Plant Molecular Biology **19**: 1057–1064, 1992.

Update section

Short Communication

Information contents and dinucleotide compositions of plant intron sequences vary with evolutionary origin

Owen White, Carol Soderlund [1], Pari Shanmugan and Chris Fields*
*Computing Research Laboratory, Box 30001/3CRL, New Mexico State University, Las Cruces, NM 88003–0001, USA (*author for correspondence); [1] Present address: Theoretical Biology and Biophysics (T-10), Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

Received 11 November 1991; accepted in revised form 10 April 1992

Key words: computation, dinucleotide, information, intron, splice site

Abstract

The DNA sequence composition of 526 dicot and 345 monocot intron sequences have been characterized using computational methods. Splice site information content and bulk intron and exon dinucleotide composition were determined. Positions 4 and 5 of 5′ splice sites contain different statistically significant levels of information in the two groups. Basal levels of information in introns are higher in dicots than in monocots. Two dinucleotide groups, WW (AA, AU, UA, UU) and SS (CC, CG, GC, GG) have significantly different frequencies in exons and introns of the two plant groups. These results suggest that the mechanisms of splice-site recognition and binding may differ between dicot and monocot plants.

The mechanism of pre-mRNA splicing is now understood in considerable detail [1, 2, 3]; however, the molecular recognition of splice sites by spliceosome components is still not well characterized. Goodall and Filipowicz have identified a bias for A and U nucleotides in plant introns, and hypothesize that a high A + U content and consensus intron-exon borders may be the only sequence requirements for plant intron processing [4]. A minimal functional length of 70–73 nucleotides has been observed for introns in monocots and dicots [5], despite a heterogeneous distribution of intron lengths between the two plant groups. Total genome dinucleotide frequencies have been measured experimentally in a variety of organisms, including plants [6, 7]. Unique distributions of dinucleotides have been observed in DNA separated by evolutionary origin [8, 9, 10], organelle [11], and function [12].

The unique ability of monocots to process introns high in G + C suggests differences in dicot and monocot pre-mRNA processing mechanisms [13]. For example, pre-mRNA of 1,5-bisphosphate carboxylase from pea, a dicot, was efficiently spliced in transgenic tobacco plants but the same gene from wheat, a monocot, was not processed as efficiently in transgenic tobacco [14]. While experimental differences in dicot-monocot splicing specificities exist, both dicot and monocot genes are spliced with similar efficiency *in vitro* by HeLa cell extracts [15], and by an autonomously replicated vector in transient expression assays of tobacco leaf disks [16]. Stable incorporation of a phaseolin gene fused to the cauliflower mosaic virus promoter resulted in equally efficient pre-mRNA processing in tobacco and rice cell lines [17].

Experimental evidence suggests that plant introns may have pre-mRNA recognition mechanisms that are different from those of vertebrate systems [13 and references therein]. Calculations of the information content of binding sites have provided some insight into site recognition [18, 19]. Information content analysis of intron splice sites in the nematode *Caenorhabditis elegans* has been previously described [20]; this analysis showed that splice sites in *C. elegans* vary with intron length. In this report we show that: (1) the information contents of splice sites differ between monocots and dicots; (2) the basal level of information is an average of 0.1 bit/base higher in introns than in exons for dicots; (3) dinucleotide usage differs between exons and introns, and between the two plant groups.

Plant DNA sequences were extracted from GenBank release 68 (June 1991). Entries containing full- or partial-length sequences from 172 loci in dicots (Magnoliopsida) and 84 loci in monocots (Liliopsida) were used in our analysis. The coding portions of these sequences encode enzymes, structural proteins, storage proteins and peptides of unknown function. Sequences were discarded in cases where intron/exon splice junctions were ambiguously determined. Overrepresenting of certain sequences due to oversampling of some gene families cannot be excluded. Sequences from 20 nucleotides upstream to 30 nucleotides downstream of the 5′ splice site were aligned, and information content, $I(n)$ in units of bits/position, was calculated as:

$$I(n) = \left(\sum_{B = \mathrm{A, C, G, U}} F(B, n) \log_2 F(B, n) \right) - \left(\sum_{B = \mathrm{A, C, G, U}} P(B, n) \log_2 P(B, n) \right)$$

where $F(B, n)$ is the observed frequency of base B in position n, and $P(B, n)$ is the prior probability of base B in position n [18]. The expression $I(n)$ represents the information contained in a single nucleotide position, as the result of elevated usage of a particular nucleotide or nucleotides at that position in a DNA sequence. The prior base probability, as measured by base frequency, has been shown to vary between introns and exons, and between dicots and monocots [4]. To visualize the differences in base composition across the exon-intron boundaries in calculations of $I(n)$, the prior base probabilities were set to the equiprobable values (i.e., $P(B, n) = 0.25$ for B = A, C, G, U in all positions). Similar information contents from sequences 30 nucleotides upstream to 20 nucleotides downstream were obtained from 3′ splice site junctions. Standard deviations were calculated using the 'exact method' [18]; error bars for information plots consist of two times the standard deviation. For basal information measurements, the 0.5% and 99.5% confidence limits were determined by numerical simulations, based on observed average nucleotide frequencies.

A dinucleotide (BB') is any two adjacent nucleotide bases. Dinucleotide frequencies were measured from both exons and introns. Two components of DNA composition are reflected in simple dinucleotide frequency measurements. One component of dinucleotide frequencies is that part strictly due to the underlying distribution of single nucleotide frequencies; for example, increases in A and U would increase the random occurrence of AA, AU, UA and UU. The second component of dinucleotide frequency is that part which reflects the correlation between the two nucleotides. To reduce the component of dinucleotide frequency that merely reflects single nucleotide distributions, the observed frequency of each dinucleotide was divided by the expected frequency of the dinucleotide, using the formula:

$$F'(BB') = \frac{F(BB')}{F(B) \times F(B')}$$

where $F(BB')$ is the observed dinucleotide frequency and $F(B)$ and $F(B')$ are the observed single nucleotide frequencies for two single nucleotides. The logarithm of $F'(BB')$ is referred to as the 'mutual information' of B and B' [21]. Expected dinucleotide frequencies for introns were calculated using measured single nucleotide frequencies from intron regions, and expected exon dinucleotide frequencies were calculated using measured single nucleotide frequencies from exon regions. Log-likelihood ratios, demonstrating differences in non-independent dinucleotide usage between dicots and monocots, were derived by the formula:

$$R\,(BB') = \log_2 \left(\frac{F'\,(BB')_{\mathrm{di}}}{F'\,(BB')_{\mathrm{mono}}} \right)$$

where $R(BB')$ is the log-likehood ratio of the dinucleotide combination, and $F'(BB')_{\mathrm{di}}$ and $F'(BB')_{\mathrm{mono}}$ are the corrected dinucleotide frequencies for dicots and monocots, respectively. Our software to generate information contents, base frequency matrices, baseline frequency simulations, dinucleotide counts, corrected dinucleotide frequencies, and log-likehood ratios is available on request.

Information contents of the splice junctions of both dicots and monocots are shown in Fig. 1. The observed basal level of information across the intronic portion of dicot sequences differs from the exon basal level of information. This asymmetric distribution of basal information is not apparent in monocot sequences. We tested whether the elevation of basal information is due solely to an enrichment of A + U in dicot introns.

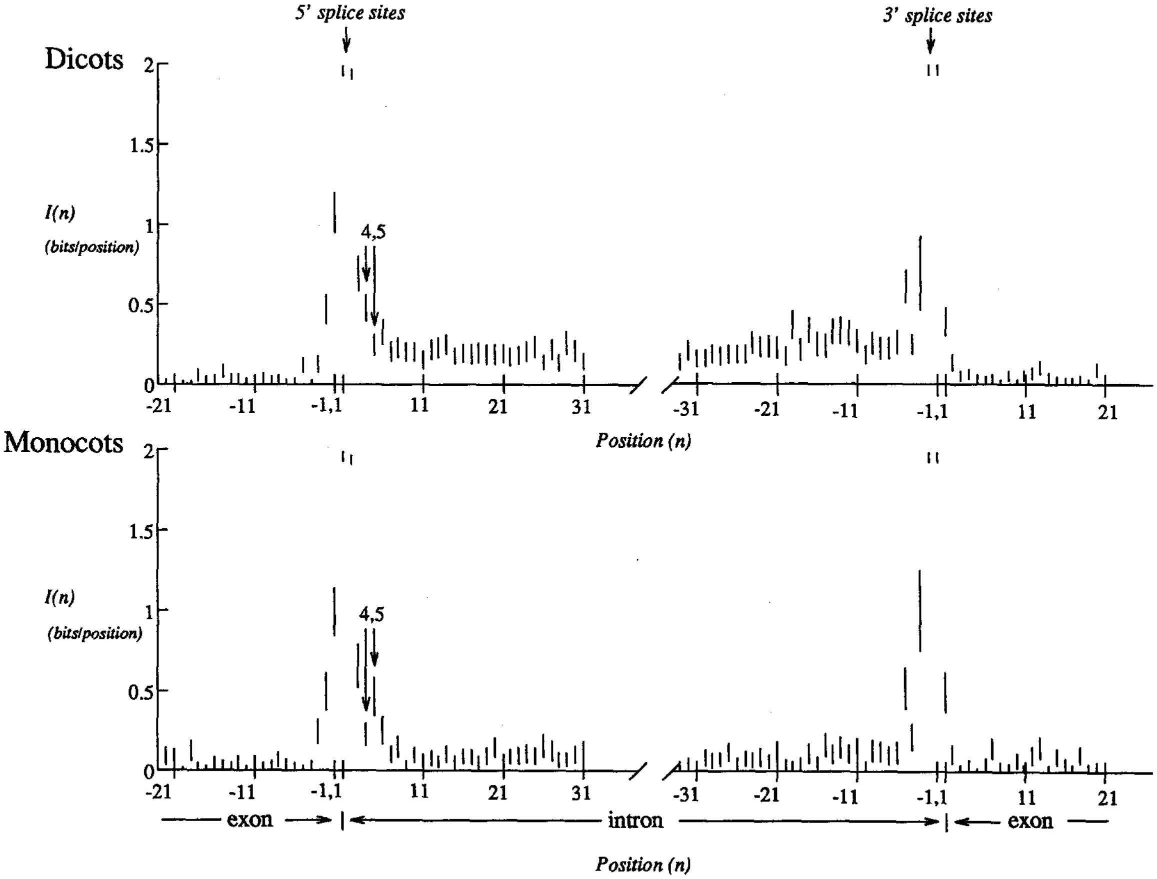

Fig. 1. Single-nucleotide information encoded around 5′ and 3′ splice sites of dicot and monocot introns. Error bars are of ± two times the standard deviation (within 95% confidence limits). Positions 4 and 5 of 5′ splice sites are marked. The error bars are exaggerated in size for visibility on positions + 1 and + 2 for 5′ splice sites and positions −1 and −2 for 3′ splice sites.

Numerical experiments examining the range of possible baseline information values that can occur from a given average frequency of nucleotides demonstrated that baseline information is highly dependent on subtle changes in nucleotide frequency. Using the average nucleotide frequencies measured between positions 10 and 50 in monocot introns, which are 60% A + U, baseline information values are expected to range from 0.047 to 0.064, within 99.5% confidence limits. Nucleotide frequencies from the same portion of dicot introns (72% A + U), gave expected baseline information values of 0.155 to 0.180, within 99.5% confidence limits. Thus the observed A + U frequencies alone are sufficient to account for the observed basal information contents in introns of both dicots and monocots.

The information contents of dicot and monocot splice sites are similar, but not identical. A shoulder of the 5′ peak extending into the coding portion of the splice site (positions −2 and −1) contains 1.53 ± 0.15 and 1.48 ± 0.20 bits of information in dicots and monocots, respectively. Not including the conserved GU, the intronic portion of the splice site (positions + 3 to + 6) encodes 1.73 ± 0.18 and 1.59 ± 0.21 bits of information in dicots and monocots, respectively. In total (positions −3 to + 6) 7.26 ± 0.24 bits are observed in dicot 5′ splice sites, while 7.07 ± 0.29 bits are contained in monocot 5′ splice sites. Statistically significant differences in information content between the two plant groups can be observed at specific nucleotide positions near the 5′ splice site. At position 4 of the 5′ splices site, 0.25 ± 0.11 more bits of information are found in dicots than in monocots, while at position 5, 0.21 ± 0.14 more bits are encoded in monocots than in dicots. Nucleotide counts at positions −3 to + 8 of the 5′ splice sites are presented in Table 1. The consensus sequence is AG|GUAAGU for both monocot and dicot 5′ splice sites.

Similar differences in information contents are found in dicot and monocot 3′ splice sites. Position + 1, the exonic portion of the 3′ splice site, contains 0.39 ± 0.09 bits of information in dicots, and 0.50 ± 0.12 bits in monocots. Significant information is encoded at positions −3 and −5 of plant 3′ splice sites (1.30 ± 0.26 and 1.53 ± 0.28 bits for dicots and monocots, respectively). This is primarily due to an enrichment of pyrimidine in position −3 and U in position −5. Monocot 3′ splice sites contain fewer U at −3 than dicots, resulting in consensus sequences of UGYAG|GG for dicots and UGCAG|GG for monocots. In total, from positions −5 to + 1, 5.93 ± 0.28 bits of information are found in dicot 3′ splice sites, and 6.24 ± 0.32 bits of information are contained in monocot 3′ splices sites. Discrete base frequency differences exist between the two plant groups in their 3′ splice sites. Dicot splice sites use C in lowest abundance in positions −8, −7, −6, −5, −4, + 2 and + 3 around the splice site (the conserved AG at −1 and −2 were not considered). The nucleotide C occurs least frequently in only position −4 in monocots (A is as infrequent as C at position −6).

Lariat branch point sequences are in greatest abundance in the window between −50 and −1 of the 3′ region in invertebrate, primate, plant and rodent introns [22]. Local information maxima are observed in dicot introns at positions −17 and −19 with respect to the 3′ splice site; these do not appear in monocot sequences. These maxima are due to an increase of U at positions −17 and −19. We did not detect a significant association of the intron branch consensus URAY [23] with Us at these positions (data not shown). In dicots, 455 strict matches to the branch site URAY, out of a possible 526 sequences (86%), were detected in the −50 to −1 3′ portion of introns. In the same region of monocot introns, 260 potential branch sites were detected in a total of 345 sequences (75%).

Raw dinucleotide frequencies and mutual information values are presented in Fig. 2 for exons and introns of both plant groups. The overall frequencies of the nucleotides A + U in introns are 71% and 61% for dicots and monocots, respectively. In exons, A + U occur at 55% and 42% in dicots and monocots, respectively. These values are in close agreement with those of Goodall and Filipowicz [4], as measured in a smaller data set. In the introns of both plant groups, SS dinucleotides (CC, CG, GC, GG) occur with the lowest

Table 1. Base number matrices for 5′ and 3′ splice sites. Base frequencies are given in parenthesis.

5′ splice sites

Dicot introns, 5′ position

	−3	−2	−1	1	2	3	4	5	6	7	8
A	196(37)	321(61)	34(07)	0(0)	0(0)	370(70)	299(56)	116(22)	117(22)	200(37)	178(33)
C	170(32)	49(09)	15(02)	0(0)	2(0)	31(05)	75(14)	51(09)	68(13)	86(17)	98(18)
G	94(18)	47(08)	434(82)	528(100)	0(0)	44(08)	17(03)	261(49)	55(10)	47(08)	37(07)
U	68(12)	111(21)	45(08)	0(0)	526(100)	83(15)	137(26)	100(19)	288(54)	195(37)	215(40)

Monocot introns, 5′ position

	−3	−2	−1	1	2	3	4	5	6	7	8
A	147(43)	217(63)	18(05)	0(0)	0(0)	238(68)	152(45)	67(19)	65(18)	122(35)	81(24)
C	113(32)	54(15)	33(09)	0(0)	2(0)	28(08)	87(24)	38(10)	78(22)	47(14)	85(25)
G	62(18)	25(07)	278(80)	346(100)	0(0)	57(16)	23(07)	211(60)	31(09)	68(19)	37(11)
U	24(06)	50(14)	17(05)	0(0)	344(100)	23(07)	84(25)	30(08)	172(50)	109(31)	143(40)

3′ splice sites

Dicot introns, 3′ position

	−8	−7	−6	−5	−4	−3	−2	−1	1	2	3
A	140(26)	109(20)	121(23)	85(16)	163(31)	25(05)	526(100)	0(0)	101(19)	108(20)	156(30)
C	57(11)	63(12)	54(10)	28(05)	30(06)	314(60)	0(0)	0(0)	71(13)	87(16)	79(15)
G	82(16)	93(18)	86(16)	60(11)	223(42)	2(0)	0(0)	526(100)	306(58)	100(19)	145(27)
U	247(47)	261(50)	265(50)	353(67)	110(21)	185(35)	0(0)	0(0)	48(09)	231(44)	146(28)

Monocot introns, 3′ position

	−8	−7	−6	−5	−4	−3	−2	−1	1	2	3
A	62(18)	57(16)	63(18)	37(10)	55(15)	11(03)	345(100)	0(0)	50(14)	55(16)	69(20)
C	64(18)	71(21)	63(18)	53(15)	42(12)	281(81)	0(0)	0(0)	43(12)	71(20)	93(27)
G	70(20)	77(22)	71(20)	32(09)	167(48)	4(01)	0(0)	345(100)	220(64)	76(22)	85(24)
U	149(43)	140(40)	148(44)	223(64)	81(23)	49(15)	0(0)	0(0)	22(09)	143(42)	98(28)

frequency. The WW dinucleotides (AA, AU, UA, UU) are the 4 most frequent in dicot introns, while they are 4 of the 5 most frequent in monocot introns (UG is more common than AA or UA). Dinucleotide abundances are reversed in monocot exons, where WW dinucleotides become the 4 least frequent dinucleotides, while SS dinucleotides are 4 out of the 5 most common dinucleotides (CA is more common than CG). This abundance reversal does not occur in dicot exons, where SS dinucleotides remain as 4 of the 6 least common dinucleotides. Quite opposite from monocot exons, AA is the most abundant dinucleotide in dicot exons. The mutual information values (lower panel, Fig. 2) show that despite the overall consistency between dinucleotide frequencies of their component single nucleotides frequencies, some dinucleotide levels are lower than expected. The dinucleotide CG, which is a potential methylation site and is rare in vertebrate genomes [24], is under-represented in introns and exons of both plant groups. The dinucleotide UA also occurs much less frequently than expected in dicot and monocot exon sequences, perhaps because in-frame UAs either encode stops or tyrosine, which is a relatively rare amino acid. Other dinucleotides, such as CA and UG, occur more frequently than expected considering their component single nucleotide frequencies. Mutual information for GC dinucleotides in monocot introns is higher than in exons and could serve as a possible distinguishing sequence feature for intron recognition.

Log-likelihood ratios for dinucleotides between exons and introns of the two plant groups are shown in Fig. 3. The largest differences are in the frequencies of the SS dinucleotides, with CC and GG preferred in dicots and CG and GC preferred

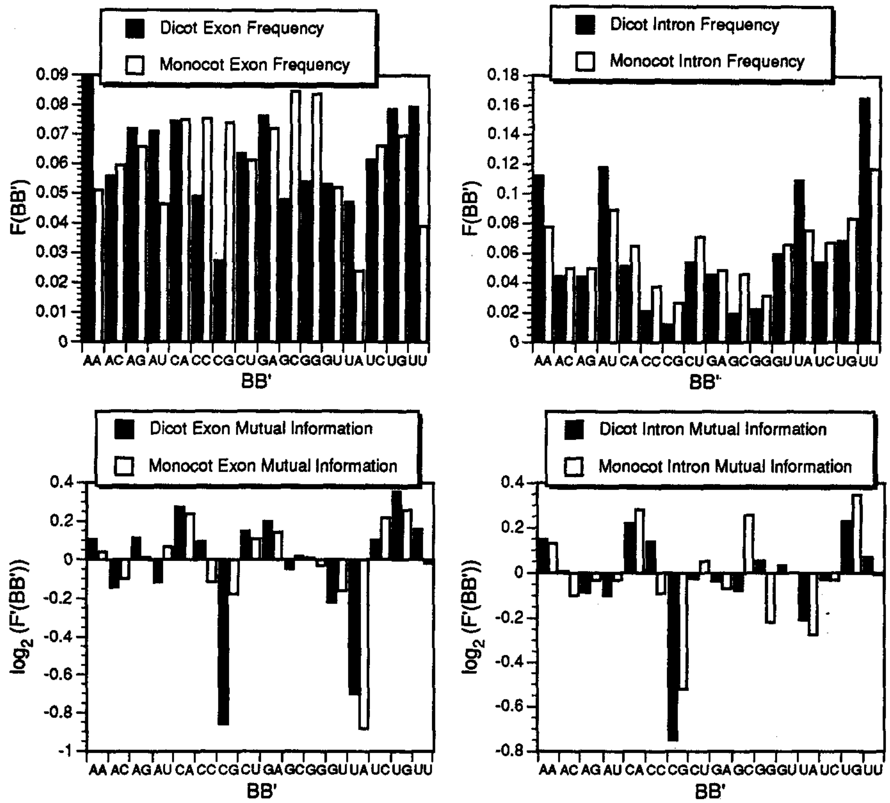

Fig. 2. Upper panel: raw dinucleotide frequency histograms for exons (left plot) and introns (right plot) from dicots and monocots. Lower panel: mutual information histograms for exons and introns. Corresponding dinucleotides are indicated below each bar. Dicot sequences contain 55% and 71% A + U in exons and introns, respectively. Monocot sequences contain 42% and 61% A + U in exons and introns, respectively. (Note the compressed y axis on the intron frequency histogram.) A mutual information value of 0.0 indicates that the observed dinucleotide frequency exactly equals the dinucleotide frequency expected from the observed single-nucleotide frequencies. Mutual information values greater than zero indicate a greater than expected dinucleotide frequency; values less than zero indicate a less than expected dinucleotide frequency.

in monocots. The SS and WW dinucleotides in exons, particularly CC, CG, UA and UU, exhibit the greatest abundance of differences of all the dinucleotides. The SS dinucleotides in introns demonstrate the largest extremes in abundance.

The differences in splice site structure and intron composition between dicots and monocots are similar to differences found between animal species. The SS and WW dinucleotides have different frequencies in exons and introns in many genomes [9, 25]. Intron sequences in *C. elegans* have high A + U content and elevated basal information content [20] similar to that found in dicots, while human intron sequences have basal information contents lower than that of monocots (data not shown). The information encoded in 5′ splice sites varies widely between species, primarily due to differences in the information encoded at positions + 4 and + 5 (the AG of the universal G|GUAAGU consensus). Position + 5 encodes 0.24 ± 0.07 bits in dicots, 0.45 ± 0.12 bits in monocots, 0.77 ± 0.22 bits in *C. elegans* [20], 1.08 ± 0.19 bits in *Drosophila* [26], and 1.15 ± 0.08 bits in primates (calculated from Table 1 in [27]). No significant differences between either dicot or monocot introns of different lengths, as reported for *C. elegans* introns [20], were observed.

The dinucleotide composition and information content differences reported here raise the possibility that alternate mechanisms for splice-site re-

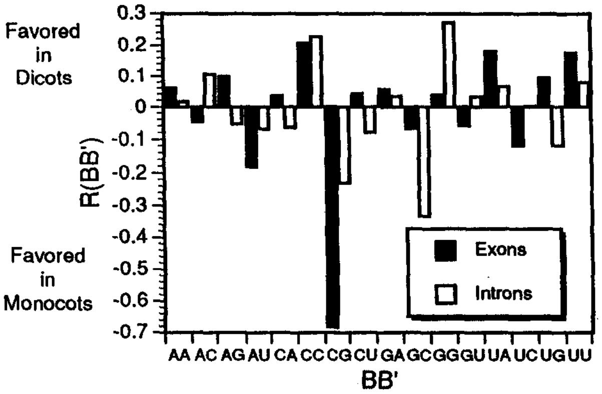

Fig. 3. Dinucleotide frequency log-likelihood ratios of dicots to monocots for exons and introns. Corresponding dinucleotides are indicated below each data bar. Values that are more positive indicate higher frequencies in dicots, while values that are more negative indicate higher frequencies in monocots. Many dinucleotides in the SS and WW dinucleotide groups exhibit large asymmetries between exons and introns.

cognition or pre-mRNA processing may be used by dicots and monocots. The sequence variation observed at the exon-intron junctions suggests that U1 snRNA recognition may not always occur at the same nucleotide positions in the 5′ splice site [28], or that additional undetermined factors may be involved in the recognition of either 5′ or 3′ splice sites. The compositional differences between monocots and dicots may also reflect the use of different mechanisms for exon or intron recognition between the two plant groups. A combination of specific splice-site recognition by snRNPs and recognition of exons by bulk composition is suggested, for example, by the exon-definition model of Berget *et al.* [29, 30]. A site- and composition-sensitive mechanism along these lines may be active in plants.

It has been shown that high mobility group (HMG) proteins bind to A + U-rich regions outside of plant genes [31]. Because of their presence in actively transcribed genes, HMGs have been implicated in transcriptional activation, perhaps by a mechanism involving conformational changes in chromatin [32]. In view of the elevated WW dinucleotide content in dicot introns, we suggest that HMG binding may not be exclusively confined to the flanking regions of genes, but may occur in intron sequences as well. Consistent with this possibility, HMGs have been demonstrated to bind to an intron portion of the N-20 gene in soybean [33]. Whether compositionally directed binding of proteins to bulk intron sequences plays any role in splice-site selection is unknown.

Acknowledgements

This work was supported in part by U.S. Department of Energy Genome Program Grant number 89ER60865 to C.F. and C.S. We also would like to thank Virginia Walbot for valuable discussion and Ted Dunning for technical assistance with the statistical modeling.

References

1. Green MR: Pre-mRNA splicing. Annu Rev Genet 20: 671–708 (1986).
2. Maniatis T, Reed R: The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. Nature 325: 673–678 (1987).
3. Sharp PA: Splicing of messenger RNA precursors. Science 235: 766–771 (1987).
4. Goodall GJ, Filipowicz W: The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. Cell 58: 473–483 (1989).
5. Goodall GJ, Filipowicz W: The minimum functional length of pre-mRNA introns in monocots and dicots. Plant Mol Biol 14: 727–733 (1990).
6. Josse J, Kaiser AD, Kornberg A: Enzymatic synthesis of deoxyribonucleic acid. J Biol Chem 236: 864–875 (1961).
7. Swartz MN, Trautner TA, Kornberg A: Enzymatic synthesis of deoxyribonucleic acid. J Biol Chem 237: 1961–1967 (1962).
8. Ohno S: Universal rule for coding sequence construction: TA/CG deficiency-TG/CT excess. Proc Natl Acad Sci USA 85: 9630–9634 (1988).
9. Kozhukhin CG, Pevzner PA: Genome inhomogeneity is determined mainly by WW and SS dinucleotides. CABIOS 7: 39–49 (1991).
10. Nussinov R: Doublet frequencies in evolutionary distinct groups. Nucl Acids Res 12: 1748–1763 (1984).
11. Boudraa M, Perrin P: CpG and TpA frequencies in the plant system. Nucl Acids Res 15: 5729–5737 (1987).
12. Bentler E, Gelbart T, Han J, Koziol JA, Beutler B: Evolution of the genome and the genetic code: selection at the dinucleotide level by methylation and polyribonucleotide cleavage. Proc Natl Acad Sci USA 86: 192–196 (1989).

13. Goodall GJ, Filipowicz W: Different effects of intron nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants. EMBO J 10: 2635–2644 (1991).
14. Keith B, Chua N: Monocot and dicot pre-mRNAs are processed with different efficiencies in transgenic tobacco. EMBO J 5: 2419–2425 (1986).
15. Brown JWS, Feix G, Frendewey D: Accurate *in vitro* splicing of two pre-mRNA plant introns in a HeLa cell nuclear extract. EMBO J 5: 2749–2758 (1986).
16. McCullough AJ, Lou H, Schuler MA: *In vivo* analysis of plant pre-mRNA splicing using an autonomously replicating vector. Nucl Acids Res 19: 3001–3009 (1991).
17. Peterhans A, Datta SK, Datta K, Goodall GJ, Potrykus I, Paszkowski K: Recognition efficiency of *Dicotyledoneae*-specific promoter and RNA processing signals in rice. Mol Gen Genet 222: 361–368 (1990).
18. Schneider TD, Stormo GD, Gold L, Ehrenfeucht A: Information content of binding sites on nucleotide sequences. J Mol Biol 188: 415–431 (1986).
19. Berg O, von Hippel P: Selection of DNA binding sites by regulatory proteins. J Mol Biol 193: 723–750 (1987).
20. Fields C: Information content of *Caenorhabditis elegans* splice site sequences varies with intron length. Nucl Acids Res 18: 1509–1512 (1990).
21. Hamming RW: Coding and Information Theory. Prentice Hall, New York (1980).
22. Harris NL, Senapathy P: Distribution and consensus of branch point signals in eukaryotic genes: a computerized statistical analysis. Nucl Acids Res 18: 3015–3019 (1990).
23. Brown JWS: A catalogue of splice junctions and putative branch point sequences from plant introns. Nucl Acids Res 14: 9549–9559 (1986).
24. Aissani B, Bernardi G: CpG islands: features and distribution in the genomes of vertebrates. Gene 106: 173–183 (1991).
25. Fields C, Soderlund CA: gm: a practical tool for automating DNA sequence analysis. CABIOS 6: 263–270 (1990).
26. Mount SM, Burks C, Hertz G, Stormo G, White O, Fields C: Splicing signals in *Drosophila*: intron size, information content, and consensus sequences. Manuscript in preparation.
27. Senapathy P, Shaprio MB, Harris NL: Splice junctions, branch point sites, and exons: sequence statistics, identification and applications to genome project. Meth Enzymol 183: 252–278 (1990).
28. Jacob M, Gallinaro H: The 5′ splice site: phylogenetic evolution and variable geometry with U1RNA. Nucl Acids Res 17: 2159–2180 (1989).
29. Robberson BL, Cote GJ, Berget SM: Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol Cell Biol 10: 84–94 (1990).
30. Talerico M, Berget SM: Effect of 5′ splice site mutations on splicing of the preceding intron. Mol Cell Biol 10: 6299–6305 (1990).
31. Pedersen TJ, Arwood LJ, Spiker S, Guiltinan MJ, Thompson WF: High mobility group chromosomal proteins bind to AT-rich tracts flanking plant genes. Plant Mol Biol 16: 95–104 (1991).
32. Spiker S: Histone variants and high mobility group non-histone chromosomal proteins of higher plants: their potential for forming a chromatin structure that is either poised for transcription or transcriptionally inert. Physiol Plant 75: 200–213 (1988).
33. Gambliel H, Feder I, Sengupta-Gopalan C: dAdT binding domains of soybean nodulin-20 and french bean β-phaseolin and phytohemagglutinin-L (Lec 2) genes are assembly sites of complex nucleoprotein structures *in vitro*. Plant Cell (submitted).