The 2nd Korea-Germany Joint Symposium in
Plant Biotechnology. p. 39-55 (1994)

CONSERVATION AND FUNCTION OF INTRONS IN PLANT GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE GENES

William Martin[1]

[1]Institut fur Genetik, TU Braunschweig, D-38023 Braunschweig, FRG

I. Introduction

Are introns as old as the genes in which they reside and did they play a role in the origin of genes *via* exon shuffling during the earliest phases of evolution? This question, though simply phrased, has remained very difficult to answer over the years (for reviews, see 21, 30, 67). From a historical perspective, protein chemists had recognized prior to the advent of DNA sequencing techniques that structural domains, *e.g.* for cofactor binding, can be shared by enzymes unrelated at the level of amino acid sequence (6, 57). Subsequent to the discovery of spliceosomal introns in 1977, Gilbert (1978) suggested that introns may have provided a means of accelerated mosiac evolution of genes by virtue of exon shuffling. Doolittle (1978) and Darnell (1978) extended that idea by reasoning that introns might have been involved in the assembly of genes in the first phases of life, later to have been lost via "streamlining" from prokaryotic genomes. Soon thereafter, Blake (1978) coupled Gilbert's idea of exon shuffling to the recombination of "building blocks" of protein structure that protein chemists had described. Thus, by the end of 1978 what is now known as the "introns early" view had crystallized, later to be christened the "exon theory genes" (27). Opponents of this view argued that introns in nuclear protein coding genes arose late in evolution through insertion into contiguous genes by means perhaps related to DNA transposition (9, 10, 51, 56) and advovated the case which in time came to be known as "introns late".

Numerous reports have since appeared in which the existence and utility of exon shuffling during the course of eukaryotic evolution has been demonstrated beyond doubt, for example in the evolution of the low density lipoprotein receptor (69, 70) or serin proteases (55). And cases have also been documented beyond doubt in which spliceosomal introns did in fact arise through insertion of transposable elements during the ephemeral course of this century (47). But an unequivocal answer to the question at hand, namely the role which introns might have played in the assembly of those genes possessed by the progenote - the postulated common ancestor eukaryotes, eubacteria and archaebacteria - is still lacking. One means to address this question is to study genes which are old enough to have been present in the genome of hte progenote, genes for glycolytic enzymes for example, and search for conservation of intron positions across distantly related genes or correlations between intron position and protein structure. Since gylcolysis is suspected to be one of the most

ancient metabolic pathways (24) and since tertiary structures for all of tis constituent enzymes have been determined from at least one source (25), it is a promising source for data on the age and origin of introns. Two enzymes of the glycolytic pathway, triosephosphate isomerase (TIM) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) have figured prominently in the "introns early vs. introns late" debate. TIM from maize was the first gene for which precise intron conservation across the plant-animal boundry could be demonstrated, at the same time, intron positions in the gene and domain boundries in the protein corresponded well (22, 28, 72). TIM has become a prime example for evidence supporting "introns early" although convincing correlations between gene structure and protein structure were also observed for amnyu other genes, *e.g.* maize ADH (3) or human GAPDH (68). Investigations of GAPDH genes of plants, on the other hand, have brought forth some intriguing data concerning both modular protein structure and intron age, the details of which are a complicated matter and will be covered in following sections. In the final section, potential functions for GAPDH introns, regardless of their age, other than exon shuffling will be addressed.

II. Plant Glyceraldehyde-3-phosphate Dehydrogenase

In higher plants, glycolytic and Calvin-cycle pathways possess a number of enzymatic reactions in common which are catalysed by dinstinct enzymes unique to each. GAPDH enzymes are integral to both pathways. Photosynthate generated by the Calvin cycle or by starch degradation in plastids is transported to the cytosol primarily in the form of divalent dihydroxyacetone phosphate in exchange (64) for inorganic phosphate *via* the phosphate translocator. The reductive step of the Calvin cycle is catalyzed by the $NADP^+$-spcific glyceraldehyde-3-phosphate dehydrogenase of chloroplasts (GapA9B) : EC 1.2.1.13). Calvin cycle GAPDH is a marker enzyme of chloroplasts (11) and exists in angiosperms as either an A_4 homotetramer or an A_2B_2 heterotetramer. Subunit A of chloroplast GAPDH (GapA) is of slightly lower molecular weight than subunit B (GapB) (12) by virtue of a highly charged 30 amino acid carboxyterminal extension in the latter (4). The chloroplast enzyme may also show activity with NAD^+ as a substrate *in vitro* (11) and its activity is regulated at the protein level *via* the thioredoxin system (15). *GapA* and *GapB* gene of higher plants are subject to transcriptional activation by light (17, 76). Transcripts of plant GAPDH genes are translated on free cytosolic ribosomes, *GapA* and *GapB* of the chloroplast are synthesized as precursors which are imported into the organelle upon cleavage of an amino-terminal transit peptide (4).

In the cytosol, photosynthate exported from chloroplasts as dihydroxyacetone phosphate or 3-phosphoglycerate can be channeled into glycolysis for the generation of metabolic energy *via* the tricarboxylic acid cycle and oxidative phosphorylation in mitochondria. The oxidative step of glycolysis is catalyzed by the NAD^+-specific glyceraldehyde-3-phosphate dehydrogenase of the cytosol (EC 1.2.1.12). Cytosolic GAPDH in angiosperms is a homotetramer of identical or electrophoretically distinguishable (58) subunits (*GapC*). *GapC* genes of angiosperms can either be constitutively expressed or are inducible by environmental stimuli (46, 76). The three dimensional

structure of GAPDH holoenzyme has been determined from both prokaryotic and eukaryotic sources (1, 29). Although no cystal structures have yet been reported for plans, codon numbering in plant GAPDH conventionally corresponds to that of the Bacillus stearothermophilus enzyme, the structure of which has been resolved at 1.8 Å (63), to permit correlation between primary and tertiary structure.

Although *GapA (B)* and *GapC* genes are nuclear encoded, both *GapA (B)* and *GapC* of eukaryotes are descendants of a prokaryotic gene family which existed in the common ancestor of extant eubacteria. As for many other genes found today in plant nuclei, GAPDH genes were acquired from organellar genomes *via* endosymbolic gene transfer during the course of eukaryotic evolution (34, 40, 45). This is an important point when considering intron conservation across the *GapA (B)/GapC* boundry.

III. Conservation of Intron Positions in Plant GAPDH Genes

Many GAPDH gene sequences from eukaryotes have been determined, and a number of introns have been found in these which lend support to the exon theory of genes. Two introns strictly conserved in nuclear genes encoding chloroplast and cytosolic GAPDH were reported in 1988 (53, 61) and were regarded as strong evidence in favor of the "introns early" hypothesis (20). The position of one of these, at G•ly 166 corresponds quite closely to the boundry between the NAD^+-binding and catalytic domains in the three dimensional structure of the enzyme (53). But for critics of the "introns early" hypothesis, this evidence was not convincing and the identity of intron positions in GAPDH was dismissed as "parallel insertion of different introns" (51). But gene structures recently determined for further GAPDH genes from plants revealed several new intron positions which are precisely conserved between chloroplast and cytosolic GAPDH genes (34). Figure 1 shows a detailed alignment of the exon sequences flanking the five introns precisely conserved between chloroplast (*GapA(B)*)- and cytosolic (*GapC*) GAPDH genes. Clearly, in all five pairwise comparisons, introns interrupt the *GapA(B)* and *GapC* coding sequences at identical positions in regions of unambiguous amino acid sequence conservation. This is also true for intron 25 (position 111-0), where the two flanking codons 110 and 111 have changed from Gln/Ala in *Chlamydomonas GapA* to Lys/Gly in maize *GapC4*. The probability of observing these five identical intron positions across the *Gap(B)/GapC* boundary by chance due to independent insertion was estimated to be roughly 2×10^{-5}. Although "introns late" supporters have argued that the probability is in fact much higher due to the higher likelihood of intron insertion into "protosplice sites" (41, 65), independent insertion appears to be a rather unlikely explanation for the conservation of intron positions observed, because insertion scenarios cannot account for other important aspects of these data (13).

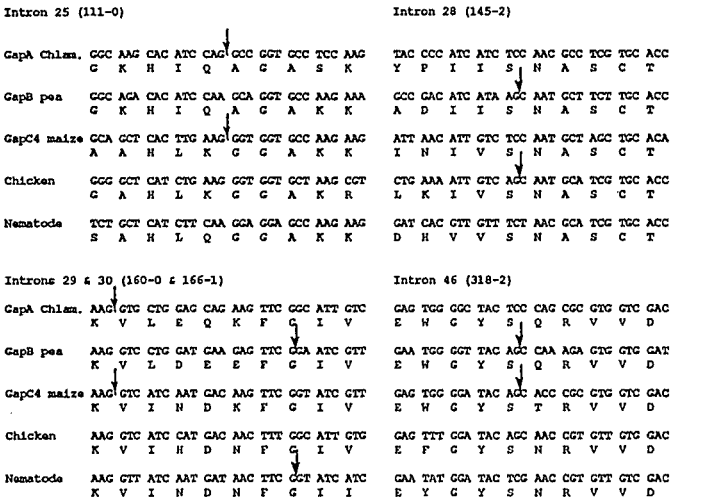

Figure 1. Conservation of intron-exon junctions between chloroplast and glycolytic GAPDH genes, *GapA(B)* and *GapC*, respectively. Arrows indicate intron positions. Codon numbering refers to the sequence of *B. stearothermophilus* (1).

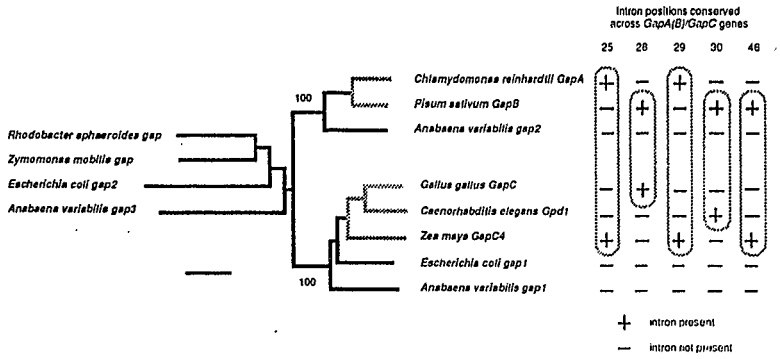

Figure 2. Phylogenetic tree for GAPDH sequences showing intron conservation patterns. The tree was inferred by the neighbor-joining method from a divergence matrix of nonsynonymous substitutions per nonsynonymous site. Branches bearing genes found in eubacterial genomes are shown as solid lines, those bearing genes found in eukaryotic genomes are shown as gray lines. The scale bar indicates 0.1 substitutions per site. The numbers above the *GapA* and *GapC* branches indicate that these were found in 100/100 bootstrap parsimony replicates (PAUP, version 3.0) for the amino acid alignment. Sources of sequences are given in (34).

In Figure 2, the same five sequences have been incorporated into a phylogenetic tree together with the GAPDH genes of the cyanobacterium *Anabaena variabilis* (genes *gap1*, *gap2* and *gap3*), the γ-purple bacterium *E. coli* (genes *gap1* and *gap2*) and the α-purple bacteria *Rhpdobacter sphaeroides* and *Zymomonas mobilis*. the tree topology and the corresponding bootstrap values clearly show that the duplication event that gave rise to chloroplast and cytosolic GAPDH genes of eukaryotes occurred long before the separation of distinct organismal lineages leading to present day eukaryotic and eubacterial chromosomes. Thus, either the duplication event which gave rise to *GapA(B)* and *GapC* genes occurred in ancient eubacteria and the eukaryotic genes are of endosymbiotic origin, or the gene duplication took place in the common ancestor of eubacteria and eukaryotes. In the former case, the five indisputably identical spliceosome intron positions in *GapA(B)/GapC* (Fig. 1) were occupied in eubacterial GAPDH genes long before the divergence of purple- and cyanobacteria (*E. coli* and *A. variabilis*). In the latter case, they were occupied in the ancestral GAPDH gene of the progenote.

Although the five introns conserved across GapA(B) and GapC genes are removed today with the aid of spliceosomes, the data do not indicate how their predecessors in the ancestral GAPDH gene were spliced. The recent discovery of grou II introns in eubacteria (23) has given support to the notion that these elements may have been the precursors from which both spliceosomal introns and snRNAs in eukaryotic nuclei arose (54). Introns in *GapA(B)*- and *GapC*-ancestors within eubacterial chromosomes may have been mechanistically group II, and if so, may have evolved into contemporary spliceosomal introns *in situ*. The conservation of five introns at identical positions in GAPDH genes which were duplicated in eubacteria (and perhaps in progenotic DNA) lends strong support to the exon theory of genes (27). However, this evidence is still not convincing to advocates of the "introns late" scenario, who recently have argued that the exon theory of genes should be abandoned because it may not be experimentally testable at all (66). Thus, the exon theory of genes is exactly where it was sixteen years ago : an enticing and elegant postulate in search of supportive data sufficiently strong as to convince even its most vehement critics.

IV. The Role of Intron 1 in Maize *GapA1* promoter Function

The majority of debate concerning the function of introns in protein coding genes of eukaryotic nuclei has focussed on the role these have played in accelerating gene evolution *via* exon shuffling, but can introns also play a role in gene expression? Although eukaryotic regulatory elements are commonly located in the promoter and 5' flanking region, sequences involved in transcriptional regulation of mammalian genomes can occasionally be found within introns, as in the case of the a2-collagen (52), troponin I (36) and *rpl*32 (14) genes.

The introns of many plant genes are also known to have a marked influence upon expression, although the mechanisms involved have not been as well-elucidated as in vertebrates. In dicots, several-fold increases in gene expression due to introns have been documented (71, 49) whereas in monocots, introns in transgenes can increase expression up to 100-fold (7, 44, 74) and thus

represent valuable basic tools for genetic engineering (38). In maize, splicing efficiency of pre mRNA correlates with increased expression mediated by some introns, but enhancer-like elements, which have often been suspected to reside in plant intervening sequences (43, 44, 62), have not yet been identified. Importantly, the mere presence of introns in plant chimaeric genes does not neccessarily lead to higher levels of expression. This is evidenced by the findings that i) different introns evoke different levels of stimulation in the context of the same promoter (7, 62), ii) one and the same intron may or may not evoke stimulation in the context of different projoters (38) and iii) introns in some cases can suppress rather than increase gene expression (62). These observations also suggest that factors in addition to splicing, such as the presence of enhancer-like elements, may also be involved in intron-mediated expression in plants. Recent investigations indicate that introns may be involved in the expression of the maize *GapA1* gene.

A schematic structure of the region surrounding the maize *GapA1* promoter is shown in Figure 3a. Since this promoter generates roughly 0.5% of the poly(A)$^+$ mRNA in maize seedlings, high levels of expression could be expected from *GapA1* promoter fusions with β-glucuronidase (GUS). pGpaGus1 contains 1.7 kb of the *GapA1* promoter and 5' UTR regions fused in frame to GUS at the *in vivo GapA1* start codon, but surprisingly, pGpaGus1 yields no detectable GUS activity in transiently transformed maize cells (Fig. 3b). Since both the *GapA1* promoter and coding regions are extremely rich in CpG dinucleotides (53, 32), one could suspect that methylation might be responsible for promotor inactivation. Yet transformation of pGpaGus into BMS cells pretreated with 5-azacytidin, a potent methylase inhibitor, did not increase expression up to detectable levels, whereby luciferase expression of positive control plasmids under the CaMV35S promoter was easily detecteable in 5-azacytidin treated cells (data not shown). Concentrations of 5-azacytidin in excess of (20 mM) were apparently toxin to BMS cells and resulted in a general decrease in gene expression in positive control plasmids.

Since silencing *cis*-elements in the 5' region can diminish gene expression (8, 37), the *GapA1* promoter was examined for the presence of 5' upstream silencing elements by construction and transformation of a deletion series in roughly 200 bp increments from pGpaGus1. None of these constructs produced an increase in GUS-expression up ti detectable levels (data not shown), indicating that 5' silencing elements are not responsible for the low level of gene expression.

The most straightforward interpretation of unsuccesful efforts to obtain GUS activity from the *GapA1* promoter was simply that pGpaGus1 lacked *cis*-elements required for *in vivo GapA1* expression, and that maize *GapA1* introns might contain cis-elements which can restore gene expression. Introns 1 and 2 of the maize *GapA1*-gene reside in theregion encoding the aminoterminal transit peptide of the GapA precursor. pGpaGusS was contructed which contains a *Stu*I cloning site in the 5' UTR of *GapA1* promoter into which introns could be introduced. Transformation of pGpaGusS does not yield detectable GUS activity (Figure 3c). Intron 1 from the *GapA1* gene was introduced into pGpaGusS to generate pGpaSIE1 which was tested in BMS cells (Figure 3c). In contrast to pGpaGusS, oGoaSUE1 us exoressed. By comparison, pGpaSEI1 yields

roughly twofold higher levels of GUS activity than the CAMV 35S promoter-driven plasmid pRT103GUS (73) and roughly eightfold lower activity than pEmuGN, in which GUS is under the control of the synthetic Emu promoter (38), one of the strongest monocot promoters currently available (Figure 3e). In order to rest for the presence of potential enhancer-like activity, intron 1 was cloned into pGpaGus in either orientation both 5' of the promoter and 3' of the OCS polyadenylation signal. None of these four constructs yielded detectable levels of GUS activity (Figure 3d), suggesting that intron 1 effects gene expression from the *GapA1* promoter in a position-dependent manner, rather than position independent, as would be expected for a general enhancer element.

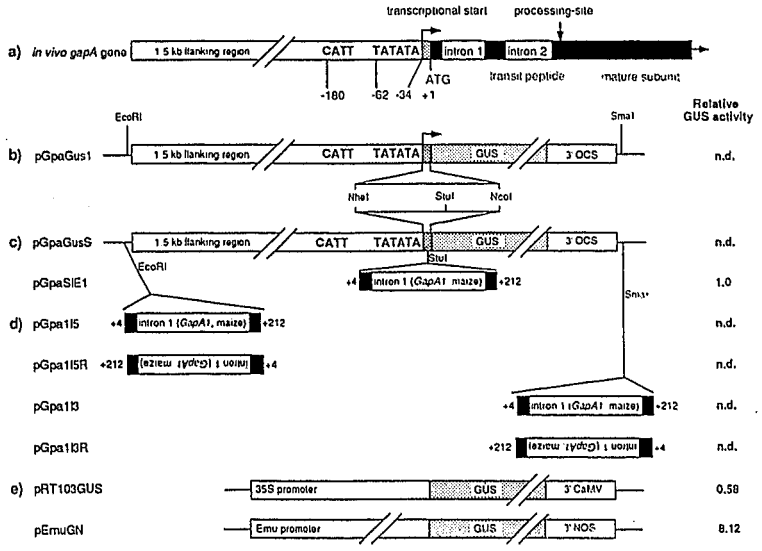

Figure 3. Manipulations of the maize *GapA1* gene. The GUS activity of pGpaSIE1 was set to 1.0, n.d. not detectable. Numbering coordinates of the *GapA1* sequence (*i.e.* +4 and +212 in pGpaSIE1) start from +1 at the adenosine residue of the maize *GapA1* start codon. Lengths of exon border sequences are given. Plasmids were constructed by standard techniques (59) and checked by sequencing (60). BMS cells wer grown in MS medium (48) and transformed by the particle gun technique. GUS activity was assayed (33) with luciferase under control of the CaMV35S promoter as an internal standard (18). a) Schematic structure of the 5' region of the gene. b) *GapA1*-promoter fusion with the GUS gene from pGUS1 (Botterman, PGS, Gent). A 39 bp *NheI/NcoI* fragment from the 5' UTR was substituted by a 46 bp *NheI/NcoI* fragment including a *StuI* cloning site. c) Intron 1 of the maize *GapA1* gene was isolated as a 217 bp *NcoI/SacI* fragment (-1 to +216), the protruding ends of which were removed by mung bean nuclease. the resulting fragment (+4 to +212) was cloned into the StuI site of pGpaGusS to yield pGpaSIE1. **d)** The same blunt *NcoI/SacI* fragment of intron 1 was also cloned in both orientations into the *Eco*RI site and into the *Sma*I site of pGpaGus1 to yield plasmids pGpa1I5, pGpa1I5R, pGpa1I3 and pGpa1I3R respectively. e) Control plasmids used for comparison and their relative GUS activity.

Previous studies of intron-mediated increase in gene expression implicated the splicing process itself as a potential mechanism, since spliced transcripts might be more efficiently transported to the cytosol for translation (5, 31). To test this possibility for *GapA1*, intron 2 of maize *GapA* with sufficient exon border regions for efficient splicing was cloned into the *StuI* site of pGpaGusS. This plasmid, pGpaSIE2 does not generate detectable levels of GUS expression in transiently transformed maize (Figure 4b), indicating that the mere presence of an intron is not sufficient to restore activity to the *GapA1* promoter. Neither pGpaSIE1R nor pGpaSIE2R, which contain intron 1 and 2 or *GapA1* in inverse orientation, respectively, confer GUS activity to maize cells (Figure 4b). Taken in total, these results suggest that intron 1 in the proper orientation specifically restores *GapA1* gene expression in a position dependent manner and is required for expression of the gene.

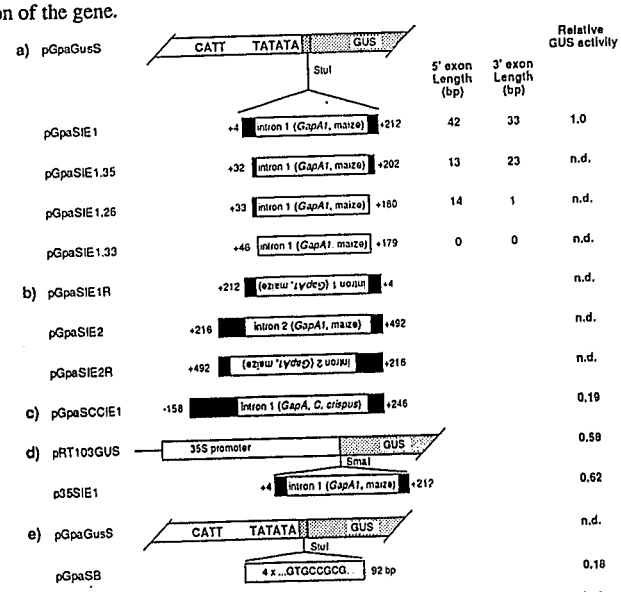

Figure 4. a) Transient expression levels of *GapA1* plasmid constructs containing intron 1 with varying exon border regions. Constructs pGpaSIE1.35, pGpaSIE1.26 and pGpaSIE1.33 were generated *via* PCR. Amplifications products were cloned blunt into the StuI site of pGpaGusS. b) Transient expression levels of *GapA1* plasmid constructs containing intron 1 in the inverse orientation (pGpaSIE1R) and intron 2 of the maize *GapA1* gene in either the correct (pGpaSIE2) or inverse (pGpaSIE2R) orientation. c) Intron 1 of the red algae Chondrus crispus *GapA* gene was isolated as a 408 bp *SacI/NaeI* fragment, the protruding ends of which were removed by mung bean nuclease. The resulting fragment (-158 to +246) was cloned into the *StuI* site of pGpaGusS (pGpaSCCIE1) in spliceable orientation. b) The same blunt *NcoI/SacI* fragment of intron 1 as described in Figure 3 was also cloned in spliceable orientation into the *SmaI* site of pRT103GUS to yield p35SIE1. c) Transient expression levels of a plasmid construct containing a tetramer of the protected sequence identified in footprint assays. The 92 bp *EcoRV/SmaI* fragment of pSK+T was cloned into the StuI site of pGpaGusS to yield pGpaSB. See text for additional details of plasmids.

The functional promoter construct pGpaSIE1 was used to generate a series of 5' deletion mutants from the original 1.4 kb *GapA1* promoter fragment in order to define the minimal promoter. GUS activity measured in the transient expression assay for these mutants revealed that plasmids containing greater than 250 bases of the *GapA1* promoter do not differ markedly in expression from pGpaSIE1. Only after removal of the CAAT box was a significant decrease in GUS activity measurable, removal of the TATA box resulted in a decrease of expression below limits of detection (data not shown). The minimal promoter so defined encompasses roughly 250 bp upsteam of the *in vivo* transcription start. The maize *GapA1* minimal promoter appears also to include intron 1.

In order to determine whether intron 1 can stimulate expression of other promoters, it was introduced into the 5' UTR of pRT103GUS to yield p35SIE-transformed cells contain only 7% greater GUS activity than the pRT103GUS control, suggesting that intron 1 does not contain a general enhancer of trnascription, but leaving open the possibility that it contains enhancer-like elements which attain function in the context of the natural promoter.

The length of bordering exon sequences is known to influence intron-mediated stimulation of gene expression at the level of splicing (62). In order to determine whether this also holds true for *GapA1* intron 1, fragments which contained shorter or no *GapA1* exon border sequences were amplified and introduced these into the StuI site of pGpaGusS (Figure 4a). By reducing the length of the *GapA1* exon 1 and exon 2 border sequences from 42 and 33 bp in pGpaSIE1 to below 14 and 23 bp respectively, GUS expression fell below detectable levels, suggesting that for *GapA1* intron 1, as for other stimulating introns in maize, exon border sequences are critical to intron mediated expression increase. Non-spliceable intron 1 derivatives lacking terminal guanosine residues at the acceptor and domor sites respectively also yielded no GUS activity (data not shown). By analogy to other similar experiments on maize introns (43, 62) it seems likely that the decrease in expression observed resulted from decreased splicing efficiency, although the possibility cannot be fully excluded that *cis* elements most straightforward interpretation of these results, in light of previous studies on exon border length requirements in maize, is that intron 1 of *GapA1* has minimum nborder requirements for proper splicing and that the loss of expression observed for pGpaSIE1.35, pGpaSIE1.26 and pGpaSIE1.33 is simply due to decreased splicing efficiency as the result of lacking exon borders.

All *GapA* and *GapB* genes of plants (including red algae) studied to date possess an intron in the transit peptide coding region at a position similar to that found for intron 1 in the maize *GapA1* gene (Figure 5), suggesting that intron 1 may have been present in the ancestral *GapA* gene of plants. All other introns in the rhodophyte *GapA* genes appear to have been lost during evolution (39, 77). Results described above suggested that selective pressure for retention of intron 1 may exist in maize and introduced the possibility that the same holds true for red algae. A fragment containing intron 1 from the *GapA* gene of the red alga *Chondrus crispus* was cloned into pGpaGusS. BMS cells transformed with this plasmid (pGpaSCCIE1) express 19% of the GUS activity found in cells transformed with pGpaSIE1 containing the maize *GapA1* intron (Figure 4a).

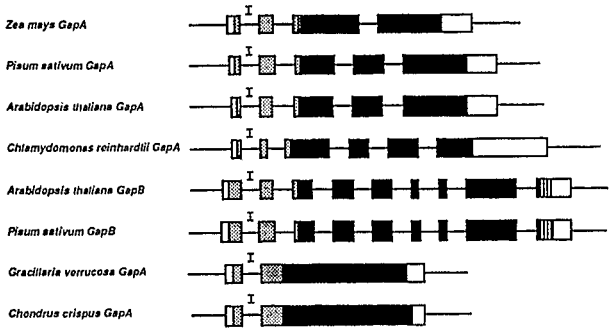

Figure 5. Schematic structure of plant *GapA(B)* genes. Boxes indicate exons. Open boxes : 5' and 3' UTRs ; shaded boxes : transit peptide coding regions ; black boxes : mature subunit coding regions, except *GapB*, where vertically hatched boxes indicate the carboxyterminal extension present in the mature subunit. Exon lengths are drawn to scale, the mature subunit of *Chondrus GapA* is 334 amino acids long. Lengths of introns are not drawn to scale. "I" indicates intron 1. Sources of sequences are given (34, 45) except *Gracillaria GapA* (77).

a)	Intron 1, maize *GapA1*	GT....8... C G T G C C G C ...114...AG
	Intron 1, maize *Adh1*	GT..100... a G T G C C G C ...
		417... C G T G C a G^C
	Intron 1, maize *Sh1*	GT..411... C G T G C C G t ...596...AG
b)	Intron 1, maize *GapA1*	GT....8... C G T G - C C G C ...114...AG
	Intron 1, *Chondrus GapA*	GT....2... C G T t t C C G C ...101...AG

Figure 6. Similarity of the GIBF motif to sequences in a) other maize introns known to increase expression and b) intron 1 of *Chonrus crispus GapA1*. Lenghts of intron sequences not shown are given in bp. GT and AG consensus intron border sequences are underlined, ^ indicates the intron exon boundry in the second GIBF motif of *Adh1* intron 1. Bases deviant from the GIBF motif are shown in lower case.

This result indicates that the *C. crispus* promoter proximal intron supplies *cis* elements required for expression which are lacking in pGpaGusS. As with maize *GapA1* intron 1, PCR-generated *C. crispus GapA* intron 1 fragments lacking exon border sequences and terminal guanosine residues at the acceptor and donor sites do not express GUS activity (data not shown). the intron splice site and exon border sequences of the *C. crispus GapA* gene conform well to the higher plant consensus (39), GUS expression emanating from pGpaSCCE1 suggests that transcripts derived from this

plasmid are spliced in maize cells.

The finding that the rhodophyte *GapA* intron can restore expression of the maize *GapA1* promoter suggests that both introns contain conserved and functionally equivalent cis elements. In order to identify sequence motifs which might be involved in protein-DNA interactions in intron 1, the end labelled 217 bp *Nco*I-*Sac*I fragment containing maize intron 1 was used for DNA footprinting experiments. Only one region could be identified which gave reproducible footprints in both upper and lower strand binding experiments (data not shown). Nuclear extracts from maize seedlings contain a protein which binds the sequence 5' CGTGCCGC 3' found 10 bp downstream of the 5' donor splice junction within the intron. This motif was synthesized and cloned in tetrameric head-to-tail form into the *Eco*RI site of pSK+ (Stratagene) to yield pSK+T as a binding substrate. The *XhoI/SpeI* insert of pSK+T binds nuclear factors specifically in gel retardation assays (data not shown). Unspecific competitor DNA does not markedly affect binding whereas addition of unlabelled substrate competes effectively for the nuclear factor which we designate GIBF (*GapA1* Intron Binding Factor). The GIBF octameric binding motif shows similarity to the human transcription factor Sp1 binding consensus (5' GCCCCGCCC 3') (75) and to the anaerobic regulatory element (ARE) of maize (5' GCCGCGCCG 3') (50). Commercially available Sp1 protein did not yield mobility shifts with XhoI/SpeI insert of pSK+T (data not shown).

In order to determine whether the protein-DNA interaction observed in gel retardation assays is functionally involved in *GapA1* gene expression, the tetrameric GIBF motif in pSK+T was excised as an *Eco*RV/*Sma*I fragment and cloned into the *Stu*I site of pGpaGusS. The resulting construct pGpaSB exhibits 18% of the GUS activity generated by pGpaSIE1 in transiently transformed BMS cells (Figure 4c). The GIBF binding motiv alone thus partially restores activity of the *GapA1* promoter. That promoter activity is not restored in pGpaSB to the same degree as intron 1 with exon border sequences can potentially be due to other factors, but the finding that the *GapA1* promoter is functional in he presence of the GIBF motiv does suggest that this element is involved in *GapA1* gene expression.

The first intron of two other maize genes, *Adh1* (7) and *Sh1* (44), were previously shown to stimulate gene expression in a manner similar to that of *GapA1*. Increase of promoter activity for *GapA1*, *Adh1* and *Sh1* through their respective first introns is, in each case, much greater than any increase conferred by the introns upon other promoters such as CaMV35S and the activating properties of *GapA1*, *Adh1* and *Sh1* introns 1 are only observed when these are located 3' of the TATA-box (42, 44). This prompted examination of the first introns and surrounding exon borders of *Adh1* and *Sh1* for the presence of sequences similar to the GIBF motif. *Sh1* intron 1 contains a single copy of the identified GIBF motif with one terminal mismatch (Figure 6) although the exon regions which also stimulate expression to some extent (44) do not. *Adh1* intron 1 contains two copies of the GIBF motif which possess a single mismatch. Interestingly, expression stimulation for internal deletion mutants of *Adh1* intron 1 were reported and it was concluded that "if an enhancer is present, it is very near the splice junctions" (42). Indeed, one of the GIBF motifs in *Adh1* intron 1

contains the 3' splice site (Figure 6). In addition, more recently reported internal deletion mutants of *Adh1* intron 1 which do not stimulate expression, although they are shown here to still contain the second GIBF motif, are not efficiently spliced because the intron is too short (43). Thus, although these GIBF-homologous motifs have not been directly shown to be functional in the *Sh1* and *Adh1* introns, available data do not exclude the possibility that they may also be involved in gene expression as in teh case of *GapA1*. Congruent with this view is the fact that sequences which do not elicit a marked increase in expression such as intron 2 of *GapA1* (Figure 4b) or introns 2 and 3 of *Adh1* and their respective exon border regions (7) also do not contain similarly conserved copies of the GIBF motif.

V. Summary

Introns have been implicated in processes governing both the evolution and the expression of plant nuclear genes. Genes for chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenases (GAPDH) of plants are descendants of an ancient gene family which existed in the common ancestor of extant eubacteria. During eukaryotic evolution, both genes were transferred to the nucleus from the antecedants of present day chloroplasts and mitochondria, respectively. Five spliceosomal introns have been found at positions which are precisely conserved between nuclear genes for this chloroplast/cytosol enzyme pair, providing strong evidence in favor of the "intros early" hypothesis, i.e. that introns were present in the earliest cells, consistent with the notion that introns facilitated the assembly of primordial genes by accelerating the rate of exon shuffling. Intron 1 is conserved in the transit peptide coding region of all known plant chloroplast GAPDH (*GapA*) genes. Through transient expression studies, it has been shown that the maize *GapA1* promoter exhibits a requirement for sequences contained within intron 1 and surrounding exon border regions for expression in BMS cells. Inactive maize *GapA1*-promoter constructs lacking intron 1 become functional upon introduction of either maize intron 1 or its homologue from the *GapA* gene of a rhodophyte, indicating functional homology of the maize and red algal introns. Through gel mobility and DNA-footprint experiments, it has been shown that nuclear extracts of maize contain factors which bind an octameric motif within intron 1 of maize *GapA1*. This binding sequence shows similarity to cis-elements bound by known eukaryotic transcription factors and is furthermore found in other maize introns known to have a stimulating effect on expression. Intron 1 and surrounding exon border regions from the maize *GapA1* gene could be functionally substituted by a tetramer of the identified octameric sequence, indicating that this intron-localized binding motif could be involved in *GapA1* promoter function.

References

1. Biesecker, G., Harris, J.I., Thierry, J.C., Walker, J.E. and A.J. Wonacott. 1977. Sequence and structure of D-glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus*. Nature 266:328-333

2. Blake, C.C.F. 1978. Do genes-in-pieces imply proteins-in-pieces? Nature 273:267-268

3. Branden C.-I., Eklund, H., Cambillua, C., and A.J. Pryor. 1984. Correlation of exons with structural domains in alcohol dehydrogenase. EMBO J. 3:1307-1310

4. Brinkmann, H., Cerff, R., Salomon, M., and J. Soll. 1989. Cloning and sequence analysis of cDNAs encoding the cytosolic precursors of subunits *GapA* and *GapB* of chloroplast glyceraldehyde-3-phosphate dehydrogenase from pea and spinach. Plant Mol. Biol. 13:81-94

5. Buchman, A.R., and P. Berg. 1988. Comparision of intron-dependent and intron-independent gene expression. Mol. Cell. Biol. 8:4395-4405

6. Buehner, M., Ford, G.C., Moras, D., Olsen, K.W., and M.G. Rossmann. 1973. D-glyceraldehyde-3-phosphate dehydrogenase: Three-dimensional structure and evolutionary significance. Proc. Natl. Acad. Sci. USA 70:3052-3054

7. Callis, J., Fromm, M., and V. Walbot. 1987. Introns increase gene expression in cultured maize cells. Genes Dev. 1:1183-1200

8. Castresana, C., Garcia-Lique, I., Alonso, E., Malik, V.S., and A.R. Cashmore. 1988. Both positive and negative regulatory elements mediate expression of a photoregulated *CAB gene* from *Nicotiana plumbaginifolia. EMBO J. 7:1929-1936*

9. Cavalier-Smith, T. 1985. Selfish DNA and the origin of introns. Nature 315:283-284

10. Cavalier-Smith, T. 19915. Intron phylogeny: A new hypothesis. Trends in Genetics 7:145-148

11. Cerff, R. 1982. Separation and purification of NAD-and NADP-linked glyceraldehyde-3-phosphate dehydrogenases from higher plants. In: **Methods in Chloroplast Molecular Biology**, M. Edelmann, R.B. Hallick and N.-H. Chua, eds., Elsevier Biomedical Press, Amsterdam. pp. 683-694

12. Cerff, R., and S. Chambers. 1979. Subunit structure of higher plant glyceraldehyde-3-phosphate dehydrogenases (EC 1.2.1.12 and 1.2.1.13). J. Biol. Chem. 254:6094-6098

13. Cerff, R., Martin, W., and H. Brinkmann. 1994. Origin of introns-early or late? Nature 369:527-528

14. Chung, S., and R.P.Perry. 1989. Importance of introns for expression of mouse ribosomal protein gene *rpl*32. Mol. Cell. Biol. 9:2075-2082

15. Cseke, C., and B.B. Buchanan. 1986. Regulation of the formation and utilisation of photosynthate in leaves. Biochim. Biophys. Acta. 853:43-64

16. Darnell. J.E. 1978. Implications of RNA-RNA splicing in evolution of eucaryotic cells. Science 202:1257-1260

17. Dewdney, J., Conley, T.R., Shih, M.-C., and H. Goodmann. 1993. Effects of blue and red light on expression of nuclear genes encoding chloroplast glyceraldehyde-3-phosphate dehydrogenase of *Arabidopsis thaliana*. Plant Physiol. 103:1115-1121

18. De Wet, J., Wood, K.V., De Luca, M., Helsinki, D.R., and S.Subramani. 1987. The firefly luciferase gene: Structure and expression in mammalian cells. Mol. Cell. Biol. 7:725-737

19. Doolittle, W.F. 1978. Genes in pieces: Were they ever together? Nature 272:581-582

20. Doolittle, W.F. 1989. Whatever happened to the progenote? In : **The Hierarchy of Life**, B. Fernholm, K.Bremer and H. Jornvall, eds., Elsevier, Amsterdam.pp. 65-72

21. Doolittle, W.F. 1990. Understanding introns: Origins and functions. In: **Intervening sequences in Evolution and Development**, E.M. Stone and R.J. Schwartz, eds., Oxford University Press, New York. pp. 43-62

22. Doolittle, W.F., and A. Stoltzfus. 1993. Genes in pieces revisited. Nature 361:403

23. Ferat, J.-L., and F.Michel. 1993. Group II self splicing introns in bacteria. Nature 364:358-361

24. Fothergill-Gilmore, L.A. 1986. The evolution of the glycolytic pathway. Trends Biochem. Sci. 11:47-51

25. Fothergill-Gilmore, L.A., and P.A.M. Michels. 1993. Evolution of glycolysis. Progr. Biophys. Mol. Biol. 59:105-238

26. Gilbert, W. 1978. Why genes in pieces? Nature 271:501

27. Gilbert, W. 1987. The exon theory of genes. Cold Spring Harbor Symp. Quant. Biol. 52:901-924

28. Gilbert, W., Marchionni, M., and G. McKnight. 1986. On the antiquity of introns. Cell 46:151-154

29. Harris, J.I., and M. Waters. 1976. Glyceraldehyde-3-phosphate dehydrogenase. In: **The Enzymes**, P.D. Boyer, ed., Academic Press, New York. Vol. XIII, pp. 1-50

30. Holland, S.K., and C.C.F. Blake. 1990. Proteins, exons and molecular evolution. In: **Intervening Sequences in Evolution and Development**, E.M. Stone and R.J. Schwartz, eds., Oxford University Press, New York. pp. 10-42

31. Huang, M.T.F., and C.M. Gorman. 1990. Intervening sequences increase efficiency of RNA 3' processing and accumulation of cytoplasmic RNA. Nucl. Acids Res. 18:937-947

32. Jansson, S., Meyer-Gauen, G., Cerff, R., and W. Martin. 1994. Nucleotide distribution in gymnosperm nuclear sequences suggests a model for GC-content change in land plant nuclear genomes. J. Mol. Evol. 39:34-46

33. jefferson, R.A., Kavanagh, T.A. and M.W. Bevan. 1987. GUS-fusion: β-Glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. 6:3901-3907

34. Kersanach, R., Brinkmann, H., Liaud, M.-F., Zhang, D.-X., Martin, W., and R. Cerff. 1994. Five identical intron positions in ancient duplicated genes of eubacterial origin. Nature 367:387-389

35. Klein, T.M., Gradziel, T., Fromm, M.E., and J.C. Sanford. 1988. Factors influencing gene delivery into *Zea mays* cells by high-velocity microporjectiles. Bio/Tech. 6:559-563

36. Konieczny, S.F., and C.P. Emerson. 1987. Complex regulation of muscle-specific contractile protein (troponin I) gene. Mol. Cell. Biol. 7:3065-3075

37. Kuhlemeier, C,m Fluhr, R., Green, P., and N.-H. Chua. 1987. Sequences in the pea *rbcS-3A* gene have homology to constitutive mammalian enhancer but function as negative regulatory elements. Genes Devel. 1:247-255

38. Last, D.I., Brettell, R.I.S., Chamberlain, D.A., Chaudhury, A.M., Larkin, P.J., Marsh, E.L., Peacock, W.J., and E.S. Dennis. 1991. pEmu an improved promoter for gene expression in cereal cells. Theor. Appl. Genet. 81:581-588

39. Liaud, M.-F., Valentin, C., Brandt, U., Bourget, F.Y., Kloareg, B., and R. Cerff. 1993. The GAPDH gene system of the red alga Chondrus crispus: promoter structure, intron/exon organization, genomic complexity and differential expression of genes. Plant Mol. Biol. 23:981-994

40. Liaud, M.-F., Valentine, C., Martin, W., Bouget, F.-Y., Klogreg, B., and R. Cerff. 1994. The evolutionary origin of red algae as deduced from the nuclear genes encoding cytosolic and chloroplast glyceraldehyde-3-phosphate dehydrogenases from *Chondrus crispus*. J. Mol. Evol. 38:319-327

41. Logsdon, J.M., and J.D. Palmer. 1994. Origin of introns: Early or late? Nature 369:526

42. Luehrsen, K.R., and V. Walbot. 1991. Intron enhancement of gene expression and the splicing efficiency of introns in maize cells. Mol. Gen. Genet. 225:81-93

43. Luehrsen, K.R., and V. Walbot. 1994. Addition of A- and U-rich sequence increases the splicing efficiency of a deleted form of a maize intron. Plant Mol. Biol. 24:449-463

44. Maas, C., Laufs, J., Grant, S., Korfhage, C., and W. Werr. 1991. The combination of a novel stimulatory element in the first exon of the maize *Shrunken1* gene with the following intron 1 enhances reporter gene expression up to 1000-fold. Plant Mol. Biol. 16:199-207

45. Martin. W., Brinkmann, H., Savona, C., and R. Cerff. 1993. Evidence for a chimaeric nature of nuclear genomes: Eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenase from maize. J. Mol. Biol. 208:551-565

46. Martinez, P., Martin, W., and R. Cerff. 1989. Structure, evolution and anaerobic regulation of a nuclear gene encoding cytosolic glyceraldehyde-3-phosphate dehydrogenase from maize. J. Mol. Biol. 208:551-565

47. Menssen, A., Hohmann, S., Martin. W., Schnable, P., Peterson, P., Saedler, H., and A. Gierl. 1990. The En/Spm transposable element of *Zea mays* contains splice sites at the termini generating a novel intron from a dSpm element in the A2 gene. EMBO J. 9:3051-3057

48. Murashige, T., and F. Skoog. 1963. A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol Plant. 15:473-497

49. Norris, S.R., Meyer, S., and J. Callis. 1993. The intron of *Arabidopsis thaliana* polyubiquitin genes is conserved in location and is a quantitative determinant of chimeric gene expression. Plant Mol. Niol. 21:895-906

50. Olive, M.R., Peacock, W.J., and E.S. Dennis. 1991. The anaerobic responsive element contains two GC-rich sequences essential for binding a nuclear protein and hypoxic activation of the maize *Adh1* promoter. Nucl. Acids Res. 19:7053-7060

51. Palmer, J.D., and J.M. Logsdon. 1991. The recent origin of introns. Curr. Opinion Genet. Devel. 1:470-477

52. Pellegrino, R., and B. De Crombrugghe. 1987. Identification of a cell-specific transcriptional enhancer in the first intron of the mouse a a (type I) collagen gene. Proc. Natl. Acad. Sci. 84:5590-5594

53. Quigley, F., Martin, W., and R. Cerff. 1988. Intron conservation across the prokaryote-eukaryote boundry: Structure of the nuclear gene for chloroplast glyceraldehyde-3-phosphate dehydrogenase from maize. Proc. Natl. Acad. Sci. USA 85:2672-2676

54. Roger, A.J., and W.F. Doolittle. 1993. Why introns-in-pieces? Nature 315:213-216

55. Rogers, J. 1985. Exon shuffling and intron insertion in serine protease genes. Nature 315:458-459

56. Rogers, J. 1989. How were introns inserted into nuclear genes? Trends in Genetics 5:213-216

57. Rossmann, M.G., Moras, D., aand K.W. Olsen. 1974. Chemical and biological evolution of a nucleotide binding domain. Nature 250:194-199

58. Russell, D.A., and M.M. Sachs. 1991. The maize cytosolic glyceraldehyde-3-phosphate dehydrogenase gene family: Organ-specific expression and genetic analysis. Mol. Gen. Genet. 229:219-228

59. Sambrook, J., Fritsch, E.F., and T.Maniatis. 1989. Molecular cloning. CSH Cold Spring Harbor Laboratory Press. Second edition

60. Sanger, F., Nickler, S., and A.R. Coulson. 1977. DNA sequenzing with chain termination inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467

61. Shih, M.-C., Heinrich, P., and H.M. Goodman. 1988. Intron existence predated the divergence of eukaryotes and prokaryotes. Science 242:1164-1166

62. Sinibaldi, R.M., and I.J. Mettler. 1992. Intron splicing and intron-mediated enhanced expression in monocots. Progr. Nucl. Acids Res. Mol. Biol. 42:229-257

63. Skarzynski, T., and A.J. Wonacott. 1988. Coenzyme-induced conformational changes in glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus*. J. Mol. Biol. 203:1097-1118

64. Stitt, M. 1990. The flux of carbon between the chloroplast and the cytosol. In: **Plant Physiology, Biochemistry and Molecular Biology**, D.T. Dennis and D.H. Turpin, eds., Longman,, Singapore. pp. 319-338

65. Stoltzfus, A. 1994. Origin of introns: Early or late? Nature 369:526-527

66. Stoltzfus, A., Spencer, D.F., Zuker, M., Logsdon, J.M. and W.F. Doolittle. 1994. Testing the exon theory of genes: The evidence from protein structure. Science, in press.

67. Stone, E.M., and Schwartz, R.J. 1990. Intron-dependent evolution of progenotic enzymes. In: **Intervening Sequences in Evolution and Development**, E.M. Stone and R.J. Schwartz, eds., Oxford University press, New York. pp. 63-91

68. Stone, E.M., Rothblum, K.N., and R.J. Schwartz. 1985. Intron dependent evolution of the chicken glyceraldehyde-3-phosphate dehydrogenase gene. Nature 313:498-501

69. Sudhog, T.C., Russel, D.W., Goldstein, J.L., Brown, M.S., and D.W. Russel. 1985. The LDL receptor gene: A mosaic of exons shared with different proteins. Science 228:815-822

70. Sudhof, T.C., Russel, D.W., Goldstein, J.L., Brown, M.S., Sanchez-Pescador, R., G.I. Bell. 1985. Cassette of eight exons shared by genes for LDL receptor and EGF precursor. Science 228:893-895

71. Tanaka, A., Mita, S., Ohta, S., Kyozuka, J., Shimamoto, K., and K. Nakamura. 1990. Enhancement of foreign gene expression by a dicot intron in rice but not in tobacco is correlated with an increased level of MRNA and an efficient splicing of the intron. Nucl. Acids. Res. 18:6767-6770

72. Tittiger, C., Whyard, S., and V. Walker. 1993. A novel intron site in the triosephosphate isomerase gene from the mosquito *Culex tarsalis*. Nature 361:470-472

73. Topfer, R., Prols, M., Schell, J., and H.H. Steinbiβ. 1988. Transient gene expression in tobacco protoplasts: II Comparision of reporter gene systems for CAT, NPTII and GUS. Plant Cell reports 7:225-228

74. Vasil, V., Clancy, M., Ferl, R.J., Vasil, I., and L.C. Hannah. 1989. Increased Gene Expression by the first intron of maize *Shrunken-1* Locus in grass species. Plant Physiol. 91:1575-1579

75. Westin, G., and W. Schaffner. 1988. Heavy metal ions in transcrioption factors from HeLa cells: Sp1, but not octamer transcrioption factor requires zinc for DNA binding and for activator function. Nucl. Acids Res. 16:5771-5781

76. Yang, Y., Kwon, H.B., Peng, H.P., and M.-C. Shih. 1993. Stress responses and metabolic regulation of glyceraldehyde-3-phosphate dehydrogenase genes in *Arabidopsis*. Plant Physiol. 101-209-216

77. Zhou, Y.H., and M.A. Ragan. 1994. Cloning and characterisation of the nuclear gene encoding plastid glyceraldehyde-3-phosphate dehydrogenase from the marine red alga *Gracillaria verrucosa*. Curr. Genet. in press.