Nucleic Acids Research, 1993, Vol. 21, No. 23 5345–5350

Structure of the gene encoding the 14.5 kDa subunit of human RNA polymerase II

Joël Acker, Marguerite Wintzerith, Marc Vigneron* and Claude Kedinger
Laboratoire de Génétique Moléculaire des Eucaryotes (CNRS), Unité 184 de Biologie Moléculaire et de Génie Génétique (INSERM), Institut de Chimie Biologique, 11 rue Humann, 67085 Strasbourg Cedex, France

Received September 3, 1993; Revised and Accepted October 19, 1993 EMBL accession no. Z23102

ABSTRACT

The structure of the gene encoding the 14.5 kDa subunit of the human RNA polymerase II (or B) has been elucidated. The gene consists of six exons, ranging from 52 to over 101 bp, interspaced with five introns ranging from 84 to 246 bp. It is transcribed into three major RNA species, present at low abundance in exponentially growing HeLa cells. The corresponding messenger RNAs contain the same open reading frame encoding a 125 amino acid residue protein, with a calculated molecular weight of 14,523 Da. This protein (named hRPB14.5) shares strong homologies with the homologous polymerase subunits encoded by the *Drosophila* (*RpII15*) and yeast (*RPB9*) genes. Cysteines characteristic of two zinc fingers are conserved in all three corresponding sequences and, like the yeast protein, the hRPB14.5 subunit exhibits zinc-binding activity.

INTRODUCTION

In eukaryotic cells, transcription of genes encoding messenger RNAs is effected by RNA polymerase II, in combination with a number of co-factors that control the selectivity and/or efficiency of transcription initiation, elongation and termination (for recent reviews see 1–3). Depending on its origin, RNA polymerase II consists of 10 to 14 polypeptides ranging from 220 to 10 kDa (4–7). The structure of these subunits as well as their contribution to the fundamental catalytic events and their interplay with the numerous co-factors that are associated with the transcriptional apparatus are still poorly characterized. The understanding in molecular terms of the mechanisms governing transcriptional regulation requires the identification of all the components of this complex machinery and ultimately their molecular cloning.

The *in vitro* systems designed for the study of transcription initiation often use human cell extracts with the corresponding endogenous or partially purified polymerase (8,9). We have therefore undertaken the isolation of sequences encoding the human RNA polymerase II subunits. The cDNA of the largest subunit (10) codes for a polypeptide (average molecular mass of 220 kDa) that is identical to the mouse equivalent (11) (with a nucleotide conservation of 89.4%) and highly homologous to the corresponding subunit from *Drosophila* (12), nematode (13), trypanosoma (14) and baker's yeast (15), as well as to the proteins encoded by the *rpo A* and *C* genes of archaeas (16,17) and to the β' subunit of bacterial RNA polymerase (18). The second largest subunit of human RNA polymerase II (molecular mass of about 140 kDa; 19) shares strong homologies with the *Drosophila* (20) and baker's yeast (21) counterparts and is also homologous to the protein encoded by the *rpo B* gene of archaeas (17) and to the β subunit of bacterial RNA polymerase (22).

In the case of baker's yeast, sequences encoding ten of the small subunits have been cloned so far: *RPB3* (subunit B44) (23), *RPB4* (subunit B32) (24), *RPB5* (subunit ABC27) (25), *RPB6* (subunit ABC23) (25), *RPB7* (subunit B16) (26),*RPB8* (subunit ABC14.5) (25), *RPB9* (subunit B12.6) (27), *RPC10* (subunit ABC10α) (28), *RPB10* (subunit ABC10α) (29) and *RPB11* (subunit B12.5) (30). Among these subunits, ABC27, ABC23, ABC14.5, ABC10α and ABC10β are present in all three types of yeast polymerase (6,25,31).

In the case of human polymerase, clones encoding hRPB25 and hRPB33 (homologous to the yeast *RPB5* and *RPB3* genes, respectively) have been isolated (32, 33). The gene of the *Drosophila* RpII15 subunit (homologous to yeast *RPB9*) has recently been described (34). Here we report the genomic sequence for the human hRPB14.5 subunit which represents the first *RPB9* homolog so far characterized in vertebrates.

MATERIALS AND METHODS

Cloning

A DNA probe was generated by specific PCR-mediated amplification of HeLa cell cDNA. A mixture of degenerate primers, designed according to the sequence of two conserved motifs of the *Drosophila* RpII15 peptide sequence (19-RFCQECNNMLY-29 and 79-DPTLPRTED-87), was used to amplify sequences from cDNAs prepared from oligodT-primed HeLa cell polyA$^+$ RNA. An amplified fragment (about 200 bp)

* To whom correspondence should be addressed

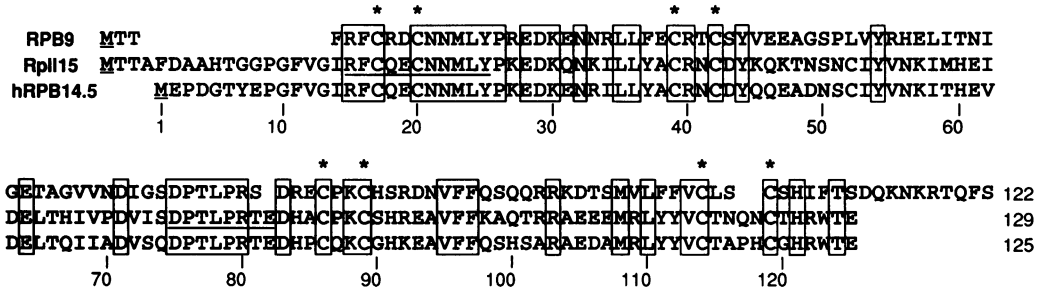

Figure 1. Peptide sequence alignments of the human, *Drosophila* and yeast RNA polymerase II homologous subunits.Sequences have been aligned according to the procedure of Higgins and Sharp (40). As indicated on the left, the top line corresponds to the sequence of the *Saccharomyces cerevisiae* B12.6 subunit (*RPB9*), the middle line to that of the *Drosophila melanogaster* RpII15 protein and the bottom line to that of the human hRPB14.5 protein. The amino acid residues (one-letter code, numbered in the case of hRPB14.5) which are conserved in all three sequences are boxed. The conserved cysteine residues are marked by an asterisk. The initiator methionine and the peptide sequences used to derive the PCR primers are underlined. The overall number of residues of each subunit is given at the bottom right.

was obtained which was ^{32}P-labelled by nick-translation and used as a probe for the screening of both lambda Zap-cDNA (derived from random-primed HeLa cell polyA$^+$ RNA) and lambda-EMBL3-genomic (derived from partial Sau3A I digests of human placental DNA) libraries.

The libraries were plated at 40,000 plaques per 13.5 cm diameter Petri dish and two filters were lifted per dish. The filters were hybridized (42°C, 35% formamide) with the labelled PCR-derived probe (see above). Independent clones were picked, purified and characterized. In the case of the lambda Zap-cDNA library, the inserts were recovered by helper-mediated excision, as Bluescript recombinants. These were subsequently characterized by Southern-blot analysis of Eco RI digests with the same PCR-derived probe. In the case of the genomic library, the positive clones were characterized by Southern-blot analysis of Sac I, Not I, Sau3A I and Sfi I digests, with the same probe. A positive 3 kb Sac I fragment, from two independent clones, was subcloned into a Bluescript vector. The resulting recombinants were purified by two successive CsCl-gradient centrifugations, before sequence analysis. Sequence analysis was performed by a combination of manual and automated sequencing procedures.

Transfections and RNA analysis

Cos-7 cells were transfected by calcium phosphate coprecipitation with the purified plasmids, as previously described (35). The medium was changed after 15 to 20 hr. After an additional 20 to 24 hr, the cells were harvested, lyzed in Nonidet P40 (0.5%) and cytoplasmic RNA was prepared by phenol−chloroform extraction.

Specific RNA transcripts were analysed by S1 nuclease mapping (36), using a PCR-amplified single-stranded fragment as probe (spanning nucleotides 253 to 446, Fig. 3A), 5′ ^{32}P-end-labelled at nucleotide 446 or by reverse transcriptase primer extension (36), using a synthetic oligonucleotide (spanning nucleotides 427 to 446, Fig. 3A) 5′ ^{32}P-end-labelled at nucleotide 446, as primer.

Zinc blot analysis

The hRPB14.5 and hRPB25 proteins were produced in *E. coli* BL21 transformed by a pET3a-based vector (37) harbouring the complete hRPB14.5 or hRPB25 (J.A., M.W. and M.V., unpublished) cDNAs, extended 5′ of their ATG initiator codon

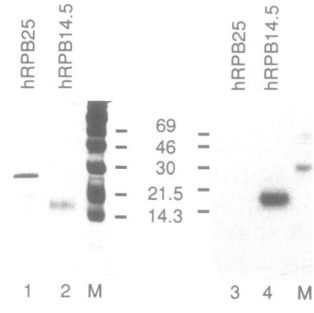

Figure 2. Zinc-binding activity of bacterially synthesized hRPB14.5 protein. About 3 μg of purified protein from bacteria expressing the histidine-tagged hRPB25 and hRPB14.5 proteins (as indicated) was electrophoresed on SDS−polyacrylamide gels and stained with Coomassie blue (lanes 1 and 2) or electrotransferred to nitrocellulose and incubated with radioactive zinc (lanes 3 and 4), as described in Material and Methods. M, rainbow molecular mass markers (kDa), among which the carbonic anhydrase (30kDa) and bovine serum albumin (69 kDa) are labelled by the zinc.

by a sequence encoding the Met-(His)$_6$-Ala-Ser tag-peptide. After IPTG-induction (2 hr, 37°C), bacteria were collected, sonicated and the histidine-tagged products were solubilized in buffer A (6 M guanidinium−HCl, 20 mM Tris−HCl, pH 8, KCl 100 mM) and purified by chromatography on metal-chelate-affinity-agarose, as described by the manufacturer (Qiagen, Pharmacia). The histidinylated products, eluted in buffer A, 100 mM EDTA, were loaded onto a 12.5% SDS−polyacrylamide gel (38), after dialysis against phosphate-buffered saline. After electrophoresis, the gel was treated as previously described (39). Briefly, the proteins were reduced in the presence of 5% 2-mercaptoethanol, before electrotransfer onto nitrocellulose membranes. The blot was incubated in the presence of $^{65}ZnCl_2$ (0.2 μCi/ml), washed and exposed for autoradiography.

RESULTS AND DISCUSSION

The cDNA encoding hRPB14.5

A PCR-amplified HeLa cell cDNA fragment spanning two domains that are conserved in the homologous *Drosophila melanogaster* (RpII15) and *Saccharomyces cerevisiae* (B12.6) RNA polymerase II subunits, was used as probe for the screening of a HeLa cell cDNA library. Several independent overlapping

A

```
   1  gagctcgcccagtccgcgaagggtaaactcgcgcatgcgcactggccttcaaccaagttaagcgccttggcttctgtcctggatgctcagtgttttgcgg
 101  tttacgatgcgcatcgttttcccggctctggcaccagatccatactattctgtaaactgaggcggcgatgcccaaaaattggaagcaattactctaagat
 201  atcaaaagattgaagcaagccatgaaacctctgcgcggagagagccgagcccgaggcggcggggagccagcaagtgtgcgcagctccgccctcttttggc

 301  gaagcacctaggcctggcccctcccgcgacctgtagcgcggcggagcaagcgca GAAGGCTGGGAGGGCTGC GCGGGCTGCGCGTCGCC ATG GAG
                                                        (b)                (a)                 M  E     2

 396  CCC GAC GGG ACT TAC GAG CCG GGC TTC GTG GGT ATT CGC TTC TGC CAG GAA TG GTGAGGCGGAGGGAGCGGGTGATGTTAGA
       P  D  G  T  Y  E  P  G  F  V  G  I  R  F  C  Q  E  C                                            20

 478  AGGCGGGCGAGGAATGGCTCCAGGACTAACCCCTGGCACCTCCCCCTCCCCACAG T AAC AAC ATG CTG TAC CCC AAG GAA GAC AAG GA
                                                              N  N  M  L  Y  P  K  E  D  K  E           31

 566  G AAC CGC ATT CTG CTC TAC GCG GTGAGCGCCGGGGTCTTCCCCGGCAGGTCCACCTGTTCCCAGCGTGGCCCCGCCCTGCCCCCTGCCTCTG
        N  R  I  L  L  Y  A                                                                            38

 658  TGATCGGACTCCCTGTTCGCCCTCCGCTGCACCTCAAGTGATCCTGACTCTCATGTCCTGCCCATCCCCAAAGAGAGTCCTGTCGTCATATCCTCCGGGA

 758  TCACCCTGCCCTCTCCTCCCAATCTTGTCTGGAATCCTTCCGTGACCCCCTTTCCCCCACACTCGCCCTCTCGTAG TGC CGG AAC TGT GAT TAC
                                                                          C  R  N  C  D  Y              44

 852  CAG CAG GAG GCC GAC AAC AGC TGC ATC TAT GTC AAC AAG ATC ACG CAC GAA GTG GA GTGAGTGGCGGGCCCTGAGCTGGGG
       Q  Q  E  A  D  N  S  C  I  Y  V  N  K  I  T  H  E  V  D                                        63

 933  GGCGGGGGTGTTGGCTCTGGAGGCTGGGTCTGAGCGTAATTTTGCACCCCCGCGTCCCTGCAG C GAA CTG ACC CAG ATT ATC GCC GAC GT
                                                                    E  L  T  Q  I  I  A  D  V           72

1023  G TCC CAG GAC CCC ACG TTG CCG CGG ACC GAG GAC CAC CCG TGC CAA AA GTGAGCGTTCAGCTGGTACTGGGGAGGTAAAGTGA
        S  Q  D  P  T  L  P  R  T  E  D  H  P  C  Q  K                                               88

1106  GGCTCGTCTCCTCACTGGGCGAAGGGAGGAGGGAAGGGTGTGAAGCCTCCGGTGACCGAGCCCCTACCCTCTCAG G TGC GGC CAC AAG GAG GC
                                                                                C  G  H  K  E  A        94

1199  T GTG TTC TTC CAG TCA CAC AGT GCG CGG GCC GAG GTAAGTGGCGCGAGGGCTTTCTAAGTATCTGGGGCTTGTTTGGTCCTTTCTGGG
        V  F  F  Q  S  H  S  A  R  A  E                                                                105

1287  GGCGGGGATATCAATGTGGGTGACCCTGCGCTTCTTGGAACTCTGCCGGGCCCTTACCTATTGGTGAAGTTCTCTTGCCTGTTCTTTTGGGGTAAGGCGC

1387  TGCTGGGTCCAAATCCTTCCTAACCATGCATCATTTCTGCCAG GAC GCC ATG CGC CTT TAC TAC GTG TGC ACA GCC CCA CAC TGC
                                                    D  A  M  R  L  Y  Y  V  C  T  A  P  H  C           119

1472  GGC CAC CGC TGG ACC GAG TGA CCTCCTCTCTCCCCCGAGTGTAATAAACACCAGATTCC atgcgtgagttttctgtgttttatttgcctcgtg
       G  H  R  W  T  E  *                                                                            125

1564  gggaggacgctggcacaccgtatcaggctgcgcggtgcagggcgcggtgctgtgcgtaggtgtcgggccgggagctggtgaagccgcagtcctcgcacac
1664  gtgcagcttctcgcggcgctcacggtaagcgtagctggcccggcctgtccatgcaccttagcaaggtgagcttccagggagcagcgctgcgtaaacgctt
1764  tcccgcaagcactggagcggaagggccggatccctgcacagatgggcgcagaggggagggctggacagagagggcaatgggctctggactaaagcgcccc
1864  caccctgctttccagtggcgagcatttctggtacctattccaagctctacagactcttccgcccatgacccaccatctgctaccgacacccagtggagac
1964  atgggacagacacctgtgtctccactctcccacaccaggatgctccctcccctgggggatggagagggaggccttcccttgcctcgcgatctggagcctg
2064  aatcgtataaccggcatccgtctacactct 2093
```

B

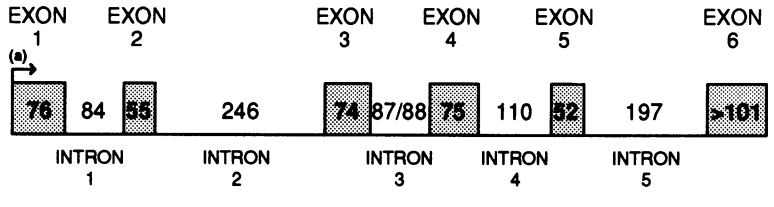

Figure 3. Sequence of the Sac I genomic fragment encoding the hRPB14.5 gene. (**A**) The nucleotidic sequence (2093 bp, numbered on the left) is presented, with uppercase characters corresponding to sequences known to be transcribed (deduced from RNA 5′ end mapping and cDNA sequencing). Bold characters denote the transcription start sites [(a) and (b)], as deduced from the data presented in Fig. 4. Sequence elements corresponding to a potential polyadelylation signal are underlined. The variable G-track at positions 929−934 is underlined. The translated sequence, as deduced from the open reading frame of the cDNA, is spaced in triplets, with the predicted peptide sequence given below (amino acid residues numbered on the right). (**B**) The structure of the smallest primary transcript (starting at initiation site 'a' and ending between the two elements of the polyadenylation signal) is summarized, with stippled boxes corresponding to the exons and simple lines to the introns. Lengths are given in nucleotides. The two alternative sizes of intron 3 are given. The minimal size of exon 6 (101 nucleotides) is deduced from the longest cDNA isolated.

clones were isolated and sequenced. The largest cDNA sequence that could be reconstructed (439 bp) comprised an open reading frame (375 bp) that encodes a polypeptide of 125 aminoacid residues (see Fig. 1). Since the calculated molecular mass of this protein is 14,523 Da, we refer to it henceforth as the hRPB14.5 subunit.

The results of a comparison of the hRPB14.5 peptide sequence with its *Drosophila* and yeast counterparts are presented in Fig. 1. After optimal sequence alignment, 40% and 71% of the hRPB14.5 residues were found to be identical at corresponding positions of the yeast and *Drosophila* sequences, respectively. It is noteworthy that the primers chosen for the PCR-mediated amplification of the probe were derived from the best conserved areas of the sequence (underlined). Among the invariant residues (boxed), two sets of four cysteine residues (positions 17/20/39/42 and 86/89/114/119) define two potential zinc fingers, as suggested previously from the yeast and *Drosophila* sequences (27). Zinc-blot experiments, in which purified yeast RNA polymerase II

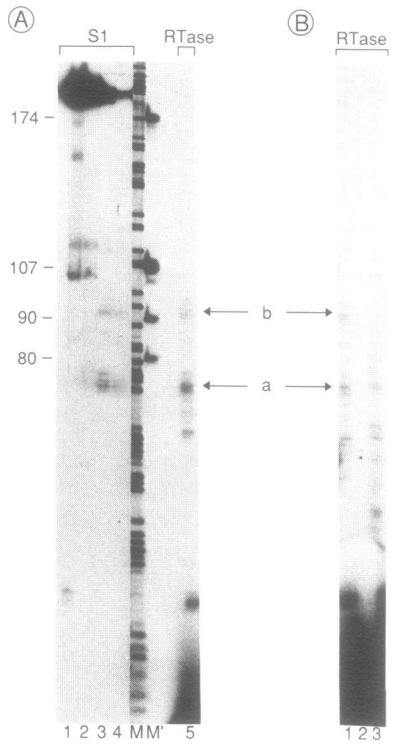

Figure 4. Transcription initiation at the hRPB14.5 gene. (**A**) Localization of the transcription start sites. PolyA$^+$ RNA was extracted from NP40-lysed HeLa cells which were exponentially growing in suspension. Total RNA was fractionated twice on oligodT−cellulose columns and recovery of polyA$^+$ RNA was 3.5% of the starting RNA. Purified polyA$^+$ RNA (15 μg) was hybridized to 5000 cpm of a 5′ end-labelled single-stranded probe (nucleotides 253 to 446, Fig. 3A) and digested with increasing concentrations of S1 nuclease: 1U (lane 1), 5U (lane 2), 25U (lane 3), 125U (lane 4). Lane M: sequence ladder of an unrelated DNA fragment used as size marker. Lane M′: PCR-amplified DNA fragments with sizes (nucleotides) indicated on the left. Lane 5: 15 μg of the purified polyA$^+$ RNA was hybridized to 25,000 cpm of a 5′ end-labelled 20 nucleotide primer (nucleotides 427 to 446, Fig. 3A), the hybrids were incubated with AMV reverse transcriptase (RTase). The samples were separated by denaturing polyacrylamide gel elecrophoresis. Arrows point to S1 nuclease-resistant probe fragments and RTase-extended primers that identify major RNA initiation sites (a and b). (**B**) The Sac I genomic fragment exhibits promoter activity. Lane 1: 15 μg of the purified polyA$^+$ RNA was analyzed by RTase primer extension as in panel A (lane 5). Lane 2: Cos-7 cells were transfected with the unloaded Bluescript vector (15 μg per 9 cm Petri dish). Thirty-six hours after transfection, total cytoplasmic RNA was extracted and an amount corresponding to half of the cells from a Petri dish was analyzed as in lane 1. Lane 3: Cos-7 cells were transfected (2 μg per 9 cm Petri dish, adjusted to 15 μg with Bluescript plasmid DNA) with the recombinant Bluescript vector containing the Sac I genomic fragment. Half of the cytoplasmic RNA obtained from a Petri dish was analyzed as in lane 2. Arrows point to the RTase-extended primers corresponding to major initiation sites (a and b).

subunits were probed with radioactive zinc (31,41), revealed a labelled band corresponding to a polypeptide with an apparent size of 12.5 kDa, at the position of the comigrating B12.5 (*RPB11*) and B12.6 (*RPB9*) subunits. Since the protein encoded by *RPB11* does not contain obvious zinc fingers, this observation supported the conclusion that the cysteine-containing domains of B12.6 are capable of binding zinc.

To directly determine whether the hRPB14.5 subunit was also able to bind zinc, we applied the zinc-blot technique to bacterially-expressed hRPB14.5. As shown in Fig. 2, radioactive zinc was bound at the position of the purified hRPB14.5 recombinant protein (compare lanes 2 and 4). As a negative control, the hRPB25 subunit, which contains no potential metal-binding element, was analyzed in parallel. As expected, no signal could be detected at the position of this protein (compare lanes 1 and 3). This also ruled out the possibility that the zinc atoms were bound by the histidine tag present at the N-terminus of both hRPB14.5 and hRPB25 proteins. Whether both cysteine-containing domains of hRPB14.5 or B12.6 are involved in metal binding is presently unknown. However, the fact that the most highly conserved region (between residues 15 and 44), in all three species, encompasses the first set of 4 cysteines, suggested that at least the N-terminal half of this subunit is responsible for zinc-binding. A definitive answer to this question clearly awaits site-directed mutagenesis.

Although the first methionine codon of the open reading frame is within a favorable context for initiating translation (42), its position as the initiator codon was not unambiguously established, since no in-frame stop codon was found further upstream. This raised the possibility that the actual N-terminus of the subunit was missing from the cloned cDNAs. To solve this question, we decided to undertake the cloning of the corresponding genomic region.

The gene for hRPB14.5

Using the same probe as for the cDNA cloning to screen a genomic library, we isolated two independent clones. A 3 kb Sac I fragment was subcloned in each case and subjected to sequence analysis (Fig. 3A; EMBL accession No. Z23102). With the exception of one nucleotide insertion (or deletion), the sequences of the two genomic clones (2093 bp determined so far) were identical. A comparison of the genomic sequences with that of the cDNA indicated that the entire hRPB14.5 coding sequence was contained within the genomic Sac I fragment and this allowed us to deduce the internal structure of the gene which appeared to be composed of 6 exons separated by 5 short introns (see Fig. 3A and B for details). The only difference between the two clones isolated is in intron 3, where the G-track between positions 929 and 934 was either 6 (Fig. 3A) or 5 bp long. Whether this difference corresponds to a natural variant of the hRPB14.5 gene or results from a cloning artifact is not known, but clearly indicates that the two clones were actually independent and therefore that the sequence established is reliable.

The longest cDNA started at nucleotide 367 and ended at nucleotide 1530. A polyadenylation signal with a perfect match with both upstream and downstream consensus sequences could be identified on either side of position 1530, strongly suggesting that the polyadenylation site, and hence the 3′ end of the 6th exon, must be very close to this position.

To define the 5′ limit(s) of the first exon, we determined the transcription start site(s) by S1 nuclease and reverse transcriptase (RTase) mapping of HeLa cell polyA$^+$ RNA. To facilitate the identification of the probe fragments that are fully protected against S1 nuclease, identical amounts of hybrids were digested with increasing amounts of S1 nuclease (Fig. 4A, lanes 1−4). The pattern of RTase-extended primers (Fig. 4A, lane 5) was nearly identical to that of the probe fragments protected against the strongest S1 nuclease digestion conditions (lanes 3 and 4). The results suggest that transcription initiation is somewhat heterogenous, the start sites being scattered over about 40 bp, with 2 major initiation regions positionned around nucleotides 373 ('a') and 355 ('b') (Fig. 3A). Additional but minor start sites can also be detected between 'a' and 'b' and further upstream.

All RNA species transcribed from these sites encode the same open reading frame as predicted from the cDNA analysis, thus confirming the position of the translation initiation site at position 390 (Fig. 3). In agreement with this conclusion, several stop codons are found 5′ to start site 'b', the closest being at nucleotide 309.

Examination of the nucleotide sequences located upstream of the transcription start sites did not reveal signatures of typical promoter elements, with the exception of a significantly higher GC content (71%) within the first 100 bp of 5′ flanking sequences (between nucleotides 236 and 337), which, together with the relative spreading of initiation sites, is a characteristic of prmoters of house-keeping genes (43). To test for the presence of promoter sequences within the cloned genomic sequence, we examined its ability to be transcribed in eukaryotic cells. A vector devoid of its own eukaryotic promoter elements, but harboring the hRPB14.5 gene, together with about 330 bp of 5′ flanking DNA was transfected into cultured cells. As revealed by the RTase assay (Fig. 4B), specific transcripts initiated at site 'a' could be detected in total cytoplasmic RNA from cells transfected with the recombinant plasmid (lane 3), but not from cells transfected with the unloaded vector (lane 2). This indicates that the isolated genomic sequence carries enough information to allow efficient transcription from the transfected gene. Strikingly however, no RNA starting at position 'b' could be visualized, suggesting either that these transcripts are not stable under the transfection conditions or that additional promoter elements required for the recognition of these sites are missing in the present construct. The precise delineation of the essential control elements will require a saturating mutation analysis of the entire promoter region.

CONCLUSION

The sequences encoding the hRPB14.5 subunit of the human RNA polymerase II are split into 6 exons, with the open reading frame being flanked by rather short 5′ and 3′ untranslated regions, spanning only 17 bp from the major start site and about 40 bp up to the presumptive polyadenylation site, respectively. The weak intensity of the signals detected by S1 and RTase analysis of polyA$^+$ RNA, most likely reflects low levels of transcription from the hRPB14.5 gene in growing cells. The corresponding mRNA must be sufficiently stable to allow translation of proper amounts of hRPB14.5 subunit. The particularly high stability of this mRNA is illustrated in the case of *Drosophila* (34) where homozygotic alteration of the homologous *RpII15* gene caused lethality only at mid-larval stages, because the wild type mRNA, that has been deposited by mother cells into the egg before fertilization, is stable enough to support early larval development.

Interspecific sequence comparisons of the hRPB14.5 protein homologs revealed peptidic elements which, due to their conservation, must play critical functions within the integral hRPB14.5 molecule. The role of these domains may now be explored by systematic site-directed mutagenesis in appropriate assay systems.

ACKNOWLEDGEMENTS

We wish to thank R. Gopalkrishnan for critical reading of the manuscript, C. Hauss for technical assistance, J.M. Garnier for providing DNA libraries, the chemistry staff for oligonucleotides, S. Vicaire for automated DNA sequencing, F. Plewniak for help with computer analyses and C. Werlé for artwork. This work was supported by the Association pour la Recherche sur le Cancer and by grant no. 92H0924 from the MRE. J.A. was a recipient of a fellowship from the Ligue Nationale contre le Cancer.

REFERENCES

1. Sawadogo, M. and Sentenac, A. (1991) Annu. Rev. Biochem. *59*, 711–754.
2. Zawel, L. and Reinberg, D. (1992) Current Biol. *4*, 488–495.
3. Ham, J., Steger, G. and Yaniv, M. (1992) FEBS Lett. *1*, 81–86.
4. Sentenac, A. (1985) CRC Crit. Rev. Biochem. *18*, 31–91.
5. Young, R.A. (1991) Annu. Rev. Biochem. *60*, 689–715.
6. Sentenac, A., Riva, M., Thuriaux, P., Buhler, J.M., Treich, I., Carles, C., Werner, M., Ruet, A., Huet, J., Mann, C., Chiannilkulchai, N., Stettler, S. and Mariotte-Labarre, S. (1992) In 'Transcriptional Regulation' Cold Spring Harbor Laboratory Press, pp. 27–54.
7. Thuriaux, P. and Sentenac, A. (1992) In 'The molecular and cellular biology of the yeast Saccharomyces: Gene expression' Cold Spring Harbor Laboratory Press, Volume II, pp. 1–48.
8. Manley, J.L., Fire, A., Cano, A., Sharp, P.A. and Gefter, M.L. (1980) Proc. Natl. Acad. Sci. USA. *77*, 3855–3859.
9. Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) Nucleic Acids Res. *11*, 1475–1489.
10. Wintzerith, M., Acker, J., Vicaire, S., Vigneron, M. and Kedinger, C. (1992) Nucleic Acids Res. *20*, 910.
11. Ahearn, J.M., Bartolomei, M.S., West, M.L., Cisek, L.J. and Corden, J.L. (1987) J. Biol. Chem. *262*, 10695–10705.
12. Jokerst, R.S., Weeks, J.R., Zehring, W.A. and Greenleaf, A.L. (1989) Mol. Gen. Genet. *215*, 266–275.
13. Bird, D.M. and Riddle, D.L. (1989) Mol. Cell. Biol. *9*, 4119–4130.
14. Smith, J.L., Levin, J.R., Ingles, C.J. and Agabian, N. (1989) Cell *56*, 815–827.
15. Allison, L.A., Moyle, M., Shales, M. and Ingles, C.J. (1985) Cell *42*, 599–610.
16. Leffers, H., Gropp, F., Lottspeich, F., Zillig, W. and Garrett, R.A. (1989) J. Mol. Biol. *206*, 1–17.
17. Pühler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H.P., Lottspeich, F., Garrett, R.A. and Zillig, W. (1989) Proc. Natl. Acad. Sci. USA *86*, 4569–4573.
18. Ovchinnikov, Y.A., Monastyrskaia, G.S., Gubanov, V.V., Guryev, S.O., Salomatina, I.S., Shuvaeva, T.M., Lipkin, V.N., Sverdlov, E.D. and Shemyakin, M.F. (1982) Nucleic Acids Res. *10*, 4035–4044.
19. Acker, J., Wintzerith, M., Vigneron, M., and Kedinger, C. (1992) J. Mol. Biol. *226*, 1295–1299.
20. Falkenburg, D., Dworniczak, B., Faust, D.M. and Bautz, E.K.F. (1987) J. Mol. Biol. *195*, 929–937.
21. Sweetser, D., Nonet, M. and Young, R.A. (1987) Proc. Natl. Acad. Sci. USA. *84*, 1192–1196.
22. Lisitsyn, S.A., Monastyrskaya, G.S., Sverdlov, E.D. (1988) Eur. J. Biochem. *177*, 363–369.
23. Kolodziej, P.A. and Young, R.A. (1989) Mol. Cell. Biol. *9*, 5397–5394.
24. Woychick, N.A. and Young, R.A. (1989) Mol. Cell. Biol. *9*, 2854–2859.
25. Woychick, N.A., Liao, S.M., Kolodziej, P.A. and Young, R.A. (1990) Genes Dev. *4*, 313–323.
26. McKune, K., Richards, K.L., Edwards, A.M., Young, R.A. and Woychik, N.A. (1993) Yeast *9*, 295–299.
27. Woychick N.A., Lane, A.S., and Young, R.A. (1991) J. Biol. Chem. *255*, 19053–19055.
28. Treich, I., Carles, C., Riva, M., and Sentenac, A. (1992) Gene Exp. *2*, 31–37.
29. Woychick, N.A. and Young, R.A. (1990) J. Biol. Chem. 265, 17816–17819.
30. Woychick, N.A., McKune, K., Lane, W.S. and Young, R.A. (1993) Gene Exp. *3*, 77–82.
31. Carles, C., Treich, I., Bouet, F., Riva, M., and Sentenac, A. (1991) J. Biol. Chem. *266*, 24092–24096.
32. Pati, U.K., and Weissman, S.M. (1989) J. Biol. Chem. *264*, 13114–13121.
33. Pati, U.K., and Weissman, S.M. (1990) J. Biol. Chem. *265*, 8400–8403.
34. Harrison, D.A., Mortin, M.A., and Corces, V.G. (1992) Mol. Cell. Biol. *12*, 928–935.
35. Zajchowski, D.A., Boeuf, H., and Kedinger, C. (1985) EMBO J. *4*, 1293–1300.
36. Zajchowski, D.A., Boeuf, H., and Kedinger, C. (1987) Gene *58*, 243–256.

37. Studier, F.W., Rosenberg, A.H., Dun, J.J. and Dubendorf, J.W. (1990) Methods Enzymol. *185*, 60–89.
38. Laemmli, U.K. (1970) Nature *227*, 680–685.
39. Mazen, A., Gradwohl, G. and De Murcia, G. (1988) Anal. Biochem. *172*, 39–42.
40. Higgins, D.G. and Sharp, P.M. (1988) Gene *73*, 237–244.
41. Treich, I., Riva, M. and Sentenac, A. (1991) J. Biol. Chem. *266*, 21971–21976.
42. Kozak, M., (1992) Crit Rev in Biochem and Mol Biol *28*, 285–402.
43. Smale, S.T. and Baltimore, D. (1989) Cell *57*, 103–113.