Volume 11 Number 10 1983 Nucleic Acids Research

Characterization of a genomic clone for rat seminal vesicle secretory protein IV

Jagannadha C.Kandala*, Malathi K.Kistler*, Robert P.Lawther+ and W.Stephen Kistler*

Departments of *Chemistry, +Biology, and *Biochemistry Section of the School of Medicine, University of South Carolina, Columbia, SC 29208, USA

Received 17 February 1983; Revised and Accepted 19 April 1983

ABSTRACT

The entire coding region for rat seminal vesicle secretory protein IV was obtained on a 3.5 kb Eco RI fragment isolated from a genomic library in λ Charon 4A. The coding sequence for SVS IV message is interrupted twice by introns. The first lies just downstream from the juncture of the 21 amino acid secretory signal peptide with the start of the mature protein, and the second lies in the 3'-nontranslated region. The major transcriptional start site was mapped by primer extention and is 22 nucleotides upstream from the translational initiation codon. S1 protection experiments indicated additional minor transcriptional starts about 27 and 50 nucleotides further upstream from the major cap site. The entire transcriptional unit comprises about 1740 nucleotides. The SVS IV gene does not belong to an obvious gene family, and it is conserved in mice and guinea pigs.

INTRODUCTION

The rat seminal vesicle synthesizes a small family of major secretory proteins, and these products are of interest as markers for study of the hormonal regulation of protein synthesis in an androgen target organ. The best characterized seminal vesicle secretory protein is SVS IV (1), a small, 90-residue polypeptide of established sequence (2) but of unknown function. Large increases occur in the relative production of SVS IV during normal sexual development (3,4) while corresponding decreases in SVS IV production follow castration of adult animals (5,6). Differential changes in the mRNA pool size seem to be the major factor accounting for changes in SVS IV production in adult animals (6,7,60). In immature rats translational level effects appear to play a significant role as well (3). Despite these studies, the detailed mechanisms regulating SVS IV production remain unknown.

As part of our program to unravel the factors that control SVS IV synthesis, we previously constructed a cDNA clone for SVS IV message (8). We have now used that clone to select the SVS IV gene from a rat genomic library. Analysis of the cloned gene has shown that it contains two intervening sequences but that virtually the entire sequence of the secreted form of the

protein is coded for within a single exon. The major cap site for SVS IV message was mapped on the gene by two different procedures. In addition to the major cap site, two additional upstream cap sites make minor contributions to the SVS IV message pool.

EXPERIMENTAL PROCEDURES

Animals

Sprague-Dawley rats were obtained from Harlan Industries or from Charles River. Inbred female C3HeB/Fe mice were from Jackson Laboratories. Hartley guinea pigs were from Charles River.

Restriction Digests

Restriction enzymes were obtained from New England Biolabs or Bethesda Research Laboratories and used acccording to the protocols supplied. Unless indicated otherwise, restriction digests were analyzed in gels buffered by 89 mM Tris, 89 mM boric acid, and 2 mM Na_2EDTA.

Probe Preparation

Restriction fragments were separated electrophoretically in polyacrylamide gels, stained with ethidium bromide, and recovered by electroelution in dialysis bags (9). Radioactive labels were incorporated by nick translation (10) using $\alpha[^{32}P]$deoxynucleotides from New England Nuclear and DNA polymerase I from Boehringer. Specific activities from 0.5 to 2 X 10^8 CPM/µg were obtained routinely.

DNA Isolation from Tissues and Genomic Blots

DNA was isolated by the procedure of Sala-Trepat et al. (11), but the starting material was a 1:10 homogenate of whole tissue (rather than isolated nuclei) and the final preparation was not sheared. Livers were obtained from animals fasted for 24 hr to decrease glycogen stores. Restriction digests of genomic DNA were monitored for completeness of digestion by addition of λ DNA to a sample removed from the main digest (12). DNA was concentrated by ethanol precipitation, separated electrophoretically, and then transferred to nitrocullulose sheets (Schleicher & Schuell, BA 85) by the method of Southern (13). Pretreatment of filters, hybridization, and post hybridization washings were as described by Maniatis et al. (14) for genomic library screens, except that poly(A) was not used in any step and the final two post hybridization washes were in 0.3X SET, 0.1% SDS, 0.1% sodium pyrophosphate at 68°C for 15 min. each. Filters were exposed to Kodak X-Omat film in the presence of Dupont Cronex Lightning Plus intensifying screens at -70°C.

Preparation of Plasmid and Viral DNA Molecules

Plasmid pSVM401 is a derivative of pBR322 with a nearly full length SVS

IV cDNA insert at the Pst I site (8). Plasmid pSVG40 contains a 3.5 kb rat genomic fragment encompassing the SVS IV coding region at the Eco RI site of pBR325 (15). Plasmids were maintained in E. coli RR1 (16) and were isolated from cleared lysates (17) by polyethylene glycol precipitation (18) and two cycles of CsCl isopycnic centrifugation in the presence of propidium iodide (19). Lambda Charon 4A-JK1 (λ JK1) was selected from a rat genomic library (20). The virus was propagated in E. coli DP50supF (21), and DNA was prepared by an unpublished procedure supplied by Dr. William Marzluff that involved digestion of liquid culture lysates by RNase and DNase, then digestion with proteinase K in the presence of SDS, and finally recovery of viral DNA by phenol extraction and ethanol precipitation. Studies with recombinant DNA were carried out in compliance with the National Institutes of Health guidelines.

Genomic Library Screens

The library was kindly provided by Dr. Tom Sargent and Professor James Bonner and consisted of a partial Eco R1 digest of Sprague-Dawley rat liver DNA inserted into the arms of lambda Charon 4A (20). The library was screened using the general procedure of Benton and Davis (22) at a density of 10^4 plaques/15 cm petri plate using the nick translated cDNA insert of pSVM401 as probe.

Nucleic Acid Sequencing

The 910 bp 5' Eco RI-Sst I fragment of pSVG40 was isolated electrophoretically and labeled at each end by successive use of calf intestinal alkaline phosphatase [Boehringer, purified as described (22)] and polynucleotide kinase (P-L) with $\gamma[^{32}P]$ ATP (ICN) as label source. Following cleavage with Hpa II, the 5'-Eco RI-Hpa II fragment was recovered by preparative electrophoresis and used for Maxam-Gilbert sequencing (24,25) employing 0.4 mm thick gels for analysis of cleavage products.

Isolation of RNA

Total RNA was extracted from seminal vesicles of adult rats by homogenization in the presence of SDS, proteinase K digestion, and phenol extraction (26). Poly(A) RNA was selected by two cycles of oligo(dT) chromatography (27).

S1 Nuclease Mapping Using Unlabeled DNA

S1 nuclease mapping as developed by Berk and Sharp (28) was employed as described by Favaloro et al. (29). Briefly, 200 ng of Eco RI cut plasmid DNA and adult rat seminal vesicle RNA (3.3 μg poly(A) RNA or 4.5 μm total RNA) or 3.3 μg yeast RNA (as blank) were dissolved in 20 μl of 40 mM piperazine-N,N'-bis[2-ethanesulfonic acid] (PIPES), pH 6.4, 400 mM NaCl, 1 mM EDTA, 80% (v/v)

formamide (Fluka, deionized by treatment with a mixed bed ion exchange resin). The nucleic acids were denatured at 85°C for 10 min and then hybridized at 52°C for 16 hr. S1 nuclease buffer (0.3 ml) containing 40 µg of denatured calf thymus DNA was added to each tube along with 5 units of S1 (P-L) and digestion allowed to take place at 37°C for 1 hr. Digests were stopped by addition of 10 µg yeast RNA, sodium acetate to a final concentration of 0.2 M and Na_2EDTA to 10 mM. Nucleic acids were precipitated by addition of two volumes of ethanol, and the precipitate dissolved in 25 µg of 50 mM NaOH, 1 mM EDTA, 2% (w/v) Ficoll, 0.015% (w/v) bromocresol green. Samples were then electrophoresed in a 10.5 cm, 2% agarose gel run with an electrolyte of 30 mM NaOH, 1 mM EDTA (9) at 40 V for 4 hr. Gels were immersed in 3 M sodium acetate, pH 5.5 for 15 min, with agitation, and DNA bands were transferred to nitrocellulose as described above.

S1 Mapping with 5'-End Labeled DNA

S1 mapping as modified by Weaver and Weissman (30) was used to locate the cap site(s) on the SVS IV gene. The 5'-Eco RI-SSt I fragment of the pSVG40 insert was isolated by preparative electrophoresis and then digested incompletely with Dde I. The partial digest was labeled at 5' ends as described above and separated electrophoretically. The 5'-238 bp Eco RI_{-145}-Dde I_{93} as well as the 102 bp Dde I_{-9}-Dde I_{93} fragment were recovered by the elution procedure of Maxam and Gilbert (25). In view of the cutting specificity of Dde I (31) each fragment contains a labeled 5'-end at position +96 on the strand coding for SVS IV message. Each labeled fragment (30-40 ng) was mixed separately with 3 µg of seminal vesicle poly(A) RNA and precipitated with ethanol. The mixtures were then hybridized at 49°C and digested with S1 nuclease as described above. Ethanol precipitated S1 resistant fragments were taken up in 80% deionized formamide, 10 mM NaOH, 1 mM EDTA and analyzed on a 7.5% acrylamide 8 M urea gel (43 cm long X 0.04 cm thick) (25).

Mapping the Cap Site(s) by Primer Extention

Primer extention studies were carried out essentially as described by Ghosh et al. (32). The cDNA insert of pSVM401 was excised by Pst I digestion and isolated by preparative electrophoresis. Dde I fragment 93-237 (Fig. 3) was subsequently isolated and labeled at 5'-ends as described earlier. Treatment of this labeled fragment with Alu I allowed isolation of a 3'-fragment that could be electrophoretically resolved into individual strands by virture of the fact that the Dde I end had a 3-nucleotide 5'-overhang while the Alu I end was flush ended (31). Thus use of a 7.5% acrylamide, 8 M urea gel allowed isolation of a 102 residue single strand that was complementary to the message

and labeled at the 5'-end corresponding to residue +240 of SVS IV message. The probe (60 ng) was annealed to 15 µg of seminal vesicle poly(A) RNA under conditions identical to those used for S1 mapping with end-labeled probes, except that 60% formamide was used. Hybridized samples were ethanol precipitated and dissolved in 50 µl of 50 mM Tris-HCl, pH 8.2, 10 mM $MgCl_2$, 50 mM KCl, 10 mM dithiothreitol, and 1 mM each of the four dNTPs. AMV reverse transcriptase (70 units, provided by Life Sciences, St. Petersburg, Florida under the auspices of the National Cancer Institute, NIH) was added, and incubation was carried out at 41°C for 3 hr. The incubation was then adjusted to 0.2 M NaOH, and RNA was destroyed by further incubation at 41°C for 1 hr. The sample was neutralized with HCl, adjusted to 0.5% (w/v) SDS and extracted with phenol. After ethanol precipitation of the DNA, samples were run on 7.5% acrylamide, 8 M urea gels.

RESULTS

Isolation of Genomic Clone for SVS IV

A nearly full length cDNA clone for SVS IV was constructed earlier (8) and served as a hybridization probe to screen for SVS IV genomic sequences. A preliminary study of a total Eco RI digest of rat liver DNA by the Southern technique (13) indicated that, depending on the DNA preparation, one or two fragments with sizes between 3kb and 4kb hybridized strongly to the probe (results not shown). Accordingly, we screened a library prepared by partial RI digestion of rat liver DNA by Sargent et al. (20). Some 200,000 plaques were analyzed, and a single positive signal was obtained. Purification yielded the viral clone λ JK1. Southern blot analysis showed that a 3.5 kb Eco R1 fragment from λ JK1 hybridized intensely to the SVS IV cDNA probe, and subsequent analysis has shown that this 3.5 kb fragment contains the entire DNA sequence necessary to code for the SVS IV message. The 3.5 kb fragment was therefore subcloned into the Eco R1 site of pBR325 (to generate pSVG40), and an extensive restriction map was generated by analysis of the results of single and double digests (Fig. 1).

Since preliminary experiments indicated some uncertainty about the number of genomic Eco RI fragments containing SVS IV sequences, we wished to establish unequivocally which genomic segment we had cloned. For that purpose Southern blot analysis was made of several rat DNA preparations, and for species comparison, of guinea pig and mouse DNA as well. While only a single 3.5 kb fragment was identified from one sample of rat liver DNA (Fig. 2), a 2nd sample of liver DNA as well as the seminal vesicle DNA gave rise to two R1

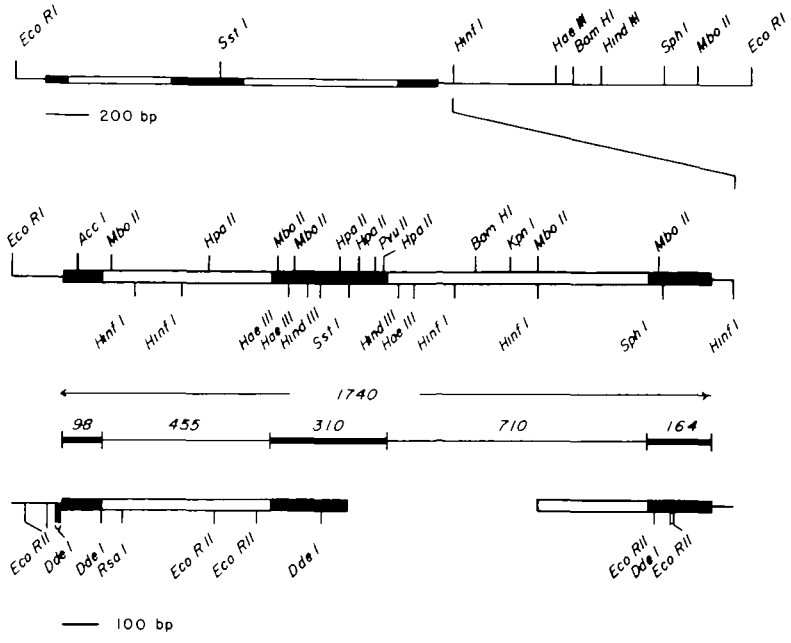

Figure 1. Restriction map of the 3.5 kb rat genomic insert of pSVG40. The location of the SVS IV gene defined by the boundaries of the major cap site and polyadenylation site are indicated by the boxed region with the 5'-end of the message to the left. Exons are shown as filled regions. Restriction sites for frequent cutting enzymes are shown for certain portions of the gene in the lowermost map. All restriction sites were experimentally defined. Distances are in base pairs. The overall length of the insert was determined with reference to the 3.6 kb Pst I-Eco RI and the 3.2 kb Pst I-Bam HI fragments of pBR322 (56). The insert is oriented in the vector so that the left end of the sequence is nearest to the single Hind III site of pBR325 (15,57).

fragments of approximately 3.5 and 3.8 kb. These results were repeated at least three times for each DNA sample under conditions in which test DNA samples were digested completely. Mixing experiments indicated that the RI fragment of the SVS IV subclone migrated indistinguishably from the shorter (3.5 kb) genomic fragment (Fig. 2). Dickson and Harris (44) have also observed these two Rl fragments, and they have concluded that they are Mendelian alleles based on their behavior during genetic crosses.

From the species comparisons it appears that the SVS IV gene is conserved among rats, mice, and guinea pigs, and that the 3.5 Eco Rl fragment is perhaps most typical of the gene in rodents as a class (Fig. 2).

The SVS IV Coding Region is Interrupted by Two Introns

Intervening sequences are generally though not invariably present in eu-

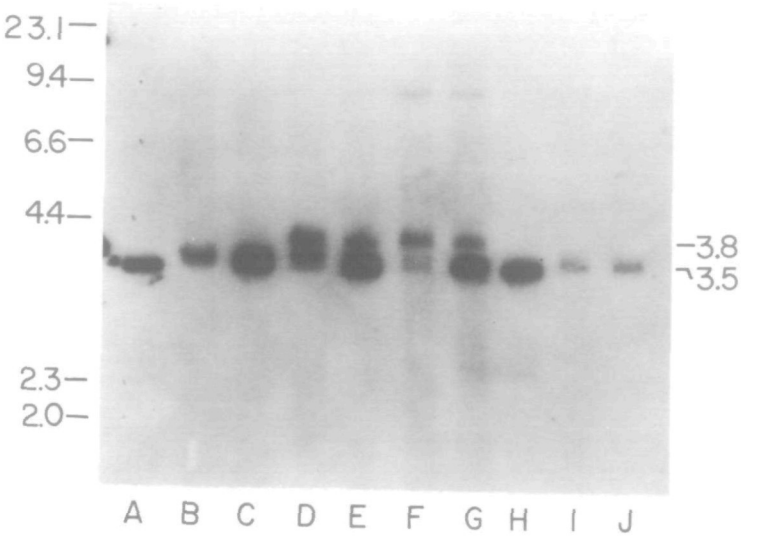

Figure 2. Southern blot analysis of genomic DNA digested with Eco RI. Samples of digested DNA (20 μg) with or without added Eco RI cleaved pSVG40 DNA (100 pg) were separated electrophoretically in an 0.8% agarose gel (0.5 cm thick X 18 cm long; 50 V X 16 hr). The gel was acid treated to enhance transfer of larger DNA fragments (58), and after Southern blotting the immobilized DNA was hybridized to the nick translated 5'-Eco RI-Sst I fragment of pSVG40 (Fig. 1). Exactly the same results were obtained when the cDNA insert of pSVM401 was used as probe (results not shown). DNA sources were Lane: A, pSVG40 alone; Lane B, rat liver from animal I; Lane C, same as B plus PSVG40; Lane D, rat liver from animal II; Lane E, same as D plus pSVG40; Lane F, rat seminal vesicle (pooled tissues); Lane G, same as F plus pSVG40; Lane H, pSVG40 alone; Lane I, guinea pig kidney from a single animal; Lane J, mouse liver (pooled tissue). Hind III fragments of lambda were run as size markers.

karyotic genes for proteins (33). Analysis of the SVS IV gene for intervening sequences has been greatly facilitated by a nearly complete sequence for cloned SVS IV cDNA reported by Mannson et al. (7). The cDNA contains a distinctive set of sites for the restriction enzymes Hind III, Sst I, Pvu II, and Sph I (See Fig. 3). The first three of these sites were readily located about ⅓ of the way in from the left end of the 3.5 kb genomic fragment we had isolated. Search for the single Sph I site in the 3'-nontranslated portion of SVS IV message indicated that it was at least 700 bp further away from the Pvu II site than predicted by the message sequence. Similarly, the 5' Dde I site at residue 93 of the message sequence was at least 450 bp further in the 5' direction from the Hind III site than predicted by the message. These dis-

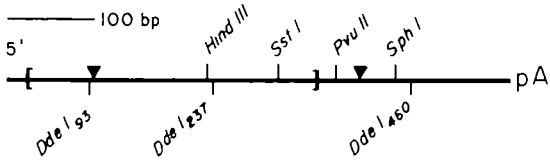

Figure 3. Map of SVS IV messenger RNA. The exact cutting sites for Dde I were assigned by joining residue 1 of the cDNA sequence of Mansson et al. (7) to residue +31 of the gene sequence of Fig. 5. The translated region is enclosed in brackets, and splice sites are indicated by triangles.

crepancies between the two maps indicated the likely presence of two introns, which have been confirmed as described below. More detailed restriction mapping showed that the message region coding for nearly the entire mature (secreted) form of SVS IV appeared in a continuous stretch of about 310 bp. However, the 5'-nontranslated and signal peptide sequences as well as most of the 3'-nontranslated portion of the message were separated from the central exon by introns of approximately 455 and 710 bp respectively (Fig. 1).

Confirmation of this structural map was made by the S1 nuclease mapping technique originated by Berk and Sharp (28) and modified by Favaloro et al. (29). Polyadenylated RNA isolated from rat seminal vesicle was hybridized to pSVG40 DNA under conditions favoring DNA-RNA hybrids, and unhybridized nucleic acids were digested with S1 nuclease. Resistant DNA fragments, which should correspond to exons in spliced RNA, or possibly to full length transcripts protected by trace amounts of nuclear mRNA precursors, were identified by alkaline electrophoresis, transfer to nitocellulose, and hybridization to either 5'-specific (Eco RI-Sst I) or 3'-specific (Sst I-Eco RI) segments of the pSVG40 insert (Fig. 4).

Both probes identified a 310 bp fragment corresponding to the large central exon. The expected small 5'-exon was detected by the 5'-probe as a roughly 100 bp fragment only by long exposure of autoradiograms (Fig. 4B), probably due to the relative instability of hybrids involving this fragment in high formamide and also to the inefficient transfer of small fragments to nitrocellulose. The 3'-probe specifically identified a roughly 170 bp fragment that corresponds to the expected 3'-exon (Fig. 4C). Both probes identified trace amounts of an approximately 1750 bp fragment that is the length expected for a continuous transcript covering all three exons and the two introns. An additional band of roughly 1300 kb corresponds to the expected product resulting from the removal of intron I from the 1750 bp fragment. This assignment is supported by the fact that the 3'-specific probe hybridized

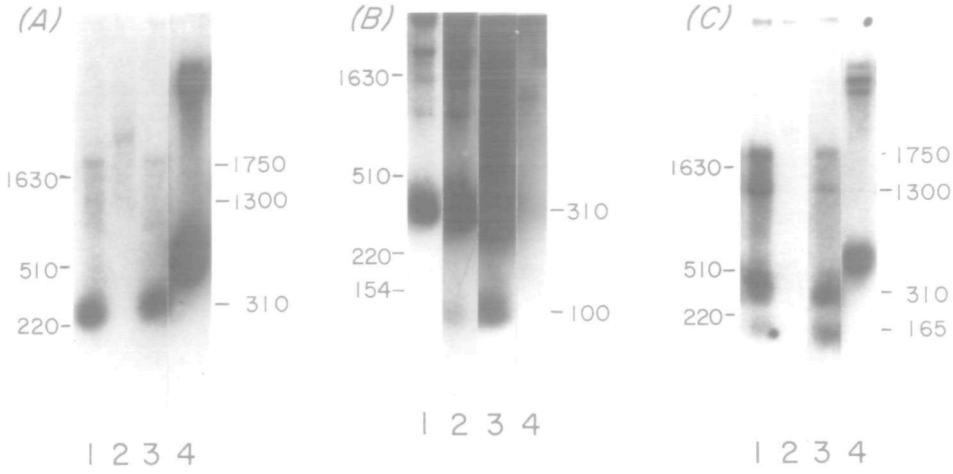

Figure 4. Segments of the SVS IV gene protected from S1 nuclease digestion by hybridization to seminal vesicle RNA. As described under "Experimental Procedures", pSVG40 DNA was hybridized to total seminal vesicle RNA or to seminal vesicle poly(A) RNA in the presence of high formamide to favor selective formation of RNA/DNA hybrids. S1 nuclease was used to digest unhybridized portions of the nucleic acids, and resulting DNA fragments were separated electrophoretically in alkaline gels and transferred to nitrocellulose. The blots were probed with either the 5'-Eco RI-Sst I (A and B) or the 3'-Sst I-Eco RI fragment (C) of pSVG40. Gel (A). 5'-Probe. Lane 1, poly(A) RNA; Lane 2, yeast RNA blank, Lane 3, total RNA; Lane 4, pSVM401 DNA hybridized to poly(A) RNA (a positive control that should leave the bulk of the 550 bp cDNA insert protected from S1). Gel (B). 5'-Probe. Lane 1, poly(A) RNA, autoradiogram exposed 1 hr.; Lane 2, same as Lane 1 but exposed 6 hr.; Lane 3, same as Lane 1 but exposed 16 hr.; Lane 4, yeast RNA blank exposed for 16 hr. Gel (C). 3'-Probe. Lanes contain the same samples as in Gel (A). Size markers indicated to the left of each set of gel tracks are a Hinf I digest of pBR322.

to this fragment much more intensely than did the 5'-specific probe. These results thus support a structural gene separated into three coding elements by two intervening sequences as predicted by the restriction map. The detection of large DNA fragments that seem to result from protection by nuclear RNA precursors means that precursors should be identifiable by "Northern" blots of nuclear polyadenylated RNA. In fact RNA samples run on formaldehyde gels (34) and transfered to nitrocellulase as described by Thomas (35) gave rise to faint bands of the expected precursor lengths detected after probing with the Eco RI-Sst I fragment of pSVG40 (results not shown). Such presumptive precursor forms showed up much more strongly following the S1 mapping procedure, however, perhaps because RNA was in excess and the long precursor RNA/DNA hybrids were significantly favored due to greater stability under the high

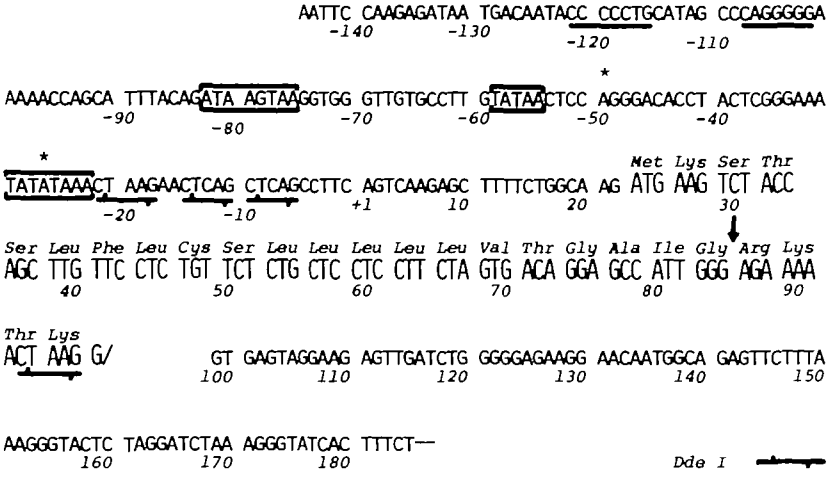

Figure 5. Nucleotide sequence in the vicinity of the cap site(s) for SVS IV. The sequence shown is for the strand with polarity the same as the message, and the first transcribed nucleotide for the major cap site is designated +1. Other cap sites are indicated by asterisks. The sequence between nucleotides +32 and +98 overlaps and confirms the SVS IV cDNA sequence reported by Mannson et al. (7). Three regions with varying degrees of homology to the TATA box component of the presumed RNA polymerase II promoter are boxed. The region -102 to -122 with a 7 bp inverted repeat possibly capable of a stem loop structure is underlined. The presumed boundary between the first exon and intron [assuming the AG/GT consensus sequence for splice sites (33)] is indicated by a slash. The cutting sites for Dde I are indicated as these have particular relevance to mapping experiments discussed in the text. A vertical arrow indicates the start of the mature protein (36).

formamide hybridization conditions used.

Sequence of the 5' Region of the SVS IV Gene

Many of the structural features that relate to transcriptional initiation are thought to be clustered in the vicinity of the 5'-terminus of the message (cap site). Accordingly the sequence of the 330 residues extending from the 5'-Eco RI site of the pSVG40 insert were established by the Maxam-Gilbert technique (Fig. 5). Inspection of this sequence revealed the entire first exon of the SVS IV gene comprising 21 residues of the signal peptide known to be present (1,36) as well as 4 residues of the mature protein. Following the codon for residue 4 is a typical exon/intron junction sequence (33) and a series of nucleotides that have no coding relationship to SVS IV, confirming the presence of an intron.

Identification of the Cap Site by Primer Extention

A number of techniques have been developed to determine the cap site of a

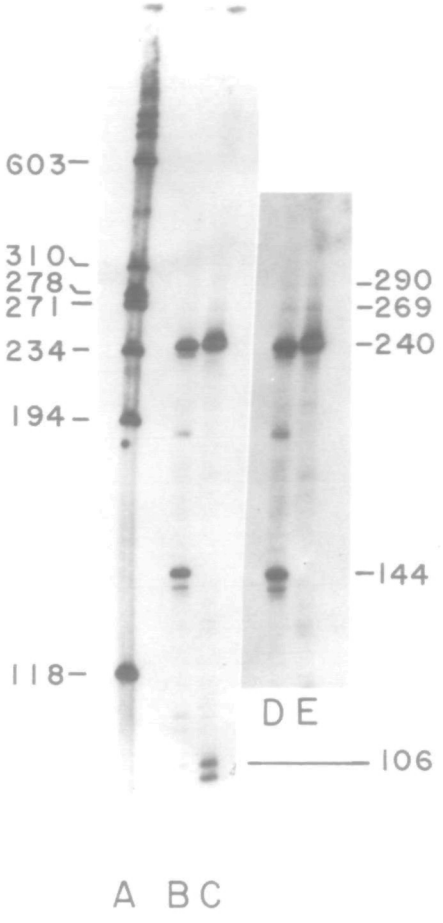

Figure 6. Mapping the cap site for SVS IV by primer extention. The 144 bp Dde I_{93}-Dde I_{237} fragment (see Fig. 3) of cDNA clone pSVM401 was isolated and 5'-end labeled. A portion of this material was converted to a 106 nucleotide single strand labeled specifically at position 240 on the strand coding for the message as described under "Experimental Procedures". Both the single strand and the original double stranded primers were hybridized to seminal vesicle poly(A) RNA, and then elongated with reverse transcriptase in the presence of all four unlabeled deoxynucleotide triphosphates. The elongated primers were freed of RNA by alkalai digestion and electrophoresed in a denaturing polyacrylamide gel. Lane A, size markers from a 5'-end labeled Hae III digest of ϕX174RF; Lane B, Product resulting from elongation of the double stranded primer; Lane C, Product resulting from elongation of the single stranded primer; Lanes D & E, Longer exposure of lanes B & C (only the portion encompassing the extended products). Locations of the single stranded primer (106 nucleotides), the double stranded primer (144 nucleotides) and the extended products are indicated to the right. In this experiment both double and single stranded primers were equally effective.

eukaryotic gene, and most evidence indicates that the cap site is normally the transcriptional initiation point (33). Based on the available restriction sites, the most useful technique for us was that of primer extention as used by Ghosh et al. (34). A single stranded primer complementary to the message and 5'-labeled uniquely at position 240 of the message was constructed as described under "Experimental Procedures".

This primer was hybridized to total seminal vesicle polyadenylated RNA and elongated by AMV reverse transcriptase. The product was then denatured and electrophoresed on a sequencing type polyacrylamide gel. After autoradiography, the primer was found to be elongated to a single length (240 bp) presumed to represent the capped end of the message (Fig. 6). The results were not dependent upon the exact DNA/RNA ratio, as reduction of the RNA concentration by a factor of 6 (to ensure DNA excess) did not change the results (results not shown). The residue identified as the capped nucleotide is an adenine preceded by pyrimidines and thus fits the pattern of apparent initiation sites observed in other RNA polymerase II dependent genes (33).

Many genes transcribed by RNA polymerase II are characterized by an AT rich sequence (so called TATA or Hogness box) with its 5'-end about 30 residues upstream from the cap site (33). This sequence appears to function as part of a polymerase II promoter. The SVS IV gene contains a typical TATA sequence in the expected position just upstream from the cap site identified (Fig. 5).

Evidence for Two Additional Minor Upstream Cap Sites

The primer extention study described above located a single cap site present in the majority of accumulated SVS IV transcripts. However, trace bands corresponding to longer extention products were visible with prolonged exposure of the autoradiogram [extended primers of approximately 269 and 290 nucleotides (Fig. 6)]. These longer products would correspond to capped transcripts presumably initiated at nucleotides -27 and -50. Upstream about 30 base pairs from each of these sites is a sequence with reasonable homology to the TATA element thought to direct precise iniation points.

Because of the apparent association of additional upstream cap sites with their own TATA sequences, further evidence for the use of these redundant start sites was sought by use of S1 mapping with specific 5'-end labeled restriction fragments (30). In this modification of the S1 technique, a restriction fragment was chosen that starts within the message sequence but which overhangs its 5'-end. Following hybridization with a message population and digestion with S1 nuclease, a specific length of protected DNA should be

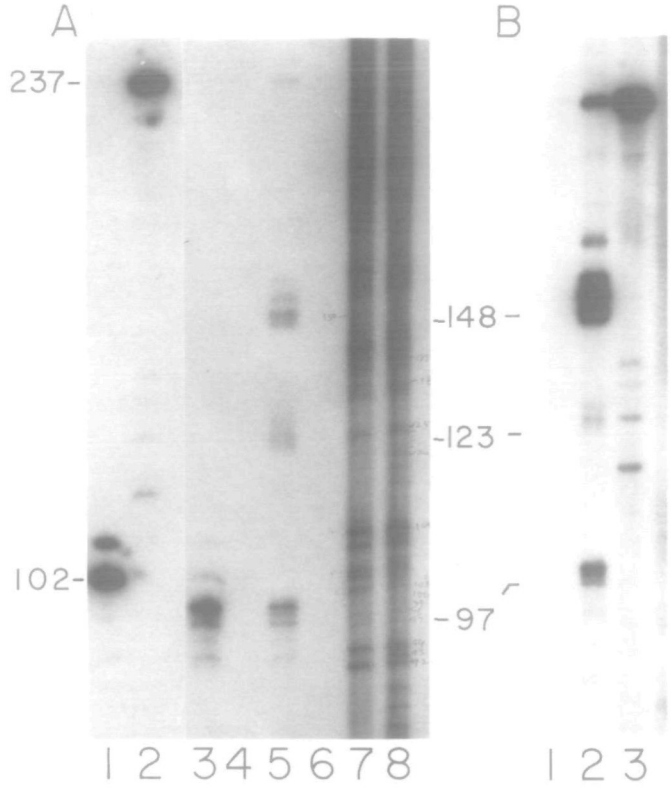

Figure 7. Mapping the cap site(s) for SVS IV by S1 nuclease digestion. Two restriction fragments expected to overlap the cap site were isolated from the genomic clone pSVG40: fragment A comprised nucleotides from Dde I_{-9}-Dde I_{+93}; fragment B comprised nucleotides from Eco RI_{-145}-Dde I_{+93}. Each thus has a 5' terminus on the strand complementary to the message at position +96 (Fig. 5), and the two probes vary only with respect to their lengths. The 5'-labeled fragments were then hybridized to seminal vesicle poly(A) RNA, digested with S1, and resistant products were analyzed on a sequencing type gel. Gel (A) Lane 1, Fragment A intact; Lane 2, Fragment B intact; Lane 3, Fragment A hybridized to seminal vesicle poly(A) RNA and S1 treated; Lane 4, Fragment A hybridized to yeast RNA and S1 treated; Lane 5, Fragment B hybridized to seminal vesicle poly(A) RNA and S1 treated; Lane 6, Fragment B hybridized to yeast RNA and S1 treated; Lanes 7 & 8, Size markers from the pyrimidine cleavage reactions from Maxam-Gilbert chemical sequencing of the lefthand end of the Eco RI-Pvu II fragment of the SVS IV gene. Sizes indicated are those estimated for the shortest protected fragment in each set of S1 products seen in the experimental lanes. Gel (B). Lane 1, Fragment B hybridized to yeast RNA and S1 treated; Lane 2, Fragment B hybridized to seminal vesicle poly(A) RNA and S1 treated; Lane 3, Fragment B run intact. Lanes displaying intact probes contained only 1/10 the material present in the S1-digested samples. The experiments shown in (A) and (B) were carried out under supposedly identical conditions but gave quite different pictures of the relative distribution of protected fragments among the 3 size classes observed.

preserved that allows estimation of the distance from the labeled end to the cap site. Two such fragments were employed. One was the Eco RI_{-145} to Dde I_{+93} fragment (resulting from a partial Dde I digest of the pSVG40 insert) and the other was the fragment Dde I_{-10} to Dde I_{93} (see Fig. 5). When such hybridizations are done under DNA excess conditions, the frequency of protected DNA fragments resulting from S1 treatment should reflect the frequency of the various cap sites. However, when RNA excess occurs, and hybridization is done under the high formamide conditions that favor DNA/RNA hybrids, long hybrids may be significantly more stable than short. Accordingly one might expect to see an over representation of any putative transcripts resulting from minor upstream redundant promoters. In fact, exactly such results were obtained in two different experiments. In the first a set of clusters of protected fragments was found corresponding to all three potential start sites (Fig. 7A). [Capped messages seem to protect several overhanging nucleotides on the labeled probe (30,37), but it is also true that careful analysis of a number of messages has revealed microheterogenity at the 5' ends extending over a distance of several nucleotides (38-40). Such factors make it impossible to interpret precisely the results of these S1 experiments, and cap sites were arbitrarily assigned to A residues corresponding to the shortest member of each set of protected fragments.] In a second experiment, the furthest upstream promoter served to generate RNA species that protected the greatest proportion of the DNA fragments surviving S1 treatment (Fig. 7B). Both experiments used the same preparation of RNA prepared from pooled seminal vesicle tissue. While the exact contribution of the 3.5 and 3.8 kb SVS IV alleles to this message population is unknown, it is unlikely that this issue has any bearing on the existence of multiple cap sites since the difference between the alleles seems to be a single insertion in the second intron (44). While the exact proportion of protected fragments was clearly subject to experimental variation, these results argue strongly that a finite series of RNA transcripts do correlate with the upstream, redundant TATA sequences.

Additional Features of the Sequence Upstream from the Cap Site(s)

While a number of eukaryotic genes contain a homology sequence related to CAAT about -85 nucleotides from the cap site (41,42), this homology region is not evident for the SVS IV gene. An inverted repeat of 7 base pairs possibly capable of a stem-loop structure is located between nucleotides -102 and -122 (Fig. 5).

DISCUSSION

While it is increasingly common to find that proteins are encoded by members of gene families, SVS IV appears to be a single copy gene. Only one positive signal was obtained from 200,000 plaques of the genomic library, a frequency appropriate for a single copy gene. While Sprague-Dawley rat DNA contains two Eco RI fragments that hybridize strongly to an SVS IV probe, as reported initially by Abrescia et al. (43), the 3.5 and 3.8 kb fragments we detect have been shown to be allelic variants by Dickson and Harris (44). Our clone is for the 3.5 kb allele. Allelic variants with altered restriction site patterns are known for other genes such as beta globin (45) and ovalbumin (46). It remains to be seen whether there might be additional versions of the SVS IV gene present in rats in a form too divergent to hybridize well with the probes used in our screenings.

With secreted proteins such as SVS IV it is common though not universal to find an intron that separates the bulk of the mature protein from the hydrophobic signal peptide usually present on the precursor forms of such proteins. Such is true for several types of immunoglobulin (47), conalbumin (48), ovomucoid (49), and α fetoprotein (50). The first intron of SVS IV fits this pattern, interrupting the message coding sequence only 4 amino acids into the mature protein. From comparison of the restriction maps of SVS IV cDNA (7) and the 3.5 kb SVS IV gene, it appears that the second exon contains all of the remaining coding sequences, so that the final intron interrrupts only the 3'-nontranslated portion of the message. This is an unusual location for an intron but has also been found for α fetoprotein (50), its relative albumin (20), and for α2u globulin (51). While this paper was under review McDonald et al. (60) reported isolation of a genomic clone for SVS IV, and their analysis of intron locations agrees with our own.

Experiments based on reverse transcriptase catalyzed primer extention indicated that the vast majority of cellular SVS IV messages have caps located at the nucleotide designated +1 (Fig. 5) and that only traces of longer messages occur. Providing that no unusual secondary structure of the message led to premature termination, the extended primers should give a quantitatively accurate picture of the message population. This major transcriptional start corresponds to a TATA box sequence at -28 that agrees completely with the consensus sequence of TATAAA believed to function in setting the transcriptional start for RNA polymerase II (33). In fact this sequence is preceeded by an additional 5 AT base pairs, leaving some uncertainty as to the actual start of the homology sequence.

S1 protection experiments designed to overemphasize the contribution of messages giving the longest hybrids with an end-labeled probe showed that there are also rare transcripts that have 5'-ends at about positions -27 and -50. Each of these cap sites is preceeded by a TATA homology sequence lying about 30 nucleotides upstream. However, neither of these redundant TATA sequences is as good a match to the consensus as that preceeding the major cap site. The use of the upstream initation sites makes only a very minor contribution to the message pool, and it remains to be seen if they have any functional significance in the overall level of expression of the SVS IV gene. None of the longer transcripts contains an additional translational initiation codon, so they do not affect the amino terminal sequence of the nascent protein chain. Cases were multiple cap sites are quantitatively more significant than with SVS IV have been reported for liver amylase (52), chicken lysozyme (53), and ovomucoid (54). Interestingly, in all these cases the most downstream cap site is the dominant one, even when the TATA sequence associated with it is not the best fit to the consensus (55). An even greater heterogeneity of cap sites is found for the late messages of papova viruses, but this is perhaps due to the lack of clear TATA homologies with any of the apparent transcriptional initiation sites (32,59).

ACKNOWLEDGMENTS

We are grateful to Robert E. Taylor, Jr. and Daniel Burch for help in preparation of hybridization probes and isolation of messenger RNA and to Nancy Biggs for preparation of the manuscript. This work was supported by grants HD 13472 and GM 28021 from the U.S. Public Health Service.

REFERENCES

1. Ostrowski, M. C. , Kistler, M. K., and Kistler, W. S. (1979) J. Biol. Chem. 254, 383-390.
2. Pan, Y. E., Silverberg, A. B., Harris, S. E., and Li, S. S.-L. (1980) Int. J. Pept. Protein Res. 16, 143-146.
3. Kistler, M. K., Ostrowski, M. C., and Kistler, W. S. (1981) Proc. Nat. Acad. Sci. U.S.A. 78, 737-741.
4. Higgins, S. J., Smith, S. E., and Wilson, J. (1982) Molec. Cell. Endocrinol. 27, 55-65.
5. Higgins, S. J., and Burchell, J. M. (1978) Biochem. J. 174, 543-551.
6. Ostrowski, M. C., Kistler, M. K., and Kistler, W. S. (1982) Biochemistry 21, 3525-3529.
7. Mansson, P.-E., Sugino, A., and Harris, S. E. (1981) Nucleic Acid Res. 9, 935-946.
8. Kistler, M. K., Taylor, R. E., Jr., Kandala, J. C., and Kistler, W. S. (1981) Biochem. Biophys. Res. Commun. 99, 1161-1166.
9. McDonell, M. W., Simon, M. N., and Studier, F. W. (1977) J. Mol. Biol. 110, 119-146.

10. Maniatis, T., Jeffrey, A., and Kleid, D. G. (1975) Proc. Nat. Acad. Sci. U.S.A. 72, 1184-1188.
11. Sala-Trepat, J. M., Sargent, T. D., Sell, S., and Bonner, J. (1979) Proc. Nat. Acad. Sci. U.S.A. 76, 695-699.
12. Jeffreys, A. J., and Flavell, R. A. (1977) Cell 12, 429-439.
13. Southern, E. M. (1975) J. Mol. Biol. 98, 503-517.
14. Maniatis, T., Hardison, R. C, Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K., and Efstratiadis, A. (1978) Cell 15, 687-701.
15. Bolivar, F. (1978) Gene 4, 121-136.
16. Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heynecker, H. L., Boyer, H. W., Crosa, J. H., and Falkow, S. (1977) Gene 2, 95-113.
17. Clewell, D. B., and Helinski, D. R. (1969) Proc. Nat. Acad. Sci. U.S.A. 62, 1159-1166.
18. Humphreys, G. O., Willshaw, G. A., and Anderson, E. S. (1975) Biochim. Biophys. Acta 383, 457-463.
19. Hudson, B., Upholt, W. B., Devinny, J., and Vinograd, J. (1969) Proc. Nat. Acad. Sci. U.S.A. 62, 813-820.
20. Sargent, T. D., Wu, J.-R., Sala-Trepat, J. M., Wallace, R. B., Reyes, A. A., and Bonner, J. (1979) Proc. Nat. Acad. Sci. U.S.A. 76, 3256-3260.
21. Leder, P., Tiemeier, D., and Enquist, L. (1977) Science 196, 175-177.
22. Benton, W. D., and Davis, R. D. (1977) Science 196, 180-182.
23. Efstratiadis, A., Vournakis, J. N., Donis-Keller, H., Chaconas, G., Dougall, D. K., and Kafatos, F. C. (1977) Nucleic Acids Res. 4, 4165-4174.
24. Maxam, A., and Gilbert, W. (1977) Proc. Nat. Acad. Sci. U.S.A. 74, 560-564.
25. Maxam, A., and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
26. Lee, D. C., McKnight, G. S., and Palmiter, R. D. (1978) J. Biol. Chem. 253, 3494-3503.
27. Aviv, H., and Leder, P. (1972) Proc. Nat. Acad. Sci. U.S.A. 69, 1408-1412.
28. Berk, A. J., and Sharp, P. A. (1977) Cell 12, 721-732.
29. Favaloro, J., Treisman, R, and Kamen, R. (1980) Methods Enzymol. 65, 718-749.
30. Weaver, R. F., and Weissmann, C. (1979) Nucleic Acids Res. 7, 1175-1193.
31. Roberts, R. J. (1982) Nucleic Acids Res. 10, γ117-γ144.
32. Ghosh, P. K., Reddy, V. B., Swinscoe, J., Lebowitz, P., and Weissmann, S. M. (1978) J. Mol. Biol. 126, 813-846.
33. Breathnach, R., and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349-383.
34. Lehrach, H., Diamond, D., Wozney, J., and Boedker, H. (1977) Biochemistry 16, 4743-4751.
35. Thomas, P. S. (1980) Proc. Nat. Acad. Sci. U.S.A. 77, 5201-5205.
36. Ostrowski, M. C., Kistler, M. K., and Kistler, W. S. (1979) Int. Cong. Biochem. 11, 207.
37. Hentschel, C., Irminger, J. C., Bucher, P., and Birnstiel, M. (1980) Nature 285, 147-151.
38. McReynolds, L., O'Malley, B. W., Nisbet, A. D., Fothergill, J. E., Givol., D., Fields, S., Robertson, M., and Brownlee, G. G. (1978) Nature 273, 723-728.
39. Malek, L. T., Eschenfeldt, W. H., Munns, T. W., and Rhoads, R. E. (1981) Nucleic Acids Res. 9, 1657-1673.
40. Baker, C. C., and Ziff, E. B. (1981) J. Mol. Biol. 149, 189-221.
41. Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissmann, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C., and Proudfoot, N. J. (1980) Cell 21, 653-668.
42. Benoist, C., O'Hare, K., Breathnach, R., and Chambon, P. (1980) Nucleic Acids Res. 8, 127-142.
43. Abrescia, P., Guardiola, J., Felsani, A., and Metafora, S. (1982) Nucleic Acids Res. 10, 1159-1174.

44. Dickson, B. A., and Harris, S. E. (1982) J. Cell Biol. 95, 218a.
45. Kan, Y. W., and Dozy, A. M. (1978) Proc. Nat. Acad. Sci. U.S.A. 75, 5631-5635.
46. Lai, E. C., Woo, S. L. C., Dugaiczyk, A., and O'Malley, B. W. (1979) Cell 16, 201-211.
47. Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O., and Gilbert, W. (1978) Proc. Nat. Acad. Sci. U.S.A. 75, 1485-1489.
48. Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F., Chambon, P. (1979) Nature 282, 567-574.
49. Stein, J. P., Catterall, J. F., Kristo, P., Means, A. R., and O'Malley, B. W. (1980) Cell 21, 681-687.
50. Eiferman, F. A., Young, P. R., Scott, R. W., and Tilghman, S. M. (1981) Nature 294, 713-718.
51. Dolan, K. P., Unterman, R., McLaughlin, M., Nakhasi, H. L., Lynch, K. R., and Feigelson, P. (1982) J. Biol. Chem. 257, 13527-13534.
52. Hagenbüchle, O., Tosi, M., Schibler, U., Bovey, R., Wellauer, P. K., and Yound, R. A. (1981) Nature 289, 643-646.
53. Grez, M., Land, H., Giesecke, K., and Schütz, G. (1981) Cell 743-752.
54. Lai, E. C., Roop, D. R., Tsai, M.-J., Woo, S. L. C., and O'Malley, B. W. (1982) Nucleic Acids Res. 10, 5553-5567.
55. Young, R. A., Hagenbüchle, O., and Schibler, U. (1981) Cell 23, 451-458.
56. Sutcliffe, J. G. (1978) Nucleic Acids Res. 5, 2721-2728.
57. Prentki, P., Karch, F., Iida, S., and Meyer, J. (1981) Gene 14, 289-299.
58. Wahl, G. M., Stern, M., and Stark, G. R. (1979) Proc. Nat. Acad. Sci. U.S.A. 76, 3683-3687.
59. Cowie, A., Tynsall, C., and Kamen, R. (1981) Nucleic Acids Res. 9, 6305-6322.
60. McDonald, C., Williams, L., McTurk, P., Fuller, F., McIntosh, E., and Higgins, S. (1983) Nucleic Acids Res. 11, 917-930.