Translation of chloroplast-encoded mRNA: potential initiation and termination signals

Peta C.Bonham-Smith¹ and Don P.Bourque^{1,2}

¹Department of Biochemistry and ²Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ 85721, USA

Received September 20, 1988; Revised and Accepted February 2, 1989

ABSTRACT

A survey of 196 protein-coding chloroplast DNA sequences demonstrated the preference for AUG and UAA codons for initiation and termination of translation, respectively. As in prokaryotes at every nucleotide position from -25 to +25 (AUG is +1 to +3) and for 25 nucleotides 5' and 3' to the termination codon an A or U is predominant, except for C at +5 and G at +22. A Shine-Dalgarno (SD) sequence (GGAGG or tri- or tetranucleotide variant) was found within 100 bp 5' to the AUG codon in 92% of the genes. In 40% of these cases, the location of the SD sequence was similar to that of the consensus for prokaryotes (-12 to -7 5' to AUG), presumed to be optimal for translation initiation. A SD sequence could not be located in 6% of the chloroplast sequences. We propose that mRNA secondary structures may be required for the relocation of a distal SD sequences to within the optimal region (-12 to -7) for initiation of translation. We further suggest that termination at UGA codons in chloroplast genes may occur by a mechanism, involving 16S rRNA secondary structure, which has been proposed for UGA termination in E.coli.

INTRODUCTION

Mechanisms which might regulate expression of chloroplast-encoded genes at the level of translation are unknown. We compared 196 published chloroplast gene sequences to identify potential signals of importance for translational regulation. In our search we assumed that sequences (and/or location of individual nucleotides) important for translational regulation are likely to be evolutionarily conserved, relative to non-regulatory regions of chloroplast transcripts. The many prokaryotic features of chloroplasts (1-3) suggest that translation of chloroplast genes may share regulatory mechanisms similar to those proposed for prokaryotic translation systems.

Prokaryotic mRNA species contain several specific sequence-related features thought to be important in initiation and

termination of translation. An initiation codon [AUG, GUG or UUG (4), AUA (5), AUU (6) or ACG (7)], is located at the start of translation (nucleotides +1 to +3). A Shine-Dalgarno (SD) sequence (8), GGAGG or a tri- or tetranucleotide variant thereof, is generally located 7 ± 4 bp upstream of the initiation codon (9). An A is predominant at position -1 (10) and -3 (9), a purine is most abundant at position +4 (10) and there is a preference for AAA (Lys) or GCU (Ala) as the second translated codon (10). Secondary structural elements of messenger RNA may also modulate the translation initiation rate by either masking the SD sequence (11,12) or the AUG initiation codon (13) or by bringing a distal SD sequence closer to an initiation codon (14). Translation of prokaryotic transcripts usually terminates at an UAA (15) but, occasionally it terminates at UGA or UAG. We have identified examples of all these prokaryotic sequence elements among the 196 chloroplast-encoded transcripts studied.

From our results we infer that initiation and termination of translation from chloroplast-encoded mRNA transcripts may occur by specific mechanisms already described for prokaryotic genes. Our results further document the close evolutionary relationship between prokaryotic and chloroplast genes.

METHODS

All nucleotide frequency analysis was carried out manually. The nucleic acid folding data was generated on an IBM PS/2 Model 80 computer, using the PCFOLD program of Zuker (16). The data in Table 2 and Figure 4 were normalized with respect to the A and T rich (62%) nature of tobacco chloroplast DNA. In addition, the data presented in Figure 1 was obtained after normalizing with respect to differences in the number of sequences available for each of the 56 genes studied.

RESULTS AND DISCUSSION

1. Initiation codon flanking sequences

Analysis of base frequency occurrence in the 25 nucleotide sequence 5' and 3' to the initiation codon was performed for 196 mRNA sequences deduced for 56 chloroplast-encoded genes (see Appendix A for references). Results show a preference for A or U

Figure 1. Number of occurrences (frequency) of A,U,G and C in the 25 nucleotide region 5' and 3' to the initiation codon of 196 sequences of 56 chloroplast-encoded genes (see Appendix A for references). The number of sequences available for comparison for each of the 56 genes varied. Therefore, the numbers (frequency) presented for each nucleotide position in this figure were normalized proportionally to a sample size of 56, the total number of genes studied. The values expected by random occurrence (14) and those observed for each nucleotide are indicated by the solid and dashed lines, respectively.

in all positions except C at +5 and G at -11 to -9, +4 and +22 (Figure 1). The normalized number of occurrences of each nucleotide within this 50 nucleotide region is 21 (38%) A, 18 (32%) U, 10 (17%) G and 7 (13%) C. A similar preference for A/U, C at +5 and G at +22, in the flanking regions (both 5' and 3') occurs in prokaryotic mRNA transcripts, but exceptions are the G-rich SD sequence at -7 ± 4 and a C at +21 (17). The occurrence of a highly conserved C at position +5 in chloroplast and prokaryote transcripts suggests requirement in assembly of the translation initiation complex. Analysis of point mutations are necessary to evaluate this hypothesis. Normalized values for chloroplast-encoded mRNAs show a predominance of U at position -1 (31

Table 1. Frequency of dinucleotides immediately 3' to the initiation codon of 196 chloroplast genes.

Position +1 +5	% of mRNA ^a
AUG GA	7
AUG GU	6
AUG GC	16
AUG GG	8
AUG AA	16
AUG AU	7
AUG AC	11
AUG AG	6
AUG CA	2
AUG CU	5
AUG CC	2
AUG UU	4
AUG UC	10

^a Values rounded to nearest %.

sequences, 56%), A at -3 (24 sequences, 43%) and a purine at +4 (43 sequences, 77%), a feature also found in eukaryotes (82% of *Drosophila* mRNAs and 81% of yeast mRNAs (9)). *In vitro* experiments with prokaryotic ribosomes and synthetic AUG-containing oligonucleotides show a higher efficiency of ribosome binding when an A, U and purine are found in positions -3, -1 and +4, respectively (18-20). The G frequency peaks at position -10, which is consistent with the location of the G-rich SD sequences in prokaryotes and chloroplasts (see Section 3).

Unlike prokaryotic mRNAs (10), the chloroplast-encoded mRNAs have a low frequency of AAA (10% of genes studied) as the second codon and GCU did not occur in this position. Thirteen different dinucleotides occupy positions +4 and +5 with AA [Asn/Lys] (16%), GC [Ala] (16%), AC [Ser/Arg] (11%) and UC [Ser] (10%) being the combinations of preference (Table 1). If codon usage for the second codon position were random, only 6% of the 196 chloroplast gene sequences would contain these dinucleotides. The dinucleotides CG, UA and UG did not occur in the 196 sequences studied. Lack of the UA dinucleotide in the second codon position, and decreased use of CG and UG dinucleotides in all codons was recently reported in both chloroplast and plant

Table 2. Comparison of the consensus nucleotide sequence from -12 to +4 (relative to the initiation codon in chloroplast genes) of prokaryotic, plant nuclear and eukaryotic genes.

<u>GENOME</u>		<u>5'-FLANKING SEQUENCE^a</u>	<u>.CODING SEQUENCE</u>
		-12	+1
C^b	Chloroplast (56 genes, 196 sequences)	G G G G A G A T C A T T	A T G G
C:P	HOMOLOGY SCORE^c	* * * * o o o * *	* * * * o
P	Prokaryotic (123 sequences)	G G G G G A T A A A T A A A A A A T A T	A T G A G
C:PN	HOMOLOGY SCORE		* * * *
PN	Plant nuclear (79 sequences)	T A A T A A T A A A C A A T T T	A T G G
C:E	HOMOLOGY SCORE		* * * *
E	Eukaryotic (vertebrate, 699 sequences)	C C C G C C G C C A C C G	A T G G

a. The chloroplast consensus sequence was determined by normalizing the sequence of each nucleotide in order to compensate for the high A/T content (62%) of tobacco chloroplast DNA. The top line of each sequence is the predominant nucleotide at this position. Bases indicated on the second line of a sequence are included when the difference in frequency between the two indicated nucleotides is small. The homology score for each pairwise comparison is given above the sequence to which the chloroplast consensus sequence is being compared.

b. Consensus sequences compared are C (chloroplast), P (procaryotic, reference 17), PN (plant nuclear, reference 22), and E (eucaryotic, reference 9).

c. Identity between the predominant nucleotide of the comparison sequence with that of the chloroplast sequence is indicated by an asterisk (*). Identity between the second most common nucleotide of either sequence being compared is indicated by a circle (o).

nuclear-encoded genes (21). Since 32% of the second codons specify the same amino acids as in *E. coli* genes, we consider this feature to be a conserved procaryotic feature of chloroplast genes in spite of the lack of GCU codons at this position. The relative abundance of chloroplast tRNAs with anticodons AAN or GCN is unknown. Our results support Gold's suggestion (14) that conserved sequences at the second codon position may be involved in 50S subunit association with the 30S initiation complex and subsequent initiation of polypeptide synthesis.

We compared (Table 2) the 5' consensus sequences (-12 to +4) of chloroplast-encoded transcripts and leader sequences proposed for plant nuclear-encoded (22), eukaryotic (vertebrate) (9) and

[illegible]

Figure 2. A comparative analysis of mRNA nucleotide sequences from -25 to +25 (AUG occupies +1 to +3) of various (A) *atpB* and (B) *rps12* transcripts. All sequences are compared to tobacco. Upper case letters indicate sequence identity and lower case letters non-identity. Conserved sequence motifs (boxed) are discussed in the text. Tobacco, pea, soybean, and spinach are dicotyledonous plants, *Marchantia* is a liverwort, and maize, barley, and wheat are monocotyledonous plants.

prokaryotic transcripts (17). The chloroplast DNA-encoded consensus sequence from -12 to +4 contains 9 identical predominant nucleotides and 4 positions in which the second most common nucleotide is identical, relative to the procaryotic consensus. The high level of homology between the chloroplast and prokaryotic consensus sequences is due, in part, to the location of G-rich SD sequences (see Section 3) within the -12 to -7 region of both prokaryotic and chloroplast-encoded transcripts. No corresponding region is present in eukaryotic mRNAs (9), and, other than the coding sequence (+1 to +4), few sequence

psbA

	-25	-1
tobacco	UGAUUAAAUAAACCAGAUUUUACC	AUG
mustard	UGAUUAAAUAAACCAGAGUUUUUACC	AUG
spinach	aaAUUAAAUAAACCAGAGUUUUUACC	AUG
s. nigrum	UGAUUAAAUAAACCAGAUUUUACC	AUG
alfalfa	UGAUUAAAUAAACCAGAGUUUUUACC	AUG
petunia	UGAUUAAAUAAACCAGAGUUUUUACC	AUG
pea	augUUAUUAAUAAACCAGAUUUUACC	AUG
rice	aaUUGUAUAUAAACCAGAGUcUUACC	AUG
barley	aaUUGUAUAUAAACCAGAGUcUUACC	AUG

Figure 3. A comparative analysis of the 25 nucleotides 5' to the initiation codon of various *psbA* transcripts. Upper case letters indicate sequence identity with tobacco. Rice and barley are monocotyledonous plants, while all other sequences are from dicotyledonous plants.

identities were found between the chloroplast and plant nuclear or vertebrate consensus sequences (Table 2).

With regard to the 5' leader sequences of atpB transcripts, an apparent low sequence identity is more evident within the dicotyledon plants (32-44%) and between dicotyledon and monocotyledon plants (44-64%) than between the monocotyledons alone (80-100%) (Figure 2A). A lack of sequence identity does not appear to correlate with an absence of a SD sequence in this region. In contrast to the situation among atpB genes, the psbA transcripts of six dicotyledonous and two monocotyledonous plants (which also lack a SD sequence in this region) show a high sequence identity (88-92%) in the 5' leader sequence (Figure 3).

Although possible translation regulatory sequences are expected to occur in the 5' leader region of chloroplast mRNAs, they could also exist 3' to the initiation codon. In the case of atpB transcripts the sequences 5' to the initiation codon (-25 to -1) from several monocotyledon and dicotyledon plants (all of which lack a SD sequence - see Section 3) are generally 40-64% conserved, but the important -3 (A) and -1 (U) positions are invariant. However, comparing the sequences 3' to the initiation codon (+4 to +25), two conserved sequence motifs (UCCU at positions +12 to +15 and CUUC at positions +20 to +23, separated by an ACUA sequence), were observed in all atpB genes examined (Figure 2A and ref 23). These two sequence motifs (CCU and UUC), again separated by ACUA, are also present in the 5' leader of rps12 transcripts from maize and Marchantia, at positions +4 to +13 (Figure 2B). Although the 5' leader region of tobacco and soybean rps12 transcripts contain an ACUA sequence, they both lack the flanking 3' CCU and UUC sequence motifs (Figure 2B). The importance of sequences 3' to the initiation codon in translation initiation, has previously been reported for bacteriophage T7 genes (24,25). Expression of various length fragments 3' to the initiation codon of the T7 early gene 1, which are cloned into a reporter plasmid, shows a requirement for sequences up to 112 nucleotides downstream of the initiation codon in efficient ribosome binding to mRNA of this gene (24). Furthermore, even sequences downstream of nucleotide +112 affected (inhibited) translation in these experiments. The

authors suggest that both effects result from secondary or tertiary structures within the mRNA sequences. Requirements for such conserved sequence or secondary structural elements 3' or 5' to the initiator AUG have yet to be explored for the 6% of chloroplast genes which lack a SD sequence in the expected position and/or which lack other 5' sequences thought to be required for translation initiation.

2. Initiation codon

Initiation of translation from 55 of 56 chloroplast-encoded genes compared in this study occurs at an AUG codon. An exception is the rps19 mRNA coding for the chloroplast small ribosomal subunit protein S19 (26). The coding sequence of this protein starts with a GUG codon in tobacco, spinach, soybean and maize (27-30). However, in Marchantia, S19 initiation occurs at an AUG codon (31). We conclude that the GUG initiation codon in the chloroplast-encoded rps19 transcript is a prokaryotic feature, since a functional GUG initiation codon has yet to be identified in an eukaryotic organism. Initiation of eukaryotic mRNA translation was thought to utilize only AUG (9); however, translation of mRNA for the c-myc proto-oncogene from human, murine, avian and amphibian cells, initiates from an AUG and a CUG codon, resulting in the translation of two different proteins from the same mRNA (32).

3. Shine-Dalgarno sequences

The Shine-Dalgarno (SD) sequence (GGAGG) in the 5' non-coding region of prokaryotic mRNAs forms a base paired region with its complementary (CCUCC) sequence at the 3' terminus of 16S rRNA prior to (and as a prerequisite for) initiation of translation (33,34). A SD sequence, or tri- or tetranucleotide variant thereof, was found within 100 bp 5' to the initiation codon in 181 (92%) of 196 chloroplast-encoded transcripts (see Appendix A). The GGANN (where N is an A, U or C) variant of the SD sequence is most common (found in 32% of the 181 sequences) followed by GGAG (25%), GGAGG (18%), AGG (9%), GAGG (8%) and GAG (7%). In sequences containing a GGAGG SD sequence, it occurs within the first 25 nucleotides 5' to the initiation codon. In 72 (40%) of these 181 sequences, the SD sequence is located within the -12 to -7 region, with a frequency of occurrence

(Figure 4) of GGAGG (23 of 72 sequences) GGAG (17) > GGA (13) > GAGG (11) > AGG (4) > GAG (4). The high incidence of a SD sequence in this region conforms to the optimal location of the SD sequence in prokaryotes (9). Thirty-nine percent of the 41 SD variant sequences found within the -100 to -25 region are of the GGANN type and 19% are AGG. The remaining 42% possess either a GAG or GGAG variant. The GGAGG and GAGG variants are not represented within this group of sequences. A SD sequence (or variant) could not be located within 100 bases 5' to the AUG codon in 11 (6%) of the 196 sequences studied. Although a rare situation, the dnaG gene of E.coli (35) and the C1 repressor proteins of lambda phage (36) and phage 34 (37) are examples of prokaryotic genes which lack a SD sequence. In these cases (both prokaryotic and eucaryotic), and in yet to be discovered examples, other sequence elements not recognizable as SD sequences and/or secondary structural features must occur which permit the formation of functional initiation complexes.

Longer prokaryotic SD sequences may increase the stability of the mRNA - 16S rRNA interaction and favor initiation of translation (14). Comparing two SD sequence variants, UAAGGAGG and AAGGA, Barrick (ref in 14) has shown, by in vitro translation initiation experiments, that the longer sequence was three times more efficiently used than the smaller sequence. Furthermore, only minor effects on efficiency were seen when the location of the long SD varied within the -12 to -4 region (ref in 14). A full copy of the long-SD (UAAGGAGG) is not present in any of the 31 chloroplast-encoded sequences containing a complete SD sequence (GGAGG). However, the 5' sequences of some chloroplast transcripts do contain single nucleotide variations of this long-SD: e.g. wheat atpH and spinach atpI (UAAGAGG), pea, maize and tobacco psa B/A2; tobacco, spinach, pea and Marchantia psbD and tobacco, petunia, spinach, barley, maize and Marchantia rbcL (UAGGAGG). Whether these longer SD sequences effect the efficiency of translation of these chloroplast transcripts requires experimental testing.

A large number of transcripts from the atpE, psaA, rbcL and psbE genes of a variety of monocotyledonous and dicotyledonous plants have been sequenced. The protein products from each of

Figure 4. Location of (A) all SD sequences and variants, and specific SD sequences (B) GGAGG, (C) GAGG, (D) GGAG, (E) GGA, (F) AGG, (G) GAG within 25 nucleotides 5' to the initiation codon (AUG) of 140 chloroplast-encoded transcripts. Bars indicate the number of nucleotides between A of the SD sequence and A of the initiation codon. (See Appendix A for sequences and references).

these genes represent a large proportion of the total protein content within the chloroplast. Members of each gene class contain a SD sequence variant, the sequence and location of which are highly conserved (Figure 5). Furthermore, the overall sequence identity of the 5' leader region is 92-96% within the monocot and dicot species, respectively, and 84% identity among the examples from these two classes of plants. This situation is in contrast to the low degree of conservation in similar comparisons for the *atpB* transcripts (compare Figure 2A and Figure 5).

An important feature of the *atpB* mRNA transcripts from tobacco, pea, spinach, barley, wheat and rice (which lack a SD sequence within the 25 nucleotides 5' to the initiation codon), is the conservation of an A, U and A/G in the nucleotide positions -3, -1, and +4 positions, respectively (Figure 2A). These nucleotides may have a more important function in the initiation of translation in the absence of a SD sequence. Their importance may be influenced by flanking nucleotides. However, considering the overall A-T richness of the chloroplast genome (60-65% A or T), the occurrence of A and U at these positions may

<i>atpE</i>	-25	-18	-15	-1
tobacco	UAGAAAU <u>CGAG</u> ACCAAUUGAGCAA	AUG		
pea	UAGAAAU <u>CGAG</u> ACCAAUUCAGAA	AUG		
spinach	UAGAAAU <u>CGAG</u> ACCAAUUAAGAA	AUG		
maize	UAGAAga <u>CGAG</u> ACCAAUUGAGCAA	AUG		
rice	UAGAAga <u>CGAG</u> ACCAAUUGAGCAA	AUG		
barley	UAGAAga <u>CGAG</u> ACCAAUcCAGAA	AUG		
wheat	UAGAAga <u>CGAG</u> ACCAAUcCAGAA	AUG		

<i>psaA</i>	-25	-10	-6	-1
tobacco	UGUUGUCCGGAAGAG <u>CGAGG</u> ACUUA	AUG		
pea	UGUUGUCCGGAAGAG <u>CGAGG</u> ACUUA	AUG		
spinach	UCCGGAAGAG <u>CGAGG</u> ACUUA	AUG		
maize	UcUUGUCCGGAAGAG <u>CGAGG</u> ACUUA	AUG		

<i>psbE</i>	-25	-10	-7
tobacco	UGUACUAGAAU <u>UUGGAGG</u> CUCAGC	AUG	
spinach	aAGAAU <u>UUGGAGG</u> CUCAGU	AUG	
O.hookeri	UACaAGAAU <u>UUGGAGG</u> CUCAGC	AUG	
wheat	UGUACAAaAAU <u>UUGGAGG</u> UUCAGC	AUG	

<i>rbcL</i>	-25	-10	-6	-1
tobacco	AUUCAGAGUUGUAG <u>CGAGG</u> CAUUU	AUG		
petunia	AUUCAGAGUUGUAG <u>CGAGG</u> CAUUU	AUG		
pea	AUUCAGUcGUUGUAG <u>CGGcG</u> CAUUU	AUG		
alfalfa	AUUCAGAGUUGUAG <u>CGGcG</u> CAUUU	AUG		
spinach	AaUuugaAGUUGUAG <u>CGAGG</u> CAUUU	AUG		
maize	AUUCAGAGUUGUAG <u>CGAGG</u> CAUUU	AUG		
barley	AUUCaAGAGUUGUAG <u>CGAGG</u> CAUUU	AUG		

Figure 5. A comparative analysis of the translation initiation region (-25 to +3) of several chloroplast-encoded *atpE*, *psaA*, *psbE* and *rbcL* transcripts. Upper case letters indicate sequence identity with that of tobacco. SD sequences are boxed.

not be statistically significant. In contrast to the *atpB* transcripts, *rps12* transcripts lack both a SD sequence (within the -25 to -1 region) and the consensus nucleotides in positions -1 and +4 (Figure 2B).

The presence, absence or length of a SD sequence does not appear to be influenced by a gene's location in a polycistronic or monocistronic operon. For instance, *psaA* and *psaB* are cotranscribed as a dicistronic message containing a GGAGG SD sequence 5' to the coding regions of both genes (38). Shorter SD variants occur in the *rps12* - *rps7* dicistronic message of tobacco; GGA is found at position -41 from the start codon of *rps12* exon 1 and AGG occurs in the mature mRNA 89 bases 5' to the *rps7* coding region (39,40), within exon 2 of the *rps12* sequence. The *atpBE* transcripts contain SD sequences quite distant from the *atpB* coding region in tobacco (-65), pea (-114) and liverwort (-

85), but there is a GGAG sequence 15 bases 5' to the distal atpE coding region (Appendix A). In another case, each of the coding regions comprising the polycistronic atpI - atpH - atpF - atpA transcript are preceded by some form of a SD sequence except for the atpA coding sequence (41). Whether the presence, absence or length of the SD sequence becomes important once a polycistronic transcript has been processed is not known. Also unknown, in cases of adjacent coding sequences such as atpF and atpA, is whether translation of the distal gene requires a new ribosome binding event or whether translation is processive, using the same ribosome required for atpF translation.

4. mRNA secondary structure and translation initiation

Secondary structure within the -12 to -1 leader region which might shield the SD sequence (at its consensus -7 $\pm$ 4 location) and/or the initiation codon, could result in a dramatic reduction in translation of a particular mRNA (42,43). However, there is evidence that secondary structure may also serve to relocate a distal SD sequence (5' to the -12 position) closer to the initiation codon. For instance, the ⁻²⁶GGAG⁻²³ SD sequence of T4 bacteriophage gene 38 mRNA can be relocated, by formation of a stem-loop structure between nucleotides -23 and -5, into a position within 5 nucleotides of the AUG initiation codon (43) - an optimal position to function in prokaryotic translation initiation.

A SD variant located 5' to the optimal -12 to -2 region is present in 18% of the 196 chloroplast-encoded sequences studied. In the case of the pea rps11 gene, which contains a GGA SD sequence 26 nucleotides 5' to the initiation codon (44), an A/U rich stem-loop structure with a G at 25°C of -1.4 kcal/mole (45) can be drawn within the -26 to -1 region of this mRNA. This structure would result in a more favourable location (-6 as opposed to -26) for the SD sequence, with respect to the initiation codon (Figure 6A). A potential stem-loop structure, with a G of -1.2 kcal/mole (45), can also be drawn between nucleotides -41 and +3 of the tobacco 5'rps12 gene (Figure 6B). The presence of such a structure would result in the relocation of the GGA SD sequence from -41 to the more optimal -7 position. Although the free energies of these structures are marginally

5. Termination codon

Table 3. Translation termination codons used by 196 chloroplast-encoded mRNA transcripts.

	UAA	UAG	UGA	Total genes
All plants ^a	70%	19%	11%	57
Tobacco	60%	26%	14%	49
<u>Marchantia</u>	91%	7%	2%	55
Maize	57%	28%	14%	14

a. See Appendix A for sequences and references.

most frequent (70% of 57 chloroplast genes) followed by UAG (19%) and UGA (11%). Both tobacco (a dicotyledon) and maize (a monocotyledon) show a similar termination codon preference. However, in the liverwort, Marchantia, there is a significant preference for UAA (91%) as a stop codon (Table 3). In contrast to both tobacco and maize chloroplast genes, the stop codon UGA is found only in the ndh6 gene of Marchantia (31). Termination of translation of plant nuclear-encoded genes also occurs most frequently at an UAA: 50% of 323 sequences (46). The UAG and UGA codons are used in 21% and 29% of plant nuclear-encoded genes, respectively. In 1243 eukaryotic (animal) sequences, termination of translation occurs at an UAA or UGA codon with equal frequency (40%) and an UAG was used by the remaining 20% of sequences (46). Both plant nuclear-encoded and chloroplast-encoded genes show a prokaryotic-like preference for UAA as a termination codon which differs from that seen in other eukaryotic sequences. In contrast to usage among chloroplast genes, both plant and animal nuclear-encoded sequences frequently use UGA as a termination codon.

A potential mechanism has been proposed for termination of translation in E.coli at an UGA codon. A spontaneous mutation (deletion of a C at position 1054 of the 16S rRNA) was responsible for suppression of translation termination at UGA in the trpA gene mRNA (47). A tandem repeat of the 5'UCA³ triplet (complementary to the 5'UGA³ termination codon) was found (nucleotides 1199 to 1204) opposite the site of the C¹⁰⁵⁴ deletion and within the same stem structure. The authors suggest that deletion of C at position 1054 weakened or disrupted a specific

RNA-RNA interaction between the UCA tandem repeat and the UGA termination codon, resulting in suppression of the UGA codon.

The tobacco (27), soybean (48) and maize (49) chloroplast 16S rRNA sequences possess a C at position 1001 (tobacco) and a 5'UCAUCA³' sequence (positions 1145 to 1150 [tobacco]) located within a potential stem structure similar to that suggested for E.coli (47) (Figure 7A). In contrast, in Marchantia chloroplast 16S rRNA, where the UGA codon is used in only the ndh6 gene (compared to 14% of tobacco and maize chloroplast genes), a C is located at position 1008 but the second U of the UCA tandem repeat, has been replaced by a G - 5'UCAGCA³' (Figure 7B). The sequence equivalent to the conserved tandem repeat is represented in maize (50), wheat (51) and soybean (52) mitochondrial 18S rRNAs by 5'UCCGCA³' (Figure 7C). In animal and yeast mitochondria, UGA codes for tryptophan and is not used as a termination codon (53) however, in both the maize F1 ATPase alpha subunit gene (54) and the apocytochrome b gene from Oenothera (55), an UGA codon is thought to act as a termination codon. Whether these genes are exceptions to the mitochondrial rule will become apparent once more plant mitochondrial genes have been sequenced. We suggest that modifications of the tandem repeat in mitochondria 18S rRNA and in Marchantia 16S chloroplast rRNA, would prevent the specific rRNA-mRNA interaction proposed as a requirement in E.coli for termination of polypeptide synthesis at an UGA codon. Thus translation would be expected to read through the UGA codon in genes from these organelles, with the Marchantia ndh6 mRNA being a unique exception.

A frequency analysis of the 25 nucleotides 5' and 3' to the termination codon shows, as in the case of sequences flanking the AUG initiation codon, a preference for A or U at every position. However, unlike the 25 nucleotide regions 5' and 3' to the initiation codon, the 25 nucleotides upstream and downstream from the termination codon do not show any unusually high occurrences of G or C (data not shown).

CONCLUSIONS

In efforts to identify potential translation regulatory sequences of chloroplast genes, one must consider that regulation

may occur at two distinct levels : (i) via interactions with conserved recognition sequences and (ii) through gene specific sequences or secondary structures. In the case of transcriptional regulation of chloroplast-encoded genes, major changes in mRNA transcript levels can occur as a result of light induction. It is known that specific changes in transcript levels of specific 'light-regulated' genes results from variations among their individual promoter strengths (56,57). In the case of translation, simple frequency analyses of the 5' and 3' flanking regions of the initiation codon reveals that certain nucleotide positions favor a 'prokaryotic-type', 30S ribosome/mRNA association required for *in vivo* initiation of translation of chloroplast mRNAs. Furthermore, our results suggest that sequences both 5' and 3' to the initiation codon could be important in initiation of translation, especially in transcripts with a 'weak' SD sequence or in the absence of a suitably-positioned SD sequence. It is possible, in those transcripts lacking a SD sequence within 100 bp upstream of the initiation codon, that RNA secondary structure within this region could result in the transposition of distal SD sequences to positions proximal to the AUG initiation codons, in such a manner as to facilitate initiation.

RNA secondary structure may also be important in termination of translation from those chloroplast-encoded transcripts containing an UGA termination codon. We suggest that the proposed mechanism for termination at a UGA codon in *E.coli*, utilizing a C and UCAUCA sequence located on the same 16S rRNA stem structure, may also function in the chloroplast.

The existence of homologous prokaryotic sequence motifs within (i) the translation initiation region of chloroplast-encoded mRNAs, (ii) the translation termination region of mRNAs, and (iii) chloroplast-encoded 16S rRNAs confirms the prokaryotic nature of chloroplast mRNA. We conclude from these observations that procaryotic-like translation regulatory mechanisms are likely to occur within chloroplasts.

ACKNOWLEDGEMENTS

This work was supported by NIH grant GM26937 to D.P.B.

REFERENCES

1. Whitfield, P.R. and Bottomley, W. (1983) *Ann. Rev. Plant Physiol.* **34**, 279-310.
 2. Weil, J.H. (1987) *Plant Sci.* **49**, 149-157.
 3. Bonham-Smith, P.C. and Bourque, D.P. (1988) In: Adolph, K. (ed.), *Chromosomes: Eukaryotic, Prokaryotic and Viral*, CRC Press, Inc., in press.
 4. Schneider, T.D., Stormo, G.D. and Gold, L. (1986) *J. Mol. Biol.* **188**, 415-431.
 5. Singer, B.S. and Gold, L. (1976) *J. Mol. Biol.* **103**, 627-646.
 6. Sacerdot, C., Fayat, G., Dessen, P., Springer, M., Plumbridge, J.A. et al (1982) *EMBO J.* **1**, 311-315.
 7. Curran, J. and Kolakofsky, D. (1988) *EMBO J.* **7**, 245-251.
 8. Shine, J and Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 1342-1346.
 9. Kozak, M. (1983) *Micro. Revs.* **47**, 1-45.
 10. Stormo, G., Schneider, T.D. and Gold, L. (1982) *Nucl. Acids Res.* **10**, 2971-2996.
 11. Simons, R.W. and Kleckner, N. (1983) *Cell* **34**, 683-691.
 12. Botterman, J., deAlneida, L., Bouguelert, L. and Zabeau, M. (1985) In: Calandar, R. and Gold, L. (eds), *Sequence Specificity in Transcription and Translation*, Alan R. Liss, Inc., pp. 397-407.
 13. Lodish, H.F. (1970) *J. Mol. Biol.* **50**, 689-702.
 14. Gold, L. (1988) *Ann. Rev. Biochem.* **57**,
 15. Maruyama, T., Gojobori, T., Aota, S. and Ikemura, T. (1986) *Nucl. Acids Res.* **14**, r151-r197.
 16. Zuker, M., and Stiegler, P. (1981) *Nucl. Acids Res.* **9**, 133-148.
 17. Ganzoa, M.C., Kofoed, E.C., Maliere, P. and Louis, B.G. (1987) *Nucl. Acids Res.* **15**, 345-360.
 18. Ganzoa, M.C., Sullivan, P., Cunningham, C., Hader, P., Kofoed, E.C. and Neilson, T. (1982) *J. Biol. Chem.* **257**, 8228-8232.
 19. Schmitt, M., Manderscheid, U., Kyriatsoulis, A, Brinckman, U. and Gassen, H.G. (1980) *Eur. J. Biochem.* **109**, 291-299.
 20. Eckhardt, H. and Luhrman, R. (1981) *Biochemistry* **20**, 2075-
 21. Boudraa, M. and Perrin, P. (1987) *Nucl. Acids Res.* **15**, 5729-5737.
 22. Joshi, C.P. (1987) *Nucl. Acids Res.* **15**, 6643-6653.
 23. Zurawski, G. and Clegg, M.T. (1987) *Ann. Rev. Plant Physiol.* **38**, 391-418.
 24. Geisen, R.M., Fatscher, H-P. and Fuchs, E. (1987) *Nucl.*
 25. Dunn, J.J., Buzasl-Pollert, E. and Studier, F.W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 2741-2745.
 26. Hallick, R.B. and Bottomley, W. (1986) *Plant Mol. Biol. Rep.* **1**, 38-43.
 27. Shinozaki, M., Ohme, M., Tanaka, T., Wakasugi, N., Hayshida, T., Matsubayasha, N., Zaita, J., Chunwongse, J., Obokata, K., Yamaguchi-Shinozaki, C., Ohto, K., Torazawa, B.Y., Meng, M. Sugita, H., Deno, T., Kamogashira, K., Yamada, J., Kusuda, J., Takaiwa, A., Kata, N., Tohdohj, H., Shimada and Sugiura, M. (1986) *Plant Mol. Biol. Rep.* **4**, 110-147.
 28. Zurawski, G., Bottomley, W. and Whitfield, P.R. (1984) *Nucl. Acids Res.* **12**, 6547-6558.
 29. Spielman, A., Roux, E., von Allmen, J.M. and Stutz, E. (1988) *Nucl. Acids Res.* **16**, 1199.
-

30. McLaughlin, W.E. and Larrinua, I.G. (1987) Nucl. Acids Res. **15**, 3932.
31. Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H. (1986) Plant Mol. Biol. Rep. **4**, 148-175.
32. Hann, S.R., King, M.W., Bentley, D.L., Anderson, C.W. and Eisenman, R.N. (1988) Cell **52**, 185-195.
33. Gold, L. Pribnow, D., Schneider, T., Shinedling, S., Singer, B.S. and Stormo, G. (1980) Ann. Rev. Microbiol. **35**, 365-403.
34. Steitz, J.A. and Jakes, K. (1975) Proc. Natl. Acad. Sci. USA **72**, 4734-4738.
35. Smiley, B., Lupaki, J., Svec, P., McMachen, R. and Godson, G.N. (1982) Proc. Natl. Acad. Sci. USA **79**, 4550-4554.
36. Ptashne, M., Backman, K., Humayum, M.Z., Jeffrey, A., Maurer, B. and Sauer, R.T. (1976) Science **194**, 156-161.
37. Pirrotta, V. (1979) Nucl. Acids Res. **6**, 1495-1508.
38. Westhoff, P., Alt, J., Nelson, N., Bottomley, W., Buneman, H. and Herrman, R.G. (1983) Plant. Mol. Biol. **2**, 95-107.
39. Hildebrand, M.M., Hallick, R.B., Passavant, C. and Bourque, D.P. (1988) Proc. Natl. Acad. Sci. USA **85**, 372-376.
40. Koller, B., Fromm, H., Galun, E. and Edelman, M. (1987) Cell **48**, 111-119.
41. Hudson, G.S. et al (1987) J. Mol. Biol. **196**, 283-298.
42. McPheeters, D.S., Christensen, A., Young, E.T., Stormo, G. and Gold, L. (1986) Nucl. Acids Res. **14**, 5813-5826.
43. Simons, R.W. and Kleckner, N. (1983) Cell **34**, 683-691.
44. Purton, S. and Gray, J.C. (1987) Nucl. Acids Res. **15**, 1873.
45. Saenger, W. (1984) Principles of Nucleic Acid Structure, pp. 116-158. Springer-Verlag new York Inc.
46. Aota, S., Gojobori, T., Maruyama, T. and Ikemura, T. (1988) Nucl. Acids Res. **16**, r315-r402.
47. Murgola, E.J., Hijazi, K.A., Goringer, H.U. and Dahlberg, A.E. (1988) Proc. Natl. Acad. Sci. USA **85**, 4162-4165.
48. von Allmen, J-M. and Stutz, E. (1988) Nucl. Acids Res. **16**, 1200.
49. Schwarz, Z. and Koessel, H. (1980) Nature **283**, 739-742.
50. Chao, S., Sederoff, R. and Levings III, C.S. (1984) Nucl. Acids Res. **16**, 6629-6644.
51. Spencer, D.F., Schnare, M.N. and Gray, M.W. (1984) Proc. Natl. Acad. Sci. USA **81**, 493-497.
52. Grabau, E.A. (1985) Plant Mol. Biol. **5**, 119-124.
53. Bonitz, S.G., Berlani, R., Corruzzi, G., Li, M., Marcino, G. Nobrego, F.G. and Nobrego, M.P. (1980) Proc. Natl. Acad. Sci. USA **77**, 3167-3170.
54. Braun, C.J. and Levings, C.S. III. (1985) Plant Physiol. **79**, 571-577.
55. Schuster, W. and Brennicke, A. (1985) Curr. Genet. **9**, 157-163.
56. Link, G. (1984) EMBO J. **3**, 1697-1704.
57. Gruissem, W. and Zurawski, G. (1985) EMBO J. **4**, 3375-3383.

APPENDIX A
DISTANCE OF THE SHINE-DALGARNO SEQUENCE FROM CHLOROPLAST PROTEIN
CODING GENE ATG START CODONS

N ^a	SD sequence	gene	plant ^b species	Reference
2	GGA	petA	T, BB, W, S	1, 3, 4, 62
	GGAG	petB	T	1
	GGA	petD	T, L, P	1, 2, 52
	GGA	rps3	T	1
3	GGA	psbF	T	1
	GGA	rbcL	BB	6
	AGG	petB	S, P	63, 52
	GGAG	rps2	S	68
4	GGA	rbcL	A	9
	GGA	rps2	W	10
5	GGAG	rps2	T	1
	GGAGG	rpl33	T	1
	GGAG	atpI	L	2
	GGAG	rbcL	P	11
	GAG	secX	T [#]	1
6	GGAG	rps4	T, S, L, M	1, 12, 2, 12
	GAG	rpoB	S	67
	GGAG	rps8	L	2
7	GGAG	psbE	T, Pr, S, W	1, 13, 64, 14
	GGAGG	psbF	L, Pr	2, 13
	GGAGG	psaA	T, P, M, S	1, 16, 17, 8
	GAGG	atpF	L, W	2, 5
	GGAG	atpH	L	2
	GGAG	atpI	S	41
	GAGG	rpoB	L	2
	GGAGG	rbcL	T, B, M, PT, S, L	1, 18, 19, 20, 15, 2
8	GGAG	psaA	L	2
	GGAG	psbG	L, M	2, 21
	GGAGG	psbD	T, S, P, B	1, 22, 23, 66
	GGAG	psbD	L	2
	GGA	rpoB	T	1
	GGAGG	ndhC	T	1
	GGAGG	ndhE	T	1
	GGA	ndh2	L	2
	GGAGG	ndh3	L	2
	GGA	ndh4L	L	2
	GAGG	rps16	T	1
	GGA	rpl14	T	1
	GGAG	rps19	T	1
9	GAGG	atpF	T, S	1, 41
	GGA	petB	L	2
	GGAG	rps15	L	2
	GGA	rps19	L	2
	GGAG	mbpX	L	2
	GGAG	frxA	L	2
	GGA	frxB	L	2
	GGAG	secX	P (L36) [#]	24
10	GAGG	atpH	T, SB	1, 25
	GGAG	atpC	P	26
	GGA	rps14	M	27
	GGAGG	rps19	M, S	28, 29
	GGAG	rps19	SB	30
	GGAGG	rpl22	T	1
	GGA	secX	M (L36) [#]	31
11	AGG	psbB	T, S, L	1, 7, 2
	GGA	rpoC	L	2
	GAGG	rpoC2	L	2
	GGA	rps14	T	1
	GAG	rps11	S, T	33, 1
	GGAG	rpl22	L	2
	GAGG	atpH	S	41
	GGA	atpB	BB	6

N ^a	SD sequence	gene	plant ^b species	Reference
12	GAG	*secX	L [#]	2
	GGAG	*atpE	L	2
	GAG	rpl20	L	2
	GAG	*ndhD	T	1
	GGA	*rpl14	M	61
13	GAGG	atpH	W	32
	GAGG	*atpI	T	1
	GGA	*rpl23	T,S	1,53
	GGAG	infA	L	2
14	GGAG	*rpl23	L	2
	GGAGG	ndh1	L	2
	GAG	atpA	W	58
	AGG	psbB	M	59
15	GGAG	atpE	T,B,P,R,S,M,W	1,34,35,36,57,37,58
	GGA	petA	P	39
	GGAG	infA	T	1
16	GGAGG	psaB	T,L,P,M,S	1,2,16,17,8
	GGAG	ndhA	T	1
	GAGG	atpF	T	1
	GGA	*petB	L	2
	GGA	*rpl14	L	2
17	GGA	*rps7	SB	40
18	AGG	*rps18	L	2
19	GAG	ndhB	T	1
20	GGA	*rps18	T	1
	GGA	*rpoC	S	67
	AGG	*rpl22	M	43
21	GGAGG	*rps8	M	61
	GGA	psbC	T,S,P,L	1,22,23,2
22	GGA	*rps7	M	42
	GAG	rpoA	P	44
23	GGAGG	*rps8	T	1
26	GGA	*rps11	P	44
28	GGA	psbE	L	2
30	AGG	ndh4	L	2
32	GGAG	*psbA	A	45
	AGG	*rpl33	L	2
33	GGA	psbA	T,L	1,2
	GAG	frxB	W,P	46,47
	GGAG	psbA	PT,Mt,R,B	48,56,60,65
34	GGAG	psbA	W	54
35	GAG	atpA	L	2
37	GGAG	psbA	S	50
38	AGG	*rps12	L	2
39	GAG	*rpoA	T	1
	GGA	*rpl20	T	1
40	GGA	*ndhF	T	1
	AGG	*rpl16	T	1
41	AGG	*atpA	T	1
	GGA	*rps12	T	1
45	GAG	rpoA	S	33
47	AGG	atpA	S	41

N ^a	SD sequence	gene	plant ^b species	Reference
50	AGG	infA	S	33
51	GGA	atpB	S	57
55	GAG	petA	L	2
56	AGG	*rps11	L	2
	GAG	*rpl2	S	29
59	GGA	psbA	Dn	49
62	GGA	rpoA	L	2
63	GGA	atpB	T	1
67	GGAG	*rps7	L	2
77	GGA	*rpl16	L	2
82	GGA	atpB	M	37
84	GGA	atpB	W	58
85	AGG	atpB	L	2
89	GAG	*rps7	T	1
90	GGA	atpB	R	36
99	GGA	*rpl2	T	1
no SD				
sequence				
within				
100 bp of				
the AUG				
codon				
		atpB	B*,P	34*,35
		psbG	T	1
		ndh5	L	2
		rps12	SB,M*	40*,42*
		rps3	M,L	47*,2
		*rpl16	M	51
		*rps2	L	2
		*rps14	L	2
		*rpl21	L	2
		*rpl2	L	2
		frxC	L	2
		psaC	T	55

a. N = the number of nucleotides between the A of the initiation codon AUG and the central A of the orthodox (GGAGG) Shine-Dalgarno sequence. Asterisk (*) indicates chloroplast ribosomal protein genes and (+) indicates that insufficient sequence data was available. The *secX*(#) gene is thought to encode protein L36 of the chloroplast ribosome large subunit.

b. plant species - A: alfalfa, B: barley, BB: broadbean, Dr: deadly nightshade, L: liverwort, M: maize, Mt: mustard, P: pea, Pr: primrose, PT: petunia, R: rice, S: spinach, SB: soybean, T: tobacco, W: wheat.

APPENDIX REFERENCES

1. Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayshida, N., Matsubayasha, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamagashire, T., Yamada, K., Kusuda, J., Takaiwa, F., Kata, A., Tohdoh, N., Shimada, H. and Sugiura, M. (1986) *Plant Mol. Biol. Rep.* **4**, 111-147.
2. Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H. (1986) *Nature* **322**, 572-574.
3. Ko, K. and Straus, N.A. (1987) *Nucl. Acids Res.* **15**, 2391.
4. Willey, D.L., Howe, C.J., Auffret, A.D., Bowman, C.M., Dyer,

- T.A. and Gray, J.C. (1984) *Mol. Gen. Genet.* **194**, 416-422.
5. Bird, C.R., Koller, B., Auffret, A.D., Huttley, A.K., Howe, C.J., Dyer, T.A. and Gray, J.C. (1985) *EMBO J.* **4**, 1381-1388.
6. Shinozaki, K., Sun, C.R. and Sugiura, M. (1984) *Mol. Gen. Genet.* **197**, 363-367.
7. Morris, J. and Herrman, R.G. *Nucl. Acids Res.* **12**, 2837-2850.
8. Herrmann, R.G., Westhoff, P., Alt, J., Tittgen, J. and Nelson, N. (1985) In: Vloten-Doting, L., Groot, G.S.P. and Hall, T.C. (eds), *Molecular Form and Function of the Plant Genome*, Plenum, New York, pp. 233-256.
9. Aldrich, J.; Cherney, B., Merlin, E. and Palmer, J. (1986) *Nucl. Acids Res.* **14**, 9535.
10. Hoglund, A.S. and Gray, J.C. (1987) *Nucl. Acids Res.* **15**, 24.
11. Zurawski, G., Whitfeld, P.R. and Bottomley, W. (1986) *Nucl. Acids Res.* **14**, 3975.
12. Ben Tahar, S.B., Bottomley, W. and Whitfeld, P.R. (1986) *Plant Mol. Biol.* **7**, 63-70.
13. Carrillo, N., Seyer, P., Tyagi, A. and Herrmann, R.G. (1986) *Curr. Genet.* **10**, 619-624.
14. Hird, S.M., Wiley, D.L., Dyer, T.A. and Gray, J.C. (1986) *Mol. Gen. Genet.* **203**, 95-100.
15. Zurawski, G., Perrot, B., Bottomley, W. and Whitfeld, P.R. *Nucl. Acids Res.* **9**, 3251-3270.
16. Lehmbeck, J., Rasmussen, O.F., Bookjans, G.B., Jepsen, B.R., Stummann, B.M. and Henningsen, K.W. (1986) *Plant Mol. Biol.* **7**, 3-10.
17. Fish, L.E., Kueck, U. and Bogorad, L. (1985) *J. Biol. Chem.* **260**, 1413-1421.
18. Zurawski, G., Clegg, M. and Brown, A.H.D. (1984) *Genetics* **106**, 735-749.
19. McIntosh, L., Poulsen, C. and Bogorad, L. (1980) *Nature* **288**, 556-560.
20. Aldrich, J., Cherney, B., Merlin, E. and Palmer, J. (1986) *Nucl. Acids Res.* **14**, 9534.
21. Steinmetz, A.A., Castroviejo, M., Sayre, R.T. and Bogorad, L. (1986) *J. Biol. Chem.* **261**, 2485-2488.
22. Holschuh, K., Bottomley, W. and Whitfeld, P.R. (1984) *Nucl. Acids Res.* **12**, 8819-8834.
23. Rasmussen, O.F., Bookjans, G., Stutmann, B.M. and Henningsen, K.W. (1984) *Plant Mol. Biol.* **3**, 191-199.
24. Purton, S. and Gray, J.C. (1987) *Nucl. Acids Res.* **15**, 9080.
25. Wintz, H., Skunca, M. and Pillay, D.T.N. (1987) *Gene* **59**, 47-53.
26. Cozens, A.L., Walker, J.E., Phillips, A.L., Huttly, A.K. and Gray, J.C. (1986) *EMBO J.* **5**, 217-222.
27. Srinivasa, B.R. and Subramanian, A.R. (1987) *Biochemistry* **20**, 3188-3192.
28. McLaughlin, W.E. and Larrinua, I.G. (1987) *Nucl. Acids Res.* **15**, 3932.
29. Zurawski, G., Bottomley, W. and Whitfeld, P.R. (1984) *Nucl. Acids Res.* **12**, 6547-6558.
30. Spielman, A., Roux, E., von Allmen, J.M. and Stutz, E. (1988) *Nucl. Acids Res.* **16**, 1199.
31. Markmann-Mulisch, U., vonKnoblauch, K., Lehmann, A. and Subramanian, A.R. (1987) *Biochem. Int.* **15**, 1057-1067.
32. Howe, C.J., Auffret, A.D., Doherty, A., Bowman, C.M., Dyer,

- T.A. and Gray, J.C. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6903-6907.
33. Sijben-Muller, G., Hallick, R.B., Alt, J., Westhoff, P. and Herrmann, R.G. (1986) *Nucl. Acids Res.* **14**, 1029-1044.
 34. Zurawski, G. and Clegg, M.T. (1984) *Nucl. Acids Res.* **12**, 2549-2559.
 35. Zurawski, G., Bottomley, W. and Whitfeld, P.R. (1986) *Nucl. Acids Res.* **14**, 3974.
 36. Moon, E., Kao, T. and Wu, R. (1987) *Nucl. Acids Res.* **15**, 4358-4359.
 37. Krebbers, E.T., Larrinua, I.M., McIntosh, L. and Bogorad, L. (1982) *Nucl. Acids Res.* **10**, 4985-5002.
 38. Herschberg, J. and McIntosh, L. (1983) *Science* **222**, 1346-1349.
 39. Willey, D.L., Auffret, A.D. and Gray, J.C. (1984) *Cell* **36**, 555-562.
 40. von Allmen, J.M. and Stutz, E. (1987) *Nucl. Acids Res.* **15**, 2387.
 41. Hennig, J. and Herrman, R.G. (1986) *Mol. Gen. Genet.* **203**, 117-128.
 42. Giese, K., Subramanian, A.P., Larrinua, I.M. and Bogorad, L. (1987) *J. Biol. Chem.* **262**, 15251-15255.
 43. McLaughlin, W.E. and Larrinua, I.M. (1987) *Nucl. Acids Res.* **15**, 4356.
 44. Purton, S. and Gray, J.C. (1987) *Nucl. Acids Res.* **15**, 1873.
 45. Aldrich, J., Cherney, B. and Merlin, E. (1986) *Nucl. Acids Res.* **14**, 9537.
 46. Dunn, P.P.J. and Gray, J.C. (1988) *Nucl. Acids Res.* **16**, 348.
 47. McLaughlin, W.E. and Larrinua, I.M. (1987) *Nucl. Acids Res.* **15**, 4689.
 48. Aldrich, J., Cherney, B., Merlin, E., Christopherson, L.A. and Williams, C. (1986) *Nucl. Acids Res.* **14**, 9536.
 49. Goloubinoff, P., Edelman, M. and Hallick, R.B. (1985) *Nucl. Acids Res.* **12**, 9489-9496.
 50. Zurawski, G., Bohnert, H.J., Whitfeld, P.R. and Bottomley, W. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7699-7703.
 51. McLaughlin, W.E. and Larrinua, I.M. (1987) *Nucl. Acids Res.* **15**, 5896.
 52. Phillips, A.L. and Gray, J.C. (1984) *Mol. Gen. Genet.* **194**, 477-484.
 53. Thomas, F., Massenet, O., Dorne, A-M, Briat, J-F. and Mache, R. (1988) *Nucl. Acids Res.* **16**, 2461-2472.
 54. Hanley-Bowdoin, L. and Chua, N-H. (1988) *Plant Mol. Biol.* **10**, 303-310.
 55. Hayashida, N., Matsubayashi, T., Shinozaki, K., Sugiura, M., Inoue, K. and Hiyama, T. (1987) *Curr. Genet.* **12**, 247-250.
 56. Link, G. and Langridge, U. (1984) *Nucl. Acids Res.* **12**, 945-958.
 57. Zurawski, G., Bottomley, W. and Whitfeld, P.R. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6260-6264.
 58. Howe, C.J., Fearnley, I.M., Walker, J.E., Dyer, T.A. and Gray, J.C. (1985) *Plant Mol. Biol.* **4**, 333-345.
 59. Westhoff, P. and Herrman, R.G. (1988) *Eur. J. Biochem.* **171**, 551-564.
 60. Wu, N-H., Cote, J-C. and Wu, R. (1987) *Dev. Genet.* **8**, 339-

61. Markmann-Mulisch, U. and Subramanian, A.R. (1988) *Eur. J. Biochem.* **170**, 507-514.
62. Alt, J. and Herrman, R.G. (1984) *Curr. Genet.* **8**, 551-557.
63. Heinemeyer, W., Alt, J. and Herrmen, R.G. (1984) *Curr. Genet.* **8**, 543-549.
64. Herrman, R.G., Alt, J., Schiller, B., Widger, W.R. and Cramer, W.A. (1984) *FEBS. Lett.* **176**, 239-244.
65. Efimov, V.A., Andreeva, A.V., Reverdatto, S.V., Jung, R. and Chakhmakhcheva, O.G. (1988) *Nucl. Acids Res.* **16**, 5685.
66. Efimov, V.A., Andreeva, A.V., Reverdatto, S.V. and Chakhmakhcheva, O.G. (1988) *Nucl. Acids Res.* **16**, 5686.
67. Hudson, G.S., Holton, T.A., Whitfeld, P.R. and Bottomley, W. (1988) *J. Mol. Biol.* **200**, 639-654.