PTGS have been proposed (Dougherty and Parks, 1995; Stam *et al*., 1997; Vaucheret *et al*., 1998; Waterhouse *et al*., 1998). Initial studies suggested that high expression of transgenes led to PTGS, suggesting that if levels of a particular RNA message exceeded a threshold, all homologous RNA molecules were degraded (Lindbo *et al*., 1993; Smith *et al*., 1994; Elmayan and Vaucheret, 1996). However, PTGS also occurs in plants that do not show high levels of transgene expression (Van Blokland *et al*., 1994). In many cases, PTGS is observed for transgene loci with tandemly linked sequences arranged in inverted repeats (Waterhouse *et al*., 1998; Hamilton *et al*., 1998; Stam *et al*., 1998); such loci could produce aberrant RNAs that trigger PTGS. The use of transgene arrangements, such as inverted repeats or sense/antisense T-DNAs, which produce double stranded RNA, leads to PTGS at high frequency (Hamilton *et al*., 1998; Waterhouse *et al*., 1998). PTGS is also often associated with transgene methylation (English *et al*., 1996; Conner *et al*., 1997; Jones *et al*., 1998a; Stam *et al*., 1998; O'Dell *et al*., 1999); this may contribute to the production of aberrant transcripts, as in *Ascobolus immersus*, transcription of methylated genes produces truncated transcripts (Barry *et al*., 1993).

Gene silencing mutants have been isolated in Arabidopsis. TGS is impaired in the *som*, *hog1* and *ddm1* mutants; these mutants reduce DNA methylation at the transgene locus and throughout the genome (Scheid *et al*., 1998; Furner *et al*., 1998). This indicates that these mutants are impaired in DNA methylation, and that methylation is needed for maintenance of TGS, as shown by Dieguez *et al*., (1998). It should be noted that some *som* mutants are allelic to *ddm1* (Scheid *et al*., 1998). Mutants *sgs1* and *sgs2* are impaired only in PTGS, not TGS, and both show reduced methylation in the transcribed region of the transgenes, but no changes in methylation in centromeric DNA or ribosomal DNA (Elmayan *et al*., 1998). In *sgs1*, methylation of the 35S promoter of the transgene was increased; this may have reduced, but did not prevent transcription (Elmayan *et al*., 1998). These results indicate that the *sgs* mutants are not impaired in the general process of DNA methylation, but specifically in post-transcriptional silencing. This impairment may be upstream of silencing-related methylation, or a methylation process specific to silencing may be impaired (Elmayan *et al*., 1998).

The epigenetic phenomenon of paramutation also shows links with chromatin remodelling and cytosine methylation. Paramutation is a heritable alteration of one endogenous (paramutable) allele induced by the presence of another endogenous (paramutagenic) allele (reviewed in Matzke *et al*., 1996; Matzke and Matzke, 1998; Wolffe and Matzke, 1999). Differing results were seen for two paramutable loci in maize. Paramutation of alleles at both the R and B loci was associated with transcriptional inactivation; however paramutation at the multigenic R locus was

associated with increased cytosine methylation (Walker, 1998) while at the B locus, no changes in methylation were seen (Patterson *et al*., 1993). Patterson *et al*. (1993) suggested that paramutation at the B locus involved heritable chromatin remodelling. Deletion of a region of DNA at the R locus containing the "sigma" sequence, derived from the doppia transposon, prevented methylation (Walker, 1998) and reduced the paramutagenicity of the allele (Kermicle, 1996). This suggests that the presence of TE sequences in endogenous genes may provoke a defence response leading to paramutation (Walker, 1998; Matzke and Matzke, 1998).

The link between cytosine methylation and chromatin structure in the epigenetic phenomena described above is supported by the recent identification of a new class of plant cytosine methyltransferases containing a chromo domain (Henikoff and Comai, 1998; Genger *et al*., 1999). Chromo domains are stretches of 30-50 amino acids conserved in many eukaryotic proteins which bind to chromatin and maintain condensed chromatin states, including the *Drosophila* Heterochromatin Protein 1 (HP1) and Polycomb group (PcG) proteins, mammalian homologues of these proteins, and the yeast SNF2/SWI2 proteins (reviewed in Koonin *et al*., 1995; Cavalli and Paro, 1998). Most of these proteins are involved in transcriptional down-regulation, but a number of exceptions activate gene expression (Koonin *et al*., 1995). The Arabidopsis *DDM1* gene encodes a SWI2/SNF2-like protein, indicating that chromatin remodelling is an essential factor in the maintenance of normal cytosine methylation patterns (Jeddeloh *et al*., 1999). The pleiotropic phenotype of the Arabidopsis *curly leaf (clf)* mutant (Goodrich *et al*., 1997) is reminiscent of the phenotype of Arabidopsis methyltransferase antisense transgenics

(Finnegan *et al*., 1996; Ronemus *et al*., 1996), and the *CLF* gene encodes a protein with homology to the Drosophila Polycomb-group proteins which regulate homeotic gene expression through chromatin remodelling (Goodrich *et al*., 1997). These links suggest that cytosine methylation acts together with chromatin remodelling proteins to silence repeated sequences including TEs, transgenes and endogenous sequences which may contain TEs. In animals, the transcriptional repressor MeCP2 binds methylated cytosine residues; silencing resulting from methylation and MeCP2 binding can be reversed by inhibition of histone deacetylase, allowing remodelling of chromatin (Jones *et al*., 1998b). MeCP2 co-fractionates with histone deacetylase and the repressor Sin3, suggesting that these three proteins form a complex binding to methylated DNA and maintaining chromatin structure that represses transcription through histone deacetylation (Jones *et al*., 1998b). Similar mechanisms could operate in plants.

1.7.5 Cytosine methylation and genomic imprinting

In mammals, cytosine methylation regulates the expression of imprinted genes, for which only one allele is expressed, depending on the parental origin (reviewed in Jaenisch, 1997). Methylation is essential for correct imprinting (Li *et al*, 1993; Beard *et al*., 1995), and disruption of imprinting can result in developmental abnormalities or cancer (Reik, 1989; Brenton *et al*., 1995). In angiosperm plants, imprinting affects development of endosperm, which contains one copy of the paternal genome and two copies of the maternal genome. Translocation experiments in maize have shown that the long arms of both chromosome 1 and of chromosome 10 must be derived from the paternal genome if

the endosperm is to develop normally (Birchler, 1979; Lin, 1982). Three maize genes that may be imprinted in endosperm have been identified: α-tubulin, the *dzr* gene, which regulates accumulation of the zein seed storage protein and the *r* gene, which encodes a transcription factor that regulates anthocyanin biosynthesis. In endosperm DNA, the maternal alleles of tubα3 and tubα4 are demethylated, while the paternal alleles are hypermethylated. These methylation changes may correlate with expression levels of specific alleles, though this has not yet been shown (Lund *et al*., 1995). The *dzr* allele derived from inbred maize line MO17 decreases zein accumulation, and is dominant when inherited from the female parent, and recessive when inherited from the paternal parent (Chaudhuri and Messing, 1994). The mode of silencing of the paternal allele is not known. In maize, certain alleles of the R gene are expressed in the endosperm, resulting in uniform pigmentation, only if inherited from the maternal parent. Paternal inheritance of the R gene results in a mottled phenotype, even if two paternal copies are transmitted, indicating that this is not a gene dosage effect (Kermicle, 1970). This paramutation phenomenon is associated with changes in cytosine methylation (Ronchi *et al*., 1995; Eggleston *et al*., 1995). These observations on the behaviour of these three genes indicate some form of genomic imprinting which could involve methylation.

An Arabidopsis gene which is imprinted in endosperm has been the subject of detailed study. Mutations at the *MEDEA* (*MEA = FIS1*; Luo *et al*., 1999) gene allow endosperm production without fertilization; fertilization with wildtype pollen produces seeds showing delayed morphogenesis, excessive embryo cell proliferation and prolonged endosperm nuclear proliferation; these seeds eventually abort (Chaudhury *et al*., 1997;

Grossniklaus *et al*., 1998; Kiyosue *et al*., 1999). This indicates that *MEA* encodes a regulator of cell proliferation. The *MEA* gene codes for a SET domain Polycomb protein (Grossniklaus *et al*., 1998; Kiyosue *et al*., 1999; Luo *et al*., 1999). In mammals, fungi and insects, SET domain Polycomb proteins have been shown to repress transcription of homeotic genes through chromatin remodelling, thus reducing cell proliferation (Pirotta, 1998); the observed *mea* phenotype suggests that *MEA* functions in a similar role in Arabidopsis. The mutant phenotype is only seen if the *mea* mutation is inherited from the maternal line; if *mea* is transmitted through pollen, normal seeds develop (Chaudhury *et al*., 1997; Grossniklaus *et al*., 1998; Kiyosue *et al*., 1999). In wildtype plants, the *MEA* maternal allele is expressed in both the embryo and the endosperm at all stages observed, while the paternal allele is silenced in the endosperm, but expressed in the embryo and other tissues (Kinoshita *et al*., 1999; Ville-Calzada *et al*., 1999). These observations indicate that *MEA* is an imprinted gene, expressed in the endosperm only from the maternal allele. The nature of the imprint is indicated by the observation that seeds carrying a maternal *mea* mutant allele and a paternal *MEA* allele survive if they are homozygous for the *ddm1* mutation (Ville-Calzada *et al*., 1999). This implies that seed survival is permitted by re-activation of the paternal *MEA* allele due to lack of *DDM1* activity (Ville-Calzada *et al*., 1999). *DDM1* encodes a SWI2/SNF2-like protein likely to be active in chromatin remodelling (Jeddeloh *et al*., 1999), and the *ddm1* mutant shows a loss of cytosine methylation (Vongs *et al*., 1993), so the imprint is likely to involve chromatin remodelling and possibly also changes in cytosine methylation.

1.7.6 Endogenous genes regulated by methylation

Cytosine methylation has been implicated in regulating the expression of other endogenous plant genes. The abnormal phenotype and disruption to cytosine methylation caused by methyltransferase (*METI*) antisense expression in Arabidopsis plants was accompanied by ectopic expression of several floral organ identity genes. *APETALA3* (*AP3*) and *AGAMOUS* (*AG*), which are normally only expressed in flowers, were expressed at low levels in leaves (Finnegan *et al*., 1996). The *SUPERMAN* (SUP) gene is normally expressed in early floral development in the third whorl, acting to maintain a boundary between the stamen and carpel whorls (Sakai *et* al., 1995). In *METI* antisense plants, *SUP* was ectopically expressed in the carpel, and plants which bore abnormal flowers similar to those seen on *sup* and *sup ag* mutants did not express *SUP* and *AG*, respectively, in developing buds (H. Sakai and E.J. Finnegan, cited in Finnegan *et al*., 1998a). Jacobsen and Meyerowitz (1997) found that in *METI* antisense plants, the *SUP* sequence is hypermethylated in both symmetrical and non-symmetrical sites, and this hypermethylation was associated with repression of *SUP* transcription. Seven unstable *sup* epi-alleles, designated *clark kent* (*clk*), also show hypermethylation at the *SUP* locus (Jacobsen and Meyerowitz, 1997). The repression of *AG* expression in *METI* antisense plants is also associated with hypermethylation at the *AG* locus (Jacobsen *et al*., 2000). It is possible that the genome-wide hypomethylation in *METI* antisense plants may have deregulated normal controls of methylation patterns, allowing localized hypermethylation; alternatively, ectopic expression of *SUP* and *AG* may have prompted silencing.

Two independent lines of the late flowering *fwa* mutant show hypomethylation of a region of at least 5 Mb spanning the *FWA* locus (W. Soppe, personal communication). *FWA* is expressed in the mutant, but not in wildtype plants, indicating that it acts as a suppressor of flowering (W. Soppe, personal communication). The sequence of the *FWA* gene is identical for wildtype and *fwa-1* mutant alleles. Bisulphite sequencing revealed that two repeats in the promoter region of *FWA* are methylated in the wildtype allele, but not in the mutant allele. This strongly suggests that hypermethylation of the repeats in the promoter is preventing expression of the wildtype *FWA* allele; hypermethylation at other sites around this locus may also play a part in silencing the *FWA* gene. A heritable late flowering epimutation, isolated in the *ddm1* mutant, maps to the *FWA* locus and shows expression of *FWA* (Kakutani, 1997; W. Soppe, personal communication). This epimutant is likely to be associated with hypomethylation of the *FWA* locus and surrounding regions.

Expression levels of a family of genes encoding phosphoribosylanthranilate isomerase (PAI), an enzyme in the tryptophan biosynthetic pathway, are regulated by methylation. Seven Arabidopsis ecotypes, including Wassilewskija (Ws), contain four *PAI* genes at three unlinked loci: two genes in a tail-to-tail inverted repeat (*PAI1*-*PAI4*), plus 2 single genes (*PAI2*, *PAI3*) (Bender and Fink, 1995; Melquist *et al*., 1999). All four copies show dense methylation in homologous coding regions (Luff *et al*., 1999; Melquist *et al*., 1999). Other ecotypes, including Columbia (Col), contain three single genes (*PAI1*, *PAI2*, *PAI3*) that show no methylation (Luff *et al*., 1999; Melquist *et al*., 1999). This

indicates that, as seen in transgene silencing, inverted repeats can trigger methylation (Stam *et al*., 1998; Luff *et al*., 1999; Mette *et al*., 1999).

A spontaneous Ws mutant in which the *PAI1-PAI4* inverted repeat was deleted showed a tryptophan deficient phenotype due to insufficient expression of the methylated *PAI2* and *PAI3* genes. The methylation pattern and lack of expression of these two genes was heritable even when segregating in crosses to an ecotype with unmethylated *PAI* genes. The mutant phenotype was unstable, and occasional revertants were produced, which showed hypomethylation and increased expression of the *PAI2* and *PAI3* genes (Bender and Fink, 1995). Introduction of the *ddm1* mutation into the *pai1-pai4* deletion mutant background led to hypomethylation and increased expression of *PAI2* and *PAI3* at high frequency (Jeddeloh *et al*., 1998). Introduction of the *PAI1-PAI4* inverted repeat from Ws into the Col background led to methylation of the three previously unmethylated Col *PAI* genes, suggesting that DNA methylation may be transmitted by a DNA-DNA pairing mechanism (Luff *et al*., 1999).

1.7.7 Cytosine methylation and the vernalization response

The modification of cytosine residues by methylation provides a mechanism for storing additional information in the base sequence of DNA. A number of the features of cytosine methylation suggest that modifications of the methylation pattern could store and transmit the information that the plant has been vernalized. Methylation patterns, at least in symmetrical sites, are inherited mitotically (Razin and Riggs, 1980; Stein *et al*.,

1982), as is the vernalization signal. This process of mitotic inheritance, or maintenance methylation, is proposed to occur by the action of DNA methyltransferases on unmethylated cytosines in the daughter strand of newly replicated, and thus hemi-methylated, DNA (Holliday and Pugh, 1975; Riggs, 1975; Razin and Riggs, 1980). Maintenance methylation occurs following DNA replication, and thus passive changes in methylation can occur at cell division if maintenance methylation does not occur; this parallels the observation that only cells which are dividing during the cold treatment are capable of showing a vernalization response (Wellensiek, 1964; Metzger, 1988a; Burn *et al*., 1993a). Unlike the vernalization signal, methylation patterns can, in at least some cases, be inherited through meiosis (Vongs *et al*., 1993; Finnegan *et al*., 1996; Jacobsen and Meyerowitz, 1997; Kakutani *et al*., 1999); this indicates that another mechanism, perhaps involving targeted methylation or demethylation, must operate to reset the vernalization signal.

Cold treatment, whether of whole plants or cell cultures, leads to a decrease in cytosine methylation levels (Burn *et al*., 1993a; Finnegan *et al*., 1998b); normal methylation levels are restored when plants are returned to higher temperatures (Finnegan *et al*., 1998b). Application of the demethylating agent 5-aza-cytidine induces early flowering in vernalization responsive ecotypes and late flowering mutants of Arabidopsis, but not vernalization unreponsive late flowering mutants (Burn *et al*., 1993a). These results strengthen the hypothesis that the vernalized state could be transmitted by changes in methylation. As discussed above, methylation is generally correlated with gene inactivation, so it is possible that reduced methylation following cold treatment could

allow activation of a gene critical in the pathway to initiation of flowering. However, since there is a time lag before initiation of flowering after vernalization, such a "critical" gene could only be promotive of flowering, rather than directly causing floral initiation.

1.8 Cytosine methyltransferases

1.8.1 Prokaryote cytosine methyltransferases

Cytosine methyltransferases catalyze the transfer of methyl groups from S-adenosyl-L-methionine (AdoMet) to the carbon 5 cytosine residues within DNA. Prokaryote methyltransferases show a high degree of conservation of structure. All prokaryote methyltransferases so far identified possess 10 conserved motifs, separated by regions more variable in length and sequence (Posfai *et al*., 1989; Lauster *et al*., 1989; Kumar *et al*., 1994). In most prokaryote methyltransferases the order of conserved motifs is conserved, but exceptions include M.*Aqu*I, in which the enzyme is encoded by two open reading frames (ORFs); both peptides encoded by these ORFs are needed for methyltransferase activity (Karreman and Waard, 1990), and M.*Bss*HII, where a circular permutation of the more usual motif arrangement has resulted in motifs IX and X preceding motifs I to VIII (Xu *et al*., 1997). The function of several conserved motifs has been established. Motif IV contains an absolutely conserved proline-cysteine doublet identified as part of the active site (Chen *et al.*, 1991; Wyszynski *et al.*, 1992; Mi and Roberts, 1993). Motifs I and X bind the methyl donor, AdoMet (Ingrosso *et al*., 1989). The variable region between motifs VIII and IX encodes the target recognition domain

(TRD), which directs the enzyme to the correct cytosine residue within its recognition site (Balganesh *et al.*, 1987; Klimasauskas *et al.*, 1991; Mi and Roberts, 1992).

The structures of two prokaryote cytosine methyltransferases, M.*Hha*I and M.*Hae*III, have been determined by X-ray crystallography (Cheng *et al*., 1993; Klimasauskas *et al*., 1994; Reinisch *et al*., 1994, 1995). The M.*Hha*I enzyme folds into a large and a small domain, separated by a hinge region; this forms a cleft in which the DNA substrate binds (Cheng *et al*., 1993; Klimasauskas *et al*., 1994). The first half of motif IX and the last half of motif X fall in the small and large domains respectively; the hinge region contains the remainder of these two motifs. The large domain contains the binding site for the methyl donor, AdoMet; it contains motif I, and residues from motifs II, III and X also interact with AdoMet (Cheng *et al*., 1993). The TRD forms most of the small domain; as this domain forms one side of the DNA-binding cleft, it is well placed to function in specific DNA sequence recognition (Cheng *et al*., 1993). The target cytosine is flipped out of the DNA helix, bringing it into close proximity with the catalytic cysteine residue in motif IV and AdoMet (Klimasauskas *et al*., 1994). A similar structure is seen for M.*Hae*III complexed with DNA (Reinisch *et al*., 1994, 1995). One notable difference is that, whereas only the pairing of the G-C base pair containing the target cytosine is disrupted when M.*Hha*I binds (Klimasauskas *et al*., 1994), all bases in the bound M.*Hae*III recognition sequence are unpaired, opening a gap in the DNA helix (Reinisch *et al*., 1995). Studies of M.*Hha*I-DNA interactions in solution indicate that the movement of the target cytosine out of the helix occurs after binding of the enzyme to the DNA helix, but is not dependent on binding of the cytosine in the catalytic pocket of the

enzyme (Klimasauskas *et al*., 1998). This suggests that the enzyme has an active role in opening the DNA helix; different enzymes may open the helix to a greater or lesser extent.

1.8.2 Mammalian cytosine methyltransferases

The first mammalian cytosine methyltransferase gene isolated was the mouse *Dnmt1* gene (Bestor *et al*., 1988). Homologues of *Dnmt1* have since been isolated from human (Yen *et al*., 1992) and rat (Deng and Szyf, 1998). Eight of the ten motifs conserved in prokaryote methyltransferases are also present in *Dnmt1*, and the order of these motifs is the same as that usually seen in prokaryote methyltransferases (Bestor *et al*., 1998). The *Dnmt1* gene also encodes a large amino-terminal domain not seen in any prokaryote methyltransferases. Motifs within this domain target the enzyme to the DNA replication fork within the nucleus during S phase (Leonhardt *et al*., 1992; Liu *et al*., 1998). Bestor and Ingram (1983) showed that Dnmt1 had a 30-50 fold preference for hemi-methylated DNA over unmethylated DNA. Cleavage of the amino-terminal domain allows the methyltransferase domain to methylate DNA *de novo* at a high rate, without altering the methylation rate for hemi-methylated DNA, indicating that the amino-terminal domain inhibits *de novo* methylation (Bestor, 1992). These results indicate that Dnmt1 would effectively maintain cytosine methylation patterns in CpG dinucleotides by methylating cytosines in daughter strands of newly replicated DNA.

Mouse embryonic stem (ES) cells with a null mutation in the *Dnmt1* gene show a stable low level of cytosine methylation, and integrated provirus DNA is methylated at a rate similar to that seen in wildtype ES cells (Lei *et al*., 1996). These results indicated the existence of a second mouse cytosine methyltransferase acting to *de novo* methylate DNA. Two methyltransferase genes, *Dnmt3a* and *Dnmt3b*, have been shown to encode enzymes with *de novo* methylation activity. These enzymes lack the amino-terminal domain (Okano *et al*., 1998a). A third methyltransferase-like gene, *Dnmt2*, also lacks an amino-terminal domain. *Dnmt2* contains the eight conserved motifs possessed by other eukaryote methyltransferases, and is expressed in all tissues examined (Yoder and Bestor, 1998). However, *dnmt2* null mutant ES cells show no changes in *de novo* or maintenance methyltransferase activity, and the Dnmt2 protein shows no methyltransferase activity (Okano *et al*., 1998b).

1.8.3 Other eukaryote methyltransferases

Dnmt1-like genes have been isolated from a number of other eukaryotes, including *Xenopus laevis* (Kimura *et al*., 1996), sea urchin (Aniello *et al*., 1996), chicken (Tajima *et al*., 1995), the fungus *Ascobolus immersus* (Malagnac *et al*., 1999), and a number of plants (discussed below). All these genes encode a large amino-terminal domain and a smaller methyltransferase domain that includes the eight conserved motifs seen in *Dnmt1* and other mouse methyltransferases. A gene essential for *de novo* methylation, *Masc1*, has been isolated from *Ascobolus immersus*; as discussed above, this gene is essential for normal sexual reproduction (Malagnac *et al*., 1997).

Methyltransferase-like genes have been isolated in fission yeast (*Schizosaccharomyces pombe*) and *Drosophila melanogaster*, two eukaryote organisms which lack detectable cytosine methylation (Wilkinson *et al*., 1995; Pinarbasi *et al*., 1996; Hung *et al*., 1999). This suggests that ancestors of these organisms possessed enzymes with the capacity to methylate DNA, but this function has been lost.

1.8.4 Plant cytosine methyltransferases

The Arabidopsis *METI* gene was the first plant cytosine methyltransferase gene cloned (Finnegan and Dennis, 1993). *METI* contains the eight motifs conserved in all eukaryote methyltransferases, and its structure is similar to mouse *Dnmt1*. Amino acid homology between *METI* and *Dnmt1* is 50% in the methyltransferase domain, and 24% in the amino-terminal domain (Finnegan and Dennis, 1993). Southern analysis of genomic DNA revealed that Arabidopsis contains a small family of *METI*-like genes (Finnegan and Dennis, 1993).

Two cytosine methyltransferases have been identified in carrot. The predicted proteins are more than 85% identical through most of their length, but diverge at the amino-terminus where Met2-21 shows five nearly perfect repeats of a 57 residue sequence present only once in Met1-5 (Bernacchia *et al*., 1998a). Both genes are expressed most strongly in proliferating cells, but *Met2-21* is expressed at much lower levels that *Met1-5*. Both genes are expressed in the vegetative shoot apex and in embryos, but the expression

of *Met2-21* is localized to specific regions, while *Met1-5* is expressed more evenly throughout the plant. These differences suggest that the two genes perform different functions (Bernacchia *et al*., 1998a). A cDNA homologous to *METI* has been isolated from a tomato apical meristem library; this gene shows greater homology to *Met2-21* than to *Met1-5*, despite lacking the five repeats found in the N-terminus of Met2-21 (Bernacchia *et al*., 1998b). Two *METI*-like genes have been identified in maize; these genes may be orthologous, as the maize genome is tetraploid (Kaepplar and Springer, cited in Finnegan and Kovac, 2000). Although two methyltransferase enzymes, with activity on CpG and CpNpG sequences respectively, have been isolated from pea (Pradhan and Adams, 1995), only one methyltransferase gene has so far been identified in pea (Pradhan *et al*., 1998). The expressed protein can methylate cytosines in CpG and CpNpG *in vitro* (Pradhan *et al*., 1998), suggesting that it may fulfill both functions *in vivo*, possibly after post-transcriptional or post-translational modifications.

Disruption of activity of the Arabidopsis *METI* gene by expression of a *METI* antisense construct lowered methylcytosine levels by as much as 90%, predominantly in CpG dinucleotides, suggesting that *METI* encodes a CpG methyltransferase (Finnegan *et al*., 1996; Ronemus *et al*., 1996). As Arabidopsis contains a small family of *METI*-like sequences, it is possible that one of these encodes a CpNpG methyltransferase, or a methyltransferase which performs a function distinct from that of *METI*.

A second class of plant cytosine methyltransferases are the chromomethylases (CMT), in which a chromo domain is inserted between methyltransferase motifs II and IV (Henikoff

and Comai, 1998). As mentioned above, chromo domains occur in many eukaryotic proteins that bind heterochromatin, most of which are involved in transcriptional silencing (reviewed in Cavalli and Paro, 1998). Arabidopsis contains at least three members of the chromomethylase class (Genger *et al*., 1999; Henikoff and Comai, cited in Finnegan and Kovac, 2000). Chromomethylases contain the 8 motifs conserved in all other eukaryote cytosine methyltransferases (Genger *et al*., 1999), but the first chromomethylase to be cloned, *CMTI*, does not possess the large amino-terminal domain characteristic of many eukaryotic methyltransferases (Henikoff and Comai, 1998; Genger *et al*., 1999). *CMTI* is expressed in vegetative and floral tissues, but is not an essential gene, as four of the thirteen Arabidopsis ecotypes surveyed do not produce the intact protein. In three ecotypes, insertion of a retrotransposon, Evelknievel, results in production of a truncated protein lacking the catalytic region; the fourth ecotype expresses mostly aberrant mRNAs, again leading to the production of a truncated protein lacking the catalytic region (Henikoff and Comai, 1998). The *CMTII* and *CMTIII* genes are expressed, and levels of *CMTII* expression are higher than either *CMTI* or *CMTIII* (Genger *et al*, 1999; Henikoff and Comai, cited in Finnegan and Kovac, 2000).

1.8.5 Cytosine demethylation

Cytosine methylation levels and patterns are set through interactions between the processes of methylation and demethylation. In mammals, genome wide demethylation occurs during preimplantation embryogenesis, with only imprinted genes and some repeat sequences retaining 5mC residues; post-blastulation, *de novo* methylation occurs

throughout the genome, such that only CpG island sequences remain unmethylated (Monk *et al.*, 1987; Kafri *et al.*, 1992; Razin and Kafri, 1994). Changes in methylation levels are seen at different stages in plant development also; methylation levels in seeds are high relative to mature tissues and pollen (Messeguer *et al.*, 1991; Palmgren *et al*., 1991), and methylation levels in the shoot increase with age (Banks and Federoff, 1989; Messeguer *et al.*, 1991; Palmgren *et al*., 1991).

Demethylation could occur passively, if maintenance methylation failed to occur after DNA replication. A passive mechanism is supported by some observations, such as the staged decrease in binding of antibodies to 5-methylcytosine during the genome-wide loss of methylation in preimplantation mouse development (Rougier *et al*., 1998). However, evidence for active demethylation has come from observations of changes in the methylation pattern within specific genes. For example, Paroush *et al*. (1990) observed demethylation of cytosines on both strands of the rat α-actin gene transfected into L8 myoblast cells in the absence of DNA replication. After one division cycle in mouse embryos (from the 8-cell to 16-cell stage), full methylation at sites in the ApoAI gene was replaced by complete demethylation (Shemer *et al*., 1991). These experiments indicate that an active demethylation mechanism occurs in mammals; by analogy such a mechanism may occur in plants also.

Several groups have demonstrated active demethylation of double-stranded oligonucleotides by glycosylase activity; in this reaction, the 5-methylcytosine is excised from the DNA, and replaced with cytosine by DNA repair enzymes (Jost, 1993; Jost *et*

al., 1995, 1997, 1999; Fremont *et al*., 1997; Vairapandi and Duker, 1993, 1996). Demethylation of hemi-methylated CpG sites by 5-methylcytosine (5-MeC) glycosylases extracted from mouse and chicken cells was sensitive to ribonuclease activity, suggesting a role for RNA in the demethylation process (Fremont *et al*., 1997; Jost *et al*., 1997, 1999). Both the chick embryo 5-MeC glycosylase and RNA were required for demethylation (Jost *et al*., 1997). The RNA present in the purified 5-MeC glycosylase was heterogeneous, and rich in CpGs; these RNAs may target the 5-MeC glycosylase to CpG islands from which they were transcribed (Jost *et al*., 1997). However, the involvement of RNA in the demethylation reaction is controversial. Work by Swisher *et al*. (1998) indicates that the demethylase activity isolated from rat myoblasts is sensitive to protein digestion, but insensitive to ribonuclease, overturning previous suggestions that that the rat demethylase activity was due to a RNA enzyme (Weiss *et al*., 1996). These two demethylase activities may represent different mechanisms of demethylation (Jost *et al*., 1999; Swisher *et al*., 1998). The chicken 5-MeC glycosylase acts more rapidly on hemimethylated DNA (Jost *et al*., 1995), whereas a human 5-MeC glycosylase activity prefers fully methylated over hemi-methylated substrate (Vairapandi and Duker, 1993); the rat 5-MeC glycosylase activity was assayed only on fully methylated DNA (Swisher *et al*., 1998). These different preferences may also indicate different mechanisms of action. Further purification and characterization of the demethylase activities, and eventual cloning of the genes encoding them, will allow comparison of these mechanisms.

It should be noted that recent reports of a mammalian protein with demethylase activity (Ramchandani *et al*., 1999; Bhattacharya *et al*., 1999) have not been confirmed independently (A.P. Bird and A.P. Wolffe, personal communication).

The methylation pattern set in plant, animal and fungal genomes may result from the overlapping action of methyltransferases and demethylases, whether glycosylases or other enzymes, as well as passive demethylation processes. While the process of methylation is relatively well understood, and the demethylation mechanism is now being elucidated, the way in which the specificity of each process is controlled is not clear, and forms a major research question for the future.

1.9 Scope of thesis

This thesis reports an investigation into the role of cytosine methylation in the vernalization response of *Arabidopsis thaliana*. Chapter Three begins with observations of the flowering time of Arabidopsis plants from the vernalization responsive ecotype C24, transformed with an antisense construct for the cytosine methyltransferase *METI* (Finnegan and Dennis, 1993). The early flowering phenotype of different transgenic families is attributable to the extent of DNA demethylation, indicating a role for methylation in flowering time control, and possibly in the vernalization response. Mutants at the *DDM1* locus (decreased DNA methylation; Vongs *et al*., 1993) also show an early flowering phenotype in conditions which allow a vernalization response, supporting the hypothesis of a role for methylation in the vernalization process. This

work was published in the Proceedings of the National Academy of Sciences of the USA under the title "DNA methylation and the promotion of flowering by vernalization" (Finnegan *et al*., 1998b).

In Chapter Four, the effect of demethylation on the flowering time of 2 late flowering mutants, *fca* and *fe*, which differ in vernalization responsiveness, is examined. DNA demethylation in plants of the vernalization responsive *fca* mutant causes early flowering. In plants of the *fe* mutant, which is only slightly responsive to vernalization, DNA demethylation does not lead to early flowering. The results reported in this chapter also demonstrate the importance of a gene segregating between two Arabidopsis ecotypes to vernalization responsiveness and flowering time, and support the involvement of the FLF floral suppressor (Sheldon *et al*., 1999) in the vernalization response. A paper reporting this work is in preparation.

Chapter 5 explores the role of GA in the vernalization response of Arabidopsis, and possible interactions with cytosine methylation levels. In the conditions used, exogenous GA, demethylation and vernalization had additive effects in reducing time to flowering. Observations of the flowering time of the progeny of crosses between *METI* antisense transgenics and the GA-insensitive mutant *gai* suggest that the promotive effect of demethylation on flowering may act partly through activation of the GA signal transduction pathway. A paper reporting this work is in preparation.

In Chapter 6, the characterization of the sequence and expression pattern of a second *METI*-like Arabidopsis cytosine methyltransferase gene is described. This gene is not linked to *METI*, and although highly similar to *METI* is expressed in *METI* antisense transgenics. This research was published in Plant Molecular Biology under the title "Multiple DNA methyltransferase genes in *Arabidopsis thaliana*" (Genger *et al*., 1999).

In Chapter Seven, the general conclusions drawn from the work reported in this thesis are brought together, and possible future directions are discussed.