# UNIVERSITY OF ILLINOIS

May 16 19 88

THIS IS TO CERTIFY THAT THE THESIS PREPARED UNDER MY SUPERVISION BY

Katherine Lorene Schaefer

ENTITLED Sequence Analysis of Three Small Nuclear RNA Genes in the

Fission Yeast Schizosaccharomyces Pombe

IS APPROVED BY ME AS FULFILLING THIS PART OF THE REQUIREMENTS FOR THE

DEGREE OF Degree of Bachelor of Science in Biochemistry

Instructor in Charge

APPROVED:

HEAD OF DEPARTMENT OF Biochemistry

# SEQUENCE ANALYSIS

# OF THREE SMALL NUCLEAR RNA GENES IN

# THE FISSION YEAST SCHIZOSACCAROMYCES POMBE

BY

KATHERINE LORENE SCHAEFER

---

THESIS

for the

DEGREE OF BACHELOR OF SCIENCE

IN

BIOCHEMISTRY

College of Liberal Arts and Sciences

University of Illinois

Urbana, Illinois

1988

**ACKNOWLEDGEMENTS:**

I would like to thank Dr. Jo Ann Wise for support, encouragement, and guidance on this project. I also extend warmest thanks to the members of the Wise lab for help in all aspects of this project, especially to Pat Brennwald for advice on data analysis and Jean Jerome for always being willing to answer my questions and to find equipment that was usually right in front of me. Finally, thanks to May Chen ("it's not too late!") for much needed humor during discouraging times during this project.

# TABLE OF CONTENTS

## ABSTRACT:

I have isolated a genomic DNA clone containing the 5' end of the Schizosaccharomyces pombe U4 snRNA gene, as well as clones containing putative genes encoding snRNAs U5 and U6. In addition, I have sequenced the 5' portion of the S. pombe U5 and U6 RNAs. Both the primary sequence and secondary structure of the 5' end of the U4 RNA gene show significant similarities to mammalian and Saccharomyces cerevisiae U4 sequences and structures. The S. pombe U5 RNA contains limited sequence similarity to mammalian U5, but is conserved to about the same degree as U5 as S. cerevisiae; in addition, fission yeast U5 is nearly identical in size to mammalian U5 RNA, while the budding yeast component is significantly larger. S. pombe U6 RNA shows significant primary sequence similarity to mammalian U6 RNA, and Northern analysis shows it to be approximately the same size as mammalian U6.

# INTRODUCTION

The U-class small nuclear RNAs (snRNAs) are a class of RNAs implicated in splicing of mRNA precursors into mature mRNA. U-class snRNAs are small, discrete, and stable RNAs ranging in size from 56-217 nucleotides. With the exception of U3, which is found in the nucleolus and is not involved in pre-mRNA splicing (20), the U-class snRNAs are localized in the nucleoplasm (28). The snRNAs also have in common, with the exception of U6, a trimethylguanosine 5' cap structure. In the eukaryotic cell, the U-class snRNAs U1 - U6 are found complexed with proteins in small nuclear ribonucleoprotein particles (snRNPs). Except for U4/U6, snRNPs consist of one snRNA molecule per snRNP, associated with six to eight different proteins, at least one of which is recognized by antisera from patients with autoimmune disease (12). Each variety of snRNP, named after its component snRNA, is found in 200,000 to 1,000,000 copies per cell.

The U-class snRNAs were first suggested to be involved in splicing in 1980 (11, 23). The phylogenetic conservation in primary sequence of the

snRNAs, the fact that the snRNAs were present in highest abundance in metabolically active cell types, the observation that the snRNAs co-sedimented with 30S particles containing hnRNA, and the complementarity of a conserved U1 sequence to splice junctions in the pre-mRNA were noted. From this indirect evidence, the authors suggested that the U-class snRNAs are involved in splicing. This idea was further advanced by the obeservations that splicing was inhibited by anti-U1 antibodies (17), and by oligonucleotide directed RNAse H digestion of U1, U2 (3), or U4/6 (2,3). In addition, using an <u>in vitro</u> genetic manipulation of cloned genes, U1 was conclusively shown to hydrogen bond to the 5' splice junction in mammalian pre-mRNA (31), and U2 was shown to recognize the branch site of yeast pre-mRNA during splicing (18). The current model (for a recent review, see (26)) proposes that U1, U2, U4/6, and U5 associate on the splicing substrate in a multicomponent assembly called a spliceosome, with U1 at the 5' splice junction, U5 at the 3' splice junction, U2 at the branchpoint, and U4/6 associated with each other and U5 (but not the precursor RNA). Splicing then proceeds through a two step mechanism, in which first the 5' splice junction is cleaved and joined to

the branch site sequence within the intron. The free 3' hydroxyl at the end of the 5' exon sequence then attacks the 3' splice juction to give joined exons and a liberated intron lariat.

The structure of the snRNA, both on the primary sequence level and the secondary structure level, appears to be a determinant of function. This determination is obvious in the RNAs that directly associate with the RNA by base pairing (U1 and U2, mentioned above). Less obvious is the relationship between structure and function of molecules that associate without direct RNA-RNA association. The idea that structure is a functional determinant of the snRNAs is supported by two lines of evidence. The first is that primary and especially secondary sequence similarities are maintained over wide phylogenetic distances. U1 of chicken, rat, mouse, (4) and fruit fly (15), have striking primary sequence similarity (86%, 89%, 94%, and 72%, respectively) to human U1. In addition, the primary sequences can all be folded into very similar structures, as can the yeast S. pombe U1 ( G. Porter, unpublished data); even more interestingly, so can at least the 5' end of U1 from the yeast S. cerevisiae, although cerevisiae primary sequence similarity to the others

is limited and its size is greatly divergent from the others' (10, 25). Similarly, U2 of human, rat, mouse, frog, and chicken (21) all have very similar primary sequences and secondary structures; frog, S. cerevisiae (1), and S. pombe (Brennwald, et. al., manuscript in preparation) U2 can be fit into the mammalian secondary structure model, although again, the cerevisiae U2 is much longer than any of the others. U4 shows a similar pattern to U1 and U2. U4 from human, rat, and mouse have identical structures; chicken deviates by one nucleotide (21). Fly U4 can be folded into a secondary structure similar to rat/ human U4 (16), and S. cerevisiae U4 has some primary sequence and secondary structure similarities in the 5' domain to mammalian U4, conserving three stem-loop structures (Thomas and Blumenthal, manuscript in prep.) U6 has not been as widely studied, but human, rat, and mouse have identical sequences (21). Finally, human, mouse, and rat U5 primary sequences are very similar ( 6 nt. changes) (21); chicken, duck, and pheasant are 62% similar to mammalian (4), fly is 80% similar to mammalian, (16); and all of them, and in addition S. cerevisiae U5 (19) can be folded into very similar secondary structures, despite very limited primary sequence conservation between the S.

caravisiae RNA and the others'. As can be seen from this data, the secondary structure especially is maintained in eukaryotic organisms as diverse as humans and yeasts.

The second line of evidence that the primary and secondary structure is a determinant of function is supported by the observation that highly or absolutely conserved regions of the molecule tend to be in single stranded regions of the molecule suitable for interacting with other components of the splicing machinery or substrates. For instance, the 5' 13 nucleotides of U1 are identical between human and yeast; this region interacts with the 5' splice junction and is in a single stranded region of the molecule (15). The highly conserved regions in U4 and U6 that are proposed to bind to each other to aid or form the U4/U6 association are likewise in single stranded regions of the molecules (22). Thus, it appears that the primary and secondary structures of the U-class snRNAs are important by virtue of determining structure. Sequence data from as many organisms as possible is helpful in determining the absolutely and highly conserved structural regions of the molecule which correspond to functionally important domains.

The organism we have chosen to work with is the fission yeast Schizosaccharomyces pombe. S. pombe is an especially interesting organism because, while it can be genetically manipulated like prokaryotic cells, it shows significant structural and functional similarities to mammalian cells, more so than the most commonly studied yeast, S. cerevisiae. This latter observation leads to the hypothesis that S. pombe can be treated as a model for a mammalian splicing system. While S. cerevisiae U1 (10) and U2 (1) are much larger than human U1 and U2, and contain only limited, although significant secondary structure similarity, S. pombe U1 (G. Porter, unpublished data) is only 4 nucleotides shorter than and has a nearly identical secondary structure to human U1; similarly, S. pombe U2 is only 2 nucleotides shorter than human U2 and can be folded into a superimposable secondary structure model. Additional evidence for the structural similarity of human and S. pombe small RNAs is the 7SL RNA (5). Human 7SL RNA is 46 nucleotides larger than S. pombe 7SL; yet, they can be folded into highly similar structures. Most relevant to the goals of my work, S. pombe has been shown to correctly splice a mammalian transcript (8), suggesting a functional similarity between the mammalian

and S. pombe splicing machineries.

This thesis presents primary sequence data for three S. pombe snRNAs, U4, U5, and U6, based on RNA sequencing (U5 and U6) and DNA sequencing of a genomic DNA clone obtained by conserved oligonucleotide hybridization strategies (U4). Possible secondary structure models are presented, and comparisons made to the snRNAs of other organisms, with emphasis on the analogous budding yeast and mammalian RNAs.

## MATERIALS AND METHODS

Materials and Enzymes: Restriction enzymes were obtained from Bestheda Research Laboratories. Reverse transcriptase was obtained from Life Sciences, Inc., St. Petersburg, Fla.; T4 polynucleotide kinase was obtained from Bestheda Research Laboratories. The nick translation kit was from Amersham; RNA sequencing enzymes, nucleotides and the DNA sequencing "Sequenase" kit were from Pharmacia, Inc., Piscataway, N.J. Nitrocellulose was from Schleicher & Schuell, Inc., Keene, N.H.; GeneScreen from New England Nuclear Corp., Boston, Mass. ; and Hybond nylon membrane from Amersham. Bactotryptone, yeast extract, and agar were all obtained form Difco Laboratories, Detroit, Mi. Oligonucleotide sequencing primers and probes were made by the University of Illinois Biotechnology Center. [ $\alpha$ -$^{32}$P]dCTP and [ $\alpha$ -$^{35}$S]dATP were obtained from Amersham Corp., Arlington Heights, Ill., and [ $\gamma$ -$^{32}$P]ATP was from ICN Pharmaceuticals, Inc., Irvine Calif.

Screening for the clone: The S. pombe U4, U5, and U6 clones were

isolated from a bank of fission yeast DNA generated by digestion with Pst I, followed by insertion into the vector pUC18 (29). The E. coli strain JM101 was transformed with the Pst bank, using $CaCl_2$ transformation (12). The bank was plated on 2XYT - ampicillin plates (16 g/L bactotryptone, 10 g/L yeast extract, 6 g/L NaCl, pH to 7.5 with NaOH, 14 g/L agar, and 100mg/L ampicillin). Surviving colonies were then gridded, and colony filters made following the protocol of Grunstein and Hogness (6). The filters were probed with 5' $^{32}P$ labelled oligonucleotides and hybridized as previously decribed (13). The filters were then washed twice, first at room temperature in 6XSSC/0.1%SDS, and then at a higher stringency temperature (32 degrees C) in the same solution, with autoradiography after each wash. Potential clones were those that were significantly above background after both washes.

Restriction mapping: Enzyme digests were done in 15 microliter volumes, with 0.5 micrograms DNA, 1/10 volume appropriate 10X reaction buffer (supplied with the enzyme), 5 micrograms RNase A, greater than four units but less than 1/5 volume enzymes (depending on enzyme concentration), and water to volume. The digests were electrophoresed in

a 1% agarose gel with 0.5 micrograms ethidium bromide.

Sequence analysis: DNA sequencing was done with the "Sequenase" sequencing kit, according to the protocol provided. RNA primer extension and primer extension sequencing were done according to a modified protocol of Wagner, et. al (26).

Northern and Southern Analysis: Northern and Southern blots were made and probed as previously described by Wise, et. al (28).

Labelled probes: Nick translated probes were made using the protocol provided with the kit. 5' $^{32}P$ labelled probes were made using T4 polynucleotide kinase in a 20 microliter volume: 150 nanograms oligonucleotide, 1/10 volume 10X buffer (50 mM Tris-HCL, pH 8.0, 10 mM $MgCl_2$), 1000 microcuries [$\gamma$-$^{32}P$]ATP, 1 microliter T4 polynucleotide kinase, and water to volume, with an additional microliter added after 0.5 hours of incubation. Incubation was carried out at 37 degrees C for 1 hour.

# RESULTS AND DISCUSSION

## U4 snRNA

Identification and isolation of the U4 clone. Potential cloned genes were identified by hybridization to a 5' $^{32}P$ labelled oligonucleotide, U4a. Fig. 1a shows comparisons of human and S. cerevisiae U4 primary sequences. A region which has 13/16 identical nucleotides (shown boxed) was chosen as the site for U4a design. Where the two sequences differed, the human nucleotide was chosen. The complement of the RNA sequence was made into the oligonucleotide U4a. The complement was chosen so that both DNA and RNA hybridization experiments could be performed with this oligonucleotide. It was expected that the oligonucleotide would be at least a 15/16 match, which is sufficient to remain hybridized at the wash temperatures used.

Two colonies remained hybridized after the 32 degrees C wash. Fig. 2 shows the two colonies. Hybridization, as visualized by autoradiography, is significantly above background with these two colonies. The corresponding colonies from duplicate plates were grown in culture and were determined to be duplicates of each other by restriction mapping

(data not shown). One colony was picked at random for further characterization.

Sequence analysis of the U4 clone. The restriction map of the insert containing the clone and the location of the gene are shown in Fig. 3. The location of the gene was determined by DNA sequencing. Fig. 4 shows the DNA sequence (determined by direct DNA sequencing) of the part of the clone containing the gene and the derived RNA sequence. The exact location of the 5' end of the gene was determined by RNA primer extension using the U4a oligonucleotide (data not shown) and phylogenetic analysis. The 3' end of the gene is missing due to an unfortunate circumstance. The clone was isolated from a library constructed with the restriction enzyme Pst I (for details, see Materials and Methods), and the gene has a Pst I site at nucleotide 99, (Fig. 4) so we do not know the gene sequence beyond nucleotide 100. However, the 5' end of the gene reveals significant sequence and structural similarities between human, S. pombe, and S. cerevisiae U4.

Fig. 1b compares the RNA sequences of human, C. elegans, S. pombe, and S. cerevisiae U4. The most significant primary sequence homology

between S. pombe and human U4 is found between nucleotides 53 and 69 (15/17), which also contains the U4a site. This block of primary sequence is also found between human and S. cerevisiae from nucleotides 54 to 69 (13/16). Homology in this region is extended between C. elegans and human from nucleotides 48 to 79 (28/32). Also worth noting is the conservation of the first 10 nucleotides: 10/10 between C. elegans and human; 8/10 between S. pombe and human, and 7/10 between S. cerevisiae and human. On the basis of this data, S. cerevisiae U4 is found to be only a slightly better match at the primary sequence level to human than is S. pombe U4. It should be noted, however, that this pertains to only the 5' end of the molecule, and that the primary sequence similarity is not striking in either case, except for the U4a region.

Fig. 5 shows a comparison of possible secondary structures of human, S. pombe and S. cerevisiae U4. The conservation of the 5' stem and loop structure is striking. The stems start at almost the same site (20, 22, and 22 nt., respectively) and are of a similar length (9, 8, and 9 base pairs, respectively). All the stems have one G - U pair and one looped out nucleotide, and all the loops are approximately the same size ( 15, 11, and

17 nucleotides, respectively). Again, the S. cerevisiae secondary structure is very slightly more similar to human's than S. pombe in that the S. cerevisiae loop is closer in size to human's than S. pombe's. The highest degree of sequence conservation is found in the single stranded region, consistent with the hypothesis of a possible role for this region in binding to another spliceosome constituent.

Fig. 6 shows a comparison of the upstream sequence of various S. pombe U-class snRNAs. The area of most significant primary sequence conservation is boxed, and includes a TATA box-like sequence. The homology is 15/15 between U2 and U3, 10/15 between U1 and U2/3, and 11/15 between U4 and U2/3. This observation of the conserved upstream region at a constant position with respect to the start of transcription leads further credence to tha hypothesis that this clone contains the 5' end of U4. It also strengthens the argument that the S. pombe U-class snRNA genes share an upstream binding activating sequence, suggesting that they may employ a common transcription factor whose function is yet to be elucidated.

On the basis of this data, I conclude that I have cloned the 5' 2/3 of

the S. pombe U4 gene. While it is difficult to make a valid comparison at this point, owing to the lack of the 3' sequence data, it can be stated that S. pombe shows U4 primary sequence and especially secondary structure similarity to human U4 comparable to that of S. cerevisiae U4. This is somewhat surprising since, on the basis of the sequence and structure data for U1 and U2 snRNAs described in the Introduction, our lab has hypothesized that S. pombe U-class snRNAs are more closely related to human U-class snRNAs than are S. cerevisiae's. This may suggest that fission yeast U4 has a unique facet to its function. Firm conclusions must await sequencing of the 3' end of the S. pombe U4 gene. Toward this end, construction of a recombinant library using a different restriction enzyme is underway.

## U5 snRNA

Identification and isolation of the U5 clone. As with U4, U5 clones were identified from our Pst I bank using hybridization to 5' $^{32}$P labelled oligonucleotide, U5a. Fig. 7 shows a comparison of primary sequences between human, Tetrahymena, and S. cerevisiae U5. The two major areas of sequence similarity are boxed. As can be seen, the area between

nucleotides 35 and 49 is conserved (15/15) between Tetrahymena and human U5, and the same region is found (11/15) in S. cerevisiae U5, although S. cerevisiae U5 is much longer than the other two snRNAs. Like U4a, U5a was designed to be complementary to the 35 to 46 region so that it could be used in both DNA and RNA hybridization experiments, and where there is a nucleotide difference between S. cerevisiae U5 and human U5, the human nucleotide was used.

Possible U5-containing colonies were obtained in the same manner as the U4 colonies. Unfortunately, hybridization with U5a identified a colony later determined not to contain the U5 gene; a false positive ( G. Porter and H. Skinner, personal communication). Therefore, a new tactic was required to isolate the U5 clone. I used U5a to obtain a direct RNA primer extension sequence of S. pombe U5 . The sequence of this RNA is shown in Fig. 8. From this sequence, a new oligonucleotide, Grimm, identical to the RNA sequence, was designed. Since this oligonucleotide is identical to the S. pombe RNA sequence, it should hybridize only to true positive colonies. Fig. 9 shows the results of hybridizing colonies with the 5' $^{32}P$ labelled Grimm oligonucleotide. As with the U4, two possible colonies were

identified by visualization with autoradiography. The U5 signal is very obviously above background, much more so than the U4 signal. This difference can be attributed to the greater sequence homology of Grimm to the S. pombe U5 RNA (13/13) than the U4a oligonucleotide to the S. pombe U4 RNA (14/15). The two colonies were grown in culture and determined to be duplicates of each other by restriction mapping (data not shown). One colony was chosen at random for further characterization.

Sequence analysis of the U5 clone. The evidence that I have actually obtained the U5 clone is not as compelling as that for U4, as the characterization is not yet complete. The restriction map containing the clone is shown in Fig. 10. The location of the gene is undetermined at this point, except that we know that it is not located at the extreme 3' end (determined by direct DNA sequencing, data not shown). Two lines of evidence suggest that the clone is a bona fide U5 gene. The first is that direct DNA sequencing with the Grimm oligonucleotide gives a legible sequence. If the clone did not contain a perfect match to the Grimm sequence, it would not be possible to get this result.

The second line of evidence that this is actually a U5 clone is the

limited primary sequence similarity to human U5. Fig. 11 shows a possible alignment of the S. pombe U5 RNA sequence with the human U5 RNA. It should be noted that the U5a binding site has not been directly sequenced; however, its presence can be inferred from the fact that U5a hybridized well enough to S. pombe U5 RNA to produce a legible RNA primer extension sequence. No major blocks of primary sequence homology exist between the two organisms outside of the U5a region. However, this is not suprising, as U5 primary sequence is not well conserved between S. cerevisiae, pea, and rat U5 (Patterson and Guthrie, 1987), although the secondary structure is quite strikingly conserved. A secondary structure analysis of the S. pombe U5 RNA cannot be attempted at this point, as insufficient sequence data has been obtained.

From this evidence, it is reasonable to suggest that the clone that I have isolated contains the U5 gene. Analysis of the complete DNA sequence and the upstream region are needed before any firm conclusions can be drawn. These experiments are underway.

## U6 snRNA

Identification and isolation of a U6 clone. The possible U6 clones

were identified from our Pst 1 bank by hybridization to a $^{32}P$ 5' labelled oligonucleotide, U6a. It was designed (by D. Brow and C. Guthrie) in the same manner as U4a and U5a. Fig. 12 shows sequence comparisons of human, broadbean, and Drosophila U6. As can be seen, the U6 primary sequence is very highly conserved between these organisms, much more so than the U4 or U5 sequences. The region from which U6a was designed is boxed.

Fig. 13 shows hybridizations of possible U6 colonies with $^{32}P$ 5' labelled U6a. Again, two colonies above background were visualized by autoradiography. No analysis of these has been performed to date, but a similar analysis to the one performed on U4 and U5 will be carried out.

Analysis of the U6 RNA sequence. Fig. 14 shows part of the RNA sequence of S. pombe U6, obtained by primer extension sequencing with U6a, compared with rat and human sequences. Two major blocks of similarity can be noted in addition to the U6a binding site: one 10/12 and one 6/7 (shown boxed). This result suggests that U6a is a successful oligonucleotide and binds to a sequence in S. pombe U6, rather than to a non-snRNA related sequence, which strengthens the hypothesis that at

least one of the clones that I have isolated contains the U6 gene.

Finally, a "zoo" blot of RNA from several different organisms was probed with $^{32}P$ 5' labelled U6a (Fig. 15). As shown, U6a hybridizes to RNAs in all of the organisms, showing that at least the U6a region is widely conserved. It also shows that the human and S. pombe U6 RNAs are roughly the same size, while S. cerevisiae's U6 RNA is significantly larger. This is consistent with the idea that S. pombe U-class snRNAs are more closely related to human U-class snRNAs. Sequence analysis of the gene encoding the S. pombe U6 RNA needs to be performed to determine if U6 does indeed support this hypothesis.

In summary, it can be stated that my evidence strongly suggests that I have isolated S. pombe genes for each of the snRNAs U4, U5, and U6, and that the snRNAs contain significant structural similarities to human Y4, U5, and U6.

## REFERENCES


(1) Ares, M. (1986) Cell 47, 49-59.

(2) Berget, S. M. and Robberson, B. L. (1986) Cell 46, 691-696.

(3) Black, D. L. and Steitz, J. A. (1986) Cell 46, 229-239.

(4) Branlant, C., Krol, A., Lazar, E., Haendler, B., Jacob, M., Galego-Diaz, L., and Pousada, C. (1983) Nucl. Ac. Res. 11, 8359-8368.

(5) Brennwald, P., Liao, X., Holm, K., Porter, G., and Wise, J. A. (1988) Mol. and Cell. Biol. 8, 1580-1590.

(6) Myslinski, E. and Branlant, C. (1984) J. Mol. Biol. 180, 927-945.

(7) Grunstein, M. and Hogness, D. (1975) Proc. Nat'l. Acad. Sci. 72, 3961-3965.

(8) Kaufer, N., Simanis, V., and Nurse, P. (1985) Nature 318, 78-80.

(9) Kiss, T., Antal. M., and Solymosy, F. (1987) Nucl. Ac. Res. 15, 543-560.

(10) Kretzner, L., Rymond, B., and Rosbash, M. (1987) Cell 50, 593-602.

(11) Lerner, M., Boyle, J. A., Mount, S. M., Wolin, S. L., and Steitz, J. A. (1980) Nature 283, 220-224.

(12) Lerner, M. and Steitz, J. (1979) Proc. Nat'l. Acad. Sci. 76, 5495-5499.

(13) Mandel, A. and Higa, T. (1970) J. Mol. Biol. 53, 154-164.

(14) Maniatis, T., Frisch, E. F., and Sambrook, J. (1982) Molecular Cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

(15) Mount, S. M. and Steitz, J. A. (1981) Nucl. Ac. Res. 23, 6351-6368.

(16) Myslinski, E. and Branlant, C. (1984) J. Mol. Biol. 180, 927-945.

(17) Padgett, R., Mount, S. M., Steitz, J. A., and Sharp, P. A. (1983) Cell 31, 101-107.

(18) Parker, R., Siliciano, P., and Guthrie, C. (1987) Cell 49, 229-239.

(19) Patterson, B, and Guthrie, C. (1987) Cell 49, 613-624.

(20) Prestayako, A. W., Tonato, M., and Busch, H. (1971) J. Mol. Biol. 47, 505-515.

(21) Reddy, R. (1986) Nucl. Ac. Res. 13 supl, r155-r163.

(22) Rinke, J., Appel, B., Digweed, M., and Luhrmann, R. (1985) J. Mol. Biol. 185, 721-731.

(23) Rogers, J. and Wall, R . (1980) Proc. Nat'l. Acad. Sci. 77, 1877-1879.

(24) Baluz, H., Choffat, Y., and Kubli, E. (1988) Nucl. Ac. Res. 16, 1202.

(25) Silicano, P., Brow, D. A., Roiha, H., and Guthrie, C. (1987) Cell 50, 585-592.

(26) Sharp, P. A. (1987) Science 235, 766-770.

(27) Wagner, M., Sharp, J., and Summers, W. (1981) Proc. Nat'l. Acad. Sci. 78, 1441-1445.

(28) Weinberg, R. and Penman, S . (1968) J. Mol. Biol. 38, 289-304.

(29) Wise, J. A., Tollervey, D., Maloney, D., Swerdlow, H., Dunn, E. J., and Guthrie, C. (1983) Cell 49, 613-624.

(30) Yanuch-Perron, C., Vierra, J., and Messing, J. (1985) Gene 33, 103-109.

(31) Zhuang, Y. and Weiner, A. (1986) Cell 46, 827-835.

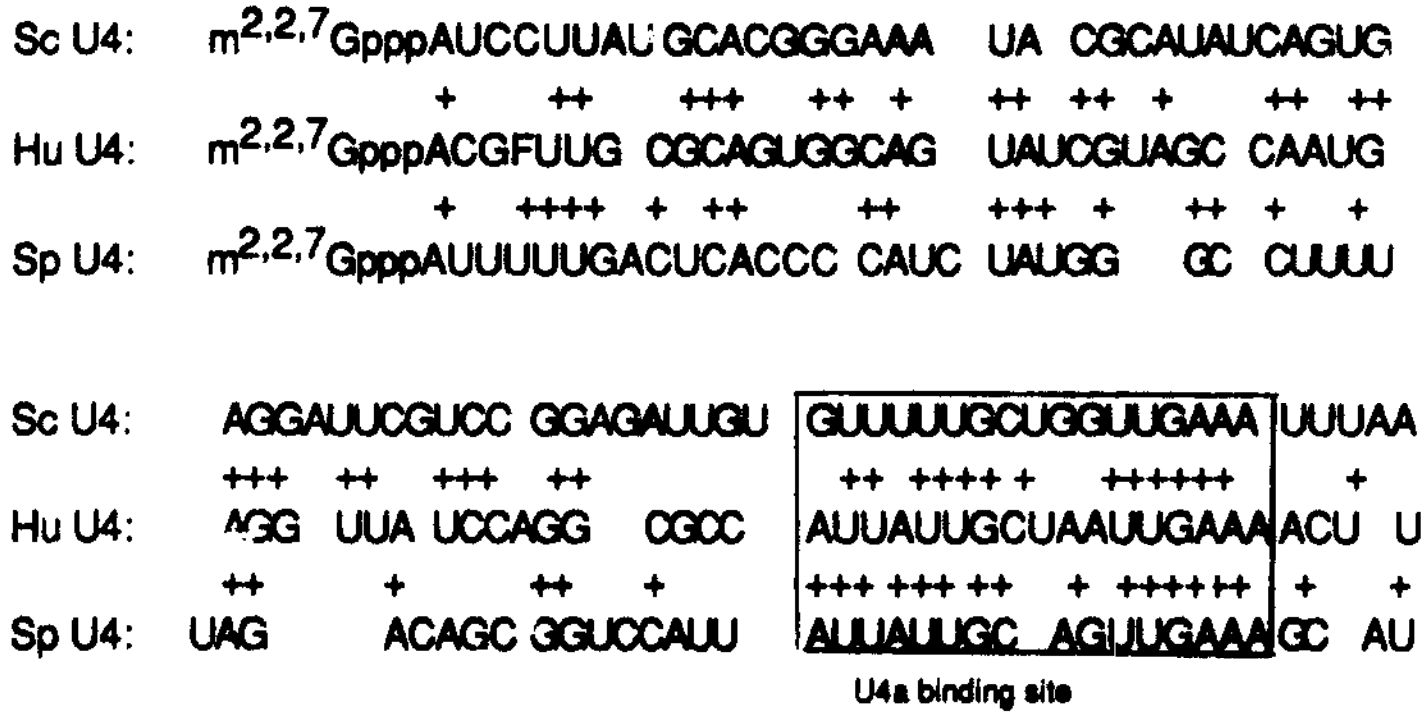



Figure 1a. Comparison of U4 sequences in three organisms. Sc = S. cerevisiae (24); Hu = HeLa U4 (21); and Sp = S. pombe U4. Identical nucleotides are denoted by (+), and the U4a binding site is boxed. U4a is complementary to this site.

```
Ce U4:   m2,2,7GpppAGCUUUG  CGCUGGGGC GAUAACGUGACCAA
                  +++ +++ +  +++  +  +++  +  ++  +++   ++ ++
Hu U4:   m2,2,7GpppAGCFUUG  CGCAGUGGCAG UAUCGUAGCCAA
                   +   ++++  +  ++      ++   +++   +  +++
Sp U4:   m2,2,7GpppAUUUUUGA CUCACC   CAUCUAUGG  GCCUU

Ce U4:   UGAGGCUUUGCCGAGGUGC   CUUUAUUGCUGGUUGAAAACUUU
         +++++  +  +  ++  +++  ++     +++ ++++ +   +++ ++++ ++++
Hu U4:   UGAGGUUAU CC  AGGCGC  GAUUAUUGCUAAUUGAAAACUUF
         +  ++     +  +       +      +++ +++ ++    +++ +++     +
Sp U4:   UUAGAC  AGC  GGUUCCAU UAUUAUUGCAG  UUGAAAGCAUA

Ce U4:   UCCC  AA  UUGCCCGCGAUGUCCCCU GAAACAUGGGGUGGCA
         ++ ++  ++  +   ++ +++  ++      ++  +  +  +  ++  +    ++ ++
Hu U4:   UCCC  AA  FACCCCGCCGUGACGACUUGCAAUAUAGUC  GGCA
         ++ ++  ++          ++      +++  ++  ++   ++         ++  +
Sp U4:   UCCCCAA            CGAA  UGA  GA  UU   AAAUAU  UCU GCAG

Ce U4:   UACGCAAUUUUUGACGCCUCU   A GGAGGCAGGA OH
         +   ++++ +++++++++ ++    +++   +  +++ + +  ++
Hu U4:   UUGGCAAUUUUUGACGAGUCUCUACGGAGACUGG OH
         ++
Sp U4:   UU
```

Figure 1b. Comparison of U4 sequences from 3 organisms. Hu = HeLa U4 (21); Ce = C. elegans U4 (T. Blumenthal, pers. comm.); and Sp = S. pombe U4. Identical nucleotides are denoted by (+).

Figure 2. Colony filters probed with 5' $^{32}P$ labelled U4a:

5' TTTCAACCAGCAATAA 3'
TT

Hybridization was carried out at 25 degrees C, and the final wash was done at 32 degrees C.

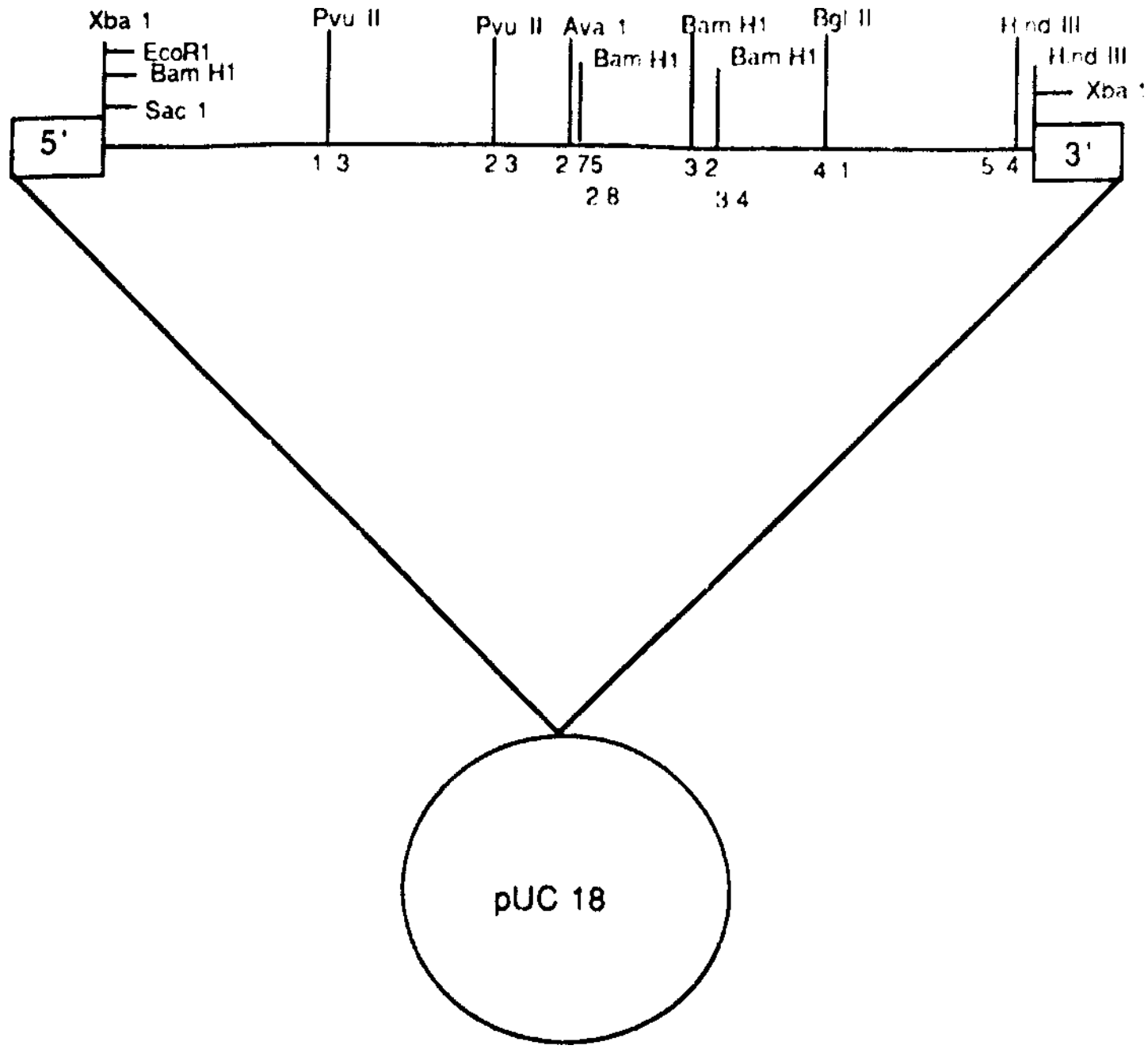



Figure 3. The restriction map of the DNA clone containing the 5' 2/3 of the U4 gene. The restriction map was determined as described in Materials and Methods. The gene is shown with an arrow. The insert is approximately 5.5 kilobases long, and numbers below the sites refer to the distance from the 5' polylinker.

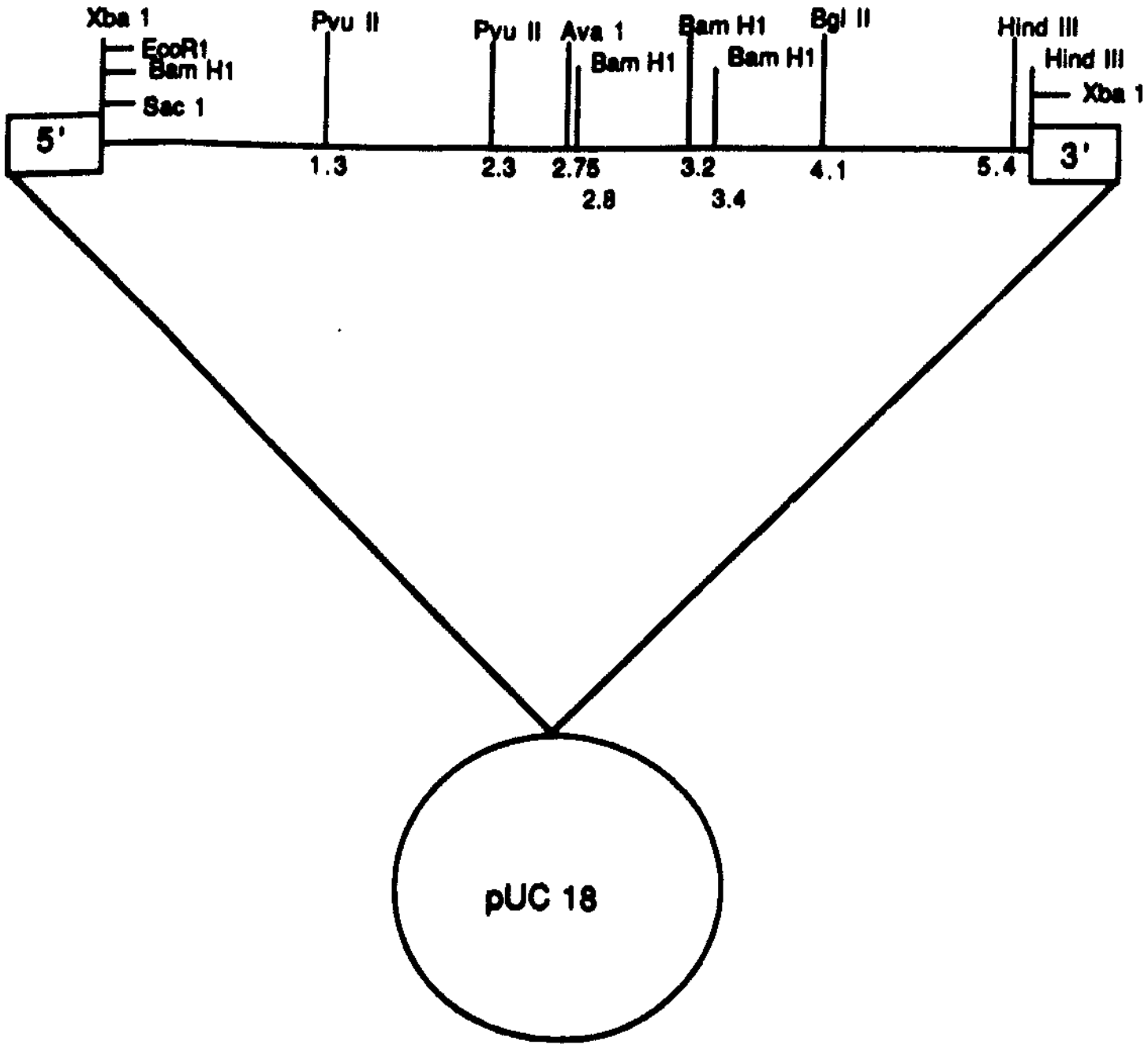



Figure 3. The restriction map of the DNA clone containing the 5' 2/3 of the U4 gene. The restriction map was determined as described in Materials and Methods. The gene is shown with an arrow. The insert is approximately 5.5 kilobases long, and numbers below the sites refer to the distance from the 5' polylinker.

Hind III

CGCTGCTGTTGGTTTAATCATTGGCGTCGGAAGCTTTCTCCATTTCTCAT

Xba I

CCCTTTATGAAGAATCGCAAACAATTTATTTCTAGAGACGCAGGTGTTAC

ACCTGGTATCGGATTTTTGACTCACCCCATCTATGGGCCTTTTTAGACAGC

$m^{2,2,7}$GpppAUUUUUGACUCACCCCAUCUAUGGGCCUUUUUAGACAGC

GGTTCCATTATTATTGCAGTTGAAAGCATATCCCCAACGAATGTGATTAA

GGUUCCAUUAUUAUUGCAGUUGAAAGCAUAUCCCCAACGAAUGUGAUUAA

Pst I

ATATTCTGAG . . .

AUAUUCUGAG . . .

Figure 4. DNA sequence of the part of the clone containing the U4 gene and the corresponding RNA sequence. The 5' end of the gene was determined by RNA primer extension and phylogenetic analysis. The gene is truncated after nucleotide 99.

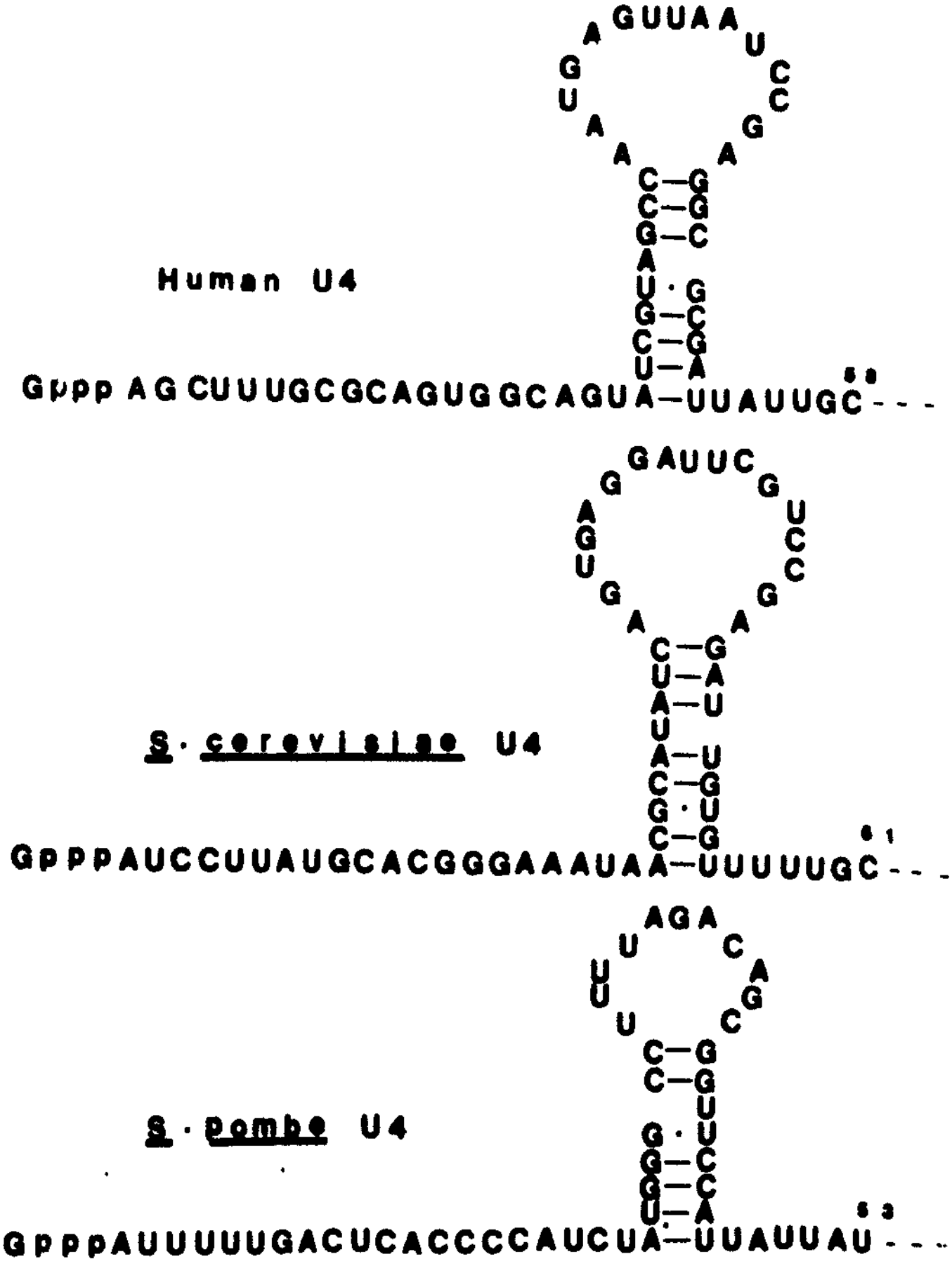



Figure 5. Comparison of the secondary structures of U4 in three organisms. Only the 5′ end is shown.

```
U4:   CCTTATGAAGAATCGCAAAC  ATTTATTTC TAGAGAC
         +  +      ++       +  +++ ++  +   ++  +  ++
U2:   GGGTTCGGTTTATTTATGAG  ATTTACTA  TATATAC
      +  +  +  +  ++  +  +  ++ ++ ++ ++  ++ +++++
U3:   GAATGTGACTCTTTCGTTAT  ATTTACTA  TATATAC
      +     ++   ++ +++     +    +++ +++  ++  +++
U4:   GCGACTGCTTCGTTCCGCGT  CGTTACTA  TAAATAT
```

```
                                          1
U4:     GCAGGTGTT ACCTGGTAUCGGAUU
        +  +     ++  +   ++ ++    +  +++
U2:   AGGTACTCGTGCATTTGGTTCGGTAUU
       +++   +   + +
U3:    GGTCATTATCCTCGAACAAAC
       ++++  +++  ++        +  +
U4:   GGGTCGTTACACTTACGGATTCTTTACU
```

Figure 6. Comparison of U6 15 homologies in S. pombe snRNAs. Identical nucleotides are denoted by (+), and the U6 15 box is outlined. Numbering is with respect to U4, and 1 refers to the start of the gene.

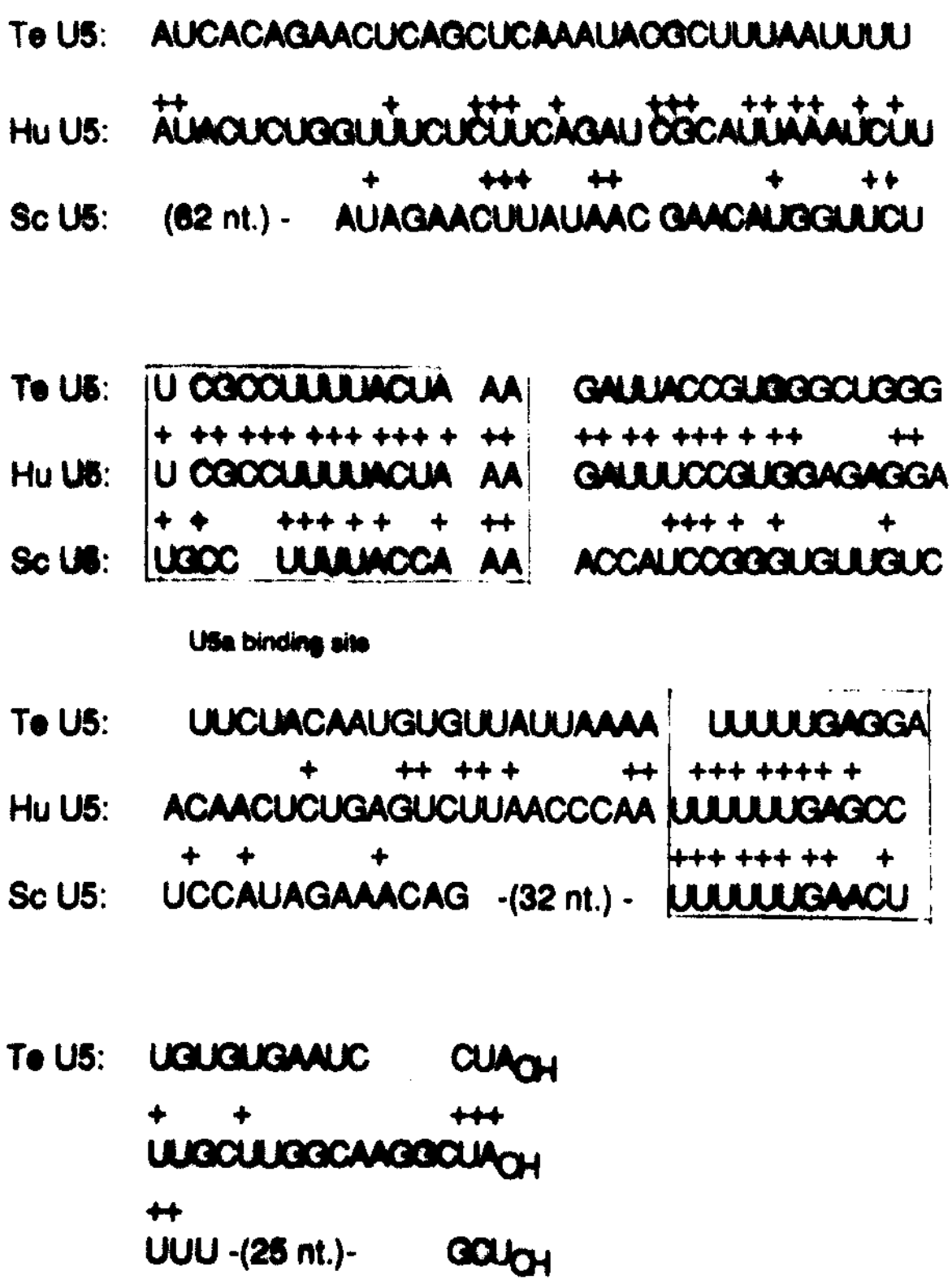



Figure 7. Conserved U5 sequences from three organisms. Hu = HeLa U5 (4); Te = Tetrahymena U5 (4); and Sc = S. cerevisiae U5 (19). Identical nucleotides have been denoted by a (+), and numbering and sequence homologies are with respect to the human sequence. Regions of significant homology are boxed. U5a is complementary to the RNA.

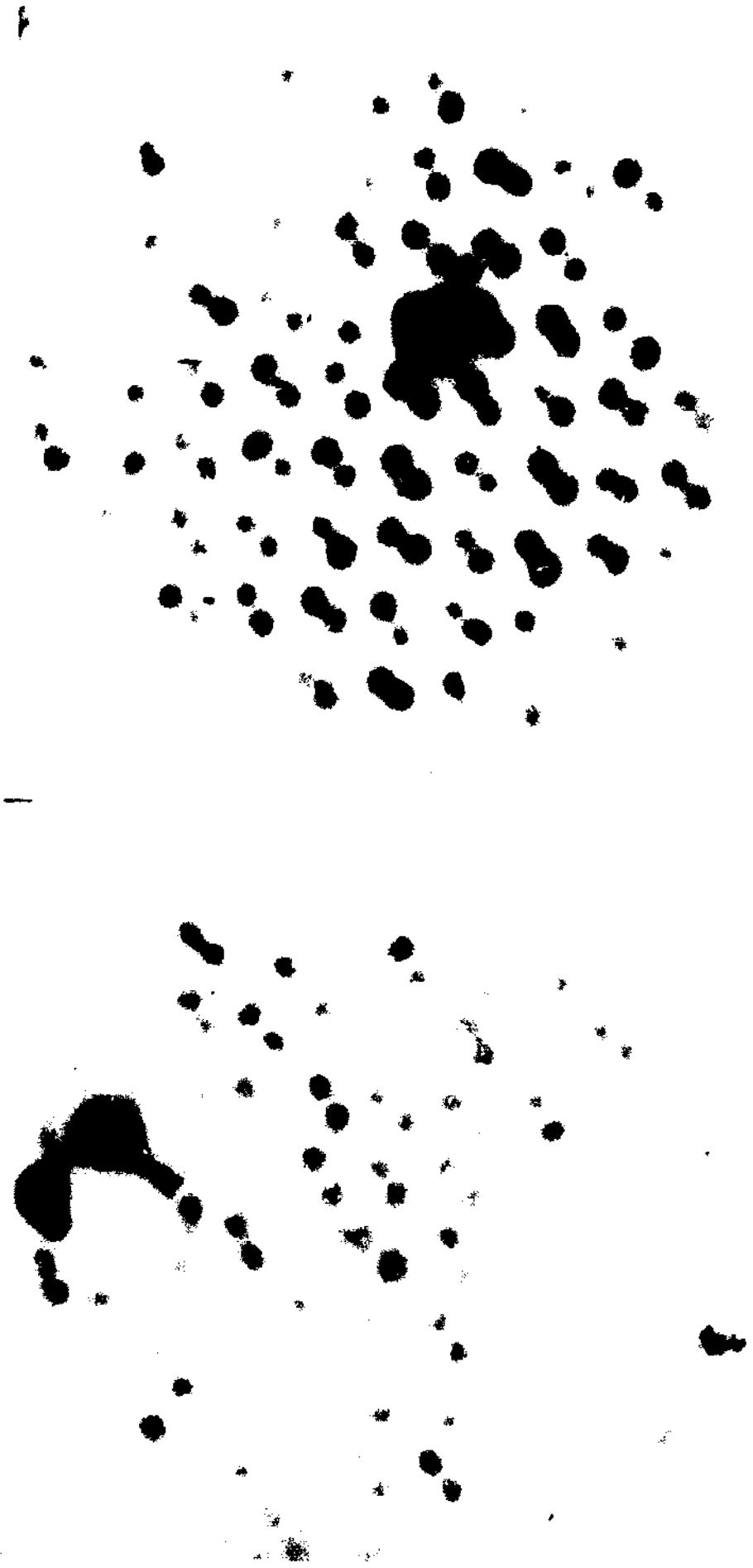
```
RNA:    5'AAAUC  GUCAAAGCACUUUGC  AAAAGCUAACGUCUC3'
Grimm:        5'  GTCAAAGCACTTTGC  3'
```

Figure 8. Design of the Grimm oligonucleotide. The RNA primer extension sequence data was obtained using the U6a oligonucleotide. Grimm was designed to be the DNA equivalent of the RNA sequence.

Figure 9. Colony filters probed with 5' $^{32}P$ labelled Grimm:

5' GTCAAAGCACTTT 3'

Hybridization was carried out at 25 degrees C, and the final wash was done at 32 degrees C.



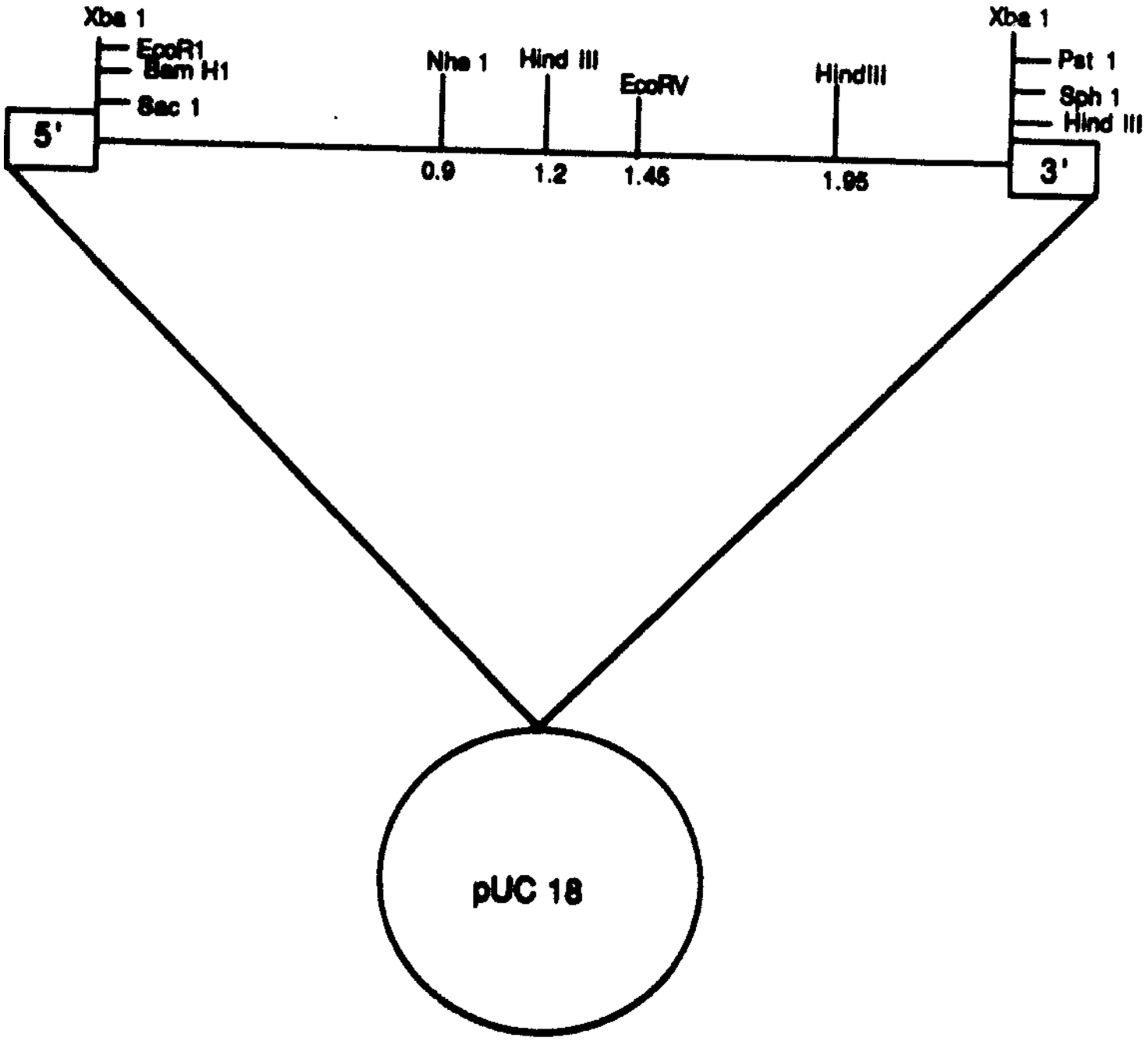



Figure 10. Restriction map of the DNA clone containing the U5 gene. Restriction mapping was done as described in Materials and Methods. The location of the gene has not yet been determined. The insert is approximately 2.35 kilobases long, and numbers below the sites refer to distances from the 5' end of the polylinker.

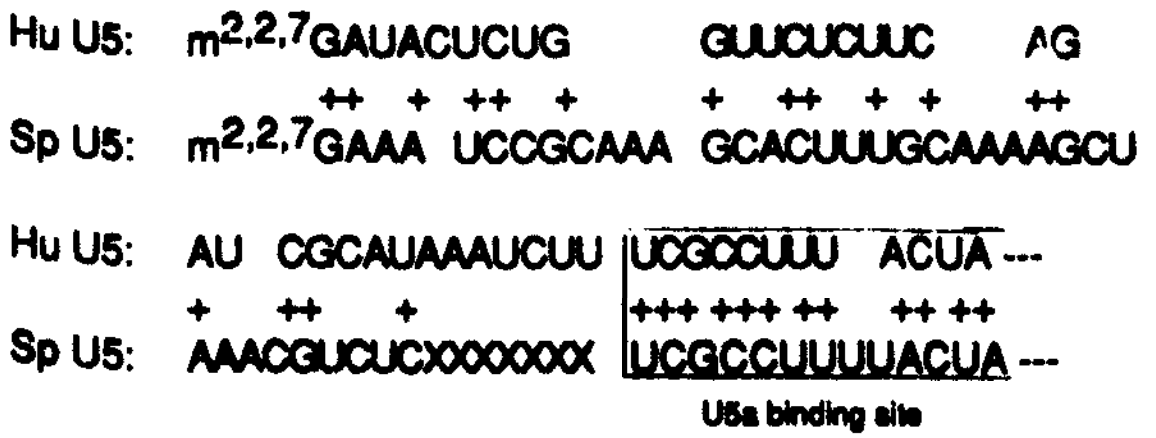



Figure 11. A possible alignment of the S. pombe U5 RNA sequence (determined by direct RNA sequencing) with the human sequence. Sp = S. pombe U5 RNA; Hu = HeLa U5 RNA sequence (4). Nucleotides that are unknown (have not been sequenced) are denoted by a (X); identical nucleotides by a (+). The S. pombe U5a nucleotides have not yet been verified by sequencing, but indirect evidence exists for their identity to human U5a region nucleotides (for details, see Discussion).

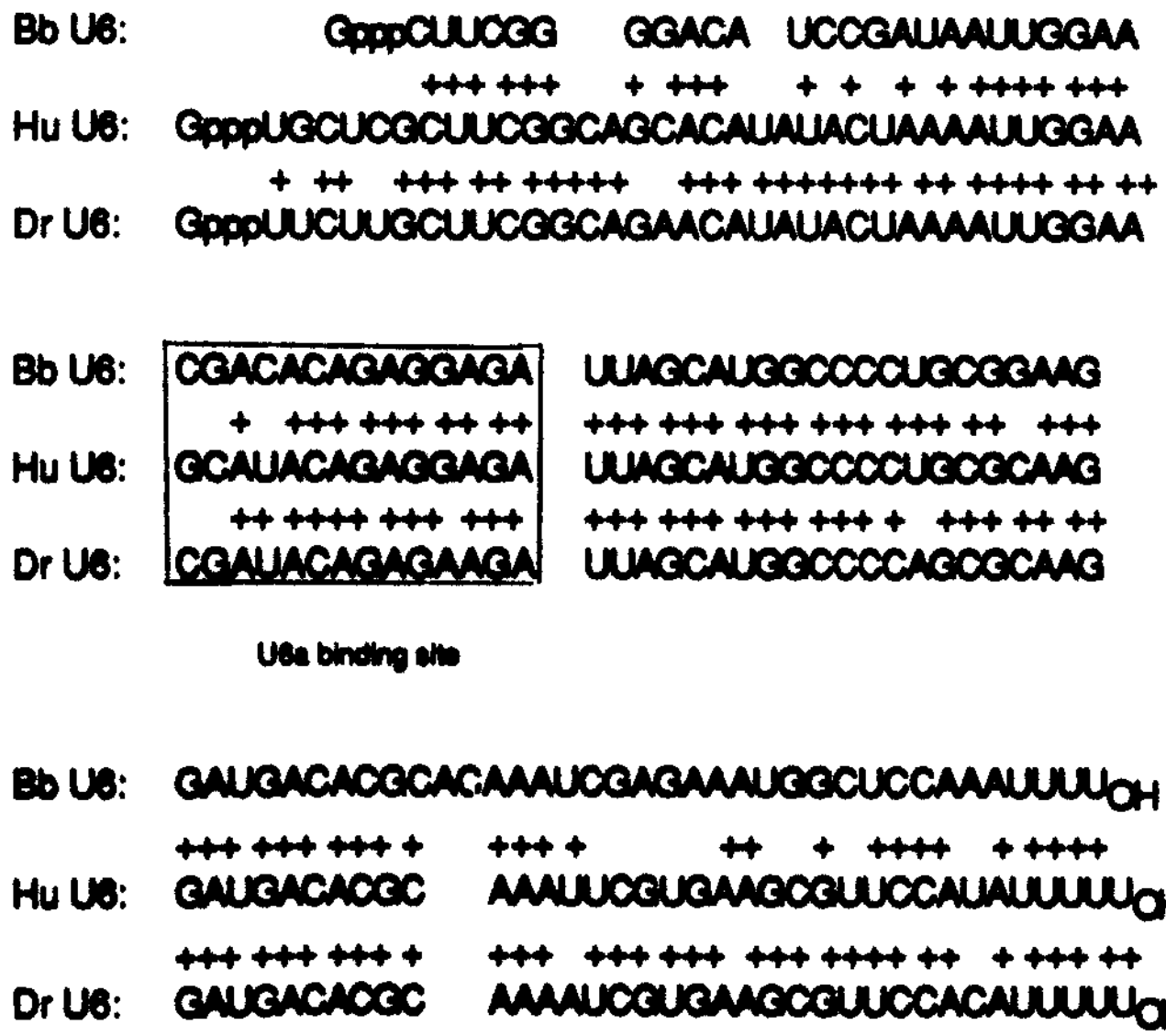


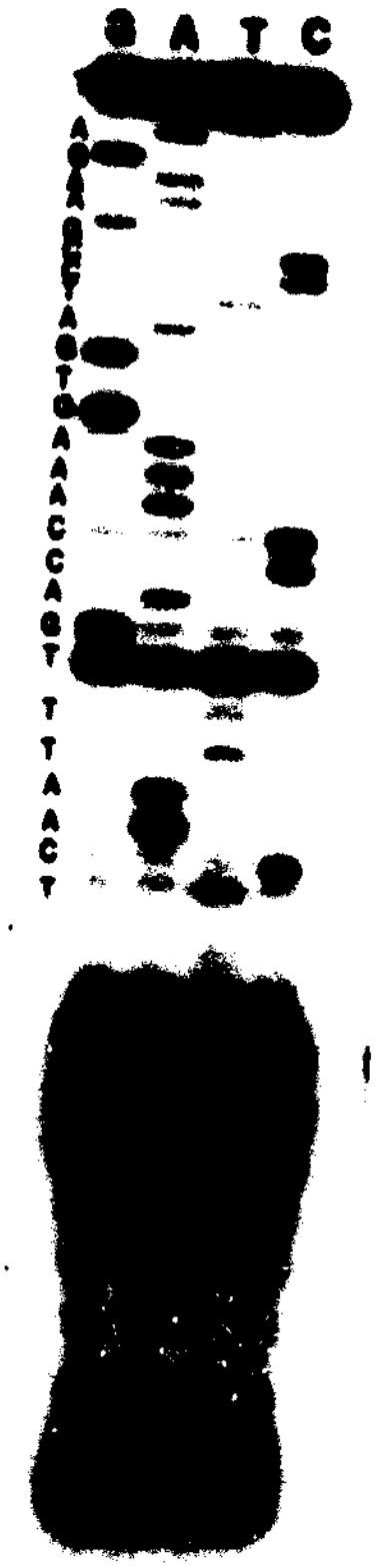

Figure 12. Conserved sequences of U6 in three species. Bb = Broad bean U6 (9); Hu = HeLa U6 (22); and Dr = Drosophila U6 (24). Identical nucleotides are denoted by (+), and all numbering and sequence homologies are determined with respect to the human sequence. The sequence from which U6a is derived is boxed. U6a is complementary to the RNA sequence, and contains an A/T degeneracy at position 3.

Figure 13. Colony filters probed with 5' $^{32}P$ labelled U6a:

5' TCATCTCTGTATTG 3'
T

Hybridization was carried out at 25 degrees C, and the final wash was done at 32 degrees C.

Figure 14a. The RNA primer extension sequence obtained using U6a oligonucleotide. The sequence shown to the right of the lanes is the complement of the S. pombe U6 sequence.



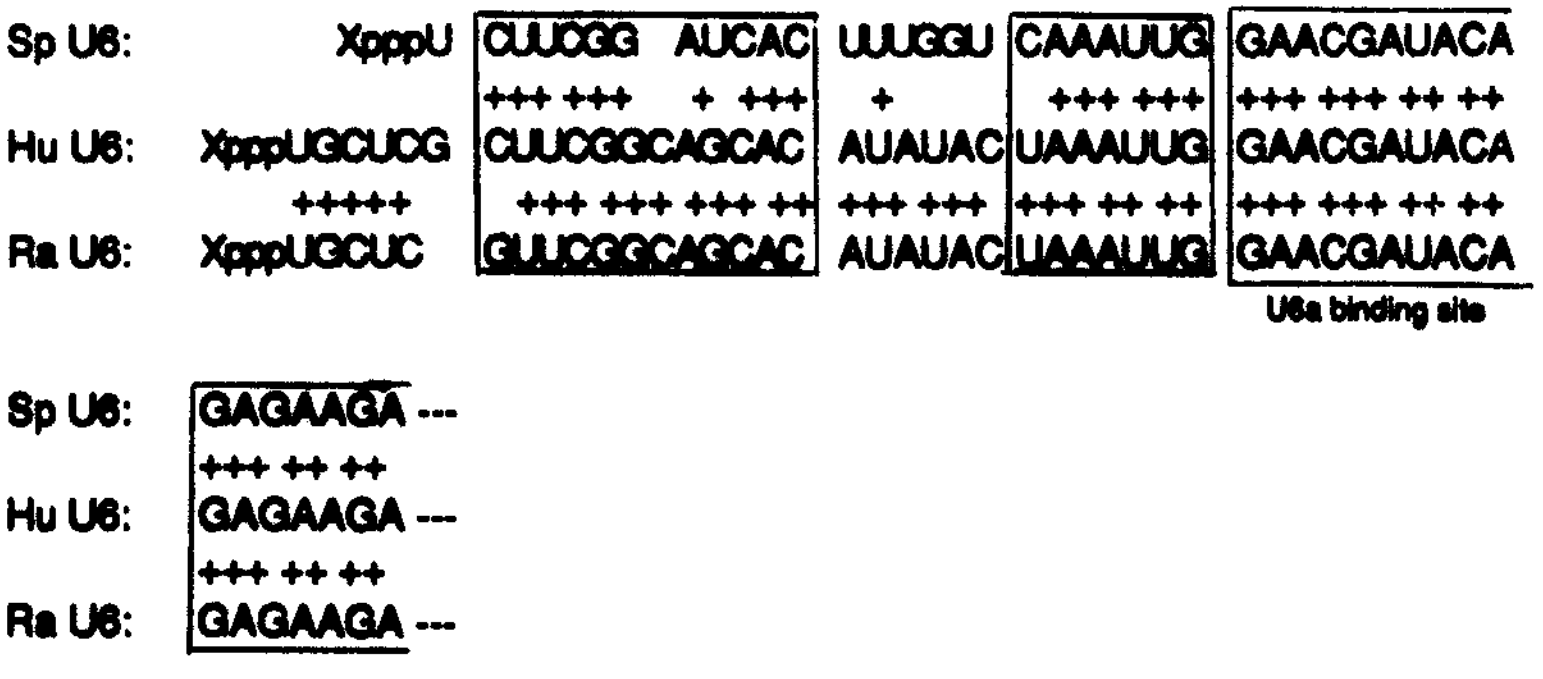



Figure 14b. Comparison of the 5' end of the U6a gene in three organisms. Identical nucleotides are denoted by a (+), and sequence homologies are determined with respect to the human sequence. Sp = S. pombe U6 RNA; Hu = HeLa U6 RNA (21); and Ra = Rat U6 RNA (21). The major regions of conservation are boxed. As with U5, the S. pombe U6a nucleotides have not yet been verified by sequencing, but indirect evidence exists for their identity to human U6a region nucleotides.


Figure 15. "Zoo" Northern containing RNAs from several species probed with U6a. Hybridization and washing were carried out at 25 degrees C