Subunits of the Schizosaccharomyces pombe RNA polymerase II: Enzyme purification and structure of the subunit 3 gene

Yoshinao Azuma¹,², Masahiro Yamagishi¹ and Akira Ishihama¹*
¹Department of Molecular Genetics, National Institute of Genetics, Mishima, Shizuoka 411 and ²School of Life Science, Graduate University for Advanced Studies, Mishima, Shizuoka 411, Japan

Received April 26, 1993; Revised and Accepted June 28, 1993

ABSTRACT

To improve our understanding of the structure and function of eukaryotic RNA polymerase II, we purified the enzyme from the fission yeast Schizosaccharomyces pombe. The highly purified RNA polymerase II contained more than eleven polypeptides. The sizes of the largest, the second-, and the third-largest polypeptides as measured by SDS-polyacrylamide gel electrophoresis were about 210, 150, and 40 killodaltons (kDa), respectively, and are similar to those of RPB1, 2, and 3 subunits of Saccharomyces cerevisiae RNA polymerase II. Using the degenerated primers designed after amino acid micro-sequencing of the 40 kDa third-largest polypeptide (subunit 3), we cloned the subunit 3 gene (rpb3) and determined its DNA sequence. Taken together with the sequence of parts of PCR-amplified cDNA, the predicted coding sequence of rpb3, interrupted by two introns, was found to encode a polypeptide of 297 amino acid residues in length with a molecular weight of 34 kDa. The S. pombe subunit 3 contains four structural domains conserved for the α-subunit family of RNA polymerase from both eukaryotes and prokaryotes. A putative leucine zipper motif was found to exist in the C-terminal proximal conserved region (domain D). Possible functions of the conserved domains are discussed.

INTRODUCTION

Knowledge of the structure and function of RNA polymerase is essential to understand the molecular mechanisms of transcriptional regulation. In eukaryotes, three classes of nuclear RNA polymerases, I(A), II(B) and III(C), exist, each responsible for transcription of large rRNA, mRNA, and small rRNA plus tRNA, respectively (for reviews see refs. 1,2). Each class of purified RNA polymerase is composed of two large and about 10 small polypeptides (1,2).

The genes coding for the largest (subunit 1) and the second-largest (subunit 2) subunits of the three classes of RNA polymerase have been cloned from a number of organisms (3,4; and other work cited there). Sequence analyses indicate that these subunits share notable homologies with the β' and β subunits, respectively, of prokaryotic RNA polymerases. The gene encoding the third-largest subunit (henceforth designated as subunit 3) has so far been cloned from only four organisms (5–8), but it appears that the subunit 3 has partial similarity in primary structure with the α subunit of prokaryotic RNA polymerases (8). The roles of these three major subunits of eukaryotic RNA polymerases can therefore be predicted from the knowledge of the functions of E.coli RNA polymerase subunits: the catalytic site is located on β, while β' has non-specific binding activity to DNA; subunit α links these two large subunits into core enzyme structure (reviewed in refs. 9,10) and in addition, carries the contact site with various transcription factors (reviewed in refs. 11,12). In addition to these three core enzyme subunits, prokaryotic RNA polymerase holoenzyme contains one of the multiple species of σ subunit, which plays a major role in promoter recognition (13,14). In eukaryotes, low levels of homology to the σ subunit were found in general transcription factor, TFIIID, and RNA polymerase II subunit RPB4 (15,16). Determination of the role(s) for each subunit of eukaryotic RNA polymerases was hitherto difficult due to the complexity in structure and to the low yield of purified enzymes, but recent progress in cloning of the genes encoding RNA polymerase subunits has provided an experimental basis for detailed characterization of each subunit polypeptide.

For the molecular anatomy of eukaryotic RNA polymerase II, we have been concerned with cloning of the genes for the putative subunits of RNA polymerase II from the fission yeast Schizosaccharomyces pombe. We chose S. pombe because genetic analysis is as easy as S.cerevisiae, nevertheless the strategies for gene expression such as transcription initiation mechanism (17) and splicing pattern (18,19) are similar to those in higher eukaryotes. Up to the present, we have isolated the genes, rpb1 and rpb2, encoding the subunit 1 and subunit 2, respectively (3,4). In this report, we describe the protein composition of purified S.pombe RNA polymerase II and cloning of the gene, rpb3, for subunit 3. By sequence analyses of both the genomic clones and

* To whom correspondence should be addressed
parts of PCR-amplified cDNA, we determined the gene organization and the protein structure of the *S. pombe* RNA polymerase II subunit 3 and compared the structure with those of the corresponding subunits from other organisms.

**MATERIALS AND METHODS**

**S. pombe** strain and medium

A wild-type strain 972h~ of *S. pombe* provided by M. Yamamoto (Univ. Tokyo) was used throughout this study, and was grown in YPD medium (2% yeast extract, 1% Bacto-peptone (Difco Laboratories), 2% glucose).

Buffers and reagents

TGED buffer contained 50 mM Tris–HCl (pH 7.8 at 4°C), 25% glycerol, 0.5 mM EDTA, and 1 mM DTT. Buffer A contained 75 mM Tris–HCl (pH 7.8 at 4°C), 6% glycerol, 0.15 mM EDTA, 1.5 mM DTT, and 150 mM ammonium sulfate, while buffer B consists of 50 mM Tris–HCl (pH 7.8 at 4°C), 0.1 mM EDTA, 1 mM DTT and 200 mM ammonium sulfate. Protease inhibitor mixture contained (per liter): 100 mg benzamidine-HCl (Wako Chemicals), 10 mg aprotinin (Sigma), 10 mg pepstatin A (Funakoshi), 10 mg leupeptin (Funakoshi), 10 mg antipain (Sigma), 50 mg N-α-p-tosyl-L-lysine chloromethylketone (Sigma), 10 mg N-tosyl-L-phenylalanine chloromethylketone (Sigma), and 174 mg phenylmethanesulfonil fluoride (PMSF) (CLABIOCHEM). α-amanitin (CALBIOCHEM) was added to a non-specific transcription assay at 50 μg/ml.

The medium used for reverse phase HPLC column chromatography was 0.1% trifloroacetic acid (Wako Chemicals) containing either 10 or 60% acetonitril. Electro-elution buffer of proteins from gels contained 0.125 M Tris–HCl (pH 6.8), 0.1% SDS and 5% glycerol. V8-digestion buffer contained 50 mM Tris–HCl (pH 7.8) and 0.1% SDS.

**Purification of RNA polymerase II**

*S. pombe* RNA polymerase II was purified by the modified method of RNA polymerase purification from wheat germ (20). In brief, *S. pombe* was grown at 30°C with aeration to 5 X10^7 cells per ml. About 250 g of the cells was harvested by centrifugation (8,000 × g for 10 min at 4°C) from 25 l of the culture, suspended in two volumes (about 500 ml) of buffer A containing protease inhibitors, and disrupted repeatedly with Mini Lab (RANNIE) under 90 MPa pressure for each 4 min. Crude extract was obtained after removing particulate materials by centrifugation at 28,000 × g for 30 min at 4°C. One percent volume of 10% Polyvin P solution was added to the crude extract and after stirring for 1 hour at 4°C, the precipitates were recovered by centrifugation at 15,000 × g for 20 min, and then extracted with 1 l of buffer B containing protease inhibitors. After centrifugation at 15,000 × g for 20 min, 36.1 g of solid ammonium sulfate per 100 ml was added to the supernatant. After stirring for 20 min at 4°C, the precipitates were collected by centrifugation (15,000 × g for 20 min) and resuspended in 200 ml of TGED containing protease inhibitors. After the ammonium sulfate concentration was reduced to 100 mM by adding TGED, the sample was centrifuged at 28,000 × g for 20 min and loaded onto a DEAE-Sephadex A-25 column (bed volume, 300 ml) equilibrated with TGED containing 50 mM ammonium sulfate and 1 mM PMSF. After washing the column with six times volume of TGED containing 50 mM ammonium sulfate and 1 mM PMSF, proteins were eluted with a 1.8 l linear gradient of ammonium sulfate from 50 to 400 mM in TGED buffer. Each fraction was assayed for non-specific transcription activity. Pooled RNA polymerase II fractions eluted at about 150 mM ammonium sulfate were applied onto a Q Sepharose FF column (20 ml; Pharmacia), and eluted with a 100 ml linear gradient of 150–450 mM ammonium sulfate in TGED. RNA polymerase II was eluted at about 250 mM ammonium sulfate. In some experiments, RNA polymerase II was further purified by Superose 6 gel filtration column chromatography in TGED containing 100 mM ammonium sulfate.

**Non-specific transcription assay**

Reaction mixture contained in 40 μl: 50 mM Tris–HCl (pH 7.8 at 37°C), 2 mM MnCl₂, 0.5 mM DTT, 50 mM ammonium sulfate, 0.5 mM each of ATP, GTP, CTP, 7 μM UTP, 0.2 μCi [³²P]UTP (Amersham), 3 μl of the enzyme in TGED, and 2 μg of heat-denatured calf thymus DNA. α-amanitin was added at 50 μg/ml. Non-specific transcription assays was carried out for 20 min at 30°C and terminated by adding 0.9 ml stop solution containing 5% trichloroacetic acid and 10 mM pyrophosphate. TCA precipitates were collected on a GF/C glass-filter, and washed three times with the stop solution and then with 100% ethanol. The filters were counted for radioactivity with a liquid scintillation spectrometer. One unit of the enzyme activity represents the incorporation of 1 nmole labeled substrate into TCA precipitates under the standard assay conditions. Protein concentration was determined according to Lowry et al. (21).

**Isolation of RNA polymerase II subunits**

About 30 μg of the RNA polymerase II subunit 3 was separated from other polypeptides by SDS-gel electrophoresis according to the method of Laemmli (22) using a discontinuous separation gel consisting of 15 (bottom) and 7.5% (top) polyacrylamide. Electro-elution of proteins from the gel was performed by the method of Toda and Ohashi (23). In brief, about 0.2 ml suspension of homogenized pieces of the subunit 3 gel band was incubated in 1 ml of the electro-elution sample buffer for 30 min at room temperature and then boiled for 5 min at 70°C. The gel suspension was poured into a disposable column (7×60 mm) sealed with dialysis tube at the bottom; after electrophoresis for 3 hours at 2 mA using the Laemmli's electrode reservoir solution (22), the subunit 3 eluted from the gel pieces was recovered in 50 μl sample buffer in the dialysis tube. The recovery of subunit 3 was more than 70%.

**Amino acid sequencing**

Proteins dialyzed against V8-digestion buffer for 2 hours at 4°C were treated with 1 mg of V8 protease per 10 mg sample protein for 12 hours at 37°C. Digested polypeptide fragments were separated by reverse phase chromatography on RPC18 column (Pharmacia). Amino acid sequences of the purified fragments were determined using an Applied Biosystem model 977 system attached on-line with HPLC model 120 protein sequencer.

**Gene cloning and sequencing**

To isolate the *S. pombe rpb3* gene encoding the RNA polymerase II subunit 3, PCR amplification of parts of cDNA was carried out using the degenerated primers designed from the amino acid sequences. Amplified cDNA fragments of *rpb3* were cloned into...
M13mp18 phage vector and sequenced. Using one of the cDNA fragments as a probe, Southern hybridization was performed, and one hybridized DNA fragment containing the entire coding region of rpb3 was cloned into M13mp18.

RNA analysis
Poly(A)^+ RNA was prepared from S. pombe by the standard RNA extraction procedure (24). Northern hybridization analysis was performed after denaturation of poly(A)^+ RNA with glyoxal and dimethyl sulfoxide (24). For primer extension analysis, primers were designed based on the sequence of genomic DNA, end-labeled at 5' termini with [γ-32P]ATP, and elongated using reverse transcriptase (25). To determine splice junctions and poly(A) sites, the test regions of cDNA were amplified by PCR and the PCR products were cloned into M13mp18 for sequencing.

RESULTS AND DISCUSSION
Purification of RNA polymerase II
Cell lysates were prepared from wild-type S. pombe by disruption with a high pressure French press (Mini Lab). Purification of the RNA polymerase II was achieved in four steps: precipitation with Polymin P and elution from the Polymin P precipitates with ammonium sulfate solution; separation of the RNA polymerase II from other class RNA polymerases by DEAE-Sepharose column chromatography (Fig. 1A); fractionation by Q Sepharose FF column chromatography; and final-step purification by Superose 6 gel filtration. Starting from 251 of cell culture, about 1.1 mg of RNA polymerase II was obtained at the step of Q-Sepharose FF chromatography (Table 1). Inhibition of 50% activity of the purified RNA polymerase II was achieved at 2.5 μg/ml α-amanitin. The α-amanitin sensitivity was as high as that of S. cerevisiae RNA polymerase II (26), but was significantly lower than that of RNA polymerase II from higher eukaryotes. Substitutions of one amino acid (aa) residue in the subunit 1 of RNA polymerase II render the enzyme resistant to α-amanitin (27,28). This essential aa residue affecting α-amanitin sensitivity of both S. pombe and S. cerevisiae RNA polymerase II is different from that of other eukaryotic RNA polymerases (for details see ref. 3).

The highly purified S. pombe RNA polymerase II contained more than eleven polypeptides as analyzed by SDS-polyacylamide gel electrophoresis. The molecular weights of these components were estimated to be 210 (subunit 1), 150 (subunit 2), 40 (subunit 3), 34, 33, 25, 20, 16, 15, 13.5 and 13 kDa (Fig. 1B), although the staining intensity of some components is different between two dye. This observation agreed with the proposal that the S. cerevisiae RNA polymerase II is composed of more than 11 subunits (29). The molar stoichiometry of the three large subunits was 1:1:2 for subunit 1:2:3, as measured with a laser densitometer ULTROSCAN XL (LKB) and for both silver- and Coomassie brilliant blue-stained SDS-polyacrylamide gels. After prolonged storage of the purified RNA polymerase II, the subunit 1 was partially converted to a degradation product of 190 kDa devoid of the C-terminal repeated domain (CTD).

Sequencing of the subunit 3 and cloning of the rpb3 gene
The S. cerevisiae RPB3 gene encoding the RNA polymerase subunit 3 did not cross-hybridize with the S. pombe gene (data not shown) as analyzed by Southern hybridization under the same conditions used for cloning of the rpb1 and rpb2 genes (3,4).

For cloning the S. pombe gene for the subunit 3, we then carried out aa sequence analysis of polypeptide fragments generated by limited digestion of the isolated subunit 3 with Staphylococcus aureus V8 protease (the sequences determined are indicated in Fig. 2B). Two species of 20-mer oligodeoxynucleotides with the sequences of (5')ATICCIACIGTICCITA(GA)3' and (5')ATCTCA(A/G)IGGIATCIC(3') were designed from the aa sequences, IPTVAID and LGMIPLD. Using these oligonucleotides as primers, we carried out RT(reverse transcription)-PCR and obtained a single major product of amplified cDNA. After DNA sequence analysis, this fragment of 97 bp (F1 in Fig. 2A) was found to have one open reading frame, which is homologous in aa sequence to a part of S. cerevisiae RNA polymerase II RPB3 (60% identity from aa 41 to 73).

For isolation of a complete genomic clone carrying the entire subunit 3-coding sequence, total S. pombe DNA was digested with various restriction enzymes and analyzed by Southern hybridization using the PCR-amplified cDNA fragment F1 as a probe and under high stringency conditions (for details see ref.
3). As shown in Fig. 3A, one major band was identified for all the restriction enzymes used. A 2.0 kbp Pstl-EcoRI fragment (lane 5 in Fig. 3A; indicated as F2 in Fig. 2A) was size-selected by agarose gel electrophoresis, and cloned into M13mp18. The sequence determination revealed the presence of all the aa sequences determined for the subunit 3 fragments within this cloned DNA fragment (see Fig. 2B). A region of high degree of aa sequence homology (56% identity; the conserved domain A in Fig. 4) was found between this major open reading frame and the S.cerevisiae RNA polymerase II RPB3 (aa 46 to 248). The result strongly suggested that the cloned DNA fragment contained a part of the RNA polymerase II subunit 3 (rpb3) gene of S.pombe. In order to clone the complete gene, we next cloned a 2.5 kbp Pstl-EcoRV genomic DNA fragment (F3 in Fig. 2A) into M13mp18 using the Pstl-EcoRI fragment as a hybridization probe.

Structure of the rpb3 gene

Nucleotide (nt) sequence was determined for the continuous genomic DNA segment of 2,348 bp between Pstl and EcoRV sites. The outline of this sequence is shown in Fig. 2. The coding frame spanning from nt position 988 to 1,986 encodes a polypeptide of 297 amino acids in length with a high degree of aa sequence homology to the S.cerevisiae RNA polymerase II subunit 3 (see Fig. 4 for aa sequence comparison). We then propose to designate this gene as rpb3, according to the nomenclature for the S.cerevisiae RNA polymerase genes proposed by Nonet et al. (30). After comparison of the sequences of the cDNA clone F1 and the genomic DNA clones covering the corresponding region, the rpb3 gene was found to contain at least one intron (from nt 1,161 to 1,206). This finding is in good agreement with the result of direct aa sequence analysis of one V8 fragment covering aa 40–48 (Fig. 2B).

Systematic analysis of the exon-intron organization was then extended to other regions by sequencing cDNA clones, which cover different parts of rpb3 mRNA, with use of primers designed from the genomic DNA sequence of rpb3. Results indicated that the coding frame of rpb3 was interrupted by two introns (another from nt 1,068 to 1,110). In both 5' and 3' boundaries of these two introns, there are the consensus sequences of the S.pombe intron-exon junctions (boxed sequences in Fig. 2B). In sharp contrast to S.cerevisiae, the rpb1, rpb2 and rpb3 genes of S.pombe all contain several introns at their N-terminal proximal regions (3,4; this paper).

Detailed Southern analysis of the genomic DNA digested with various restriction enzymes showed that this gene is present as a single copy in the S.pombe genome (Fig. 3A).

Transcription organization

Northern analysis demonstrated that the size of rpb3 transcript is about 1.2 kb in length (Fig. 3B). The start site of the transcript determined by primer extension analysis using primer #1 (nt 877–894) was located at nt 792 (Fig. 3C). We concluded that ATG at nt 988 is the start codon because: 1) the predicted start codon is located within the first exon (no intron was found between the transcription start site and this putative start codon) (see Fig. 2B); 2) no other ATG codon exists upstream the putative start codon on the same open reading frame; 3) the N-terminal proximal region of subunit 3 from the initiator Met to the first conserved domain (domain A in Fig. 4) is as large as that of S.cerevisiae RNA polymerase II subunit 3, even though homology is not so high in this region between S.pombe and S.cerevisiae; and 4) the sequence near the putative initiation codon fits well to the Kozak's rule (31) of the consensus sequence for translation initiation.

Finally, we determined the nt sequence near the 3' end of rpb3
Figure 3. [A] Southern hybridization of *S. pombe* genomic DNA. Three μg of total *S. pombe* DNA was digested with various restriction enzymes and subjected to Southern analysis under high stringency hybridization conditions using the PCR-amplified cDNA fragment F1 as a probe. Restriction enzymes used were: *EcoRI* (lane 1); *BamHI* (lane 2); *PstI* (lane 3); *EcoRI/BamHI* (lane 4); and *EcoRI/PstI* (lane 5). [B] Northern blot analysis of *rpb3* transcript. Seven μg of poly(A)⁺ RNA was subjected to Northern analysis. The blot was hybridized with the cDNA fragment F1 containing a part of *rpb3*. [C] Primer extension analysis. 32P-labeled primer #1 was hybridized with 3 μg of poly(A)⁺ RNA and then elongated by AMV reverse transcriptase. Arrowhead indicates the major start site of *rpb3* transcription. The DNA sequence ladders were obtained using a cloned *rpb3* DNA fragment.

transcript. For this purpose, we amplified a 3’-terminal region from one of our cDNA clones using primer #2 (nt 1833 – 1852 in Fig. 2B) and the primer b with the sequence (5’)-GCCGCGGCGTAAAGTTT(3’), which hybridizes to poly(A)⁺ tail (see ref. 3 for details of the primer b). From sequence analysis of 3 independent clones generated from the PCR products, the poly(A)⁺ site was identified at nt 2,043 or 2,044 (as marked in Fig. 2B). The putative poly(A)⁺ signal may be located from 2,029 to 2,037 (double-underline in Fig. 2B).

The length of mature mRNA (1,172 b plus poly(A)⁺ tail) predicted from the cDNA sequence analysis is in good agreement with the result (1.2 kb) obtained by Northern analysis (see Fig. 3B).

Structure of the subunit 3

The predicted aa sequence of *S. pombe* RNA polymerase II subunit 3 was compared with those of other organisms: *S. cerevisiae* RPB3 (5) and RCP40 (6); *H. sapiens* RPB33 (7); *T. thermophila* CnJC (8); and *E. coli* α subunit (32). As illustrated in Fig. 5A, we discovered four structurally conserved regions (domains A to D) within all these proteins except for *S. cerevisiae* RCP40, which lacks the domain B [the conserved regions I, II and III proposed by Martinale (8) correspond to domains A, C and D, respectively]. The domains A and D exist even in the α subunit of *E. coli* RNA polymerase, suggesting that these two domains are involved in some common and essential functions associated with the RNA polymerase. On the other hand, the domains B and C are conserved only in eukaryotic RNA polymerases and thus considered to be involved in function(s) specific for eukaryotes.

A temperature-sensitive *E. coli* mutant defective in RNA polymerase assembly has a mutation *rpoA112* in domain A of the α-subunit gene (33, 34), while a *S. cerevisiae* mutant *rpb3-1* carries double mutations in the *RBP3* gene, one in domain A and another in domain C (35). Thus, the domain A may play an important role in subunit-subunit contact of RNA polymerase. As shown in Fig. 4B, the domain B has a putative metal-binding sequence, CXCX,CX,C (from aa 90 to 99 in the *S. pombe* sequence). Since this domain B sequence is not present in the corresponding subunits of RNA polymerases I and III, this motif
may be related to function(s) specific for RNA polymerase II. According to the recent studies with the \( \alpha \)-subunit C-terminal deletion mutants of \textit{E. coli} RNA polymerase (36), the N-terminal fragments of \( \alpha \) subunit containing the domain D can still be assembled into enzymatically active pseudo-core enzymes but further deletion from the C-terminus renders mutant \( \alpha \) inactive in the core assembly. These observations suggest that the domain D is required for the formation and/or stability of RNA polymerase. In the domain D sequence of \textit{S.pombe} (and also \textit{H.sapiens}), we discovered a leucine zipper-like motif (aa 249–263). This finding may support a role of the domain D in subunit-subunit contact. The C-terminal region of \textit{E. coli} RNA polymerase \( \alpha \) subunit downstream from aa 235 carries the protein–protein contact site I with the class I activator proteins. In the domain D sequence of \textit{E.coli} RNA polymerases plays a role in contact with transcription factors. This leucine zipper-like motif in the domain D of eukaryotic RNA polymerase and transcription factors may be different between prokaryotes and eukaryotes. It is not excluded yet that this leucine zipper-like motif in the domain D of eukaryotic RNA polymerases plays a role in contact with transcription factors.

**ACKNOWLEDGEMENTS**

We thank N.Fujita and N.Shimamoto for advice and discussion. This study was supported by Grants-in-Aid from the Ministry of Education, Science and Culture of Japan, Joint Research Programs of The Graduated University for Advanced Studies, and NIG (National Institute of Genetics) Cooperative Research Programs.

**REFERENCES**