RESEARCH COMMUNICATION

Effect of CWG methylation on expression of plant genes

Sriharsa PRADHAN[1], Nigel A. R. URWIN[2], Gareth I. JENKINS and Roger L. P. ADAMS[3]

Division of Biochemistry and Molecular Biology, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, Scotland, U.K.

The presence of two DNA methyltransferases in *Pisum* raises the possibility that they serve different functions. *In vitro* methylation of CWG sequences in the strong cauliflower mosaic virus *35S* promoter had no effect on reporter gene expression. In contrast, *in vitro* methylation of CWG sequences in the relatively weak, CG-deficient *Phaseolus vulgaris rbcS2* promoter inhibited transcription. Expression of both constructs was strongly inhibited by extensive CG methylation. A search of published plant promoter sequences revealed that the CG content of promoters is very variable, with some promoters having typical CG islands. In contrast, the distribution of CWG sequences is more even with little evidence for CWG islands.

Key words: CG island, *in vitro* methylation, methylcytosine, methyltransferase.

INTRODUCTION

The methylation of CG dinucleotides is well known in the vertebrate genome, but the genomes of higher plants contain additional methylcytosine in the trinucleotide sequence mCNG, where N represents any of the four common DNA bases [1]. Since there is little depletion of the CG dinucleotides, and even less depletion of CNG trinucleotides in plants [2], this results in methylcytosine representing as much as one-third of the cytosine bases in plant DNA.

Little is known about the localization or function of methylcytosine in plant DNA although, as found with animals, methylation of CG dinucleotides can interfere with transcription factor binding [3]. Although Antequera and Bird [4] have shown that, as with vertebrates, angiosperm genomes contain clusters of unmethylated CG dinucleotides (CG islands), the resistance of plant DNA to cleavage by restriction enzymes such as *Hpa*II indicates that the majority of the CG dinucleotides are methylated and are spread throughout the genome [5]. In a similar manner, restriction enzyme studies have shown that most cytosines in CWG sequences (where W is A or T) are methylated, although there is controversy concerning methylation of the 5′-cytosine in CCG sequences [5].

We have reported previously the purification of separate CG and CWG DNA methyltransferases from *Pisum sativum* [5], raising the possibility that the action and control of these enzymes might be different, even to the extent that methylation of CG and CWG sequences might serve completely different functions. Herskovitz et al. [6] have shown that the cauliflower mosaic virus (CaMV) *35S* gene promoter can be silenced in transient expression assays by methylation of all CG dinucleotides. It is therefore of interest to identify whether CWG methylation is also able to silence gene expression in a similar manner.

In this paper we show that transcription from the *Ph. vulgaris rbcS2* promoter, rich in CWG sequences, is inhibited by CWG methylation, but that expression from the *35S* promoter is resistant to CWG methylation.

MATERIALS AND METHODS

Chimaeric promoter chloramphenicol acetyltransferase (CAT) constructs

pCAMVCN (Pharmacia) is a 35S–CAT construct containing 400 bp of the CaMV *35S* gene promoter fused to the coding region of the CAT gene of Tn9 and the termination sequence of the nopaline synthase (NOS) gene.

The rbcS–CAT construct contained a promoter fragment (-1433 bp to $+26$ bp) of the bean *rbcS2* gene fused to the CAT coding sequence and the NOS terminator of pCAMVCN [7].

In vitro methylation of reporter gene constructs

Prokaryotic DNA methyltransferases were used according to the manufacturer's instructions. The pea CWG methyltransferase was purified as described previously [5], and was used to methylate plasmid DNA by incubation overnight at 30 °C in the presence of 0.5 mM *S*-adenosylmethionine. The DNA was re-purified, and the amounts were quantified using a microfluorimeter (Hoefer Scientific Instruments, San Francisco, CA, U.S.A.) before use in transfection.

Plant growth and protoplast isolation

This was performed essentially as described by Urwin and Jenkins [7]. *Ph. vulgaris* L cv. Tendergreen plants were grown for 10–15 days in continuous white light provided by warm white fluorescent tubes at a fluence rate of 10 μmol·m^{-2}·s^{-1}. Plants were transferred to darkness for 2 days (dark-adapted). All further steps prior to incubation of protoplasts were performed under a green safe light. The upper surfaces of primary leaves

Abbreviations used: CaMV, cauliflower mosaic virus; CAT, chloramphenicol acetyltransferase; NOS, nopaline synthase.

[1] Present address: New England Biolabs, Tozer Road, Beverly, MA 01915, U.S.A.

[2] Present address: Farrer Centre, Faculty of Agriculture and Science, Charles Sturt University, Wagga Wagga, NSW 2678, Australia.

[3] To whom correspondence should be addressed, at the Davidson Building, University of Glasgow, Glasgow G12 8QQ, U.K. (e-mail r.adams@bio.gla.ac.uk).

were rubbed with alumina, which was subsequently removed by washing thoroughly in distilled water. Leaves were floated (upper-side down) on 0.8% (w/v) cellulase R10, 0.25% (w/v) macerozyme R10 (both from Yakult Honsha, Nishinomiya, Japan) in culture buffer (0.4 M mannitol with CPW salts [8], and 50 $\mu g \cdot ml^{-1}$ ampicillin, pH 5.7) for 20 h at 22 °C in the dark. Protoplasts were released from leaves by brief, gentle agitation, and cells were filtered through one layer of muslin and a steel sieve (70-μm pore size). Protoplasts were allowed to pellet under gravity on ice for 1–1.5 h, after which they were washed twice in cold 0.4 M mannitol/CPW salts, pH 5.7.

Transient expression assay

Protoplasts were resuspended in electroporation buffer [0.4 M mannitol/4 mM Mes (pH 5.7)/20 mM KCl] at $(1–3) \times 10^6$ cells $\cdot ml^{-1}$. Aliquots of 0.5 ml were dispensed into cuvettes on ice, and 25 μg of plasmid DNA was added in < 30 μl. Cells were electroporated with a single pulse at a capacitance of 100 μF and a field strength of 325 $V \cdot cm^{-1}$. Cells were then returned to ice for 5 min. Aliquots of 200 μl were added to 800 μl of culture buffer, and cells were then incubated in 24-well culture plates for 20 h in white light, as described above, at a fluence rate of 80 $\mu mol \cdot m^{-2} \cdot s^{-1}$. Protoplasts were pelleted at 100 ***g*** for 5 min, and resuspended in 50 μl of 20 mM Tris/HCl, pH 8.0/2 mM $MgCl_2$. After freezing at −80 °C and heating at 65 °C for 15 min, debris was removed by centrifugation at 14000 ***g*** for 5 min. CAT activity was assayed by the method of Seed and Sheen [9] using 40 μl of extract for all constructs, apart from 35S–CAT, where 40 μl of a 1:100 dilution of the extract was used.

RESULTS

Effect of methylation on expression from different promoters

We have investigated the effect of *in vitro* methylation on the transient expression of plasmids containing promoters with differing CG and CWG contents. We have used the CAT reporter gene under the control of either the *CaMV35S* promoter (a typical CG-island promoter) or the bean *rbcS2* promoter (a promoter lacking a CG island, but containing the expected number of CWG sequences). Transient expression was assayed in bean-leaf protoplasts. In order to prevent possible confusion with any pre-existing methylation, the plasmids were grown in an *Escherichia coli* strain (GM2163; New England Biolabs, Beverly, MA, U.S.A.) that was dam$^-$ and dcm$^-$ (i.e. lacking methylation of GATC and CCWGG sequences). Complete plasmids were methylated *in vitro* by using either the purified pea CWG methyltransferase [5] or one of a variety of prokaryotic methyltransferases (M.*Hpa*II, M.*Msp*I or M.*Sss*I) that would lead to the introduction of methyl groups into the following sequences: C^mCGG (M.*Hpa*II); ^{m}CCGG (M.*Msp*I); or ^{m}CG (M.*Sss*I). The prokaryotic CG methyltransferase, i.e. M.*Sss*I, was used in preference to the pea CG enzyme because (i) it is a much more active methyltransferase *de novo* and (ii) it is more readily available. Compared with the *rbcS2* promoter, the *35S* promoter confers a 1000-fold greater level of CAT expression with the control, unmethylated plasmids (i.e. the *rbcS2* is a weak promoter relative to the *CaMV35S* promoter). A diagram of the proximal 425 bp of the promoter regions is shown in Figure 1 (upper panel).

Methylation of CCGG sequences has little effect on gene expression from either promoter (Figure 1, lower panel). This is not surprising, since neither promoter contains CCGG target sites, and what sites there are in the *CAT* gene and the vector are well dispersed. In contrast, methylation of all CGs with M.*Sss*I leads to a very strong inhibition of expression from both promoters. Whether this is a direct effect of promoter methylation or an indirect effect mediated partly by vector methylation [10] is discussed below.

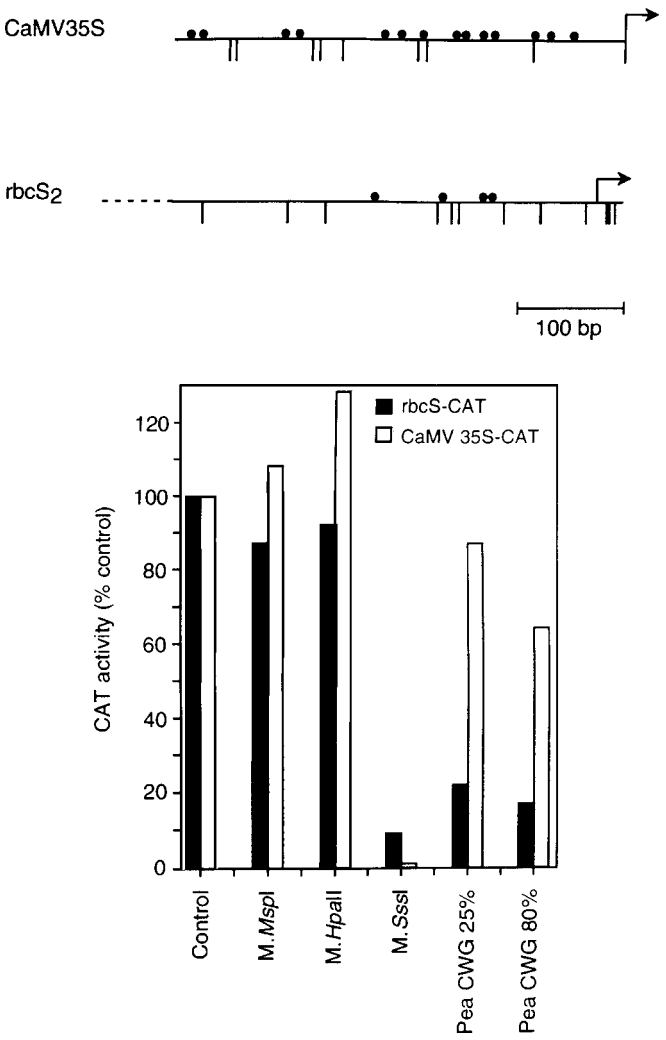

Figure 1 Transient expression of methylated plasmids

The plasmids were grown in dam$^-$, dcm$^-$ bacteria and methylated *in vitro*, as described in the Materials and methods section. They contained the *CAT* reporter gene linked to either the *CaMV35S* promoter (CaMV 35S–CAT) or the bean *rbcS2* promoter (rbcS–CAT). The proximal 425 bp of the promoters is illustrated in the upper panel, where the vertical lines indicate CWGs and the dots indicate CGs. Results of the transient expression studies are shown in the lower panel; the data represent an average of two experiments, each performed in duplicate, and are expressed as a percentage of expression with a mock-methylated control plasmid. Methylation with the pea CWG methyltransferase was carried out for 4 h (25% saturation) or 24 h (80% saturation). In one experiment, the 100% values obtained using the unmethylated *CaMV35S* promoter or the bean *rbcS2* promoter were 8250 c.p.m. and 2040 c.p.m. respectively (averages of duplicate values).

Methylation of CWG sequences leads to a dramatic inhibition of expression from the *rbcS2* promoter, but has only a limited effect on expression from the *35S* promoter (Figure 1, lower panel). Since the plasmids used are very similar, differing substantially only in their promoter regions, this difference must be attributable to promoter methylation.

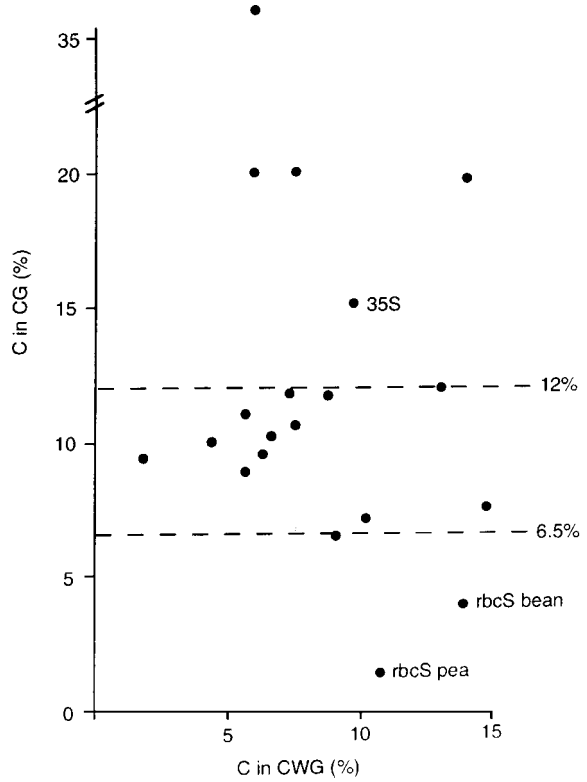

Figure 2 Plant promoter analysis

The Figure shows the percentage of cytosine in CG dinucleotides plotted against the percentage in CWG trinucleotides for 20 plant promoters. Up to 1 kb was analysed upstream of the translation start site. The dotted lines enclose the majority of sequences. Although not labelled specifically in the Figure, the promoter sequences and their respective accession numbers analysed were as follows: tomato phenylammonia lyase (*PAL*, M83314); tomato anionic peroxidase tap-1 (S63739); pea *UBC4* (L39921); pea *rbcs3A* (M21356); pea plastocyanin *PETE* (S66544); *Arabidopsis cdc2a* (U19862); *Arabidopsis* plastocyanin (S67901); *Arabidopsis PAL* (S45847); Maize zein (K00543); *Nicotiana* promoter (X55365); *Nicotiana* seed coat protein (U08931); *Phaseolus* hydroxyproline-rich glycoprotein (U18991); *Phaseolus* glutamine synthetase (S44882); rice actin (S44221); petunia *Adh2* (U25536); petunia chalcone synthase A (S52984); soya-bean leghaemoglobin Iba (X06438); soya-bean auxin-responsive promoter (D11429); *Ph. vulgaris rbcS2* (*AF028707*); and CaMV35S.

Database analysis of plant promoter sequences

We performed an analysis of 40 plant promoter sequences that are present in the GenBank/EMBL database, and Figure 2 shows the relationship between the frequency of occurrence of CG and CWG sequences for a representative subset of these promoters. The proportion of cytosines in CWG trinucleotides varies between 2 and 17%, and appears to be independent of the frequency of occurrence of CG dinucleotides, which, in most cases, is fairly constant, with 6.5 to 12% of cytosines being in CG sequences. The *rbcS* pea and bean promoters are unusual in having a very low CG content, and 12 promoters, including that of the *CaMV35S* gene, have a high CG content.

We have also examined the putative promoter regions of a number of genes on chromosome 4 of *Arabidopsis thaliana* [11] but, whereas a number of CG islands are obvious, these are not associated with clusters of CWG trinucleotides. There is no parallel or reciprocal relationship between the CWG and CG contents of promoters examined.

The *rbcS* promoter is clearly unusual in that it has a deficiency in CG (the CG:GC ratio of the pea *rbcS3A* promoter is < 0.1) and a higher-than-average CWG content. A low CG content is not, however, a characteristic of all *rbcS* promoters, since the tomato *rbcS3A* promoter has 11.5% of its cytosines in CG dinucleotides (accession number S44160). In contrast, the 12 promoters with 15% or more of their cytosines in CG dinucleotides show no deficiency in CG (average CG:GC ratio of 1.1), and a range of values for the proportion of cytosines in CWG trinucleotides that have an average value (9.6%) that is very close to that found for all the promoters (9.3%).

DISCUSSION

Ribulose bisphosphate carboxylase/oxygenase (or Rubisco) is an essential plant enzyme and the expression of the small subunit (RbcS) is controlled by light. The bean *rbcS2* promoter contains a CG dinucleotide in the sequence CACGTG (known as a G-box) that binds the transcription factor GBF, which is implicated in regulation by a variety of environmental signals [2,12] and is required for high levels of expression [7,13]. Southern blot analysis, using a probe stretching from -1433 to -4 bp relative to the transcription start site of the bean *rbcS2* gene, indicates an absence of CG methylation of the G-box sequence in genomic DNA from dark-grown, dark-adapted or light-grown leaves, roots or seeds. This analysis was made possible since we have shown that *Eco*72I (Promega) is sensitive to methylation of the internal cytosine in the sequence CACGTG (results not shown). The pea *rbcS3A* promoter contains only a single CG dinucleotide, and this is not in the G-box, which therefore lacks the CG dinucleotide discussed above [14]. It is therefore very unlikely that inhibition of expression is a result of interference in the binding of transcription factors (e.g. GBF), but it is likely that the very strong inhibitory effect of CG methylation in transient expression assays is mediated via the formation of an inactive chromatin structure [10]. This could also be true for the *35S* promoter, and Diéguez et al. [15] have shown that none of the promoter CG sites are essential for expression, even though their methylation status correlates inversely with expression.

In contrast, the selective effect of CWG methylation on the weak *rbcS2* promoter implies a direct effect on the promoter, and this might well involve inhibition of transcription factor binding. Although the bean *rbcS2* promoter contains very few CG dinucleotides (CG:GC ratio is 0.2), it has very similar levels of CWG sequence relative to the *35S* promoter. As the expression of the 35S–CAT plasmid is resistant to methylation of over 80% of its CWG trinucleotides, it is unlikely that CWG methylation of either the vector or the CAT gene can bring about the formation of inactive chromatin [10], perhaps indicating an alternative role for CWG methylation.

Southern blot analysis indicates that there is no more than a very low level of *in vivo* methylation of the *Eco*RII site (CCWGG) at -850 bp in the *rbcS2* promoter in DNA from all bean plant tissues tested. A similar complete lack of methylation was observed at the *Bal*I site (TGGCCA*G*) at -138 bp that overlaps a CAG trinucleotide (results not shown). Thus although we have shown that, in transient expression assays, both promoters are sensitive to CG methylation, and the *rbcS2* promoter is sensitive to CWG methylation, we have not found a situation where the endogenous bean *rbcS2* promoter is methylated, despite examining tissues that show a wide range of expression. This is in contrast with the $> 90\%$ methylation of *Pst*I sites (CTGCAG) and *Hpa*II sites (CCGG) in total pea DNA [5].

We acknowledge the assistance of the British Commonwealth Commission for a scholarship to S. P. and the Biotechnology and Biological Sciences Research Council for support.

REFERENCES

1 Gruenbaum, Y., Naveh-Many, T., Cedar, H. and Razin, A. (1981) Nature (London) **292**, 860–862
2 Gardiner-Garden, M., Sved, J. A. and Frommer, M. (1992) J. Mol. Evol. **34**, 219–230
3 Staiger, D., Kaulen, H. and Schell, J. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 6930–6934
4 Antequera, F. and Bird, A. P. (1988) EMBO J. **7**, 2295–2299
5 Pradhan, S. and Adams, R. L. P. (1995) Plant J. **7**, 471–481
6 Herskovitz, M., Gruenbaum, Y., Renbaum, P., Raxin, A. and Loyter, A. (1990) Gene **94**, 189–193
7 Urwin, N. A. R. and Jenkins, G. I. (1997) Plant Mol. Biol. **35**, 929–942
8 Power, J. B. and Chapman, J. V. (1985) in Plant Cell Culture: a Practical Approach (Dixon, R. A., ed.), pp. 37–65, IRL Press, Oxford
9 Seed, B. and Sheen, J.-Y. (1988) Gene **67**, 271–277
10 Kass, S. U., Goddard, J. P. and Adams, R. L. P. (1993) Mol. Cell. Biol. **13**, 7372–7379
11 Bevan, M., Bancroft, I., Bent, E., Love, K., Goodman, H., Dean, C., Bergkamp. R., Dirkse, W., Van Staveren, M., Stiekerna, W. et al. (1998) Nature (London) **391**, 485–488
12 Williams, M. E., Foster, R. and Chua, N.-H. (1992) Plant Cell **4**, 485–496
13 Donald, R. G. K. and Cashmore, A. R. (1990) EMBO J. **9**, 1717–1726
14 Gilmartin, P. M., Sarokin, L., Memelink, J. and Chua, N.-H. (1990) Plant Cell **2**, 369–378
15 Diéguez, M. J., Vaucheret, H., Paszkowski, J. and Mittelsten Scheid, O. (1998) Mol. Gen. Genet. **259**, 207–215

Received 14 May 1999/6 June 1999; accepted 15 June 1999