J. Mol. Biol. (1989) **208**, 551–565

Structure, Evolution and Anaerobic Regulation of a Nuclear Gene Encoding Cytosolic Glyceraldehyde-3-phosphate Dehydrogenase from Maize

Pascal Martinez[1], William Martin[2] and Rüdiger Cerff[1]†

[1]*Laboratoire de Biologie Moléculaire Végétale*
CNRS UA 1178, Université Joseph Fourier
F-38041 Grenoble, France

[2]*Max-Planck-Institut für Züchtungsforschung*
D-5000 Köln 30, F.R.G.

(Received 18 November 1988, and in revised form 3 April 1989)

A nuclear gene encoding cytosolic glyceraldehyde-3-phosphate dehydrogenase from maize (subunit GAPC1, gene *Gpc1*) and $2{\cdot}2 \times 10^3$ base-pairs of its 5′ flanking region have been cloned and sequenced. The structure of the maize *Gpc1* gene (10 introns) is different from that of the maize gene encoding subunit GAPA of chloroplast glyceraldehyde-3-phosphate dehydrogenase (1 intron) and relatively similar to that of the chicken gene (11 introns). Introns in the *Gpc1* gene show a positional polarity; the more 3′ their position, the more they are displaced relative to introns in the chicken gene. The *Gpc1* gene and other nuclear genes from maize are associated with CpG islands, the relative size of which determines the degree of codon bias in the gene. The promoter of the maize *Gpc1* gene contains an anaerobic regulatory element and a pyrimidine box upstream from the TATA box and within intron 1. Southern blotting analyses and Northern hybridizations suggest that there are three functional *Gpc* genes in maize whose transcript levels are controlled differentially by anaerobiosis. In spite of its "typical" anaerobic promoter, the *Gpc1* gene does not seem to be an anaerobic gene *in vivo*.

1. Introduction

Glyceraldehyde-3-phosphate dehydrogenase (GAPDHase‡) is present in both prokaryotes and eukaryotes, and is highly conserved across all species with respect to sequence (for a review, see Martin & Cerff, 1986) and three-dimensional structure (Rossmann *et al.*, 1975; Biesecker *et al.*, 1977). This indicates that all modern GAPDHase variants arose from a single ancestral enzyme without major rearrangements such as deletions and insertions of large polypeptide segments.

Since all known GAPDHase enzymes are similar to the ancestral protein, one might expect that the intron/exon structures of eukaryotic GAPDHase genes also may have remained relatively undisturbed over this long evolutionary period. However, a comparison shows that these structures are different and fall into three separate categories: (1) the continuous GAPHDase genes without introns (type 1) such as those from yeast (Holland & Holland, 1979), *Drosophila melanogaster* (Tso *et al.*, 1985) and *Trypanosoma brucei* (Michels *et al.*, 1986); (2) the simple GAPDHase genes with few introns (type 2) such as the nematode gene (2 introns, Yarbrough *et al.*, 1987); (3) the complex GAPDHase genes with many introns (type 3), of which so far only the chicken gene is known, with 11 introns (Stone *et al.*, 1985*a*).

Higher plants contain three different phosphorylating GAPDHase species: two chloroplast GAPDHase isoenzymes I and II (EC 1.2.1.13), for which the subunit structures A_2B_2 and A_4, respectively, have been suggested (however, see Brinkmann *et al.*, 1989); and a single cystolic GAPDHase (EC 1.2.1.12), with the proposed subunit structure C_4 (Cerff, 1979; Cerff & Chambers, 1979; for a review, see Cerff, 1982). All subunits are encoded in the nucleus (Cerff & Kloppstech, 1982) and cDNAs have been characterized encoding subunits A and C

† Author for correspondence.

‡ Abbreviations used: GAPDHase, glyceraldehyde-3-phosphate dehydrogenase; cDNA, complementary DNA; kb, 10^3 bases or base-pairs; bp, base-pair(s); ARE, anaerobic regulatory element; ADHase, alcohol dehydrogenase.

0022-2836/89/160551-15 $03.00/0

(GAPA and GAPC, for GAPDHase terminology, see Materials and Methods) from mustard and maize (Martin & Cerff, 1986; Brinkmann *et al.*, 1987), subunit C (GAPC) from barley (Chojecki, 1986), subunits A, B and C (GAPA, GAPB and GAPC) from tobacco (Shih *et al.*, 1986) and subunits A and B (GAPA and GAPB) from pea and spinach (Brinkmann *et al.*, 1989). These studies revealed that the sequences of GAPA and GAPB are different from that of GAPC and similar to that of the GAPDHase from thermophilic eubacteria, suggesting that GAPA and GAPB are of prokaryotic origin (see Brinkmann *et al.*, 1987, 1989).

We have characterized a genomic clone encoding maize GAPA1 (gene *Gpa1*; Quigley *et al.*, 1988). The maize *Gpa1* gene is a type 2 gene with three introns, two within the region encoding the transit peptide and one (intron 3) separating the mature subunit into its two constituent domains. A comparison of the three interrupted GAPDHase genes from maize, nematode and chicken revealed identical positions of introns 2 and 11 (Tr·p310†) for nematode and chicken, and of introns 1 and 3 (G·ly166) for nematode and maize chloroplast, respectively. This suggests that these introns were present in the parental GAPDHase genes from which these modern descendants originated about 700 million years ago (Tr·p310, nematode/chicken) and two billion years ago (G·ly166, nematode/chloroplast). Both intron positions, G·ly166 and Tr·p310, coincide with junctions between important structural units of the GAPDHase catalytic domain suggesting their implication in early GAPDHase evolution (Yarbrough *et al.*, 1987; Quigley *et al.*, 1988).

Here, we report the primary structure of a genomic clone encoding GAPC1 from maize (gene *Gpc1*). The maize *Gpc1* gene is a type 3 gene with a complex intron/exon organization similar to that of the chicken gene. We further demonstrate that the *Gpc1* gene and other nuclear genes from maize are associated with CpG islands, the relative size of which determines the degree of codon bias in the gene. Finally, we show that there are several (probably 3) functional *Gpc* genes in maize whose transcript levels are differentially controlled under anaerobic conditions.

2. Materials and Methods

(a) *Plant material*

The maize plants used for preparation of DNA (genomic library and Southern blots) originated from a genetic stock of P. A. Peterson (Ames, IO) and were grown in Cologne under the accession number 906. Maize seedlings used for RNA preparations belonged to the French variety RP704 (Rhône Poulenc). These seedlings were grown in 2 different ways. (1) Seedlings used for the preparation of poly(A)$^+$ mRNA in primer extension assays were grown on soil for 10 days at 25 °C under white fluorescent light. (2) Seedlings used for anaerobic treatment were grown on wet paper towels for 115 h at 25 °C in the dark. For anaerobic induction, seedlings were submerged for 20 h in 10 mM-Tris·HCl (pH 7·0) (Springer *et al.*, 1986). After anaerobic treatment, primary roots and shoots were removed, frozen in liquid nitrogen and stored at −70 °C.

(b) *Construction and analysis of genomic clones*

Genomic DNA for cloning and Southern blots was prepared as described (Schwarz-Sommer *et al.*, 1984). Maize DNA (180 μg) was partially digested with *Mbo*I (New England Biolabs) and size-fractionated on 0·7% (w/v) agarose. The 17 to 24 kb fraction was electroeluted and purified by chromatography on Whatman DE-52. Lambda EMBL4 vector arms were prepared by digestion with *Bam*HI and *Sal*I (Frischauf *et al.*, 1983) and subsequent centrifugation through potassium acetate gradients (Maniatis *et al.*, 1982). *Mbo*I partials (1 μg) were ligated for 12 h at 16 °C to 1·5 μg of EMBL4 vector arms in a volume of 10 μl containing 0·005 unit phage T4 DNA ligase (Boehringer). The ligation mixture was heated at 55 °C for 5 min and *in vitro* packaged according to the method of Hohn (1979). *Escherichia coli* strain K803 (Federoff, 1983) was employed as a host. Recombinant clones specific for cytosolic GAPC were identified and isolated by plaque hybridization. DNA from CsCl-purified phage was isolated and digested with *Eco*RI. Hybridizing *Eco*RI fragments were subcloned into pUC18. Purification of recombinant plasmids was performed as described (Maniatis *et al.*, 1982).

(c) *Sequencing of the maize* Gpc1 *gene*

Suitable restriction subfragments of genomic *Eco*RI fragments were subcloned into phages M13mp10, mp11, mp18 and mp19, and sequenced by the dideoxy chain termination method following the protocol supplied by Amersham. For the 2·5 kb region upstream from exon I of the maize *Gpc1* gene, an ordered set of M13 deletion clones was prepared before sequencing by using exonuclease III following the manufacturer's (STRATAGENE) protocol. Both strands of DNA fragments were sequenced by using the M13 primers as well as *Gpc1*-specific oligonucleotides prepared in the laboratory by means of the automatic DNA synthesizer (Applied Biosystems, model 381A).

(d) *Southern hybridizations*

Maize DNA (10 μg) was digested to completion with 10 units of the respective restriction enzyme, electrophoresed on 0·8% agarose and depurinated before capillary transfer and coupling with ultraviolet to Hybond-N (Amersham) nylon filters according to the manufacturer's specifications. Filters were hybridized for 24 h at 65 °C in 35 ml of 3× SSPE (SSPE is 0·18 M-NaCl, 0·010 M-NaH_2PO_4, 0·001 M-EDTA, pH 7·4), 0·1% (w/v) SDS, 0·2% (w/v) polyvinylpyrrolidone, 0·2% (w/v) Ficoll containing 50 μg denatured salmon sperm DNA/ml and 50 ng of the 102 bp *Hin*dIII–*Sau*3A fragment of cDNA pZm9 (encoding maize GAPC1; see Results) random-prime labelled (Feinberg & Vogelstein, 1984) to a specific activity of 5×10^8 cts/min per μg. Filters were washed twice for 20 min in 2× SSPE, 0·1% SDS at 65 °C and exposed for 72 h at −70 °C.

† Tr·p310, G·ly166, etc. indicate that the intron interrupts the specified codon after the second or first base, respectively.

(e) *RNA preparations*

Poly(A)$^+$ mRNA from shoots of green maize seedlings used in primer extension assays was isolated as described (Cerff & Kloppstech, 1982). Total RNA for Northern Blotting was prepared from 2 g of maize primary roots of dark-grown seedlings or shoots of light-grown seedlings before and after anaerobic treatment, essentially as described by Westhoff *et al.* (1981), except for the following modifications. After precipitation with isopropanol, the nucleic acids were dissolved in 6 ml of TE buffer (TE is 10 mM-Tris·HCl, 1 mM-EDTA, pH 8·0). LiCl was added to a final concentration of 2 M and RNAs were precipitated overnight at 0 °C. After a 2nd precipitation with LiCl, RNAs were taken up in TE buffer and stored at −20 °C.

(f) *Primer extension assay*

The 17mer oligonucleotide GCGTAGGCGACGGAGGC located 72 bases upstream from the AUG codon was synthesized (381A DNA synthesizer, Applied Biosystems). The reverse transcriptase reaction was carried out with this primer and 5 μg of poly(A)$^+$ mRNA, as described by the manufacturer of the cDNA synthesis kit (Boehringer). The cDNA was electrophoresed on an 8% polyacrylamide sequencing gel. The sequence of a 229 bp *Nco*I–*Nco*I fragment spanning the entire leader region and primed with the same oligonucleotide was used as reference.

(g) *Northern hybridization*

Total RNA (15 μg/slot) extracted from maize primary roots before and after anaerobic treatment was electrophoresed in 1% (v/v) formaldehyde agarose gels (Maniatis *et al.*, 1982) and transferred to Hybond-C extra nylon filters (Amersham). Filters were prehybridized at 65 °C in 3 × SSC (SSC is 0·15 M-NaCl, 0·015 M-trisodium citrate, pH 7), 5 × Denhardt's solution (0·02% Ficoll, 0·02% polyvinylpyrrolidine, 0·02% bovine serum albumin), 0·5% SDS, 10% (w/v) dextran sulphate, 200 μg denatured calf thymus DNA/ml. After 1 h, hybridization was started by addition of a DNA probe (50 ng) labelled by random priming (Boehringer). Hybridization was continued at 65 °C for 48 h. Filters were washed twice for 2 min at 65 °C in 2 × SSC, 0·5% SDS, then once for 30 min at room temperature in 0·2 × SSC, 0·1% SDS.

(h) *GAPDHase nomenclature*

The nomenclature of the plant GAPDHase system has been defined as follows. Gene products (mRNAs, cDNAs, proteins) encoding or corresponding to subunits A and B of chloroplast GAPDHase and subunit C of cytosolic GAPDHase are specified as GAPA, GAPB and GAPC transcripts or proteins, respectively, and products from different members of the same gene family are numbered consecutively, e.g. GAPA1, GAPA2 . . ., GAPB1, GAPB2 . . ., GAPC1, GAPC2 . . . The corresponding genes (gene families) are designated: *Gpa1*, *Gpa2* . . ., *Gpb1*, *Gpb2* . . ., *Gpc1*, *Gpc2* . . ., respectively.

(i) *Computer analysis*

Sequence data were processed on a Multics computer (CICG Grenoble) by using the program developed by Greaves and Ware (University of Bristol, England, unpublished) and on a CII-Honeywell-Bull DPS8 computer of the computer service center CIT12 at Paris by using the program BISANCE.

3. Results

(a) *GAPC is encoded by a small multigene family in maize*

The number of genes and pseudogenes encoding glycolytic GAPDHase in vertebrates varies between a single copy in chicken (Stone *et al.*, 1985*a*), ten to 30 copies in man, hare, guinea pig and hamster, and over 200 copies in mouse and rat (Hanauer & Mandel, 1984; Piechaczyck *et al.*, 1984). For barley, the only higher plant so far analysed, a single-copy gene for cytosolic GAPDHase has been reported (Chojecki, 1986).

To enumerate the genes and pseudogenes encoding GAPC in maize, we analysed Southern blots (Fig. 1) of maize DNA digested with *Eco*RI, *Bam*HI and *Hin*dIII with a 102 bp *Hin*dIII–*Sau*3AI fragment of cDNA pZm9 (Brinkmann *et al.*, 1987), spanning codons 186 to 219. At the nucleotide level, this probe shows only 56% similarity to the corresponding cDNA sequence encoding chloroplast GAPA from maize (Brinkmann *et al.*, 1987) and, hence, does not cross-hybridize with the *Gpa* genes under our hybridization conditions (see also Quigley *et al.*, 1989). In the *Gpc1* gene, this coding sequence is interrupted by intron 8 (Fig. 2(a)), which is 237 bases long (Table 1) and which does not contain recognition sequences of the three restriction enzymes used. Furthermore, CpG and CpXpG methylation sites (Gruenbaum *et al.*, 1981) are absent from the recognition sequences of the three restriction enzymes. Therefore, each fragment detected should represent a single maize *Gpc* gene. The hybridization patterns in Figure 1 show three strong bands for the *Eco*RI and *Hin*dIII digests and five bands for the *Bam*HI digest. These results suggest that there may be three separate (functional) *Gpc* genes in maize (*Gpc1*, *Gpc2* and *Gpc3*; see Discussion, section (c)). The fastest migrating band in the *Eco*RI digest corresponds to a genomic *Eco*RI fragment of 5·3 kb carrying the 5′ part of the *Gpc1* gene (see below).

(b) *Isolation and sequencing of the maize* Gpc1 *gene*

Approximately 1×10^6 recombinant phage from a maize genomic ligrary were screened with a nick-translated cDNA probe encoding cytosolic GAPDHase from maize (clone pZm9; see Brinkmann *et al.*, 1987). Eight positive plaques were purified. DNA was isolated from each of these phages and digested with *Eco*RI restriction endonuclease. The resulting *Eco*RI fragments were fractionated on a 0·8% agarose gel and transferred to a nylon filter that was hybridized with the same probe as that used for screening the library. Four clones out of these eight contained two strongly hybridizing bands of 5·3 kb and 4·1 kb. The two fragments of one of these clones λGpc1, were submitted to sequence analysis by the M13 dideoxy chain termination technique (see Materials and Methods). It turned out that the 5·3 kb and 4·1 kb fragments carry the 5′ and 3′ parts, respectively, of the maize

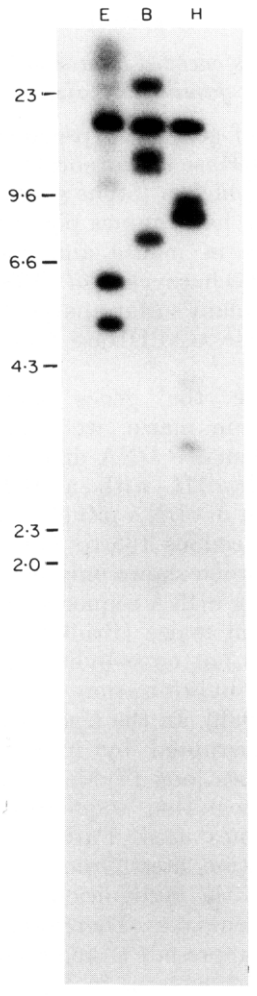

Figure 1. Counting of maize *Gpc* genes by Southern blotting. Portions (15 μg) of genomic maize DNA were digested with *Eco*RI (E), *Bam*HI (B), and *Hin*dIII (H). The fragments were separated by electrophoresis on a 0·8% agarose gel and blotted onto a nylon membrane. The filter was probed with a radioactive cDNA fragment spanning codons 186 to 219, as described in the text. Autoradiography was for 4 days at −70 °C, with intensifying screen. Positions of molecular weight markers (in kb) are indicated.

Gpc1 gene, which are separated by an internal *Eco*RI site in intron 10 (see Fig. 2(a)). The sequence of the complete maize *Gpc1* gene and 2·2 kb of its 5′ flanking sequence upstream from the promoter has been established. This sequence (together 5955 bases) has been entered in the EMBL/GenBank library (accession number X15596) and is available from the authors upon request.

The 5′ end of the *Gpc1* transcription unit was determined by extension with reverse transcriptase of a synthetic primer annealed to the 5′-leader region of GAPC1 mRNA (see Materials and Methods). A major extension product of 73 bases was detected (not shown), corresponding to a transcription start-point 47 bases downstream from the presumptive TATA box region ((TAATTATTTGTAATTA, see promoter structure in Fig. 6(a)) and 118 bases upstream from the AUG codon.

The *Gpc1* gene contains ten introns (Fig. 2(a)) and differs in a number of base positions from the previously determined cDNA sequence pZm9 (Brinkmann *et al.*, 1987). Within the coding part, there are nine silent substitutions (7 transitions and 2 transversions) and one replacement substitution leading to a conservative amino acid change (serine to threonine) at codon position 331. In the 3′ non-coding region between the stop codon and the poly(A) addition site, genomic clone λGpc1 contains nine base substitutions (5 transitions and 4 transversions) and two deletions of one base and 15 bases, respectively. These sequence differences probably indicate that cDNA pZm9 and genomic clone λGpc1 represent two allelic variants of the same genetic locus *Gpc1*.

(c) *Conservation, displacement and differential loss of introns in the maize* Gpc1 *gene*

The sizes of the ten introns of the maize *Gpc1* gene, their 5′ and 3′ splice junctions and the putative branchpoint sequences are given in Table 1. It can be seen that the consensus sequences for the maize *Gpc1* gene agree well with those published for plant genes in general (Brown, 1986). Except for intron 3 (382 bases) and intron 5 (529 bases), all introns are relatively short, between 84 and 237 bases long. They all have a stop codon in-frame, except for intron 1, which is inserted one base after the initiation codon AUG (G·ly 1, see Fig. 3). Intron 1 could be translated in-frame with the exonic AUG up to the beginning of exon II, which starts with a different frame.

A comparison of the intron positions in the maize *Gpc1* gene with those in the chicken gene (Figs 2(a) and (b) and 3) shows that introns 2 and 3 are precisely conserved while introns 1, 4 and 5 are slightly displaced. Introns 6 to 10, interrupting the catalytic domain, are more strongly displaced, and intron 11 of the chicken gene is absent in maize (see Discussion).

(d) *The maize* Gpc1 *gene is associated with an ancient truncated pseudogene*

As shown in Figure 4, there is a truncated *Gpc* pseudogene in the 5′-flanking region of the active gene between base positions 1370 and 1720, 1135 bases upstream from the AUG codon. This partial pseudogene comprises the 3′ end of exon IX (codons 250 to 260), intron 9, exon X (codons 261 to 288) and the beginning of intron 10. Its coding sequence (codons 250 to 288) is considerably more conserved than intron 9 (79% *versus* 57% sequence similarity), suggesting that the gene was functional before its inactivation.

(e) *Distribution of CpG, TpG and CpA dinucleotides in the maize Gpc1 gene and its 5′ flanking sequence*

We have shown (Quigley *et al.*, 1988) that CpG dinucleotides are non-randomly distributed in the genomic DNA region carrying the gene for subunit

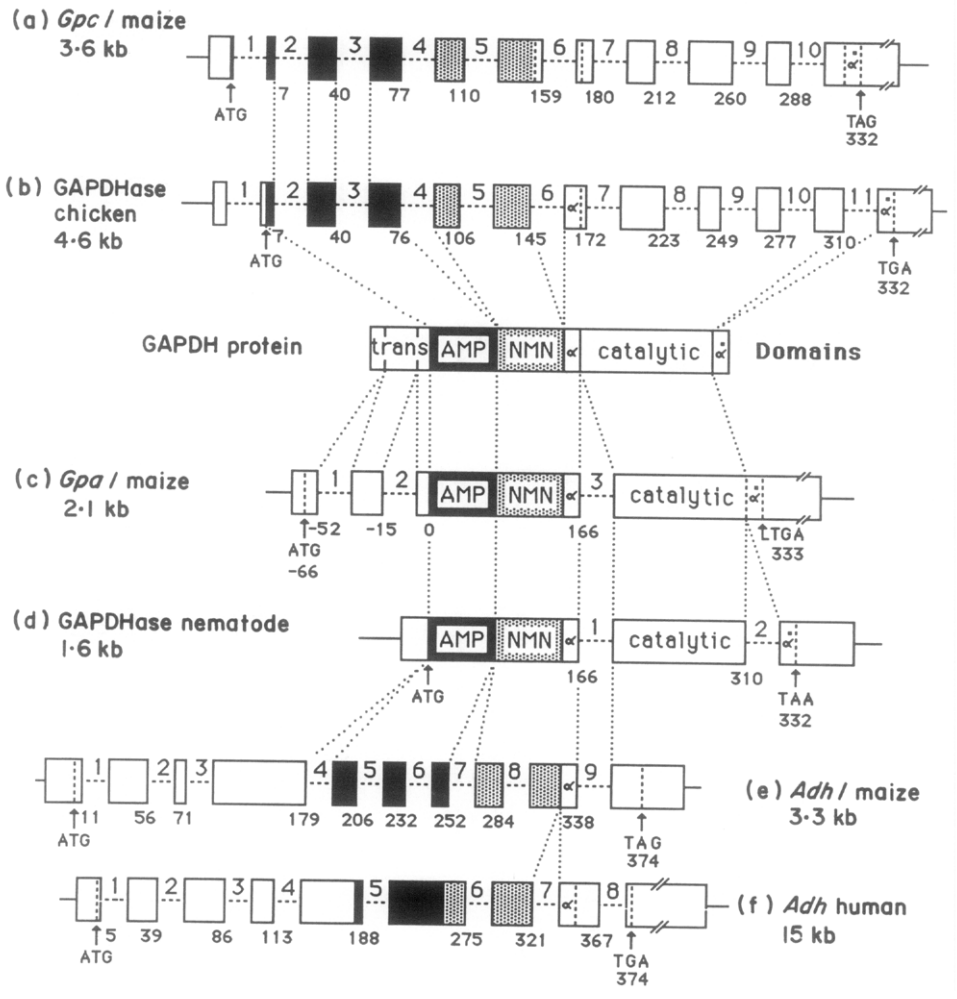

Figure 2. Schematic comparison of the exon–intron organization of the genes encoding various GAPDHase ((a) to (d)) and ADHase ((e) and (f)) enzymes. The exon–intron arrangement of the genes encoding maize cytosolic GAPDHase (*Gpc1*), chicken GAPDHase (Stone *et al.*, 1985*a*), maize chloroplast GAPDHase (*Gpa1*; Quigley *et al.*, 1988), nematode GAPDHase (Yarbrough *et al.*, 1987), maize ADHase (Brändén *et al.*, 1984) and human ADHase (Duester *et al.*, 1986) are aligned on the basis of the corresponding structural homologies of proteins as defined by X-ray diffraction studies. Exons are indicated by boxes, introns by broken horizontal lines and arabic numerals. Numbers below the exons indicate their terminal codons. The lengths of all exons are true to scale. trans, transit peptide; AMP and NMN, adenosine monophosphate and nicotinamide mononucleotide subdomains of the NAD binding domain; α, α′, helices alpha-1 and alpha-3 of the catalytic domain (see Fig. 3); catalytic, central part of the catalytic domain.

Table 1

Intron positions and consensus sequences for the maize Gpc1 *gene*

No.	Position (codon no.)	Length (bases)	5′ Donor		Putative branch point		3′ Acceptor
1	−1:G·ly	184	TGG:GTAATT	...127...	TGCTGAC	...39...	TGCAG:G
2	7:G·ly	117	ACG:GTGAGT	....53...	ATCTGAG	...46...	CTCAG:G
3	40/41:Met/Tyr	382	ATG:GTACGC	...333...	TGCTCAT	...31...	CGTAG:A
4	77:Ar·g	84	CAG:GTGCTC	....47...	GGCTAAT	...19...	TACAG:G
5	110/111:Lys/Gly	529	AAG:GTATAA	...490...	TTCTGAT	...21...	GACAG:G
6	159/160:Lys/Val	91	AAG:GTGCGC	....59...	TAATGAT	...14...	ACCAG:G
7	180:A·la	87	CTG:GTTAGT	....53...	AAATGAC	...16...	TTCAG:C
8	212/213:Lys/Ala	237	AAG:GTATAG	...176...	TTCTGAT	...43...	TGCAG:G
9	260:Ly·s	146	TAA:GTAAGT	...115...	TGGTAAT	...13...	TACAG:G
10	288:Ar·g	111	CAG:GTATGG	....69...	GATTCAT	...24...	CTCAG:G
Gpc1 consensus			AAG:GTAAGT C C		TGCTGAT T		TNCAG:G C
Plant consensus			CAG:GTAAGT A		TTCTRAY RR		TGCAG:G

Lengths of introns and distances between consensus blocks are indicated as number of bases. Letters R and Y designate purine and pyrimidine bases, respectively.

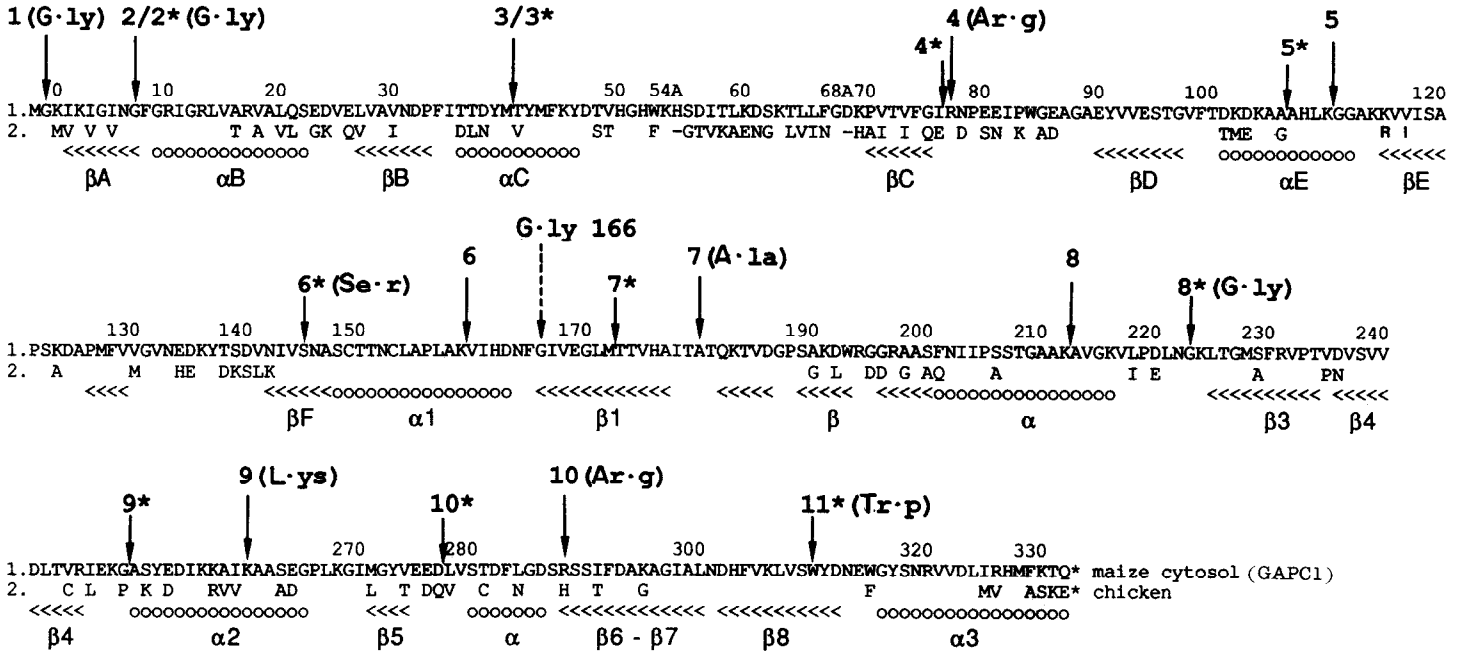

Figure 3. Comparison of intron positions in the GAPDHase genes from maize (*Gpc1*) and chicken (Stone *et al.*, 1985*a*). Amino acid sequences of maize GAPC1 (line 1) and chicken GAPDH (line 2) are aligned to maximize similarity and sequence elements forming alpha-helices (○○○○○) and beta-structures (< < < < <) are indicated below the alignment (Rossmann *et al.*, 1975). Intron positions are indicated by arrows and chicken introns are marked with an asterisk (*). Ancient intron G·ly166, present in the genes for maize chloroplast GAPA and nematode GAPDHase (Quigley *et al.*, 1988), is indicated by a broken arrow. Chicken intron 1*, interrrupting the leader region, is not shown.

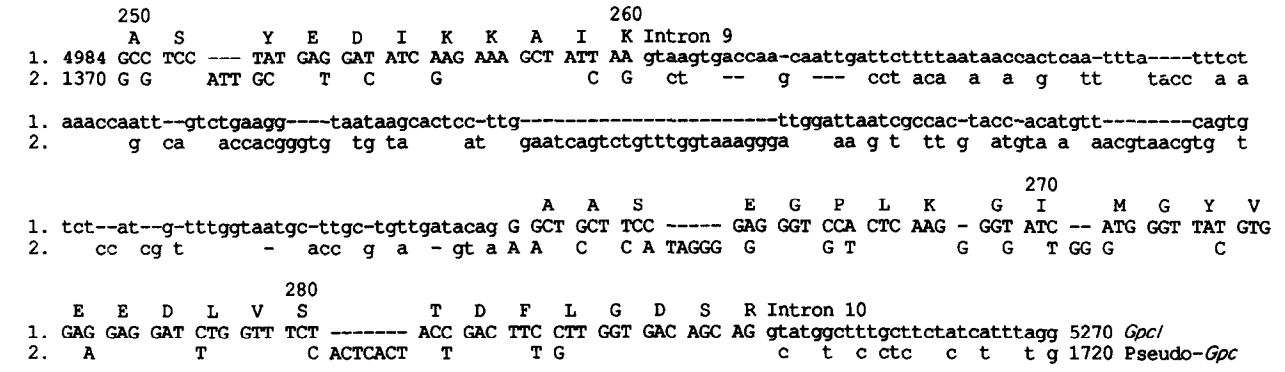

Figure 4. The sequence of a truncated *Gpc* pseudogene (bases 1370 to 1720, line 2) is aligned with the corresponding gene region of the functional *Gpc1* gene (bases 4984 to 5270, line 1), comprising codons 250 to 258, intron 9 and the beginning of intron 10. The sequence differences are 21% and 43% for codons and intron 9, respectively. Insertions and deletions were scored as single mutation events equivalent to single nucleotide substitutions.

GAPA of chloroplast GAPDHase from maize (gene *Gpa1*). The CpG profile of the maize *Gpa1* gene and its surrounding sequences, recently established (Quigley *et al.*, 1989), is shown in Figure 5(a). Compared to the *Gpa1* gene, the constitutive *Gpc1* gene from maize has a relatively moderate codon bias; 67% G+C in the triplet third base position (Brinkmann *et al.*, 1987) and 54% total G+C in the coding sequences. Nevertheless, as shown in Figure 5(b), the distribution of CpG dinucleotides within the *Gpc1* gene is strongly asymmetric. Most CpG doublets are clustered in 1·6 kb region at the 5′ end of the gene (region 1). The central part, comprising exons V to X (region 2 of 1·7 kb) is CpG-poor and there is a second small peak at the 3′ end of the gene (region 3 of 0·6 kb). As shown in Table 2, the CpG frequencies in the three regions are 7·2% (7·2% expected), 1·3% (4·5% expected) and 3·9% (5·0% expected), respectively. Hence, there is a 5·5-fold difference in CpG frequency between the 5′ end and the central part of the gene, due to a combination of G+C-richness in region 1 and a 3·5-fold CpG suppression in region 2. CpG suppression is observed also in the 5′ region upstream from the *Gpc1* gene (2·4% CpG *versus* 4·0% expected; see Table 2 and region 0 in Fig. 5(a)). Suppression of CpG in regions 0 and 2 is correlated with an excess of TpG and CpA (see Table 2). Similar CpG profiles are found for the maize genes encoding triose phosphate isomerase (TIMase; Marchionni & Gilbert 1986) and sucrose synthase A (Werr *et al.*, 1985), as shown in Figure 5(c) and (d) and Table 2 (see Discussion).

Table 2

*G+C content and frequencies of CpG, TpG and CpA dinucleotides of the genomic DNA regions carrying the maize genes encoding GAPA (*Gpa1*), GAPC1 (*Gpc1*), triosephosphate isomerase (*TIM*), sucrose synthase A (*Sh1*) and the chicken GAPDHase gene*

			Values (%)					
			CpG		TpG		CpA	
Gene	Region	G+C	A	B	A	B	A	B
Gpa1	0	43	2·5	4·6	6·5	5·6	8·2	6·7
	1	62	11·2	9·5	5·3	5·3	7·4	6·4
	2	41	2·8	4·1	6·3	5·8	7·8	6·2
Gpc1	0	40	2·4	4·0	7·6	6·5	6·5	5·5
	1	54	7·2	7·2	6·3	6·9	6·0	5·3
	2	43	1·3	4·5	10·2	7·5	7·0	4·8
	3	45	3·9	5·0	8·4	6·3	7·4	5·9
TIM	1	51	6·1	6·3	5·9	6·4	4·0	5·8
	2	39	1·0	4·2	10·1	7·5	6·8	4·7
	3	45	3·9	5·2	9·5	6·9	5·6	5·5
Sh1	1	52	5·2	6·9	8·9	7·5	5·2	4·8
	2	42	1·9	4·4	9·1	6·9	6·9	5·3
	3	51	5·7	6·5	8·1	6·7	6·4	5·8
Chicken	1	71	10·8	12·6	6·0	5·6	5·1	4·7
	2	50	1·1	6·1	10·7	7·7	7·0	5·0

The genomic sequences are subdivided into regions that are poor (regions 0 and 2) or rich (regions 1 and 3) in G+C and CpG, respectively, as defined in Fig. 5. The observed (A) and expected (B) values are given.

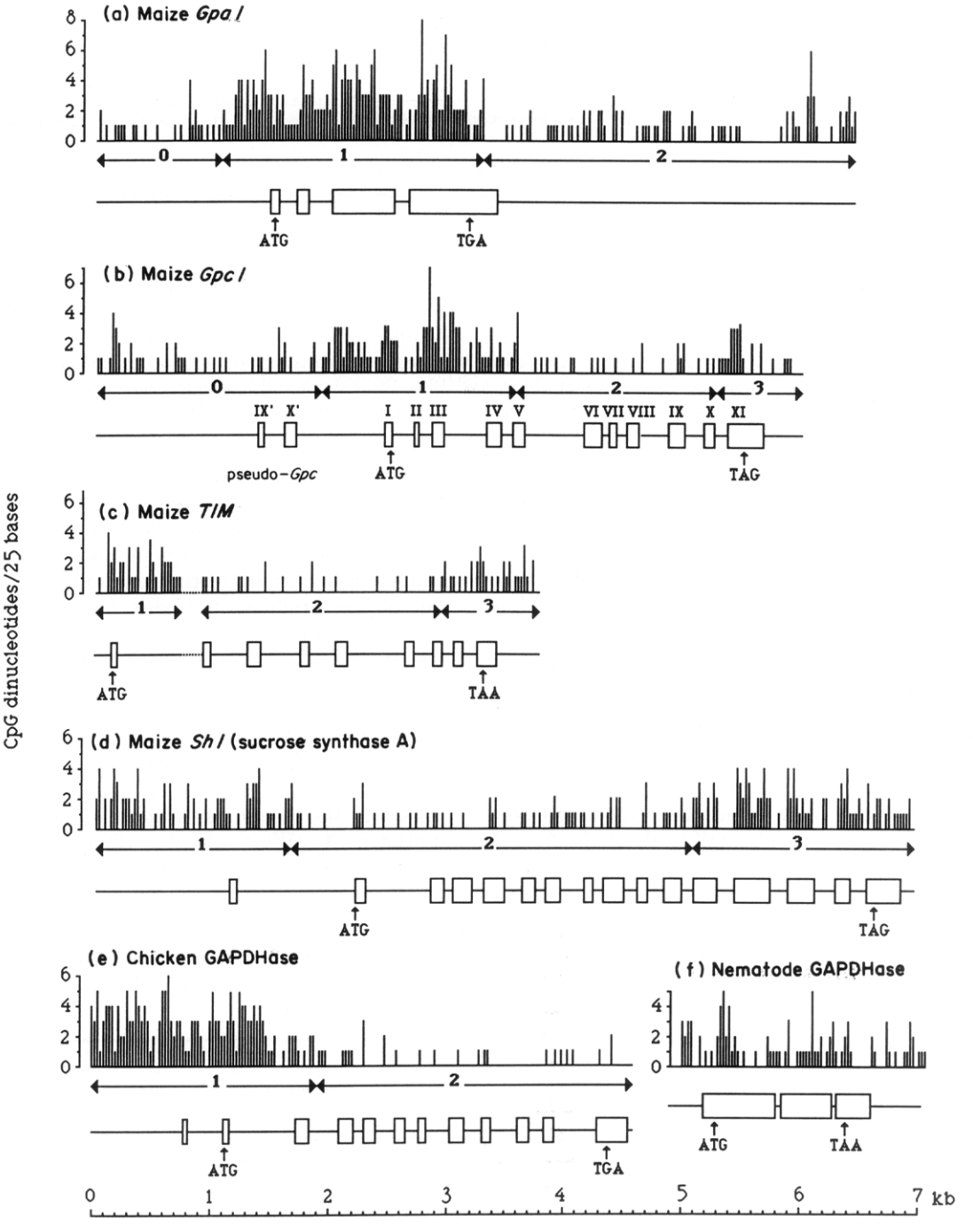

Figure 5. CpG profiles of nuclear genes from maize ((a) to (d)), chicken (e) and nematode (f). Each vertical line represents the number of CpG doublets per 25 bases. CpG-rich and CpG-poor regions of the genes are numbered consecutively from 0 to 3 (see Table 2). They are shown together with the intron–exon arrangements below the CpG profiles. Introns (continuous lines) and exons (boxes) are true to scale. For sources of sequence information, see the legend to Fig .2.

(f) *Promoter structure and anaerobic control of the maize* Gpc1 *gene*

Maize and other higher plants can survive anaerobiosis for prolonged periods of time, due to about 20 specific proteins and enzymes whose expression in seedlings is induced or maintained under anaerobic conditions (see Discussion). In order to determine whether the maize *Gpc1* gene may be regulated by anaerobiosis, we compared its promoter structure with the promoter structures of the known anaerobic genes for alcohol dehydrogenase in maize (*Adh1*; Walker *et al.*, 1987) and pea (Llewellyn *et al.*, 1987). As shown in Figure 6, there are several regions of nucleotide conservation. The region surrounding the TATA box, known to be important

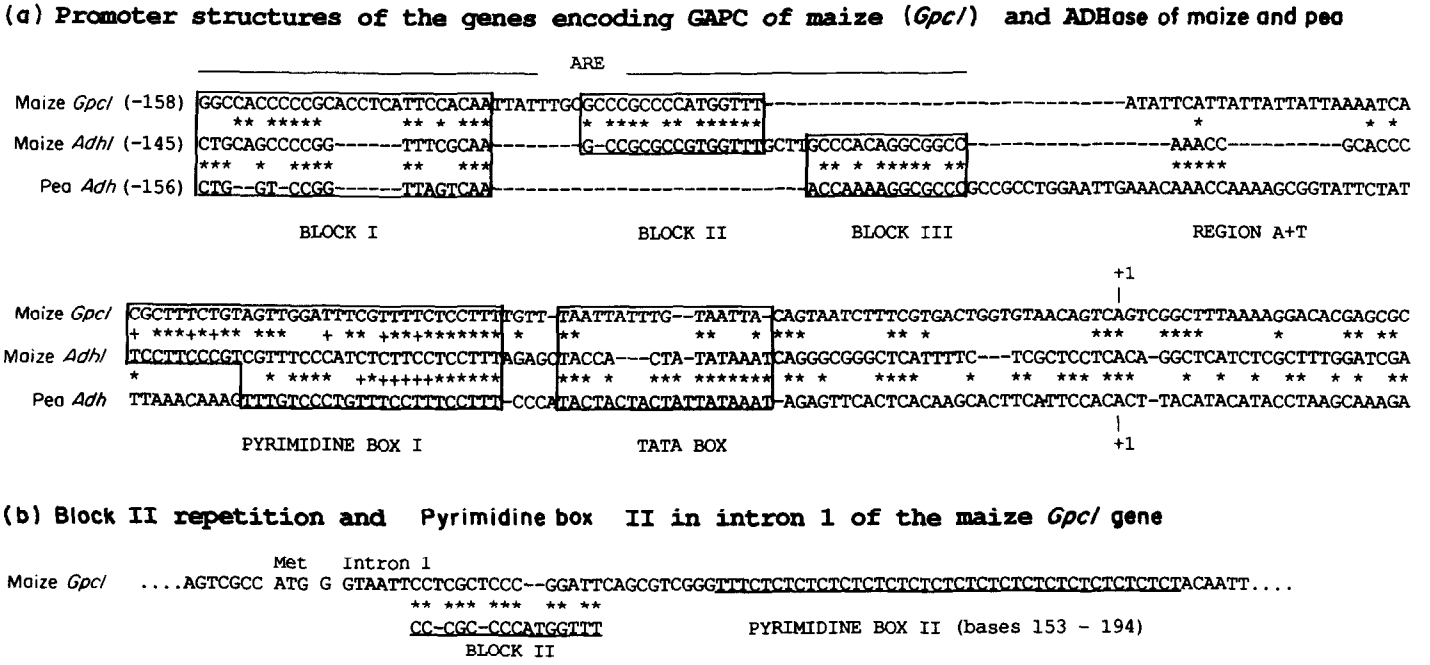

Figure 6. The maize *Gpc1* gene has an "anaerobic" promoter similar to those of the *Adh* genes from maize and pea. (a) The promoters of the genes encoding maize GAPC1 (*Gpc1*, line 1), maize ADH1 (*Adh1*, Walker *et al.*, 1987; line 2) and pea ADH (Llewellyn *et al.*, 1987; line 3) were aligned by introducing gaps to maximize sequence similarity. Identical bases in the pairwise comparisons maize *Gpc1*/maize *Adh1* and maize *Adh1*/pea *Adh* are indicated by an asterisk (*), coincident pyrimidine bases in pyrimidine box I by a plus (+). Similar sequence elements are boxed. Numbering is relative to the transcription start as +1. (b) Incomplete repetition of ARE block II and pyrimidine box II in intron 1 of the *Gpc1* gene.

for accurate transcription initiation, is relatively conserved. Just upstream from the TATA box there is an extensive pyrimidine-rich region in both maize genes and a somewhat shorter stretch in the pea *Adh* gene (pyrimidine box I). Further upstream, a putative anaerobic regulatory element (ARE; Walker *et al.*, 1987) is present in all three genes. The putative ARE is composed of three homology blocks that are G+C-rich, especially in the maize genes. Block I is present in all three genes, block II is shared by maize *Gpc1* and maize *Adh1*, and block III by maize *Adh1* and the pea *Adh*. The hexanucleotide "core element" TGGTTT (Dennis *et al.*, 1987) is found in block II of both maize genes, the complementary copy ACCAAA in block III of the pea *Adh* gene. In maize *Gpc1* and pea *Adh*, the bipartite ARE and the pyrimidine box are separated by an A+T-rich region. In intron 1 of the maize *Gpc1* gene there is an incomplete repetition of ARE block II (12/17 matches) and just downstream there is a second pyrimidine-rich region (pyrimidine box II) containing 19 CT repetitions (Fig. 6(b)).

In order to study the anaerobic control of maize *Gpc* genes *in vivo*, we analysed GAPC transcript levels in primary roots of etiolated maize seedlings before and after anaerobic treatment (see Materials and Methods). Total levels of GAPC transcripts were analysed with a non-specific probe, a 1029 bp *Xho–Sal*I fragment covering most of the coding sequence of clone pZm9 (Brinkmann *et al.*, 1987). Individual levels of GAPC1 transcripts were analysed with a specific probe containing the 3′ non-coding region of cDNA pZm9, a 250 bp *Sal*I–*Eco*RI fragment starting five nucleotides upstream from the TGA stop codon. The results of this Northern blot analysis are shown in Figure 7. Both probes detect a single band, suggesting that all GAPC transcripts are similar in size. They comigrate with 18 S rRNA, which is about 1·8 kb in size. The calculated size of the GAPC1 transcript without the poly(A) tail is 1349 bp. It can be seen clearly that the steady-state level of total maize GAPC transcripts increases or is at least maintained after 20 hours of anaerobiosis (see Fig. 7(a)). Surprisingly, however, GAPC1 transcript levels clearly decrease in anaerobic maize roots (see Fig. 7(b)). This decrease is comparable to that found for GAPA transcripts (encoding subunit A of chloroplast GAPDHase) in green anaerobic maize shoots (see Fig. 7(c)).

4. Discussion

The plant GAPDHase system represents the first example of a pair of cytosolic/chloroplast enzyme homologues (Weeden, 1981) for which sequence data and gene structures have become available. The structure of the maize *Gpc1* gene (10 introns) contrast with that of the gene encoding chloroplast GAPA (gene *Gpa1*, 1 intron within the mature subunit) and is relatively similar to that of the chicken gene (11 introns). These results provide strong corrobarative evidence for the separate evolutionary history of the *Gpa1* gene and for its apparent prokaryotic (symbiotic) origin (Martin & Cerff, 1986; Shih *et al.*, 1986; Brinkmann *et al.*, 1987; Quigley *et al.*, 1988). At least five but probably more introns of the maize *Gpc1* gene have homologous counterparts in the chicken gene (see below), suggesting that the maize *Gpa1* gene may have lost most of its introns before its transfer into the plant nucleus. The present results further suggest that the continuous and simple GAPDHase genes of invertebrates and yeast (types 1 and 2, see Introduction) evolved by "structural streamlining"; that is, by elimination of introns after their divergence 700 million (invertebrates) and over 1 billion (yeast) years ago (Gilbert *et al.*, 1986; Brinkmann *et al.*, 1987).

(a) *The placement of introns in the maize* Gpa1 *gene follows a gradient*

In Figure 3 the intron positions in the genes for cytosolic GAPDHase of maize (*Gpc1*) and chicken are shown with regard to the sequence and secondary structure of the protein. They are numbered consecutively, and chicken introns are marked with an asterisk to distinguish them from the corresponding introns in maize. Chicken intron 1*, which inter-

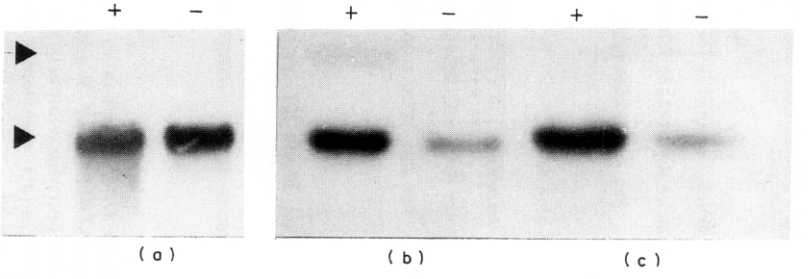

Figure 7. Accumulation of (a) total GAPC transcripts and (b) individual levels of GAPC1 mRNAs under aerobic (+) and anaerobic (−) conditions in primary roots of dark-grown maize seedlings. For anaerobic induction, seedlings were submersed for 20 h in 10 mM-Tris·HCl, pH 7·0 (Springer *et al.*, 1986). Filters were hybridized with (a) a non-specific GAPC probe and (b) a specific GAPC1 probe, respectively, as described in Materials and Methods, and in Results. Northern blots were performed after electrophoresis of total RNA (15 μg/slot). Arrows show the positions of the 28 S and 18 S rRNAs. (c) GAPA transcript levels in aerobic (+) and anaerobic (−) shoots of light-grown maize seedlings. The filter was hybridized with a nearly full-length GAPA cDNA (clone pZm57; Brinkmann *et al.*, 1987).

rupts the non-coding leader region, is not shown in Figure 3. Introns 2*/2 (G·ly7) and 3*/3 (after Met40) are at identical positions, and introns 1*/1, 4*/4 and 5*/5 are slightly displaced, i.e. 16, 2 and 12 bases, respectively. The distances between the individual members of the other intron pairs vary between 22 (7*/7) and 34 (6*/6) bases. There is no maize intron corresponding to intron 11* in chicken. This ancient intron, identical with intron 2 in the nematode gene (see Fig. 2(b) and (d)), has probably been eliminated in maize. The identical position of maize intron 3 and, hence, its conservation over the plant/animal distance is not consistent with a suggestion made by Stone *et al.* (1985*b*) that this intron would be the remnant of former "parasitic DNA" inserted at random into the chicken gene.

The placement of introns in the maize *Gpc1* gene relative to the chicken gene follows a gradient; the more 3′ their position, the more they are displaced relative to the chicken introns (see Fig. 3 and below). As a consequence of this, it is difficult to judge whether the relative positions of introns 6 to 10, interrupting the catalytic domain of the maize *Gpc1* gene, are due to displacement or differential loss of introns. Also, the differential insertion of "new" introns into the gene region encoding the catalytic domain cannot be excluded completely. However, the gene region between intron 6* (Se·r145, chicken) and intron 7 (A·la180, maize) merits particular attention. It encodes the element alpha-helix 1, which carries the strongly conserved catalytic centre in GAPDHase and which probably played an important role during early GAPDHase and ADHase evolution in joining the NMN subdomain to the catalytic domain (Brändén *et al.*, 1984; Quigley *et al.*, 1988). This interpretation is based on the observation that alpha-helix 1 is flanked by introns in several divergent species; its N terminus is flanked by intron 6* in chicken GAPDHase and intron 7 in human ADHase (see Fig. 2(b) and (f)), its C terminus by intron G·ly166, present at identical positions in the genes for maize chloroplast GAPDHase (*Gpa1*, intron 3) and nematode GAPDHase (intron 1) and by the homologous intron 9 in the gene for maize ADHase (compare Fig. 2(c), (d) and (e)). Surprisingly, intron G·ly166 is absent from the cytosolic GAPDHase genes from both chicken and maize (see Fig. 2(a) and (b)). However, because of its extremely ancient origin (Quigley *et al.*, 1988), one might expect that it was present in the common ancestor. It seems possible, therefore, that one or both of the present introns 6 (Lys159/Val160) and 7* (Met172/Tyr173) flanking position G·ly166 are, in fact, descendants of ancient intron G·ly166, which moved 19 bases upstream in maize and 20 bases downstream in chicken.

The present results suggest that introns can be displaced along a coding sequence without changing either the length or the sequence of the corresponding polypeptide. This distinguishes the gene for cytosolic GAPDHase from the genes for serine proteases and dihydrofolate reductases (Craik *et al.*, 1983), where intron positions often coincide with internal length differences in the polypeptide, a phenomenon that has been interpreted in terms of sliding of a single intron/exon junction (Craik *et al.*, 1983).

The displacement of an intron along a conserved coding sequence is difficult to explain by intron sliding and probably requires a different mechanism. On the basis of the earlier considerations of Sharp (1985) and Cech (1986), it could be explained formally by an "intron reinsertion–homologous recombination" model comprising four consecutive steps; excision of the intron, reinsertion into a nearby site of the pre-mRNA *via* reversible transesterification (Sharp, 1985; Cech, 1986), reverse transcription of the modified pre-mRNA and homologous recombination *via* gene conversion. Omission of step 2 would lead to the elimination of the intron and precisely such a "homologous recombination model" has been proposed by Fink (1987) to explain the paucity of introns in yeast genes. Fink's model also explains why the few introns that have been retained by yeast are mainly localized at the 5′ end of the genes; reverse transcription always initiates at the 3′ end but rarely extends completely to the 5′ end of the RNA, thereby favouring the conservation of introns in the 5′ part of the gene (see Fink, 1987). Similarly, reverse transcription, which is functional in higher plants (for processed pseudogenes and retroposons in plants, see Drouin & Dover, 1987; Voytas & Ausubel, 1988; Grandbastien *et al.*, 1989), may be responsible for the polarity we observe with respect to intron displacements/deletions in the maize *Gpc1* gene (see Fig. 3 and above).

We realize that this intron reinsertion–homologous recombination idea is speculative and needs further substantiation. For example, intron reinsertion is less likely than intron excision and, therefore, introns should be more frequently deleted than displaced. However, this would be expected only if introns are neutral and do not have an immediate selective value, as may be the case for most yeast genes but possibly not for most mosaic genes of higher organisms.

The present finding of an intron displacement polarity in the *Gpc1* gene may be related to our observation (Quigley *et al.*, 1989) that, in the *Gpa* gene family of maize, sequence divergence between individual introns increases in the 5′ to 3′ direction, possibly due to preferential accumulation of reverse transcription errors in the introns located in the 3′ part of the gene (see Quigley *et al.*, 1989). It may be hypothesized, therefore, that the two phenomena, sequence divergence polarity and position polarity of introns, reflect the short-term and long-term effects, respectively, of the same genetic mechanism; that is, the homologous recombination of reverse transcribed (modified) pre-mRNAs.

(b) *CpG islands and codon bias in maize genes*

One characteristic feature of certain vertebrate housekeeping genes is the asymmetric distribution of CpG dinucleotides. CpG doublets are clustered in

so-called CpG islands (Bird, 1986, 1987) at the 5′ end of these genes. These islands are rich in G+C and CpGs, which occur at their expected frequencies. The rest of the gene is abnormally CpG-poor; that is CpGs are suppressed and may be present at less than one-quarter of their expected frequency. The combination of G+C-richness and lack of CpG suppression means that CpG islands of vertebrate housekeeping genes may contain ten to 20 times more CpGs than an equivalent stretch of non-island DNA. One major cause for CpG suppression in the vertebrate genome is probably CpG methylation *in vivo* leading to enhanced CpG mutability. The expected products of 5mCpG mutation are TpG and CpA, and these doublets are present in excess in CpG-deficient regions. CpG islands of active vertebrate genes are non-methylated due to an unknown protective mechanism (see below) and non-methylation of these islands is thought to be a functional prerequisite or consequence of gene expression (for a review, see Cedar, 1988). The gene encoding chicken GAPDHase is a typical example of a vertebrate gene with a CpG island. Its CpG profile is shown in Figure 5(e) and the dinucleotide frequencies for the CpG-rich and CpG-poor regions of this gene are given in Table 2.

We have shown (Brinkmann *et al.*, 1987) that the maize gene encoding chloroplast GAPA (Gene *Gpa1*) and other inducible genes from monocotyledonous plants are extremely G+C-rich at the third base position of codons. Originally, we interpreted these findings in terms of possible constraints exerted at the translational level; for example, by major tRNA species best adapted to codons with G or C at the third base position (see Brinkmann *et al.*, 1987). However, by characterizing the genomic DNA carrying the maize *Gpa1* gene (Quigley *et al.*, 1988, 1989), we realized that the G+C-richness of this gene is not restricted to the coding sequences but extends, in the form of a CpG island, at least 400 bases to the 5′ side of the AUG codon (see Fig. 5(a)). Antequera & Bird (1988) have demonstrated the presence of CpG islands in three other maize genes, including the gene for sucrose synthase A (see Fig. 5(d)). They further showed that these CpG islands are non-methylated, in contrast to the zein gene, which is entirely CpG-suppressed and methylated.

The present results show clearly that the maize *Gpc1* gene also is associated with a CpG island extending from about 500 bases upstream from the AUG down to exon 5, and there is a second small peak at the 3′ end of the gene (see Fig. 5(b)). This bimodal CpG distribution is found also in the gene encoding triosephosphate isomerase (Fig. 5(c)) and it is especially pronounced in the maize gene encoding sucrose synthase A (Fig. 5(d), *Sh1*). The G+C content at the third base position of *Gpc1* codons is 84% and 51% for exons I to V and exons VI to X, respectively (see Fig. 5(b)). In the *Sh1* gene, which has a CpG island in its 3′ part (see Fig. 5(d)), these values are 60% and 82% for exons 2 to 11 (region 2) and exons 12 to 16 (region 3), respectively. These results seem difficult to reconcile with constraints at the translational level and suggest that codon bias (G+C preference in the 3rd base position) is a consequence of G+C and CpG clustering in the 5′ and 3′ parts of the genes rather than an adaptation to efficient (rapid) translation. As shown in Figure 5, maize genes may be associated with CpG islands of variable sizes. If the CpG island is large (rich in G+C) and the gene is small, as in the case of the *Gpa1* gene, the whole gene fits into the CpG island (see Fig. 5(a)) and the G+C content at the third base position of codons approaches saturation (see Brinkmann *et al.*, 1987). If the CpG island is small (less rich in G+C) and the gene is relatively large, as in the case of the maize *Gpc1* gene (Fig. 5(b)), the average codon bias (G+C preference) of the gene is moderate. Hence, the average codon bias of a particular maize (monocotyledonous) gene seems to be determined by the relative size and G+C content of its associated CpG island. The observation that within CpG islands, introns usually have lower G+C values than exons (see CpG profiles (a) and (d) in Fig. 5; Quigley *et al.*, 1988; Werr *et al.*, 1985) may be explained by intron-specific constraints, possibly exerted at the level of premRNA splicing.

The potential functions of CpG islands in maize (monocotyledonous) genes are unknown. However, in view of the rough correlation between increasing codon bias and gene expressivity *in vivo* (Brinkmann *et al.*, 1987), it may be speculated that CpG islands in vertebrate genes somehow stimulate gene transcription. The potential roles of CpG islands have been discussed in detail (see Bird, 1987). One interesting suggestion was that CpG islands may be preferred sites for interaction between DNA and ubiquitous nuclear transcription factors. On the basis of this hypothesis, one may suspect that in maize and other monocotyledonous genes, ubiquitous transcription factors interact cooperatively with CpG islands, possibly leading to an altered chromatin structure. In addition, a constitutive association of DNA-binding proteins with CpG islands denying access to the methylase would provide an easy explanation for the observation that CpG islands are not methylated (not CpG-suppressed; see Table 2 and Bird, 1987). As a consequence of this, one would expect that non-transcribed pseudogenes from maize are methylated at CpG islands, as has been shown for vertebrates in the case of the human α-globin pseudogene (Bird *et al.*, 1987). Our recent findings (Quigley *et al.*, 1989) showing enhanced turnover of CpG and CpXpG methylation sites in two strongly conserved *Gpa* pseudogenes from maize are consistent with this idea.

Although CpG islands may modulate transcriptional activity in maize (monocotyledonous) plants, they do not seem to be an essential prerequisite for gene function. As already mentioned, CpG islands are absent from the functional zein genes, which are entirely CpG-suppressed (Antequera & Bird, 1988). They may be absent also from most dicotyledonous

genes, which have comparatively low G+C contents (Brinkmann *et al.*, 1987). We analysed the genes encoding nodulin 35 from soybean (32% G+C; Nguyen *et al.*, 1985) and glutamine synthase from *Medicago sativa* (35% G+C; Tischer *et al.*, 1986). Both genes have normal CpG frequencies in their 5′ parts (2·1 and 2·8%, respectively, as expected) but are five- to tenfold CpG-suppressed in their central parts (data not shown). This suggests that protection of the 5′ part of a gene against CpG methylation does not automatically lead to G+C enrichment in this region. Hence, differential CpG methylation alone does not seem to be sufficient to explain the origin of CpG islands in maize (monocotyledonous) genes. This is particularly evident for the CpG islands of the maize *Gpa1* gene, which is almost saturated in G+C content (97% G+C in the 3rd base position of codons), while the flanking sequences are only slightly CpG-suppressed (Fig. 5(a) and Table 2). Hence, the primary cause for CpG island formation in maize (monocotyledonous) genes seems to be differential G+C enrichment in the 5′ and 3′ parts of the gene, possibly due to selective constraints exerted at the level of transcription.

It has been suggested by Salinas *et al.* (1988) that the G+C-richness of certain monocotyledonous genes may reflect the overall mosaic structure of the genome; that is, G+C-rich "isochores" of several hundred thousand bases surrounding these genes. The present data show clearly that this is not the case for the *Gpa1* and *Gpc1* genes of maize, where G+C enrichment is a local phenomenon, maintained independently of the surrounding (non-coding) sequences, which are G+C-poor (Fig. 5(a) and (b), Table 2; see also Quigley *et al.*, 1989).

(c) *Promoter structure of the* Gpc1 *gene and anaerobic control of GAPC transcript levels* in vivo

If seedlings from maize and other higher plants are exposed to low oxygen pressure (anaerobiosis), for example during flooding, they maintain their energy supply by increasing their glycolytic capacity and by inducing fermentation to regenerate glycolytic cofactors. About ten major and ten minor proteins are selectively synthesized under these conditions (Sachs *et al.*, 1980), six of which have been functionally identified; the two alcohol dehydrogenase polypeptides ADH1 and ADH2 (Sachs & Freeling, 1978; Ricard *et al.*, 1986), pyruvate decarboxylase (Wignarajah & Greenway, 1976), glucose phosphate isomerase and aldolase (Kelley & Freeling, 1984*a*,*b*), and the shrunken locus enzyme sucrose synthase A (Springer *et al.*, 1986). Anaerobic control is exerted at the transcriptional level (and possibly also at the translational level; Sachs *et al.*, 1980; McElfresh & Chourey, 1988) and potential *cis*-acting sequence elements have been identified by functional analysis and sequence comparisons in the promoter region of the *Adh* genes from maize and pea (Walker *et al.*, 1987; Llewellyn *et al.*, 1987).

These elements, the anaerobic regulatory element (ARE, block I and block II with the hexanucleotide "core" sequence TGGTTT; Dennis *et al.*, 1987) and pyrimidine box I (see Fig. 6(a)), are conserved in the promoter region of the maize *Gpc1* gene. Considering this sequence conservation, it was surprising to find that GAPC1 transcripts decrease in maize primary roots after they have been immersed for 20 hours (see Fig. 7(b)). This suggests that the *Gpc1* gene, in spite of its "typical" promoter structure, is not an anaerobic gene *in vivo*. The ARE element with the core sequence TGGTTT, shown to stimulate anaerobic expression of reporter genes in transgenic maize protoplasts (Walker *et al.*, 1987), apparently is not sufficient to stimulate GAPC1 transcript levels *in vivo*. Since the total level of GAPC transcripts increases, or is at least maintained, under anaerobic conditions (see Fig. 7(a)), our results further predict that at least one of the two additional *Gpc* genes found in maize (see Fig. 1) should be induced by anaerobiosis. The cDNA corresponding to this anaerobic gene has been isolated by Russell & Sachs (1989), and has been termed GAPC3 (anaerobic gene *Gpc3*). The same authors characterized another cDNA, GAPC2, corresponding to a second constitutive gene, *Gpc2*, whose transcript levels decrease under anaerobic conditions. The cloned coding regions of cDNAs GAPC3 and GAPC2, starting at codons Phe99 and Glu86, respectively (see Fig. 3), show 88% and 98% amino acid sequence similarity, respectively, with the constitutive gene *Gpc1* and no or little similarity in the 3′ non-coding regions (Russell & Sachs, 1989). These findings support our view that there are three functional *Gpc* genes in maize (see above and Fig. 1). Experiments are in progress in our laboratory to isolate genes *Gpc3* and *Gpc2* and to compare their promoter structures with that of the *Gpc1* gene.

We thank G. Margérie and colleagues (Laboratory of Hematology, CENG, Grenoble) for providing synthetic oligonucleotides and for their valuable experimental help. We thank M. Sachs (Washington University, St Louis) for sending us the cDNA clones encoding GAPC2 and GAPC3 from maize. This work was funded by grants from the Centre National de la Recherche Scientifique (UA 1178), the Ministère de la Recherche et Technologie (MRT, program: "Génétique et Physiologie des Végétaux supérieurs"), and the Ministère de l'Education Nationale (specific program: "Essor des Biotechnologies"). P.M. is a recipient of a doctoral grant from the MRT.

References

Antequera, F. & Bird, A. P. (1988). *EMBO J.* **7**, 2295–2299.

Biesecker, G., Harris, J. I., Thierry, J. C., Walker, J. E. & Wonacott, A. J. (1977). *Nature (London)*, **266**, 328–333.

Bird, A. P. (1986). *Nature (London)*, **321**, 209–213.

Bird, A. P. (1987). *Trends Genet.* **3**, 342–347.

Bird, A. P., Taggart, M. H., Nicholls, R. D. & Higgs, D. R. (1987). *EMBO J.* **6**, 999–1004.

Brändén, C.-I., Eklund, H., Cambillau, C. & Pryor, A. J. (1984). *EMBO J.* **3**, 1307–1310.
Brinkmann, H., Martinez, P., Quigley, F., Martin, W. F. & Cerff, R. (1987). *J. Mol. Evol.* **26**, 320–328.
Brinkmann, H., Cerff, R., Salomon, M. & Soll, J. (1989). *Plant Mol. Biol.* **13**, 81–94.
Brown, J. W. S. (1986). *Nucl. Acids Res.* **14**, 9549–9559.
Cech, T. R. (1986). *Cell*, **44**, 207–210.
Cedar, H. (1988). *Cell*, **53**, 3–4.
Cerff, R. (1979). *Eur. J. Biochem.* **94**, 243–247.
Cerff, R. (1982). In *Methods in Chloroplast Molecular Biology* (Edelmann, M, Hallick, R. B. & Chua, N.-H., eds), pp. 683–694, Elsevier Biomedical Press, Amsterdam.
Cerff, R. & Chambers, S. E. (1979). *J. Biol. Chem.* **254**, 6094–6098.
Cerff, R. & Kloppstech, K. (1982). *Proc. Nat. Acad. Sci., U.S.A.* **79**, 7624–7628.
Chojeki, J. (1986). *Carlsberg Res. Commun.* **51**, 203–210.
Craik, C. S., Rutter, W. J. & Fletterick, R. (1983). *Science*, **220**, 1125–1129.
Dennis, E. S., Walker, J. C., Llewellyn, D. J., Ellis, J. G., Singh, K., Tokuhisa, J. G., Wolstenholme, D. R. & Peacock, W. J. (1987). In *Plant Molecular Biology 1987* (Wettstein, D. & Chua, N.-H., eds), pp. 407–417, Plenum Press, New York.
Drouin, G. & Dover, G. A. (1987). *Nature (London)*, **328**, 557–558.
Duester, G., Smith, M., Bilanchone, V. & Hatfield, G. W. (1986). *J. Biol. Chem.* **5**, 2027–2033.
Federoff, N. (1983). *Plant Mol. Biol. Rep.* **1**, 27–29.
Feinberg, A. & Vogelstein, B. (1984). *Anal. Biochem.* **137**, 266.
Fink, G. R. (1987). *Cell*, **49**, 5–6.
Frischauf, A.-M., Lehrbach, H., Poustka, A. & Murray, N. (1983). *J. Mol. Biol.* **170**, 827–842.
Gilbert, W., Marchionni, M. & McKnight, G. (1986). *Cell*, **46**, 151–154.
Grandbastien, M.-A., Spielmann, A. & Caboche, M. (1989). *Nature (London)*, **337**, 376–380.
Gruenbaum, Y., Naveh-Many, T., Cedar, H. & Razin, A. (1981). *Nature (London)*, **292**, 860–862.
Hanauer, A. & Mandel, J. L. (1984). *EMBO J.* **3**, 2627–2633.
Hohn, B. (1979). *Methods Enzymol.* **68**, 299–309.
Holland, J. P. & Holland, M. J. (1979). *J. Biol. Chem.* **254**, 9839–9845.
Kelley, P. M. & Freeling, M. (1984*a*) *J. Biol. Chem.* **259**, 673–677.
Kelley, P. M. & Freeling, M. (1984*b*) *J. Biol. Chem.* **259**, 14180–14183.
Llewellyn, D. J., Finnegan, E. J., Ellis, J. G., Dennis, E. S. & Peacock, W. J. (1987). *J. Mol. Biol.* **195**, 115–123.
Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). Editors of *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Marchionni, M. & Gilbert, W. (1986). *Cell*, **46**, 133–141.
Martin, W. & Cerff, R. (1986). *Eur. J. Biochem.* **159**, 323–331.
McElfresh, K. C. & Choury, P. S. (1988). *Plant Physiol.* **87**, 542–546.
Michels, P. A. M., Poliszczak, A., Osinga, K. A., Misset, O., Van Beumen, J., Wierenga, R. K., Borst, P. & Opperdoes, F. R. (1986). *EMBO J.* **5**, 1049–1056.
Nguyen, T., Zelchowska, M., Foster, V., Bergmann, H. & Verma, D. P. S. (1985). *Proc. Nat. Acad. Sci., U.S.A.* **82**, 5040–5044.
Piechaczyk, M., Blanchard, J. M., Sabouty, S. R.-E., Dani, C., Marty, L. & Jeanteur, P. (1984). *Nature (London)*, **312**, 469–471.
Quigley, F., Martin, W. F. & Cerff, R. (1988). *Proc. Nat. Acad. Sci., U.S.A.* **85**, 2672–2676.
Quigley, F., Brinkmann, H., Martin, W. F. & Cerff, R. (1989). *J. Mol. Evol.* in the press.
Ricard, B., Mocquot, B., Fournier, A., Delseny, M. & Pradet, A. (1986). *Plant Mol. Biol.* **7**, 321–329.
Rossman, M. G., Liljas, A., Brändén, C.-I. & Banaszak, L. J. (1975). In *The Enzymes* (Boyer, P. D., ed.), 3rd edit., vol. 11, pp. 67–102, Academic Press, New York.
Russell, D. A. & Sachs, M. M. (1989). *The Plant Cell*, in the press.
Sachs, M. & Freeling, M. (1978). *Mol. Gen. Genet.* **161**, 111–115.
Sachs, M., Freeling, M. & Okimoto, R. (1980). *Cell*, **20**, 761–767.
Salinas, J., Matassi, G., Montero, L. M. & Bernardi, G. (1988). *Nucl. Acids Res.* **16**, 4269–4285.
Schwarz-Sommer, Z., Gierl, A., Klösgen, R. B., Wienand, U., Peterson, P. A. & Saedler, H. (1984). *EMBO J.* **3**, 1021–1028.
Sharp, P. A. (1985). *Cell*, **42**, 397–400.
Shih, M.-C., Lazar, G. & Goodman, H. M. (1986). *Cell*, **47**, 73–80.
Springer, B., Werr, W., Starlinger, P., Bennett, D. C., Zokolica, M. & Freeling, M. (1986). *Mol. Gen. Genet.* **205**, 461–468.
Stone, E. M., Rothblum, K. N., Alvey, M. C., Kuo, T. M. & Schwartz, R. J. (1985*a*). *Proc. Nat. Acad. Sci., U.S.A.* **82**, 1628–1632.
Stone, E. M., Rothblum, K. N. & Schwartz, R. J. (1985*b*). *Nature (London)*, **313**, 498–500.
Tischer, E., DasSarma, S. & Goodman, H. M. (1986). *Mol. Gen. Genet.* **203**, 221–229.
Tso, J. Y., Sun, X.-H. & Wu, R. (1985). *J. Biol. Chem.* **260**, 8220–8228.
Voytas, D. F. & Ausubel, F. M. (1988). *Nature (London)*, **336**, 242–244.
Walker, J. C., Howard, E. A., Dennis, E. S. & Peacock, W. J. (1987). *Proc. Nat. Acad. Sci., U.S.A.* **84**, 6624–6628.
Weeden, N. F. (1981). *J. Mol. Evol.* **17**, 133–139.
Werr, W., Frommer, W. B., Maas, C. & Starlinger, P. (1985). *EMBO J.* **4**, 1373–1380.
Westhoff, P., Nelson, N., Bünemann, H. & Herrmann, R. G. (1981). *Curr. Gen.* **4**, 109–120.
Wignarajah, K. & Greenway, H. (1976). *New Phytol.* **77**, 575–584.
Yarbrough, P. O., Hayden, M. A., Dunn, L. A., Vermersch, P. S., Klass, M. R. & Hecht, R. M. (1987). *Biochim. Biophys. Acta*, **908**, 21–33.

Edited by N.-H. Chua

Note added in proof. After submission of this paper John C. Rogers (Washington University, School of Medicine, St Louis) sent us CpG, GpC and G+C profiles of the barley genes encoding α-amylase (Amy32b: 3 introns, about 2 kb) and aleurain (7 introns, about 4 kb), two genes that are strongly and moderately induced, respectively, by giberellic acid in barley aleurone cells (see Whittier *et al.* (1987). *Nucl. Acids Res.* **15**, 2515–2535). The two genes, similar in structure to maize genes *Gpa1* and *Gpc1*, respectively, have also strikingly similar profiles: for Amy32b the whole gene region can be considered a CpG island as in the case of *Gpa1* (Fig. 5(a)), while the aleurain gene has an asymmetric bimodal G+C and CpG distribution comparable to that of gene *Gpc1* (Fig. 5(b)). The G+C content in codons and, hence, codon bias variation along the aleurain gene follows this bimodal distribution. This suggests that our findings of a causal relationship between CpG islands and codon bias in maize genes may be extended to monocotyledonous genes in general. It is likely that this relationship also exists for CpG island carrying vertebrate genes (see M. Gardiner-Garden & M. Frommer (1987). *J. Mol. Biol.* **196**, 261–282), although long-range G+C fluctuations in the genomes of warm-blooded vertebrates (G. Bernardi *et al.* (1985). *Science*, **228**, 953–958; S. Aota & T. Ikemura (1986). *Nucl. Acids Res.* **14**, 6345–6355) may be a major and independent cause for codon usage variation in these organisms.