The complete primary structure of the human snRNP E protein

David R.Stanford, Maria Kehl¹, Caroline A.Perry, Eileen L.Holicky, Scott E.Harvey,
Anne M.Rohleder, Kai Rehder Jr, Reinhard Luhrmann² and Eric D.Wieben*

Department of Biochemistry and Molecular Biology, Mayo Clinic/Foundation, Rochester, MN 55905, USA, ¹Max-Planck-Institut für Biochemie, Genzentrum and ²München Max-Planck-Institut für Molekulare Genetik, Otto-Warburg Laboratorium, Ihnestrasse 73, D-1000 Berlin 33, FRG

Received July 28, 1988; Revised and Accepted October 17, 1988

ABSTRACT

The snRNP E protein is one of four "core" proteins associated with the snRNAs of the U family (U1,U2,U4,U5, and U6). Screening of a human teratoma cDNA library with a partial cDNA for a human autoimmune antigen resulted in the isolation of a cDNA clone containing the entire coding region of this snRNP core protein. Comparison of the 5' end of this cDNA with the sequences of two processed pseudogenes and primer extension data suggest that the cDNA is nearly full length. The longest open reading frame in this clone codes for a basic 92 amino acid protein which is in perfect agreement with amino acid sequence data obtained from purified E protein. The predicted sequence of this protein reveals no extensive similarity to other snRNP proteins, but contains regions of similarity to a eukaryotic ribosomal protein.

INTRODUCTION

Small nuclear RNAs of the U family have been shown to be required for a variety of RNA processing reactions in eukaryotic cells (1,2). In vivo, these RNAs are found complexed with a small number of discrete proteins. The importance of the protein components of snRNPs in modulating the function of the snRNP is underscored by the finding that snRNP proteins are required to obtain specific interaction of U1 RNA with 5' splice junctions in vitro (3). However, the specific roles of individual snRNP proteins in RNA processing reactions are unknown. Progress towards understanding the function of each protein will be facilitated by detailed structural information.

Interest in the structure of these proteins has been heightened by the discovery that many snRNP proteins are recognized by antibodies produced by patients with autoimmune disorders (4,5). In at least one case, structural studies have revealed that an autoimmune antigen, the U1 specific 70K protein, has structural similarity to a viral coat protein, suggesting a possible mechanism for the initiation of the autoimmune response (6). Thus, there are compelling reasons to investigate the structure of the snRNP proteins, and considerable progress has been made in this area. cDNA clones corresponding to several snRNP proteins have been cloned and sequenced (7-13), and biochemical and

immunological studies have yielded valuable data concerning the variety and relatedness of these proteins (14-16).

We report here the complete sequences of a cDNA clone for the snRNP E protein and a second member of the E protein multigene family. The E protein is both a snRNP core protein (17) and a known autoimmune antigen (12,18,19). Although the E protein exhibits little amino acid sequence similarity to other snRNP proteins, it does have regions of similarity to a ribosomal protein suggesting possible functional homologies between the spliceosome and the ribosome.

MATERIALS AND METHODS

Materials

Restriction enzymes were purchased from Boehringer Mannheim, New England Biolabs, and Bethesda Research Laboratories. Radioisotopes were supplied by Amersham and ICN. T4 polynucleotide kinase was obtained from New England Biolabs or Boehringer Mannheim. AMV reverse transcriptase was purchased from ProMega and Bethesda Research Laboratories. The vectors M13mp18/19 and reagents for dideoxy-sequencing were obtained from Pharmacia P-L Biochemicals, Inc. The plasmid vectors pGem3 and pGem4 were purchased from ProMega while the vectors BS+/- were purchased from Stratagene. The human teratoma cDNA library was a gift from Dr. J. Skowronski, a human genomic library in Charon 4A was a gift from Dr. T. Maniatis, and a human genomic library in EMBL-3 was purchased from Clontech.

Screening of a human teratoma cDNA library

A gt10 cDNA library from human teratoma cells (20) was screened with nick-translated p281 (10). A single positive plaque was isolated and DNA was prepared as described (21) using the C600 strain of *E. coli* as host. A 1700 nucleotide *Hind*III-*Bgl*II restriction fragment from the lambda clone was gel purified and subcloned into *Hind*III-*Bam*HI digested pUC13. This plasmid clone was subsequently digested with *Eco*RI and a 550 nucleotide fragment was gel purified. The *Eco*RI-*Eco*RI fragment was cloned into *Eco*RI digested pGem4 to yield cDNA clone p11HB1.

Screening of human genomic libraries

Nick-translated p281 was used to screen a lambda Charon 4A library of human genomic DNA (22). Positive clones were plaque-purified and DNA was prepared as described (21). Positive lambda clones were analyzed by restriction mapping and southern blots (23). Four distinct but overlapping clones were isolated (EJ,EK,24,137). Selected restriction fragments from the lambda clones were subcloned into plasmid vectors (either pGem3/4 or pBS+/-). The plasmid clones were restricted and subcloned

into M13mp18 or mp19 for sequence analysis. The 550 nucleotide *EcoRI-EcoRI* insert of p11HB1 was nick-translated and used to screen a lambda EMBL-3 library of human genomic DNA. Restriction fragments from positive clones were subcloned into plasmid vectors and processed for sequencing as described above.

Nucleotide sequencing and analysis

Single-stranded sequencing was as described previously (24). The annealing step for double-stranded sequencing was performed by mixing 10 ng oligodeoxynucleotide with 1 μ g of plasmid DNA, heating to 100°C for 3 min, and then rapid cooling to 4°C for 10 min. After the annealing step double-stranded sequencing followed the protocol for single-stranded sequencing. Sequence manipulation and analysis was accomplished using the following software packages: Cornell Sequence Analysis Package, DNASTAR, and University of Wisconsin Genetics Computer Group (UWGCG) package. Database searching was accomplished using the PROSCAN program of the DNASTAR package with a k-tuple of 2. This program uses the same algorithm and similarity matrix as the FASTP program described by Lipman and Pearson (25). The mean similarity score for the search of E protein versus 8588 sequences in the NBRF protein database was 18.7, with a standard deviation of 5.94.

Figure 1: Partial restriction map of genomic clones representing the 137 family. The filled boxes represent the pseudogene region and the horizontal lines represent human genomic DNA. Clone LH75 was isolated from an EMBL-3 library and clones 137, 24, EJ, EK were isolated from a Charon 4A library. Restriction sites are: E, *EcoRI*; H, *HindIII*; P, *PstI*; Sal, *SalI*; Sst, *SstI*.

Primer extension studies

Primer extensions utilized poly(A)⁺ RNA isolated from HeLa cells (12). An oligodeoxynucleotide (DRS-7, 5'-AAGAGCACACCCGACGCTG-3') was end-labeled with T4 polynucleotide kinase (21) and annealed to poly(A)⁺ RNA at 42°C for 8-12 hr. The resulting hybrids were incubated in the presence of 5 units AMV reverse transcriptase and 1 mM dNTPs for 2 hr at 41°C. Following this reaction the products were ethanol precipitated and electrophoresed on denaturing polyacrylamide gels (7M urea-8% polyacrylamide).

Amino acid sequence analysis

E protein was isolated by electroelution following preparative sodium dodecyl sulfate polyacrylamide gel electrophoresis of a mixture of total proteins from anti-m₃G affinity purified HeLa snRNPs (16,18). Approximately 50 µg of electroeluted E protein was further purified by reverse phase HPLC applying inverse gradient conditions (26) to remove contaminants from gel electrophoresis and staining. The protein containing peaks were pooled, evaporated to dryness and dissolved in 250 µl of freshly prepared 70% formic acid. An equal volume (250 µl) of 20% cyanogen bromide (w/v) in 70% formic

Figure 2: Sequencing strategies and partial restriction maps for cDNA clones p281 and p11HB1 and for genomic clone 137. The heavy lines for the cDNA clones represent the cDNA insert and the lighter lines represent vector. The entire length of clone 137 represents human genomic DNA and the heavier region denotes the pseudogene portion. Each of the clones is aligned with respect to the original cDNA (p281). The arrowed lines represent the extent of sequencing reactions, which proceed 5' to 3' in the direction of the arrows. Restriction sites are: A, *AluI*; E, *EcoRI*; H, *HindIII*; Hc, *HincII*; Nco, *NcoI*; P, *PstI*; Pvu, *PvuII*; R, *RsaI*; S, *Sau3AI*; Sma, *SmaI*; Ssp, *SspI*; Sst, *SstI*; T, *TaqI*; X, *XmnI*.

acid was added and left for 4 hr in the dark at room temperature. Thereafter, the sample was diluted 10-fold with water, concentrated to about 200 μ l, acidified with 0.1% trifluoroacetic acid and applied to reversed phase HPLC. Protein containing peaks were analyzed by amino acid analysis. N-terminal amino acid sequence analysis was performed on a gas-phase sequencer 470 A from Applied Biosystems. The phenylhydantoin amino acid derivatives were analyzed by an HPLC system which separates all components isocratically (27).

RESULTS

A partial cDNA clone (p281) for a human autoimmune antigen (10,12) was used to screen both a human genomic library and a lambda gt10 cDNA library from human teratoma cells.

Four overlapping clones (EJ,EK,24,137) were isolated from a genomic library made from human fetal liver DNA, and another overlapping clone (LH75) was isolated from a human leukocyte genomic library. The five genomic clones were determined to be overlapping by restriction (Fig. 1) and sequence analysis. These genomic clones are collectively referred to as clone 137, after the first isolate of this group. Partial restriction maps and sequencing strategies are shown for 137 and the two cDNA clones in Fig. 2.

Sequence data indicated that the 137 family of clones represent a second pseudogene for the E protein. The restriction map of 137 differs from the cDNA restriction map at several locations (Fig. 2). These differences were confirmed at the nucleotide level (Fig. 3). Clone 137 shows 90% similarity to the full length cDNA described below and 88% similarity to clone 63, a previously described pseudogene (10). Although clone 137 contains a consensus translation initiation sequence (position 43-49, Fig. 3) and a polyadenylation signal (position 443-449, Fig. 3) at appropriate locations, the open reading frame of clone 137 is interrupted by a two base deletion at position 169 which results in a shorter open reading frame than the cDNA. Further evidence that clone 137 represents a processed pseudogene comes from the finding that there is a long tract of As interrupted by several Gs 23 nucleotides downstream of the polyadenylation signal. Furthermore, a short direct repeat (9/10 matching) is found on each side (positions -20 to -11 and 610 to 618, Fig. 3) of the putative mRNA sequence. An Alu repetitive element is found directly upstream (position -258 to -26, Fig. 3) of the 5' direct repeat.

Interestingly, nine of the mutations found in clone 137 are also present in the previously characterized pseudogene (clone 63). However, clone 137 contains a number of mutations which differentiated this pseudogene from pseudogene 63 as well as the cDNA (Fig. 3).

[illegible]

	800	850
137	CATTGAGGTTGACCAAGATTTACCCACATATCAGGCCGATTTTTGAGGTTCCGACCAGGGAACTACCTAAACCAAGTGTGAGCCATTTCTATGTCC	
	900	950
137	TTTATATGCATCAAACTACTGTTCTGGAGCAGAGGCACCTTTAGATGCCTCTCTGGATGACCCACTGAAAGGCCTCTGCTACTCTTGTCTCTCCCTG	
	1000	1050
137	CCATAATTGCTGCCTCCTGCATGAGGAAGCAAAACAAACATAATGCATGTGGTCTGAAATGTGAGGAAGAGGAAAGTTCCTGGCTTCACAATGCTTGCT	
	1100	1150
137	CACATGAAACACAATCATGTACTTCATCATTCTTGGTGAAGAGTGTGTGGTCTGCCACGCCTTCGCCAAGTGAGTCAGTCAAGAGGAAATTCCT	
	1200	1250
137	GTGTATTGCATGGTGGTGTCTAGGCAGGATAACTGACCCACTGACCAcTGGTGAGGATAAGACATGACTTACTTCTCTCACTGTGTTgCACAgCTTgT	
	1300	1350
137	AAAGCgAACgTCCcAGAAGACAgAACTAGAAgTTCgGGCAGgTGGCCAAGgCTGGCAGGCACAGGCTGGAGCAGACCTAAAGTCCCCAGGCTGCACATT	
	1400	1450
137	CCTTTTTTGTCTATTCCACTCTCCCTTAGGGTAGGCTCCTGAAAAAAGAAAAAAGGTGCTGGCGTTAACTTTCTGGTTGTGTACTCTTCATCCAAC	
	1500	1550
137	AAATACCTGTGGAGCCCTGGCATGGAGAGGCGCTGCGCTGGTGGTGGGGTAGGAATGATGCAGCATGCCAATTCTTGTTCCTGGGAGCTCAGTCCAG	
	1600	1650
137	GGAGGAGAATATCGACGTgGGCAcTAATGACAgTGTGGTCTGATGACTCCATAGGCTcAgAAgTCTATCCCGGG	

Figure 3: Comparison of nucleotide and amino acid sequences of genomic clones 137 and 63 with the cDNA clone p11HB1. Identical bases and amino acids are indicated by a (-). Insertions/deletions are denoted by a (.). Direct repeats flanking the two pseudogenes are overlined. A consensus translation initiation site is underlined (43-49) and two polyadenylation signals are underlined (444-449 and 468-473). Numbering begins at the first nucleotide in the cDNA insert of p11HB1 and does not reflect insertions/deletions in the pseudogene sequences. 11aa, amino acid sequence deduced from the nucleotide sequence of p11HB1; 11HB1, nucleotide sequence of the cDNA insert of p11HB1; 63, nucleotide sequence of a previously described pseudogene; 137, nucleotide sequence of genomic clone 137; 137aa, amino acid sequence deduced from the nucleotide sequence of genomic clone 137.

Screening of a human teratoma cDNA library with p281 yielded one positive lambda cDNA clone. The insert from this clone was subcloned into pGem4 to generate p11HB1 (Fig. 2). This new cDNA clone contains a 492 nucleotide cDNA insert plus a poly(A) tract of 52 nucleotides (Fig. 3). Two lines of evidence suggest that p11HB1 is nearly full length. Extension of an end-labeled oligodeoxynucleotide (DRS-7) by reverse transcriptase yielded two predominant products of 41 and 45 nucleotides (Fig. 4). Since the 5' end of DRS-7 is complementary to position 31 of p11HB1, these extension products map 10 and 14 nucleotides 5' to the end of p11HB1 (Fig. 4). This localization of the cap site of E protein mRNA is in good agreement with the end point (ATT position -12, Figs. 3 and 4) defined by the extent of similarity between pseudogenes 63 and 137.

Figure 4: Cap site determination of E protein mRNA by primer extension analysis. End labeled oligodeoxynucleotide DRS-7 was annealed to HeLa cell poly(A)⁺ RNA. Reverse transcriptase and dNTPs were added and the final reaction products were loaded on a denaturing polyacrylamide gel. Lane 1, end labeled *Hpa*II cut pBR322 markers; lane 2, primer extension products. The primer extension end points (vertical arrows) are diagrammed below the data. The relevant sequences of the two pseudogenes, the cDNA, and the oligodeoxynucleotide are shown in the diagram. The horizontal arrow represents the extension reaction.

The sequence of p11HB1 is identical to that of the partial cDNA clone (p281) over the entire length of the insert in p281 (295 nucleotides). Clone p11HB1 contains 77 additional nucleotides at the 5' end and 119 additional nucleotides at the 3' end as compared to p281. The new cDNA clone contains a translation initiation consensus sequence (position 43-49, Fig. 3) and one consensus polyadenylation signal (position 444-449, Fig. 3). The poly(A) regions of the two pseudogenes begin 23 and 24 nucleotides downstream of the AATAAA signal. However, the poly(A) tail in p11HB1 begins 43 nucleotides downstream of the AATAAA signal. Interestingly, there is a variant poly-

Figure 5: Predicted Secondary Structure of E Protein. A secondary structure prediction (32) of the E protein amino acid sequence was plotted using a program from the University of Wisconsin Genetics Computer Group software package. The line represents the peptide backbone with turns corresponding to 180° changes in the direction of the line. The sine wave represents regions of predicted α -helix, the compressed saw tooth line represents β -sheet, and the expanded saw tooth represents random coil. Hydrophobic regions are indicated with a diamond and hydrophilic regions are indicated with an octagon. Both the hydrophobicity and hydrophilicity thresholds were set to 1.7 using the PLOTSTRUCTURE program of the UWGCG package. Potential N-linked glycosylation sites are indicated with a triangle.

adenylation signal (ATTAAA) located 19 nucleotides upstream of the poly(A) tail in p11HB1.

The longest open reading frame of p11HB1 begins at nucleotide position 46 and codes for a protein of 92 amino acids. The deduced polypeptide coded by this cDNA would have a molecular weight of 10800 daltons and an approximate pI of 9.5. The predicted amino acid sequence differs at one position from that predicted earlier (10). A histidine residue was predicted for amino acid position 4 based on the nucleotide sequence of pseudogene 63, whereas, an arginine residue is predicted for amino acid position 4 based on the sequence of p11HB1.

The predicted secondary structure of the E protein coded by p11HB1 is presented in Fig. 5. Hydrophobic and hydrophilic regions are indicated in Fig. 5 as are potential N-linked glycosylation sites.

Previous experiments have shown that the original cDNA clone (p281) coded for an in vitro translation product which co-migrated with authentic E protein and was

Cycle	Amino acids		
1	R	N	L
2	I	L	K
3	E	V	G
4	G	L	D
5	?	D	N
6	I	D	(D)
7	I	A	T
8	G	E	L
9	F	E	L
10	D	I	(Q)
11	E	H	(S)
12	Y	(S)	(V)
13		K	
14		T	
15		K	
16		?	
17		R	

predicted cyanogen bromide fragments from E protein :

Figure 6: Amino Acid Sequence of CNBr fragments of the snRNP E Protein. Three amino acids were determined for each cycle of the sequencing run as shown in the table. These amino acids can be arranged into three peptide sequences which correspond to three of the five cyanogen bromide fragments (boxed amino acids in the diagram) predicted from the cDNA sequence. Yields for most of the amino acids were sufficient for unambiguous identification. Amino acids for which yields were low have been placed in parentheses.

recognized by anti-Sm sera (12). To confirm that the open reading frame of p11HB1 does code for the E protein, purified E protein was subjected to amino acid analysis.

HeLa cell snRNPs were affinity purified on an anti-m₇G affinity column and E protein was isolated by preparative gel electrophoresis. Electroeluted E protein was further purified by HPLC. Direct sequence analysis on the purified E protein was unsuccessful, suggesting that the amino terminus was blocked. Therefore, the intact protein was subjected to cleavage with cyanogen bromide, and the resulting peptides were fractionated by HPLC. Material from one major peptide peak was then subjected to amino acid sequence analysis on a gas-phase sequencer. Surprisingly, three PTH amino acids were detected for each cycle as shown in Fig. 6. However, the three amino acid

residues from each cycle can be aligned to match perfectly the cyanogen bromide peptides predicted from the cDNA sequence (Fig. 6). Thus, the protein sequence predicted from the DNA sequence of p11HB1 is entirely consistent with the amino acid analysis of highly purified E protein isolated from HeLa cell snRNPs.

Previously, we reported the cloning of a cDNA for a human Sm antigen which co-migrated with the snRNP E protein (12). The results presented here provide solid confirmation of the identification of this autoimmune antigen as E, and provide a good basis for structure/function analysis of this protein. The primary translation product produced from E protein mRNA would have a molecular weight of 10800 which agrees well with our estimates from SDS gel electrophoresis. However, we cannot exclude the possibility that a few of the N-terminal amino acids are removed to generate the functional form of the E protein. We have not noted sequence similarities at either the nucleic acid or amino acid level with other nucleic acid binding proteins. The RNP consensus sequence found in several other snRNP proteins and nucleic acid binding proteins (28) is not present in the E protein. This may indicate that E does not directly interact with the snRNA in the RNP complex. If this is true, it would suggest that the E protein is not the Sm antigen which binds to the consensus Sm binding site A(U)_nG (n=3 to 6) in snRNAs.

10603

structural or functional relationship between a ribosomal protein and a spliceosomal protein adds further interest to these comparisons. Unfortunately, the role of the var1 protein in the yeast mitochondrial ribosome is not clear, so the observed sequence similarity does not provide any useful clues as to the function of the E protein in the RNA processing machinery. The isolation of the full-length cDNA for the E protein should facilitate further studies of the role of this protein in snRNP structure and function.

We have not yet begun to investigate the regulation of E protein expression in mammalian cells, but the sequence data reveal a potential diversity in the structure of E protein mRNAs from different sources. Current theory dictates that the structure of the processed pseudogenes should reflect the structure of mRNAs synthesized in the germ line tissues where they originated (30,31). Thus, the different locations of the poly(A) tails in the pseudogenes and the cDNA may reflect differential usage of alternative poly(A) signals between germ line cells and the teratoma cell line which gave rise to p11HB1. Since the two mRNA species would differ only by 20 nucleotides of 3' noncoding sequence, it is not clear what the functional significance of differential polyadenylation would be in this case.

The primer extension data suggest that there is also some length heterogeneity at the 5' end of the E protein mRNA. At least two major extension products are obtained from a single oligonucleotide primer. Further characterization of the expression of the E protein gene in different cells may reveal a possible function for this micro-heterogeneity in the structure of E protein mRNA.

ACKNOWLEDGEMENTS

We appreciate the generous gifts of the human genomic library from Dr. T. Maniatis and the human teratoma cDNA library from Dr. J. Skowronski. We would also like to thank J. Greene for assistance with primer extensions.

This work was supported by NIH Grants AR38554 and CA09441 to E.D.W. and D.R.S., a grant from the Deutsche Forschungsgemeinschaft to R.L. (Lu 294/2-2) and in part by the Mayo Foundation (E.D.W.).

*To whom correspondence should be addressed

REFERENCES

1. Maniatis, T., and Reed, R. (1987) *Nature* **325**, 673-678.
2. Sharp, P.A. (1987) *Science* **235**, 766-771.
3. Mount, S.M., Pettersson, I., Hinterberger, M., Karmas, A., and Steitz, J.A. (1983) *Cell* **33**, 509-518.
4. Lerner, M.R., and Steitz, J.A. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5495-5499.

5. Tan, E.M. (1982) *Adv. Immunol.* **33**, 167-240.
6. Query, C.C., and Keene, J.D. (1987) *Cell* **51**, 211-220.
7. Habets, W.J., Sillekens, P.T.G., Hoet, M.H., Schalken, J.A., Roebroek, A.J.M., Leunissen, J.A.M., van de Ven, W.J.M., and van Venrooij, W.J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 2421-2425.
8. Sillekens, P.T.G., Habets, W.J., Beijer, R.P., and van Venrooij, W.J. (1987) *EMBO J.* **6**, 3841-3848.
9. Spritz, R.A., Strunk, K., Surowy, C.S., Hoch, S.O., Barton, D.E., and Francke, U. (1987) *Nucleic Acids Res.* **15**, 10373-10391.
10. Stanford, D.R., Rohleder, A., Neiswanger, K., and Wieben, E.D. (1987) *J. Biol. Chem.* **262**, 9931-9934.
11. Theissen, H., Etzerodt, M., Reuter, R., Schneider, C., Lottspeich, F., Argos, P., Luhrmann, R., and Philipson, L. (1986) *EMBO J.* **5**, 209-3217.
12. Wieben, E.D., Rohleder, A.M., Nenninger, J.M., and Pederson, T. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7914-7918.
13. Yamamoto, K., Miura, H., Moroi, Y., Yoshinoya, S., Goto, M., Nishoika, K., and Miyamoto, T. (1988) *J. Immunol.* **140**, 311-317.
14. Bringmann, P., and Luhrmann, R. (1986) *EMBO J.* **5**, 3509-3516.
15. Reuter, R., and Luhrmann, R. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8689-8693.
16. Reuter, R., Rothe, S., and Luhrmann, R. (1987) *Nucleic Acids Res.* **15**, 4021-4034.
17. Fisher, D.E., Conner, G.E., Reeves, W.H., Wisniewolski, R., and Blobel, G. (1985) *Cell* **42**, 751-758.
18. Bringmann, P., Rinke, J., Appel, B., Reuter, R., and Luhrmann, R. (1983) *EMBO J.* **2**, 1129-1135.
19. Pettersson, I., Hinterberger, M., Mimori, T., Gottlieb, E., and Steitz, J.A. (1984) *J. Biol. Chem.* **259**, 5907-5914.
20. Skowronski, J., Fanning, T.G., and Singer, M.F. (1988) *Mol. Cell. Biol.* **8**, 1385-1397.
21. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
22. Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G., and Maniatis, T. (1978) *Cell* **15**, 1157-1174.
23. Southern, E. (1975) *J. Mol. Biol.* **98**, 503-517.
24. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
25. Lipman, D.J., and Pearson, W.R. (1985) *Science* **227**, 1435-1441.
26. Simpson, R.J., Moritz, R.L., Nice, E.E., and Grego, B. (1987) *Eur. J. Biochem.* **165**, 21-29.
27. Lottspeich, F. (1985) *J. Chromatogr.* **326**, 321-327.
28. Swanson, M.S., Nakagawa, T.Y., LeVan, K., and Dreyfuss, G. (1987) *Mol. Cell. Biol.* **7**, 1731-1739.
29. Steitz, J.A., (1987) In Inouye, M. and Dudock, B.S. (eds), *Molecular Biology of RNA: New Perspectives*, Academic Press, New York, pp. 69-80.
30. Lee, M.G.S., Lewis, S.A., Wilde, C.D., and Cowan, N.J. (1983) *Cell* **33**, 477-487.
31. Ponte, P., Gunning, P., Blau, H., and Kedes, L. (1983) *Mol. Cell. Biol.* **3**, 1783-1791.
32. Chou, P.Y., and Fasman, G.D. (1978) *Adv. Enzymol.* **47**, 45-147.
33. Hudspeth, M.E.S., Ainley, W.M., Shamard, D.S., Butow, R.A., and Grossman, L.I. (1982) *Cell* **30**, 617-626.