Volume 15 Number 23 1987 Nucleic Acids Research

Putative polyadenylation signals in nuclear genes of higher plants: a compilation and analysis

C.P.Joshi

Division of Biochemical Sciences, National Chemical Laboratory, Pune 411 008, India

Received August 17, 1987; Accepted October 14, 1987

ABSTRACT

In animal and viral pre-mRNAS, the process of polyadenylation is mediated through several cis-acting poly (A) signals present upstream and downstream from poly (A) sites. The situation regarding polyadenylation of higher plant pre-mRNAS, however, has remained obscure so far. In this paper, a search for putative poly (A) signals is made by considering the published data from 46 plant genomic DNA sequences. Certain domains in the 3' untranslated regions from nuclear genes of higher plants were compiled and occurrence of sequence motifs such as AATAAA, CAYTG, YGTGTTYY and YAYTG was scored in relation to poly (A) sites. Moreover, consensus sequences for important regions in the 3' untranslated sequences and poly (A) signals were also deduced from the data. It was inferred that sequence motifs similar to poly (A) signals exist around poly (A) sites but some of them are in entirely different spatial relationship than observed in other eukaryotes. This indicates their probable non-involvement in the process of polyadenylation in higher plants necessitating a functional analysis approach to define the plant specific poly (A) signals.

INTRODUCTION

In eukaryotes, the primary transcripts (pre-mRNAs) of nuclear genes are longer than the mature and translatable mRNAs present in the cytoplasm. The synthesis of pre-mRNA is followed, therefore, by extensive processing which includes 5' capping, intron splicing and polyadenylation. Our limited knowledge regarding the mechanism of polyadenylation stems from studies of animal and viral systems. Higher plants have been investigated for this purpose only recently. The polyadenylation or 3' processing involves two steps, namely endonucleolytic cleavage at a specific point in the 3' untranslated region (termed the poly (A) site) and addition of several A residues to the cleaved end. The selection of the poly (A) site is presumed to be regulated by cis-acting poly (A) signals located upstream and downstream from the poly (A) site (41, 42).

The hexanucleotide AATAAA lying 10 to 30 bases upstream from the poly (A) site represents the almost ubiquitous poly (A) signal in animal or viral pre-mRNAs (43). Deletion of this sequence motif (44), point mutations substituting one or more bases in this signal (45, 46) and varied spatial relationship between poly (A) signal and normal poly (A) site (44) resulted in abnormal or inefficient processing of the 3' termini of such pre-mRNAs. Moreover, it has been suggested that this string of bases is involved in the cleavage site selection through association with small nuclear ribonucleoproteins (snRNP) (47) or some factor in the crude nuclear extract (48). It would, however, be difficult to imagine that this signal alone can be sufficient for the complex process of polyadenylation in eukaryotes. Recently, Kessler and co-workers showed that the presence of the consensus AATAAA sequence and the downstream sequences are obligatory for efficient and correct polyadenylation (49).

The experimental evidence, obtained from deletion analysis in the 3' untranslated region, has indicated that a region of 30 to 80 bases downstream of the poly (A) site is required for efficient post-transcritional 3' processing (50-53). On the other hand, Mason et al. assigned the role of cleavage site selection to these downstream sequences (54). The signal CAYUG, flanking the poly (A) site has been generally observed in over 50 animal pre-mRNAs (55). On the basis of sequence complementarity, it has been suggested that the U4 snRNP forms a complex with AATAAA for the primary site selection and with CAYUG for precise cleavage site selection (See (56) for the controversial views). A second signal YGTGTTYY located approximately 30 bases downstream from the AATAAA signal has been observed in 67% of the mammalian pre-mRNAs (57). Alteration in the spatial relationship between these signals reduced the levels of processing of the mRNA 3' termini while the deletion of the YGTGTTYY motif caused a marked drop in the efficiency of polyadenylation to below 5% (57).

The generalizations made on the basis of animal studies are usually regarded as 'eukaryotic features' assuming that plants would follow the same mechanistic rules. How far this dogma of similarity in gene structure, function and regulation between animals and plants is valid has remained unanswered due to the paucity of corresponding information on plant nuclear genes. Over the past five years, however, a large number of genomic DNA sequences encoding various proteins in plants

have been isolated and sequenced. Recent compilation and analysis studies with plant leader sequences (58) and exon-intron junctions (59) have suggested that even though plant and animal genes show overall conservation in their structural characteristics, several minute differences do exist to enable one to postulate a difference in their 'modus operandi'. No attempts, however, have been made previously to compile and analyse data regarding polyadenylation signals from the available published information and to assess if poly (A) signals in plants are present in the same manner as in animal pre-mRNAs.

In this paper, certain domains of the 3' untranslated regions in forty-six nuclear genes of higher plants have been examined and a search made for putative poly (A) signals. Further, the structural features of the region immediately downstream of translational stop codon and the context sequences of poly (A) sites have also been analysed. Such studies can be of value in formulating the general norms in polyadenylation of plant pre-mRNAs. Extensive deletion and mutation analyses would be required before a functional significance can be attributed to these poly (A) signals in plants and these are, therefore, designated as putative in this paper.

MATERIALS AND METHODS

1. Selection of plant genes for the present compilation :

a) Only nuclear genes of higher plants have been included.

b) Genes with abnormal structure or unknown function have been excluded.

c) Since the interest of this analysis is centered on upstream and downstream sequences from poly (A) site, only those genomic DNAs, analogous to pre-mRNAs except that they have T instead of U, having known or predicted poly (A) site have been included (Table 1). Poly (A) sites had been proposed in published papers on the basis of S_1 nuclease mapping or cDNA/closely related gene comparisons.

2. Collection of data and methods of analysis : The 3' sequences of forty-six published nuclear genes fulfilling the above criteria have been collected. The features of certain domains of these sequences are presented in Table 1:-

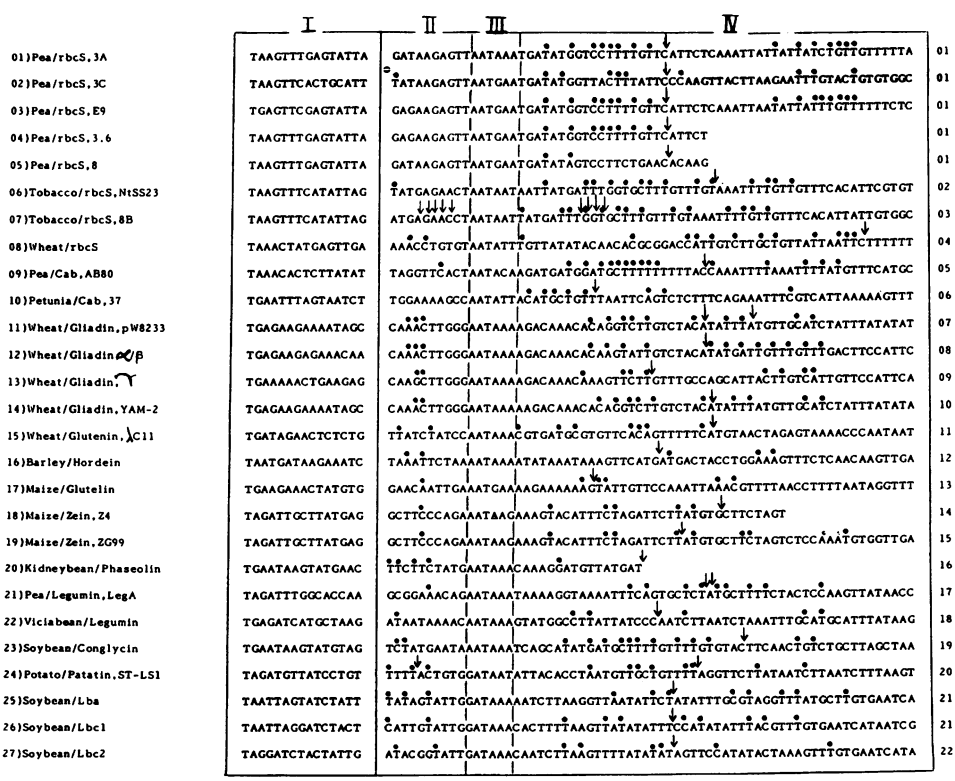
a) Domain I:- The twelve bases downstream from the translational stop codons were aligned with the TAA/TAG/TGA triplet as a reference point. The consensus sequence for downstream context of stop codon

(Table 2) was deduced using the method of Cavener (60) as given below.

b) Domain II:- Taking aligned AATAAA or its derivative as a reference point, 10 bases upstream have been listed and a search for YAYTG signal was made (window size 5, minimum 60% match). (Fig. 1 a). Moreover, a consensus for upstream context of AATAAA was also deduced (Table 3).

c) Domain III:- The AATAAA signal (minimum 4 out of 6 bases match) lying upstream of known/predicted poly (A) site was aligned and a consensus was worked out (Fig. 1b).

d) Domain IV:- The fifty bases (when available) downstream of AATAAA signal were compiled and screened for putative poly (A) downstream signal YGTGTTYY (minimum 5/8 matches) (Fig. 1 d). In addition, a consensus sequence for 10 bases following an AATAAA motif was also derived (Table 4). Moreover, 12 bases flanking the poly (A) site on either side were aligned taking the poly (A) site as a



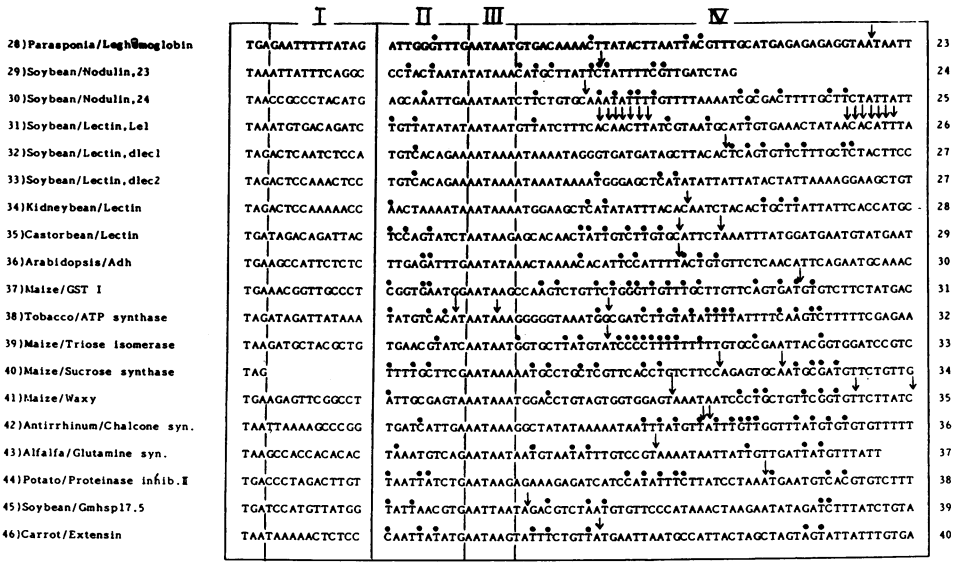


Table 1 : Compilation of certain domains in 3' untranslated regions of 46 higher plant genomic DNA sequences.

a) Domain I : Twelve bases downstream from stop codons (Boxed).

b) Domain II : Ten bases upstream from AATAAA like motif. Dots above the nucleotide, mark the beginning of a YAYTG like motif (minimum 60% match).

c) Domain III : AATAAA like motif.

d) Domain IV : Fifty nucleotides downstream from AATAAA. Arrows indicate the locations of suggested poly (A) sites. Dots above the nucleotide denote the beginning of YGTGTTYY like motif (minimum 60% match).

The references numbered at the right end of each sequence are listed in reference section.

reference point and a search for CAYTG was made and scored (Fig. 2). The consensus sequences for CAYTG signals and context of the poly (A) site were deduced (Figs. 1 c & 2). It should be noted that some ambiguity does exist regarding the exact poly (A) site and AATAAA signal location when multiple poly (A) sites and AATAAA signals were present in 3' untranslated regions.

3. Criteria for consensus sequences : The following guidelines, as suggested by Cavener (60), were used for deciding the bases to be included in the consensus:

a) The relative frequency of a single nucleotide at a certain position should be greater than 50% and greater than twice the relative frequency of the second most frequent base. A single base satisfying these conditions was assigned the status of consensus nucleotide.

b) When no single base fulfilled the above mentioned condition, a pair of bases was suggested as co-consensus nucleotides at a position

Table 2: Frequency of nucleotide occurrence at 12 positions following a stop codon (arrowed).

	▼	▼	▼	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	+11	+12
A%	--	61	78	44	27	36	27	33	44	24	40	33	27	33	20
G%	--	39	22	33	9	11	24	22	13	20	9	11	11	18	33
C%	--	--	--	4	20	11	22	24	13	11	27	18	22	18	27
T%	100	--	--	18	44	42	27	20	29	44	24	38	40	31	20

Total number of observations : 45.

Consensus : T A A A t T t a a t a t t a g
(Plant) G G A a

Table 3 : Nucleotide frequencies at 10 positions, upstream from AATAAA like motif.

	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1
A%	17	48	33	43	35	28	50	26	17	26
G%	17	17	17	15	20	26	2	33	24	37
C%	17	13	9	11	26	17	11	11	15	9
T%	48	22	41	30	20	28	37	30	43	28

Total number of observations : 46

Consensus : t a t a a a A g t g
(Plant) t T

if the sum of the relative frequencies of those two nucleotides exceeded 75%.

c) If no single nucleotide or pair of nucleotides satisfied these conditions, the position was denoted by the most frequent nucleotide in lower case.

RESULTS AND DISCUSSION

1. Domain I: Out of the three possible translational stop codons, the TGA and TAA were each present in 39% of the genes while TAG was under-represented. There was obviously no clear bias for any one

Table 4 : Nucleotide frequencies at 10 positions downstream from AATAAA like motif.

	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
A%	46	37	48	30	46	37	39	35	33	28
G%	13	37	15	24	13	13	17	26	15	17
C%	15	7	9	11	24	11	11	20	9	22
T%	26	20	28	35	17	39	33	20	43	33

Total number of observations : 46

Consensus : a $\begin{matrix} a \\ g \end{matrix}$ $\begin{matrix} A \\ T \end{matrix}$ t a $\begin{matrix} T \\ A \end{matrix}$ a a $\begin{matrix} T \\ A \end{matrix}$ t

(Plant)

of them. The downstream context of the stop codons was examined to see if any particular order of base occurrence could be deduced similar to the upstream context of translational start codon (58) where adenine was conserved at -3 position from ATG. As evident from Table 2, the first downstream position (+1) was occupied by a purine in 77% of the genes, while no consensus or co-consensus nucleotide could be assigned for position +2 onwards, except at +3 showing preference for thymine or adenine. However, all these positions had adenine or thymine as the high frequency nucleotide except at +12 position where guanine was preferred. The reason for this AT-rich tail following the stop codon is not apparent although the presence of less labile secondary structures, if formed, can be proposed (58). It can, therefore, be inferred that nucleotide sequences following stop codons (up to +12 position) show non-conserved bias except at +1 and +3 positions.

2. Domain II: The consensus sequence for ten nucleotides lying upstream of AATAAA, as decided on the basis of Table 3, was

$$t_{48}\ a_{48}\ t_{41}\ a_{43}\ a_{35}\ \begin{matrix} a_{28} \\ t \end{matrix}\ \begin{matrix} A_{50} \\ T37 \end{matrix}\ g_{33}\ t_{43}\ g_{37}.$$

This was remarkable because of the occurrence of adenine or thymine at most of the positions. The cause and consequence of such a bias remains to be elucidated. In addition, this region was also scanned to see if any particular sequence motif occurred commonly. Interestingly, a motif similar to CAYTG, namely YAYTG, was observed in 40 out of 46 genes (Table 1). This sequence was scored 75 times in an overlapp-

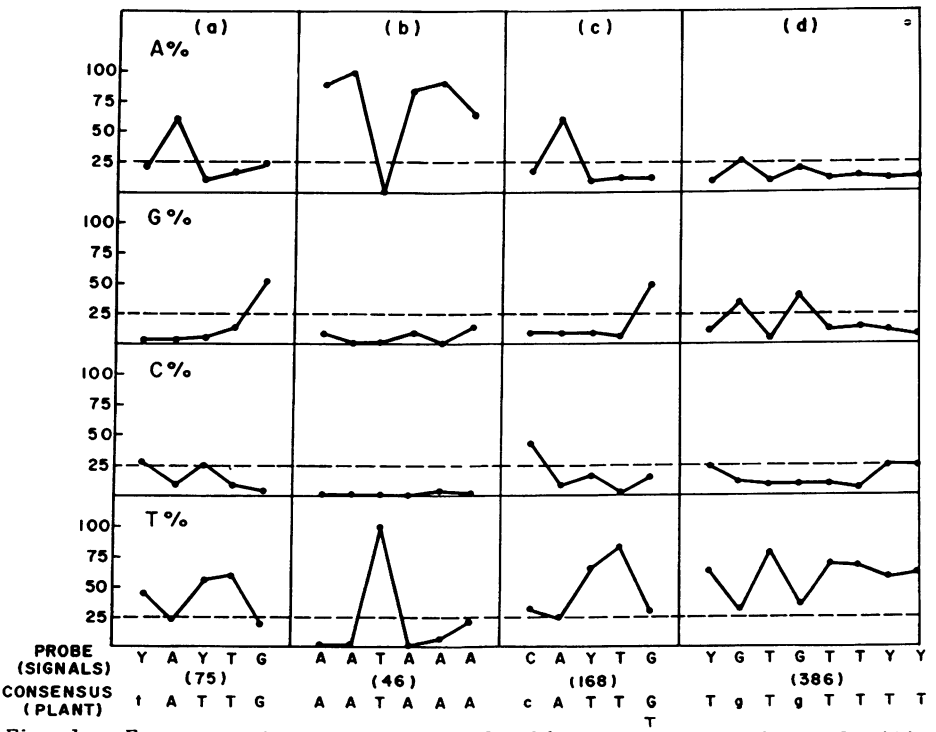


Fig. 1 : Frequency distribution of nucleotides at some putative poly (A) signals in higher plant genes based on data in Table 1.

ing manner and the consensus sequence being

tATTG.....(fig. 1 a).

On the basis of sequence analysis, Berget (55) had suggested the probable pairing between U4 snRNP and the CAYTG motif to be important for cleavage site selection prior to polyadenylation. The presence of a similar motif at an unexpected place such as upstream from AATAAA, raises doubts about the real significance of these signals being in the cleavage process. The common occurrence of this signal in 3' region as discussed below suggests that this motif could have some hitherto unrecognised function.

3. Domain III: The consensus AATAAA or its close derivatives are present most generally in the 3' regions of animal pre-mRNAs in a definite spatial relationship to the poly (A) site. Fig. 1b clearly indicates that the consensus for this signal in nuclear genes of higher plants is identical to that in animal pre-mRNAs. In fact, the consensus nucleo-

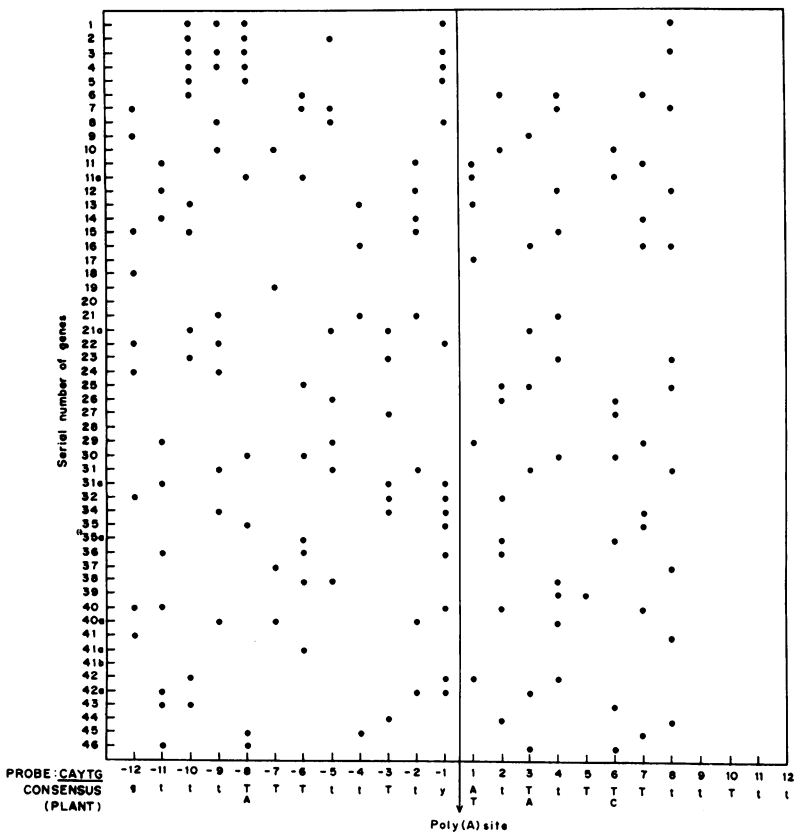


Fig. 2 : Distribution of CAYTG motif (marked by dots) around putative/ observed poly (A) sites present downstream of AATAAA motif. The suggested consensus sequence is mentioned below the figure. When more than two sites are located in the vicinity, it is regarded as one site by taking the central nucleotide as approximate site of polyadenylation.

tides were highly recurrent in the first five positions while adenine occurred at the 6th position in 65% of the genes. A critical examination of these signals, however, showed that the unaltered AATAAA motif was present in only 18 out of 46 genes (39%) whereas one base substitution occurred in 54% of the genes. Two base mismatches, a rare event, appeared in only 3 genes in the present compilation. When one or more base substitutions were observed to affect the polyadenylation efficiency in animal pre-mRNAs (45, 46), it was highly intriguing question how a majority of the plant genes ($\approx$ 61%) work efficiently (?) without an AATAAA motif. The two possible explanations are: a) plant genes

are less selective than animal genes for the obligatory presence of AATAAA, and b) there are other sequence motifs or nuclear factors which compensate for any point mutations in AATAAA. The location of AATAAA at 27 ± 9 bases from poly (A) site, however, appears to be similar in both plants and animals.

4. Domain IV: a) The region 10 bases downstream of AATAAA (Table 4) showed the following consensus

$$\underset{}{a_{46}}\ \underset{g}{a_{37}}\ \underset{T28}{A_{48}}\ t_{35}\ a_{46}\ \underset{A37}{T_{39}}\ a_{39}\ a_{35}\ \underset{A33}{T_{43}}\ t_{33}$$

which continued its less conserved adenine-thymine bias. The co-consensus nucleotides $\begin{matrix}A\\T\end{matrix}$ were present at +3, +6 and +9 positions. Whether this A-T richness has any functional significance or not needs to be investigated.

b) Context sequences of poly (A) sites: Twelve bases on either side, as shown in Fig. 2, delineate the thymine richness of the region encompassing the poly (A) sites. This span of 24 bases was also scanned for the CAYTG motif (as described under Materials and Methods) and a total of 168 putative signals were found around 53 poly (A) sites situated downstream of an AATAAA-like motif. These signals, however, were not exclusively distributed around poly (A) sites but were detected upstream (up to -12 position) and downstream (up to +8 position) of poly (A) sites. A general survey of domain IV has shown a common presence of CAYTG like sequence all over (data not shown). This raises the doubt if U4 RNPs are really involved in cleavage site selection and therefore, CAYTG were located in the vicinity of the poly (A) sites (55). Recently, Berget and Robberson (56) used cleaved U4 RNAs for *in vitro* studies to see if polyadenylation was affected. Contrary to the earlier suggestion (55), they did not find any involvement of U4 RNA in process of polyadenylation. However, the nearly ubiquitous occurrence of this or similar motifs in 3' untranslated region cannot be without significance even though this is not presently understood. As deduced from the data in Fig. 1c, this motif has cATT$\begin{matrix}G\\T\end{matrix}$ sequence as consensus. A word of caution is necessary before I conclude these observations. This paper deals solely with the structural features of certain domains in 3' untranslated regions of higher plant genes and has no bearing on actual functional analysis. The features that emerged from structural and functional analysis in animal and

viral pre-mRNAs were simply extrapolated to plants keeping in mind that they might not be identical.

c) YGTGTTYY motif: The presence of this signal was examined in the 50 bases (when available) downstream from AATAAA. In 46 genes, a total of 386 signals were recognised, as shown in Table 1, and were found to be distributed throughout domain IV. A signal can have some significance if it occurs in a particular relationship with a fixed reference point. For example, the YGTGTTYY motif was detected in 67 of the animal genes at about 30 bases downstream of the AATAAA motif (57). When attributing the recognition of signals by nuclear protein factors involved in polyadenylation, their absence at other unexpected positions should be the basic premise. Otherwise, several nuclear factors can be imagined to form complexes with several signals thus apparently requiring some additional factor(s) or signal(s) to decide which of them are in proper orientation with respect to poly(A) site selection. Considerable overlapping of YGTGTTYY - like signals and their clustered arrangement around the poly (A) site were also observed. The consensus sequence for this signal in plants was

TgTgTTTT.....Fig. 1d

which interestingly matches with the TTGTTT motif observed in the 3' region of some rbcs and cab genes in higher plants except at a few positions (61-63). The wide distribution of this signal throughout domain IV raises a doubt about its specific role in accurate cleavage site selection and efficient polyadenylation.

In conclusion, after surveying the available data from 46 plant genes, it appears that the features in the 3' untranslated regions of plant genes are distinctly different in the minute specific details from those of the animal genes. Even though similar motifs are present in both these sub-groups of eukaryotes, their distributional pattern show significant differences. This is expected for two kingdoms which diverged early in the evolutionary time scale. A detailed functional analysis would be highly desirable to clarify the structural and functional relationship between putative poly (A) signals and the process of polyadenylation in plants. Recently, Hunt et al. (64) had come to similar conclusions when they observed that tobacco cells could not properly and efficiently recognize the animal and viral poly (A) signals. Moreover within angiosperms, wheat rbcS genes representative of monocotyledons, and pea rbcS genes representative of dicotyledons, showed

deviant behaviour in relation to polyadenylation when expressed in tobacco cells (65). This supports the hypothesis that even though 3' processing might be a universal feature in eukaryotes, the mechanisms by which it operates might follow different specific paths in different eukaryotic species.

ACKNOWLEDGEMENTS

I wish to thank my colleagues at NCL, Dr. C. SivaRaman, Dr. A.F. Mascarenhas, Mr. H.B. Singh, Mr. G. Balu and Mrs. Rama Joshi for the help rendered during preparation of this manuscript. NCL communication number of this paper is 4280.

REFERENCES

1. Fluhr, R., Moses, P., Morelli, G., Coruzzi, G. and Chua, N-H. (1986) Embo J. 5, 2063-2071.
2. Mazur, B.J. and Chui, C.F. (1985) Nucleic Acids Res. 13, 2373-2386.
3. Poulsen, C., Fluhr, R., Kauffman, J.M., Boutry, M. and Chua, N-H. (1986) Mol. Gen. Genet. 205, 193-200.
4. Broglie, R., Coruzzi, G., Lampaa, G., Keith, B. and Chua, N-H. (1983) Biotechnology 1, 55-60.
5. Cashmore, A.R. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 2960-2964.
6. Stayton, M.M., Black, M., Bedbrook, J. and Dunsmuir, P. (1986) Nucleic Acids Res. 14, 9781-9786.
7. Rafalski, J.A., Scheets, K., Metzler, M., Peterson, D.M., Hedgeoth, C. and Soell, D.G. (1984) Embo J. 3, 1409-1415.
8. Sumner-smith, M., Rafalski, J.A., Sugiyama, T., Stoll, M. and Soell, D. (1985) Nucleic Acids Res. 13, 3905-3916.
9. Sugiyama, T., Rafalski, J.A. and Soell, D. (1986) Plant Sci. 44, 205-209.
10. Anderson, O.D., Litts, J.C., Gautier, M.F. and Greene, F.C. (1984) Nucleic Acids Res. 12, 8129-8144.
11. Thompson, R.D., Bartels, D. and Harberd, N.P. (1985) Nucleic Acids Res. 13, 6833-6846.
12. Forde, B.G., Heyworth, A., Pywell, J. and Kreis M. (1985) Nucleic Acids Res. 13, 7327-7339.
13. Boronat, A., Martinez, M.C., Rein, A.M., Puigdomenech, P. and Palau, J. (1986) Plant Sci. 47, 95-102.
14. Hu, N.T., Peifer, M.A., Heidecker, G., Messing, J. and Rubenstein, I. (1982) Embo J.1, 1337-1342.
15. Pedersen, K., Devereux, J., Wilson, D.R., Sheldon, E. and Larkins B.A. (1982) Cell 29, 1015-1026.
16. Slightom, J.L., Sun, S.M. and Hall, T.C. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 1897-1901.
17. Lycett, G.W., Croy, R.R.D., Shirsat, A.H. and Boulter, D. (1984) Nucleic Acids Res. 12, 4493-4516.
18. Baumlein, H., Wobus, U., Pustell, J. and Kafatos, F.C. (1986) Nucleic Acids Res. 14, 2707-2720.
19. Schuler, M.A., Schmitt, E.S. and Beachy, R.N. (1982) Nucleic

Acids Res. 10, 8225-8244.
20. Eckes, P., Rosahl, S., Schell, J. and Willmitzer, L. (1986) Mol. Gen. Genet, 205, 14-22.
21. Hyldig-Nielsen, J.J., Jensen, E.O., Paludan, K., Wiborg, O., Garrett, R., Jorgensen, P. and Marcker, K.A. (1982) Nucleic Acids Res. 10, 689-701.
22. Wiborg, O., Jorgen, J., Hyldig - Nielsen, J.J., Jensen, E.O., Paludan, K. and Marcker, K.A. (1982) Nucleic Acids Res. 10, 3487-3502.
23. Landsmann, J., Dennis, E.S., Higgins, T.J.V., Appleby, C.A., Kortt, A.A. and Peacock, W.J. (1986) Nature 324, 166-168.
24. Mauro, V.P., Nguyen, T., Katinakis, P. and Verma, D.P.S. (1985) Nucleic Acids Res. 13, 239-249.
25. Katinakis, P. and Verma, D.P.S., (1985) Proc. Natl. Acad. Sci. U.S.A. 82, 4157-4161.
26. Vodkin, L.O., Rhodes, P.R. and Goldberg, R.B. (1983) Cell 34, 1023-1031.
27. Hoffman, L.M. and Donaldson, D.B. (1985) Embo J. 4, 883-889.
28. Hoffman, L.M. (1984) J. Mol. Appl. genet. 2, 447-453.
29. Halling, K., Halling, A.C., Murray, E.E., Ladin, B.F., Houston, L.L. and Weaver, R.F. (1985) Nucleic Acids Res. 13, 8019-8033.
30. Chang, C. and Meyerowitz, E.M., (1986) Proc. Natl. Acad. Sci. U.S.A. 1408-1412.
31. Shah, D.M., Hironaka, C.M., Wiegand, R.C., Harding, E.I., Krivi, G.G. and Tieweier, D.C. (1986) Plant Mol. Biol. 6, 203-211.
32. Boutry, M. and Chua, N-H. (1985) Embo J. 4, 2159-2165.
33. Marchionni, M. and Gilbert, W. (1986) Cell 46, 133-141.
34. Werr, W., Frommer, W.B., Maas, C. and Starlinger, P. (1985) Embo J. 4, 1373-1379.
35. Klossgen, R.B., Gierl, A., Schwarz-Sommer, Z. and Saedler, H. (1986) Mol. Gen. Genet. 203, 237-244.
36. Sommer, H. and Saedler, H. (1986) Mol. Gen. Genet. 202, 429-434.
37. Tischer, E., DasSarma, S. and Goodman, H.M. (1986) Mol. Gen. Genet. 203, 221-229.
38. Keil, M., Sanchez-serrana, J., Schell, J. and Willmitzer, L. (1986) Nucleic Acids Res. 14, 5641-5650.
39. Czarnecka, E., Gurley, W.B., Nagao, R.T., Mosquera, L.A. and Key, J.L. (1985) Proc. Natl. Acad. Sci. U.S.A. 82, 3726-3730.
40. Chen. J. and Varner, J.E. (1985) Embo J. 4, 2145-2151.
41. Birnstiel, M.L., Busslinger, M. and Strub, K. (1985) Cell 41, 349-359.
42. Proudfoot, N.J. (1984) Nature 307, 412-413.
43. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.
44. Fitzgerald, M. and Shenk, T. (1981) Cell 24, 251-260.
45. Higgs, D.R., Goodbourn, S.E.Y., Lamb, J., Clegg, J.B., Weatherall, D.J. and Proudfoot, N.J. (1983) Nature 306, 398-400.
46. Wickens, M. and Stephenson, P. (1984) Science 226, 1045-1051.
47. Hashimoto, C. and Steitz, J.A. (1986) Cell 45, 581-591.
48. Zarkower, D. and Wickens, M. (1987) Embo J. 6, 177-186.
49. Kessler, M.M., Beckendorf, R.C., Westhafer, M.A. and Nordstorm, J.L. (1986) Nucleic Acids Res. 14. 4939-4952.
50. Conway, L. and Wickens, M. (1985) Proc. Natl. Acad. Sci. U.S.A. 82, 3949-3953.
51. Gil, A. and Proudfoot, N.J. (1983) Nature 312, 473-474.
52. McDevitt, M.A., Imperiale, M.J., Ali, H., Nevins, J.R., (1984) Cell 37, 993-999.

53. Woychik, R.P., Lyons, R.H., Post, L. and Rottman, F.M. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 3944-3948.
54. Mason, P.J., Elkington, J.A., Lloyd, M.M., Jones, M.B. and Williams, J.G. (1986) Cell 46, 263-270
55. Berget, S.M. (1984) Nature 309, 179-182.
56. Berget, S.M. and Robberson, B.L. (1986) Cell 46, 691-696.
57. McLauchlan, J., Gaffney, D., Whitton, J.L. and Clements, J.B. (1985) Nucleic Acids Res. 13, 1347-1368.
58. Joshi, C.P. (1987) Nucleic Acids Res. 15, 6643-6653.
59. Brown, J.W.S. (1986) Nucleic Acids Res. 14, 9549-9559.
60. Cavener, D.R. (1987) Nucleic Acids Res. 15, 1353-1361.
61. Dean, C., Van den Elzen, P., Tamaki, S., Dunsmuir, P. and Bedbrook, J. (1985) Embo J. 4, 3055-3061.
62. Dunsmuir, P. (1985) Nucleic Acids Res. 13, 2503-2518.
63. Leutwiler, L.S., Meyerowitz, E.M. and Tobin, E.M. (1986) Nucleic Acids Res. 14, 4051-4064.
64. Hunt, A.G., Chu, N.M., Odell, J.T., Nagy, F. and Chua, N-H (1987) Plant Mol. Biol. 8, 23-35.
65. Keith, B. and Chua, N-H. (1986) Embo J. 5, 2419-2425.