Molecular Cloning and Characterization of cDNA for Eukaryotic Transcription Factor S-II*

Shoji Hirashima, Hiroshi Hirai, Yoshinobu Nakanishi, and Shunji Natori

From the Faculty of Pharmaceutical Sciences, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

Complementary DNA (cDNA) clones encoding a transcription factor S-II were isolated and characterized. The primary structure of S-II was determined by nucleotide sequence analysis of these clones. The predicted primary structure was consistent with the model that we proposed previously from the results of biochemical analyses of S-II. Using these clones as probes, we analyzed the mRNA for S-II. RNA blot analysis demonstrated the presence of four species of mRNA that hybridized with S-II cDNA in Ehrlich ascites tumor cells. This is the first evidence of polymorphism of mRNA encoding a transcription factor of RNA polymerase II. The results of analysis of the genomic structure suggested that the polymorphism of mRNA may be due to alternative splicing, or differences in initiation or termination of transcription.

Transcription is one of the most important steps in regulation of gene expression. In eukaryotic cells, RNA polymerase II is responsible for the transcription of protein-coding genes (class II genes). Studies on cell-free transcription systems have demonstrated that RNA polymerase II transcribes class II promoters accurately and efficiently only in the presence of additional factors contained in crude cellular extracts.

One way to identify these transcription factors is fractionation of cellular extracts to separate the components that are essential for accurate transcription and reconstitution of a transcription system from these components. At least four factors (TFIIA, TFIIB, TFIIIE, and TFIID) have been shown to be required for accurate transcription initiation of the adenovirus 2 major late promoter (2-5). The initiation factor TFIIA (or (AB)) has been purified from calf thymus and shown to be replaceable for human TFIIA (6). Recently, Reinberg and Roeder (7) reported the purifications of the two initiation factors, TFIIIB and TFIIIE, from HeLa cells. They also partially purified two other factors, TFIIA and TFIID, and characterized these four factors in a highly purified reconstitution system (8). These four factors appear to be "general" factors required for accurate initiation of transcription.

Mutation analysis in cell-free transcription systems or in exogenous gene expression systems using living cells has shown that the promoters of protein-coding genes consist of several distinct elements. The TATA element, found in most promoters of the genes transcribed by RNA polymerase II at positions -25 to -30, appears to be responsible for determination of the initiation site of transcription (9). Usually one or several upstream elements are required for maximum expression of a particular gene (10). These upstream elements seem to be recognized and activated by sequence-specific DNA binding factors. Recently, several such factors have been identified and a few of them have been purified (11, 12). Most of these factors appear to activate transcription by stimulating the rate of transcription initiation.

Many studies on initiation factors have been reported, but little is known about the factors affecting elongation. We purified three proteins from Ehrlich ascites tumor cells, named S-II, S-II', and S-I(b), which specifically stimulate RNA polymerase II (Refs. 13-15; see Ref. 16 for a review), and demonstrated that antibody to S-II selectively inhibited a-amaminatin-sensitive RNA synthesis in isolated nuclei (17) and also accurate transcription of truncated adenovirus 2 DNA containing the major late promoter in a HeLa cell lysate (18). Moreover, we showed that when S-II was added to an in vitro transcription system with purified RNA polymerase II, it was incorporated into both the initiation and the elongation complex (19). These results strongly suggested that S-II is a transcription factor, and is effective in the initiation and/or elongation step of transcription.

Recently, Reinberg and Roeder (20) reported that a factor named TFIIS stimulates specific transcription from the adenovirus 2 major late promoter by increasing the rate of RNA chain elongation. Rappaport et al. (21) purified S-II from calf thymus and demonstrated that it could replace the TFIIS fraction in a HeLa cell transcription system reconstituted with purified factors and RNA polymerase II. These findings suggest that S-II, which is the only factor so far purified that affects elongation of transcription mediated by RNA polymerase II, is in fact a transcription factor.

Besides the adenovirus major late gene, some cellular genes may also be regulated at the elongation step of transcription. Thus, studies on elongation factors in regulation of gene expression seem important.

This paper reports the isolation and characterization of cDNA clones of S-II. We screened a cDNA library of Ehrlich ascites tumor cells using an oligonucleotide probe corresponding to the partial amino acid sequence of S-II, and isolated three hybridization positive clones. The primary structure of S-II was determined by nucleotide sequence analyses of these clones. We analyzed the mRNA for S-II in Ehrlich ascites tumor cells and the genomic sequence of the S-II gene using these clones as probes, and demonstrated the presence of...
multiple mRNAs of different sizes that hybridized to cDNA of S-II.

**MATERIALS AND METHODS**

**RESULTS**

**Determination of Partial Amino Acid Sequences of S-II—**
To synthesize oligodeoxyribonucleotide probes for isolation of S-II cDNA, we first determined the partial amino acid sequence of S-II. For this, highly purified S-II was digested with trypsin and the resulting peptides were separated by reverse-phase high performance liquid chromatography. Of the various peaks shown in Fig. 1a that were obtained reproducibly, peaks I and II were each pooled, lyophilized, and their partial sequences were determined (Fig. 1b). We then synthesized two sets of oligodeoxyribonucleotides corresponding to the cDNA sequences of parts of the peptide of peak II; namely, oligodeoxyribonucleotide set I (Oligo I) corresponding to Asp-Asp-Tyr-Val-Ala and Oligo II corresponding to Asp-Glu-Glu-Glu-Leu. Each oligonucleotide probe was a mixture of 32 different tetradecamers, and represented all possible complementary sequences excluding the third nucleotide residues for the codon of the carboxyl-terminal amino acid, as shown in Fig. 16.

**Isolation of cDNA Clones for S-II—** Oligo I was used to screen a cDNA library of poly(A)* RNA of Ehrlich ascites tumor cells constructed in the vector of λgt10. We screened 3,000 recombinants, probing duplicate filters with Oligo I. Of these, 10 hybridization positive recombinants were picked up and screened again. We selected seven independent clones, extracted their DNAs, and analyzed them by Southern blot hybridization using both Oligo I and Oligo II as probes. Only one clone, ASII-1, contained an insert that hybridized to both probes. This insert was 0.8 kb long. We subcloned it into a plasmid vector (pUC8) after cutting it with EcoRI, and designated the plasmid pSII-1. RNA blot hybridization analysis of total cellular RNA from Ehrlich ascites tumor cells revealed that the length of the major mRNA for S-II was about 2.8 kb. We, therefore, tried to isolate clones with a longer insert. Five hundred thousand recombinants were screened again using the 0.8-kb insert of pSII-1 as a probe, and 7 hybridization positive clones were picked up. One of them, ASII-2, had a 2.8-kb insert, and another clone, ASII-3, had a 1.3-kb insert. Of the seven clones, two were identical to ASII-1, and four to ASII-3. The inserts of ASII-2 and ASII-3 were subcloned into the plasmid vector pBSM13 and designated as pSII-2 and pSII-3, respectively.

**Characterization of cDNA Clone for S-II—** The restriction map and sequencing strategy of cDNA clones for S-II were as shown in Fig. 2. Nucleotide sequences were determined by the dye-sequencing method (30) after subcloning each DNA fragment into M13 vector (31). The complete nucleotide sequences of the inserts of pSII-1, -2, and -3 are shown in Fig. 3. No significant homology was found between the first 78 bases of pSII-2 and the corresponding regions of pSII-1 and pSII-3 by direct sequence comparison. It is uncertain whether this difference reflects the presence of different mRNAs or whether pSII-2 is an artifact formed during cDNA cloning. Apart from this region, the sequences of pSII-1 and pSII-3 were included in the sequence of pSII-2. There was a long open reading frame and the two amino acid sequences corre-

* Portions of this paper (including "Materials and Methods," Table 1, and Figs. 1, 2, 5, and 7) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full-size photocopies are included in the microfilm edition of the Journal that is available from Waverly Press.
S-I(b) is a part of S-II including the 21-kDa domain, and this domain should contain most of the Tyr, Phe, and Cys residues in S-II, since S-I1 and S-I(b) have almost the same numbers of these residues. This domain was shown to have the same activity as S-I1 and S-I(b) in stimulating RNA polymerase II.

In contrast, the 18-kDa domain was shown to contain phosphorylation sites, which are phosphorylated in S-II'.

The amino acid sequence predicted from the nucleotide sequences were compared with the proposed structure. As shown schematically in Fig. 5, the stretch of 125 amino acids from the amino-terminal residue contains no Tyr, Phe, or Cys residue, these amino acid residues being concentrated in the carboxy-terminal half of the protein. Moreover, most sites that are potentially cleavable by chymotrypsin are also located in this 21-kDa region. This was also expected from our previous results, since this domain was much more sensitive to chymotryptic digestion than the 18-kDa domain. These structures fit well to our previous model. Thus, it is likely that the 21-kDa domain with activity to stimulate RNA polymerase II is located in the carboxy-terminal half of the S-II molecule.

Analysis of S-II mRNA in Ehrlich Ascites Tumor Cells—We analyzed poly(A)⁺ RNA of Ehrlich ascites tumor cells by Northern blot hybridization to determine how many species of mRNA can hybridize to the cDNA clone for S-II. Total RNA was prepared from Ehrlich ascites tumor cells in the presence of guanidine HCl, and poly(A)⁺ RNA was purified on oligo(dT)-cellulose and separated by formaldehyde-agarose gel electrophoresis. The RNA was then transferred to a nitrocellulose filter and hybridized with ³²P-labeled fragment A, which covers the whole coding region, as shown at the bottom.
FIG. 4. Analysis of in vitro translation products generated after in vitro transcription of cloned DNA. pSII-2 (lane 1), pSII-12 (lane 2), and pBSM15* (lane 3) DNAs were digested with BamHI and transcribed at 37 °C for 30 min with nucleotide triphosphates by T7 RNA polymerase. The reaction mixture was extracted with phenol/chloroform and RNA was recovered by ethanol precipitation. The resulting RNA was translated in the presence of [35S]methionine in a rabbit reticulocyte lysate translation system. The in vitro translation products were subjected to SDS-polyacrylamide gel electrophoresis followed by fluorography. Lanes 1–3 were exposed for 1 h. Lane 4 shows 1 µg of authentic S-II electrophoresed on the same gel and stained by Coomassie Brilliant Blue. Bovine serum albumin (BSA) was derived from the storage buffer of S-II. Arrowheads indicate the following molecular mass markers: bovine serum albumin (69 kDa), ovalbumin (45 kDa), α-chymotrypsinogen (25 kDa).

FIG. 6. Analysis of S-II mRNA in Ehrlich ascites tumor cells. Total cellular RNA and poly(A)+ RNA were prepared from Ehrlich ascites tumor cells as described under "Materials and Methods." Total cellular RNA (40 µg) (lane 1) and poly(A)+ RNA (5 µg) (lanes 2–4) were subjected to formaldehyde-agarose gel electrophoresis, blotted onto a nitrocellulose filter, and hybridized with 32P-labeled EcoRI-insert of pSII-1. The results are shown in Fig. 7. When digested with EcoRI (lane E), which does not cut the known cDNA sequences, one major band (7.6 kb) and two minor bands (14 and 2.1 kb) were observed. When digested with HindIII (lane H), which does not cut the EcoRI-insert of pSII-1 but cuts at nucleotides 1661 and 2143 in the 3′-noncoding region, one major band (8.3 kb) and two minor bands (5.4 and 4.6 kb) were observed. The intensities of the major bands were compared quantitatively with that of a cDNA clone equivalent to a single copy per haploid genome, as shown in Fig. 7. The intensities indicated that the copy number of the S-II gene was about one per haploid genome.

These results suggest that the four mRNA species that hybridized with the cDNA clone for S-II were derived from the same gene. Different sized mRNAs may be produced from a single gene by alternative splicing, or use of different initiation or termination sites for transcription. For investigation of the mechanisms involved, the genomic sequences of the S-II gene must be characterized.

DISCUSSION

We have isolated cDNA clones for transcription factor S-II, and determined the primary structure of S-II by analyzing the nucleotide sequence of these clones. Recent studies showed that the S-II factor affects the elongation step of transcription from the major late promoter of adenovirus serotype 2 in a cell-free accurate transcription system (20,
21). This seems to be the first report of isolation of a cDNA clone for RNA polymerase II transcription factor in a higher eukaryote, since purification of other general transcription factors has proved difficult. We recently developed a method for large-scale purification of S-II from Ehrlich ascites tumor cells (15), and Rappaport et al. (21) also reported a method for large-scale purification of S-II from calf thymus. Establishment of such methods is important because a large amount of protein is needed for further characterization of this factor. In a preliminary computer search, we could not find any sequence that was significantly homologous to that of S-II in either a nucleic acid or protein sequence library.

The primary structure of S-II predicted from cDNA was consistent with our previous results obtained by analyzing the two domains of the S-II molecule. The structural characteristics of the 21-kDa domain can be summarized as follows. (i) This domain, which has activity to stimulate RNA polymerase II, contains most of the Tyr, Phe, and Cys residues. (ii) It should contain many more sites than the other 18-kDa domain, which contains phosphorylation sites. The present study showed that the 21-kDa domain was located in the carboxyl-terminal half of the S-II molecule. The presence of 8 Cys residues in this area suggests that this protein has a unique local secondary structure.

Recently, TFIIIA, an RNA polymerase III transcription factor, and several other DNA-binding proteins have been reported to contain unique structures, so-called the “finger” structure. Thus, S-II migrates slower in gel than other proteins encoding regions having DNA binding activity.

The molecular mass of S-II predicted from the nucleotide sequence was 32.9 kDa, whereas that of the translation product synthesized in vitro in a transcription-translation system was 38 kDa, as determined by SDS-polyacrylamide gel electrophoresis. Thus, S-II migrates slower in gel than other proteins with the same molecular mass.

RNA blot hybridization experiments showed that Ehrlich ascites tumor cells contain at least 4 species of mRNA of different sizes that hybridize with the isolated cDNA clone for S-II. Therefore, it is possible that there are several species of mRNA encoding proteins related to S-II. We found different nucleotide sequences on the amino-terminal side of the protein encoding regions of pSII-2 and two other clones. This difference might reflect the polymorphism of mRNA for S-II-related proteins. However, a possibility remains that this difference is due to an artifact in cDNA cloning. Previously, we purified a similar transcription factor termed S-I(b) with a molecular mass of 24 kDa (14), which seems to be a part of S-II. Although, a possibility remains that S-I(b) is produced from S-II post-translationally, this protein might be encoded by one of these mRNAs. For more precise insight into the structure of S-II-related proteins and their mRNAs, it will be important to isolate cDNA clones for these mRNAs and analyze their sequences. The polymorphism of S-II mRNA may be important in understanding the regulation of eukaryotic transcription at the elongation level.

Analysis of genomic sequences revealed that the S-II gene is probably a single copy gene per haploid. This suggests that four mRNA species hybridizable to the cDNA of SII are derived from the same gene. Various mechanisms for production of mRNAs of different sizes from the same gene are possible. However, care must be taken in estimating the copy number of a specific gene. Various quantitative factors must be taken into consideration, such as the efficiency of transfer of DNA from gel to a nitrocellulose filter and the length of exons hybridizable with the probe. Thus, more precise experiments are needed before definite conclusions can be drawn.

Acknowledgments—We wish to thank Drs. D. Engel and P. Lengyel for providing us with a cDNA library of Ehrlich ascites tumor cells.

REFERENCES

Molecular Cloning of S-II cDNA

Molecular Cloning and Characterization of cDNA for Erythrocrypt Transcription Factor II

Koji Inohara, Hiroshi Hira, Yoshio Nakanishi, and Shoji Satoro

MATERIALS AND METHODS

Gel filtration and determination of its partial amino acid sequences — PI-1 was purified from Ehrlich ascites tumor cells as described previously [15]. A purified PI-1 solution (1 ml) was digested with trypsin (trypsin, grade I, Worthington) by the method of Isshii et al. [16]. The digested solution was applied to a Sephadex G-150 column eluted with 0.1 M sodium pyrophosphate, 0.05 M sodium chloride, 0.05 M sodium borate, and 0.005% sodium azide in 0.005 M Tris-HCl buffer, pH 7.4 at 37°C at an enzyme-protein ratio of 1:10 by weight. This sample was directly applied to a Sephadex G-150 column of Synchropak HP-IC-181 (4.1 x 150 cm, Synchrom Inc., Liden, MA) connected to a Gurney high-performance liquid chromatography (HPLC) system. The column was eluted with a linear gradient of 3 to 40% acetonitrile in 0.1% trifluoroacetic acid at a flow rate of 0.5 ml/min. The absorbance of the eluate at 280 nm was monitored.

The purified peptides were collected, lyophilized, and subjected to automated sequence analysis in an Applied Biosystems 477A Protein Sequencer using the procedures described by Mankouri et al. [17]. The generated amino acid sequences were identified unambiguously in a region of a proteinase digest of S-II (Fig. 1).

Cloning procedure — Oligodeoxyribonucleotides designed to correspond to partial amino acid sequences of S-II were synthesized by a modified trityl cation method [18], and were isolated from a preparation of E. coli cells by the method of Kawasaki and Nakamura [19].

A cDNA library prepared from poly(A)^+ RNA of Ehrlich ascites tumor cells in a vector, lambda gt11, was a generous gift from Drs. T. Engel and P. Hayashizawa (Takai University). Screening was done by standard procedures (21) with several modifications. About 2,500 recombinant plates were plated with 982 cells at a density of about 10,000 plaques per 150-mm plate. Two nitrocellulose replica filters (Schleicher and Schuell, BA851) were made per 150-mm plate by the application method described by Woo [22]. The filters were baked at 80°C for 2 h. Then they were incubated in 1xSSC (15% 1xSSC, 0.05 M Na-phosphate, pH 7.0) containing 0.1% SDS at 65°C, prehybridized for 3 h at 65°C in a water bath (Schleicher and Schuell, BA851). Polyvinylpyrrolidone, 0.02 M NaCl, containing 0.1 mg/ml of single stranded salmon sperm DNA and 0.05% sodium pyrophosphate and hybridized for 16 h at 65°C in the same solution containing 16 g/ml of radiolabeled probe. The specific activity of the probe was 0.5-1x10^6 cpm/µg. After hybridization, the filters were washed twice in 1xSSC, 0.1% SDS at 65°C for 5 min each. They were then dried briefly at room temperature, and exposed in Kodak XAR film in the presence of two intensifying screens at -80°C.

Positive plaques were picked out and re-screened under the same conditions using two different probes.

To screen clones containing long inserts, we used a 2.4 kb insert DNA of the pG-11 plasmid as a probe, which is a cDNA clone of S-II, which was obtained from the first screening. For this, 5x10^6 recombinant plaques were plated at a density of 50,000 plaques per 150-mm plate, and each plaque filter was made according to the same method. Hybridization was performed in 1xSSC, 0.1% SDS at 65°C, 50% formamide, Schleicher and Schuell, BA851, 0.5 mg/ml single stranded salmon sperm DNA, 0.05% sodium pyrophosphate. The filters were washed twice in 1xSSC, 0.1% SDS at 65°C and then hybridized in the same solution containing 6xSSC (1xSSC, 0.1% SDS labeled DNA. The probe was washed subsequently at 50°C and then exposed to XAR film in the same condition as above.

In situ hybridization and translation — This was done using pG-11, a cDNA clone containing the 5'-end, and pG-12, a newly constructed fragment containing the 3'-end, by the method of Koprowski et al. [23] using the EcoRI-Hind III fragment of a cDNA clone of S-II. (Fig. 1). The translation of the open reading frame was performed in a rabbit reticulocyte lysate (STRATAGEN) containing two promoters for viral RNA polymerase, pG-12 and pG-11 DNA, which were linearized with Bg II. Translation was stopped by the addition of 0.2 M of a buffer (24) containing 0.1 M of sodium chloride, 0.05 M of potassium chloride, 0.01 M of magnesium chloride, and 0.002 M of diethylenetriaminepentaacetic acid recommended by the manufacturer (STRATAGEN), except that the concentration of DTT was reduced to 1.2 M of sodium chloride, 0.01 M of potassium chloride, 0.002 M of diethylenetriaminepentaacetic acid recommended by the manufacturer (STRATAGEN), except that the concentration of DTT was reduced to 1.2 M of potassium chloride, 0.002 M of diethylenetriaminepentaacetic acid. 1xSSC containing 6xSSC (1xSSC, 0.1% SDS labeled DNA. The probe was washed subsequently at 50°C and then exposed to XAR film in the same condition as above.

In situ hybridization and translation — This was done using pG-11, a cDNA clone containing the 5'-end, and pG-12, a newly constructed fragment containing the 3'-end, by the method of Koprowski et al. [23] using the EcoRI-Hind III fragment of a cDNA clone of S-II. (Fig. 1). The translation of the open reading frame was performed in a rabbit reticulocyte lysate (STRATAGEN) containing two promoters for viral RNA polymerase, pG-12 and pG-11 DNA, which were linearized with Bg II. Translation was stopped by the addition of 0.2 M of a buffer (24) containing 0.1 M of sodium chloride, 0.05 M of potassium chloride, 0.01 M of magnesium chloride, and 0.002 M of diethylenetriaminepentaacetic acid recommended by the manufacturer (STRATAGEN), except that the concentration of DTT was reduced to 1.2 M of sodium chloride, 0.01 M of potassium chloride, 0.002 M of diethylenetriaminepentaacetic acid recommended by the manufacturer (STRATAGEN), except that the concentration of DTT was reduced to 1.2 M of potassium chloride, 0.002 M of diethylenetriaminepentaacetic acid.

In situ hybridization and translation — This was done using pG-11, a cDNA clone containing the 5'-end, and pG-12, a newly constructed fragment containing the 3'-end, by the method of Koprowski et al. [23] using the EcoRI-Hind III fragment of a cDNA clone of S-II. (Fig. 1). The translation of the open reading frame was performed in a rabbit reticulocyte lysate (STRATAGEN) containing two promoters for viral RNA polymerase, pG-12 and pG-11 DNA, which were linearized with Bg II. Translation was stopped by the addition of 0.2 M of a buffer (24) containing 0.1 M of sodium chloride, 0.05 M of potassium chloride, 0.01 M of magnesium chloride, and 0.002 M of diethylenetriaminepentaacetic acid recommended by the manufacturer (STRATAGEN), except that the concentration of DTT was reduced to 1.2 M of sodium chloride, 0.01 M of potassium chloride, 0.002 M of diethylenetriaminepentaacetic acid recommended by the manufacturer (STRATAGEN), except that the concentration of DTT was reduced to 1.2 M of potassium chloride, 0.002 M of diethylenetriaminepentaacetic acid.