Mol Gen Genet (1990) 223:65–75

MGG

Homologous domains of the largest subunit of eucaryotic RNA polymerase II are conserved in plants

Christiane Nawrath [1], Jeff Schell [1], and Csaba Koncz [1,2]

[1] Max-Planck-Institut für Züchtungsforschung, D-5000 Köln 30, Carl-von-Linne-Weg 10, Federal Republic of Germany
[2] Institute of Plant Physiology, Biological Research Center of Hungarian Academy of Sciences, H-6701 Szeged, P.O. Box 521, Hungary

Received April 2, 1990

Summary. Genomic and cDNA clones homologous to the *RpII215* gene of *Drosophila* were isolated from *Arabidopsis thaliana* and assigned to a single copy gene encoding a transcript of 6.8 kb. Nucleotide sequence analysis of *Arabidopsis* genomic and cDNAs revealed a striking homology to yeast, *Caenorhabditis*, *Drosophila* and mouse genes encoding the largest subunit of RNA polymerase II. The *Arabidopsis* gene *rp*II215 contains 13 introns, 12 of which interrupt the coding sequence of a protein of 205 kDa. The position of the first intron is conserved between plant and animal genes, while an intron located in the 3′ untranslated region of the *rp*II215 gene is unique to *Arabidopsis*. Common domains present in all known largest subunits of eucaryotic RNA polymerase II were identified in the predicted sequence of the *Arabidopsis Rp*II215 protein. Both the order and the position of N-terminal Zn^{2+} finger and of DNA and α-amanitin binding motifs are conserved in *Arabidopsis*. The C-terminal region of the *Arabidopsis* protein contains 15 consensus and 26 variant YSPTSPS repeats (CTDs). Highly conserved structure among the various C-terminal domains suggests that the largest subunit of RNA polymerase II in plants may also interact with transcription factors and with protein kinases that control the cell cycle as in other organisms.

Key words: Largest subunit of *Arabidopsis* RNA polymerase II – α-amanitin binding – Zn^{2+} finger motif – C-terminal repeated domain – Transcription and cell cycle regulation

Introduction

Transcriptional activation of diverse genes is mediated by common events involving the interaction of promoter elements with transcription regulatory proteins and RNA polymerases (for reviews see Maniatis et al. 1987; Ptashne 1988; Johnston and McKnight 1989). While genes have been characterized by specific interaction of promoter elements and transcription factors (Kuhlemeier et al. 1987; Schell 1987; Benfey and Chua 1989), the role of RNA polymerase subunits in the control of transcription in plants has remained largely obscure. Apart from biochemical and immunological characterization of plant enzymes (Guilfoyle et al. 1974; Guilfoyle and Dietrich 1986), most of our knowledge of RNA polymerases comes from other organisms.

RNA polymerase II (RpII) is a complex enzyme composed of 9–11 smaller and 2 larger subunits. The largest subunit is well-studied because it binds a transcription inhibitor, α-amanitin. Isolation of an α-amanitin resistant mutant and P-element tagging of the *RpII215* gene in *Drosophila* (Greenleaf et al. 1979; Searles et al. 1982) provided a key for cloning the genes for the largest subunits of RpII from fruitfly and other species (Biggs et al. 1985; Jokerst et al. 1989; Allison et al. 1985; Ahearn et al. 1987; Bird and Riddle 1989). Partial homology between the largest subunits of archaebacterial, eubacterial and eucaryotic RpII indicated two functional domains (Allison et al. 1985; Pühler et al. 1989). The N-terminal (or β′-like) domain carries at least seven evolutionarily conserved regions including Zn^{2+}-finger and DNA-binding motifs. A mutation resulting in α-amanitin resistance is located in one of these regions (Bartolomei and Corden 1987). A unique feature of the largest subunits of eucaryotic RpII is a C-terminal YSPTSPS consensus domain (CTD) that occurs in variably repeated units, 26 in yeast, 42 in *Caenorhabditis*, 44 in *Drosophila* and 54 in mouse. Mutations in the consensus sequence may cause altered developmental phenotypes, while about half of the CTD repeats are required for viability (Nonet et al. 1987; Bartolomei et al. 1988). Recent reports have demonstrated that the CTD repeats interact with transcription factors and serve as substrates for *cdc*2 protein kinases involved in cell cycle control (Brandt and Struhl 1989; Allison and Ingles 1989; Cisek and Corden 1989).

Offprint requests to: C. Koncz at Max-Planck-Institut, FRG

To explore the role of the largest subunit of RpII in the regulation of transcription and cell cycle in plants, an *rp*II 215 gene was isolated from *Arabidopsis thaliana* using a *Drosophila RpII215* probe. Here we report that common domains in the largest subunits of RpII are highly conserved in *Arabidopsis*, suggesting an evolutionary maintenance of regulatory functions of this protein in plants.

Materials and methods

Cloning of rp*II215 genomic and cDNAs from* Arabidopsis

Nuclear DNA was prepared from wild-type *A. thaliana* var. Columbia plants as described (Koncz and Schell 1986). Purification of plasmid DNAs, isolation and labeling of DNA fragments by random priming, preparation of phage lifts on nitrocellulose filters, physical mapping, treatments with alkaline phosphatase, blotting of RNA and DNA gels, generation of genomic and cDNA libraries in lambda phage vectors and other recombinant DNA techniques were according to Sambrook et al. (1989). After cloning 16–20 kb *Mbo*I fragments of *Arabidopsis* nuclear DNA into the *Bam*HI site of the lambda EMBL 4 vector, 200000 recombinant phages were screened using a 9.4 kb *Xba*I DNA fragment of plasmid pP*C*4 carrying the *RpII215* gene of *Drosophila* (Jokerst et al. 1989). A buffer containing 30% formamide, 5 mM EDTA, 50 mM NaH_2PO_4 (pH 7.0), 0.9 M NaCl, 250 µg/ml salmon sperm carrier DNA, 0.1% SDS and 0.2% each of polyvinylpyrrolidone, bovine serum albumin and Ficoll was used for low stringency hybridizations at 37° C followed by washing with 50 mM NaCl, 20 mM NaH_2PO_4 (pH 7.0), 1 mM EDTA and 0.1% SDS at the same temperature. For hybridizations at high stringency (at 42° C) the above buffer contained 50% formamide and the washing step was done at 68° C.

RNA was purified by extraction of plant tissues ground in liquid N_2 with 2 ml/g of buffer (8 M guanidium hydrochloride, 20 mM MES (2-(N-morpholino)-ethanesulfonic acid pH 7.0), 25 mM EDTA and 50 mM 2-mercaptoethanol) followed by incubation at 65° C for 20 min. After phenol-chloroform (1:1) extraction and pelleting of debris, 0.2 M potassium acetate (pH 5.0) and 10% ethanol was added to remove polysaccharides by precipitation for 15 min on ice, then the RNA was precipitated by 0.7 volume of isopropanol. The crude RNA was dissolved in 20 mM MES (pH 7.0), 20 mM EDTA, 0.5% SDS buffer, and extracted with phenol-chloroform. After repeating the polysaccharide removal step by addition of 0.3 M sodium acetate (pH 6.0) and 10% ethanol, the RNA was precipitated with isopropanol. To remove contaminating DNA, the RNA was dissolved in 10 mM TRIS/HCl (pH 7.5) and 1 mM EDTA buffer and precipitated at 4° C with 3 M LiCl. After washing the pellet with 3 M LiCl, the RNA was dissolved in water and applied onto Hybond mAP (messenger affinity) paper to isolate poly(A)$^+$ RNA according to Amersham instructions. RNA gels containing 0.8% formaldehyde were vacuum-blotted from 50 mM NaOH to Zeta-Probe membranes treated as described (Biorad instruction manual) and hybridized with diverse fragments of *rp*II215 genomic and cDNA probes. A cDNA library was constructed in lambda gt10 from poly(A)$^+$ RNA prepared from plants grown to maturity (a kind gift of A. Bachmair, MPI, Köln). Primer extension of poly(A)$^+$ RNA with an oligonucleotide (5′-GGCCGGAGAGAACGGAAACC-3′) was performed as described (Dean et al. 1987).

Nucleotide sequence analysis of rp*II215 genomic and cDNA clones*

Three genomic and six cDNA clones isolated from lambda EMBL 4 and gt10 libraries were mapped by standard methods and by hybridization with the *RpII215 Drosophila* probe. Overlapping fragments of both genomic and cDNAs were cloned in pUC18 and 19 as well as in M13mp18 and 19 vectors (Yanisch-Perron et al. 1985) to determine their nucleotide sequences. T7 DNA polymerase and single or double-stranded templates were used for DNA sequencing with the universal forward and reverse primers (Tabor and Richardson 1987). Genomic and cDNA fragments whose sequences were determined are indicated in Fig. 2. DNA sequences were analysed using a WISGEN program package (Devereux et al. 1984) adapted to a version 6.1 VAX/VCM computer. Nucleotide sequence data reported in this paper will appear in the EMBL, Genbank and DDJB Nucleotide Sequence Databases under accession number X 52954 A. thaliana rpII 215 gene.

Results

Cloning of Arabidopsis *genomic and cDNAs homologous to the* RpII215 *gene of* Drosophila

A 9.4 kb *Xba*I fragment of plasmid pP*C*4 carrying the entire *RpIIC4* gene of *Drosophila* (Jokerst et al. 1989) hybridized to several *Hin*dIII fragments of *Arabidopsis* nuclear DNA at low stringency (Fig. 1a). One of these fragments was cloned by screening of a lambda EMBL 4 genomic library with the *Drosophila RpII215* probe using similar hybridization conditions. From 200000 recombinant phages, 3 genomic clones (GC2.1, 3.0 and 4.3) were isolated. Overlapping physical maps of these clones defined a 24.3 kb region of the *Arabidopsis* genome carrying a *Hin*dIII fragment of 6.1 kb (Fig. 2) that hybridized to the *Drosophila RpII215* probe as well as to other nuclear DNA fragments at low stringency. Notwithstanding its partial homology to other fragments, hybridization mapping at high stringency revealed that the 6.1 kb *Hin*dIII fragment contained DNA sequences represented only once per haploid genome in *Arabidopsis* (Fig. 1b).

Using this fragment as a homologous probe, a 6.8 kb transcript of low abundance was detected in poly(A)$^+$ RNA prepared from fully developed plants (Fig. 1c). Screening of a lambda gt10 cDNA library with the probe yielded one homologous clone in 4×10^5 recombinant

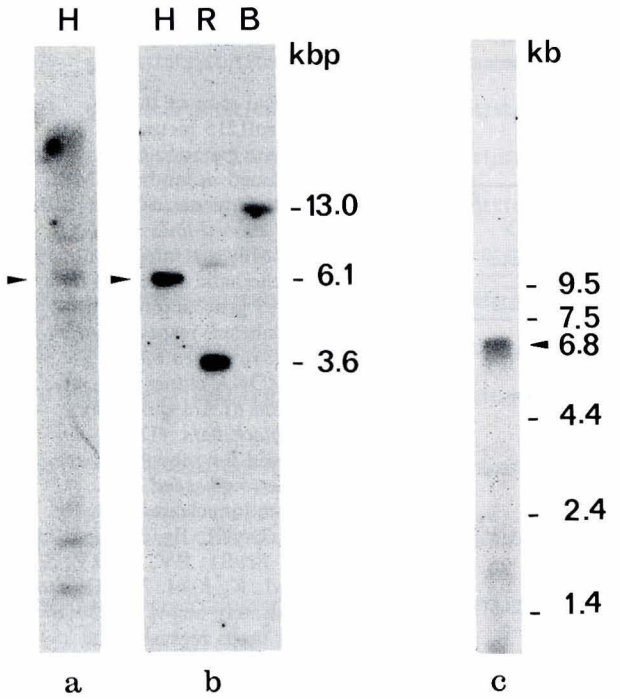

Fig. 1. a Southern hybridization of a 9.4 kb *Xba*I fragment of plasmid pPC4 DNA carrying the *RpII215* gene of *Drosophila* as probe to *Hin*dIII digested *Arabidopsis* nuclear DNA at low stringency (30% formamide, 37° C). **b** Southern hybridization of *Hin*dIII (H), *Eco*RI (R) and *Bam*HI (B) digested *Arabidopsis* nuclear DNA with a *Hin*dIII (222) – *Sal*I(1179) fragment of genomic DNA clone 3.0 as probe at high stringency (50% formamide, 42° C). (Numbers in brackets indicate the position of endonuclease cleavage sites within the *rp*II215 genomic DNA sequence. Note a weak hybridization of an *Eco*RI fragment carrying *rp*II215 homologous DNA sequences unlinked to the gene). **c** Northern hybridization of 10 µg poly(A)$^+$ RNA prepared from mature plants with a mixed probe containing *Eco*RI(−199) – *Pst*I(902) and *Bam*HI(3344) – *Bam*HI(5068) fragments of genomic DNA clone 3.0

phages. Six cDNA clones were isolated (C31, 41, 42, 43, 53 and 61) and mapped by hybridization with diverse fragments of genomic clones. Although all cDNAs were expected to posses identical 3′ ends, the ends of only two cDNAs were mapped to an identical position (Fig. 2) suggesting that internal priming occurred during reverse transcription of the long transcript. Comparison of genomic and cDNA maps indicated that the 3′ end of the identified gene is located outside the 6.1 kb *Hin*dIII fragment and contains several introns.

Identification of the rp*II215 gene of* Arabidopsis

The use of the *RpII215 Drosophila* probe for cloning related genes from other organisms demonstrated that the largest subunits of RNA polymerases II and III are substantially homologous (Allison et al. 1985; Bird and Riddle 1989). To identify to which RNA polymerase the largest subunit gene isolated from *Arabidopsis* corresponds, the nucleotide sequences of cDNA clones were determined as outlined in Fig. 2. Close to the 3′ end of cDNAs C42 and C61 a number of direct repeats were found that encoded multiple copies of a YSPTSPS peptide (Fig. 3a). This unequivocally demonstrated that the cDNAs were complementary to the 3′ end of the gene for the largest subunit of *Arabidopsis* RNA polymerase II, since similar repeats have only been identified in the carboxy-terminal domains of the largest subunits of eucaryotic RpII. Physical mapping and nucleotide sequence data showed that all cDNAs corresponded to the same transcript, a 4437 bp region which was represented in the longest cDNA clone.

To analyse the structure of the *rp*II215 gene, the nucleotide sequence of an 8676 bp region of genomic DNA was determined using subcloned fragments of lambda EMBL 4 clones (Fig. 4). By comparison of genomic and cDNA sequences, the positions of six introns were defined (Fig. 4). The nucleotide sequence of identified exon-intron boundaries followed the general GT-AG rule described for splice junctions of eucaryotic genes (Mount 1982; Fig. 3b). A computer program was used to analyse genomic DNA sequences located upstream of the 5′ end of the cDNA C61 (position 2225). First the nucleotide and derived amino acid sequences of known genes for the RpII largest subunit were compared pairwise to establish overlaps. A consensus sequence generated by the program showed that in scattered positions about 50% of amino acids are identical (Figs. 3c, 5). Alignment of the consensus DNA sequence with the genomic and overlapping cDNA sequences indicated that the program is well-suited for identification of exon-intron boundaries. To localize the missing introns, all possible reading frames and putative splice junction sequences were examined in the remaining genomic DNA sequence. Only one alignment was found which showed overlaps with all reference points of the consensus sequence. This indicated that the *rp*II215 gene of *Arabidopsis* contains 13 short introns (Fig. 4).

Analysis of the nucleotide and derived amino acid sequence of the rp*II215 gene*

1. Exon-intron structure and DNA sequence homology. The complexity of sequence organization of genes encoding RpII largest subunit correlates with the number of introns which increases from 3 in *Drosophila* to 11 in *Caenorhabditis* and 28 in mouse. In comparison of these species the position of the first intron of 190 bp is conserved in *Arabidopsis*. In a second region (position 3370), where *Drosophila* intron 2, *Caenorhabditis* intron 6 and mouse intron 14 overlap, the *Arabidopsis* gene shows homology to the uninterrupted *RPO21* gene of yeast. Similarly to the *ama*-1 gene of *Caenorhabditis*, the 3′ end of the *Arabidopsis* gene also contains three introns, albeit in different positions. Two of these interrupt the coding region of YSPTSPS repeats while, uniquely among all species, one intron is located within the 3′ untranslated region of the *Arabidopsis rp*II215 gene.

In contrast to the diversity of introns, a high level of sequence conservation was observed among the coding regions of the genes. The homology broke down

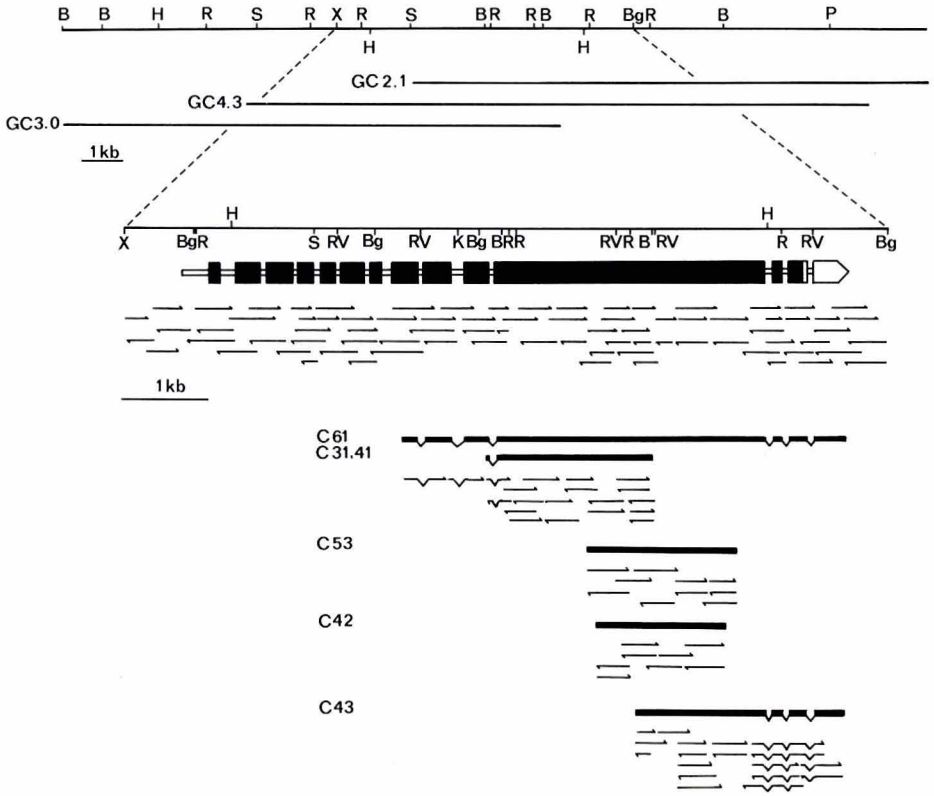

Fig. 2. Physical map of the *Arabidopsis rp*II215 locus. GC 2.1, 3.0 and 4.3 are genomic DNA fragments cloned in lambda EMBL 4. Nucleotide sequence of the region indicated by *dashed lines* was determined using a strategy outlined below the schematic structural map of the *rp*II215 gene. Exons and introns are labeled, respectively, by *black* and *narrow open boxes*. Position of cDNA clones C31, 41, 42, 43, 53 and 61 are shown by *interrupted black bars*. cDNA fragments used for sequencing are marked by *arrow-headed lines*. Restriction endonucleases cleavage sites are B, *Bam*HI; Bg, *Bgl*II; H, *Hin*dIII; R, *Eco*RI; RV, *Eco*RV; S, *Sal*I; P, *Pvu*I; K, *Kpn*I. *Hin*dIII sites marked separately below or above the physical maps represent the boundaries of the 6.1 kb *Hin*dIII fragment detected by Southern DNA hybridization with the *RpII215 Drosophila* probe

CARBOXY TERMINAL REPEATS

	Y	S	P	T	S	P	S GA
1	F	S	P	S	S S	P	G
2	Y	S	P	S	S	P	G
3	Y	S	P	T	S	P	G
4	Y	S	P	T	S	P	G
5	Y	S	P	T	S	P	G
6	Y	S	P	T	S	P	T
7	Y	S	P	S	S	P	G
8	Y	S	P	T	S	P	A
9	Y	S	P	T	S	P	S
10	Y	S	P	T	S	P	S
11	Y	S	P	T	S	P	S
12	Y	S	P	T	S	P	S
13	Y	S	P	T	S	P	S
14	Y	S	P	T	S	P	S
15	Y	S	P	T	S	P	A
16	Y	S	P	T	S	P	A
17	Y	S	P	T	S	P	A
18	Y	S	P	T	S	P	S
19	Y	S	P	T	S	P	S
20	Y	S	P	T	S	P	S
21	Y	S	P	T	S	P	S
22	Y	S	P	T	S	P	S
23	Y	S	P	T	S	P	A
24	Y	S	P	T	S	P	G
25	Y	S	P	T	S	P	S
26	Y	S	P	T	S	P	S
27	Y	S	P	T	S	P	S
28	Y	G	P	T	S	P	S
29	Y	N	P	Q	S	A	K
30	Y	S	P		S	I	A
31	Y	S	P		S	N	A R
32	L	S	P	A	S	P	
33	Y	S	P	T	S	P	N
34	Y	S	P	T	S	P	S
35	Y	S	P	T	S	P	S
36	Y	S	P	S	S	P	T
37	Y	S	P	S	S	P	
38	Y	S S	G	A	S	P	D
39	Y	S	P		S	A	G
40	Y	S	P	T	L	P	G
41	Y	S	P	S	S	T	G Q
42	Y	T	P	H	E	G	DKKDKTGKKDASKDDKGNP

a

INTRON/EXON-BOUNDARIES

1	CGATGAGATC gtaatgctcc	ttttttcag AGGCAAATGT
2	CTTATTTCGT gtcagcattg	taagtttaag CAGGCTATGA
3	TGCTGATAGG gttcgtcttt	attcttgcag GTTTTGAGTG
4	CAGGAGTGAG gtaagtcgac	gttgatgcag GATGACTTGA
5	ACAGCCAAGG gtaatgcttt	ttttcttcag GCTACTCAGA
6	TTTTCTGAGA gtaatttttc	aagactggag CGAAATATAT
7	TGGATACAAG gtatgtacta	ttggtcacag GTGGAGCGGC
8	CATAGAGAAG **gt**agcaattt	ccattttc**ag** GATGTATTCA
9	TCTCGTGCAT **gt**aagttgca	tggcttcc**ag** GGAAGAGGTT
10	GGTTGACCAG **gt**atattcaa	ttccttgt**ag** GTTTTGAATA
11	CCAACTACAG **gt**aagtagta	tactctgc**ag** CCCGACATCT
12	GTCCCAGCAG **gt**ttgacttt	tctgtctc**ag** GTCATGATTT
13	CATGTTTTAG **gt**aacgtcta	taacccgc**ag** GTCATGATTT

b

Amino Acid Identity [%]

	A.th.	yeast	C.e.	D.m.	mouse
A.th.	100.0	-	-	-	-
yeast	57.2	100.0	-	-	-
C.e.	59.0	54.5	100.0	-	-
D.m.	58.0	53.9	66.7	100.0	-
mouse	60.9	55.6	70.5	73.6	100.0

c

Fig. 3a. Comparison of the amino acid sequence of carboxy-terminal domains of *Arabidopsis* RpII215 protein. **b** Exon-intron boundaries identified in the *rp*II215 genomic DNA sequence bold letters; deduced from cDNA sequencing (see Fig. 2), all others; deduced from genomic sequence comparisons. **c** Comparison of amino acid identity among RpII largest subunit proteins of *Arabidopsis thaliana* (A.th.), yeast, *Caenorhabditis elegans* (C.e.), *Drosophila melanogaster* (D.m.) and mouse

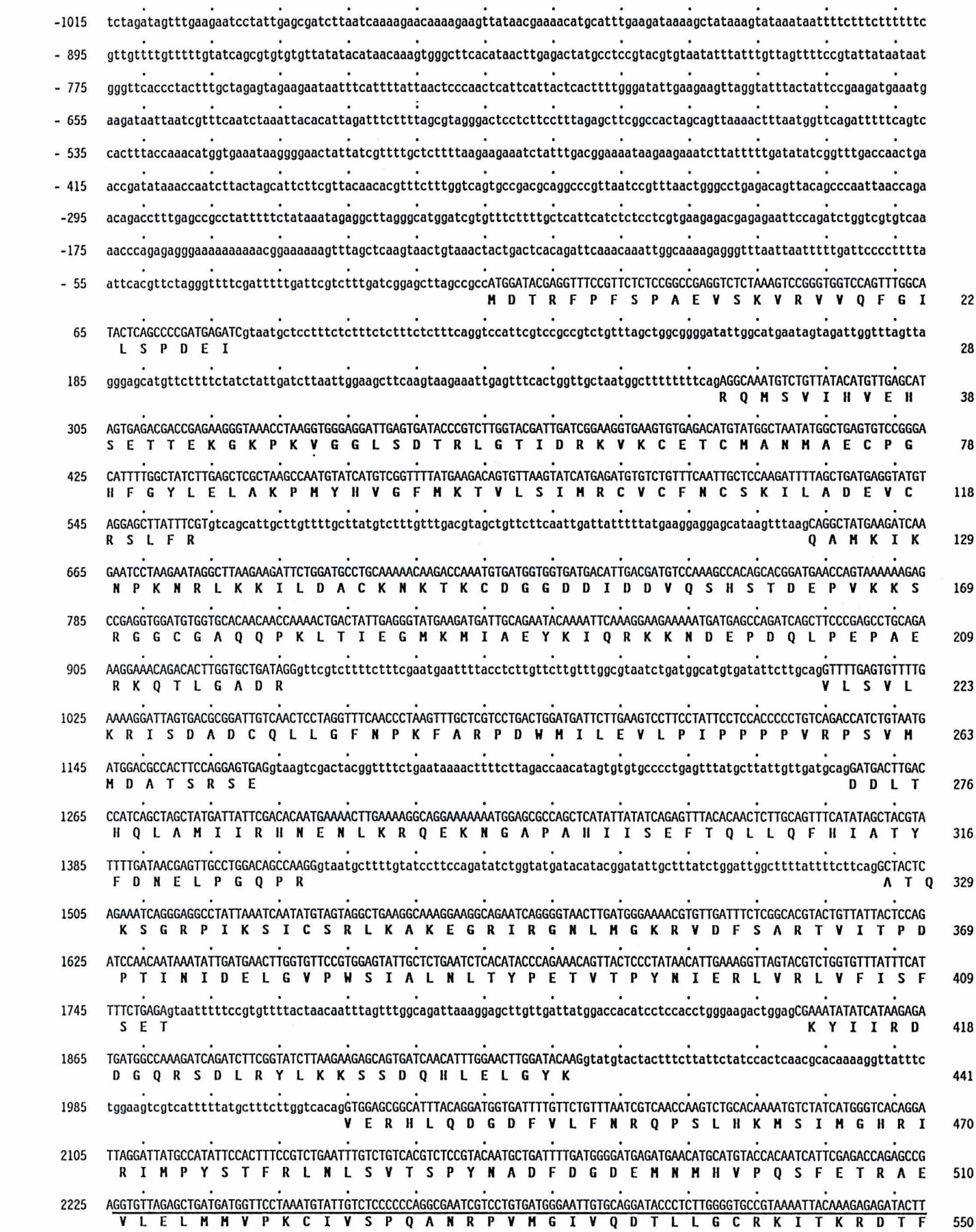

Fig. 4 (continuation, see page 70)

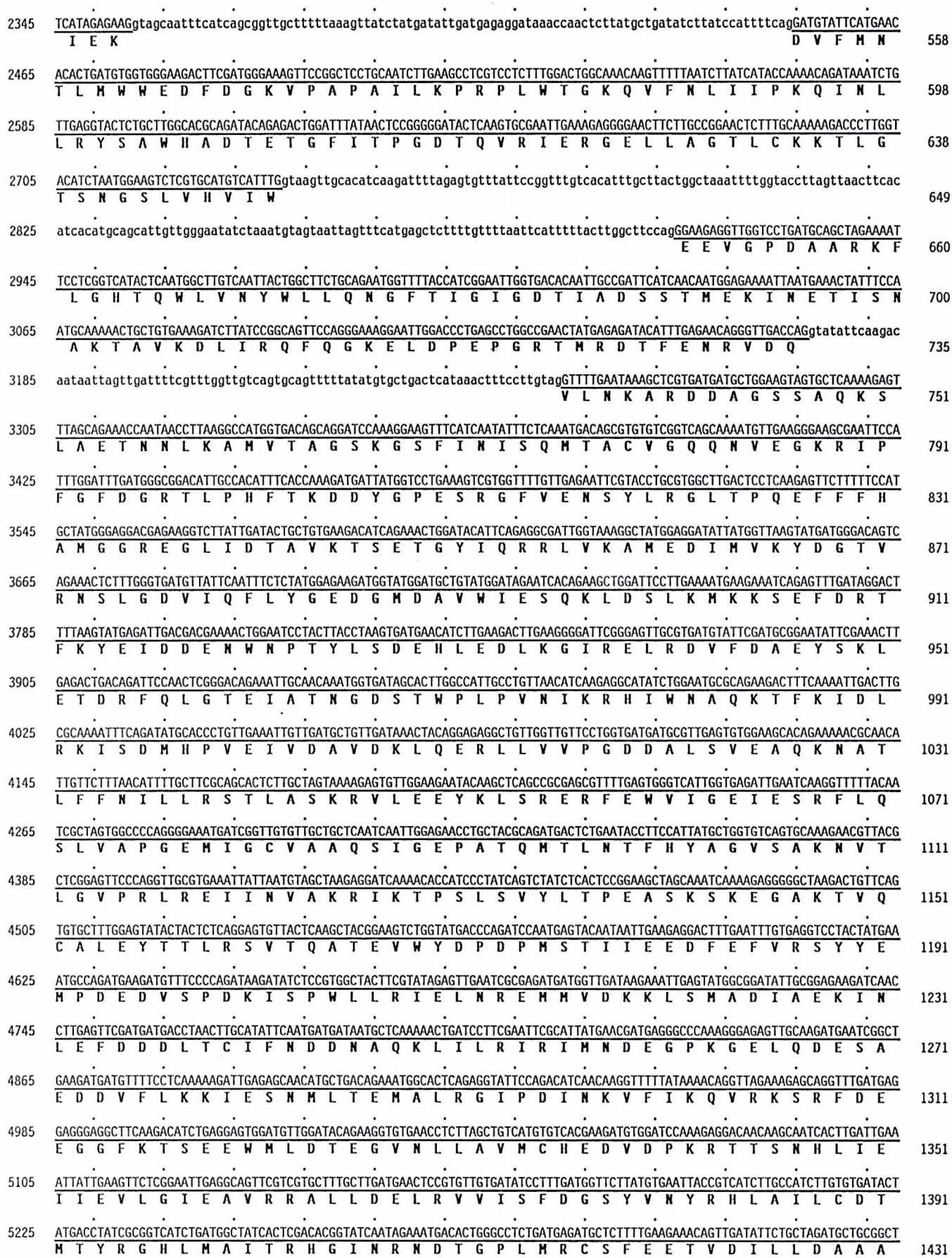

Fig. 4 (continued)

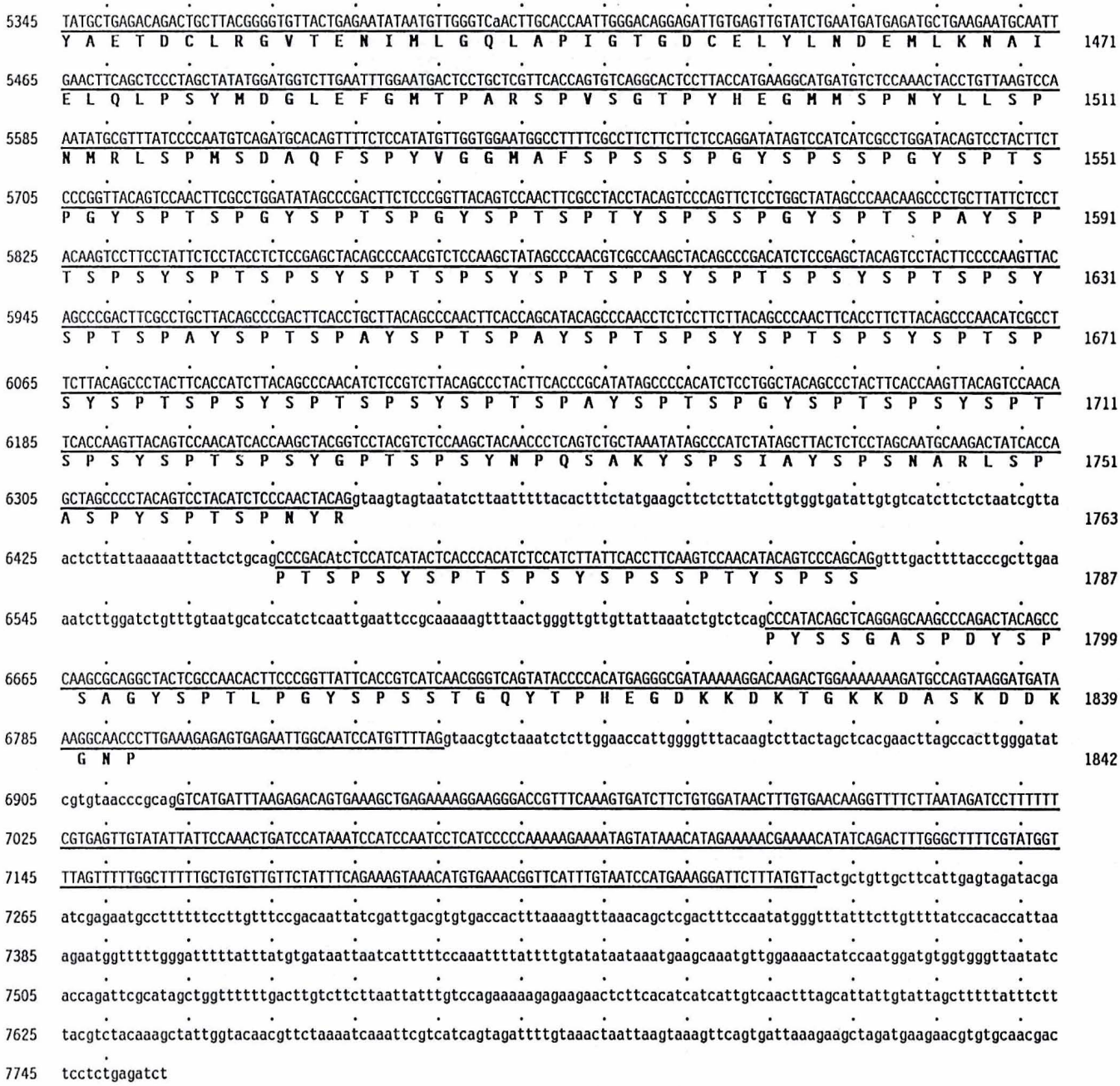

Fig. 4. The complete nucleotide and derived amino acid sequence of the *Arabidopsis rp*II215 gene. Sequences represented by cDNA clones are *underlined.* Introns are printed in *lower case*

sharply outside the predicted start and stop codons of the *Arabidopsis* gene. Lack of longer open reading frames (ORFs) and putative splice junctions upstream of the coding region suggested that the localization of the start codon was correct. Transcription of the first exon was confirmed by primer extension of poly(A)$^+$ RNA using an oligonucleotide that terminated upstream of the ATG codon (data not shown).

Translation of the common ORF of cDNA sequences defined a long 3′ untranslated region extending 436 bp downstream of the stop codon. Except for a sequence (5′-TATAA-3′) located 130 bp upstream of the poly(A) tail, no homology to the consensus (5′-AATAA-3′) polyadenylation signal was detected within this region.

2. Conserved domains in the largest subunits of RpII. In contrast to the other RpII215 proteins, the sequences of which were deduced by the analysis of genomic DNA sequences, 1332 amino acid residues of the largest subunit of *Arabidopsis* RpII were deduced from the cDNA sequence. Translation of the coding region distributed over 13 exons of the *rp*II215 gene resulted in a putative protein product of 1841 amino acids and 205 kDa.

Comparison of amino acid sequences encoded by *rp*II215 genes of *Arabidopsis* and other species indicated that the *Arabidopsis* protein contains domains that are highly conserved between the largest subunits of all eucaryotic RpII. Nine of these (Fig. 5, D1–D9) were previously identified in the N-terminal region by compari-

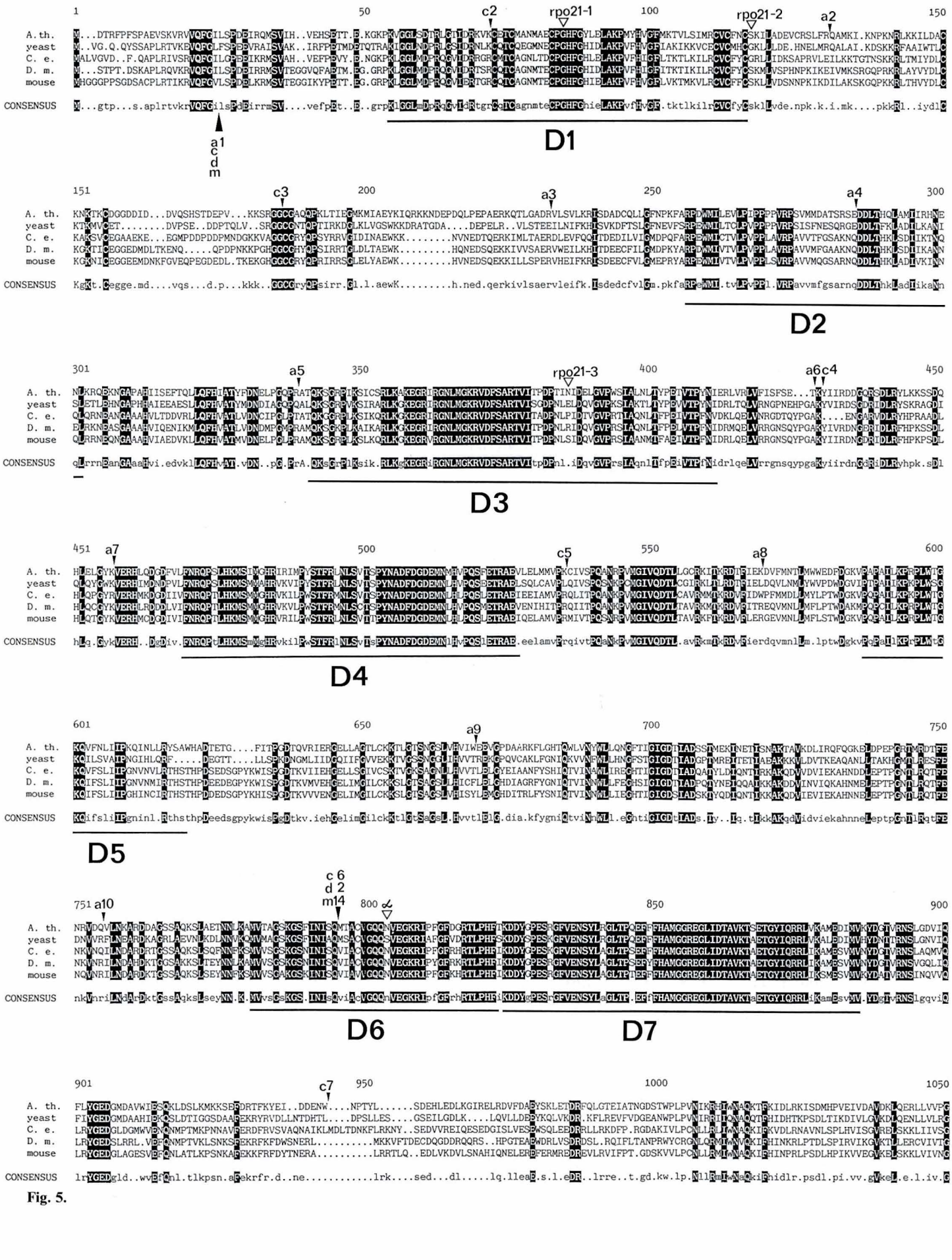
c2
rpo21-1
rpo21-2
a2
a1
c
d
m
D1
c3
a3
a4
D2
a5
rpo21-3
a6
c4
D3
a7
c5
a8
D4
a9
D5
c 6
d 2
m14
a10
D6
D7
c7

Fig. 5.

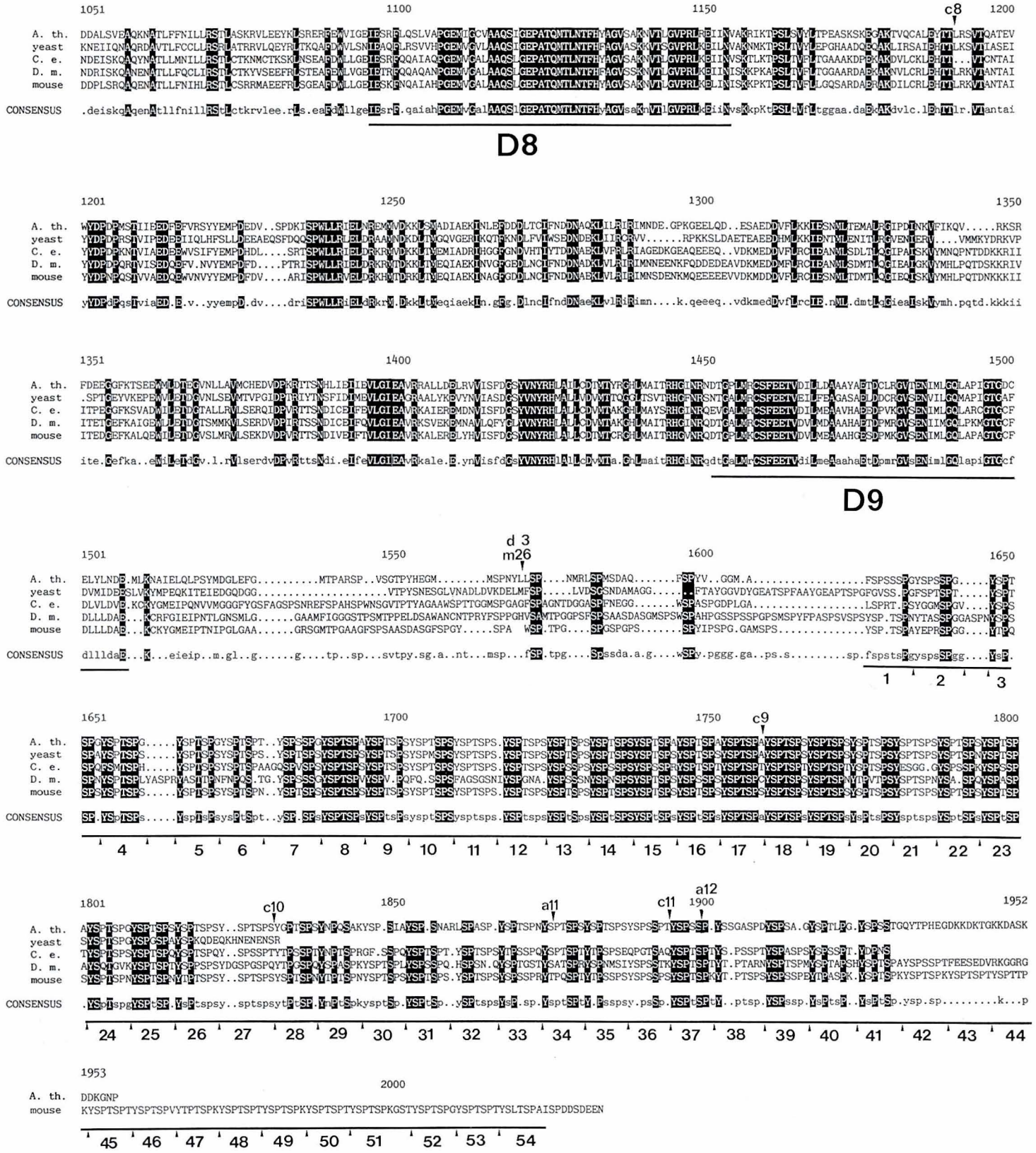

Fig. 5. A consensus sequence of eucaryotic RpII largest subunit protein generated by the VAX-WISGEN computer program. Regions containing identical amino acids are printed in *black* and shown by *capital letters* in the consensus sequence. Amino acids common in at least three species are printed in *lower case*. Homologous domains (D1 and D9) determined previously by sequence comparison with the β' subunit of *Escherichia coli* RNA polymerase (Jokerst et al. 1989) are *underlined*. Corresponding positions of *Arabidopsis* (a1–a13), *Caenorhabditis* (c1–c11), *Drosophila* (d1–d3) and mouse introns (m1, m14 and m26) are indicated by *black arrows* above the sequence. Common position of the first intron is shown by the *large black arrow* below the sequence. *Open triangles* label the position of thermosensitive yeast mutations (*rpo*21 −1 to −3, Himmelfarb et al. 1987) shown to influence RNA polymerase II function. Position of a mutation causing α-amanitin resistance in mouse is marked by α. Repeated C-terminal consensus domains are *underlined* and *numbered*

son of yeast, *Drosophila* and mouse RpII largest subunit sequences to that of the β' subunit of *Escherichia coli* RNA polymerase (Allison et al. 1985; Jokerst et al. 1989; Ahearn et al. 1987). The fact that a consensus sequence could be generated using only a minimal number of gaps, suggested that the number of regions where substitutions or deletions are permitted is very limited. Such regions occur between domains D7 and D8 (region of intron 7 in *Caenorhabditis*) and downstream of domain 9 (region of *Drosophila* intron 3) separating the N- and C-terminal parts of the protein.

Following the common region flanking the first intron, all genes contain the highly conserved domain D1. The amino acid sequence of this domain in *Arabidopsis* also fits to a consensus $C-X_{2-4}-C-X_{2-15}-C-X_{2-4}-H-X_{2-15}-C-X-C-X_2-C-X_4-L$ representing a Zn^{2+} finger motif. Thermosensitive mutations *rpo*21-1 and *rpo*21-2 in this domain in yeast resulted in a defective RpII and in decreased viability, indicating the functional importance of this sequence (Himmelfarb et al. 1987). Domain D3, identified by the *rpo*21-3 yeast mutation yielding a similar phenotype, contains 12 basic lysine and arginine residues within 36 amino acids in *Arabidopsis* as well as in *Drosophila* and mouse. Domains D3 and D4 are thought to be parts of the DNA binding site (Ahearn et al. 1987). Domains D6 and D7 represent the central and most conserved part of the largest subunits of eucaryotic RpII. A VGQQ n VEGK-RI motif located in domain D6 is identical in all species except yeast and probably corresponds to the α-amanitin binding site. A mutation changing the Asn residue to Asp in mouse has been demonstrated to cause α-amanitin resistance (Bartolomei and Corden 1987). The fact that Ser residue replaces this Asn in yeast correlates with the α-amanitin tolerance of yeast RpII.

The C-terminal domain of the *Arabidopsis* protein contains 15 consensus and 26 variant YSPTSPS repeats. The codon usage for the first six amino acids of these repeats is conserved, while in the last position a serine is encoded by either AGX or TCX codons. Mutations in the first position to G result in replacement of Ser by Gly or Ala residues. From repeat 2 to 28 all YSPTSPS domains are identical except for the last amino acids that form a characteristic G, S, A, S, A/G, S domain pattern (Fig. 5). Towards the C-terminal end (overlapping the position of *Arabidopsis* introns 11 and 12) the diversity of repeats increases. Of the C-terminal repeat domains (CTDs), 18 are highly conserved among *Arabidopsis* and other species (Allison et al. 1985; Ahearn et al. 1987). This supports the notion that the presence of 15 CTDs in the largest subunits of RpII are absolutely required for viability (Bartolomei et al. 1988; Allison et al. 1988).

Discussion

Characterization of the *rp*II215 gene of *Arabidopsis* answers several important questions. The analysis described above demonstrates that all homologous domains of functional importance in the largest subunits of RpII are conserved between plants and other eucaryotes. This implies that much of our knowledge based on studies of possible regulatory functions of these proteins in other organisms may be extrapolated to plants.

As in other species, the largest subunit of RpII is encoded by a single gene in *Arabidopsis*. Low stringency hybridization data obtained with both *Drosophila* and *Arabidopsis* probes indicated that the *rp*II215 gene is, however, homologous to other loci of the *Arabidopsis* genome. This is compatible with the data showing immunological cross-reaction between the largest subunits of plant RNA polymerases (Guilfoyle 1980). Characterization of genomic clones currently in progress will determine whether these sequences correspond to any of the largest subunit genes of RNA polymerase I and III (C. Nawrath, unpublished).

Preliminary studies of transcription indicated that the *rp*II215 gene is expressed only at a low level in differentiated plant tissues. In fully developed plants the *rp*II215 transcript represents less than 0.001% of the total amount of steady-state mRNAs. This may be explained by the lack of a proper polyadenylation signal at the 3′ end of the transcript. A possible involvement of the last intron in post-transcriptional regulation (i.e. differential splicing) also cannot be excluded, although the cDNA mapping data do not support this assumption.

The *rp*II 215 gene of *Arabidopsis* encodes a protein of 205 kDa. This value is smaller than that of the estimated 220–240 kDa molecular mass found for the corresponding RpII largest subunits in other plants. In vivo phosphorylation of Ser and Thr residues of CTD repeats, converting the protein from form IIA to IIO, probably accounts for this difference (Cadena and Dahmus 1987; Guilfoyle et al. 1984; Guilfoyle 1989). Recent reports have demonstrated that the CTD repeats are specifically phosphorylated by *cdc*2 protein kinases suggesting that post-translational modification of the largest subunit of RpII plays an important role in cell cycle control (Cisek and Corden 1989). Immunological identification of protein p34 of the *cdc*2 kinase complex (John et al. 1989) and the highly conserved structure of the CTD repeats in *Arabidopsis* indicate that similar regulation may exist in higher plants.

A major function of CTD repeats in the regulation of transcription could result from the interaction of the largest subunit of RpII with transcription activator proteins, such as GCN4 in yeast (Allison and Ingles 1989; Brandt and Struhl 1989). This observation may hold for other organisms since GCN4 is structurally related to the *jun* oncoprotein and the vertebrate transcription factor AP-1 (Jones et al. 1988). It is intriguing that certain plant transcription factors, such as TGA1a and b in tobacco and HBP-1 in wheat (Katagiri et al. 1989; Tabata et al. 1989), are homologous to GCN4 and can mediate tissue-specific regulation of transcription by binding close upstream to the TATA box. Whether any of these transcription factors interact with the largest subunit of RpII may be determined by using protein fusions and antibodies which can be generated with the available cDNA clones of the *rp*II215 gene of *Arabidopsis*.

Acknowledgements. The authors thank T. Etzold for discussions and advice on the use of computer programs, A.A. Greenleaf for providing plasmid pPC4 and Gisela Kobert for typing the manuscript. This work was supported as part of a joint project between the MPI, Köln and BRC, Szeged by the Deutsche Forschungsgemeinschaft and by the Hungarian Academy of Sciences.

References

Ahearn JM, Bartolomei Jr MS, West ML, Cisek LJ, Corden JL (1987) Cloning and sequence analysis of the mouse genomic locus encoding the largest subunit of RNA polymerase II. J Biol Chem 262:10695–10705

Allison LA, Ingles CJ (1989) Mutations in RNA polymerase II enhance or suppress mutations in GAL4. Proc Natl Acad Sci USA 86:2794–2798

Allison LA, Moyle M, Shales M, Ingles CJ (1985) Extensive homology among the largest subunits of eucaryotic and procaryotic RNA polymerases. Cell 42:599–610

Allison LA, Jerry KC, Wong W, Fitzpatrick D, Moyle M, Ingles J (1988) The C-terminal domain of the largest subunit of RNA polymerase II of *Saccharomyces cerevisiae*, *Drosophila melanogaster* and mammals: a conserved structure with an essential function. Mol Cell Biol 8:321–329

Bartolomei MS, Corden JL (1987) Localization of an α-amanitin resistance mutation in the gene encoding the largest subunit of mouse RNA polymerase II. Mol Cell Biol 7:586–594

Bartolomei MS, Halden NF, Cullen CR, Corden JL (1988) Genetic analysis of the repetitive carboxy-terminal domain of the largest subunit of mouse RNA polymerase II. Mol Cell Biol 8:330–339

Benfey DN, Chua N-H (1989) Regulated genes in transgenic plants. Science 244:174–181

Biggs J, Searles LL, Greenleaf AL (1985) Structure of the eucaryotic transcription apparatus: features of the gene for the largest subunit of *Drosophila* RNA polymerase II. Cell 42:611–621

Bird DM, Riddle DL (1989) Molecular cloning and sequencing of ama-1, the gene encoding the largest subunit of *Caenorhabditis elegans* RNA polymerase II. Mol Cell Biol 9:4119–4130

Brandt CJ, Struhl K (1989) Yeast GCN4 transcriptional activator protein interacts with RNA polymerase *in vitro*. Proc Natl Acad Sci USA 86:2652–2656

Cadena LD, Dahmus ME (1987) Messenger RNA synthesis in mammalian cells is catalyzed by the phosphorylated form of RNA polymerase II. J Biol Chem 262:12468–12474

Cisek LJ, Corden JL (1989) Phosphorylation of RNA polymerase by the murine homologue of the cell cycle control protein *cdc* 2. Nature 339:679–684

Dean C, Farreau M, Dunsmuir P, Bedbrook J (1987) Confirmation of the relative expression levels of *Petunia* (Mitchell) *rbc*S genes. Nucleic Acids Res 15:4655–4668

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

Greenleaf AL, Borsett LM, Jiamachello PF, Coulter DE (1979) α-amanitin-resistant *D. melanogaster* with an altered RNA polymerase II. Cell 18:613–622

Guilfoyle TJ (1980) Purification subunit structure and immunological properties of chromatin-bound ribonucleic acid polymerase I from cauliflower inflorescence. Biochemistry 19:5966–5972

Guilfoyle TJ (1989) A protein kinase from wheat germ that phosphorylates the largest subunit of RNA polymerase II. Plant Cell 1:827–836

Guilfoyle TJ, Dietrich MA (1986) Plant RNA polymerases: Structures, regulation and genes. In: Bruening G, Harada J, Kosuge T, Hollaender A (eds) Tailoring genes for crop improvement. Plenum press, New York, pp 87–100

Guilfoyle TJ, Hagen G, Malcolm S (1984) Immunological studies on plant DNA-dependent RNA polymerases with antibodies raised against individual subunits. J Biol Chem 259:640–648

Himmelfarb HJ, Simpson EM, Friesen JD (1987) Isolation and characterization of temperature-sensitive RNA polymerase II mutants of *Saccharomyces cerevisiae*. Mol Cell Biol 7:2155–2164

John PCL, Sek FJ, Lee MG (1989) A homology of the cell cycle control protein p34cdc2 participates in the division cycle of *Chlamydomonas* and a similar protein is detectable in higher plants and remote taxa. Plant Cell 1:1185–1193

Johnston PF, McKnight SL (1989) Eucaryotic transcriptional regulatory proteins. Annu Rev Biochem 58:799–839

Jokerst RS, Weeks JR, Zehring WA, Greenleaf AL (1989) Analysis of the gene encoding the largest subunit of RNA polymerase II in *Drosophila*. Mol Gen Genet 215:266–275

Jones NC, Rigby PWJ, Ziff EB (1988) *Trans*-acting protein factors and the regulation of eucaryotic transcription: lessons from studies on DNA tumor viruses. Genes Dev 2:267–281

Katagiri F, Lam E, Chua N-H (1989) Two tobacco DNA-binding proteins have homology to CREB. Nature 340:727–730

Koncz C, Schell J (1986) The promoter of T_L-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. Mol Gen Genet 204:383–396

Kuhlemeier C, Green PJ, Chua N-H (1987) Regulation of gene expression in higher plants. Annu Rev Plant Physiol 38:222–257

Maniatis T, Goodbourn S, Fischer JA (1987) Regulation of inducible and tissue specific gene expression. Science 230:1237–1245

Mount SM (1982) A catalogue of splice junction sequences. Nucleic Acids Res 10:459–472

Nonet M, Sweetser D, Young RA (1987) Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II. Cell 50:909–915

Ptashne M (1988) How eucaryotic transcriptional activators work. Nature 335:683–689

Pühler G, Leffers H, Gropp F, Palm P, Klenk H-P, Lottspeich F, Garrett RA, Zilling W (1989) Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eucaryotic nuclear genome. Proc Natl Acad Sci USA 86:4569–4573

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Schell J (1987) Transgenic plants as a tool to study the molecular organization of plant genes. Science 237:1176–1183

Searles L, Jokerst RS, Bingham PM, Voelker RA, Greenleaf AL (1982) Molecular cloning of sequences from a *Drosophila* RNA polymerase II locus by P element transposon tagging. Cell 31:585–592

Tabata T, Takase H, Takayama S, Mikami K, Nakashika A, Kawata T, Nakayama T, Iwabuchi M (1989) A protein that birds to a cis-eating element of wheat histone genes has a leucine zipper motif. Science 245:965–967

Tabor S, Richardson CC (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc Natl Acad Sci USA 84:4767–4771

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:109–119

Communicated by R. Herrmann