Additional Characterization and Safety Assessment of the DNA Sequence Flanking the 3' End of the Functional Insert of Roundup Ready® Soybean Event 40-3-2

February 1, 2002

Monsanto Company
700 Chesterfield Parkway North
St. Louis, MO, USA

Table of Contents

Abbreviations

bp	Base pairs
CP4 EPSPS	5-enol-pyruvylshikimate-3-phosphate synthase isolated from *Agrobacterium sp.* Strain CP4
Da	Dalton, a unit of mass
DNA	Deoxyribonucleic acid
E35S	Cauliflower mosaic virus (CaMV) 35S promoter containing a duplication of the -90 to -300 bp region
ELISA	Enzyme linked immunosorbant assay
FAO	Food and Agriculture Organization of the United Nations
FDA	United States Food and Drug Administration
kb	Kilo-base
kDa	kilo-Daltons
LOD	Limit of detection
mRNA	Messenger ribonucleic acid
NOS	Nopaline synthase
ORF	Open reading frame
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
RR	Roundup Ready
RT-PCR	Reverse transcriptase - Polymerase chain reaction
tRNA	Transfer RNA
WHO	World Health Organization of the United Nations

1.0 Summary

The purpose of this document is to summarize the results of additional molecular characterization studies of the soybean genomic DNA sequence flanking the 3' end of the functional insert in Roundup Ready® soybean event 40-3-2 (RR soybean) and to assess the potential impact of these results on the food, feed and environmental safety of this product.

Roundup Ready® soybean event 40-3-2 was evaluated extensively in numerous food, feed and environmental studies prior to the submission of regulatory dossiers. This safety information, including a molecular characterization, was reviewed and this product has been approved by many regulatory agencies around the world.

Subsequent molecular analyses of the DNA inserted into Roundup Ready soybean, using more sensitive and precise methods, identified a 250 bp portion of the CP4 EPSPS sequence adjacent to the 3' end of the primary, functional insert and a second, small, non-functional insert consisting of a 72 bp portion of the *cp4 epsps* DNA which co-segregates with the functional insert. The DNA sequence flanking each side of the functional insert was also generated to facilitate the development of event-specific detection methods. These studies confirmed that Roundup Ready soybean event 40-3-2 contains one functional insert, which produces the full-length EPSPS protein that provides for tolerance to Roundup® herbicide. These characterization data, an assessment of the possible functionality and safety of the newly described segments or any putative products derived from these segments, and evidence that the full-length CP4 EPSPS protein was the only detectable protein product resulting from the inserted DNA were provided to regulatory agencies in May, 2000.

A report subsequently published by Windels *et al*. (2001) confirmed the soybean genomic DNA sequences flanking the functional insert and suggested that a segment of DNA at the 3' end of the insert may have been rearranged during the insertion process. A rearrangement at the insertion site is consistent with the results of PCR analysis previously reported by Monsanto. Windels *et al.* (2001) did not identify any homology with soybean DNA sequences when comparing this 3' flanking genomic DNA sequence with the public DNA database. However, the segment of DNA flanking the 3' end of the functional insert has been shown by Monsanto to be soybean genomic DNA by establishing that these sequences are homologous to soybean DNA sequences in it's proprietary soybean genomic sequence collection and can be amplified by PCR in the non-transgenic parental line.

Northern blot analyses were recently conducted which indicate the presence of several secondary RNA transcripts that contain the 3'end of the functional insert. These secondary transcripts were at least 75 times lower in abundance than that of the ~1.4 kb CP4 EPSPS transcript produced by the functional insert. Additional northern blotting and

® Roundup and Roundup Ready are registered trademarks of Monsanto Technology, LLC, St. Louis, MO

RT-PCR analyses indicate that these transcripts are produced in the same orientation as the primary CP4 EPSPS transcript and that the secondary transcripts likely initiate at the E35S promoter, just as the ~1.4 kb CP4 EPSPS transcript, but continue transcription through the NOS 3' polyadenylation signal at the end of the *cp4 epsps* gene into the soybean genomic sequence flanking the 3' end of the insert. The incomplete termination or use of alternative termination sites and resulting production of multiple transcripts has been reported for endogenous genes in plants, including soybean (Rothnie, 1996; Hunt, 1994; Gallie, 1993).

Since RNA, including mRNA, is generally recognized as safe (GRAS), the presence of these secondary RNA transcripts themselves raises no safety concern. The potential of the secondary transcripts to produce a protein product other than, or in addition to, the full-length CP4 EPSPS protein has been assessed. It was determined that the ability of the detected secondary transcripts to produce a translated protein other than, or in addition to, the full-length CP4 EPSPS protein is extremely remote. Since translation of mRNA transcripts in plants, with very rare exceptions, only produces a single polypeptide (Fütterer and Hohn, 1996), the only protein product expected to be produced from the 40-3-2 insert is the full-length CP4 EPSPS. The only CP4 EPSPS-containing protein detected by western blots (with a limit of sensitivity of 100 pg) in Roundup Ready soybean is the well-characterized 46 kDa full-length CP4 EPSPS protein. This is expected since RNA transcripts produced in plants that contain consecutive coding regions, although rare in occurrence, generally result in the production of only one protein. If sequences downstream from the stop codon of the CP4 EPSPS transcript were translated, the most likely translation product would be a larger hybrid/fusion protein that contains the full-length CP4 EPSPS protein. No immunoreactive proteins of molecular weight greater than 46 kDa were detected in extracts of Roundup Ready soybeans. In addition, bioinformatics analyses on the open reading frames (ORFs) present in the region beyond the 3' end of the *cp4 epsps* coding region, including the region of soybean genomic DNA flanking the 3' end of the functional insert, do not show any homology to known toxins or allergens.

This additional characterization of the genomic DNA sequence flanking the 3' end of the primary insert of event 40-3-2 does not alter the previous conclusion that this event contains a single functional gene cassette encoding the 46 kDa CP4 EPSPS protein. The DNA inserted into event 40-3-2, the DNA sequences flanking these inserts, and any transcripts containing the 3' flanking region were constituents in the test material used to establish the food, feed and environmental safety of Roundup Ready soybeans. Roundup Ready soybeans were previously established to be as safe as conventional soybeans based on: the safety of the genetic elements contained on the transformation vector used to produce Roundup Ready soybean event 40-3-2; the history of safe use of the EPSPS family of proteins present in all plants, fungi and bacteria; the functionality and safety assessment of the CP4 EPSPS protein; the assessment of compositional and nutritional equivalence of event 40-3-2; a comparison of crop agronomic characteristics of event 40-3-2 to the parental and conventional soybeans; and a comparison of the safety and

nutritional properties of event 40-3-2 to parental and conventional soybean varieties in animal feeding studies.

A review of this safety assessment confirms the conclusions reached previously that: 1) Roundup Ready soybean event 40-3-2 contains one functional *cp4 epsps* gene cassette which is responsible for the production of the CP4 EPSPS protein; 2) no immunoreactive proteins other than the expected 46 kDa CP4 EPSPS protein are detected in Roundup Ready soybean; 3) Roundup Ready soybeans are as safe and as nutritious as conventional soybean varieties and; 4) Roundup Ready soybeans do not pose a plant pest risk or otherwise pose a risk to the environment. Any risks associated with the production and consumption of Roundup Ready soybean event 40-3-2 to human health or the environment are no different than those associated with conventional soybean varieties.

2.0 Introduction

Roundup Ready (RR) soybean event 40-3-2 was produced by transformation of non-transgenic soybean line A5403. Roundup Ready soybean event 40-3-2 was produced using the particle-acceleration method of transformation utilizing DNA derived from plasmid PV-GMGT04. The primary insertion of DNA includes a single *cp4 epsps* gene cassette: the E35S promoter, the chloroplast transit peptide and *cp4 epsps* coding sequences, and the NOS 3' transcriptional terminator. The *cp4 epsps* coding sequence encodes 456 amino acids, is terminated by tandem stop codons, and results in the synthesis of a full-length and functional 46 kDa CP4 EPSPS (5-enol-pyruvylshikimate-3-phosphate synthase) protein in Roundup Ready soybean event 40-3-2 as confirmed by western blotting, ELISA and EPSPS activity assays (Harrison *et al*., 1996; Rogan *et al*., 1999).

In addition to the functional *cp4 epsps* gene cassette, Roundup Ready soybean event 40-3-2 contains two additional small segments of inserted DNA: a 250 base pair segment of *cp4 epsps* DNA located adjacent to the NOS 3’ transcriptional termination element of the primary, functional insert and a 72 base pair segment of *cp4 epsps* DNA that co-segregates with the primary functional insert. Although two *cp4 epsps* segments are present, they are not functional genes as demonstrated by western blot analysis, in which only the predicted, full-length CP4 EPSPS protein encoded by the functional *cp4 epsps* gene casette was detected (Rogan *et al*., 1999). This additional molecular characterization data that had been developed using more sensitive and precise methodology (Lirette *et al*., 2000) to facilitate the development of event specific detection methods was provided to regulatory agencies in May, 2000.

PCR and nucleotide sequence analysis carried out in the regions flanking the primary and secondary inserts, indicated that intact genomic DNA sequences are present around both ends of the 72 bp insert, as well as at the 5’ end of the functional insert. However, PCR analysis with a primer pair designed to span the functional insert did not generate a PCR

product when DNA from the parental line was used as a template (Lirette *et al.*, 2000). This result suggests that a deletion or rearrangement occurred at the site where the functional insert was integrated into the soybean genome. It was also determined that the region flanking the 3' end of the insert is not derived from, nor is homologous to, the plasmid vector used to create event 40-3-2.

A report subsequently published by Windels *et al*. (2001) confirmed the Monsanto data describing the soybean genomic DNA sequences flanking the functional insert of Roundup Ready soybean event 40-3-2 and concluded that the region flanking the 3' end of the functional insert was likely to be rearranged. PCR and nucleotide sequence analyses by Windels *et al*. (2001) indicated intact plant genomic DNA sequence at the 5' end of the functional insert, while at the 3' end of the functional insert, they identified a region of 534 bp that was not amplified by PCR using DNA from the parental line. This region, for which no homology to publicly available DNA sequences could be found, is followed by intact plant DNA that could be amplified in the parental genome. The lack of sequence homology between the 534 bp region and genomic soybean plant DNA was interpreted to indicate that the genomic plant DNA at the pre-integration site may have been rearranged.

Monsanto provided an assessment of the significance of a rearrangement in the 3' genomic flanking region for human and animal health and the environment to regulatory authorities in 2001. It was concluded that this information was not significant from a safety standpoint since the DNA sequence 3' to the functional insert in Roundup Ready soybeans: (1) is not new DNA; (2) the DNA is not from the transformation vector; (3) the DNA is rearranged soybean genomic DNA and does not represent a new insert; (4) the data is consistent with molecular characterization information provided by Monsanto to regulatory authorities in 2000; and (5) the soybean DNA which flanks the 3' end of the functional insert was present in all the safety studies conducted by Monsanto.

This document summarizes the additional characterization and transcriptional activity of the soybean genomic DNA sequence flanking the 3' end of the functional insert in Roundup Ready soybean event 40-3-2. This information was evaluated within the context of what is known regarding the transcription and translation of plant genes and the potential for the production of protein products other than the full-length functional CP4 EPSPS protein in Roundup Ready soybeans. These data have been assessed within the context of the established food, feed and environmental safety of this product, and lead to the conclusion that the additional molecular characterization data does not alter the previous conclusions that Roundup Ready soybeans are agronomically, compositionally and nutritionally comparable to conventional soybeans, except for the Roundup Ready trait.

3.0 Characterization of DNA Sequence Flanking the 3' End of the Functional Insert in Roundup Ready Soybean Event 40-3-2

Windels *et al*. (2001) reported that the 534 bp of DNA immediately flanking the 3' end of the functional insert in Roundup Ready soybean event 40-3-2 could not be amplified by PCR using DNA from the parental line (A5304) and a primer pair from within the region. When the sequence of this 534 bp region was compared to the GenBank public DNA sequence database no meaningful homologies to any known sequences were observed. These findings led Windels *et al*. to conclude that during integration of the insert DNA into the soybean genome, DNA rearrangements or a deletion may have occurred.

The phenomenon of modification, deletions or rearrangements at the point of insertion is widely recognized within the plant transformation field (Christou *et al*., 1993; Gheysen *et al*., 1987; Christou, 1996; Kholi *et al.*, 1998, 1999; Pawlowski and Somers, 1998). The impacts of any such modification were assessed from both the food, feed and environmental safety perspective, during the conduct of numerous agronomic and safety studies where the potential for pleiotropic effects in Roundup Ready soybeans was evaluated. This would include detailed compositional analyses (Padgette *et al*., 1996) and animal feeding studies (Hammond *et al*., 1996) that were conducted on Roundup Ready soybean seed in which the inserted DNA and any rearrangements were present as constituents in the materials.

To confirm that the 534 base pairs of DNA sequence flanking the 3' end of the functional insert are soybean genomic DNA, Monsanto compared this DNA sequence against it's proprietary soybean genomic DNA sequence collection using BLASTN analysis (Altschul *et al*., 1990). This analysis (Goley *et al*., 2002) identified a clone from a BAC (*B*acterial *A*rtificial *C*hromosome) library that contains DNA sequences identical to 455 of the 534 base pairs in this segment. The BAC clone containing this homologous sequence was derived from genomic DNA of the conventional soybean variety A3244 (Tomkins *et al*., 2000), and thus represents a small portion of the genome of that variety. Two distinct and non-contiguous DNA regions from the BAC clone share 100% sequence homology to most of the 534 base pair 3' flanking sequence. This finding confirms the hypothesis that the DNA flanking the 3' end of the functional insert in Roundup Ready event 40-3-2 is indeed soybean genomic DNA that is rearranged relative to these sequences in conventional soybeans.

To confirm that the DNA sequence flanking the 3' end of the functional insert of Roundup Ready soybean event 40-3-2 consists of soybean genomic DNA, PCR analysis was performed on conventional soybean using oligonucleotide primers based on the 3' flanking DNA sequence (Goley *et al*., 2002). Three different DNA sources were used as templates, including genomic DNA from Roundup Ready soybean event 40-3-2, genomic DNA prepared from the parental line A5403, and purified BAC DNA, which was derived from the conventional variety A3244. In all three DNA samples, the amplification resulted in a PCR product of the predicted size, indicating that this DNA segment from

the 3' flanking region was present in the genomes of conventional soybean lines. These three amplified PCR products were subsequently sequenced and found to be 100% identical to one another. These data support the conclusion that the DNA flanking the 3' end of the functional insert in Roundup Ready soybean event 40-3-2, including the 534 bp region described by Windels *et al*. (2001), is rearranged soybean genomic DNA.

4.0 Summary of Gene Transcription in Higher Plants

Gene transcription in eukaryotic organisms is a dynamic multistage process that leads to the accumulation of mRNA molecules. A summary of transcription termination in higher plants is provided here; however, a more detailed description of the process and the significant differences that exist between vertebrate organisms and plants is provided in Appendix A.

Overall, transcription can be divided into three stages: the synthesis of the primary RNA transcript (also referred to as heterogeneous nuclear RNA), post-transcriptional RNA processing to yield the mature mRNA that can be subsequently translated into protein, and mRNA degradation. The interplay and interaction between these processes is responsible for producing the measurable, steady state levels of a particular mature mRNA transcript in a given tissue. As a biochemical process, it should be noted that the observed steady-state effect of these steps represents the most favored, probabilistically and thermodynamically, condition where lower frequency processes occur regularly. As a consequence, all genes produce a population of transcripts, some of which are more probable and abundant than others, and the dominant gene expression products are most relevant when considering biochemical and physiological consequences.

Although many aspects of transcription are shared by all eukaryotes, subtle mechanistic differences have been identified that distinguish vertebrate and plant transcription (reviewed in Rothnie, 1996). Through sequence analysis of mRNAs from vertebrates and plants, one such mechanistic difference that has been identified in plants is the process of 3' cleavage of primary RNA transcripts. In contrast to genes from vertebrates, plants do not recognize or utilize a single highly conserved polyadenylation signal sequence to direct transcript termination and cleavage (Dean *et al*., 1986; Hunt, 1994; Rothnie, 1996). As a result of this difference, the transcription of some plant genes leads to the accumulation of a fairly heterogeneous population of mature mRNAs that have a common coding sequence, but which differ in the sequence and length of their 3' untranslated regions exclusive of the poly(A) tail. These differences in the 3' untranslated region distinguish plant mRNAs from vertebrate mRNAs, such that a given plant gene displays a greater degree of heterogeneity in the length of its 3' untranslated region.

Relative to transcription of the *cp4 epsps* gene in Roundup Ready soybeans, the untranslated 3' end of the *Agrobacterium tumefaciens* nopaline synthase (*nos*) gene has been shown to contain two 3' polyadenylation signals (AAUAAA and AAUAAU)

(Depicker *et al*., 1982; Bevan *et al*., 1983). Both of these sequences are present in the functional insert in Roundup Ready soybeans and have been shown to be capable of directing the polyadenylation and cleavage of the transgene transcripts (Depicker *et al*., 1982; Bevan *et al*., 1983; Sanfaçon *et al*., 1991) such as the *cp4 epsps* transgene.

5.0 Transcript Analysis of the DNA Sequence Flanking the 3' End of the Functional Insert in Roundup Ready Soybean Event 40-3-2

To further assess the functional activity of the region flanking the 3' end of the functional insert, studies were conducted to determine if RNA transcripts containing the genomic DNA sequence adjacent to the 3' end of the functional insert could be identified in Roundup Ready soybean event 40-3-2 (Mittanck *et al*., 2002). Results of northern blot analyses performed using more sensitive methodology and probes specific to the 3' flanking region indicate that secondary RNA transcripts are produced which contain soybean genomic DNA sequences flanking the 3' end of the insert. Quantitation experiments revealed that the levels of the most abundant ~7.4 kb secondary transcript was at least 75 times lower than that of the CP4 EPSPS transcript produced by the functional insert, while other secondary transcripts were produced at even lower levels. Further analyses by strand-specific northern blotting indicate that these secondary transcripts are transcribed in the same orientation as the CP4 EPSPS transcript produced by the functional insert. Based on these results, it is likely that these secondary transcripts are products of transcriptional termination at sites beyond the NOS 3' polyadenylation signal. As described above in Section 4, the use of varied termination/cleavage recognition sites and production of a heterogeneous population of mRNA transcripts have also been found for endogenous plant genes (Rothnie, 1996; Hunt, 1994; Gallie, 1993).

RT-PCR analyses were performed for the sequences at both the 5' and 3' ends of the functional insert to further characterize the secondary transcripts. Based on the results of the northern blot and RT-PCR analyses (Mittanck *et al*., 2002), it can be concluded that additional secondary transcripts containing the functional *cp4 epsps* coding sequences as well a portion of the soybean genomic sequence flanking the 3' end of the functional insert are produced in Roundup Ready soybean event 40-3-2. Further, as demonstrated by RT-PCR experiments, none of these secondary transcripts are initiated from within the genomic sequence flanking the 5' end of the functional insert. Therefore, these minor transcripts appear to be initiated from the E35S promoter of the insert and continue through the NOS 3' polyadenylation sequence into the soybean genomic sequence flanking the 3' end of the insert, utilizing polyadenylation signals within the downstream 3' soybean genomic flanking DNA.

5.1 Translation in Higher Plants and Relevance to Secondary Transcripts in Roundup Ready Soybean Event 40-3-2

A substantial body of published data shows that eukaryotic translational initiation occurs via a 'ribosome scanning mechanism' (Kozak, 1992). In eukaryotes, including plants,

ribosomal subunits identify a unique structure at the 5' end of the mRNA called a "cap". Once bound to the 5' end of the mRNA transcript, a ribosome scans the mRNA until the first contextually correct AUG start codon is identified and translation is initiated. Translation then continues until a stop codon is encountered. A central feature of the ribosome scanning mechanism is that only one open reading frame contained within an mRNA is translated and that translation of this open reading frame is initiated at the AUG nearest the 5' end of the mRNA. Additional detail regarding the translation process in plants is provided in Technical Appendix B.

Thus, given our current knowledge of plant mRNA translation, it is predicted that if the observed secondary transcripts in Roundup Ready soybean described above are competent to be translated, the only protein that should be produced is the full-length functional CP4 EPSPS. All other open reading frames in the secondary transcripts would also be fully contained within the primary transcript or would be located downstream of the *cp4 epsps* coding sequence translation stop codon. The translation of these alternate open reading frames would require the secondary transcripts to function as polycistrons, an occurrence that is extremely rare in eukaryotic systems (Rothnie, 1996), or for the secondary transcripts to be translated via a mechanism that would be a significant and unprecedented exception to the ribosome scanning model (Kozak, 1992).

Although ribosomal scanning accounts for the vast majority of eukaryotic translation initiation events, several exceptions to this mechanism have been documented (Gallie, 1993; Fütterer and Hohn, 1996). These exceptions are related to polycistronic mRNAs that by definition contain two or more sequential and independent ORFs that are transcribed from a single transcription unit. Functional polycistronic transcripts, while fairly common among prokaryotic organisms, are exceedingly rare in eukaryotes and presently there is only one example of a non-viral polycistronic mRNA in plants. This polycistron is transcribed from the tomPRO1 locus of tomato and encodes gamma-glutamyl kinase (GK) and gamma-glutamyl phosphate reductase (GPR) (Garcia-Rios *et al*., 1997). The mRNA encoding these two proteins is organized such that the ORFs are separated by five nucleotides. Western blot analysis of tomato tissues using either anti-GK polyclonal or monoclonal antibodies identified either a ~60 or 70 kDa polypeptide but not the expected 44 kDa GK protein. In this instance it is believed that the ribosome is engaging in '-1 translational frameshifting' or 'ribosome hopping' in order to bypass the GK stop codon and to produce a hybrid polyprotein that displays both GK and GPR activities (Garcia-Rios *et al*., 1997). As described below, there is no reason to believe that the *cp4 epsps* gene and sequences downstream of it would be capable of functioning as a polycistron.

5.2 Potential for Translation of Secondary Transcripts

If the secondary transcripts which extend beyond the NOS 3' polyadenylation sequence were to be viewed as potential polycistronic mRNA transcripts that are subject to translation (either through stop codon bypass or frame shifting near the stop codon), predictable changes in the apparent mobility of CP4 EPSPS would be observed

experimentally. Inspection of the sequence immediately downstream of the CP4 EPSPS stop codons (the *cp4 epsps* coding sequence has tandem translational stop codons) reveals that in the event of bypass of both stop codons the next stop codon would be encountered 22 codons downstream. The translation of these codons would add ~2900 Da to the apparent molecular weight of CP4 EPSPS. Likewise, frame shifting during translation would either append 22 amino acids for a +1 shift or 44 amino acids for a -1 shift to the C-terminal of CP4 EPSPS, resulting in apparent molecular weight changes of ~2500 and ~5000 Da respectively. No such deviations in the mobility of CP4 EPSPS purified from Roundup Ready soybean event 40-3-2 or additional immunoreactive protein bands have been observed (Harrison *et al*., 1996; Rogan *et al*., 1999). A second potential outcome of the translation of a polycistronic mRNA transcript is the production of two independent polypeptides. Although such translation of multiple ORFs is at times observed (reviewed in Fütterer and Hohn, 1996), it requires that the mRNA possess two distinct structural characteristics. The first characteristic is related to the spatial organization of the translated open reading frames. Of the two translated open reading frames, the upstream open reading frame must be short. Studies have demonstrated that the short upstream open reading frames can act to diminish the efficiency of translation of longer downstream open reading frames (Fütterer and Hohn, 1996 and references therein). In the instance of a test dicistronic mRNA, a single codon upstream ORF was capable of attenuating the translation of a downstream reporter gene two fold. A 30-codon open reading frame attenuated translation of a downstream reporter five-fold, and a 100 codon open reading frame abolished translation of a downstream reporter (Fütterer and Hohn, 1992). A second characteristic of the downstream open reading frame is that it must contain a contextually correct AUG codon to initiate translation.

Inspection of the predicted arrangement of open reading frames in the sequences of the secondary CP4 EPSPS-containing transcripts which extend beyond the NOS 3' polyadenylation sequence reveals that the upstream open reading frame encoding CP4 EPSPS (CP4 EPSPS + CTP = 527 codons) greatly exceeds the 100 codons necessary to abolish translation of any downstream open reading frames. Likewise, inspection of all AUG trinucleotides contained in the 1438 nucleotides that are found downstream of the CP4 EPSPS tandem stop codons reveals that only one of the AUG trinucleotides could be considered contextually optimal for translational initiation. Specifically, such trinucleotides would have to be flanked by a G at position 4 (where the A in AUG is position 1) and an A at position –3 (Gallie, 1993). This contextually optimum start codon is located 533 nucleotides downstream of the CP4 EPSPS stop codon and is contained within the 250 nucleotide CP4 EPSPS segment that is present at the 3' end of the functional insert. If translation were to initiate at this codon, it would yield a polypeptide containing a segment of CP4 EPSPS that would be in the same frame as the full-length CP4 EPSPS. Given the overall arrangement of open reading frames in the secondary mRNA transcripts containing CP4 EPSPS, it is unlikely that any open reading frame other than the full-length CP4 EPSPS would be translated.

Western blot analyses that were conducted with crude protein preparations from Roundup Ready soybean event 40-3-2 tissue using polyclonal antibodies raised against the CP4 EPSPS protein confirm that no protein product other than the full-length 46 kDa CP4 EPSPS is detected (Harrison *et al*., 1996; Rogan *et al*., 1999). Since polyclonal antibodies react with numerous antigenic sites (amino acid sequences), if a hybrid polyprotein were synthesized that contains sufficient amino acid sequences derived from the *cp4 epsps* gene coding sequence, this CP4 EPSPS containing-protein would be expected to be detected by western blotting. No protein was detected in the initial studies other than the full-length CP4 EPSPS protein that was encoded by the full length CP4 EPSPS gene contained within the functional insert (Padgette *et al*., 1995; Harrison *et al*., 1996). Subsequent analyses using western blotting methods that were at least ten-fold more sensitive (LOD for CP4 EPSPS = 100 picograms) than the initial studies confirm that only the expected 46 kDa CP4 EPSPS is produced, and that no larger or smaller protein fragments are detected (Rogan *et al*., 1999).

5.3 Safety Assessment of ORFs 3' of CP4 EPSPS Coding Sequence

To assess the safety of any putative polypeptides that could be translated from the region beyond the NOS 3' polyadenlyation sequence, including the soybean genomic DNA which is beyond the 3'end of the functional insert, bioinformatics analyses have been conducted on all predicted polypeptides (ORFs) lying within this region. These analyses were conducted to determine whether any potential translated ORFs from the region beyond the NOS 3' polyadenylation sequence are homologous to known toxins or allergens. No evidence exists to indicate that any of the predicted polypeptides analyzed are produced in line 40-3-2. This assessment indicated that the theoretical open reading frames would be very short, highly unlikely to produce any protein and, if produced, show no sequence similarity to any known protein toxins or allergens.

These analyses represent a theoretical safety assessment that uses bioinformatic tools to predict the characteristics of a polypeptide. In order to minimize bias, no attempt is made to assess that probability that any one reading frame will be translated preferentially relative to any other reading frame. Furthermore, bias that may result from the arbitrary selection of a start codon is eliminated by translating the predicted polypeptides from stop codon to stop codon. This bioinformatic assessment demonstrates that for the 28 predicted polypeptides encoded by the sequence that lies beyond the 3' end of the NOS polyadenylation sequence, there are no homologies or immunologically significant sequence similarities to proteins identified as allergens. Moreover, the predicted polypeptides do not display sufficient identity or similarity at the level of primary structure to indicate homology with any known toxic or pharmacologically active proteins. These findings are consistent with previous studies that have demonstrated the food and feed safety of Roundup Ready soybean line 40-3-2.

5.4 Conclusions on Transcript Analysis and Probable Function of Secondary Transcripts

In the previous Roundup Ready soybean safety assessment (Astwood *et al*., 2000), the possibility that transcriptional termination occurred beyond the NOS 3' polyadenylation site and that a larger than expected "read-through" mRNA transcript could be produced was considered. A theoretical secondary mRNA was considered to have the potential to include multiple open reading frames including the full-length CP4 EPSPS protein and any putative ORFs in the NOS 3' region or the 250 bp segment of CP4 EPSPS DNA located at the 3' end of the insert. As discussed above, any read-through mRNA produced in Roundup Ready soybean event 40-3-2 would not be expected to produce additional proteins since eukaryotes, such as plants, typically do not translate multiple open reading frames (ORFs) from single mRNAs. In the exceptionally rare instance where second ORFs contained on a single mRNA have been detected in plants (Garcia-Rios *et al*., 1997; Fütterer and Hohn, 1996), a single hybrid polyprotein has been observed, rather than the synthesis of two independent proteins. The presence of multiple translation stop codons in the sequences lying downstream of the full-length *cp4 epsps* coding sequence would be expected to prevent the synthesis of a hybrid polyprotein containing elements encoded by the DNA beyond the *cp4 epsps* coding region. Western blotting confirmed that only the expected 46 kDa CP4 EPSPS protein is detected in event 40-3-2. No unexpected protein, including any large molecular weight protein potentially produced by the secondary read-through mRNA, was detected by western blotting when using more sensitive western blot methods (Rogan *et al*., 1999).

6.0 Review of Food, Feed and Environmental Safety Assessment

Based on northern blot, RT-PCR and western blot analyses, combined with the detailed evaluation of the sequences of the DNA flanking the 3' end of the primary insert in Roundup Ready soybean, as well as the understanding of translation initiation and termination in plants, it is concluded that no new proteins are expected or observed to be produced which would be derived from the DNA flanking the 3' end of the primary insert. The northern and western blotting data confirm the conclusion that Roundup Ready soybean event 40-3-2 contains one functional *cp4 epsps* gene that produces the full-length 46 kDa CP4 EPSPS protein that confers the glyphosate tolerance trait.

Furthermore, the DNA flanking the primary insert, and any secondary mRNA transcripts, or putative products of these transcripts, regardless of abundance, were constituents of the regulatory samples that were used in all the safety assessment studies which addressed the food, feed and environmental safety of event 40-3-2.

6.1 Applicability of Safety Data to Secondary Transcripts

A series of animal feeding studies have been completed previously using diets incorporating seed or processed fractions from Roundup Ready soybean event 40-3-2 (Hammond *et al*., 1996). These studies address the nutritional equivalence of Roundup

Ready soybean event 40-3-2 when used as animal feed, the relative safety of any expressed proteins or peptides (or any other newly produced constituent), the potential for any pleiotropic effect caused by the insertion process or site of insertion, and any other constituent that would result from the insert derived from plasmid PV-GMGT04. These feeding studies confirmed the food and feed safety and nutritional equivalence of diets from Roundup Ready soybean event 40-3-2 to diets from the control soybean varieties. If the presence of the secondary transcripts were to result in gene function or create potential pleiotropic effects, no measurable effects were observed. The nutritional value or wholesomeness of Roundup Ready soybean event 40-3-2, even when fed to animals at levels much higher than humans would encounter in the diet, was the same as that of conventional varieties of soybeans.

If a protein or peptide other than CP4 EPSPS were to be unexpectedly produced from the secondary transcripts that contain the region of soybean genomic sequence adjacent to the 3' end of the functional insert, these proteins or peptides would be likely to contain amino acid sequences of the CP4 EPSPS protein. The safety of such CP4 EPSPS-containing products has previously been addressed through bioinformatics, digestive fate and acute oral toxicity studies performed with the CP4 EPSPS protein, establishing that CP4 EPSPS protein and its proteolytic fragments are not homologous to toxins or allergens, are rapidly digested and are non-toxic (Harrison *et al*., 1996; Padgette *et al*., 1996a).

Roundup Ready soybean event 40-3-2 has been approved for planting and/or consumption in many countries worldwide. An assessment of the food and feed safety data developed for soybeans containing the 40-3-2 event established that these soybeans are as safe and nutritious as other commercial soybean varieties. Neither the presence of a rearranged segment of soybean genomic DNA in the region flanking the 3' end of the functional insert nor the presence of additional *cp4 epsps*-containing mRNA transcripts which extend beyond the NOS 3' polyadenylation sequence change the conclusion of this assessment for the following reasons: a) consumption of DNA and RNA *per se* is generally recognized as safe by food safety experts, including the FDA (US FDA, 1992) and the FAO/WHO (FAO/WHO, 1996); b) the food and feed safety studies were conducted on soybeans that contained genomic rearrangements in the 3' flanking region (Lirette *et al*., 2000; Goley *et al*., 2002) as well as the detected secondary transcripts (Mittanck *et al*., 2002); c) results of compositional and nutritional analyses established that these soybeans are comparable to other soybean varieties and are not affected by the 40-3-2 insert itself or any other effects brought about by this insertion (Padgette *et al*., 1994; Padgette *et al*., 1996; Taylor *et al*., 1999; Burks and Fuchs, 1995); and d) a series of animal feeding studies showed that Roundup Ready soybeans are as safe and as nutritious as conventional soybeans (Hammond *et al*., 1996).

7.0 Conclusions

The region immediately flanking the 3' end of the primary, functional insert in Roundup Ready soybean event 40-3-2 has been found to contain rearranged soybean genomic DNA. Northern blot and RT-PCR analyses indicate that secondary, less abundant transcripts that contain the soybean genomic sequence flanking the 3' end of the functional insert are produced. Since translation of competent mRNA transcripts in plants, with very rare exceptions, only produces a single polypeptide, the only protein product expected to be produced from the 40-3-2 insert is the full-length CP4 EPSPS. Indeed, the only CP4 EPSPS-containing protein detected in Roundup Ready soybean is the 46 kDa full-length CP4 EPSPS protein.

This additional characterization data of the DNA sequence flanking the 3' end of the primary insert of event 40-3-2 does not alter the previous conclusion that this event contains a single functional gene cassette encoding the 46 kDa CP4 EPSPS protein, which is responsible for glyphosate tolerance. The DNA derived from the transformation vector used to produce event 40-3-2, the DNA flanking the inserts and any transcripts containing the 3' flanking region, or other putative products derived from it, are present as constituents in all the soybean materials used to establish the food, feed and environmental safety of Roundup Ready soybean event 40-3-2. Therefore, any changes or rearrangements that may have occurred during the process of insertion of the DNA into the soybean genome were constituents of the test material used in the safety assessment of this product.

The consequences of the rearrangement and transcription of a segment of the DNA flanking the 40-3-2 insert have been assessed as part of the larger safety assessment conducted on Roundup Ready soybeans. The studies led to the conclusion that Roundup Ready soybeans are agronomically, compositionally and nutritionally comparable to conventional soybeans, except for the Roundup Ready trait. The information regarding the nature and transcriptional activity of the genomic DNA flanking the 3' end of the functional insert does not change the conclusions reached previously that: (1) Roundup Ready soybeans contain one functional *cp4 epsps* gene cassette which produces the full-length CP4 EPSPS protein; (2) Roundup Ready soybeans are as safe and nutritious as conventional soybean varieties; and (3) Roundup Ready soybeans do not pose a plant pest risk or otherwise pose an increased risk to the environment relative to conventional soybean varieties.

8.0 Literature Cited

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol*. **215**: 403-410.

Astwood, J.D., Hammond, B.G., Dobert, R.C., and Fuchs, R.L. 2000. Updated molecular characterization and safety assessment of Roundup Ready soybean event 40-3-2. Monsanto Technical Report: MSL-16712. St. Louis, Missouri.

Bevan, M., Barnes, W.M., and Chilton, M-D. 1983. Structure and transcription of the nopaline synthase gene region of T-DNA. *Nucl Acid Res* 11:369-385.

Burks, A.W. and R.L. Fuchs. 1995. Assessment of the endogenous allergens in glyphosate-tolerant and commercial soybean varieties. *Journal Allergy Clinical Immunology* **96**: 1008-1010.

Christou P., McCabe D.E., Swain W.F. *et al.* Legume Transformation. In: Verma DPS, ed. *Control of Plant Gene Expression.* New Jersey: Telford Press, 1993:547-564

Christou P. In: *Particle Bombardment for Genetic Engineering of Plants.* Texas: Academic Press, 1996:135-145

Czarnecka, E., Gurley, W.B., Nagao, R.T., Mosquera, L.A. and Key, J.L. 1985. DNA sequence and transcript mapping of a soybean gene encoding a small heat shock protein. *Proc Natl Acad Sci USA* **82**: 3726-3730.

Dean, C., Tamaki, S., Dunsmuir, P., Favreau, M., Katayama, C., Dooner, H., and Bedbrook, J. 1986. mRNA transcripts of several plant genes are polyadenylated at multiple sites in vivo. *Nucl Acid Res* **14**(5):2229-2240.

Depicker, A., Stachel, S., Dhaese,P., Zambryski, P., and Goodman, H. 1982. Nopaline Synthase: Transcript Mapping System and DNA Sequence. *J. Molec. Appl. Genet.* **1**:561-573.

Enriquez-Harris, P., Levitt, N., Briggs, D., and Proudfoot, N.J. 1991. A pause site for RNA polymerase II is associated with termination of transcription. *EMBO J* **10**:1833-1842.

FAO/WHO. 1996. *Biotechnology and Food Safety*. The Food and Agriculture Organization and The World Health Organization, Geneva, pp 1-27.

Fütterer, J., and T. Hohn. 1992. Role of an upstream open reading frame in the translation of polycistronic mRNAs in plant cells. *Nucl Acid Res* **20**:3851-3857.

Fütterer, J., and T. Hohn. 1996. Translation in plants - rules and exceptions. *Plant Mol Biol* **32**:159-189.

Gallie, D.R. 1993. Posttranscriptional regulation of gene expression in plants. *Ann Rev. Plant Physiol Plant Mol Biol* **44**:77-105.

Garcia-Rios M., Fujita, T., LaRosa, P.C., Locy R.D., Clithero, J.M., Bressan, R.A., and Csonka, L.N., 1997. Cloning of a polycistronic cDNA from tomato encoding gamma-glutamyl kinase and gamma-glutamyl phosphate reductase. *Proc Natl Acad Sci USA* **94**:8249-8254.

Gheysen, G., Van Montagu, M., and Zambryski, P. 1987. Integration of *A. tumefaciens* T-DNA involves rearrangement of target plant DNA sequences. *Proc Natl Acad Sci USA* **84**:6169-6173

Goley, M.E., Reiser S.E., Cavato, T.A., Beazley, K.A. and Lirette, R.P. 2002. The DNA Sequence Flanking the 3' End of the Functional Insert of Roundup Ready Soybean Event 40-3-2 is Identical to DNA Sequences from the Wild-type Soybean Lines A5403 and A3244. Monsanto Technical report: MSL-17561.

Hammond, B.G., Vicini, J.L., Hartnell, G.F., Naylor, M., Knight, C.D., Robinson, E.H., Fuchs, R.L., and Padgette, S.R. 1996. The feeding value of soybeans fed to rats, chickens, catfish and dairy cattle is not altered by genetic incorporation of glyphosate tolerance. *Journal of Nutrition* **126**: 717-727.

Harrison, L.A., Bailey, M.R., Naylor, M.W., Ream, J.E., Hammond, B.G., Nida, D.L., Burnette, B.L., Nickson, T.E., Mitsky, T.A., Taylor, M.L., Fuchs, R.L., and Padgette, S.R. 1996. The expressed protein in glyphosate-tolerant soybean, 5-enolpyruvylshikimate-3-phosphate synthase from *Agrobacterium* sp. strain CP4, is rapidly digested in vitro and is not toxic to acutely gavaged mice. *Journal of Nutrition* **126**: 728-740.

Hunt, A. 1994. Messenger RNA 3' end formation in plants. *Ann Rev Plant Physiol Plant Mol Biol* **45**:47-60.

Hunt, A.G. 1988. Identification and characterisation of cryptic polyadenylation sites in the 3' region of a pea ribulose-1,5-bisphosphate carboxylase small subunit gene. *DNA* **7**:329-336.

Klahre, U., Hemmings-Mieszczak, M., and Filipowicz, W. 1995. Extreme heterogeneity of polyadenylation sites in mRNA encoding chloroplast RNA-binding proteins in *Nicotiana plumbaginifolia*. *Plant Mol Biol*. **28**:569-574

Kohli, A., Griffiths, S., Palacios, N., Twyman, R.M., Vain, P., Laurie, D.A., and Christou, P. 1999. Molecular characterization of transforming plasmid rearrangements in

transgenic rice reveals a recombination hotspot in the CaMV 35S promoter and confirms the predominance of microhomology mediated recombination. *Plant J* **17**:591-601.

Kohli, A., Leech, M., Vain, P., Laurie, D., and Christou, P. 1998. Transgene organization in rice engineered through direct DNA transfer supports a two-phase integration mechanism mediated by the establishment of integration hotspots. *Proc Natl Acad Sci USA* **95**:7203-7208.

Kozak, M. 1992. A consideration of alternative models for the initiation of translation in eukaryotes. *Crit Rev Biochem Mol Biol* **27**:385-402.

Lirette, R.P., Cavato, T.A., Reiser, S.R., Beazley, K.A., Borovkova, I.G., Higgins, M.R., Albee, L.D., and Dixon, D.A. 2000. Further molecular characterization of Roundup Ready soybean event 40-3-2. Monsanto Technical report: MSL-16646.

Mittanck, D.W., Palmer, G.M., and Lirette, R.P. 2002. Transcript analysis of the sequence flanking the 3' end of the functional inset in Roundup Ready soybean event 40-3-2. Monsanto Technical report: MSL-17342.

Padgette, S.R., Re, D.B., Hammond, B.G., Fuchs, R.L., Rogers, S.R., Harrison, L.A., Nida, D.L., Naylor, M.W., Kolacz, K.H., Taylor, N.B., and Ream, J.E. 1994. *Safety, Compositional and Nutritional Aspects of Glyphosate-tolerant Soybeans: Conclusions Based on Studies and Information Evaluated According to FDA's Consultation Process.* Monsanto Company, St. Louis, Missouri, pp 1-107.

Padgette, S.R., Kolacz, K.H., Delannay, X., Re, D.B., LaValee, B.J., Tinius, C.N., Rhodes, W.K., Otero,. Y.I., Barry, G.F., Eichholtz, D.A., Peschke, V.M., Nida, D.L., Taylor, N.B., and Kishore, G.M. 1995. Development, identification, and characterization of a glyphosate-tolerant soybean line. *Crop Science* **35**: 1451-1461.

Padgette, S.R., Taylor, N.B., Nida, D.L., Bailey, M.R., MacDonald, J., Holden, L.R., and Fuchs, R.L. 1996. The composition of glyphosate-tolerant soybean seeds is equivalent to that of conventional soybeans. *Journal of Nutrition* **126**: 702-716.

Padgette, S.R., Re, D.B., Barry, G.F., Eicholtz, D.E., Delannay, X., Fuchs, R.L., Kishore, G.M., and Fraley, R.T. 1996a. New weed control opportunities: Development of soybeans with a Roundup Ready gene. *In* Duke, S.O. (ed.) *Herbicide-Resistant Crops. Agricultrual, Economic, Regulatory and Technical Aspects*. CRC Lewis Publishers, pp 53-84.

Pawlowski, W.P. and Somers, D.A. 1998. Transgenic DNA integrated into the oat genome is frequently interspersed by host DNA. *Proc Natl Acad Sci USA* **95**:12106-12110.

Proudfoot, N. 1989. How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem Sci* **14**:105-110.

Rogan, G.J., Dudin, Y.A., Lee, T.C., Magin, K.M., Astwood, J.D., Bhakta, N.S., Leach, J.N., Sanders, P.R., and Fuchs, R.L. 1999. Immunodiagnostic methods for detection of 5-enolpyruvylshikimate-3-phosphate synthase in Roundup Ready soybeans. *Food Control* **10**: 407-414.

Rothnie, H.M. 1996. Plant mRNA 3'-end formation. *Plant Molec Biol* **32**:43-61

Sanfaçon, H., Brodmann, P., and Hohn, T. 1991. A dissection of the cauliflower mosaic virus polyadenylation signal. *Genes Devel* **5**:141-149.

Singer, M., and Berg, P. 1991. *Genes and Genomes - A Changing Perspective*. University Science Books, Mill Valley, CA.

Sullivan, M.L. and Green, P.J. 1993. Post-transcriptional regulation of nuclear-encoded genes in higher plants: the roles of mRNA stability and translation. *Plant Mol Biol* **23**:1091-1104.

Taylor, N.B., Fuchs, R.L., MacDonald, J., Shariff, A.R., and Padgette, S.R. 1999. Compositional analysis of glyphosate-tolerant soybeans treated with glyphosate. *Journal of Agricultural and Food Chemistry* **47**: 4469-4473.

Tran, D.P., Kim, S.J., Park, N.J., Jew, T.M., and Martinson, H.G. 2001. Mechanism of poly(A) signal transduction to RNA polymerase II in vitro. *Mol Cell Biol* **21**:7495-7508.

Tomkins, J.P., D.A. Frisch, J.R. Byrum and R.A. Wing. 2000. Construction and Characterization of a BAC Library for the Soybean Cultivar A3244. *Soybean Genetics Newsletter.* 27 [Online journal]. URL http://www.soygenetics.org/articles/sgn2000-002.htm (posted 10 April 2000.)

United States Food and Drug Administration. 1992. Foods derived from new plant varieties. *Federal Register* **57**: 22984-23005.

Windels, P., Taverniers, I., Depicker, A., Van Bockstaele, E., and M. De Loose. 2001. Characterization of the Roundup Ready soybean insert. *Eur Food Res Technol* **213**:107-112.

Signature Page

APPENDIX A

Gene Transcription in Higher Plants and Cleavage of the 3' Ends of Primary RNA Transcripts.

Gene transcription in eukaryotic organisms is a dynamic multistage process that leads to the accumulation of mRNA molecules. Overall, transcription can be divided into three stages: the synthesis of the primary RNA transcript (also referred to as heterogeneous nuclear RNA), post-transcriptional RNA processing to yield the mature mRNA that can be subsequently translated, and mRNA degradation. Each of the aforementioned stages of transcription is affected by both *cis*- and *trans*-acting factors and each stage may be further subdivided into defined events. The interplay and interaction between these processes is responsible for producing the measurable, steady-state levels of a particular mature transcript in a given tissue. As biochemical processes, it should be noted that the observed steady-state effect of these steps represents the most favored, probabilistically and thermodynamically, condition where lower frequency processes occur regularly. As a consequence, all genes produce a population of transcripts, some of which are more probable and abundant than others, and the dominant gene expression products are most relevant when considering biochemical and physiological consequences.

Although many aspects of transcription are shared by all eukaryotes, subtle mechanistic differences have been identified that distinguish vertebrate and plant transcription (reviewed in Rothnie, 1996). Through sequence analysis of mRNAs from vertebrates and plants, one such mechanistic difference that has been identified in plants is the process of 3' cleavage of primary RNA transcripts. As a result of this difference, the transcription of some plant genes leads to the accumulation of a fairly heterogeneous population of mature mRNAs that have a common coding sequence, but which differ in the sequence and length of their 3' untranslated regions exclusive of the poly(A) tail. These differences in the 3' untranslated region are unrelated to differential primary RNA splicing and distinguish plant mRNAs from vertebrate mRNAs, where a given gene displays a greater degree of homogeneity in the length of its 3' untranslated region.

Transcription initiation of eukaryotic genes is also a multi-step process and minimally requires the basal transcription machinery composed of RNA polymerase II and several general transcription factors. To initiate transcription, these general transcription factors, RNA polymerase II, and various cofactors must assemble at the upstream promoter DNA region of a gene to form a multi-protein transcription initiation complex. Following the formation of the transcription initiation complex, the DNA strands at the transcription start site are separated (promoter opening), RNA polymerase II initiates synthesis of the RNA transcript (transcription initiation), leaves the promoter region (promoter clearance), and translocates along the DNA template under continuous extension of the RNA chain (transcription elongation) (Singer and Berg, 1991). Primary RNA transcript elongation then proceeds to a location downstream of the coding sequence's translation stop codon, where transcriptional termination occurs.

Transcriptional termination is characterized by the cessation of RNA chain elongation coupled with the endonucleolytic cleavage of the nascent transcript and subsequent addition of a poly(A) tail. In vertebrates, transcriptional termination appears to be the result of RNA polymerase II pausing at a site located downstream of the highly conserved AAUAAA polyadenylation signal (Enriquez-Harris *et al*., 1991). The precise mechanism for RNA polymerase II pausing is not known; however, recent studies have demonstrated that two proteins associated with the RNA polymerase II complex may be responsible for transduction of the AAUAAA motif and signaling RNA polymerase to pause. This pause in transcription occurs in a gene-dependant manner some 100 to >4000 nucleotides downstream of the AAUAAA polyadenylation motif (Proudfoot, 1989). Concomitant or prior to RNA polymerase II pausing, cleavage of the primary RNA transcript occurs at a location approximately midway between the conserved AAUAAA polyadenylation signal and U/GU-rich elements (typically 10 to 30 nucleotides downstream of the AAUAAA polyadenylation signal). This cleavage of the nascent transcript is coupled to the addition of up to 250 adenosines that comprise the "poly(A) tail" (Tran *et al*., 2001). The segment of the primary RNA that is located downstream of the 3' cleavage site is degraded through the combined activities of a helicase that resolves the primary RNA-template DNA hybrid and a 5' to 3' ribonuclease (Proudfoot, 1989). Once degradation of the primary transcript has extended to the paused RNA polymerase II, RNA polymerase II detaches from the template DNA. The segment of the primary transcript that is located upstream of the cleavage site is also processed to yield the mature mRNA. In addition to 3' polyadenylation, transcript processing includes the addition of the 5' cap structure and the removal of introns through splicing. Capping and splicing may occur concomitant with primary RNA transcript elongation and prior to primary RNA cleavage, while polyadenylation is by definition a post-cleavage event.

Although less is known about transcriptional termination in plants relative to the understanding of this process in vertebrates, several key differences between vertebrates and plants have been observed. The most significant distinction is related to the configuration of elements downstream of the coding sequence translation stop codon. Specifically, plant genes do not contain *per se* a U- or GU-rich element and, more importantly, they do not utilize a single highly conserved polyadenylation signal motif to direct transcript termination/cleavage. Rather, plant genes have two less conserved elements that seem to be critical in directing transcript termination and processing of nascent transcripts. These elements are described as far upstream elements (FUEs) that enhance processing efficiency, and several AU-rich sequence motifs known as near upstream elements (NUE) that resemble the vertebrate AAUAAA polyadenylation motif (*i.e.,* AAUUAAA, AUAUAA, AAUAAU). Although cleavage and polyadenylation is observed 10 to 40 nucleotides downstream of NUEs in a manner that parallels signaling of polyadenylation in vertebrates, heterogeneity exists among the 3' ends of mRNAs for many plant genes because no single NUE is exclusively used as the 3' cleavage signal (Hunt, 1994). Thus, in plants, virtually all genes have multiple potential poly(A) sites and, as a result, the position of cleavage of the primary transcript is quite heterogeneous,

often producing a population of mRNAs that have varied endpoints (Rothnie, 1996, and references therein). For instance, the genes encoding the small subunit of the pea ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbcS*) protein and soybean small heat shock (*Gmhsp17.5-E*) protein contain three and four 3' cleavage sites, respectively (Hunt, 1988; Czarnecka *et al*., 1985). Likewise, an extreme example of this phenomenon is observed with a chloroplast-RNA binding protein gene in *Nicotiana plumbaginifolia* that contains fourteen 3' cleavage sites with a corresponding number of transcripts (Klahre *et al*., 1995).

Unlike vertebrate genes that have a consensus polyadenylation signal and unique polyadenylation site, plant genes have evolved to have multiple polyadenylation signals and sites. Moreover, the transcript processing machinery in plants is highly flexible and capable of identifying cryptic polyadenylation signals. Although the evolutionary driver for this flexibility in plant transcript processing is not known, it has been proposed that post-transcriptional control mechanisms are of added significance in plants due to their sessile lifestyle (Sullivan and Green, 1993). Since it is known that 3' untranslated regions and poly(A) tails can be determinants of mRNA stability and degradation, and the control of translation, heterogeneity among the 3' untranslated regions of mRNAs derived from a single gene may provide the added speed necessary for plants to respond to changing environmental conditions from which they are unable to flee. Likewise, transgenes that are assembled using elements such as transcriptional promoters and polyadenylation signals that are derived from plant, bacterial or plant sources, would be expected to display the same 3' untranslated region heterogeneity as endogenous plant genes.

APPENDIX B

Translation in Higher Plants

Eukaryotic translation is controlled primarily at the initiation level. The nature of the 5'- and 3'- non-coding sequences (leaders and trailers, including poly(A) tails) and a set of translation initiation factors are the major determinants of this control. With a limited number of exceptions (described below), the initiation of translation in eukaryotes occurs via a 5' cap dependent scanning mechanism (Kozak, 1992). Initially, components of the 40S ribosomal subunit, in combination with the poly(A) tail bound poly(A) binding protein, identify the "cap", a unique methylated, phosphorylated nucleotide located at the 5' end of a transcript. Upon identification of the cap, the transcript is scanned by the 40S subunit to find the first contextually correct AUG start codon where assembly of the ribosome complex is completed and translation is initiated. Translation then continues through the open reading frame (ORF) that is defined by this AUG as the ribosome successively translates codons with aminoacyl-tRNAs to generate the encoded polypeptide. Translation continues until the ribosome complex reaches a stop codon (UAA, UAG or UGA) for which no complementary tRNA is available. At that time, release factors associate with the ribosome, triggering the release of the nascent polypeptide and termination of translation (Fütterer and Hohn, 1996). A central feature of the ribosome scanning mechanism is that only one open reading frame contained within an mRNA is translated and that translation of this open reading frame is initiated at the AUG nearest the 5' end of the mRNA.