Mol Gen Genet (1990) 221:379–394

MGG

Nucleotide sequence of the maize chloroplast *rpo* $B/C_1/C_2$ operon: Comparison between the derived protein primary structures from various organisms with respect to functional domains

Gabor L. Igloi, Andreas Meinke*, István Döry, and Hans Kössel**
Institut für Biologie III der Universität Freiburg, Schänzlestrasse 1, D-7800 Freiburg, Federal Republic of Germany

Summary. The genes (*rpo* $B/C_1/C_2$) coding for the β, β', β'' subnits of maize (*Zea mays*) chloroplast RNA polymerase have been located on the plastome and their nucleotide sequences established. The operon is part of a large inversion with respect to the tobacco and spinach chloroplast genomes and is flanked by the genes *trn*C and *rps*2. Notable features of the nucleotide sequence are the loss of an intron in *rpo*C_1 and an insertion of approximately 450 bp in *rpo*C_2 compared to the dicotyledons tobacco, spinach and liverwort. The derived amino acid sequence of this additional monocotyledon specific sequence is characterized by acidic heptameric repeat units containing stretches of glutamic acid, tyrosines and leucines with regular spacing. Other structural motifs, such as a nucleotide binding domain in the β subunit and a zinc finger in the β' subunit, are compared at the amino acid level throughout the RNA polymerase subunits with the enzymes from other organisms in order to identify functionally important conserved regions.

Key words: *Zea mays* – Chloroplasts – *rpo* genes – RNA polymerase – Heptameric repeat

Introduction

RNA polymerase from chloroplasts, being a key enzyme in maintaining the semiautonomous nature of the organelle, has been the subject of investigations for some years (for review see Briat et al. 1986). A number of purification schemes for the enzyme from various sources have been reported but the subunit composition, multiplicity and site of transcription of its genes (imported proteins vs plastid DNA encoded proteins) have been the subject of some discussion (Little and Hallick 1988).

Attention was directed to these issues recently with the discovery of open reading frames in the chloroplast genomes of tobacco (Shinozaki et al. 1986), liverwort (Ohyama et al. 1986) and rice (Hiratsuka et al. 1989) coding for proteins bearing a strong sequence resemblance to the subunits of *Escherichia coli* RNA polymerase. The expression was demonstrated of the α-subunit (from the *rpo*A gene) in maize (Ruf and Kössel 1988) and pea (Purton and Gray 1989) and of the β, β' and a novel β'' subunit (from *rpo*B, C_1 and C_2, respectively) in spinach (Hudson et al. 1988) and adds support to the view that this enzyme belongs to the class of proteins encoded and expressed in the organelle. While the existence was noted of an *rpo*BC operon in tobacco and liverwort, a detailed analysis of the operon was carried out for the dicotyledon spinach (Hudson et al. 1988), revealing the division of the gene corresponding to *E. coli rpo*C into two entities C_1 and C_2, with the C_1 sequence interrupted by an intron. It seemed, therefore, of interest to compare the corresponding operon in the monocotyledon maize, in order to establish the evolutionary conservation of these and other structural features.

At the time that this work was completed and presented in a short report (Kössel et al. 1989), the sequence of the rice chloroplast genome became available (Hiratsuka et al. 1989), of which the *rpo*BC operon is analysed in more detail in the accompanying paper (Shimada et al. 1990). The structural comparison of this region of the maize genome could, therefore, be extended to the *rpo*BC operon of the closely related monocotyledon rice.

Materials and methods

The maize chloroplast genomic library in pBR329 was constructed in this laboratory and made available by Dr. E. Fritzsche (Fritzsche 1988). Restriction enzymes, DNA polymerase (Klenow) and T7 DNA polymerase sequencing kits were obtained from Pharmacia-LKB (Freiburg) and used according to the manufacturers' instructions. Deoxyoligonucleotides were synthesized on an ABI 380A DNA synthesizer and could usually be used as sequencing primers without further purification.

Restriction fragments for sequence analysis were cloned into M13mp18 or pUC19. The fragments generated by the Sanger method using [α-^{35}S]dATP (Amersham) were separated on a Macrophor electrophoresis workstation (Pharmacia-LKB, Freiburg) at 55° C on 6% polyacrylamide gels containing 7 M urea. Gels were fixed in 10% MeOH/10%

* *Present address:* Department of Microbiology, University of British Columbia, Vancouver, Canada
** *Permanent address:* Agricultural Research Institute of the Hungarian Academy of Sciences, P.O. Box 19, H-2462 Martonvásár, Hungary

Offprint requests to: G.L. Igloi

The sequence data presented in this paper will appear in the EMBL/Gen Bank/DDBJ Nucleotide Databases under the accession number X17318

acetic acid (v/v) and dried. Autoradiography was accomplished by exposure to x-ray film (Fuji), usually overnight at room temperature.

The raw data were analysed using a collection of computer programs made available by W. Bottomly, CSIRO, Canberra, or with the aid of PCGENE (Genofit, Heidelberg). Multiple alignments and specific lay-out of sequences was achieved by purpose-written programs.

Results and discussion

Location and orientation of the genes

To locate the region of DNA coding for the RNA polymerase subunits, a maize chloroplast genomic DNA library (Fritzsche 1988) was screened using synthetic oligonucleotide hybridization probes designed on the basis of conserved sequences present in the tobacco and liverwort genes.

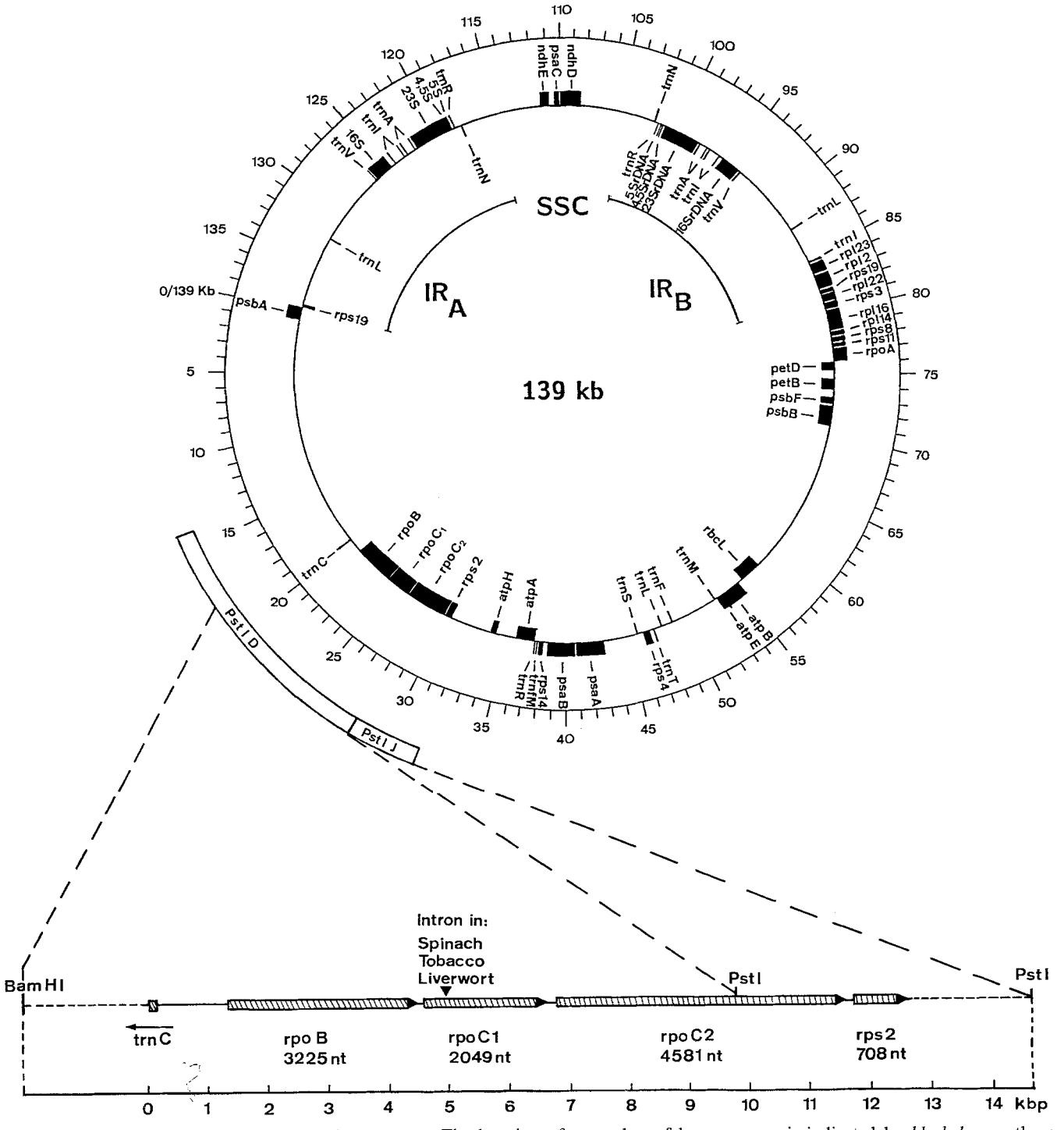

Fig. 1. Genetic map of the maize chloroplast genome. The location of a number of known genes is indicated by *black boxes*: those on the outside of the circle are transcribed clockwise, those on the inside, counterclockwise. The *rpo*B/C_1/C_2 operon and its flanking genes are contained in the two *Pst*I fragments D and J indicated at the lower left side of the map. In the *lower* part, the *rpo* genes relative to these restriction fragments, are shown in schematically enlarged linear form

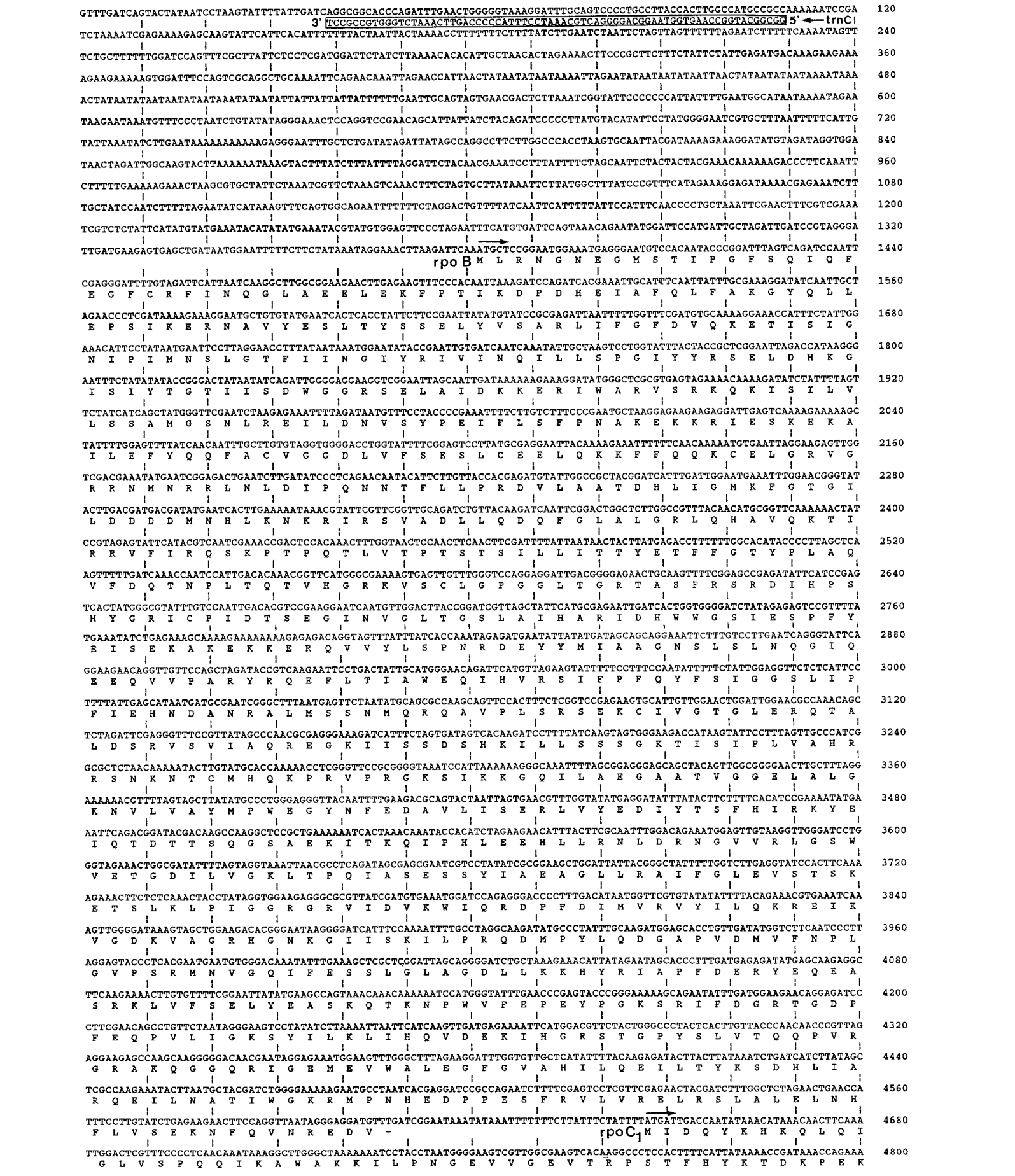

Fig. 2 (Continuation, see page 382)

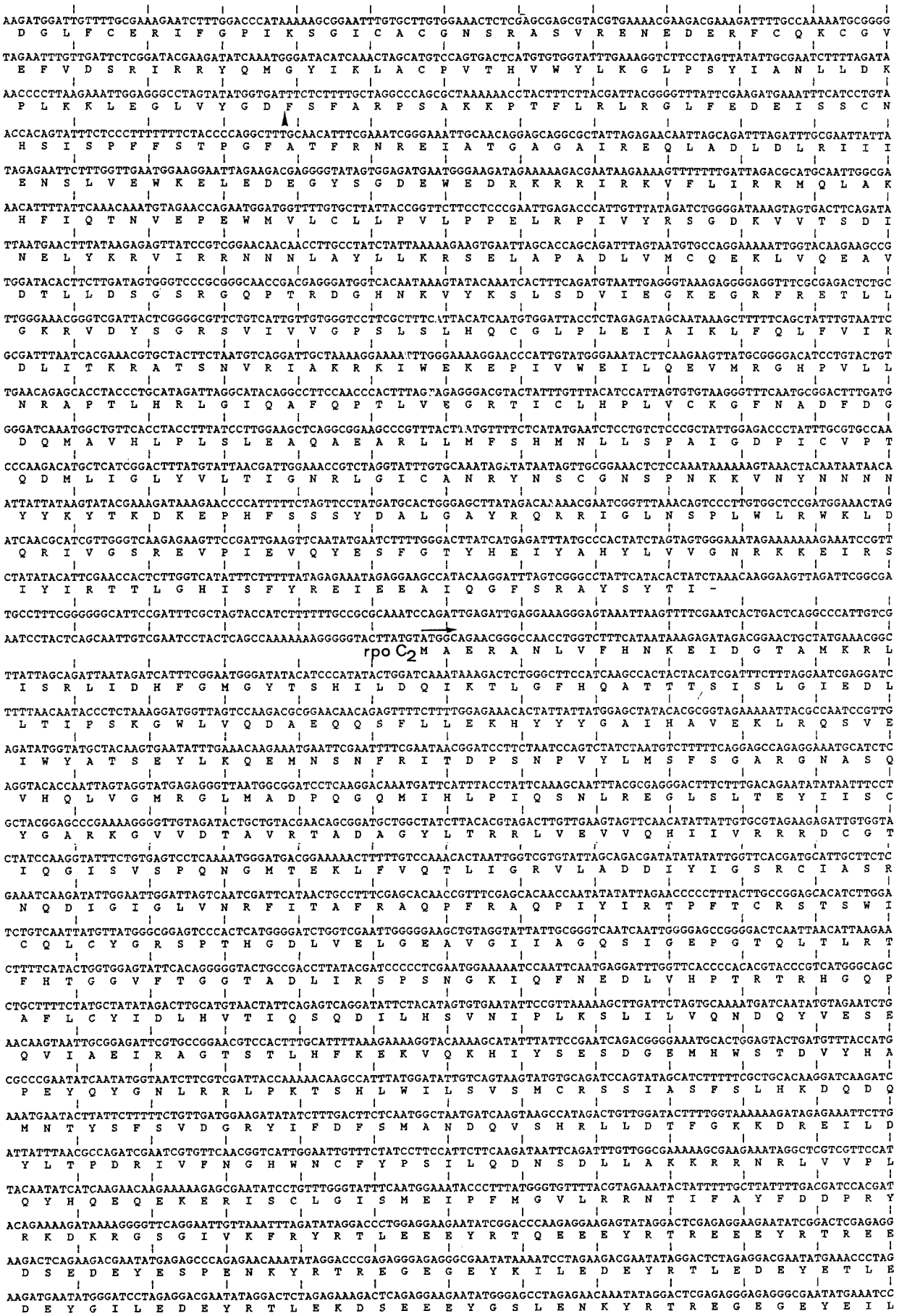

Fig. 2

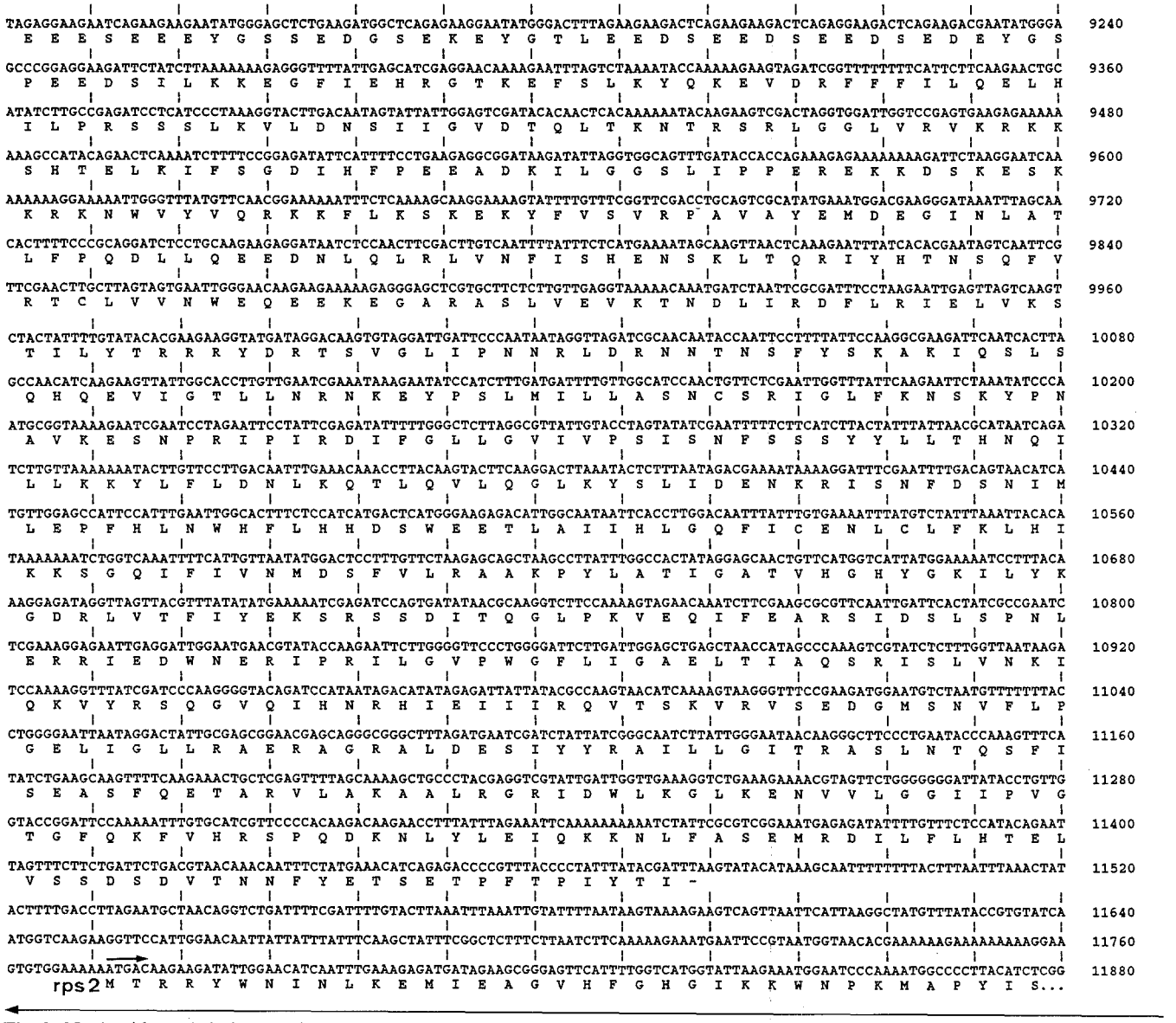

Fig. 2. Nucleotide and derived amino acid sequence of the maize chloroplast *rpo*B/C_1/C_2 operon. The start site of each of the coding regions is indicated. Note that *trn*C upstream of *rpo*B is transcribed from the opposite DNA strand to that of the protein genes. The position of the intron in spinach, tobacco and liverwort is indicated by a *vertical arrow*. The 5′-terminal region of the *rps*2 gene is presented at the distal end of the sequences. The complete sequence of the latter gene has been described (Igloi et al. 1990)

Probes for the 5′ end of *rpo*C and the 3′ end of *rpo*B gave signals within a 14 kb *Pst*I D fragment (plasmid pZmc354) of which the position in the genome is shown in Fig. 1. This fragment was subcloned after *Bam*HI/*Bgl*II or *Eco*RI digestion into M13 or pUC vectors and subjected to sequence analysis using either the universal M13 primer or one of a set of specific primers. It was found that the orientation of the genes is opposite to that observed in spinach and tobacco but is similar to that of liverwort. As a consequence, the terminal 1791 bp of *rpo*C_2 together with *rps*2 lay within the plasmid pZmc355 which contains the neighbouring genomic *Pst*I J fragment of 4.7 kb (Fig. 1). Unlike all other fragments, no overlapping sequence could be obtained for the *Pst*I site separating the two fragments inserted in the plasmids pZmc354 and pZmc355. However, the nucleotide sequence at this point shows no break in the reading frame, and a high degree of similarity between the plastids means that it is unlikely to be a source of sequence uncertainty. The sequence, encompassing 11 880 bp (Fig. 2) includes the flanking *trn*C gene, 1272 bp upstream of but of opposite orientation to four closely spaced open reading frames which by homology to the other plastid DNAs correspond to *rpo*B, *rpo*C_1, *rpo*C_2 and *rps*2.

These results show that, as is the case for wheat and rice, the region of the maize chloroplast DNA encompassing *rpo*B, C_1, C_2 is inverted with respect to spinach and tobacco, of which the gene order is reported to be evolutionarily older (Palmer 1985). In view of the second inversion of 10 kb in wheat and rice, it is likely that the closely related maize will also have this feature (Rodermel and Bogorad 1988).

The rpo*B gene and its translation product*

The ORF separated by 1272 bp from the oppositely oriented *trn*C gene (Meinke et al. 1988) represents the first

Table 1. Comparative summary of the characteristics of the *rpo*B/C_1/C_2 genes and their translation products

		Maize	Rice	Spinach	Tobacco	Liverwort	*E. coli*	Cyanobact.
*rpo*B	bp	3225	3225	3210	3210	3195	4026	–
	aa	1075	1075	1070	1070	1065	1342	–
	pI	9.35	9.45	8.11	8.14	9.16	4.93	–
	Mol. Mass (kDa)	121.6	121.6	120.9	120.5	120.4	150.6	–
	Intergenic Dist. to trnC	1272	1084	1099	1281	138	–	–
	Homology*: nt		96.6	83.9	83.9	68.5	46.0	–
	aa		98.0	81.0	81.2	63.6	37.5	–
rpoC_1	bp	2049	2046	2031	2040	2052	1734[1]	1863
	aa	683	682	677	680	684	578[1]	621
	pI	9.41	9.08	8.43	8.93	9.98	(9.19)[1]	6.31
	Mol. Mass (kDa)	78.3	78.1	78.1	78.4	78.9	65.5	70.2
	Intergenic Dist. to rpoB	40	40	28	29	33	79	134
	Homology*: nt		96.1	80.4	80.9	69.5	–	–
	aa		96.5	78.0	77.7	61.9	36.2[1]	47.2
	Length of Intron bp	0	0	756	739	596	–	0
rpoC_2	bp	4581	4539	4038	4173	4158	2487[2]	≠
	aa	1527	1513	1361	1391	1386	829[2]	≠
	pI	6.18	6.87	9.69	6.96	10.27	(5.36)[2]	–
	Mol. Mass (kDa)	176.1	173.8	154.7	157.3	160.1	89.6	–
	Intergenic Dist. to rpoC_1	201	202	151	149	76	–	93
	Homology*: nt		92.0	77.3	76.7	57.0	–	–
	aa		88.5	71.0	69.7	44.7	27.6	–
	Insert Rel. to *E. coli* (aa)	754	739	602	618	604	–	–

[1] N-terminal fragment of β'
[2] C-terminal fragment of β'
≠ incomplete sequence
* % rel. to Maize

potential messenger on this restriction fragment, having 1075 codons and corresponding to a basic protein (pI 9.35) of 121.6 kDa, compared to the acidic β subunit of *E. coli* RNA polymerase of 155 kDa (pI 4.93) (Ovchinnikov et al. 1981). The similarity at the nucleotide and amino acid level with *rpo*B genes of other organisms (Table 1) strongly supports its identity as the *rpo*B gene. Preparations of RNA polymerase from maize chloroplasts contain polypeptides with apparent molecular masses of 120 and 140 kDa (Kidd and Bogorad 1980). It remains to be seen which of these components is the *rpo*B encoded product with the calculated molecular mass of 121.6 kDa.

Figure 3 shows an alignment of the derived amino acid sequences of five plastid β subunits (maize, rice, spinach, tobacco, liverwort) together with that of *E. coli* and, by extrapolation, with the *rpo*B gene product of *Salmonella typhimurium* which shows 98.5% homology with the *E. coli* β subunit (Lisitsyn et al. 1988). In the schematic comparison with the *E. coli* protein (Fig. 4), additional regions are indicated of homology with the corresponding proteins of archaebacteria (Berghöfer et a. 1988; Leffers et al. 1989), yeast (Sweetser et al. 1987) and *Drosophila* (Falkenburg et al. 1987). Compared with the bacterial β subunit, nine deleted segments in the chloroplast sequence are clearly visible (Fig. 4), which are located towards the termini of the protein and result in a reduction in the size of the plastid protein by 270 amino acids. The largest deletion is responsible for the loss of more than 100 amino acids and corresponds to a C-terminal domain which has been demonstrated to the functionally redundant (Glass et al. 1986).

An examination of the amino acid sequence for features that have been made responsible for certain functional properties has focussed on a number of sites of potential interest. The N-terminal third is rather featureless in this respect, having a number of small deletions and, in general, being less similar to the *E. coli* β subunit than the remaining molecule. There are, nevertheless, two short homology blocks (A and B) with the yeast RPB2 protein (the second largest subunit of RNA polymerase II; Sweetser et al. 1987) which together with the overall conservation of the N-terminal domain tends to suggest a structural role for this polypeptide segment.

From approximately amino acid 390 (maize numbering) to the C-terminus the chloroplast sequence is closely aligned with that of *E. coli*, with the exception of one long and several short deletions. There are also numerous homology blocks in common with the corresponding subunits of yeast (Sweetser et al. 1987), *Drosophila* (Falkenburg et al. 1987) and archaebacteria (Berghöfer et al. 1988; Leffers et al. 1989). Some of these are characterized by amino acid motifs which are highly conserved between organisms and to which certain functional properties have been ascribed. In maize chloroplasts, at position 418, one finds a dodecapeptide whose two histidine and single cysteine residues are conserved in all plastids, in *E. coli*, archaebacteria and in yeast (homology block D). A juxtaposition of imidazole/thiol groups (also found in *E. coli* DNA ligase and T7 RNA polymerase) (Lisitsyn et al. 1988) has been implicated in the catalytic activity of kinases and aminoacyl-tRNA synthetases (Pai et al. 1977; Carter et al. 1984). On the other hand, one conserved histidine is replaced in *Drosophila* by leucine. The only other conserved cysteine at position 627 in the maize sequence does not fall within one of the eukaryotic homology boxes. Nevertheless, a corresponding cys-

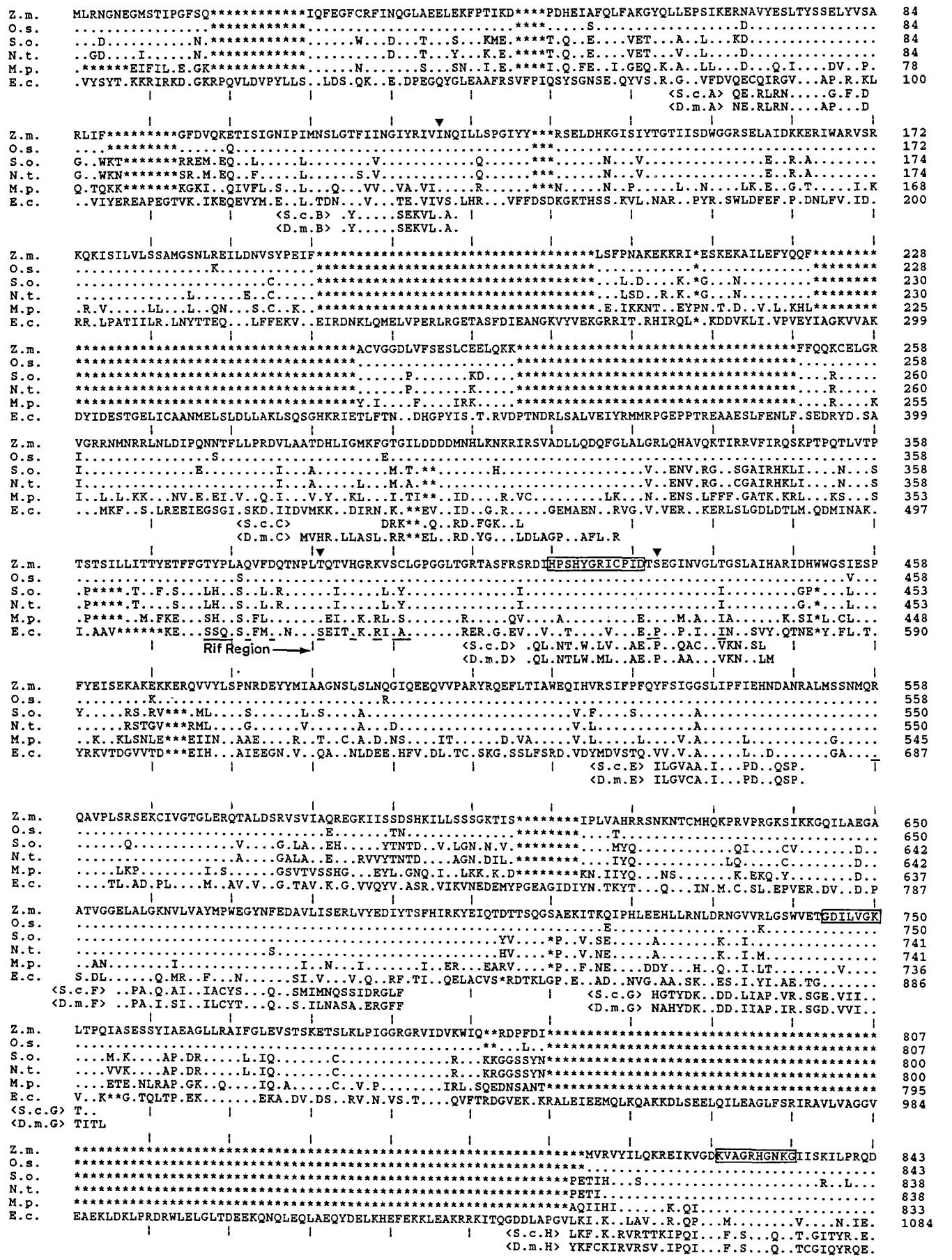

Fig. 3 (Continuation, see page 386)

```
Z.m.    MPYLQDGAPVDMVFNPLGVPSRMNVGQIFESSLGLA**********************************GDLLKKHYRIAPFDERYEQEASRKL********   904
O.s.    .......T.....................................**********************************.....................********   904
S.o.    .......R....I.................C.....**********************************.S..DR.....................********   899
N.t.    .......RS....................C.....**********************************.S..DR.....................********   899
M.p.    ..F....T.I..ILS..............CL....**********************************.SF.H.N...I.......R......********   894
E.c.    ...DEN.T...I.L.........I...L.TH..M.AKGIGDKINAMLKQQQEVAKLREFIQRAYDL.ADVRQKVDLST.SDEEVMRLAEN.RKGMPIAT   1184
<S.c.H> ..FTAE.IVP.LII..HAI....T.AHLI.CL.S
<D.m.H> .AFTCE.LAP.III..HAI....TI.HLI.CLQ.

Z.m.    *VFSELYEASKQTKNPWVFEPEYPGKSRIFDGRTGDPFEQPVLIGKSYILKLIHQVDEKIHGRSTGPYSLVTQQPVRGRAKQGGQRIGEMEVWALEGFGV   1003
O.s.    *.....................................................................................................   1003
S.o.    *..........A.....................I..NP.........D......S.H.A......L.........V.............               998
N.t.    *..........A.....................N......I...P.........D......S.H.A......L.........V.............               998
M.p.    *......K...K.T...L...DN...N.LI.....EI....IT...A.M........D...A..S...A......L...SRR....V.............               993
E.c.    P..DGAK..EIKELLK*LGDLPTS.QI.LY.....EQ..R..TV.YM.M...N.L..D.M.A....S........LG.K.QF....F.........AY.A   1283
                <S.c.I> GYQSR.FEVMYN.H..KKLMAQIFF.PT.YQR.R.M..D...A.AR..MQVL.R...E..SRD..L.F....RDCMIAH.A
                <D.m.I> GYHLR.NEVMYN.H..RKINAQ.FL.PT.YQR.K.M..D...S.AR..VQILVR..ME...RD..L.F....RDCQISH.A

Z.m.    AHILQEILTYKSDHLIARQEILNATIWGKRMPNHEDPPESFRVLVRELRSLALELNHFLVSEKNFQVNREDV   1075
O.s.    ..............................V.............................Q........E.   1075
S.o.    ......M.......IK....V.GT..I.GTI..P..A.....L...................R...I..MEA   1070
N.t.    ......M.......IR....V.GT..I.GTI..P..A.....L...................I..KEA   1070
M.p.    .Y....M..I....IR..Y.V.G.IVT.EPI.KPNTA....KL...........I..VIIC...LKLKLKEI   1065
E.c.    .YT...M..V...DVNG.TKMYKNIVD.NHQMEP*GM....N..LK.I...GINIELEDE              1342
<S.c.I> .SF.K.R.MEA..
<D.m.I> .QF.R.R.FEV..
```

Fig. 3. Alignment of the derived amino acid sequences for the RNA polymerase β subunit. Included in the alignment are the proteins from maize chloroplasts (Z.m.), rice (Hiratsuka et al. 1989) (O.s.), spinach (Hudson et al. 1988) (S.o.), tobacco (Shinozaki et al. 1986) (N.t.), liverwort (Ohyama et al. 1986) (M.p.), and *Escherichia coli* (Ovchinnikov et al. 1981) (E.c.). Also shown are regions of homology with the eukaryotic sequence of the second largest subunit of RNA polymerase II from yeast (Sweetser et al. 1987) (S.c.) and from *Drosophila* (Falkenburg et al. 1987) (D.m.). Amino acid identity is denoted by *points* and deletions (for the purpose of optimal alignment) by *asterisks*. The numbering of actual amino acid positions is given at the right margins. Sequence motifs discussed in the text are *boxed*. Positions that by mutation lead to rifampicin resistance in *E. coli* (Jin and Gross 1988) are *underlined* in the Rif region (positions 508–572) and at position 687. Positions 121, 389 and 431 of the maize sequence which coincide with *E. coli rif*R positions but deviate from the *E. coli* wild type are marked by ▼

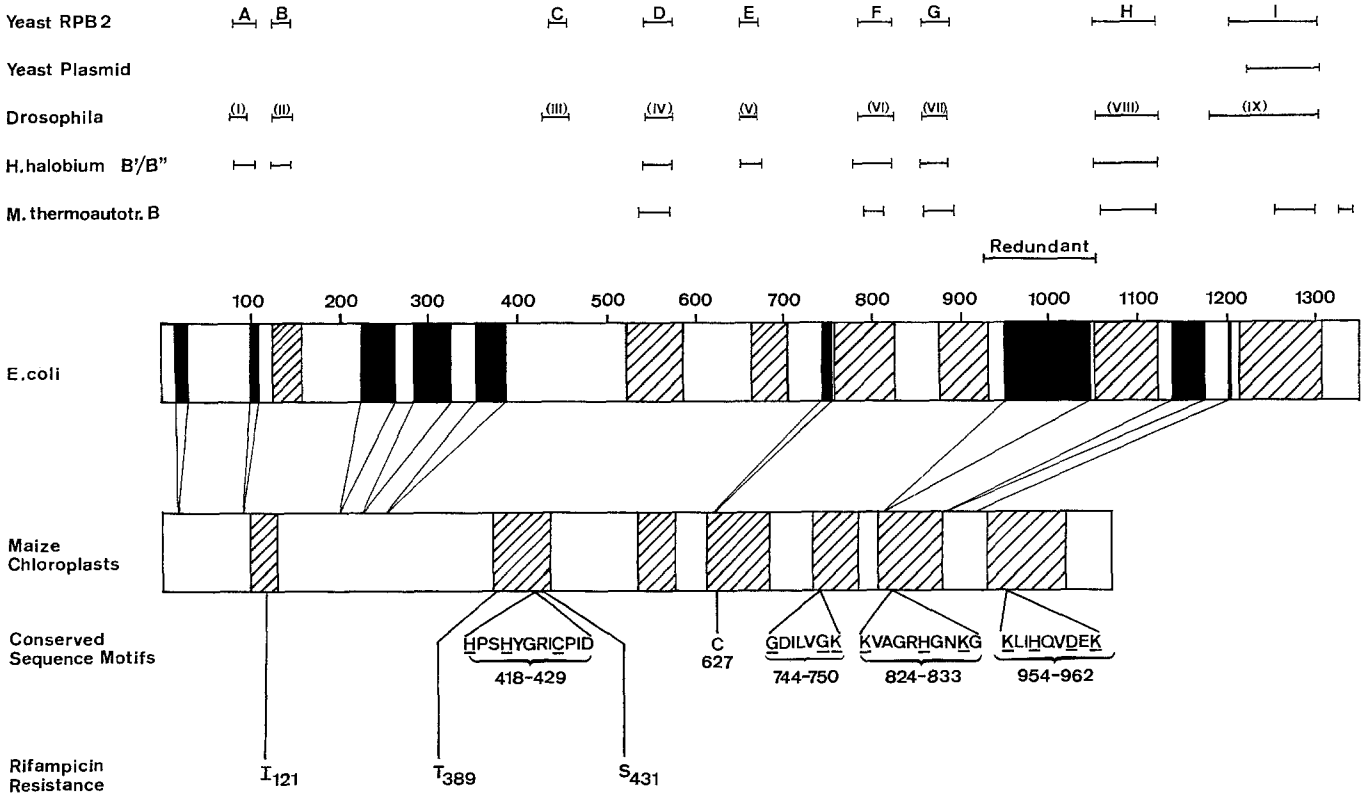

Fig. 4. Schematic comparison of the maize chloroplast RNA polymerase β subunit with the corresponding proteins from other organisms. Insertions derived from the alignments in Fig. 3 are marked by *black boxes*. Regions of significant homology between maize chloroplast and *E. coli* are denoted by *shading*. Other blocks of homology are included as *bars* and labelled according to the yeast (or *Drosophila*) nomenclature A–I (I–IX). Individual amino acids within the conserved sequence motifs, which are relevant to the discussion are *boxed*. The sources of the compared sequences are as given in the legend to Fig. 3

teine may be present in these organisms and may be involved in the interaction with DNA, although disulphide bridge formation with cysteine 426, destroying the potential thiol/imidazole catalytic centre, cannot be ruled out. As would be expected from the experimental evidence pointing to the binding of DNA by the β' subunit (Fukuda and Ishihama 1974), a zinc finger motif, discovered in the yeast β subunit (Sweetser et al. 1987) is not observed in either the *E. coli* or chloroplast β subunit.

At position 744 in the maize sequence, and falling within box G in the yeast comparison, we find the nucleotide-binding-type of motif GXXXXGK which is present in all plastids, *E. coli* (Glass et al. 1986), methanobacteria (Berghöfer et al. 1988) and, with an additional amino acid between the glycines, in yeast and *Drosophila*. This pattern of amino acids may represent a G-protein-type of nucleotide binding site (Wooley and Clark 1989). Of the other sites mentioned in the literature (Rozen et al. 1989; Hartmann et al. 1988) as possible NTP binding sites or as being involved in phosphodiester bond formation, after eliminating those not conserved in the plastid sequences, two pairs of lysines (824, 832 and 954, 962), each flanking a histidine, are found to be conserved in plastids, *E. coli* and methanobacteria. The first set of lysine/histidine/lysine is also found in yeast and *Drosophila* (homology block H) while the second is present with a lysine to arginine substitution in both these species (homology block I).

The second motif is also found in the extrachromosomal RNA polymerase of the yeast *Kluyveromyces lactis* (Wilson and Meacock 1988). Within this second motif, the aspartic acid-glutamic acid pair in maize and rice chloroplasts corresponds to the aspartic acid-aspartic acid sequence in the other plastids, *E. coli*, yeast, *Drosophila* and the extrachromosomal yeast polymerase, and has been suggested as being important in the action of many polymerases (Argos 1988) (although in methanobacteria the corresponding sequence is threonine-aspartic acid; Berghöfer et al. 1988). The histidine in this section of the sequence has been identified as reacting with a nucleotide affinity label (Lisitsyn et al. 1988). Lysine 816, which was labelled in *E. coli* (position 1057) is conserved in all plastids but is not found in yeast or *Drosophila*. On the other hand, lysine 824, also a target for affinity modification in *E. coli* (position 1065; Lisitsyn et al. 1988) is conserved in all organisms examined here. Other lysines, labelled by site-specific reagents in *E. coli* RNA polymerase, are not conserved in the plastid protein.

Despite the overall similarity of the chloroplast β-subunits to the *E. coli* protein, plastid RNA polymerases are resistant to the antibiotic rifampicin (Briat et al. 1979; Kidd and Bogorad 1980). Mutations in *E. coli*, leading to rifampicin resistance are localized exclusively on the β-subunit (Hartmann et al. 1988) and have been mapped in detail (Jin and Gross 1988). Of the 11 sites whose mutation individually results in rifampicin resistance only two (threonine 389 and serine 431) have amino acids that are conserved between the chloroplast protein but are, at the same time, "mutated" with respect to *E. coli*. Apart from the cluster of mutations between amino acids 508 and 687 in *E. coli* which result in rifampicin resistance and other related phenotypes (Jin and Gross 1989), there is a single site at position 146 that may also be involved in rifampicin binding (Lisitsyn et al. 1988). This amino acid is also present in all the plastid sequences in a "mutated" form as isoleucine 121 (in maize). It remains to be tested, by site directed mutagenesis, which of these positions (I_{121}, T_{389} or S_{431}) confer the rif resistance to plastid RNA polymerase or whether perhaps other mutated positions also contribute to this resistance in a concerted mode.

The rpoC_1 *gene and its translation product*

The second ORF of the *rpo* gene cluster, coding for 683 amino acids is initiated by an AUG codon located 40 bp downstream of the 3′ terminus of the *rpo*B gene (Fig. 2). The molecular mass of 78.3 kDa calculated for this polypeptide is in the vicinity of the apparent molecular masses of 70, 75 and 85 kDa of polypeptides observed in RNA polymerase preparations from maize chloroplasts (Kidd and Bogorad 1980). Further analysis is necessary to decide which of these polypeptides is encoded by the *rpo*C_1 gene. The 40 bp intergenic region between *rpo*B and C_1 is comparable in all examined organisms except cyanobacteria (Xie et al. 1989) where the distance is slightly greater (Table 1). In fact, the same reading frame in maize extends back through the intergenic *rpo*B-*rpo*C_1 region, overlapping in an out-frame-mode the two terminal *rpo*B amino acids and initiating with a Met. Similarly, in tobacco, the ORF extends four codons upstream from the AUG used for the sequence alignments. However, by analogy with the other plastid C_1 sequences, it is assumed that the maize C_1 gene product does not bear this N-terminal extension. A remarkable feature of the *rpo*C_1 gene is the loss of the intron found in spinach, tobacco and liverwort. The closely related rice with its high homology to maize, also lacks this intron (Shimada et al. 1990). Splicing of the precursor RNA in spinach, tobacco and liverwort leads to a derived protein sequence across the intron boundaries which is virtually identical to maize and rice in this region. This also confirms the more favoured of the two potential splice sites in spinach (Hudson et al. 1988). The site of the excision of the intron during evolution has, it seems, been so precisely conserved, that not only has the reading frame remained intact but the amino acid sequence, itself, has been undisturbed by insertions or deletions. Furthermore the sequence conservation during evolution, on going from dicots to monocots suggests, despite the loss of the intron, a highly specific mechanism such as reverse transcription of the spliced RNA followed by reintegration into the genome at the same location.

Despite the fact that the derived amino acid sequences can be well aligned with each other (Fig. 5), a number of blocks of variability exist of which the most obvious lies between amino acids 557 and 582 where the dissimilarity extends from a slight deviation in the rice sequence to a significant deletion in the liverwort. It is also a region where the maize and rice sequences are characterized by a stretch of 15 basic amino acids (including asparagines) out of 28. Clustering of basic amino acids has also been observed in the *E. coli* β' subunit (Ovchinnikov et al. 1982) but to a more limited extent and at different locations.

The similarity of the chloroplast sequence with that of the N-terminal section of *E. coli* β' subunit was previously noted by Hudson et al. (1988). The similarity ends prior to this section of the sequence which, in fact, also corresponds to the eukaryotic homology block E (Jokerst et al. 1989) and is the only eukaryotic consensus block with no counterpart in prokaryotes (Figs. 5, 6). On the other hand, the alignments are virtually colinear with only one signifi-

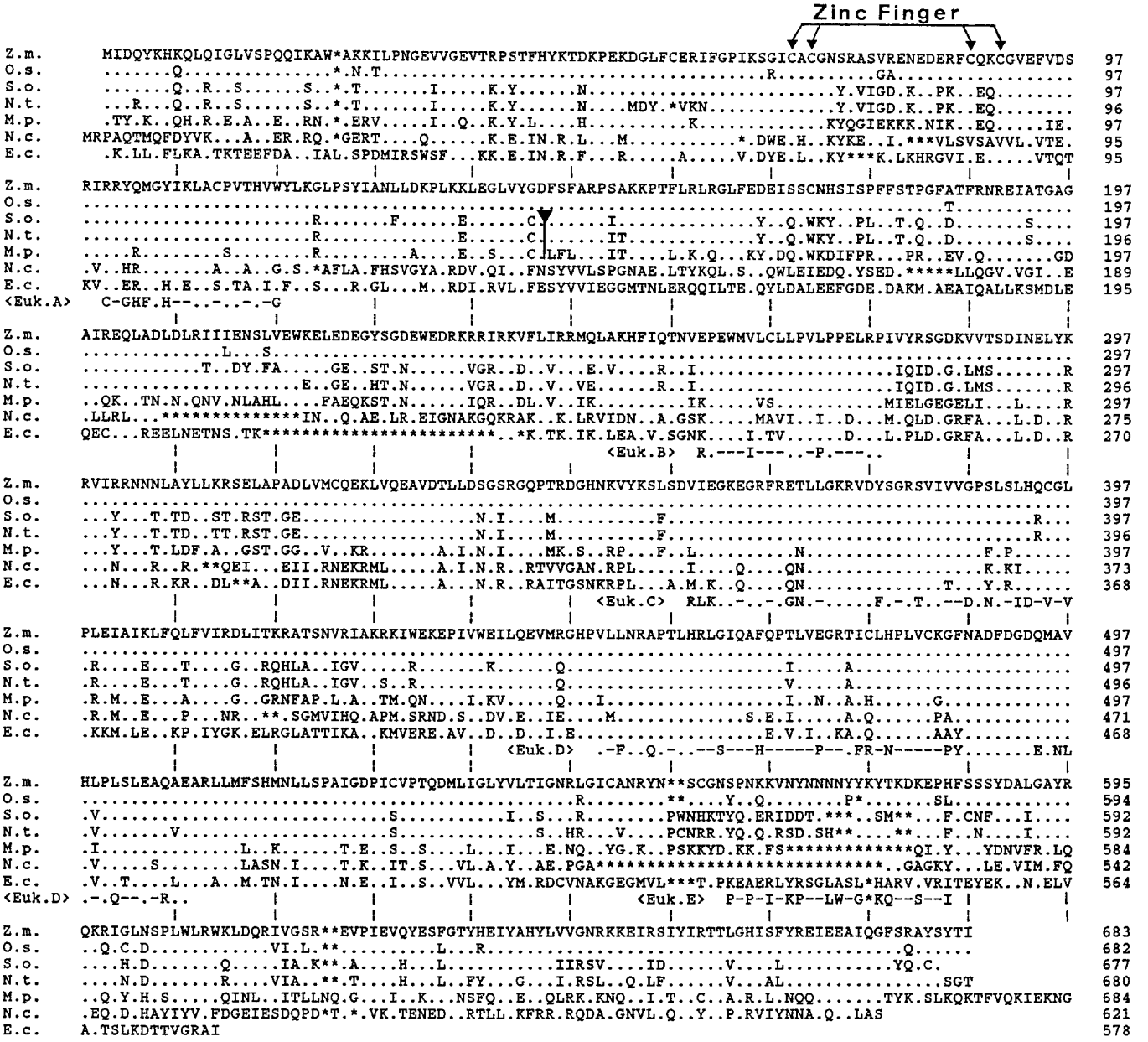

Fig. 5. Alignment of the derived amino acid sequences for the RNA polymerase β' subunit. In addition to the proteins from the organisms described in the legend to Fig. 3, the sequence from the cyanobacterium *Nostoc commune* (Xie et al. 1989) (N.c.) is also included. The comparison covers the first 578 amino acids of the *E. coli* β' subunit (Ovchinnikov et al. 1982). The eukaryotic homology consensus blocks (Jokerst et al. 1989) A–E are depicted at the appropriate positions with *dashes* representing gaps in the eukaryotic consensus sequences. The exon borders of the tobacco and spinach sequences are marked by ▼. Other symbols are as in Fig. 3

cant deletion in *E. coli*. In cyanobacteria, which have been associated with evolutionary relationships to chloroplasts (Gray 1989) and where the gene for the subunit γ was recently characterized as being homologous with the N-terminal region of β' and, therefore, corresponds to the chloroplast *rpo*C_1 gene (Xie et al. 1989), the homology with the plastid sequence is 47.2% and with *E. coli*, 36.2% (Table 1). The amino acid composition, however, gives rise to a much more acidic protein than in the other organisms compared. The cyanobacterial sequence in characterized further by two deletions, with respect to chloroplasts, of which one overlaps with the deletion in *E. coli* β'. While, as is the case for the monocots, no intron is found in cyanobacteria, the protein sequence at the equivalent position is not dissimilar, particularly at the 5′ end of the splice site (Fig. 5).

Although the specific function of the *E. coli* β' subunit still remains to be defined, it is, on the basis of available experimental evidence, frequently associated with an involvement in DNA binding (Fukuda and Ishihama 1974). It may, therefore, be significant that near the *N*-terminus, between positions 69 and 90 in maize, two pairs of cysteines are located, which have been conserved within the plastids and also in *E. coli* (Fig. 6, insert a). It has been postulated (Hudson et al. 1988) that his motif could represent a Zn-finger-type of structure, as found in other eukaryotic tran-

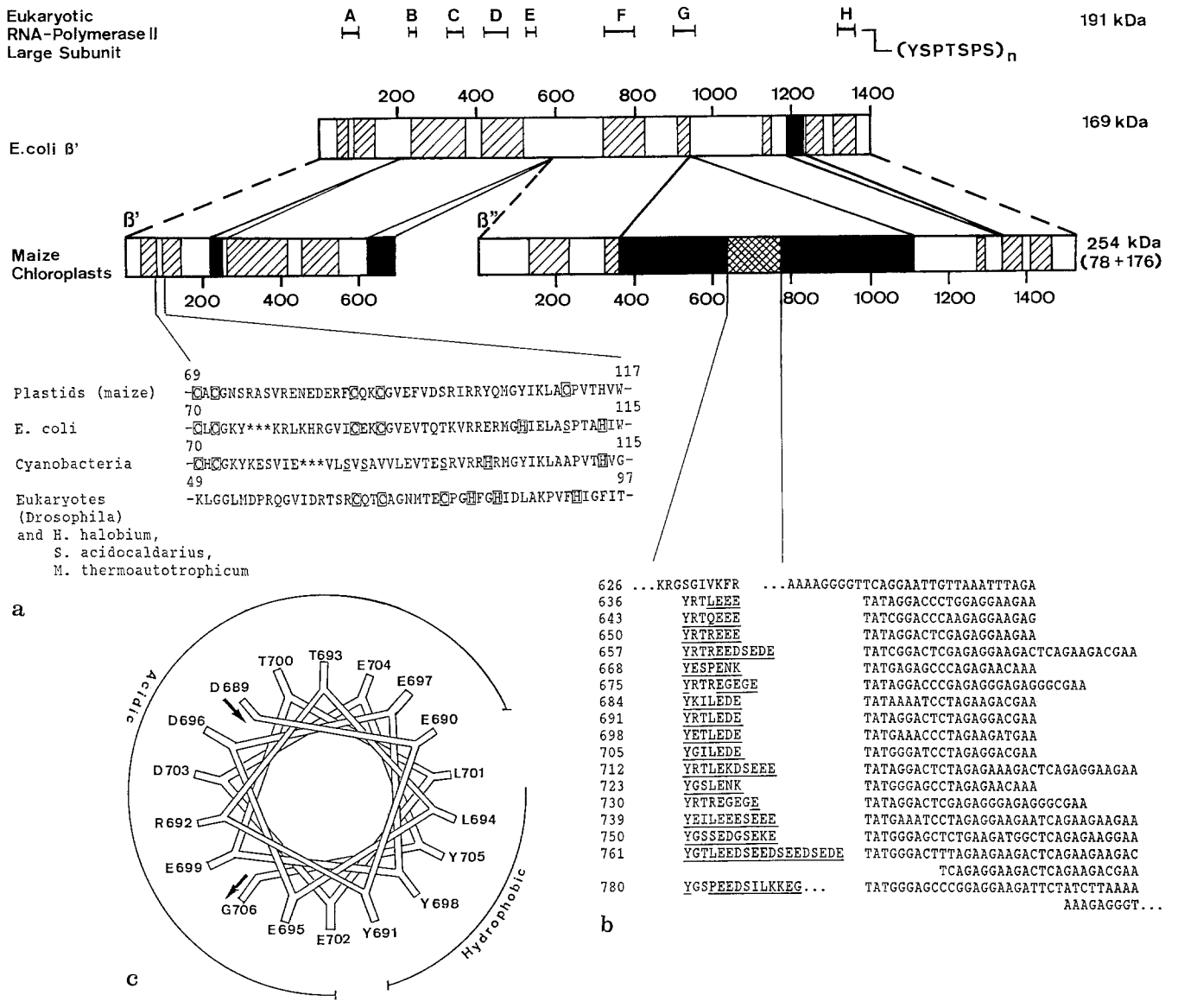

Fig. 6a–c. Schematic comparison of the maize chloroplast RNA polymerase *β′* and *β″* subunits with the corresponding proteins from other organisms. Insertions derived from Figs. 5 and 7 are marked by *black boxes*. Regions of significant homology between maize chloroplast and *E. coli* are denoted by *shading*. Eukaryotic homology blocks are indicated by *bars* and labelled A–H. The monocotyledon specific insert in *β″* is marked by *cross-hatching*. **a** The position of the potential zinc finger structures in *β′* is indicated and enlarged to permit sequence comparisons. Within this stretch of amino acids, cysteines and histidines are *boxed* and certain serines (see text) are *underlined*. In the three archaebacteria *Halobacterium halobium* (Leffers et al. 1989), *Sulfolobus acidocaldarius* (Pühler et al. 1989) and *Methanobacterium thermoautotrophicum* (Berghöfer et al. 1988), the position of the cysteines and histidines correspond exactly to the eukaryotic (*Drosophila*) case. Deletions are denoted by *asterisks*. **b** The heptameric repeats within the monocot specific insert together with the corresponding nucleotide sequence. Predicted α-helical stretches (Garnier et al. 1978) are *underlined*. **c** Helical wheel representation of the central portion of the heptameric repeat structures, covering amino acids 689–706. The position of the hydrophobic and acidic faces of the helix are indicated

scription factors (Klug and Rhodes 1987; Struhl 1989). However, the strength of the homology argument is weakened by the loss of the second pair of cysteines in cyanobacteria where they are replaced by valine. The pair of serines that appear in cyanobacteria within this stretch of amino acids could be considered as taking on the role of the missing cysteines, although the existence of serines, while postulated in some cases (Berghöfer et al. 1988) has yet to be demonstrated in true Zn finger proteins. This region of the C_1 sequence is part of a weak homology box in terms of eukaryotic polymerases. However, a limited degree of similarity does exist between *E. coli β′* and a zinc-finger-type motif suggested for the A subunit of *Halobacterium halobium* (Leffers et al. 1989) together with a number of eukaryotic polymerase subunits (Jokerst et al. 1989) (encompassing cysteines 85, 88 and histidine 104 in *E. coli*). The only available partner in such a structure for histidine 104 would seem to be either Ser 109 (which, in fact, corresponds to a conserved Cys in plastids) or a conserved His 113. In plastids and cyanobacteria even the residue corresponding to His 104 has been replaced by tyrosine, although in the latter organism a single histidine residue does

exist in the vicinity at position 100. An additional pair of cysteines in eukaryotes and archaebacteria located 15 positions downstream of His 91 is not found in plastids and *E. coli*.

Whether the alignment summarized in insert (a) of Fig. 6 has a fundamental significance (in view of a rather well-defined consensus motif for Zn fingers) remains to be seen since it is not altogether compatible with a highly specific protein-DNA interaction. On the other hand, DNA binding of RNA polymerase during the elongation phase would not allow any sequence specificity and the reduced consensus of the zinc finger motif may reflect the necessity for relaxation from sequence specificity.

Another potential site of interaction with DNA was postulated for eukaryotic RNA polymerases (Allison et al. 1985) on the basis of alignment of a section of these sequences with the helix-turn-helix domain of DNA polymerase I whose tertiary structure is known (Ollis et al. 1985a). The prokaryotic DNA polymerase I seems to have little significant homology with eukaryotic DNA polymerase (Ollis et al. 1985b). Nevertheless, certain conserved amino acids have been identified in eukaryotic RNA polymerases (Allison et al. 1985), which suggests that this part of the polypeptide chain (homologous to the regions between amino acids 379–409 and 501–509 in maize) could form a homologous tertiary structure and, accordingly, have a similar DNA binding function. However, by including these plastid sequences in the comparison with the helix domain of bacterial DNA polymerase I, the homology is diminished to a point where its significance becomes questionable. Also, the secondary structure prediction for the maize chloroplast sequence by the method of Garnier et al. (1978) does not envisage a helix-turn-helix motif in this region (data not shown).

On the other hand, a remarkably strong homology (considering the evolutionary distance involved) between the regions at positions 360–403 and 449–509 (in maize) and a segment of the extrachromosomal yeast RNA polymerase from *K. lactis* (approximately equivalent in position to the eukaryotic homology blocks C and D; see Fig. 5) is confirmed by the corresponding sequence from liverwort (Wilson and Meacock 1988). These sections of the sequence are also highly conserved in cyanobacteria (Xie et al. 1989) and indicate with considerable certainty the functional importance of this structure.

The rpoC$_2$ *gene and its translation product*

The gene for the largest subunit of chloroplast encoded RNA polymerase, *rpo*C$_2$, follows on from *rpo*C$_1$ with an intergenic region of 201 bp (Fig. 2). This compares with 151 bp in spinach (Hudson et al. 1988) (Table 1) whereby the initiation codon in maize corresponds to the second methionine codon in the spinach sequence, four codons downstreams of the first AUG in the spinach ORF. Translational initiation at this second AUG is favoured by Hudson et al. (1988) in view of the homology with the liverwort sequence and because of the availability of a good Shine-Dalgarno sequence. However, in maize, this putative Shine-Dalgarno sequence, or any similar sequence, is not conserved although in this case there is no alternative site for translational initiation prior to the first AUG of the ORF. The molecular mass of 176 kDa predicted for the *rpo*C$_2$ encoded polypeptide correlates well with the 180 kDa polypeptide observed in RNA polymerase preparations from maize chloroplasts (Kidd and Bogorad 1980).

The chain length of the protein derived from the *rpo*C$_2$ genes varies dramatically from 1527 and 1513 in maize and rice, respectively, through the dicots spinach, tobacco and liverwort with 1361, 1391 and 1386, to the corresponding section of the *E. coli rpo*C with only 829 amino acids and is indicative of a characteristic feature with evolutionary aspects. The section of the sequence responsible for the variability lies in *E. coli* at position 369 (Fig. 7), where all plastid C$_2$ gene products carry a long insert of approximately 600 amino acids, followed by a region of approximately 130 amino acids with a complete lack of significant homology compared to *E. coli*. Within the plastid specific insertion, the monocots rice and maize have an additional insertion of about 150 amino acids. The actual existence of these exceptional regions within the *rpo*C$_2$ gene, as opposed to a sequencing/cloning artefact, is confirmed not only by the conservation in two independently sequenced organisms, but in the case of maize, by amplification of the appropriate segment from total chloroplast DNA by polymerase chain reaction (PCR) and analysis of the product by restriction nuclease digestion. The length of the amplified product and the restriction patterns are in complete agreement with the determined nucleotide sequence (data not shown). Apart from these major structural differences, together with a chloroplast specific deletion with respect to *E. coli* of about 40 amino acids near the C-terminus, the alignment of the plastid sequences with *E. coli* and the available fragment of the cyanobacterial protein (Xie et al. 1989), is straightforward. Once again, regions of homology are clustered and tend to occur in parallel with the eukaryotic homology blocks (Fig. 6). There are, however, two regions of homology between chloroplasts and *E. coli* which do not have counterparts in the eukaryotic system (between homology blocks G and H; Jokerst et al. 1989).

In view of the conservation of, presumably structurally and functionally significant sections, it is all the more remarkable that such a large additional portion of sequence has evolved in the central part of the plastid *rpo*C$_2$ product by which the eukaryotic homology block G is divided into two sections. Further evolution from dicots to monocots seems to have led to a yet greater expansion of this insert by ca. 150 amino acids. The function of the chloroplast insert, let alone that of the monocot specific insert, is not known. Indeed, there are in the literature very few reports concerned with a functional analysis of the C-terminal domain of the β' subunit in prokaryotes. In eukaryotes, the C-terminal domain of the largest subunit of RNA polymerase II has been the subject of intense study (Allison et al. 1985; Jokerst et al. 1989; Evers et al. 1989a, b) largely in view of the unusual heptameric repeat structure found at the very end of the protein and the potential for transcriptional control by phosphorylation of this tail (Cisek and Corden 1989). The additional 150 amino acids in maize can be broken down into a set of 13 almost identical repeated heptameric units (Fig. 6, insert b). These are characterized by (i) tyrosine as the first amino acid (always – 16 times – coded by TAT, cf. the codon usage in *rpo*B and *rpo*C$_1$ which would predict four TAC codons), (ii) five leucines at heptameric intervals, and (iii) an extremely high content of glutamic and aspartic acids whose negative charges are partially offset by the presence of an arginine in nearly all the heptamers. The acidic residues of the hepta-

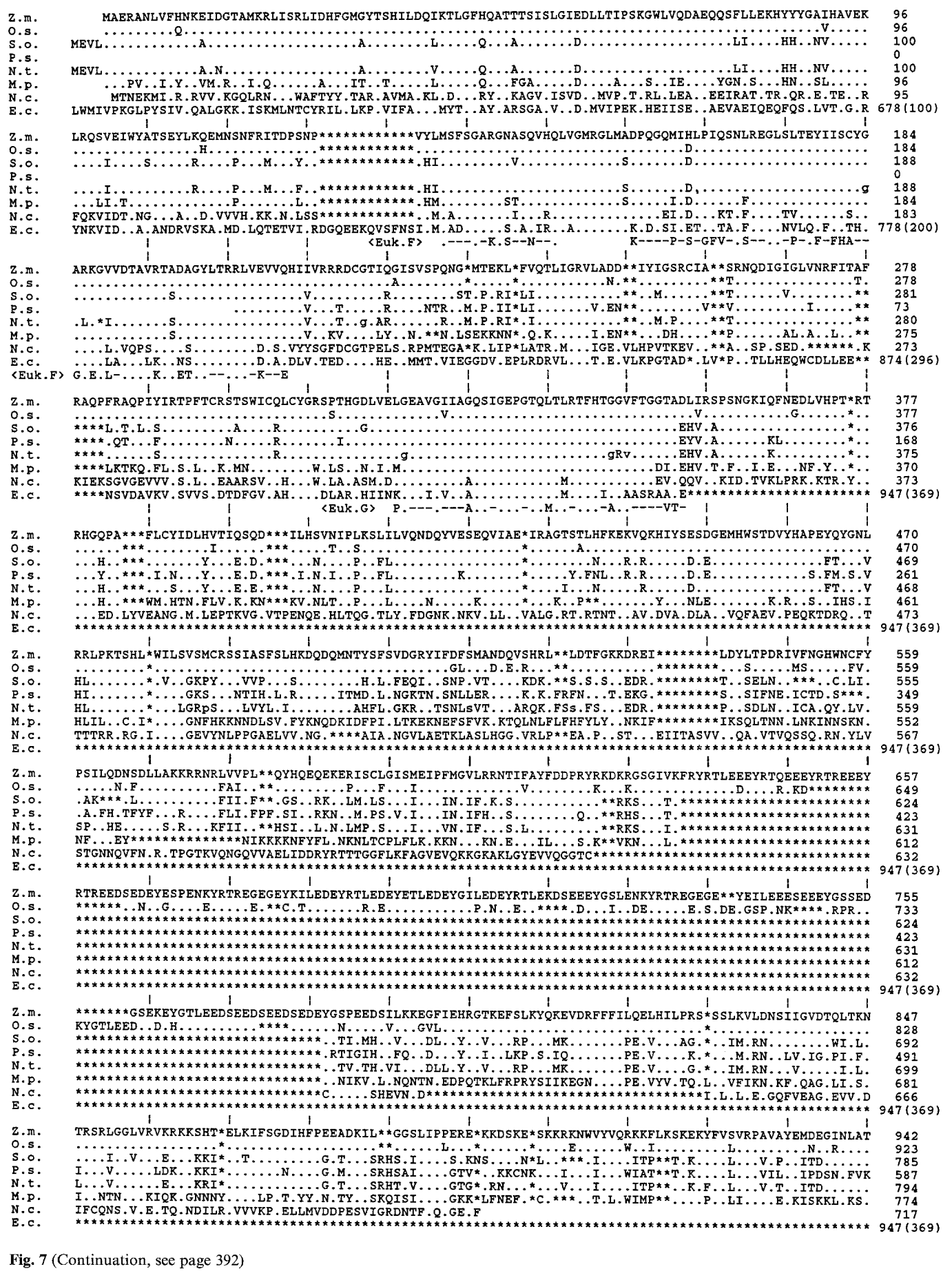

Fig. 7 (Continuation, see page 392)

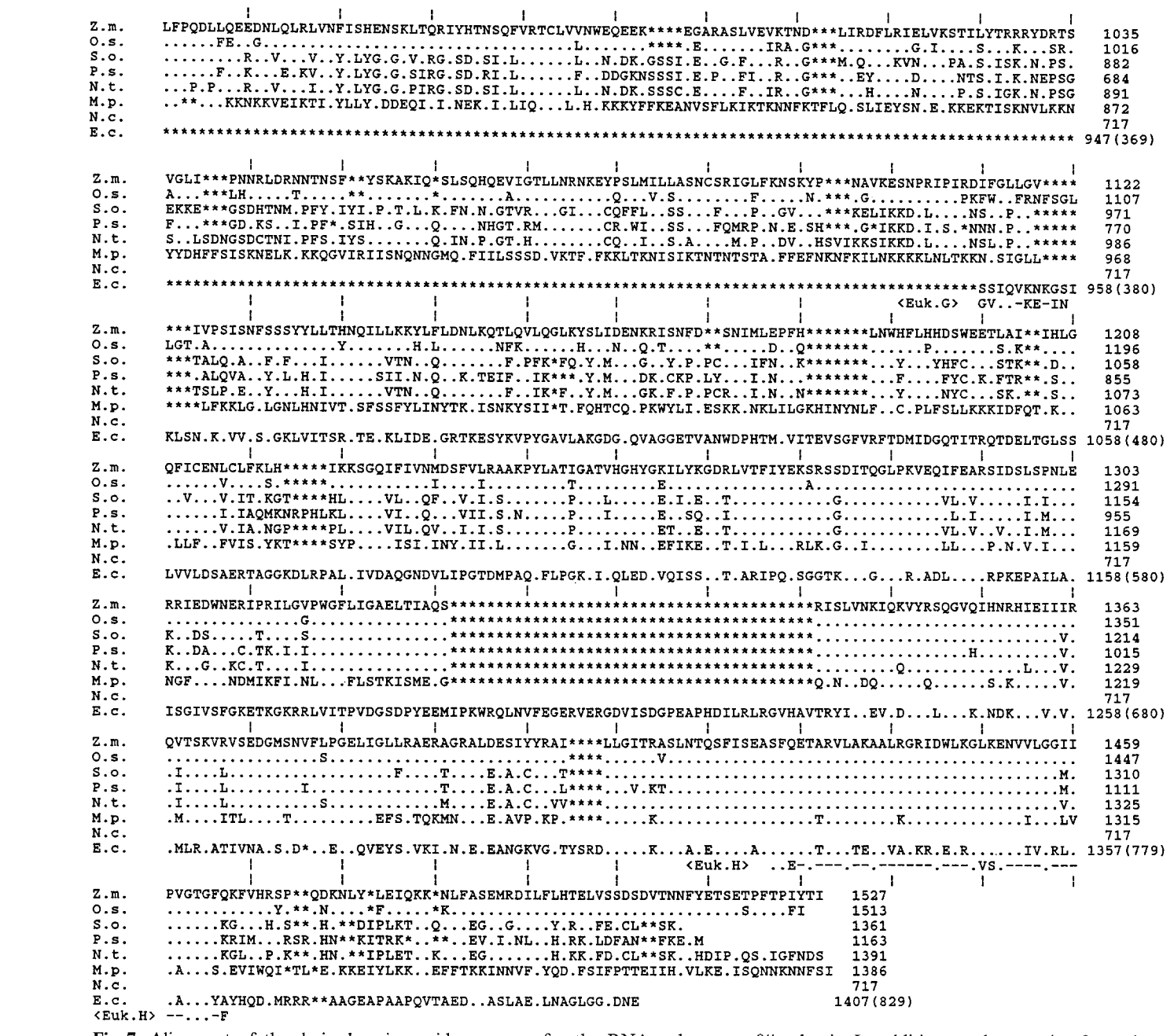

Fig. 7. Alignment of the derived amino acid sequences for the RNA polymerase β'' subunit. In addition to the proteins from the organisms described in the legend to Fig. 3, the sequences for the C-terminal fragment of the protein from pea (Cozens and Walker 1986) (P.s.) and the N-terminal fragment of the protein from the cyanobacterium *N. commune* (Xie et al. 1989) (N.c.) are also included. Due to minor corrections of the original tobacco sequence (Shinozaki et al. 1986) six positions (*small letters*) have been changed. The numbering in *brackets* for *E. coli* corresponds to the C-terminal fragment of β'' used for this alignment

meric repeat contribute to the low pI of the monocot β'' subunit compared to the dicot β'' subunits (Table 1). In rice (Shimada et al. 1990), despite the variability in this region, the initial tyrosine (coded by TAT in 13 out of 14 occasions) is retained but the leucine repeat in maize, which may be seen there as a potential zipper motif (Landschulz et al. 1988; Struhl 1989) is not well conserved. The high negative net charge is also a feature of the rice repeats. In view of the lack of success in finding any similar pattern of amino acids by a search of the EMBL data bank, one is left to speculate over a possible function, bearing in mind that the chloroplasts of dicots lack this structure.

Heptameric repeats are, apparently, not common in nature although, perhaps significantly, a further heptameric repeat of four leucines, which is partially conserved in rice and spinach, is found towards the C-terminus, starting at position 1136 in maize. Such units are clearly designed to align certain amino acids along one face of the α-helix (as, for example, in leucine zippers) and thus to promote helix-helix interaction. The segregation of tyrosines to one face and negatively charged amino acids to the opposite face as is the case of the heptamers under consideration here, could lead to an interaction of the protein with DNA through intercalation of the aromatic side chains. Such in-

teractions of DNA with tyrosines or repeating phenylalanines were recently shown to play a central role in the binding of single-stranded DNA (Merrill et al. 1988; Shamoo et al. 1989). Furthermore, directly upstream of the first heptameric unit (at position 627, see Fig. 7 and insert b of Fig. 6) the sequence motif RGSGIVKF, preserved with some conservative substitutions, in all plastids, is reminiscent of the consensus sequence for RNA binding proteins of (K,R)G(F,Y)(G,V)(F,Y)VX(F,Y) (Schwemmle et al. 1989). The aromatic side chains alternating with the small aliphatic residues are thought to facilitate interaction with the nucleic acid. Although such an alternation has been lost in the *rpo*C_2 sequences above, the agreement in five out of eight positions with the consensus sequence does not seem to be a random event: *rpo*B, of the same dimensions as *rpo*C_2, contains no sequence that agrees with the consensus sequence in even four out of eight positions. In view of the location, adjacent to the heptameric repeats, one could speculate that the alternating aromatic residues required by the consensus sequence have been replaced by the potentially more effective tyrosine repeats. The heptameric arrangement would minimize electrostatic repulsions between the backbone phosphates of the nucleic acid and the negatively charged amino acids at the distal surface of the α-helix, predicted by the method of Garnier et al. (1978) as existing virtually uninterrupted from positions 639–729; this is exemplified by the helix wheel representation of residues 689–706 in insert c of Fig. 6. This, in turn, could permit a further interaction with charged domains of other proteins such as specific transcription factors.

Developing this line of thought further, leads one to an analogy with transcriptional activators (Ptashne 1988). These proteins also bear two distinct surfaces as, for example, in an amphipathic α-helix; the hydrophobic surface binding to the DNA and the activating surface, consisting to a large extent of negatively charged amino acids, interacting with other proteins. The acidic residues and the alpha-helical conformation seem to be essential elements for the activation of transcription. Interestingly, it has been proposed that the activating region may interact directly with the RNA polymerase (Sigler 1988). If this is the case, one may reason that in RNA polymerases of monocot chloroplasts a transcriptional activator, or at least its functional domain has, for presumably evolutionary advantage, become incorporated into the β″ subunit. The addition of such a domain would represent the opposite situation to the loss of redundant parts of a protein sequence, as in the chloroplast β subunit, or an alternative to the separation of functions by fragmentation, as in the RNA polymerases of archaebacteria and in the case of β′ and β″ in chloroplasts. On the other hand, the extrachromosomal RNA polymerase from *K. lactis* (Wilson and Meacock 1988) would be another example where essential functional domains have been fused into one protein. However, it is conceivable that the large chloroplast specific insert, in general, and the monocot specific additional insert, in particular, have an as yet unknown but essential function which may lie even further outside the generally accepted role of RNA polymerases.

Acknowledgements. I.D. was recipient of a grant from the Deutscher Akademischer Austauschdienst. We are grateful to E. Schiefermayer and S. Krien for valuable technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 206) and the Fonds der Chemischen Industrie.

References

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

Argos P (1988) A sequence motif in many polymerases. Nucleic Acids Res 16:9909–9916

Berghöfer B, Kröckel L, Körtner C, Truss M, Schallenberg J, Klein A (1988) Relatedness of archaebacterial RNA polymerase core subunits to their eubacterial and eukaryotic equivalents. Nucleic Acids Res 16:8113–8128

Briat JF, Laulhere JP, Mache R (1979) Transcription activity of a DNA-protein complex isolated from spinach plastids. Eur J Biochem 98:285–292

Briat JF, Lescure AM, Mache R (1986) Transcription of the chloroplast DNA: a review. Biochimie 68:981–990

Carter PJ, Winter G, Wilkinson AJ, Fersht AR (1984) The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (*Bacillus stearothermophilus*). Cell 38:835–840

Cisek LJ, Corden JL (1989) Phosphorylation of RNA polymerase by the murine homologue of the cell-cycle control protein cdc2. Nature 339:679–684

Cozens AL, Walker JE (1986) Pea chloroplast DNA encodes homologues of *Escherichia coli* ribosomal subunit S2 and the β′-subunit of RNA polymerase. Biochem J 236:453–460

Evers R, Hammer A, Köck J, Jess W, Borst P, Mémet S, Cornelissen AWCA (1989a) *Trypanosoma brucei* contains two RNA polymerase II largest subunit genes with an altered C-terminal domain. Cell 56:585–597

Evers R, Hammer A, Cornelissen AWCA (1989b) Unusual C-terminal domain of the largest subunit of RNA polymerase II of *Crithidia fasciculata*. Nucleic Acids Res 17:3403–3413

Falkenburg D, Dworniczak B, Faust DM, Bautz EKF (1987) RNA polymerase II of *Drosophila*. Relation of its 140,000 M_r subunit to the β subunit of *Escherichia coli* RNA polymerase. J Mol Biol 195:929–937

Fritzsche E (1988) Untersuchungen zur molekularen Feinstruktur plastidärer DNA: Alkalilabile Stellen, replikative Intermediate. Kaliumpermanganat als Agens zur Identifizierung von 5-Methylcytosinpositionen. Dissertation, Albert-Ludwigs-Universität Freiburg, FRG

Fukuda R, Ishihama A (1974) Subunits of RNA polymerase in function and structure. V. Maturation *in vitro* of core enzyme from *Escherichia coli*. J Mol Biol 87:523–540

Garnier J, Osguthorpe DJ, Robson B (1978) Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J Mol Biol 120:97–120

Glass RE, Jones ST, Nene V, Nomura T, Fujita N, Ishihama A (1986) Genetic studies on the β subunit of *Escherichia coli* RNA polymerase VIII. Localisation of a region involved in promoter selectivity. Mol Gen Genet 203:487–491

Gray MW (1989) The evolutionary origins of organelles. Trends Genet 5:294–299

Hartmann GR, Biebricher C, Glaser SJ, Grosse F, Katzameyer MJ, Lindner AJ, Mosig H, Nasheuer H-P, Rothman-Denes LB, Schäffner AR, Schneider GJ, Stetter K-O, Thomm M (1988) Initiation of transcription – a general tool for affinity labeling of RNA polymerases by autocatalysis. Biol Chem Hoppe-Seyler 369:775–788

Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun C-R, Meng B-Y, Li Y-Q, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol Gen Genet 217:185–194

Hudson GS, Holton TA, Whitfeld PR, Bottomley W (1988) Spinach chloroplast *rpo*BC genes encode three subunits of the chloroplast RNA polymerase. J Mol Biol 200:639–654

Igloi GL, Döry I, Kössel H (1990) Nucleic Acids Res 18:683

Jin DJ, Gross CA (1988) Mapping and sequencing of mutations in the *Escherichia coli rpoB* gene that lead to rifampicin resistance. J Mol Biol 202:45–58

Jin DJ, Gross CA (1989) Characterization of the pleiotropic phenotypes of rifampicin-resistant *rpoB* mutants of *Escherichia coli*. J Bacteriol 171:5229–5231

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

Kidd GH, Bogorad L (1980) A facile procedure for purifying maize chloroplast RNA polymerase from whole cell homogenates. Biochim Biophys Acta 609:14–30

Klug A, Rhodes D (1987) 'Zinc fingers': a novel protein motif for nucleic acid recognition. TIBS 12:464–469

Kössel H, Igloi GL, Meinke A, Ruf M (1989) Structure and expression of genes coding for RNA polymerase subunits from maize chloroplasts. In: Galling G (ed) Proceedings of the Braunschweig Symposium on Applied Plant Molecular Biology. Technischen Universität Braunschweig, FRG, pp 91–97

Landschulz WH, Johnson PF, McKnight SL (1988) The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. Science 240:1759–1764

Leffers H, Gropp F, Lottspeich F, Zillig W, Garrett RA (1989) Sequence, organization, transcription and evolution of RNA polymerase subunit genes from the archaebacterial extreme halophiles *Halobacterium halobium* and *Halococcus morrhuae*. J Mol Biol 206:1–17

Lisitsyn NA, Monastyrskaya GS, Sverdlov ED (1988) Genes coding for RNA polymerase β subunit in bacteria. Structure/function analysis. Eur J Biochem 177:363–369

Little MC, Hallick RB (1988) Chloroplast *rpoA, rpoB* and *rpoC* genes specify at least three components of a chloroplast DNA-dependent RNA polymerase active in tRNA and mRNA transcription. J Biol Chem 263:14302–14307

Meinke A, Igloi GL, Kössel H (1988) Nucleotide sequence of tDNA(Cys)GCA and its flanking regions from *Zea mays* chloroplasts. Nucleic Acids Res 16:5696

Merrill BM, Stone KL, Cobianchi F, Wilson SH, Williams KR (1988) Phenylalanines that are conserved among several RNA-binding proteins form part of a nucleic acid-binding pocket in the A1 heterogeneous nuclear ribonucleoprotein. J Biol Chem 263:3307–3313

Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322:572–574

Ollis DL, Brick P, Hamlin R, Xuong NG, Steitz TA (1985a) Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. Nature 313:762–766

Ollis DL, Kline C, Steitz TA (1985b) Domain of *E. coli* DNA polymerase showing sequence homology to T7 DNA polymerase. Nature 313:818–819

Ovchinnikov YA, Monastyrskaya GS, Gubanov VV, Guryev SO, Chertov OY, Modyanov NN, Grinkevich VA, Makarova IA, Marchenko TV, Polovnikova IN, Lipkin VM, Sverdlov ED (1981) The primary structure of *Escherichia coli* RNA polymerase. Nucleotide sequence of the *rpoB* gene and amino acid sequence of the β-subunit. Eur J Biochem 116:621–629

Ovchinnikov YA, Monastyrskaya GS, Gubanov VV, Guryev SO, Salomatina IS, Shuvaeva TM, Lipkin VM, Sverdlov ED (1982) The primary structure of *E. coli* RNA polymerase. Nucleotide sequence of the *rpoC* gene and amino acid sequence of the β'-subunit. Nucleic Acids Res 10:4035–4044

Pai EF, Sachsenheimer W, Schirmer RH, Schultz GE (1977) Substrate position and induced-fit in crystalline adenylate kinase. J Mol Biol 114:37–45

Palmer JD (1985) Comparative organization of chloroplast genomes. Annu Rev Genet 19:325–354

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

Pühler G, Lottspeich F, Zillig W (1989) Organization and nucleotide sequence of the genes encoding the large subunits A, B and C of the DNA-dependent RNA polymerase of the archaebacterium *Sulfolobus acidocaldarius*. Nucleic Acids Res 17:4517–4534

Purton S, Gray JC (1989) The plastid *rpoA* gene encoding a protein homologous to the bacterial RNA polymerase alpha subunit is expressed in pea chloroplasts. Mol Gen Genet 217:77–84

Rodermel S, Bogorad L (1988) Genetic map of the *Zea mays* plastid chromosome. Maize Genet Coop Newslett 62:148–149

Rozen F, Pelletier J, Trachsel H, Sonenberg N (1989) A lysine substitution in the ATP-binding site of eucaryotic initiation factor 4A abrogates nucleotide-binding activity. Mol Cell Biol 9:4061–4063

Ruf M, Kössel H (1988) Structure and expression of the gene coding for the α-subunit of DNA-dependent RNA polymerase from the chloroplast genome of *Zea mays*. Nucleic Acids Res 16:5741–5754

Schwemmle M, Görlach M, Bader M, Sarre TF, Hilse K (1989) Binding of mRNA by an oligopeptide containing an evolutionarily conserved sequence from RNA binding proteins. FEBS Lett 251:117–120

Shamoo Y, Ghosaini LR, Keating KM, Williams KR, Sturtevant JM, Konigsberg WH (1989) Site-specific mutagenesis of T4 gene 32: The role of tyrosine residues in protein-nucleic acid interactions. Biochemistry 28:7409–7417

Shimada H, Fukata M, Ishikawa M, Sugiura M (1990) Rice chloroplast RNA polymerase genes: the absence of an intron in *rpoC1* and the presence of an extra sequence in *rpoC2*. Mol Gen Genet 221:395–402

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5:2043–2049

Sigler PB (1988) Transcriptional activation. Acid blobs and negative noodles. Nature 333:210–212

Struhl K (1989) Helix-turn-helix, zinc-finger and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins. TIBS 14:137–140

Sweetser D, Nonet M, Young RA (1987) Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. Proc Natl Acad Sci USA 84:1192–1196

Wilson DW, Meacock PA (1988) Extranuclear gene expression in yeast: evidence for a plasmid-encoded RNA polymerase of unique structure. Nucleic Acids Res 16:8097–8112

Wooley P, Clark BFC (1989) Homologies in the structures of G-binding proteins: An analysis based on elongation factor EF-Tu. Bio/technology 7:913–920

Xie W-Q, Jäger K, Potts M (1989) Cyanobacterial RNA polymerase genes *rpoC1* and *rpoC2* correspond to *rpoC* of *Escherichia coli*. J Bacteriol 171:1967–1973

Communicated by R. Hagemann

Received December 1, 1989

Note added in proof: Amino-terminal sequences of polypeptides corresponding to RNA polymerase subunits of 180 kDa and 120 kDa [Hu J, Bogorad L (1990) Maize chloroplast RNA polymerase: The 180-, 120-, and 38 kilodalton polypeptides are encoded in chloroplast genes. Proc Natl Acad Sci USA 87:1531–1535] have confirmed the primary structure predicted for the translation products from the *rpoC*$_2$ and *rpo*B genes, respectively, reported here.