The Cloning and Characterisation of the U1, U2, U4 and U6 snRNA's of *Schistosoma mansoni*

BY

TERRY BLAKE BALL

A Thesis
Submitted to the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree of

MASTERS OF SCIENCE

Department of Medical Microbiology
University of Manitoba
Winnipeg, Manitoba

ISBN 0-612-12971-3

Canada

Name Terry Blake Ball

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

Microbiology
SUBJECT TERM

0410
SUBJECT CODE

U·M·I

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS
Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION
General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523
Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS
Language
- General 0679
- Ancient 0289
- Linguistics 0290
- Modern 0291

Literature
- General 0401
- Classical 0294
- Comparative 0295
- Medieval 0297
- Modern 0298
- African 0316
- American 0591
- Asian 0305
- Canadian (English) 0352
- Canadian (French) 0355
- English 0593
- Germanic 0311
- Latin American 0312
- Middle Eastern 0315
- Romance 0313
- Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY
Philosophy 0422
Religion
- General 0318
- Biblical Studies 0321
- Clergy 0319
- History of 0320
- Philosophy of 0322

Theology 0469

SOCIAL SCIENCES
American Studies 0323
Anthropology
- Archaeology 0324
- Cultural 0326
- Physical 0327

Business Administration
- General 0310
- Accounting 0272
- Banking 0770
- Management 0454
- Marketing 0338

Canadian Studies 0385
Economics
- General 0501
- Agricultural 0503
- Commerce-Business 0505
- Finance 0508
- History 0509
- Labor 0510
- Theory 0511

Folklore 0358
Geography 0366
Gerontology 0351
History
- General 0578
- Ancient 0579
- Medieval 0581
- Modern 0582
- Black 0328
- African 0331
- Asia, Australia and Oceania 0332
- Canadian 0334
- European 0335
- Latin American 0336
- Middle Eastern 0333
- United States 0337

History of Science 0585
Law 0398
Political Science
- General 0615
- International Law and Relations 0616
- Public Administration 0617

Recreation 0814
Social Work 0452
Sociology
- General 0626
- Criminology and Penology ... 0627
- Demography 0938
- Ethnic and Racial Studies 0631
- Individual and Family Studies 0628
- Industrial and Labor Relations 0629
- Public and Social Welfare 0630
- Social Structure and Development 0700
- Theory and Methods 0344

Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES
Agriculture
- General 0473
- Agronomy 0285
- Animal Culture and Nutrition 0475
- Animal Pathology 0476
- Food Science and Technology 0359
- Forestry and Wildlife 0478
- Plant Culture 0479
- Plant Pathology 0480
- Plant Physiology 0817
- Range Management 0777
- Wood Technology 0746

Biology
- General 0306
- Anatomy 0287
- Biostatistics 0308
- Botany 0309
- Cell 0379
- Ecology 0329
- Entomology 0353
- Genetics 0369
- Limnology 0793
- Microbiology 0410
- Molecular 0307
- Neuroscience 0317
- Oceanography 0416
- Physiology 0433
- Radiation 0821
- Veterinary Science 0778
- Zoology 0472

Biophysics
- General 0786
- Medical 0760

EARTH SCIENCES
Biogeochemistry 0425
Geochemistry 0996
Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES
Environmental Sciences 0768
Health Sciences
- General 0566
- Audiology 0300
- Chemotherapy 0992
- Dentistry 0567
- Education 0350
- Hospital Management 0769
- Human Development 0758
- Immunology 0982
- Medicine and Surgery 0564
- Mental Health 0347
- Nursing 0569
- Nutrition 0570
- Obstetrics and Gynecology .. 0380
- Occupational Health and Therapy 0354
- Ophthalmology 0381
- Pathology 0571
- Pharmacology 0419
- Pharmacy 0572
- Physical Therapy 0382
- Public Health 0573
- Radiology 0574
- Recreation 0575
- Speech Pathology 0460
- Toxicology 0383

Home Economics 0386

PHYSICAL SCIENCES
Pure Sciences
Chemistry
- General 0485
- Agricultural 0749
- Analytical 0486
- Biochemistry 0487
- Inorganic 0488
- Nuclear 0738
- Organic 0490
- Pharmaceutical 0491
- Physical 0494
- Polymer 0495
- Radiation 0754

Mathematics 0405
Physics
- General 0605
- Acoustics 0986
- Astronomy and Astrophysics 0606
- Atmospheric Science 0608
- Atomic 0748
- Electronics and Electricity 0607
- Elementary Particles and High Energy 0798
- Fluid and Plasma 0759
- Molecular 0609
- Nuclear 0610
- Optics 0752
- Radiation 0756
- Solid State 0611

Statistics 0463

Applied Sciences
Applied Mechanics 0346
Computer Science 0984
Engineering
- General 0537
- Aerospace 0538
- Agricultural 0539
- Automotive 0540
- Biomedical 0541
- Chemical 0542
- Civil 0543
- Electronics and Electrical 0544
- Heat and Thermodynamics ... 0348
- Hydraulic 0545
- Industrial 0546
- Marine 0547
- Materials Science 0794
- Mechanical 0548
- Metallurgy 0743
- Mining 0551
- Nuclear 0552
- Packaging 0549
- Petroleum 0765
- Sanitary and Municipal 0554
- System Science 0790

Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY
General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

THE CLONING AND CHARACTERISATION OF THE U1,U2,U4 and U6 snRNA's of *Schistosoma mansoni*

BY

TERRY BLAKE BALL

A Thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

Abstract

The maturation of eukaryotic messenger RNA (mRNA) involves a number of processes including alterations of the 5′ and 3′ termini of the pre-mRNA, as well as the removal of intervening sequences or "introns". The mRNA in the trematode *S. mansoni* undergo both "cis" (intramolecular) and "trans" (intermolecular) splicing. This splicing is thought to be catalyzed by a macromolecular splicing complex, the spliceosome. The spliceosome is made up of a number of small nuclear ribonucleoprotein particles (snRNP's). Each snRNP consists of one or more small nuclear RNA's (snRNA) as well as protein cofactors. The snRNA's are believed to be responsible for the catalysis and recognition (by base pairing) of "target" consensus sites present on transcripts to be spliced. There has been no mechanism identified that targets a specific message in *S. mansoni* to be trans-spliced. It is possible that the snRNA's in *S. mansoni* may play a role in the regulation of trans-splicing. Degenerate oligonucleotide probes for the U1, U2, U4 and U6 snRNA's were designed based upon published data. These probes were used to screen a genomic library and perform Northern and Southern blots. Positive clones were isolated, mapped, and subcloned into a sequencing vector. Putative snRNA's were identified for the U2, U4 and U6 snRNA's from these data. The sequence and possible secondary structure of the snRNA's were compared to the snRNA's from other species. The snRNA's of *S. mansoni* shared many features with the snRNA's of other organisms including transcript length, possible genomic organization, and secondary structure. However, areas of divergence were also noted, and the role that these areas of divergence may play in selection of a transcript to be trans-spliced was discussed.

Dedication

First and foremost this thesis is dedicated to wife Monique Ball for her love, support and understanding. Thank You Monique, for standing by my side through the good times as well as the bad. This thesis is also dedicated to my parents, Brian and Leslie Sammons, who have taught me the value of perserverance and hard work.

Monique, Mom, and Brian this Thesis is dedicated to You.

Acknowledgements

I would like to thank everyone who has played a part in getting this thesis completed. There have been many people who with their suggestions, support, and friendship who deserve credit for helping me finish this work.

First, I would like to thank my supervisor Neil Simonson for his advice, support and direction. He has taught me how to think, present data and most importantly question in a clear, concise and rational manner.

I would also like to thank all those who have helped me learn new techniques and skills necessary to complete this thesis, and more importantly, thank them for their support and friendship. Thanks to Terry Langalier, Zeena Mohammed, Leslie Slaney, Ian Maclean, John Rutherford and Sue Ramdahin.

Thanks to my fellow students and colleagues, Keith Fowke, Jody Berry, Trevor Fudyk, Omu Anzala, and Dwayne Stupack for going through all of this with me. Thanks to "The Boys" for showing me that science can be fun and exiting as well as rewarding.

A final thanks to Nigel and Chester who managed to brighten many a bad day.

Table of Contents:

List of Figures

List of Tables

Introduction

The maturation of eukaryotic messenger RNA (mRNA) involves a number of processes including alterations of the 5′ and 3′ termini of the pre-mRNA, as well as the removal of intervening sequences or "introns". The removal of introns is a complex process resulting from the intramolecular or "cis" splicing of introns along a single transcript. The removal of introns is thought to be catalysed by a macromolecular splicing complex, or spliceosome[1]. All of the necessary cofactors involved in the formation, and catalytic activity of the spliceosome have yet to be determined[2].

Trans-Splicing

Until 1985 it was thought that all splicing reactions in eukaryotic cells occurred via a cis splicing mechanism[3]. However, the study of the Variable Surface Glycoprotein genes (VSG's) of *Trypanosoma brucei* revealed that all of the mRNA for these VSG genes contained an identical 39 nucleotide (n.t.) sequence at their 5′ ends[4]. Further characterisation showed that these "leader" sequences were not encoded contiguously with the genomic copies of these genes. It was found that these leader sequences were encoded at the 5′ end of a small non-polyadenylylated RNA (the Spliced Leader RNA or SL-RNA), transcribed from a multicopy, tandemly repeated array removed from the genomic locus of the VSG genes[5]. It was later found that all *T. brucei* mRNA's contained this sequence at their 5′ ends[6]; thus it appeared that *T. brucei* mRNA's are the product of two independently transcribed RNA molecules spliced together by a "trans"

or intermolecular reaction. If trans-splicing is mechanistically similar to cis-splicing then specific predictions about the end products of such a reaction could be made. The predicted end products of a trans-splicing reaction are shown in figure #1. The presence of these predicted products has been shown in every trypanosome species so far examined, suggesting that all trypanosomes carry out trans-splicing reactions[7]. The 39 n.t. spliced leader found at the 5′ end of trypanosome mRNA is conserved in all species so far examined. However, the SL-RNA's are considerably more divergent, although the predicted secondary structure is remarkably conserved[8]. Trypanosomes are also unusual in a number of other respects. They contain no intervening sequences (introns) in their pre-mRNA, thus they do not carry out any form of cis-splicing reactions[9]. Polycistronic pre-mRNA transcripts have also been detected, which is unusual for a eukaryotic organism as a highly modified 5′ cap structure is usually required for the translation of mRNA transcripts. In trypanosomes this cap structure would be absent on any "internal" message in such a polycistronic transcript. Trans-splicing may serve as a mechanism to "mature" and cap the ends of the pre-mRNA regions in these polycistronic transcripts by trans-splicing the capped 39 n.t. SL to the 5′ ends of each distinct coding region maturing the polycistronic transcript into a number of monocistronic, capped mRNA's[10]. Trypanosomes have also been shown to transcribe some pre-mRNA by RNA Polymerase III, which is mainly involved in the transcription of ribosomal RNA and is not usually known to cap its transcripts[11]. Trans-splicing may play a role in maturing the 5′ ends of these transcripts as well. Another unusual RNA processing event in trypanosomes is the finding that some mitochondrial RNAs are not directly encoded within the genome

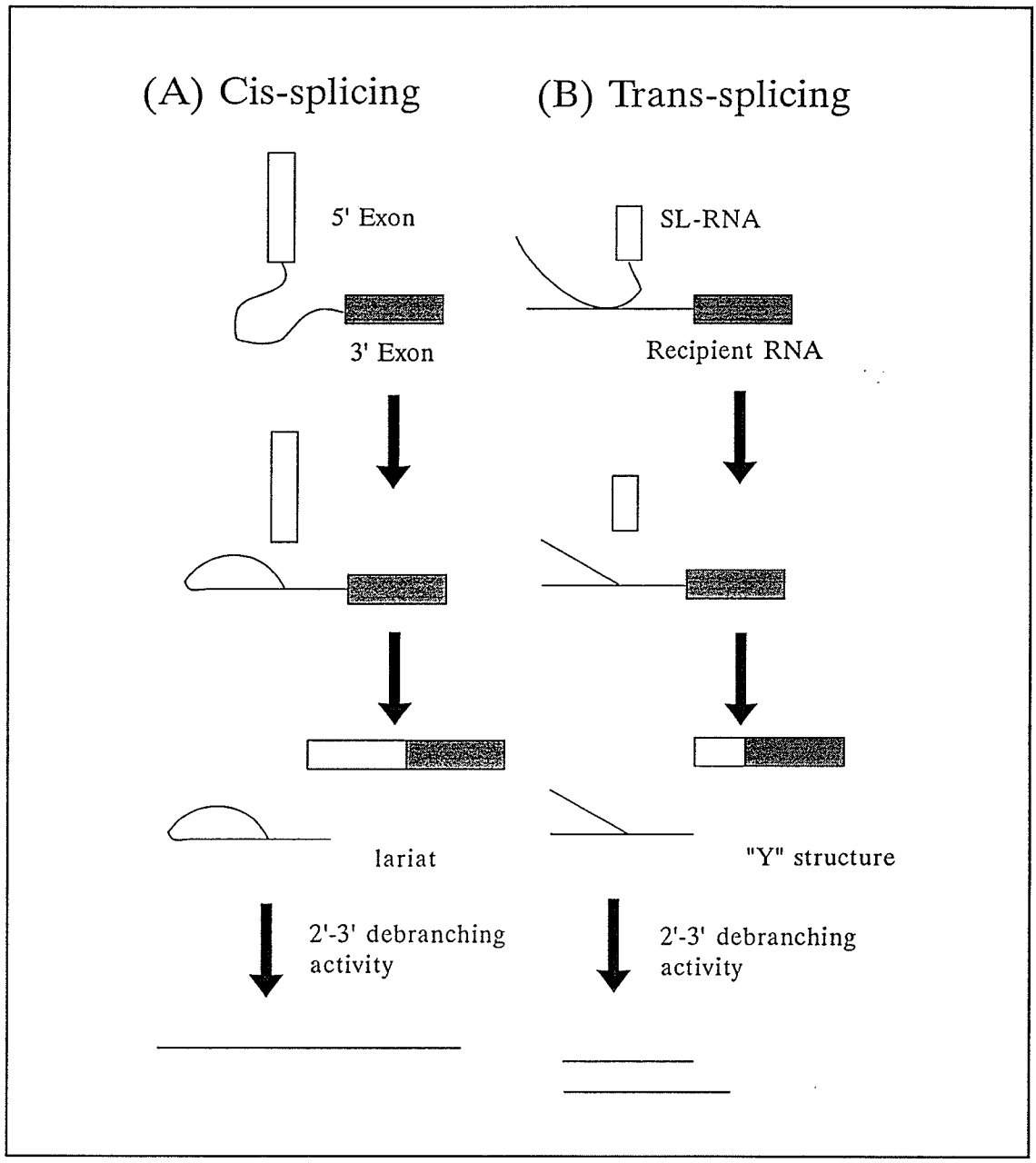


Figure 1: Comparison of End Products of Cis and Trans-Splicing Reactions

In a cis-splicing reaction (A) the intron to be removed forms a characteristic "lariat" structure, which if treated with a de-branching enzyme produces a single product. If trans-splicing is mechanistically similar, (B) then instead of the "lariat" structure one would expect a "Y" shaped structure, that if treated with a debranching enzyme, would produce two intron end products.

instead the encoded pre-mitochondrial RNA is subject to modification (RNA editing) by independently transcribed, non-polyadenylylated RNA's called guide RNA's (gRNA's). The apparent function of the gRNA's are to insert or occasionally delete uridine residues within the coding regions of pre-mRNA transcripts[12].

Trans splicing was subsequently detected in the nematode *Caenorhabditus elegans* during the characterisation of the actin genes. It was found that three of the four actin genes had an identical 22 n.t. sequence at their 5′ end that was not encoded contiguously with the rest of the genes. This 22 n.t. sequence was found to be encoded at the 5′ end of a 100 n.t. non-polyadenylylated transcript originating from a coding region within the tandemly repeated 5S ribosomal RNA array, but transcribed in the opposite direction[13]. Thus it appeared that the 22 n.t. sequence was acting similar to the leader sequence found in trypanosomes, and was trans-spliced to certain actin gene pre-mRNA's. This finding was considerably more interesting than the situation in trypanosomes as *C. elegans* appeared to utilize both cis and trans splicing, in some cases even within the same transcript[14]. Further characterisation confirmed the presence of the predicted trans-splicing intermediates and showed that the nematode SL sequence was present on 10-15% of all *C. elegans* messages[15]. The SL sequence in nematodes appears to be 100% conserved among all the nematodes so far examined[16]. This high degree of conservation (compared to the relatively lower degree of conservation found in the trypanosomes) may be due to its role as an internal promoter element for its own transcription[17]. The SL-RNA, and its genomic organisation have not been conserved, although the predicted secondary structure has been remarkably conserved[18]. In the nematodes *Ascaris*

lumbricoides var sum and *Brugia malayi* the SL gene is linked to the 5S ribosomal repeat like in *C. elegans*, but transcription occurs in the same direction as the 5S rRNA[19], not opposite as is the case for the *C. elegans* SL. Further evidence for the differences in trans-splicing between trypanosomes and nematodes is the identification of a second SL sequence found in *C. elegans*. This second SL-RNA does not appear to be conserved outside the genus, and is not linked to the 5S ribosomal locus[20]. Recent evidence suggests that nematodes may also produce some polycistronic transcripts (similar to trypanosomes) and that these "internal" polycistronic transcripts receive the second SL sequence[21]. The mechanism for this selection and the precise role played by trans-splicing in the maturation of nematode mRNA are still unknown.

The Identification of Trans-Splicing in *Schistosoma mansoni*

During the characterisation of the mRNA's encoding the 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGCoA) genes from *S. mansoni* it was found that the first exon was not encoded contiguously with the other exons in the genome. Southern blot analysis showed that exon I was transcribed from a multicopy repeat unit, while the other exons were single copy. Exon I was shown to be 36 n.t. in length and was encoded by a 90 n.t. non-polyadenylylated transcript. Further examination of other mRNA in *S. mansoni* showed that a small number also contained a spliced leader sequence at their 5′ ends[22]. These features were suggestive of the intermolecular or trans-splicing phenomena identified in trypanosomes and would very closely resemble the situation in nematodes where transcripts can undergo both cis and trans-splicing. It was

therefore apparent that trematodes such as *S. mansoni* may undergo both cis and trans-splicing in a manner similar to nematodes. Unlike nematodes, the trematode SL does not appear to be conserved either within the trematode phylum, or even within the family of schistosomes[23], however the predicted secondary structure of the SL-RNAs are highly conserved. This evidence suggested that trans splicing occurred in trematodes as well as nematodes and trypanosomes, and considerably broadened the range of organisms in which trans-splicing occurs.

The Spliceosome and its Role in Catalysis of Splicing Reactions

Chemically it has been determined that pre-mRNA's are spliced by a two step trans-esterification reaction pathway. The first trans-esterification reaction involves a nucleophilic attack on the 5′ phosphate of the intron by a 2′ hydroxyl group on an adenine residue in what is known as the branch point acceptor sequence. This reaction produces two splicing intermediates, the released 5′ exon, and the 3′ exon and intron in a lariat formation (See figure #2) in which the 5′ end of the intron is covalently linked to the 2′ hydroxyl on an adenine residue upstream of the splice acceptor site within an imperfect consensus sequence of (PyNCUGAC). The second transesterification reaction involves an attack by the 3′ hydroxyl of the released exon on the 5′ phosphate of the 3′ exon. This joins the two exons, and releases the intron in a characteristic lariat form that, if treated by a debranching enzyme, produces a single product[24].

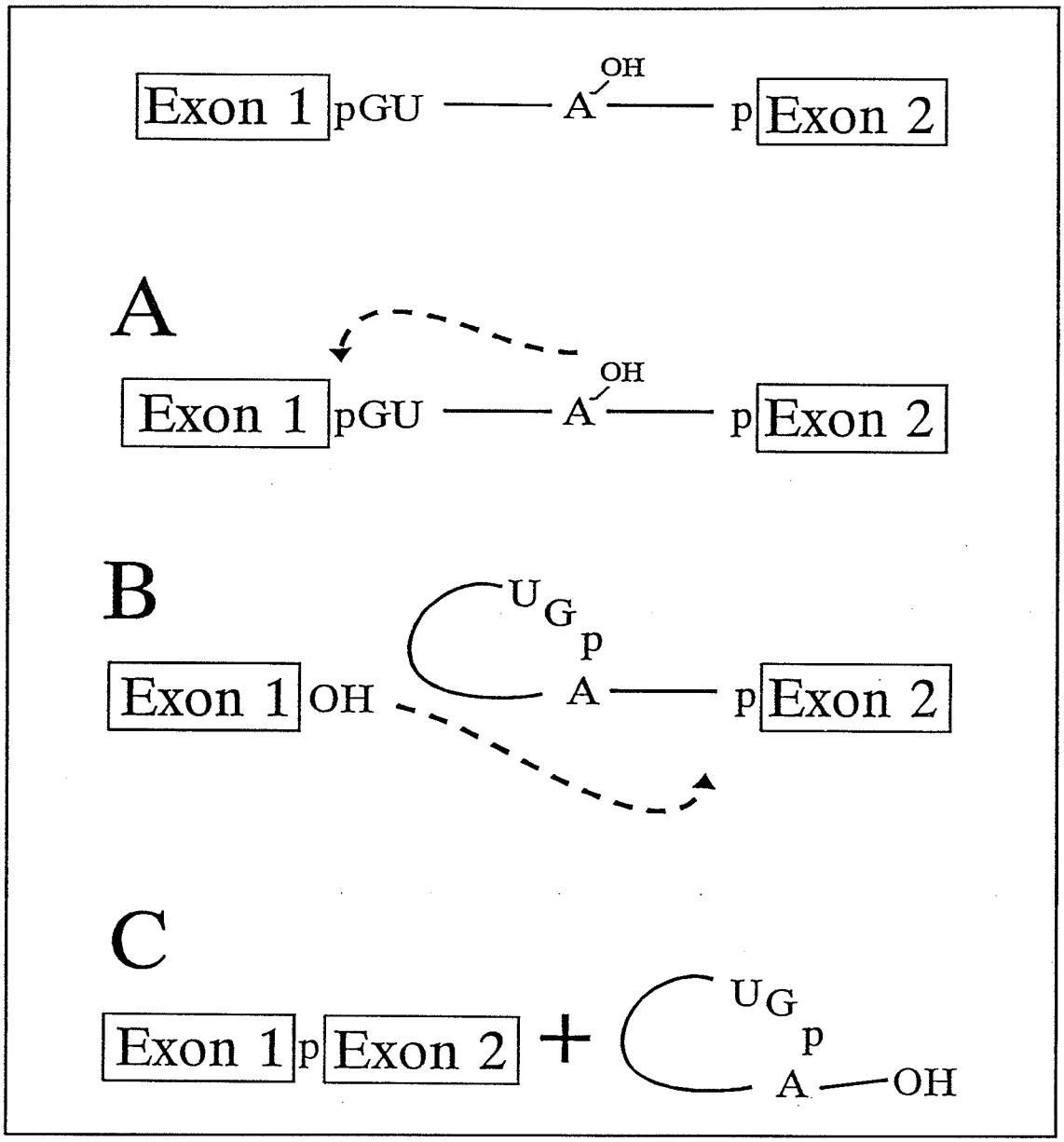


Figure 2: Chemical Intermediates In The Cis-Splicing Reaction

Splicing occurs in a two step trans-esterification reaction; first (A) the 5′ end of the intron to be removed is brought into close association with a 2′ hydroxyl group on an adenine residue in the branch point accepter sequence. The hydroxyl group "attacks" the 5′ phosphate causing the first trans-esterification reaction that releases the 3′ hydroxyl group of the first exon, leaving the intron in a characteristic "lariat" structure. Next (B) the free 3′ hydroxyl group is brought into association with 5′ phosphate of exon #2, the second trans-esterification reaction occurs, joining the two exons and releasing the intron (C).

Small Nuclear Ribonucleoproteins (snRNP's) and Their Role in Splicing

The splicing reaction in eukaryotic organisms is believed to be catalysed by a large macromolecular complex (the spliceosome) which consists of a number of small nuclear ribonucleoproteins (snRNP's) and a number of accessory proteins. The snRNP's that form the intact spliceosome consist of one or more small nuclear RNA's (snRNA's), a group of "core" proteins common to all of the snRNP particles, as well as snRNP specific protein factors (as an example the mammalian snRNP's involved in splicing are summarised in Table#1). The core group of proteins appear to be mainly structural in function while the snRNP specific proteins have been suggested to have a number of properties such as RNA binding, RNA helicase, and protein-protein interaction properties. At least four snRNP particles have been identified as being necessary for the formation of the intact spliceosome in most eukaryotic organisms. Each particle is designated according to its constituent snRNA (ie. U1 snRNP is named according to the presence of the U1 snRNA), these particles are U1 snRNP, U2 snRNP, U5 snRNP and U4/6 snRNP. There are a number of features common to the snRNA's present in the snRNP particles. Each of the snRNA's involved, with the exception of U6 (which is found complexed with U4 snRNA in the U4/U6 snRNP) contains a binding site for a common set of snRNP structural proteins (The "Core" or Sm proteins). This site is referred to as the Sm binding site and is highly conserved among the U snRNA's; the consensus being $PuAU_{3-4}NUGPu$. The core proteins that recognise this site are precipitable by the human, autoimmune, anti-Sm class of antibodies, first identified in patients with systemic lupus erythematosus[25]. The second feature common to each

snRNP	snRNA	Proteins	Notes
U1	U1	"core" 3 other snRNP specific	Sequence complementary to 5′ splice site
U2	U2	"core" 2 other snRNP specific	Sequence complementary to BPA site
U5	U5	"core" 2 other snRNP specific	Necessary for active spliceosome
U4/U6	U4	"core" 3 other snRNP specific	Negative regulator of U6 snRNA function, binds U6 snRNA
	U6	3 other snRNP specific	Active component of spliceosome

Table #1: Mammalian Small Nuclear Ribonucleotide Particles (snRNP's) Involved in Formation of an Intact Spliceosome

snRNA (except U6) is the presence of a highly modified (2,2,7) trimethylguanosine cap structure at their 5′ ends; important in completing a nuclear localisation signal necessary for the transport of the snRNP particle back into the nucleus after assembly in the cytoplasm[26]. Like pre-mRNA's, the snRNA's are transcribed by RNA polymerase II, with the exception of the U6 snRNA which is transcribed by RNA polymerase III[27].

The spliceosome is believed to function as a framework around which the splicing reaction takes place in a manner analogous to the group II self-splicing introns carried out by mitochondrial RNA[28]. The snRNA's are believed to be the active catalytic subunits that carry out the trans-esterification reactions based upon sequence recognition of conserved sequences on the target transcript to be spliced and the alignment of specific snRNA's. The snRNA's that form the spliceosome can be functionally viewed as a fragmented group II intron that are brought together by the spliceosomal components to catalyze the splicing reaction. The majority of pre-mRNA splicing substrates have a common set of conserved sequence elements necessary for the splicing reaction to take place. These include; a 5′ splice donor site, a 3′ splice accepter site, and a branch point recognition sequence flanked by a polypyrimidine tract (See figure #3). The snRNA's are believed to mediate the recognition and assembly of the intact spliceosome around these points.

Each snRNP particle is believed to play a specific role in the stepwise formation of the intact spliceosomal complex (See figure #3). The first step in the formation of the spliceosome is the binding of the U1 snRNP to the 5′ splice donor site of the pre-mRNA,

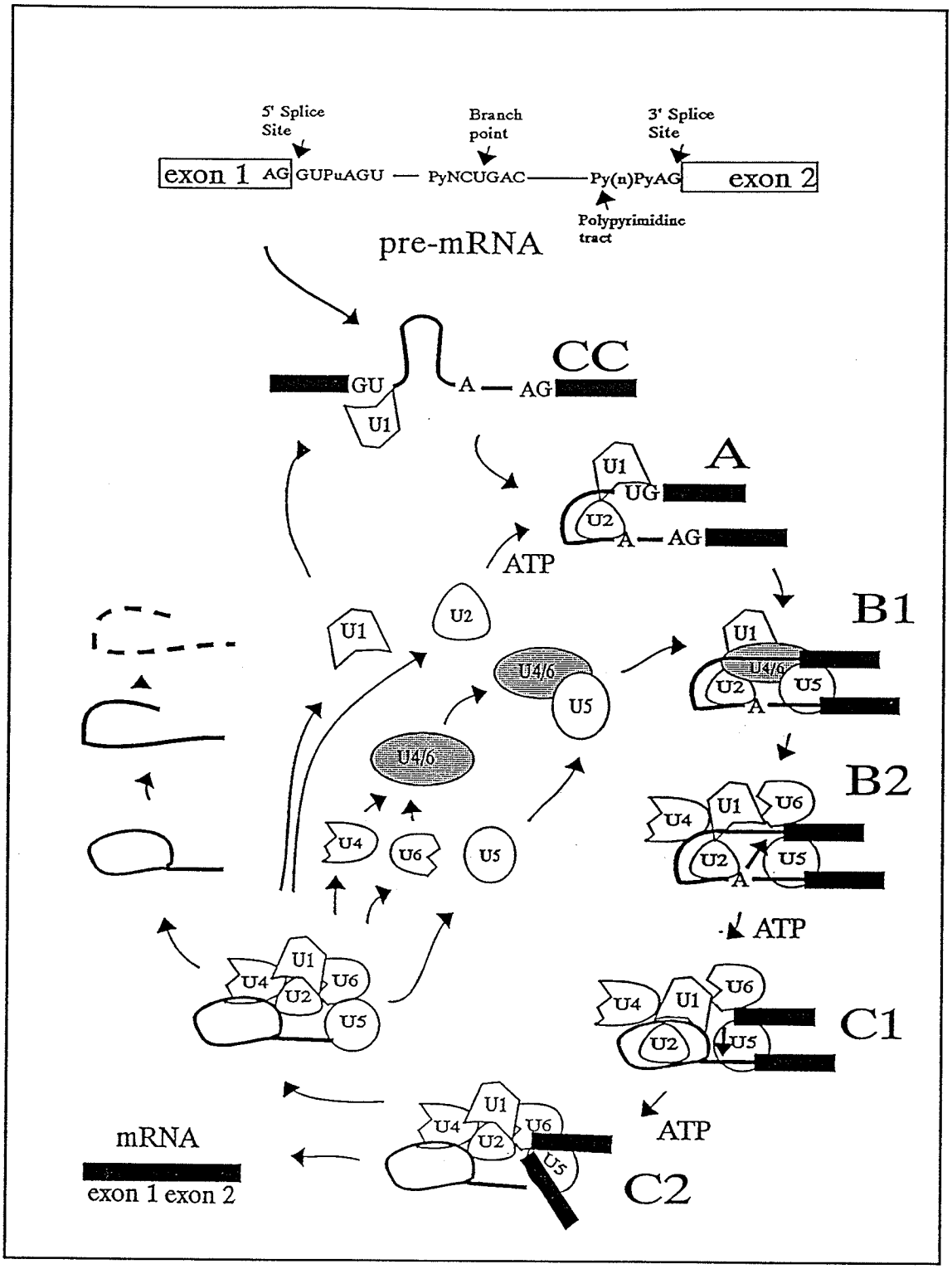


Figure 3: Model For the Involvement of snRNP's in the Catalysis of a Cis-Splicing Reaction

Figure 3: Model for the Involvement of snRNP's in the Catalysis of a Cis-Splicing Reaction

In a cis-splicing reaction, the first event is the recognition of the 5′ splice site by the U1 snRNA forming the "commitment complex" (CC). The U2 snRNA then recognises the branch point region (A), which is followed by the incorporation of the U4/5/6 complex (B1). Shortly after the formation of the intact spliceosome the U4 snRNA dissassociates from the U6 snRNA which catalyzes the two trans-esterification reactions (B1+C1) that join the two exons, releasing the snRNP's and the lariat-intron, which is then degraded. The snRNP's are recycled to catalise another splicing reaction.

this binding occurs in the absence of other snRNP particles or ATP hydrolysis[29]. This binding is reversible, and binding of the U1 snRNP to the pre-mRNA forms what is termed the "commitment complex" or CC. Based on sequence comparisons it was predicted that the U1 snRNA binds to the 5′ splice site by sequence complementarity[30]. The necessity of base pairing between the U1 snRNA and the 5′ splice site was confirmed by mutation studies in both mammalian and yeast systems[31]. The selection of a particular splice site has been shown to be a complex process, not only involving the U1 snRNA, but involving a number of different factors. A large number of cell specific and non cell specific protein cofactors have been identified as being important in 5′ splice site selection[32]. Also important in splice site selection is the context or relative positioning of elements of the pre-mRNA itself, the selection of alternatively spliced products has been shown to be dependent on intramolecular base pairing of pre-mRNA regions surrounding the 5′ and 3′ splice sites[33]. Of all the factors involved in 5′ splice site recognition, the importance of the U1 snRNA remains a constant. Recent reports have strengthened the argument for the role of the U1 snRNA by showing the importance of sequence complementarity between the 5′ splice donor site and the U1 snRNA. It was shown that mutations that increase the base pairing between the U1 snRNA and the 5′ splice site in the yeast *S. cerevisiae* can bypass the need for cell specific accessory proteins needed for proper splicing of certain mRNA's[34]. This evidence suggests that the U1 snRNA is one of the main recognition signals responsible for 5′ splice site selection. Other factors, such as accessory proteins or sequence context may play a role by influencing the efficiency of this recognition. The role of the U1 snRNA is therefore

believed to insure the proper recognition of the 5′ splice site, and thereby facilitate the positioning of the rest of the spliceosomal complex components.

Structurally the U1 snRNA has been most closely studied in mammalian systems. The snRNA is generally around 164 n.t. in length and contains four stem loop structures, stem-loops I, II and III are 5′ of the Sm site, while stem-loop IV flanks then Sm site at the 3′ end (for a model please see figure #4)[35]. It appears that only the 5′ end of the U1 snRNA that interacts with the 5′ splice donor site has been identified as absolutely necessary for the formation of an intact spliceosome. The only other region identified as important is Stem-loop II which has been identified as important in interaction of the U1 snRNA with the so called core (or Sm proteins) or snRNP specific protein factors[36].

The next step in the formation of the intact spliceosome is the binding of the U2 snRNP to the branch point acceptor sequence (BPA) of the pre-mRNA. This step is ATP dependent, and irreversibly commits the pre-mRNA to the splicing pathway[37]. The binding of the U1 snRNP and subsequent binding of the U2 snRNP forms the pre-spliceosomal complex. In this step, base pairing between the U2 snRNA and the branch point acceptor sequence of the pre-mRNA is necessary for assembly of the spliceosome[38]. Recognition of the BPA sequence in the pre-mRNA by the U2 snRNP has been shown to be dependent on a number of factors, including one or more accessory proteins[39], and possibly the presence of the U1 snRNA[40]. The sequence context of the branch point acceptor and the adjacent polypyrimidine regions may also play an important role in recognition. One hypothesis for the necessity of the U1 snRNP being present before the binding of the U2 snRNP is that this may serve as a mechanism for

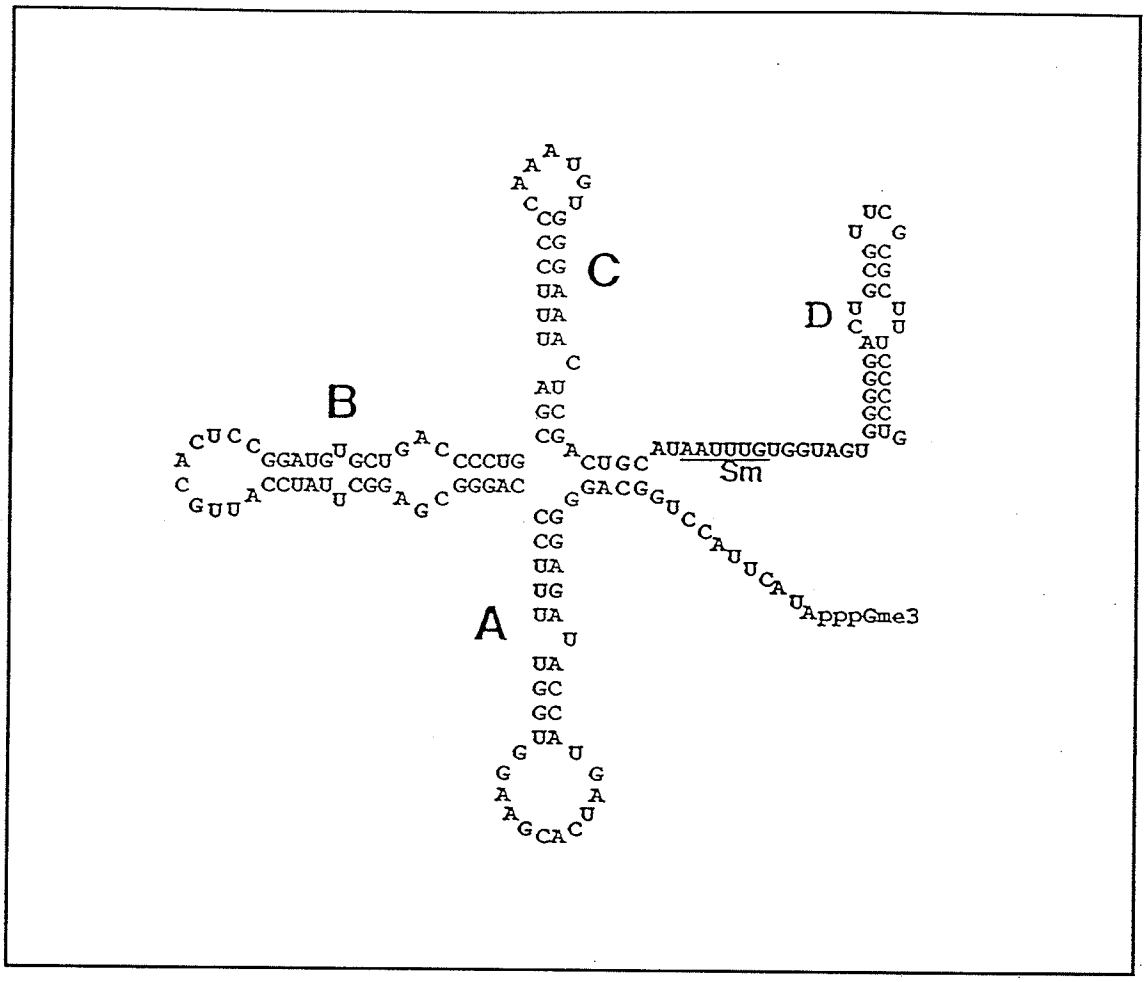


Figure 4: Proposed Secondary Structure Model of The Human U1 snRNA

the juxtaposition of the 5′ and 3′ splice sites. The precise roles played by either the U1 or U2 snRNP in the positioning of the 5′ and 3′ splice sites has been investigated in a number of recent reports. Mutational studies of the yeast *Schizosaccharomyces pombe* suggests that the U1 snRNA base pairs with both the 5′ and 3′ splice sites of the intron, and that this interaction is necessary for splicing[41]. However in the yeast *Saccharomyces cerevisiae* no important interactions between the U1 snRNA and the 3′ splice site were found[42]. The full range of interactions between the U1 and U2 snRNP's and their association with the 5′ and 3′ splice sites has yet to be fully delineated. In any case, the alignment of the two splice sites is likely to be a complex process possibly involving other protein or snRNP components. If this alignment is indeed mediated by the U1 and U2 snRNP's, as the data suggests for the two yeast species, then this suggests that there may be a number of alternate mechanisms for the alignment of these splice sites.

The U2 snRNA has been well characterised in mammalian systems. The U2 snRNA is generally around 188 n.t. and contains four stem loop structures, stem-loops I and II are 5′ of the Sm site, while stem-loops III and IV flank the Sm site at the 3′ end (for a model please see figure #5)[43]. Besides the Sm site necessary for the attachment of the "core" structural proteins, a number of regions have been identified as important for the U2 snRNA's function. These include: The 5′ end of the U2 snRNA that interacts with the branch point acceptor region, Stem-loop IV which has been identified as interacting with U2 snRNP specific proteins[44] and Stem-loop II which has been identified as important in assembly of the complete spliceosomal complex[45].

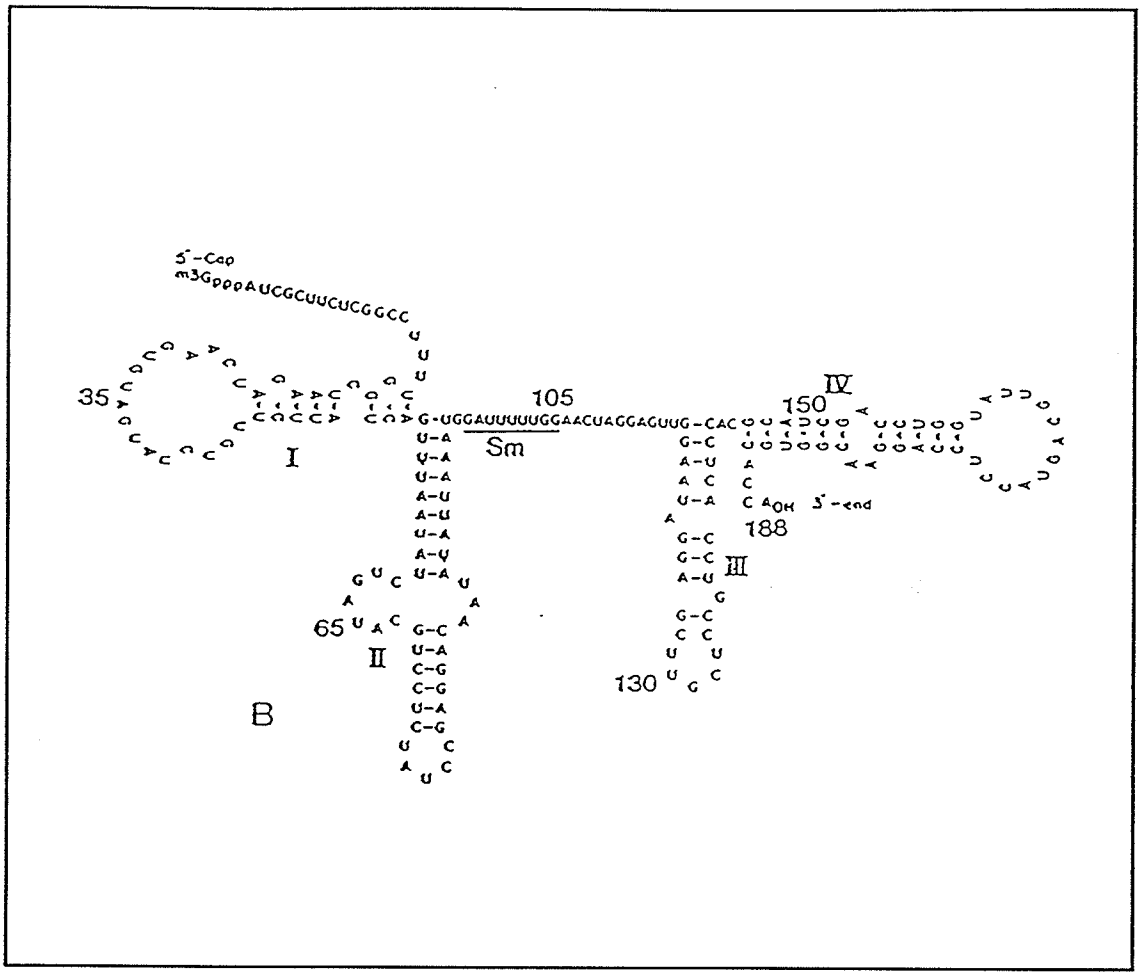


Figure 5: Proposed Secondary Structure Model for Rat U2 snRNA

After the binding of the U2 snRNP to the BPA region, the next step in the formation of an intact spliceosome is the incorporation of a multi-snRNP structure made up of the U5 snRNP and the U4/U6 snRNP. The incorporation of this multi-snRNP structure forms the intact spliceosomal complex which catalyzes the removal of the intervening intron. The spliceosome does not remain intact for long, soon after the binding of the U4/U5/U6 complex there is a loss of the U4 snRNA prior to, or simultaneously with the cleavage of the 5′ splice site and formation of the intron-lariat[46]. Initial reports suggested that the U4/U5/U6 complex contacts only the U1 and/or the U2 snRNP's and not the pre-mRNA[47]. However recent investigation has identified a number of potential interactions, among the snRNA's themselves and including interactions with the pre-mRNA. These RNA-RNA interactions play critical roles in the accuracy and efficiency of the splicing process. Not all of these interactions are permanent, or even established at the same time. Rather these interactions may be established, modified, disrupted and displaced a number of times during the splicing process. The recently proposed model for the U5 snRNP's actions during the splicing process gives an example of such RNA-RNA interactions. For a number of years no clearly defined role for the U5 snRNP had been identified. Recent evidence has shown that the U5 snRNA is involved in a number of base-pairing interactions mainly involving the 5′ and 3′ splice sites of the pre-mRNA[48]. This finding suggests a role for the U5 snRNA during the splicing process; likely as a factor involved in the juxtaposition of the 5′ and 3′ splice sites. One problem with these findings was that the U5 snRNA was shown to interact with a region of the 5′ splice site previously identified as being

involved in base pairing with the U1 snRNA. Recent reports have shown that the base pairing of the U1 snRNA with this region becomes disrupted sometime during the splicing process[49]. This disruption suggests a mechanism for the interaction of the 5′ splice site with both the U1 and U5 snRNA's during the splicing process. The interaction of the U5 snRNA at the 5′ splice site has been shown by cross-linking studies to be maintained prior to, and throughout the splicing reaction while the interaction with the 3′ splice acceptor site only occurs during the formation of the intron-lariat[50]. It is still unknown what precise role the U5 snRNA plays in the splicing process, but as the above evidence suggests, the U5 snRNA may be responsible for the alignment of the 5′ and 3′ splice sites. The U5 snRNA along with other snRNA's may also play a role in splice site selection. Mutational analyses of conserved U5 sequences have shown that mutations in these sequences can result in the use of cryptic 5′ splice donor sites[51]. Further characterisation of the structure and role of the U5 snRNA/snRNP will provide further information about its possible role in the splicing process.

The potential roles of the U4/U6 snRNP in the splicing reaction have been fairly well established. It has long been hypothesized that the U6 snRNA is highly involved in the formation of the active catalytic core of the spliceosome. The U4 snRNA has been thought to act as a negative regulator of the U6 snRNA's activity[52] and play a role in the assembly of the U4/U5/U6 multi-snRNP particle with the rest of the spliceosome[53]. The role of the U4 snRNA as a negative regulator has been supported by studies showing that the U4 apparently is lost, or disassociates from the intact spliceosome, just prior to the first splicing step[54]. The role of U6 as a main part of the catalytic element is

supported by a number of findings. The first finding is that the U6 snRNA is the most highly conserved of all of the eukaryotic snRNA's[55]. Further evidence comes from mutational studies showing that single point mutations of the U6 snRNA can inhibit splicing or even alter splice site specificity[56]. The intimate involvement of the U6 snRNA in catalysing the splicing reaction has been demonstrated by studies of reconstituted splicing systems in trypanosomes showing that mutations of the U6 snRNA can induce some startling changes in the splicing reaction. These mutations led to the inappropriate positioning of the 5′ splice site in relation to the U6 snRNA and resulted in the use of the U6 snRNA as a surrogate splicing substrate[57]. Thus, interference in the normal function of the U6 snRNA changed its role from a catalytic element to its use as a substrate for the splicing reaction. This finding provides a strong argument for the importance of the U6 snRNA in the splicing reaction. Additional evidence of the U6 snRNA's involvement comes from the identification of interactions between the U6 snRNA and the 5′ splice site by cross-linking studies[58]. More speculative evidence comes from the studies showing that the yeast *S. pombe* U6 snRNA contains an intron[59]. One interpretation of this finding is that an abortive or "reverse" splicing reaction caused an intron to be spliced into the U6 snRNA. This would be followed by some sort of reverse transcription event, and re-insertion of this altered U6 element by homologous recombination back into the genome[60].

Recently a similar mRNA-type intron was found in the U2 snRNA gene from the yeast *Rhodotorula hasegawae*. This suggests that the reverse splicing explanation for the U6 snRNA may be wrong, and that there may be some other explanation for the presence

of introns in these two snRNA genes. One possible explanation is that these genes contained introns at some ancestral level, and these introns were removed over time in all but a few cases. There are other explanations, but one of the most intriguing is that the *R. hasegawae* U2 snRNA also obtained its intron through a reverse splicing event. If this explanation is true, then one would expect to find the U2 snRNA in close proximity to the U6 snRNA, the 5′ splice site and the catalytic centre of the spliceosome[61]. A novel base pairing model has been proposed that would bring both the U2 and U6 snRNA's into close association with the 5′ splice site. Other studies confirmed that the U2 snRNA and the U6 snRNA do indeed interact via base pairing[62], and it has been proposed that this interaction may trigger a series of conformational rearrangements necessary for the catalytic activity of the spliceosome[63].

A further speculative point that may suggest an active role by the U6 snRNA comes from mutational analysis indicating that the 3′ end of the snRNA is absolutely necessary for splicing activity[64]. Recent findings have shown that the U6 snRNA population in a cell undergoes modifications of its 3′ termini, which no other snRNA appears to undergo. This modification consists of the addition of uridine monophosphate groups and the formation of 2′, 3′ cyclic phosphate groups at the 3′ ends of U6 snRNA's[65]. The role of these additions and modifications is unknown, they may play a role in the assembly or recognition of the U6 snRNA by the U4/U6 snRNP components, or they may play some sort of role in the splicing activity of the spliceosome. It is thought that these modifications may play more of a role in the splicing activity of U6 snRNA as the modifications have been shown to occur mainly in the intact

spliceosome. However, other explanations for these modifications may exist.

The U4 and U6 snRNA's have been best characterised in mammalian systems. The U4 snRNA was found to have a size of 142 n.t. while the U6 snRNA has been found to be about 107 n.t. long. The U4 snRNA has three stem loop structures all near the 3′ end of the molecule. Stem loop I and II are 5′ of the Sm site while stem loop III lies just 3′ of this site. The U6 snRNA, which as mentioned lacks an Sm site, contains 3 stem loops (for a model please see figure #6)[66].

Physically the U4 and U6 snRNA's are found in an extensive base paired arrangement. Besides these well characterised intermolecular U4/U6 base pairing regions a number of regions specific to each snRNA have been shown to be important in their function. In addition to the Sm binding region the U4 snRNA has a number of regions identified as important by mutational analysis. One region identified is the loop sequence of stem-loop I, that has been shown to be important in interactions with U4 snRNP specific protein factors necessary for spliceosome stability. Another non-base paired (folded) region that has been identified as important in splicing activity is Stem loop II and the regions surrounding it[67]. The U6 snRNA has also had a number of regions identified as important by mutational studies. These studies have shown that Stem-loops II and III appear to be important in assembly of a functional U4/U5/U6 snRNP[68]. It is not surprising that these structural regions have been identified as important considering the number of secondary structure changes, and interactions that the U6 snRNA appears to go through as a main catalytic element.

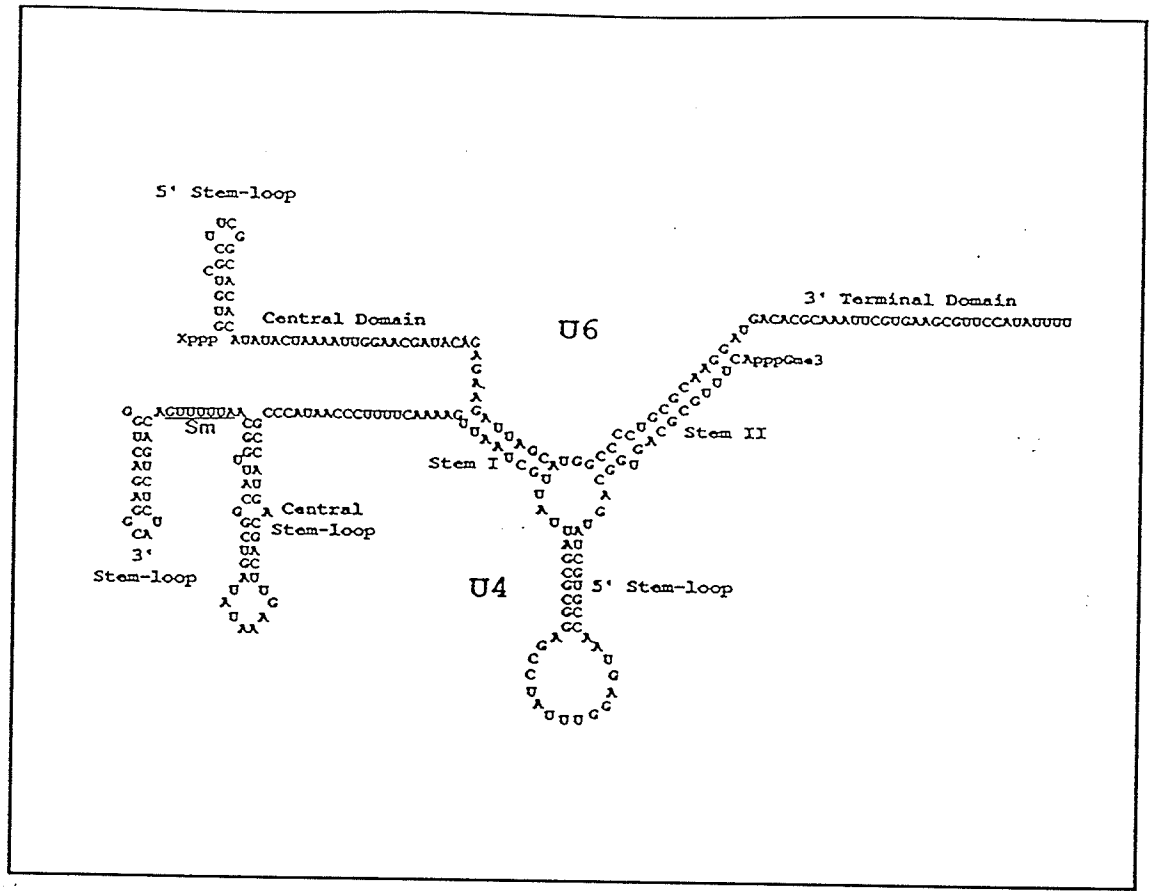


Figure 6: Proposed Secondary Structure Model for Human U4 and U6 snRNA's

A simplified outline of the role of each snRNA during the splicing reaction would be as follows: U1 snRNA recognises sequences at the 5′ and possibly the 3′ end of the intron and forms the commitment complex. U2 snRNA recognises the branch point region and a stable ATP-dependent complex is formed. The U4/U5/U6 tri-snRNP enters with an initial involvement of the U5 snRNA with the 5′ and 3′ exonic sequences followed by a series of conformational changes; U1's binding to the pre mRNA is destabilised; U5's binding changes from the 5′ exon to the 5′ end of the intron; the U4/U6 base pairing is destabilised; and a U2/U6/5′ splice site complex forms. Splicing then begins, and the two-step trans-esterification reaction occurs rapidly, possibly with further conformational changes. This results in the joining of the 5′ and 3′ exons and the release of the intron in a characteristic lariat structure. For a model of such a reaction, again, please see figure #3

Trans-Splicing is Mechanistically Similar to Cis-Splicing Reactions.

The characterisation of the cis-splicing pathway in mammalian and yeast systems has provided a well defined model of the splicing reaction and some of the protein and snRNA factors that help catalyze this reaction. The presence of a number of conserved elements in both the SL-RNA and the splice acceptor RNA, as well as the similarities of cis and trans splicing substrates have suggested that trans-splicing is mechanistically similar to cis-splicing. All organisms that carry out trans-splicing reactions have sequences similar to conserved regions identified as important in cis-splicing. These regions include the 5′ splice donor site (present on the SL-RNA instead of the 5′ exon),

and the branch point acceptor sequence and the 3′ splice acceptor site on the target pre-mRNA[69]. The examination of the predicted secondary structures of SL-RNA's from trematodes has shown a remarkable similarity between these SL-RNA's and the SL-RNA's of trypanosomes. Each SL-RNA contains three stem-loops with the 5′ splice site adjacent to the 5′ most loop. This secondary structure has been conserved in all of the trypanosomes and nematodes examined, and is reminiscent of the structure of group II self splicing introns[70]. The finding that nematode SL-RNA's contain both a 2,2,7-trimethylguanosine cap and an Sm binding site consensus sequence (both present on U snRNA's) suggest that the nematode SL-RNA may be acting as an snRNP during the trans-splicing process[71]. Indeed it was found that the SL-RNA in *C. elegans* does exist as a small nuclear ribonucleoprotein (SL-RNP) and acts analogous to a snRNP in the trans-splicing reaction[72]. The SL-RNA may participate in the splicing reaction in a manner similar to the U1 snRNA by aligning the SL-RNP particle in the spliceosome and obviating the need for the U1 snRNP in a trans-splicing reaction by autoactivating the 5′ splice site in a manner similar to group II self splicing introns. If this model for the role of a SL-RNA is true then an apparent problem is the absence of a TMG cap and Sm site on the SL-RNA of trypanosomes. This argument was countered by the finding that SL-RNA can be found in the form of an snRNP in *Trypanisoma brucei*[73]. Closer examination of the trypanosome SL-RNA does reveal a modified 7-methylguanosine cap structure[74], and the presence of a degenerate Sm site conserved between the trypanosome SL-RNA and any trypanosome U snRNA's. The finding of analogs of the U2, U4 and U6 snRNA's in *T. brucei* supports the similarity in mechanisms between cis

and trans-splicing[75]. More interesting however, was the finding that trans-splicing can be abolished by the destruction of the U2, U4 and U6 snRNA's in permeabilised trypanosomes[76]. The proposed model for trans-splicing therefore has the SL-RNA acting in two roles; firstly the SL-RNA acts similarly to the U1 snRNA, aligning the substrates, and possibly autoactivating the 5′ splice site, and secondly the SL-RNA is consumed as a substrate in the reaction (see figure #7). The SL-RNA could therefore be viewed as a fusion of an exon with a small nuclear RNA. In agreement with this model is the apparent absence of U1 snRNA's in any trypanosome species so far examined; since trypanosomes only carry out trans-splicing there would be no need for a U1 snRNA.

One argument against the trans-splicing model being similar to the cis-splicing model is the apparent absence of a U5 snRNA in trypanosomes. However, a small RNA has been recently identified that interacts with the 5′ splice site on trypanosome SL-RNA's in a manner similar to the interaction between U5 snRNA's and the 5′ splice sites in other organisms. This small RNA has a secondary structure strongly resembling structural motifs identified as being important in the interaction of U5 snRNA's with the 5′ splice site[77]. The identification of this potential U5 analog would complete the snRNA requirements in trypanosomes if trans-splicing functions similarly to cis-splicing with the SL-RNA replacing the U1 snRNA.

Characterisation of the snRNA requirements in nematodes have strongly supported the model proposed; substrates that undergo cis splicing were found to require all 5 snRNA's (U1,U2,U5,U4/U6), while substrates that undergo trans-splicing were shown not to require the U1 snRNA[78]. The model of the SL RNA acting independently of the

U1 snRNA has been further supported by the findings that constructs containing SL-RNA sequences linked to an exon can be efficiently spliced in the absence of the U1 snRNA[79]. Thus the presence of the SL sequences on the constructs obviated the need for a U1 snRNA or snRNP particle. If the SL-RNA behaves as an exon linked to a snRNA then the SL-RNA would be expected to have base pairing interactions with spliceosomal components similar to the interactions shown by both the U1 snRNA, and the 5′ splice site. One example of these interactions is an experiment showing that a region of the SL-RNA of trypanosomes does indeed interact with the U6 snRNA (similar in the interaction between the 5′ splice site and the U6 snRNA in cis-splicing)[80]. One final observation that supports the similarity in function between the cis and trans splicing systems is the finding that the introduction of a SL-RNA transcript into a mammalian cell line results in the production of trans-spliced products[81]. This demonstrates the functional conservation that must exist in the splicing machinery between cis and trans-splicing systems. For a model of this proposed activity please see figure #7.

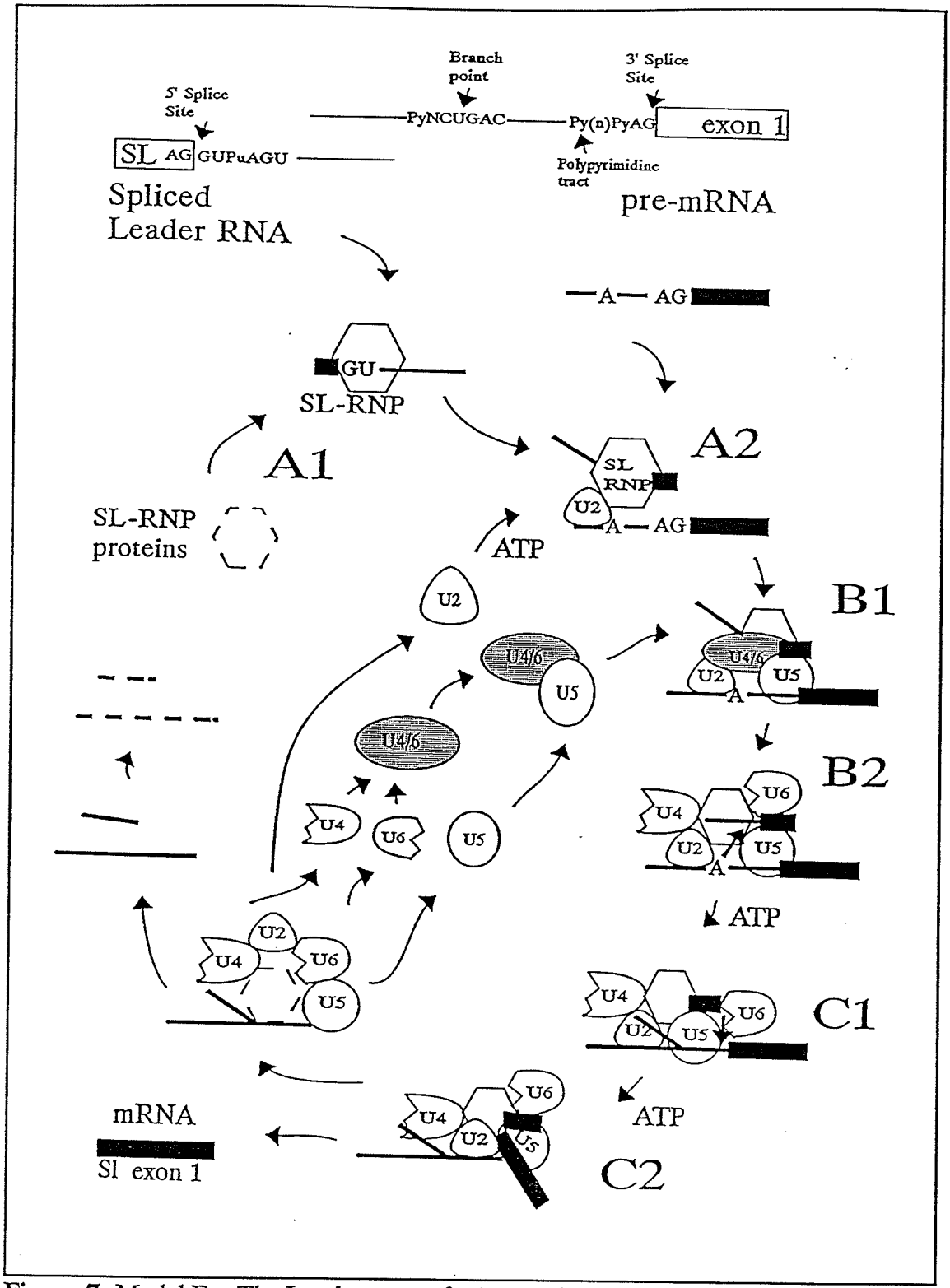


Figure 7: Model For The Involvement of snRNP's in the Catalysis of a Trans-Splicing Reaction

Figure 7: Model For The Involvement of snRNP's in the Catalysis of a Trans-Splicing Reaction

In a trans-splicing reaction, the SL-RNA acts as the U1 snRNA during the formation of the spliceosome by entering the reaction as a SLRNP (A1). The U2 snRNA acts similar to the cis-splicing system by recognition of the branch point region (A2), which is followed by the incorporation of the U4/6 complex and possibly the U5 snRNP, or a U5 like structure (B1). Shortly after the formation of the intact spliceosome the U4 snRNA dissassociates from the U6 snRNA which catalyzes the two trans-esterification reactions (B1+C1). These reactions join the SL and the target exon, releasing the snRNP's and the lariat-intron, which is then degraded. The snRNP's are recycled to catalyze another splicing reaction with the exception of the SL-RNP which is consumed in the reaction.

Control and Regulation of Trans-splicing

In all organisms that undergo trans-splicing there have yet to be any sequence requirements or control elements identified as being responsible for regulating this process. Trypanosomes would not appear to need a selection or control mechanism to target pre-mRNA transcripts to be trans-spliced since all trypanosome messages are trans-spliced. Trypanosomes appear to use trans-splicing as a mechanism to cap and mature the 5′ ends of their pre-mRNA for subsequent translation. However, the regulation of the temporal order of trans-splicing appears to be tightly regulated with other RNA processing events. It has been reported that trans-splicing and polyadenylylation of pre-mRNA is an ordered, tightly coupled process and interference with trans-splicing inhibits polyadenylylation[82]. While there is no apparent need for a control mechanism to determine which transcripts are trans-spliced there would appear to be a need to regulate the timing of trans-splicing. Specific sequence elements, or other factors that may regulate this process have yet to be identified in this organism.

The nematodes are considerably more interesting as they contain pre-mRNA that becomes both cis and trans-spliced, sometimes within the same transcript. Therefore, nematodes would appear to need some sort of mechanism to target what transcripts are to become trans-spliced. Nematodes are also interesting in that one species, *C. elegans*, contains two SL sequences (SL-1 and SL-2). The mechanism that selects for cis or trans-splicing, or even selection between the two SL's is still not completely understood. Studies have shown that the trans-splicing system may be a "default" mechanism where a pre-mRNA splice acceptor site is targeted for trans-splicing if it lacks a 5′ splice donor

upstream of its location. This would result in any transcript without an intron at its 5′ end (an "outron") being targeted for trans-splicing[83]. This finding was supported by an experiment showing that the insertion of an exon and 5′ splice donor site to such an "outron" converted it from a trans-spliced to a cis-spliced gene[84]. The question of what mechanism selects between the SL-1 and SL-2 spliced leader sequences in *C. elegans* may be answered by reports indicating some transcripts in *C. elegans* may be polycistronic. The report suggests that SL-2 is trans-spliced to the downstream gene of two co-transcribed genes orientated in the same direction (or the second pre-mRNA in a polycistronic transcript)[85]. Thus any polycistronic transcript would receive SL-2 by trans-splicing, while any transcript without a 5′ splice donor would receive a SL-1 spliced leader. The molecular mechanism of recognition and selection of either of these two genomic arrangements still has not been elucidated.

Whatever mechanisms have been proposed for the regulation and control of trans-splicing in trypanosomes and nematodes, there is evidence that regulation of trans-splicing in trematodes (ie. *S. mansoni)* may be different. The first evidence comes from a comparison of the SL-RNA's of schistosomes, trypanosomes and nematodes. SL-RNA's usually contain the consensus Sm binding site necessary for binding of the snRNP "core" or Sm proteins flanked by two stem-loop structures. This arrangement is highly conserved among the SL-RNA's so far studied. However the schistosome SL-RNA is structurally quite different as it is missing one stem loop structure[86]. This indicates the possibility of a modified group of core proteins that would bind to this region and may have an effect on the regulation of splicing. The second piece of evidence comes from

the identification of alternative trans-splicing in certain schistosome mRNA's[87]. That is, the SL-RNA is trans-spliced to either the 5′ most exon, or trans-spliced to an interior, downstream exon. This which would invalidate the nematode "default" mechanism for schistosomes as the model does not account for splicing to an "interior" or downstream exon (ie. an exon that contains a 5′ splice donor site upstream). The nematode model could still hold true if this *S. mansoni* gene contained a transcription start signal within the first exon, thus producing two "outron" transcripts. However, there has been no evidence to suggest the possibility of alternate transcription start sites in any trematode species. Thus it would appear that the regulation and control of trans-splicing in schistosomes is different than models proposed for the trypanosomes or the nematodes. The absence of clear consensus sequence elements that would identify a transcript as a target for trans-splicing suggest that the control and regulation of trans-splicing in schistosomes may occur at the level of essential cofactors such as the snRNA's or SL-RNA's that form the intact spliceosome. Therefore characterisation of these elements may lead to an understanding of a possible mechanism for the control and regulation of trans-splicing. Of all the factors possibly involved, the snRNA's involvement in recognition and formation of an intact spliceosome is absolute. Therefore in order to understand the possible mechanism for the regulation of trans-splicing in schistosomes it would first be necessary to characterise the snRNA's involved. The goal of this research was to clone and characterize both by sequence analysis and determination of secondary structure the U1, U2, U4 and U6 snRNA's of *S. mansoni* in order to provide an understanding of possible mechanisms for the control and regulation of trans-splicing.

Materials and Methods

Common molecular cloning techniques (restriction enzyme digestions, ligation reactions, etc.) were carried out as indicated in Sambrook et al. (unless otherwise indicated)[88]. Also, unless otherwise indicated chemical reagents were purchased from Sigma Chemicals (St. Louis, Mo.). The formulae for various solutions used are given in appendix I.

A) Electroelution of DNA Fragments from Agarose Gels

Electroelution of DNA fragments from agarose gels was carried out using a unidirectional electroeluter (Model UEA) as recommended by the supplier of the device (International Biotechnologies, Inc. NewHaven, CN) with slight modifications. Briefly; The device was set up with 0.5X TBE used as running buffer, while a 100 uL salt cushion of 3M NaOAC with 0.01% bromophenol blue was placed in the "V" shaped trap. Gel fragments were excised from agarose gels with razor blades. The fragments were placed in the electro-elution chamber and eluted for 15 to 30 minutes at 100 Volts. Migration of the ethidium bromide stained DNA fragments from the gel into the salt trap was visualized with a hand-held UV light (Ultra-Violet Products Inc., San Gabriel Ca.). When the majority of the DNA band had left the gel fragment the salt cushion was removed and a phenol/chloroform extraction was performed on the solution. The DNA was ethanol precipitated (as per standard protocol) and resuspended in T10E1.

B) Construction of an *S. mansoni* Genomic Library in Lambda FIX II™

Thirty μg of *S. mansoni* (egg), genomic DNA (isolated as previously described[87]) was partially digested with *Sau* 3A restriction endonuclease and resolved on a 0.8% agarose gel (0.5X TBE). High molecular weight DNA (>4 Kb) was cut from the gel and electroeluted (as described above). The first two nucleotides of the *Sau* 3A site (CTAG) were end-filled using the Klenow fragment of DNA Polymerase I (GIBCO/BRL, Burlington Ont.) and dGTP and dATP as per standard protocol. The reaction was inactivated by phenol/chloroform extraction and the DNA recovered by ethanol precipitation.

The end-filled DNA was then ligated into pre-digested Xho-1 Lambda FIX II™ vector arms (Stratagene, La Jolla CA) using T4 DNA ligase (USB Cleveland, Ohio) under standard reaction conditions. After the reaction was complete, the ligation products were packaged using Gigapack™ II Gold packaging extract (Stratagene, La Jolla, CA). Packaged bacteriophage were serially diluted and plated onto a host bacterial strain (as described below) to titer the phage produced. The titer of the packaged phage was then compared to a positive control.

C) Design of Consensus Oligonucleotides U1, U2, U4 and U6 snRNA Gene Probes

Consensus sequences were generated for the U1, U2, U4 and U6 snRNA's based on the analysis of previously published snRNA sequences using the MICROGENIE software package (Beckman Instruments, Mississauga, Ont.). These consensus sequences were constructed at an 80% homology level and were used to design complementary

oligonucleotide probes that would recognize the consensus sequence. If a base location had less than the 80% homology necessary to designate a consensus nucleotide, a degenerate mixture of bases most representative of the bases present was used to best represent that nucleotide location in the oligonucleotide probe. The following figures (#8, #9, #10, and #11) indicate the alignment and snRNA's used to design the screening probes. The probes and their estimated melting temperatures (Tm's) are as indicated in table #2. Tm's were calculated using the following formula

$$Tm = 69.3 + 0.41(\%G+C) - 650/L$$

Where %G+C is the percent of G+C in the oligonucleotide and L is the length of the oligonucleotide. The formula was derived from Sambrook et al.[88]

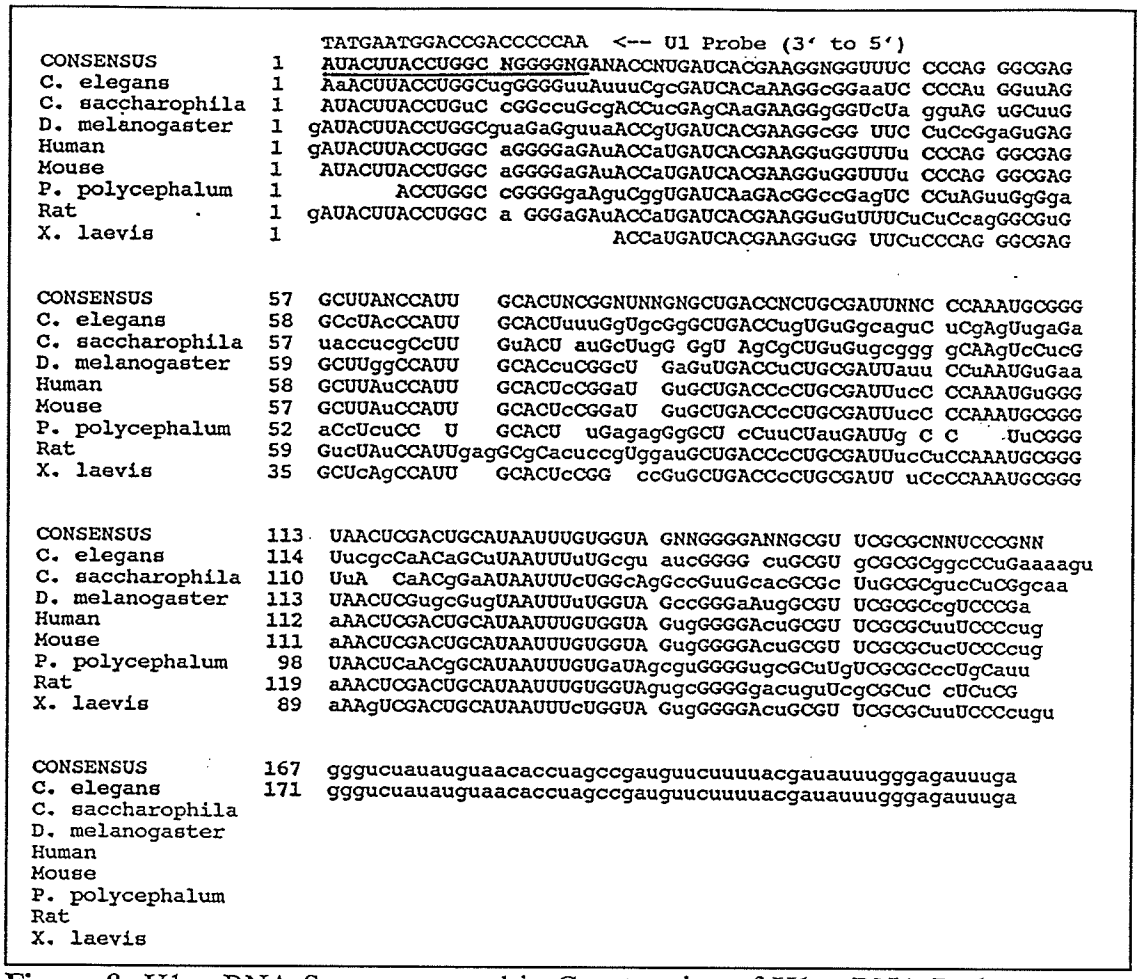


Figure 8: U1 snRNA Sequences used in Construction of U1 snRNA Probe

Nucleotides in agreement with the consensus sequence are in capital letters, while those not in agreement are in lower case. Underlined sequences indicate the regions used to design the oligonucleotide probes. The oligonucleotide probe is noted above these underlined regions.

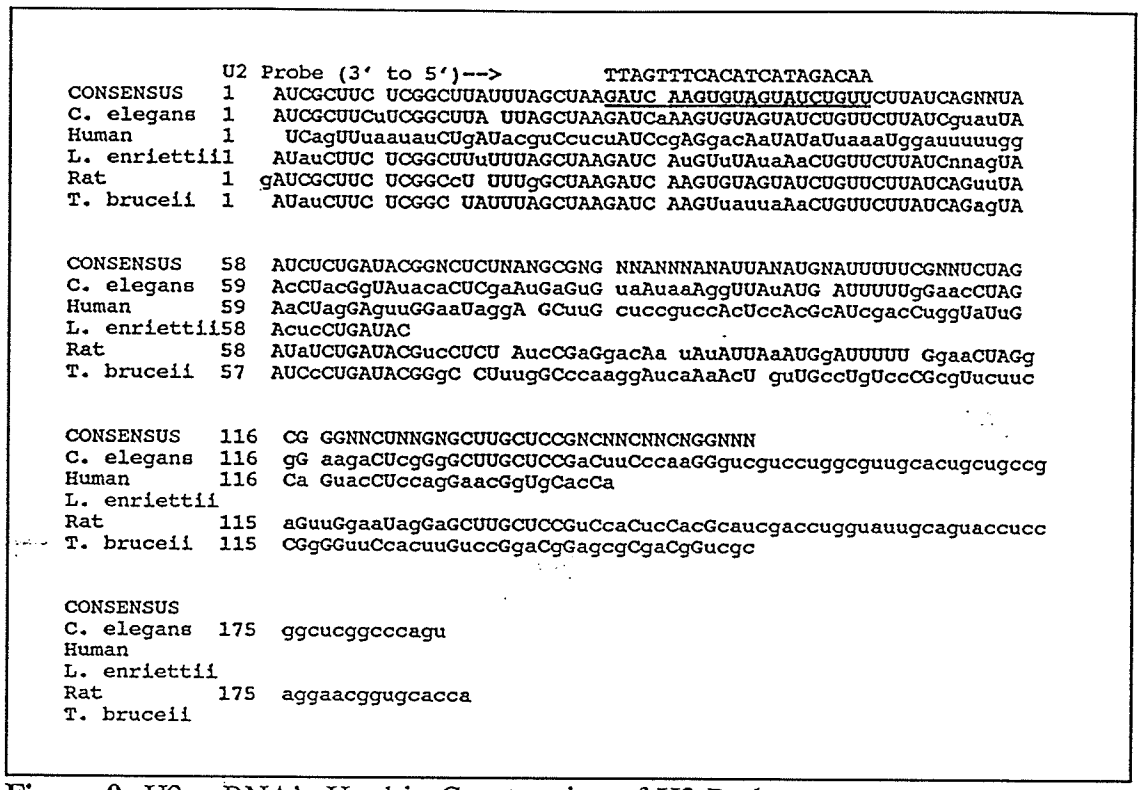


Figure 9: U2 snRNA's Used in Construction of U2 Probe

Nucleotides in agreement with the consensus sequence are in capital letters, while those not in agreement are in lower case. Underlined sequences indicate the regions used to design the oligonucleotide probes. The oligonucleotide probe is noted above these underlined regions.

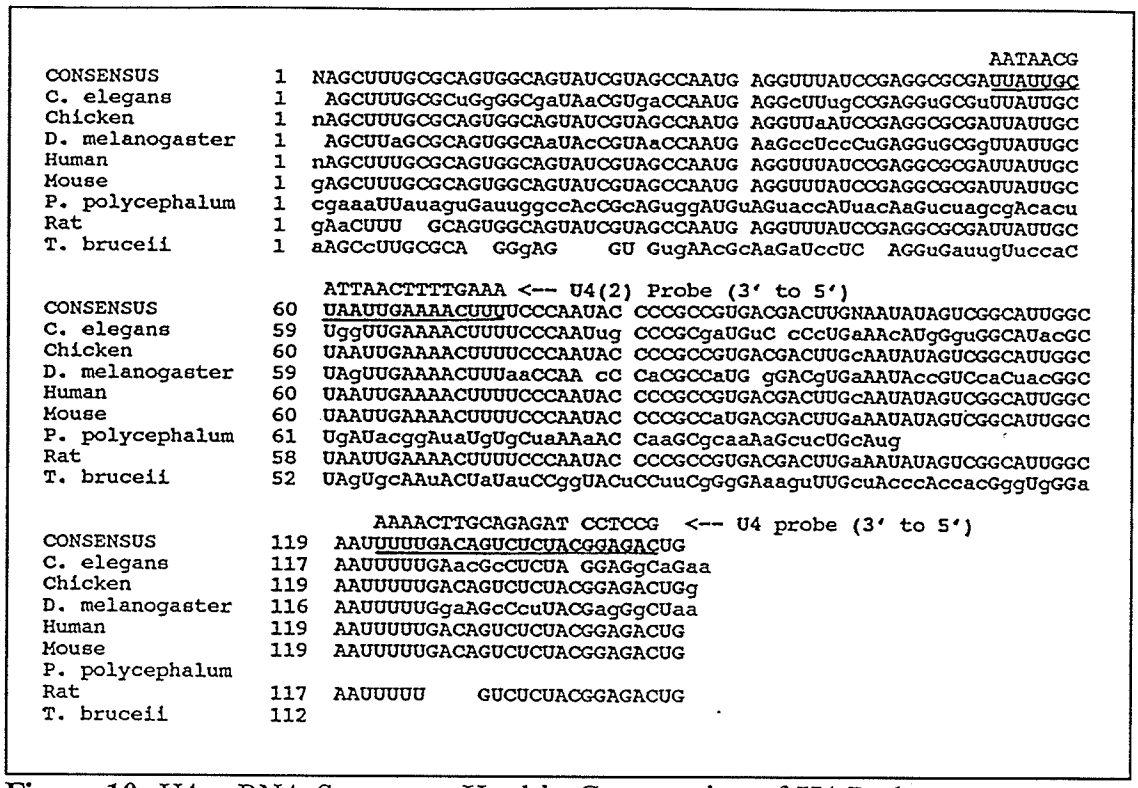


Figure 10: U4 snRNA Sequences Used in Construction of U4 Probe

Nucleotides in agreement with the consensus sequence are in capital letters, while those not in agreement are in lower case. Underlined sequences indicate the regions used to design the oligonucleotide probes. The oligonucleotide probe is noted above these underlined regions.

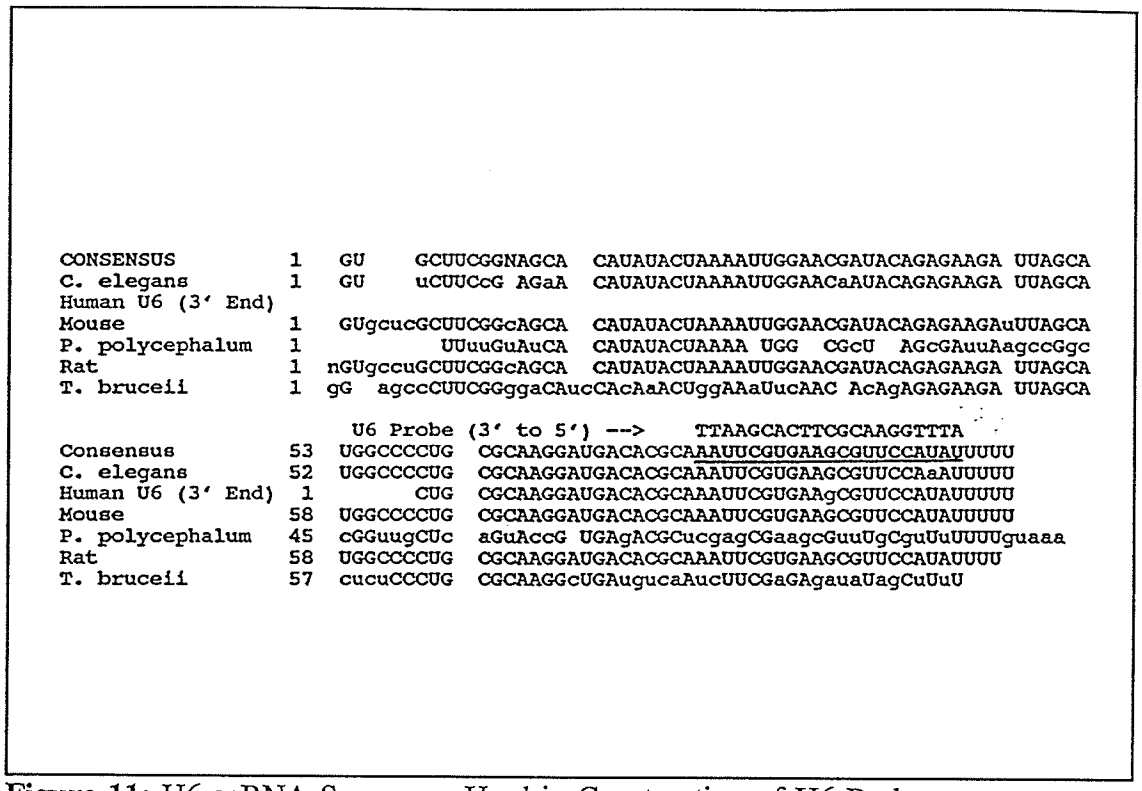


Figure 11: U6 snRNA Sequences Used in Construction of U6 Probe

Nucleotides in agreement with the consensus sequence are in capital letters, while those not in agreement are in lower case. Underlined sequences indicate the regions used to design the oligonucleotide probes. The oligonucleotide probe is noted above these underlined regions.

Table #2 snRNA Oligonucleotide Probes

Probe Name	Sequence of snRNA Probe	Sequence complementary to oligonucletotide probe	Tm
U1	AACCCCAGCCAGGTAAGTAT	nucleotides 1-20 of the U1 snRNA	60°C
U2	AACAGATACTACACTTTGATT	nucleotides 27-46 of U2 snRNA	54°C
U4	GCCTCCTAGAGAC(G/T)TTCAAAA	nucleotides 119-140 of U4 snRNA	60°C
U4(2)	AAAGTTTTCAA(C/T)(C/T)AGCAATAA	nucleotides 52-72 of U4 snRNA	48°C
U6	ATTTGGAACGCTTCACGAATT	nucleotides 78-98 of U6 snRNA	58°C

D) Bacteriophage plating

An overnight culture of the host bacterial strain (SRB) in NZY, was used to inoculate a 50 ml solution of NZY supplemented with sterile $MgSO_4$ and maltose to a final concentration of 10 mM and 0.2% respectively. The culture was grown at 37°C with shaking (200 rpm) to mid-log phase (OD_{600}= 1.0 ODU) and pelleted at 2000 rpm for 10 minutes. The cells were resuspended to a final OD_{600} of 0.5 with sterile 10 mM $MgSO_4$.

From 1 x 10^2 to 5 x 10^4 plaque forming units (PFU's) of diluted bacteriophage (in SM phage dilution buffer) were added to 200 μl of the host bacteria and adsorbed at 37°C for 15-20 minutes. 3 ml of melted (but cooled to 48°C) NZY top agar was then added, mixed and poured onto a dry, prewarmed NZY agar plate and incubated overnight at 37°C.

E) Amplification of Bacteriophage Libraries

Bacteriophage plating was performed using 50 000 PFU of recombinant bacteriophage and 600 μl of host cells prepared as above. The phage were absorbed onto the host cells for 15 minutes at 37°C, then 6.5 ml of melted top agar was added. The phage/bacteria mixture was poured onto a dry, prewarmed 150 mm NZY agar plate and incubated for 6-8 hours (until plaques were readily visable and nearly confluent). The plates were overlaid with 10 ml of SM buffer and stored at 4°C overnight with gentle shaking. In the morning the bacteriophage suspension was recovered and pooled in a test tube, 2 ml of SM was used to wash the plates, and was then pooled with the

bacteriophage suspension in the test tube. Chloroform was added to 5% of total volume and incubated with shaking at room temperature for 15 minutes. Cell debris was removed by centrifugation for 5 minutes at 2000 rpm, and the supernatant was transferred to a sterile tube and titered (as described above) generally producing 10^{10}-10^{13} PFU/ml.

F) Plaque Screening of Bacteriophage Libraries

i) Screening

The screening of genomic libraries was performed using a procedure modified from the instructions provided by the manufacturer of Colony/Plaque Screen™ Hybridisation Transfer Membrane (NEN DuPont Boston MA.). From 5 000 - 10 000 PFU's of bacteriophage were plated onto the appropriate host strain (SRB) on a 150 mM NZY agar plate as described above, but with a slight modification. Top agarose (instead of top agar) was used to suspend the bacteria/phage mixture before plating to reduce non-specific background. The plated phage were incubated overnight then cooled at 4°C for 2-3 hours. A single 150 mM Colony/Plaque Screen™ disk was overlaid onto the plaques formed in the agarose top layer. The disk was labelled with orientation markers and left for 2-3 minutes and gently removed. A second disk was then placed on top and left for 5 minutes. The disks were placed in a pool of 0.5 N NaOH x 2 for 2 minutes each and neutralized in a pool of 1.0 M Tris-HCl (pH 7.5) x 2 for 2 minutes each. The disks were baked for 2 hours at 80°C under vacuum. The disks were ready for hybridisation (as described below) and subsequent identification of positive clones. Positives were identified by the presence of duplicate radioactive signals on both plaque lifts.

ii) Plaque purification

Once a positive plaque was identified it was "plugged" out of the plate using a sterile Pasteur pipette and resuspended in 1 ml of SM buffer. This resuspension of phage was titered and replated on 82 mM plates at a lower dilution then that used for screening (approximately 200-1000 PFU's). Screening was repeated (as above) until a readily recognizable well isolated clone was identified. A positive clone was further "plugged" and re-plated until all of the resulting plaques on the lawn were positive.

G) Northern and Southern Transfer of Nucleic Acids

i) Northern Transfer of RNA

The northern transfer method used is a variation on the capillary transfer method first described by Southern et. al.[89]. Briefly, 6 μg of *S. mansoni* egg, non-polyadenylylated (A(-)) RNA (isolated as previously described[87]) in DEPC treated H_20 was heated at 95°C for 3 minutes in 10X northern gel-loading buffer. The RNA was resolved on a vertical 7M urea, 8% (19:1) bis-acrylamide gel. The samples were loaded onto the gel and run for 6-8 hours at 50 mAmps and 1500 volts. The gel was disassembled, leaving the gel on the bottom plate. The gel was overlaid by a single sheet of Genescreen Plus® hybridisation transfer membrane (NEN DuPont, Boston, Ma), followed by 4 sheets of filter paper. The entire overlay was weighed down and directly transferred for 12-16 hours. The hybridisation membrane was UV crosslinked for 3 minutes, then baked at 80°C for 2 hours.

ii) Southern Transfer of DNA

Southern transfer was carried out as follows: 2 μg of *S. mansoni* (egg) DNA mixed with 10X Gel loading buffer dyes was loaded onto a 0.8-2% agarose gel (0.5X TBE) and resolved for 2-16 hours. The DNA was nicked by UV irradiation for 10-15 minutes. After UV treatment, the gel was washed 3 X 20 minutes in 0.5N NaOH/1.5M NaCl, rinsed briefly in distilled H_20, and washed twice more for 20 minutes each in 0.5 M Tris (pH 7.5)/1.5 M NaCl and rinsed in distilled H_20. The transfer was performed in 10X SSC as follows: A glass plate was suspended over a glass dish containing 10X SSC. A filter paper wick was draped over the plate into the 10X SSC, 3 layers of filter paper (slightly larger than the gel) were placed over the wick. The agarose gel was placed on top of the filter paper layer. Genesceen Plus® hybridisation membrane was placed directly over the gel, with three layers of filter paper laid on top of the transfer membrane. The whole overlay was then topped with a stack of absorbent towels with a weight placed on top of them. Capillary transfer was allowed to take place for 16-72 hours.

H) Hybridisation of Probes to Northern and Southern Blots or Plaque Lifts

Hybridisation of end-labelled oligonucleotide probes to target DNA or RNA was carried out in the following manner. The oligonucleotide was 5′ end-labelled using T4 Polynucleotide Kinase (USB Cleveland, Ohio) under the following reaction conditions 100 ng of oligonucleotide, 5-10 μl of 10 mCi/ml γ^{32}P-ATP, 5.3 units of T4 Kinase and the appropriate buffer were incubated for 1 hour at 37° C. The reaction was then diluted

to 1 ng/μl in T10E1 and spun through a Sephadex G-50 (Pharmacia) column for two minutes in a clinical table top centrifuge to remove unincorporated γ^{32}P-ATP. The specific activity was determined by adding 1 μl of the reaction to 5 ml scintillation cocktail (Ecolume, ICN) and counted on a scintillation counter (LS 500CE, Beckman). Average specific activities were in the range of 2 x 10^8 to 1 x 10^9 counts per minute per μg of oligonucleotide. The hybridisation membrane was first pre-hybridised in a low stringency hybridisation solution (Low Hybridisation Solution) for 30 minutes at a restrictive temperature (basically the Tm of the probe -5°C). The hybridisation temperatures used are as follows; U1-54°C, U2-50°C U4(2)-45° and U6-53°C. Labelled probes were added at a concentration between 2-5 x 10^6 CPM per ml of Low Hybridisation Solution and incubated overnight at the restrictive temperature. After hybridisation the blot was washed once for 10 minutes followed by two 30 minute washes in 2X SSC/0.1% SDS at the restrictive temperature. The blot was exposed to X-Ray Film (Kodak X-O-matAR5) and placed between two intensifying screens (NEN DuPont) at -80°C for 6-72 hours.

I) Isolation of Bacteriophage DNA

i)Large scale preparation of bacteriophage lambda

Standard methods for recovery of large amounts of bacteriophage (Sambrook et. al.[88]) were generally used with slight modifications. A 5 ml culture of NZYCM was inoculated from a single colony of SRB and grown overnight at 37°C on a rotary shaker (200 rpm/minute). One ml of this culture was used to further inoculate a 500 ml culture

of NZYCM which was grown at 37°C (with shaking) to mid-log phase (OD_{600}= 0.5), 10^{10} PFU of bacteriophage was added and further incubated until lysis of the bacteria occurred (3-4 hours). Ten ml of chloroform was added to the flask and incubated a further 10 minutes with shaking lysing the bacterial culture. The flask was stored overnight at 4°C. The cultures were warmed to room temperature and 500 μg of pancreatic DNAse and RNAse (Sigma Chemicals, St. Louis Mo.) added and allowed to incubate at room temperature for 30 minutes. NaCl was added to a final concentration of 1M (29.2 g) and allowed to dissolve. The solution was cooled on ice for 1 hour. Debris was isolated by centrifugation at 11 000 X g for 10 minutes at 4°C. PEG 8000 (Sigma) was added to the recovered supernatant at a final concentration of 10%(w/v) and allowed to dissolve for 1 hour with slow stirring. The solution was cooled on ice for 1 hour to allow bacteriophage particles to precipitate. The particles are recovered by centrifugation at 11 000 x g for 10 minutes at 4°C. The supernatant was removed and the pellet resuspended in 8 ml of SM. The PEG 8000 was removed by a v:v phenol/chloroform extraction. The aqueous phase was recovered, and CsCl added (0.5 g per ml), giving a final density of 1.05 g/ml. This solution was layered onto the top layer of a discontinuous CsCl step gradient. The step gradient was prepared by layering 10 ml each of 3 different densities of CsCl in SM (1.7 g/ml, 1.5 g/ml, 1.45 g/ml). The step gradient was centrifuged at 22 000 rpm for 2 hours in an SW 28 swinging bucket rotor (Beckman Instruments). Four to five ml of bacteriophage particles were recovered from the 1.45/1.5 g/ml interface.

ii) Extraction of Bacteriophage Lambda DNA

The recovered phage was placed in Spectrapor 4 dialysis membrane (Spectrum Medical Instruments Inc. Los Angeles CA.) and dialysed twice for 1 hour each against 4 L of a 10 mM NaCl, 50 mM Tris, 10 mM $MgCl_2$ solution. The dialysed solution was transferred to a centrifuge tube, and EDTA was added to a final concentration of 20 mM. Proteinase K was added at 50 μg/ml followed by SDS to a final concentration of 0.5% (w/v). The solution was incubated at 65°C for 1 hour and gently extracted using V:V phenol/chloroform. The aqueous phase was dialysed overnight at 4°C against 4 litres of T10E1. The DNA was precipitated by EtOH precipitation and quantified by spectrophometric analysis yielding between 100-1500 μg of bacteriophage DNA.

J) Mapping and Subcloning of Lambda Phage Inserts

i) Mapping of Lambda Phage Inserts

The quality of the bacteriophage DNA was assessed by restriction enzyme digestion and agarose gel electrophoresis. The DNA was judged to be of good quality if intact bands were observed, and an insert of the appropriate size could be liberated by restriction enzyme digestion. Once the DNA was judged suitable, 5-8 μg of phage DNA was digested with 10 units of the restriction enzyme *Not* I for 1 hour, removing the recombinant insert. Once this digestion was complete 3-5 units of either *Nde* I, *Eco* RV, or *Hinc* II restriction enzymes were added to the reaction and an aliquot immediately removed and transferred to a 0.5 M EDTA solution, effectively stopping the digestion. Samples were taken in this manner at times of 1,3,5,10,15, and 60 minutes, producing

a set of partial digestions. The resulting digestions were resolved on a 1% agarose gel (0.5X TBE) and transferred to a support matrix as described. The Southern blots were probed by radiolabelled T3 oligo (as described) and exposed to X-ray film. The Southern blot was stripped of the T3 probe (30 minutes at 95°C in 1% SDS, 0.1% SSC) and re-probed with the T7 oligo. The T7 probe was then stripped (as above) and the blot re-probed using the specific screening oligonucleotide.

ii) Subcloning of Lambda Phage Inserts

Once the restriction map was determined, inserts suitable for subcloning were identified based upon the size of fragment, and the compatability of the digested ends relative to the multiple cloning site of the vectors used. From 5-8 μg of phage DNA was digested by the appropriate restriction enzymes. The digestion was resolved on an agarose gel, and the appropriately sized fragments isolated and eluted as described. Eluted DNA was ligated into either pBluescript, or pBluescribe (Stratagene). From 50-150 ng of recovered insert DNA was mixed with 50 ng of Calf Intestinal Phosphatase (CIP) (Boehringer Mannheim) treated plasmid and incubated with T4 DNA Ligase (USB) under standard reaction conditions, either at 16°C overnight or at 4°C for 72 hours.

K) Transformation of Competent Escherichia coli

i) Heat-Shock transformation of DH5α

a) Preparation of Competent Cells

A 5 ml overnight culture of SOC was inoculated from a single colony of DH5α and grown overnight at 37°C with shaking. One ml of this overnight was used to

inoculate a 100 ml culture of SOC which was grown for 2-3 hours with shaking to an OD_{600} of approximatly 0.46. The culture was incubated on ice for 5 minutes, pelleted for 10 minutes at 3K rpm, and the supernatant poured off. Pellets were resuspended immediately in 33 ml FSB and centrifuged for 5 minutes at 1.5K rpm. The supernatant was discarded and the pellet resuspended immediately in 8 ml FSB. The cells were centrifuged for 5 minutes at 1.5 K rpm, and the supernatant discarded. A 280 μl aliquot of DMSO was added and the cells placed on ice for 5 minutes. A further 280 μl of DMSO added, and placed on ice for 10 more minutes. Six hundred μl aliquots of the cell suspension were flash-frozen in a dry ice/ethanol bath.

b) Transformation of DH5α by Heat Shock.

One μl of a ligation reaction (generally 5 ng) was added to 20 μl of competent cells (previously thawed on ice) and incubated on ice for 30 minutes. The cells were heat shocked by placing them at 42°C for 60 seconds, and placed back on ice. Eighty uL of SOC media was added and incubated for 1 hour at 37°C with shaking. From 10-100 μl of the transformation was plated on selective media (LB-AMP) and incubated overnight at 37°C. The selective media also contained 5 mg IPTG and 1 mg X-gal per plate. White transformants were selected for further evaluation.

ii) Electroporation of DH5α

a) Preparation of Competent Cells

A 5 ml culture of LB was inoculated from a single colony of DH5α and

grown overnight at 37°C with shaking. One ml of this culture was used to inoculate a 1L culture of LB which was grown with shaking to an OD_{600}= 0.5 - 0.7. The culture flask was chilled on ice for 15-30 minutes and centrifuged at 4000 x g for 15 minutes. The media was poured off, the culture resuspended in 1 L of 10% Glycerol and centrifuged again (as above). The cells were resuspended in 500 ml of 10% glycerol, centrifuged again, resuspended in 30 ml of 10% glycerol, spun down a final time and resuspended in 4-6 ml of 10% glycerol. The cells were frozen in aliquots on dry ice.

b) Electroporation of competent cells

Electroporation was performed in a Gene Pulser™ device (Bio-Rad Richmond CA.). Fourty μl of cells and 1-2 μl of DNA (5-10 μg) were mixed in a sterile, chilled 0.1 cM cuvette. The cells were pulsed at 2.5 kV, with a resistance of 200 ohms. The charging voltage was 1.8 kV, with a direct field strength of 12.25 kV/cM and a pulse length of 4-5 mSec. The cuvette was removed from the chamber and the cells resuspended in 1 ml of SOC media. The cells were transferred to a sterile test tube and incubated at 37°C for 1 hour. Ten to twenty μl of cells were plated on selective media.

L) Boiled Preps for Plasmid Preparation

A 5 ml culture of BHI-Amp or LB-Amp was inoculated from a single colony and grown overnight at 37°C with shaking. A 1.5 ml aliquot of the culture was centrifuged at 12 000 rpm for 1 minute, the supernatant removed and the pelleted bacteria

resuspended in 350 μl of Boiled Prep Buffer (BPB). Twenty-five μl of freshly prepared lysozyme (10 mg/ml in BPB) was added, vortexed, and incubated at room temperature for 1 minute. The tube was placed at 95°C for 1 minute, and centrifuged at 12 000 rpm for 15 minutes. The resulting slimey, insoluble pellet was removed with a sterile toothpick. 200 μl of 7.5 M NH_4OAc, and 600 μl of isopropanol were added and incubated for 15 minutes at room temperature. The mixture was centrifuged for 15 minutes at 12 000 rpm, the supernatant discarded, and the pellet allowed to air dry. The pellet was resuspended in 50 μl of T10E1 containing 10 μg/μl of RNAse A.

M) Small Scale Purification of Plasmid DNA for Sequencing

A 5 ml culture of BHI-Amp or LB-Amp was inoculated from a single colony and grown overnight at 37°C with shaking. A 1.5 ml aliquot of the culture was spun down at 12 000 rpm for 2 minutes and the supernatant removed, leaving the pellet as dry as possible. The pellet was resuspended in 100 μl ice cold plasmid preparation solution I, mixed by inversion and incubated for 5 minutes at room temperature. Two hundred μl of plasmid preparation solution II was added, gently mixed and incubated on ice for 5 minutes. Finally, 150 μl of plasmid preparation solution III was added, mixed and incubated on ice for 5 minutes. The mixture was centrifuged for 5 minutes at 12 000 rpm, the supernatant transferred to a new tube, and centrifuged again for 5 minutes at 12 000 rpm. The supernatant was transferred to a new tube, RNAse A added to a final concentration of 20 ug/ml and incubated for 20 minutes at 37°C. After RNAse A treatment the solution was extracted using v:v phenol/chloroform and precipitated by the

addition of 2-2.5 volumes of 100% EtOH followed by freezing on dry ice. The solution was centrifuged at 12 000 rpm for 15 minutes, the supernatant removed and the pellet allowed to air dry. The pellet was resuspended in 50 μl T10E1 and 30 μl of a 20%PEG/2.5M NaCl solution added. This was placed on ice for 20 minutes, centrifuged for 10 minutes, and the supernatant removed. The pellet was rinsed in 70% EtOH, frozen, centrifuged and air-dried. The pellet was resuspended in 20 μl H_20 yielding between 9-15 μg of DNA suitable for sequencing.

N) Large Scale Preparation of Plasmid DNA

A 5 ml culture of BHI-Amp or LB-Amp was inoculated from a single colony and grown overnight at 37°C with shaking. One ml of this culture was used to inoculate 500 ml broth culture of BHI-Amp which was grown at 37°C with shaking to an OD_{600} of 0.6 to 0.8. Chloramphenicol was added to a final concentration of 25 μg/ml and incubated overnight. The culture was centrifuged for 20 minutes at 4000 rpm, the supernatant removed, and the pellet resuspended in 10 ml plasmid preparation solution I containing lysozyme (5 mg/ml). This was incubated for 5 minutes at room temperature. Ten ml of plasmid preparation solution II was added, gently mixed, and placed on ice for 10 minutes. Finally, 7.5 ml of plasmid preparation solution III was added, gently mixed again, and incubated on ice for a further 10 minutes. The solution was centrifuged at 15 000 rpm for 30 minutes at 4°C and the supernatant transferred to another tube. Thirteen and a half ml of isopropyl alcohol was added and incubated at room temperature for 15 minutes. The solution was centrifuged at 12 500 rpm for 30 minutes at room temperature

and the supernatant removed. The pellet was resuspended in 8 ml T10E1, to which 8.4 grams of CsCl, 500 μl of ethidium bromide, and 250 μl of 10% sarkosyl was added. The mixture was then heat sealed in a 15 ml Beckman ultracentrifuge tube and centrifuged at 50 000 rpm for 18-24 hours on a Beckman L8-70M model ultracentrifuge. The plasmid band was extracted and brought up to 5 ml volume with sterile H_20. An equal volume of isoamyl alcohol was added, mixed, and the organic layer removed and discarded. This was repeated until both phases were clear and all the ethidium bromide was removed from the aqueous phase. Two volumes of 100% EtOH were added and precipitated at -20°C for 30 minutes. The DNA was recovered by centrifugation at 8 000 rpm for 15 minutes. The DNA pellet was resuspended in 500 μl of T10E1 and a final EtOH precipitation was done to remove any excess salt.

O) Sequencing Reactions/Acrylamide Gel Preparation

The sequencing of plasmid DNA was carried out following the instructions of the Sequenase™ kit (USB Cleveland Ohio) with slight modifications, briefly the reactions were carried out as follows.

i) Denaturing Supercoiled Plasmid

Alkali denaturation of 3-4 μg of plasmid was performed in 0.2 N NaOH for 5 minutes at room temperature. The reaction was neutralized/precipitated with 1/10 volume 2M NH_40Ac and 2 volumes of 100% EtOH. The reaction was frozen on dry ice, centrifuged at 12 000 rpm for 15 minutes, the supernatant discarded, and the pellet washed with 70% EtOH.

ii) Sequenase™ reactions

The denatured DNA was resuspended in the presence of 1 picomole of primer, 2 μl of Sequenase buffer, and 7 μl of H_20. The primer/template mixture was incubated at 65°C for 15 minutes and cooled to room temperature for 15 minutes. The annealed primer/template was labelled by the addition of 1 μl 0.1 M DTT, 2 μl (1/10 dilution) labelling mix, 0.5 μl of α^{35}S dATP (10 μCi/μl), and 2 μl (1/8 dilution: 1.6 units final) Sequenase™ V2.0. The reaction was mixed and incubated for 5 minutes at room temperature. 3.5 μl of this labelling reaction was transferred to 2.5 μl of each of the four ddNTP termination tubes (ddCTP, ddTTP, ddATP, ddGTP) at 8 μMolar ddNTP/ 80 μM dNTP, mixed, and the reaction allowed to proceed for 5 minutes at 37°C. The reaction was stopped by the addition of 4 μl of stop solution. Each of the four ddNTP termination reactions was heat denatured for 3 minutes at 95°C before resolution on an acrylamide gel.

iii) Acrylamide Gel Electrophoresis

The reactions were resolved on a combination of 4% and 8% 7M urea, 19:1 bis-acrylamide gels. Each reaction usually allowed one 8% run along with a 4% double run, allowing the reading of sequence 500-600 bp away from the primer. Gel's were prepared as described below:

	8% Acrylamide	4% Acrylamide
Urea	63 grams	63 grams
10X TBE	7.5 ml	7.5 ml
40% Acrylamide Stock	30 ml	15 ml
H_20	60 ml	75 ml

The gel preparation was mixed by stirring for approximately 30 minutes to allow the urea to dissolve. Then, 1.5 ml of 10% Ammonium persulfate (APS) and 75 μl of TEMED was added to the gel preparation and poured between the cleaned and taped IBI vertical electrophoresis plates (International Biotech Inc. New Haven CN.). The gel was allowed to polymerize from 1 to 16 hours before the gel apparatus was set up. An 8% gel was usually run at 30-50 mAmps for 3-4 hours. A 4% double run gel was run at 30-50 mAmps for 3-4 hours, the samples re-loaded into empty lanes and run for an additional 3-4 hours. The gel was disassembled and fixed for 30 minutes in 5% acetic acid/7% methanol, followed by rehydration for 30 minutes in ddH_2O. The gel was lifted from the plate by absorption to a sheet of filter paper and vacuum dried at 80°C for 1 hour. The dried gel was exposed to X-ray film overnight at room temperature.

P) Sequence Analysis of Subcloned Fragments

Sequence data obtained was entered using the MICROGENIE sonic digitiser and analyzed using the MICROGENIE software package provided with the instrument. Analysis of the data usually consisted of merging sequence data obtained from a single

primer, and merging of opposite strands. Sequence data was also screened against possible contaminant organisms (lambda phage DNA, plasmid DNA, etc.).

Q) Secondary Structure Analysis of Putative snRNA's

Secondary structure analysis of putative snRNA's were carried out using the computer program PC-Fold. This computer program uses the mathematical free energy calculations of Zuker et al[90]. in designating potential secondary structures of both DNA and RNA. The putative snRNA's were analyzed using this program and by comparison with the previously published secondary structures of snRNA's from other organisms.

R) Polymerase Chain Reactions

Reaction conditions for PCR were as follows: from between 10-100 ng of *S. mansoni* (egg) DNA was used as a template in the following final concentration of cocktail reagents. 1 X PCR buffer : (1 mM Tris-Cl, 50 mM KCl, 1.5 mM $MgCl_2$, 0.001 % (w/v) gelatin), 0.2 mM deoxynucleotide triphosphate, 0.5 μM of primers, and 2.5 units AmpliTaq DNA polymerase (Perkin Elmer Cetus) in a final volume of 100 μl. The entire mixture was overlaid with mineral oil, and placed in a Model 7200 thermal cycler (Perkin Elmer Cetus). The amplification protocol for the thermal cycler was as follows: 30 cycles of denaturing at 94°C for 1 minute, annealing at 45°C for 1 minute, and extension at 72°C for 1 minute, followed by a final extension at 72°C for 7 minutes.

S) Primer Extension Reactions

i) Labelling of Primers

Radiolabelling of 100 ng of primer was completed as described for hybridisation reactions, with the following changes; the final kinase reaction was resuspended in a 20 uL volume (5 ng/ml rather than 1 ng/ml), and the reaction was not run through a spun column (unincorporated γ^{32}P-ATP was not removed).

ii) RNA Sequencing technique

Five ng (1 μl) of the oligonucleotide primer was mixed with 10 μg of total RNA in an annealing buffer (250 mM KCl, 10 mM Tris-HCl (pH 8.3)) in a total volume of 12 μl. The mixture was heated at 80°C for 3 minutes then incubated at a restrictive temperature (Tm - 5°C) of the oligonucleotide; 48°C for U2-probe, 43°C for U4(2)-probe, and 53° for U6-probe. Two μl of the annealed primer-template combination was added to 4 separate termination reaction tubes that contained 3.3 μl of reverse transcription buffer (24 mM Tris-Hcl (pH 8.3), 16 mM $MgCl_2$ 8 mM dithiothreitol, 0.4 mM dNTP's, 100 μg/ml actinomycin D, and 5 units AMV Reverse Transcriptase Superscript (Gibco-BRL)), and 1 μl each of four ddNTP's (ddCTP, ddTTP, ddATP, ddGTP) at 1 mMolar. The reaction was incubated for 45°C for 45 minutes. The reaction was stopped by the addition of 2 μl of stop solution (100% formamide with 0.3% bromophenol blue and 0.03% xylene cyanol). The reaction mixture was then heated for 3 minutes at 95°C and loaded on a polyacrylamide sequencing gel (as described).

Results

Cloning and Characterisation of the U1 snRNA

Evaluation of U1-probe for Screening a Library

To assess the usefulness of the U1-Probe in screening the *S. mansoni* genomic library (constructed as described earlier) a set of northern and southern analyses were performed using this oligonucleotide. As can be seen in figure #12 the U1-probe clearly recognised a single band of 172 nucleotides in length in the Northern blot of *S. mansoni* A(-) RNA. This demonstrated that the probe was recognizing a single transcript of non-polyadenylylated (A-) RNA of 172 nucleotides in length. The size of the RNA fragment recognized fell within the size range of U1 snRNA's already identified[35]. In figure #13 the U1-probe oligonucleotide was used to probe a Southern blot of genomic *S. mansoni* DNA digested to completion with the restriction endonucleases; *Bam* H1, *Eco* R1, *Sau* 3A and *Sma* 1. Digestion was estimated to be complete as a vast excess of restriction endonuclease was used, as well digestion was allowed to take place for an extended period of time (overnight). The probe clearly recognized multiple bands with varying intensities. This suggested that the DNA sequence recognised by the U1-probe existed in multiple copies in the genome. Therefore the U1-probe was used to screen the genomic library for DNA fragments containing U1 or U1-like snRNA genes.

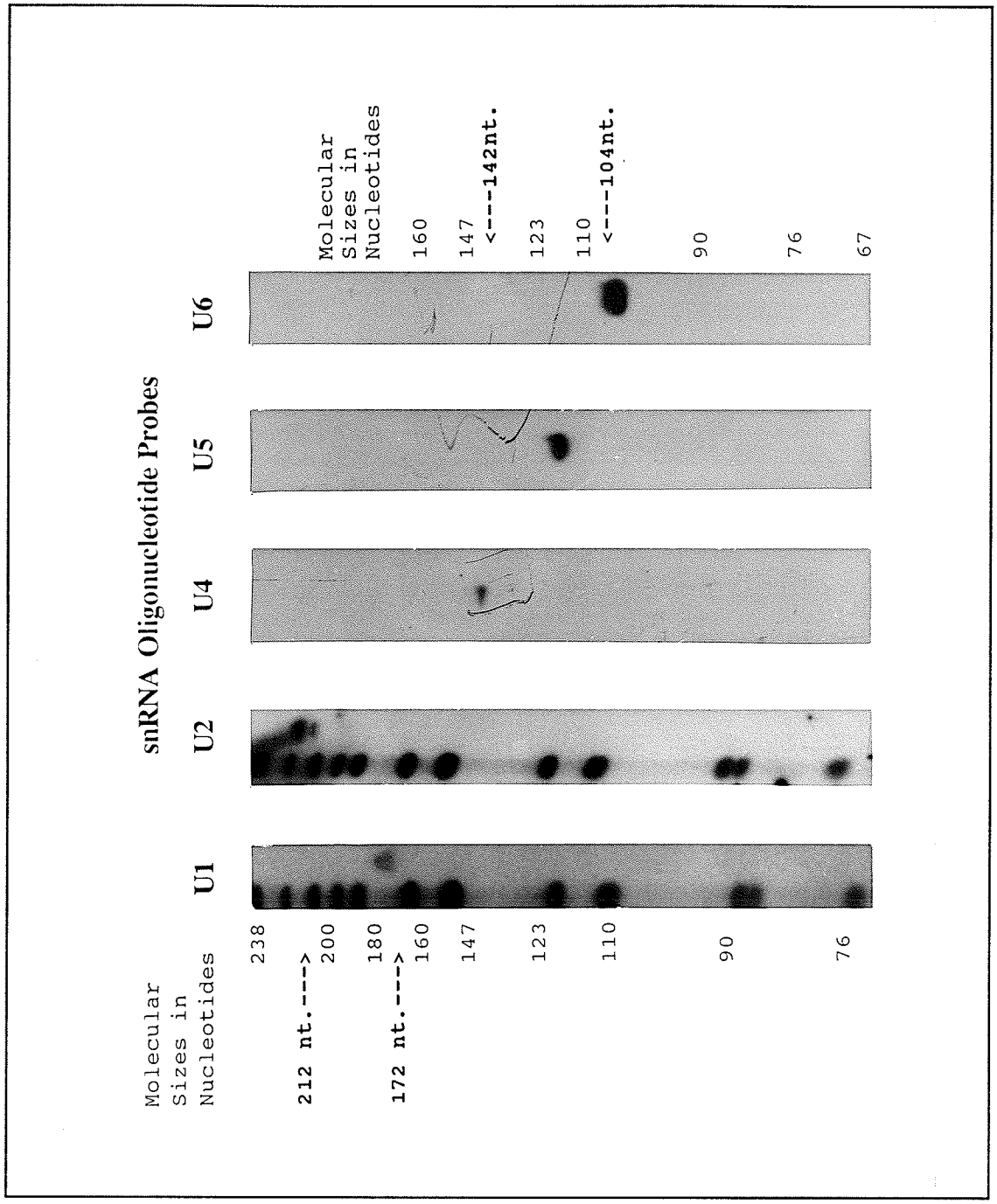


Figure 12: PAGE Northern Analysis of snRNA Probes

6 μg of A(-) RNA was resolved on an acrylamide gel, transferred and probed with the indicated screening oligonucleotides. Molecular weight markers are indicated on the left for the U1 and U2 probes, and on the right for U4, U5 and U6 probes.

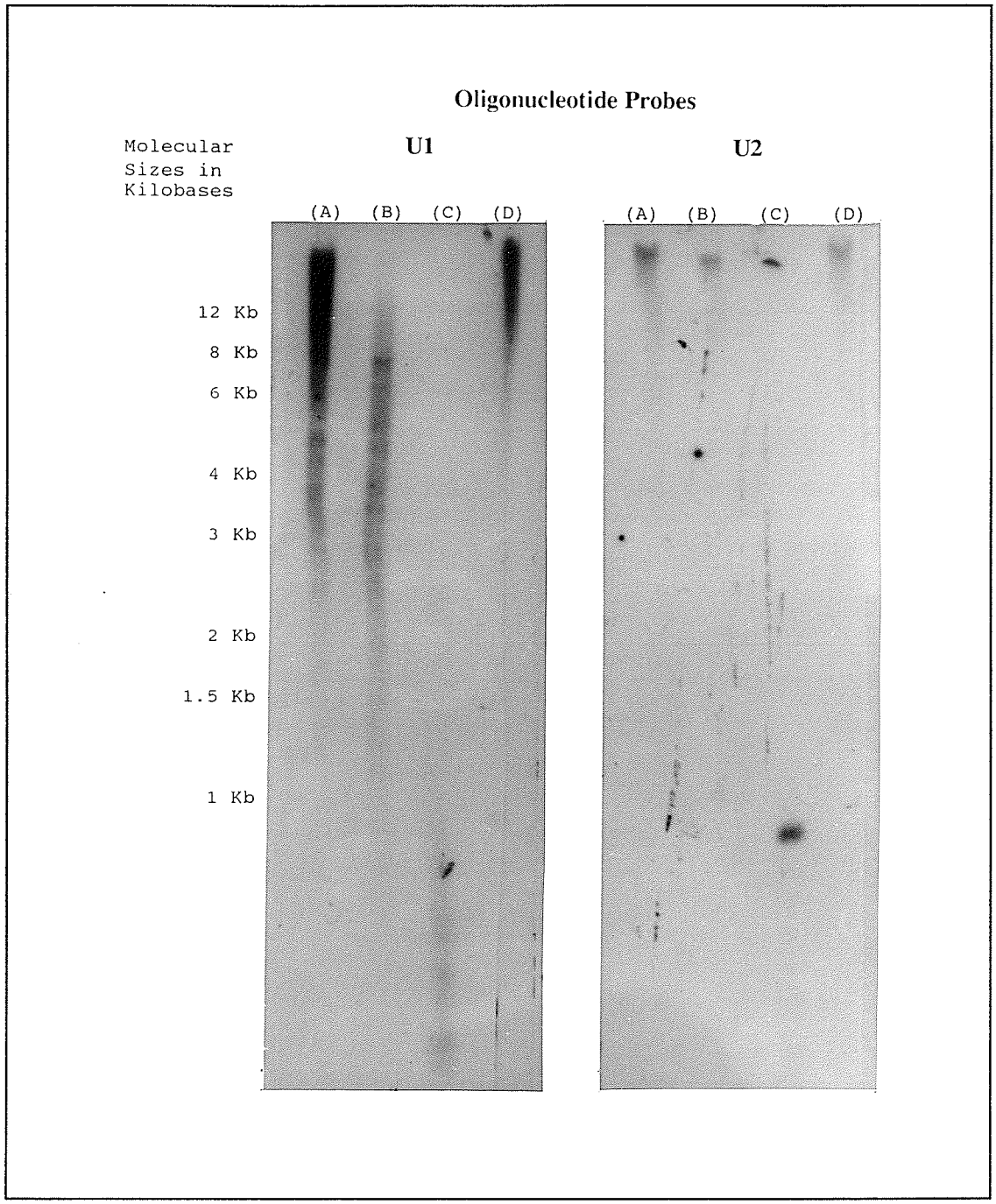


Figure 13: Southern Analysis of U1 and U2 Oligonucleotide Probes

2 μg of DNA was digested to completion with restriction enzymes; Lane A: *Bam* H1, Lane B: *Eco* R1, Lane C: *Sau* 3A, Lane D: *Sma* 1. The DNA was resolved on an agarose gel, and transferred. The blot was then probed with the indicted oligonucleotide probe. Numbers at left are Molecular weight markers.

Screening a Genomic Library for U1 snRNA clones

With the utility of the probe established by the Northern and Southern analyses, the genomic library was screened. Approximately 600 000 plaque forming units (PFU's) of recombinant lambda phage were screened using the U1-Probe. Four positive clones were identified and purified. They were labelled as U1c1, U1c2, U1c3, and U1c4. These bacteriophage clones were amplified, grown in large scale, and the recombinant phage DNA extracted and purified as described.

Mapping of U1 Genomic Clones

The insert of the U1c1 clone was initially characterised by restriction enzyme digestion and agarose gel electrophoresis. The insert was found to be 14 KB in length. Restriction enzyme mapping of this insert was carried out as described. The insert DNA within the multiple cloning site of the Fix II vector was removed from the bacteriophage arms by a complete *Not*-1 restriction endonuclease digestion. This digestion resulted in the release of the insert DNA flanked at its 5′ and 3′ ends by the characterised T3 and T7 bacteriophage promoter binding regions. The 5′ and 3′ regions could be readily identified by oligonucleotide probes complementary to either the T3 or T7 promoter sites (T3 or T7 probe). A series of incomplete digestions by different restriction enzymes (*Hinc* II, *Eco* RV and *Nde* I) produced a nested group of fragments of various lengths. Resolution on an agarose gel, followed by southern transfer, and subsequent probing by the T3 or the T7 probe demonstrated the size of each fragment containing the T3 or T7 region. Since these T3 and T7 regions were located at the 5′ and 3′ ends of the released

insert, the location of specific restriction enzyme sites from either end of the insert were identified. Based upon the sizes of the incompletely digested restriction enzyme fragments a complete restriction enzyme map of the insert was generated. The predicted result was the presence of larger fragments at the shorter digestion times (with fewer cuts, more chance of obtaining larger fragments) with the smallest fragments being present at the longer digestion times (longer time to cut, more sites digested, producing shorter fragments). Southern blots of these digests were used to pinpoint the possible location of potential snRNA genes by stripping them and probing with the snRNA probe. By comparing bands recognised by the snRNA probes with bands identified by the T3 or T7 mapping of restriction enzyme sites, the location of the snRNA gene from the T3 or from the T7 end was determined. The map of U1c1 is presented in figure #14. This map was used to identify regions suitable for subcloning and sequencing. Two different sized fragments were targeted for subcloning; a 0.5 Kb *Hinc* II fragment, and a 1.9 Kb *Hinc* II/*Not* I fragment. These fragments were isolated by agarose gel electrophoresis and subcloned into the sequencing vector pBluescript SK+. The sequencing vector pBluescript SK+ contains a multiple cloning site inserted into a β-galactosidase gene which allows for blue/white colour selection of any recombinant plasmids. Any bacterial colonies containing inserts were selected and the recombinant plasmids tested for the presence of the expected insert by restriction enzyme digestion, followed by agarose gel electrophoresis.

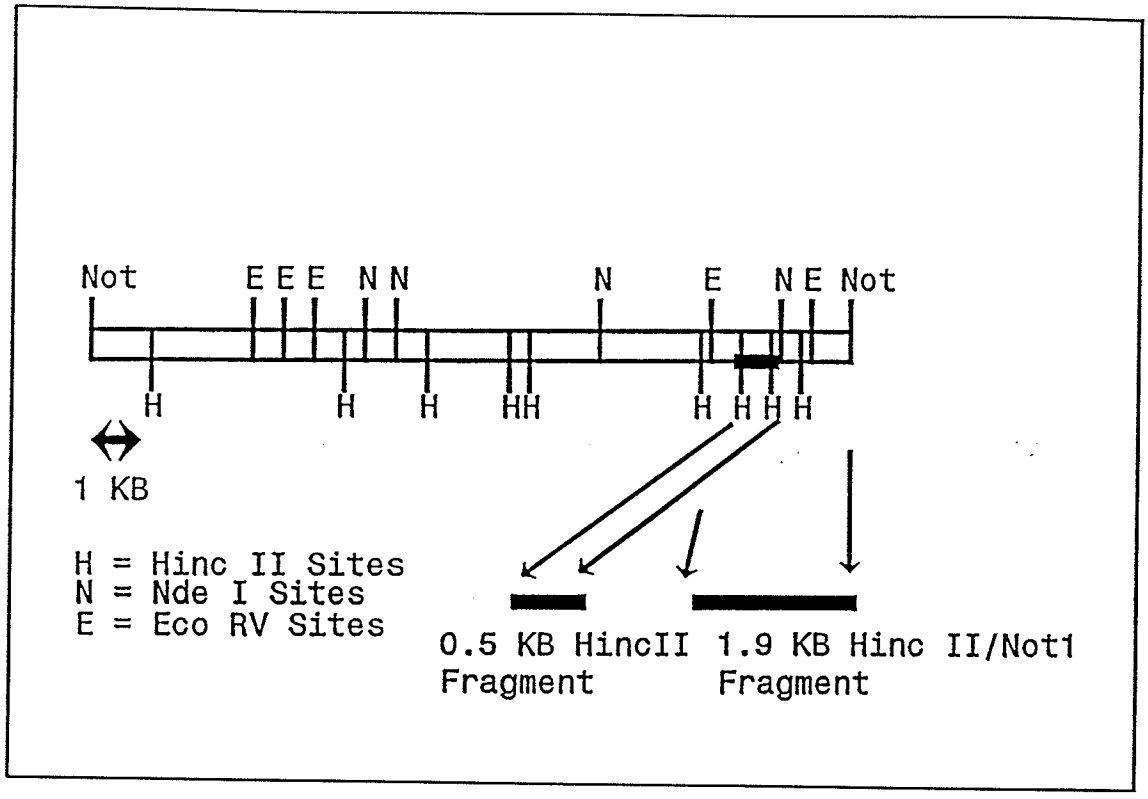


Figure 14: Restriction Map of Clone U1c1 Lambda Phage Insert (14 Kb)

The restriction map of U1c1 is indicated above, the letters indicate the specific restriction enzyme sites, while the dark box indicates the region of the insert recognised by the U1-Probe. Regions targeted for subcloning are indicated by the arrows (not to scale).

Inserts of the expected size were then sequenced. However, after repeated ligations it proved difficult to obtain fragments of the expected size. Some of the aberrantly sized fragments were sequenced and these data suggested that an unexplainable recombination event was occurring. The positive clones often contained elements of the lambda phage arms along with the expected S. mansoni insert DNA. The reason for this recombination was unclear, as the host *E. coli* strain was recombination deficient in the standard *E. coli* recombination mechanisms. Due to time constraints, further characterisation of U1 clones was abandoned so that more attention could be placed on the characterisation of the other snRNA's of interest.

Cloning and Characterisation of the U2 snRNA

Evaluation of U2-probe for Screening a Library

To assess the utility of the U2-Probe for screening the *S. mansoni* genomic library a set of Northern and Southern analyses were performed. As indicated in figure #12, the U2-probe recognized a single band of 212 nucleotides in the Northern blot of *S. mansoni* A(-) RNA. This indicates that the probe recognized a single non-polyadenylylated transcript of 212 nucleotides in length. The size of the RNA fragment recognized was within the range of U2 snRNA's already described[35]. In figure #13 the U2-probe oligonucleotide was used to probe a Southern blot of genomic *S. mansoni* DNA digested to completion with the restriction endonucleases *Bam* H1, *Eco* R1, *Sau* 3A and *Sma* 1. The probe recognises multiple bands with varying intensities. Similar to the results from the U1-probe discussed earlier the data suggests that the fragment recognised exists in

multicopy, and that the product recognised may exist in some sort of repeat motif (described in discussion). These data suggested that the U2-probe recognised the U2 snRNA, as well as genomic copies of either the U2 snRNA gene or U2 snRNA pseudogenes and was therefore useful for screening the *S. mansoni* genomic library.

Screening a Genomic Library for U2 snRNA Clones

With the utility the U2-probe established by the Northern and Southern analysis, screening of the constructed genomic library was completed. Approximately 200 000 plaque forming units (PFU's) of recombinant lambda phage were screened (as described) using the U2-Probe. Six positive clones were identified and purified. They were labelled as U2c1 to U2c6. These clones were amplified, grown in large scale, and recombinant phage DNA extracted and purified as described.

Mapping of U2 Genomic Clones

The insert of the U2-3 clone was characterised by restriction enzyme digestion and agarose gel electrophoresis. The insert was found to be 13 KB in length. Restriction enzyme mapping of this insert was carried out as described for the U1 genomic clones. The map of U2c3 is presented in figure #15. This map was used to identify fragments suitable for subcloning and sequencing. Two sizes of fragments were targeted for subcloning a 1 KB *Xho* I fragment, and a 0.7 KB *Nde* I fragment. These fragments were isolated and subcloned into the sequencing vector pBluescribe SK+. Recombinant

plasmids were identified by colour selection and the presence of the expected sized insert confirmed by restriction enzyme digestion.

Sequence Analysis and Identification of a Putative U2 snRNA

The clones resulting from the ligation of the inserts into pBS were sequenced using the T3 or T7 oligonucleotides as primers. The resulting sequences were analyzed using the software described and used to generate a complete sequence of the insert. (See Figures #16 and #17. These sequences were then used to identify regions corresponding to the U2-Probe oligonucleotide. The sequence of the putative U2 snRNA (pU2 snRNA) was determined based upon the following; i)the predicted size of the U2 snRNA from the northern analysis data, ii)the location recognised by the U2 screening oligonucleotide which could be determined by the published U2 sequence data and iii)the primer extension sequence of the 5′ end of the U2 snRNA.

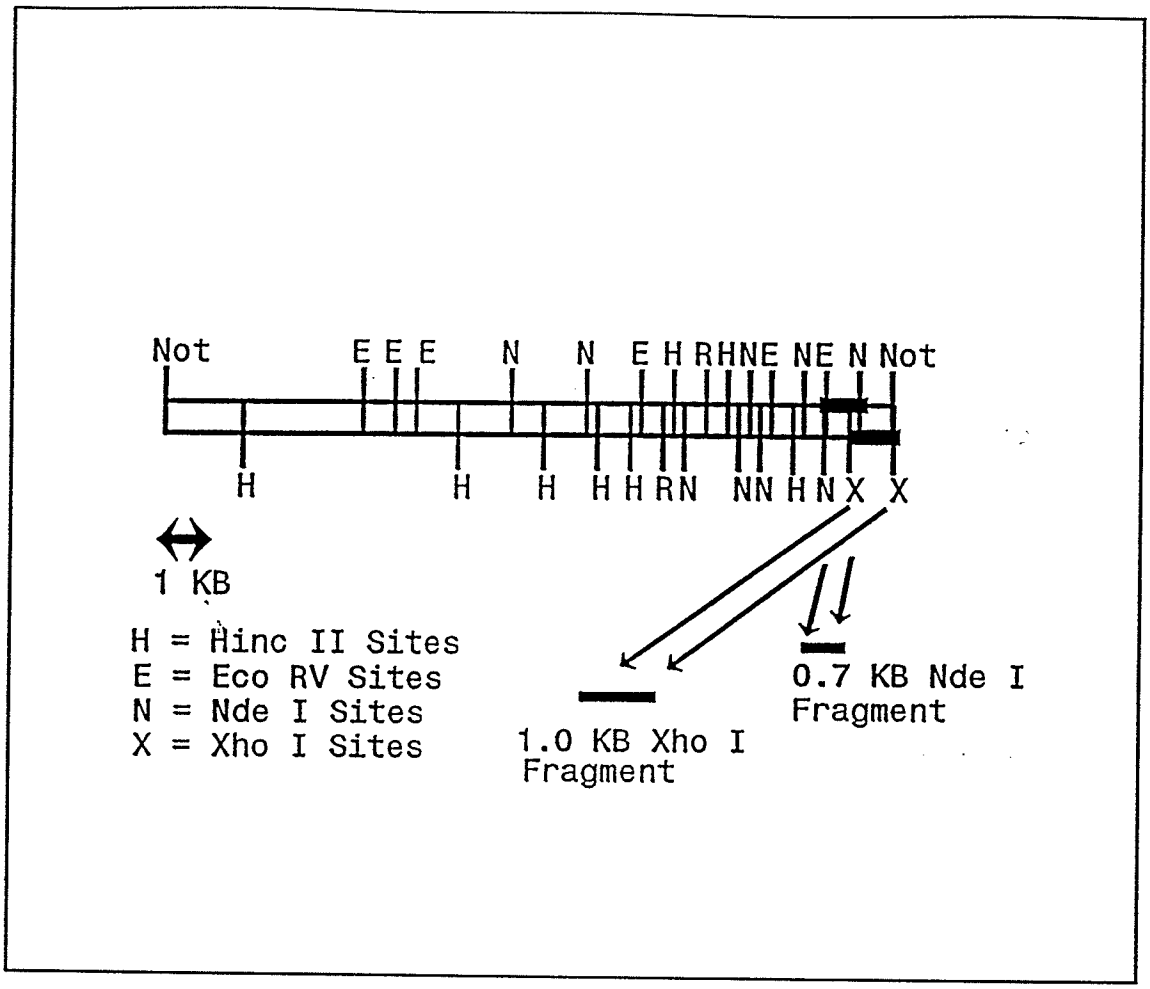


Figure 15: Restriction Map of Clone U2c3 Lambda Phage Insert (13 Kb)

The restriction map of U2c3 is indicated above, the letters indicate the specific restriction enzyme sites, while the dark box indicates the region of the insert recognised by the U2-Probe. Regions targeted for subcloning are indicated by the arrows (not to scale).

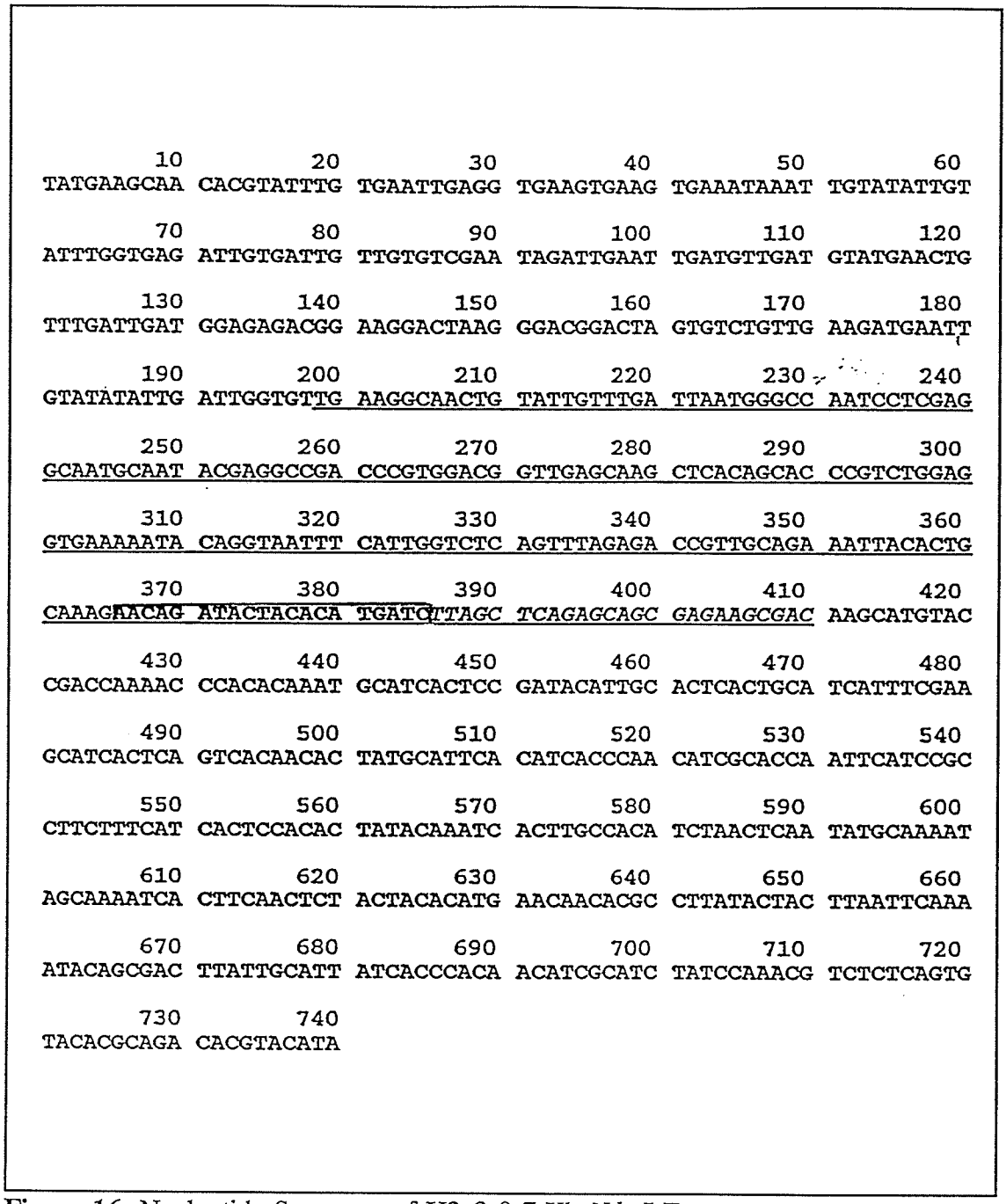


Figure 16: Nucleotide Sequence of U2c3 0.7 Kb *Nde* I Fragment

Underlined regions are sequence complementary to the pU2-snRNA. The region is complementary to nucleotides 212 to 1 of the pU2 snRNA. Boxed regions correspond to the sequence of the screening oligonucleotide, while sequence in italics (385-410) is complementary to the primer extension sequence data.

```
        10          20          30          40          50          60
TCGAGGCAAT GCAATACGAG GCCGACCCGT GGACGGTTGA GCAAGCTCAC AGCACCCGTC

        70          80          90          100         110         120
TGGAGGTGAA AAATACAGGT AATTTCATTG GTCTCAGTTT AGAGACCGTT GCAGAAATTA

        130         140         150         160         170         180
CACTGCAAAG AACAGATACT ACACATGATC TTAGCTCAGA GCAGCGAGAA GCGACAACGA

        190         200         210         220         230         240
TGTACCGACC AAAACCCACA CAAATGCATC ACTGATCATA CATTGCACTC ACTGCATCAT

        250         260         270         280         290         300
TTGCAAGCAT CACTCAGTCA CAACACTATG CATTCACATC ACCCAACATC GCACCAATTC

        310         320         330         340         350         360
ATCCGCCTTC TTTCATCACT CCACACTATA CAAATCACTT GCCACATCTA ACTCAATATG

        370         380         390         400         410         420
CAAATAGCAA ATCACTTCAA CTCTACTACA CATGAACAAC ACGATCCTTA TACTACTTAA

        430         440         450         460         470         480
TTCAAATACA GCGACTTATT GCATTATCAC CACACATCGC ATCTATCAAA TCGTCTCTCA

        490         500         510         520         530         540
GTGTACACGC AGACACGTAC ATATGCTCAC AAGTCATTGA CTCACTCTCA CCTACACATA

        550         560         570         580         590         600
ACACAACCAT TATCACTACC GATAACAATG CCACAACCAA TTAGAACACC ACAAAAAGCA

        610         620         630         640         650         660
CAACTAACAA CACATGTAGC ACTCTACAAC TATCAATACC AATACCACCA ACACCATGAC

        670         680         690         700         710         720
TATGAACACC AACAGAACAC CGACACGATA GCGATGACAA GTGTTCAGTG TTCAATTGCA

        730         740         750         760         770         780
CATAGAATCG TGCGCATATG AAGCAACACG TATTTGTGAA TTGAGGTGAA GTGAAGTGAA

        790         800         810         820         830         840
ATAAATTGTA TATTGTATTT GGTGAGATTG TGATTGTTGT GTCGAATAGA TTGAATTGAT

        850         860         870         880         890         900
GTTGATGTAT GAACTGTTTG ATTGATGGAG AGACGGAAGG ACTAAGGGAC GGACTAGTGT

        910         920         930         940         950         960
CTGTTGAAGA TGAATTGTAT ATATTGATTG GTGTTGAAGG CAACTGTATT GTTTGATTAA

        970
TGGGCCAATC TCGA
```

Figure 17: Nucleotide Sequence of U2c3 1 Kb *Xho* I Fragment

Underlined regions are complementary to the pU2-snRNA. The first region is complementary to nucleotides 174 to 1 of the pU2 snRNA, while the second region is complementary to nucleotides 212 to 174 of the pU2. Boxed regions correspond to the sequence of the screening oligonucleotide, while sequence in italics (151-175) is complementary to the primer extension sequence data.

Primer Extension Analysis of U2-probe

In an attempt to generate some preliminary data it was decided to perform primer extension sequencing using the U2-probe screening oligonucleotide as a primer. The primer extension analysis using the U2-Probe was expected to produce sequence for the first 26 or 27 nucleotides of the U2 snRNA at the 5′ end, based upon the binding position of the U2-probe (27-46). Readable sequence was generated for the first 25 nucleotides. These data validated the sequence data from the two subclones (See Figures #16 and #17) and also confirmed the sequence of the 5′ end of the pU2 snRNA. An example of such sequence data is presented in figure # 18.

Secondary Structure Analysis of pU2 snRNA

The secondary structure and possible folding patterns of the pU2 snRNA was determined by comparison to previously published U2 snRNA folding patterns, as well as computer assisted modelling based on the method of Zucker[90]. As can be seen in Figure #19 the pU2 snRNA could be folded with a number of stem-loop structures. The overall shape of this folded pU2 is strongly suggestive of the folding pattern of U2 snRNA's from other species. In addition to this structural analysis the sequence of the U2 snRNA was compared to published sequences of the U2 snRNA's from other species (See Figure #20).

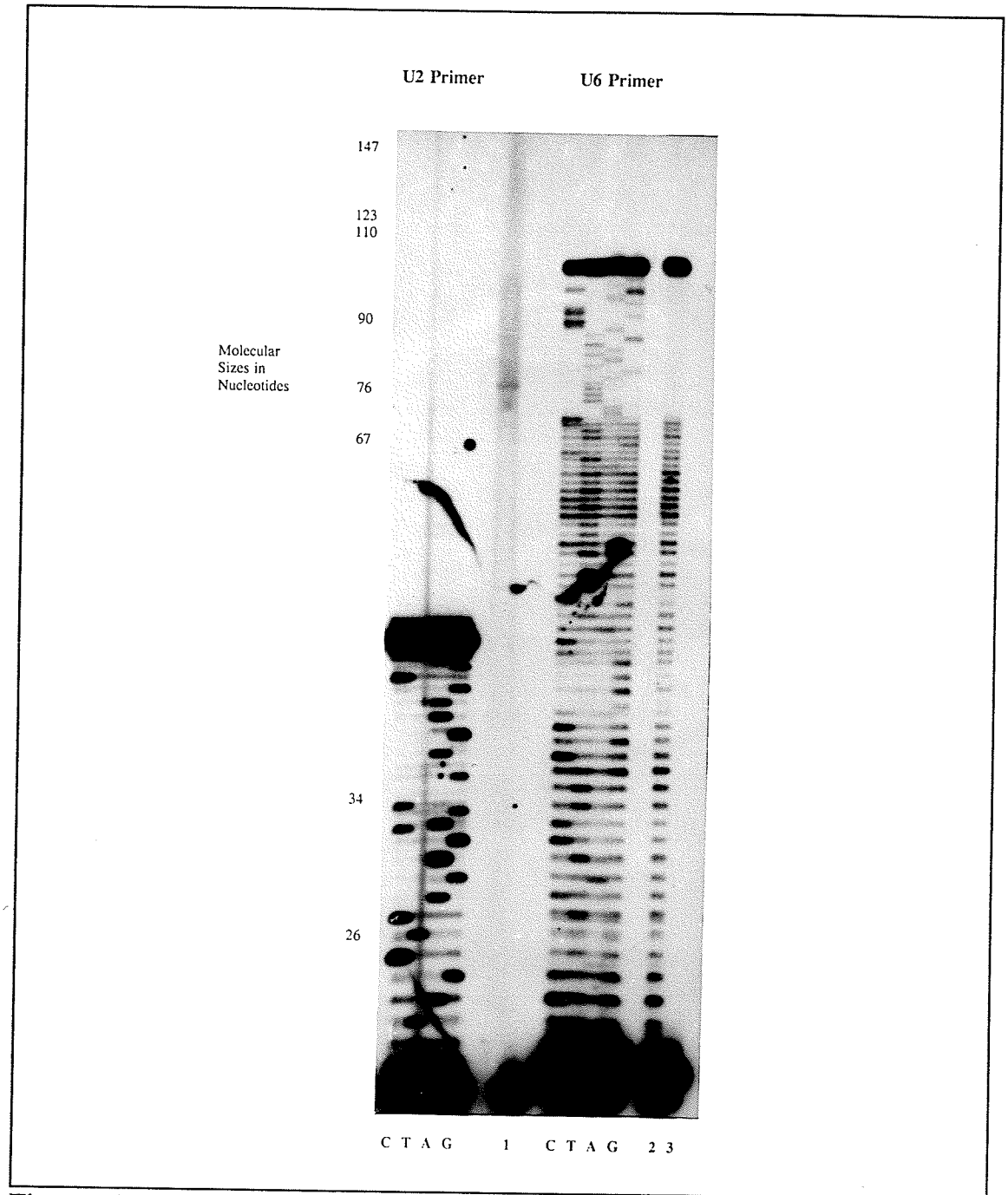


Figure 18: Primer Extension Sequencing Using The U2 and U6 Oligonucleotide Probes.

Primer extension sequencing for the U2 Primer is indicated on the left, and on the right for the U6 Primer. CTAG indicate the presence of the specific ddNTP's in the reaction mix. (1) and (2) are the labelled U2 and U6 probe respectively. (3) is a primer extension reaction using the U6 probe without ddNTP's in the mix.

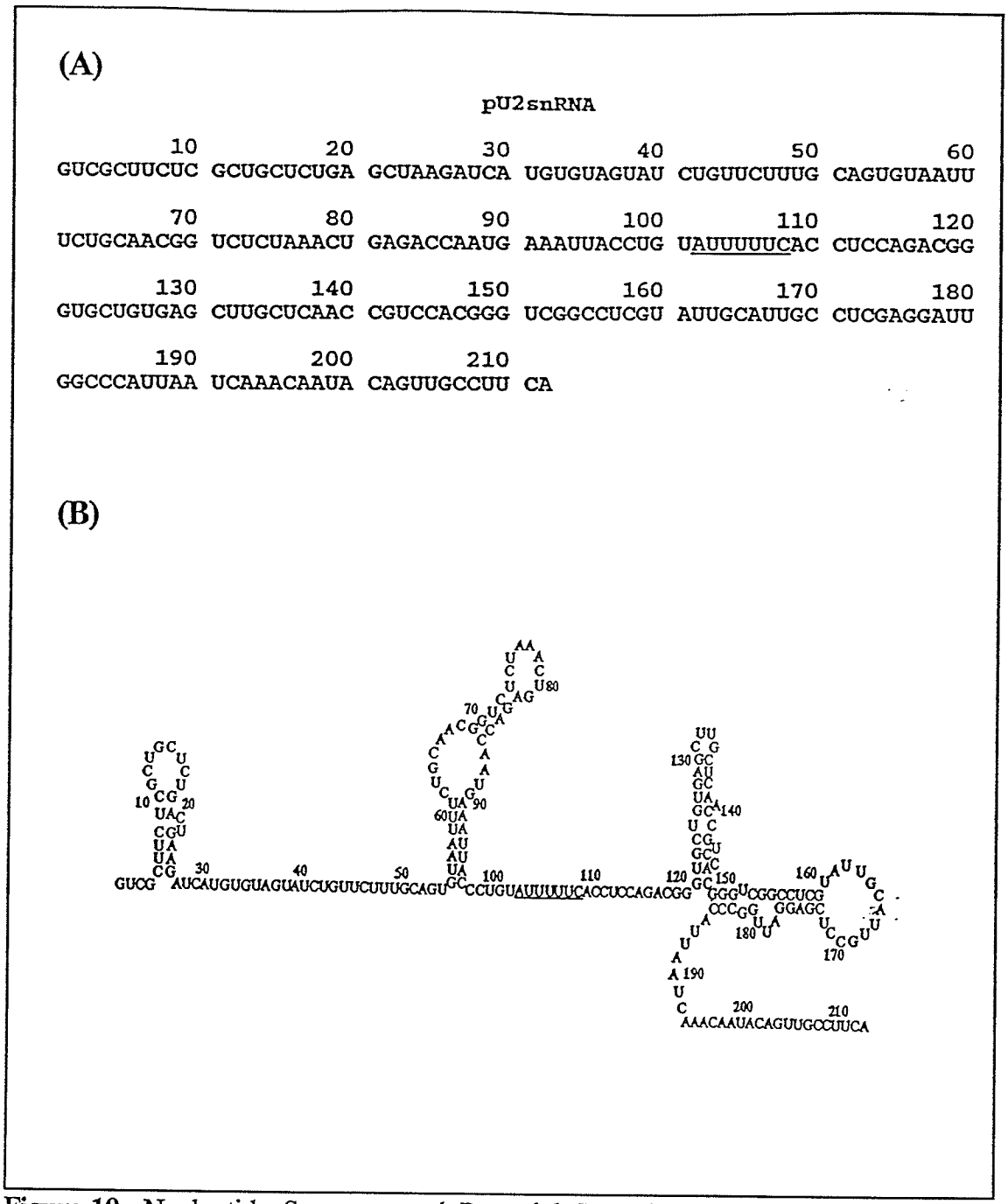


Figure 19: Nucleotide Sequence and Potential Secondary Structure of Putative *S. mansoni* U2 snRNA

(A) Shows the sequence for the pU2 snRNA. Bold sequence is confirmed by primer extension analysis, while underlined sequence is the putative Sm site. (B) Shows the predicted secondary structure of the pU2 snRNA.

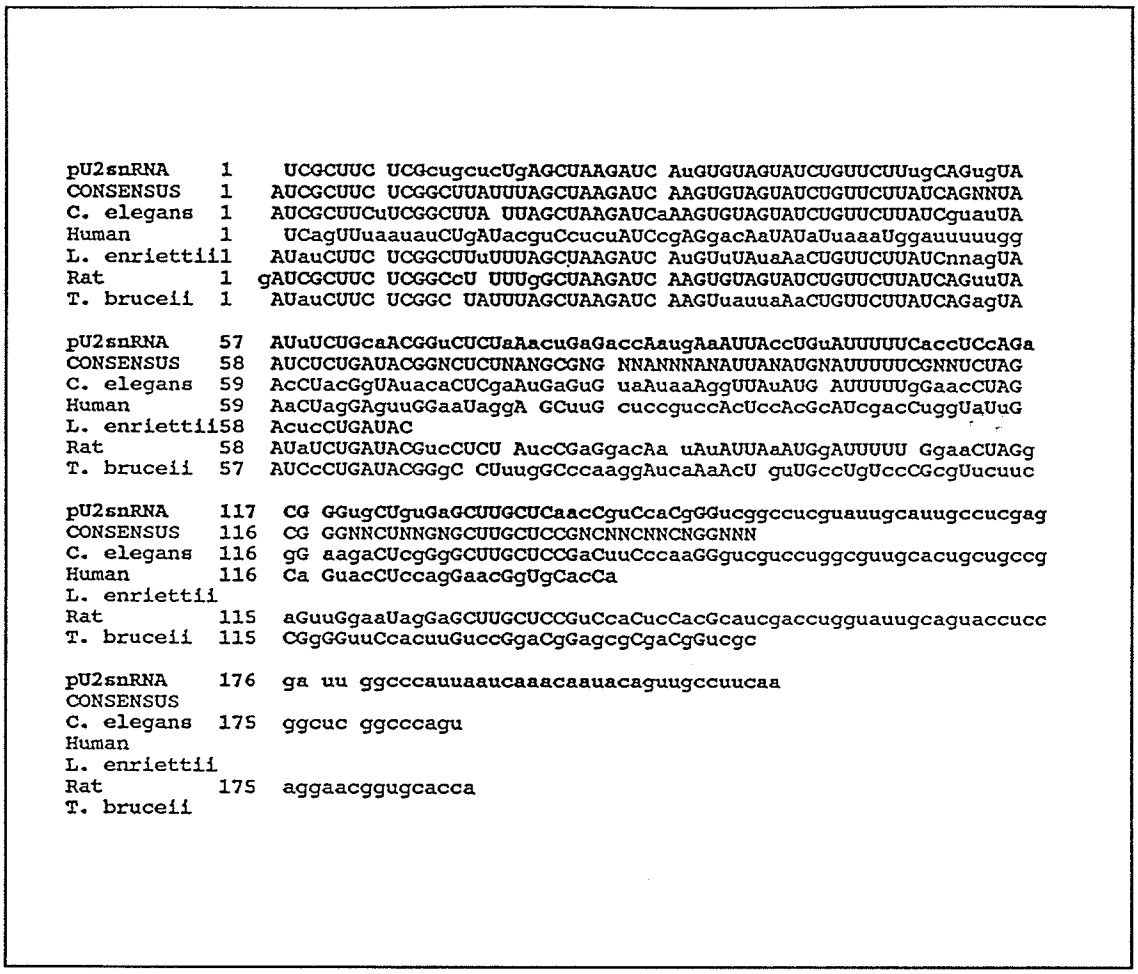


Figure 20: Sequence Comparison of pU2 snRNA and Published snRNA's

A comparison of the pU2 snRNA (Top) with previously published U2 snRNA's. Sequences in agreement with the pU2 are in upper case, while sequences in disagreement are in lower case.

Preliminary Evaluation of Genomic Organisation of U2 snRNA Genes

Based on data obtained from the sequence analysis of the 1 Kb *Xho* I fragment it was determined that this fragment contained the sequence of two portions of pU2-snRNA's. Based upon this, and the suggestive evidence from the Southern analysis of the U2-probe (described in discussion) that the U2-snRNA gene may be encoded in repeat units, it was determined that the U2 snRNA gene may be present in multiple copies in the *S. mansoni* genome. Based on the sequence data of the 1 Kb *Xho* I fragment it was thought that this repeat unit was in a head to tail array. Knowing the sequence of the pU2 snRNA and the surrounding area, oligonucleotide primers were designed to amplify a complete U2 snRNA gene repeat unit. Primer U2-3′ (AGAATTCTGTTGAAGGCAACTGTATTGTTTG) was complementary to the 3′ end of the pU2-snRNA, primer U2-5′ (AGGATCCAATCAATATATACAATTCATC) corresponded to the sequence on the opposite strand, just downstream of the 3′ end of the U2 snRNA. Using these two primers in a PCR reaction permitted the amplification of a region of genomic *S. mansoni* that corresponded to an entire U2 snRNA gene, if the genomic organisation was a head-to-tail tandom array. Preliminary PCR was attempted using these primers along with T3 and T7 primers to better characterize the genomic inserts in the bacteriophage vector. The resulting PCR products were slightly anomalous in that the 5′ primer in combination with the T3 primer gave the same sized product as the 3′ primer and the T7 primer. This was unexpected as the resulting products should have been quite different in size (based on the genomic map the 5′/T7 product should be smaller than the 3′/T3 product. This led to the testing of the individual primers in the

PCR reaction. The use of either the U2-3′ primer, or the U2-5′ primer alone led to the generation of a 700 bp PCR product (see figure #21). The only explanation for this finding is that the U2 snRNA genes do exist in a repeat organisation, but rather than a head to tail array, these repeat units exist in some sort of head to head type of array being encoded on opposite strands. For a diagram of expected PCR products see figure #22.

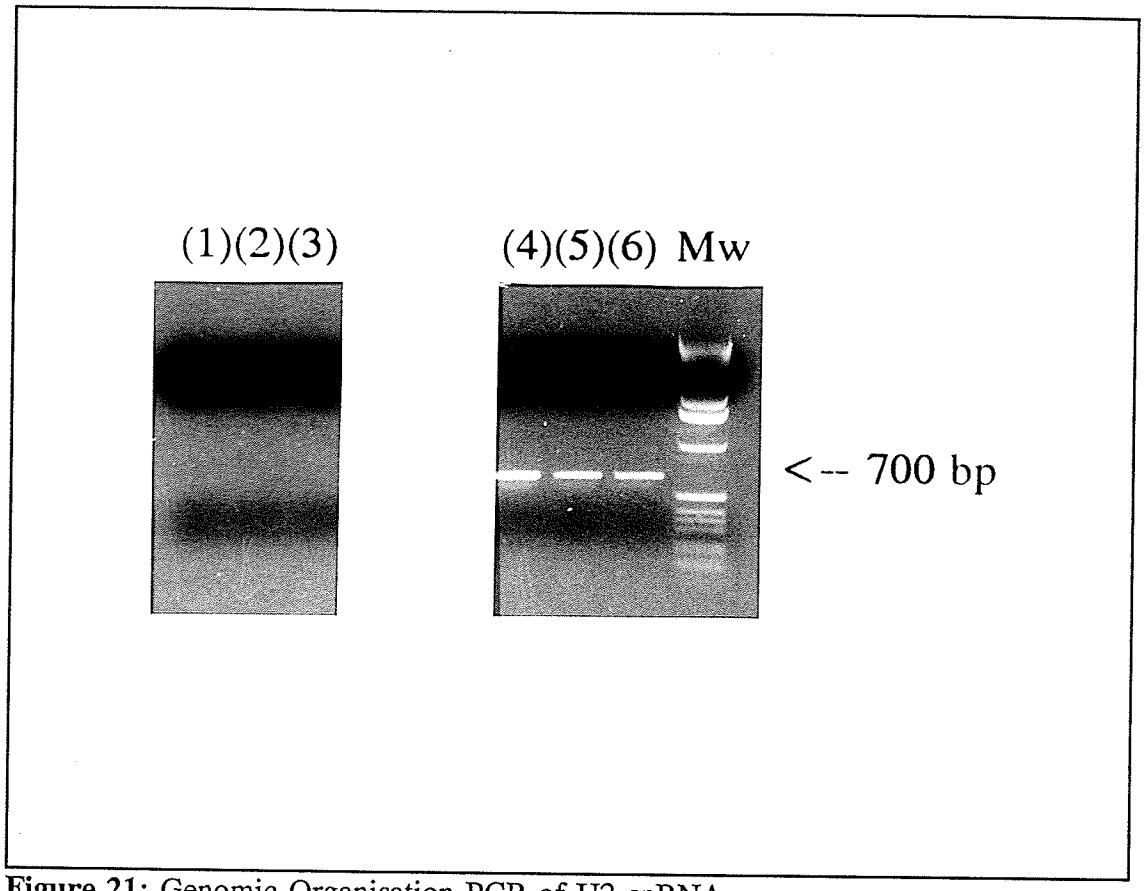


Figure 21: Genomic Organisation PCR of U2 snRNA

PCR reactions were performed as described. Lanes (1-3) are negative controls without template DNA (H_2O) using the U2-5′ oligonucleotide, U2-3′ oligonucleotide. or U2-5′ oligonucleotide and U2-3′ oligonucleotide together. Lanes (4-6) are the same oligonucleotide combinations with genomic *S. mansoni* DNA as template. MW are molecular weight markers (1 Kb Ladder(Gibco)).

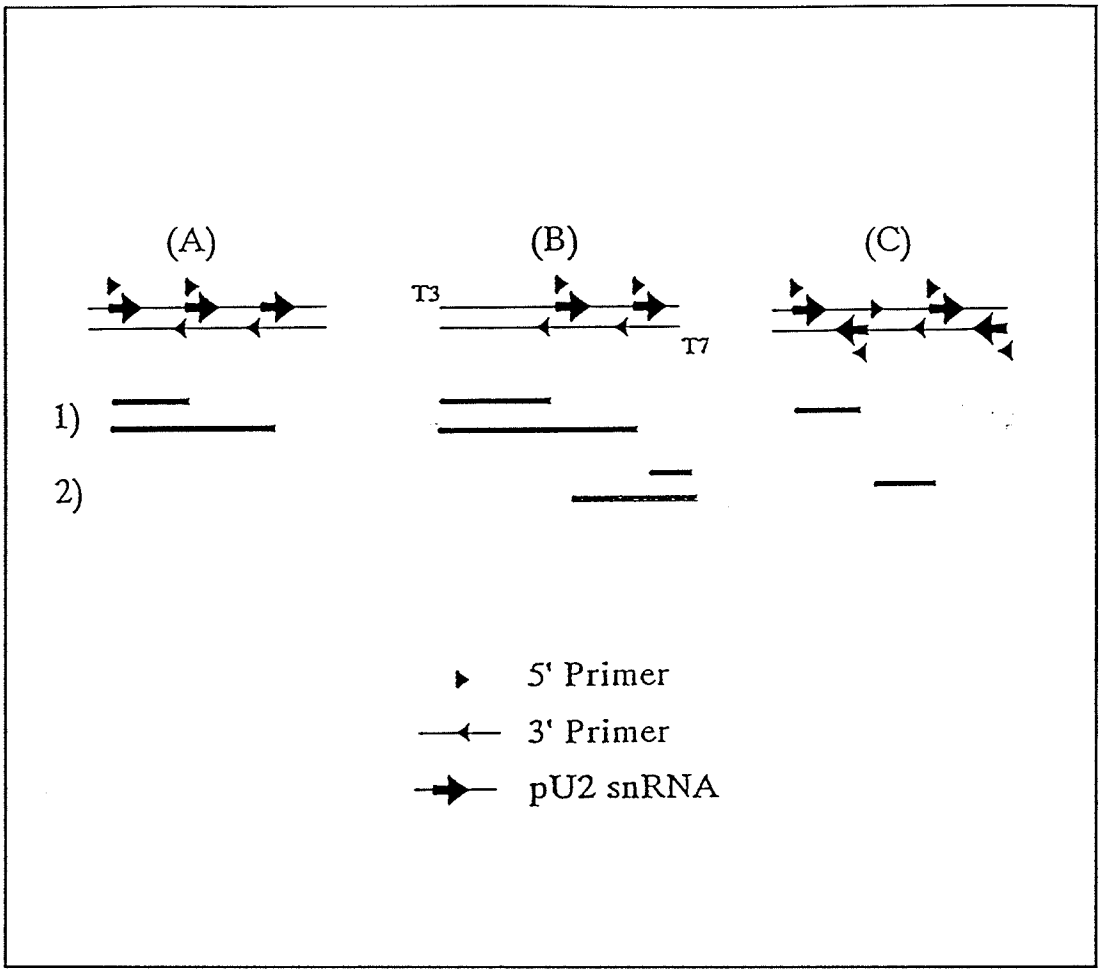


Figure 22: Predicticted End Results of pU2 snRNA PCR

(A) represents the expected products from the pU2 snRNA PCR if genomic organisation of the pU2 snRNA is in a head to tail repeat. The PCR product would include an entire repeat unit, as well one would expect a "ladder" of multiple repeat units being amplified (1). (B) represents the expected end products using the bacteriophage vector (plus insert) as template using the T3 and the 5′ primer (1), or T7 and the 3′ primer. The expected results were not obtained and this led to the testing of the 3′ primer or the 5′ primer alone as amplification primers. The result was PCR product using a single primer. (C) represents the head to head genomic arrangement that is one possible explanation for how PCR product could be generated by either the 3′ (1), or 5′ (2) primer.

Cloning and Characterisation of the U4 snRNA

Evaluation of U4-probe for Screening a Library

To assess the utility of the U4-Probe for screening the *S. mansoni* genomic library a set of Northern and Southern analyses were performed. Initial screening using the U4-probe was unable to detect any *S. mansoni* DNA or RNA. It was established that the U4-probe was not suitable for screening a library, and a second probe was designed (U4(2)-probe). As indicated in Figure #12, this probe clearly recognized a single band of 142 nucleotides in length in the Northern blot of *S. mansoni* A(-) RNA. This indicated that the U4(2) probe recognized a single non-polyadenylylated transcript of 142 nucleotides in length. The size of the RNA fragment recognized was within the range of U4 snRNA's already described[35]. Figure #23 shows the U4(2) oligonucleotide used to probe a Southern blot of genomic *S. mansoni* DNA digested to completion with the restriction endonucleases Sal 1, *Eco* R1, and *Sau* 3A. The probe recognised multiple bands with varying intensities. Similar to the U2-probe, the fragments recognised existed in multiple copies, and these multiple copies may exist in some sort of repeat motif. These data suggested that the probe recognised the U4 snRNA, as well as genomic copies of either the U4 snRNA gene or U4 snRNA pseudogenes and wass therefore judged useful for screening the genomic library.

Screening a Genomic Library for U4 snRNA Clones

With the utility of the probe established by the Northern and Southern analyses, screening of the genomic library was completed. Approximately 400 000 plaque forming

units (PFU's) of recombinant lambda phage were screened using the U4(2) probe. Four positive clones were identified and purified. They were labelled as U4c1 to U4c4. These clones were amplified, grown in large scale, and recombinant phage DNA extracted and purified as previously described.

Mapping of U4 Genomic Clones

The insert of the U4ç3 clone was characterised by agarose gel electrophoresis. The insert was 15 KB in length. Restriction enzyme mapping of this insert was carried out as previously described. The map of U4c3 can be seen in figure #24. This map was used to identify a region suitable for subcloning and sequencing. A single 0.7 Kb *Sau* 3A fragment was targeted for subcloning. This fragment was isolated and subcloned into the sequencing vector pBluescribe SK+. Recombinant plasmids containing the insert of interest were isolated as described.

Sequence Analysis and Identification of a Putative U4 snRNA

The clones resulting from the ligation of the inserts into pBluescript SK+ were sequenced as described using the T3 or T7 oligonucleotides as primers (see figure #25). The sequence of this insert was used to identify regions corresponding to the U4-Probe oligonucleotide. Knowing the predicted size of the U4 snRNA from the northern analysis data, and knowing the sequence recognized by the U4-probe, the sequence of a putative U4 snRNA was determined (figure #26).

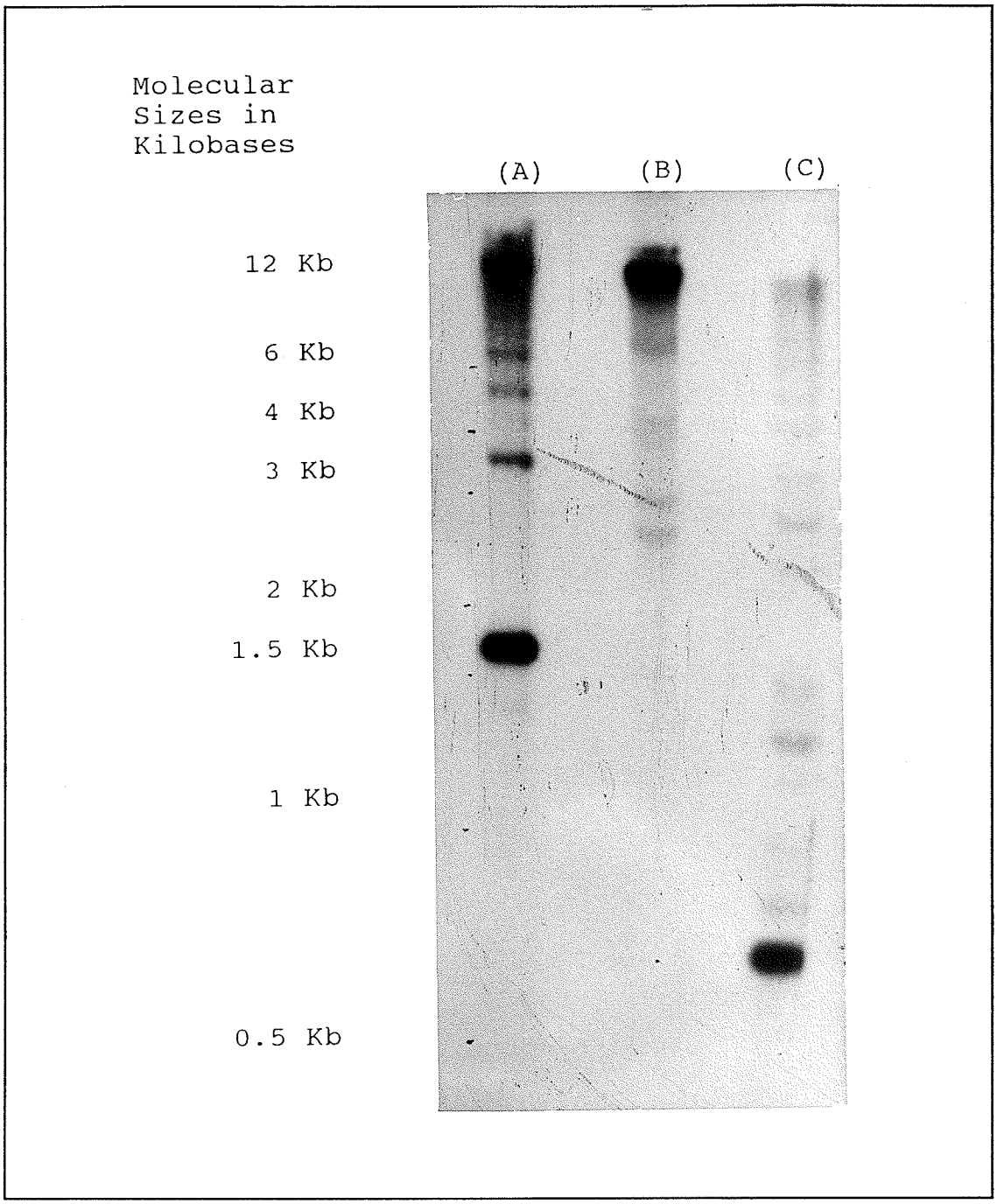


Figure 23: Southern Analysis of U4 probe

2 μgrams of DNA was digested to completion with restriction enzymes, Lane A: *Sal* 1, Lane B: *Eco* R1, Lane C: *Sau* 3A. The DNA was resolved on an agarose gel and transferred. The blot was then probed with the U4(2) oligonucleotide probe. Numbers at left are MW markers.

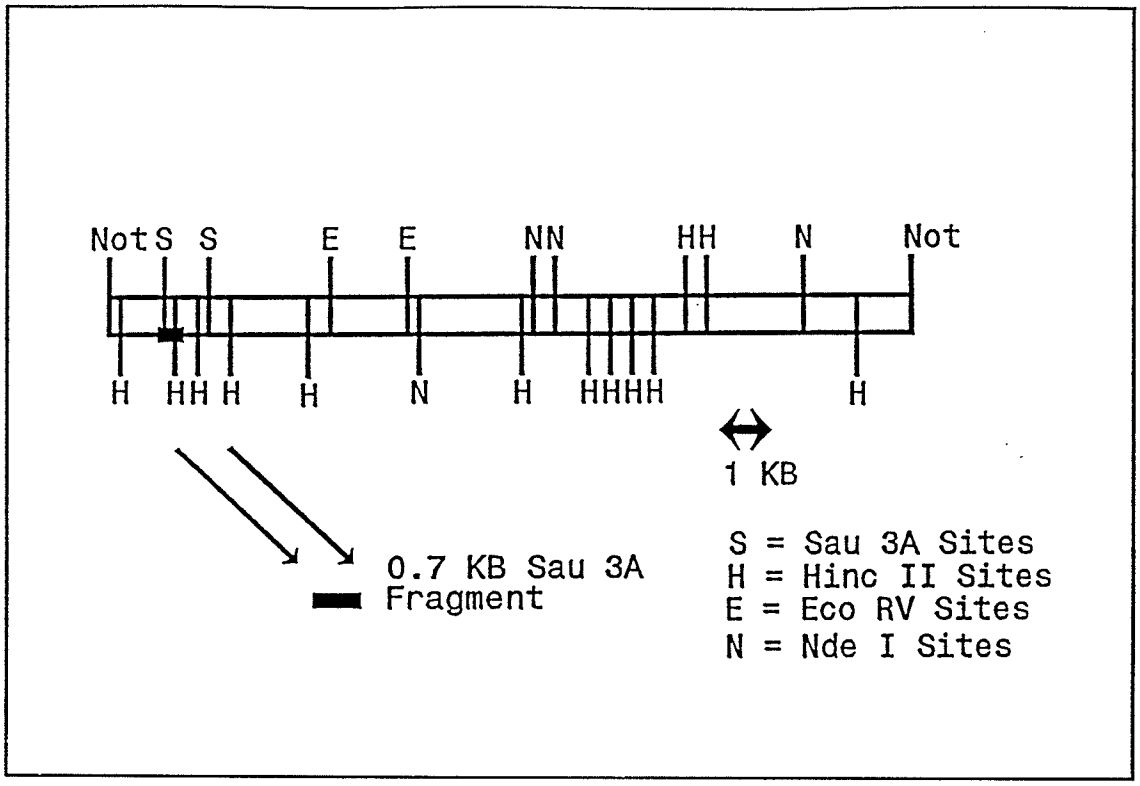


Figure 24: Restriction Map of U4c3 Lambda Phage Insert (15 Kb)

The restriction map of U4c3 is indicated above, the letters indicate the specific restriction enzyme sites, while the dark box indicates the region of the insert recognised by the U4(2)-Probe. Regions targeted for subcloning are indicated by the arrows (not to scale).

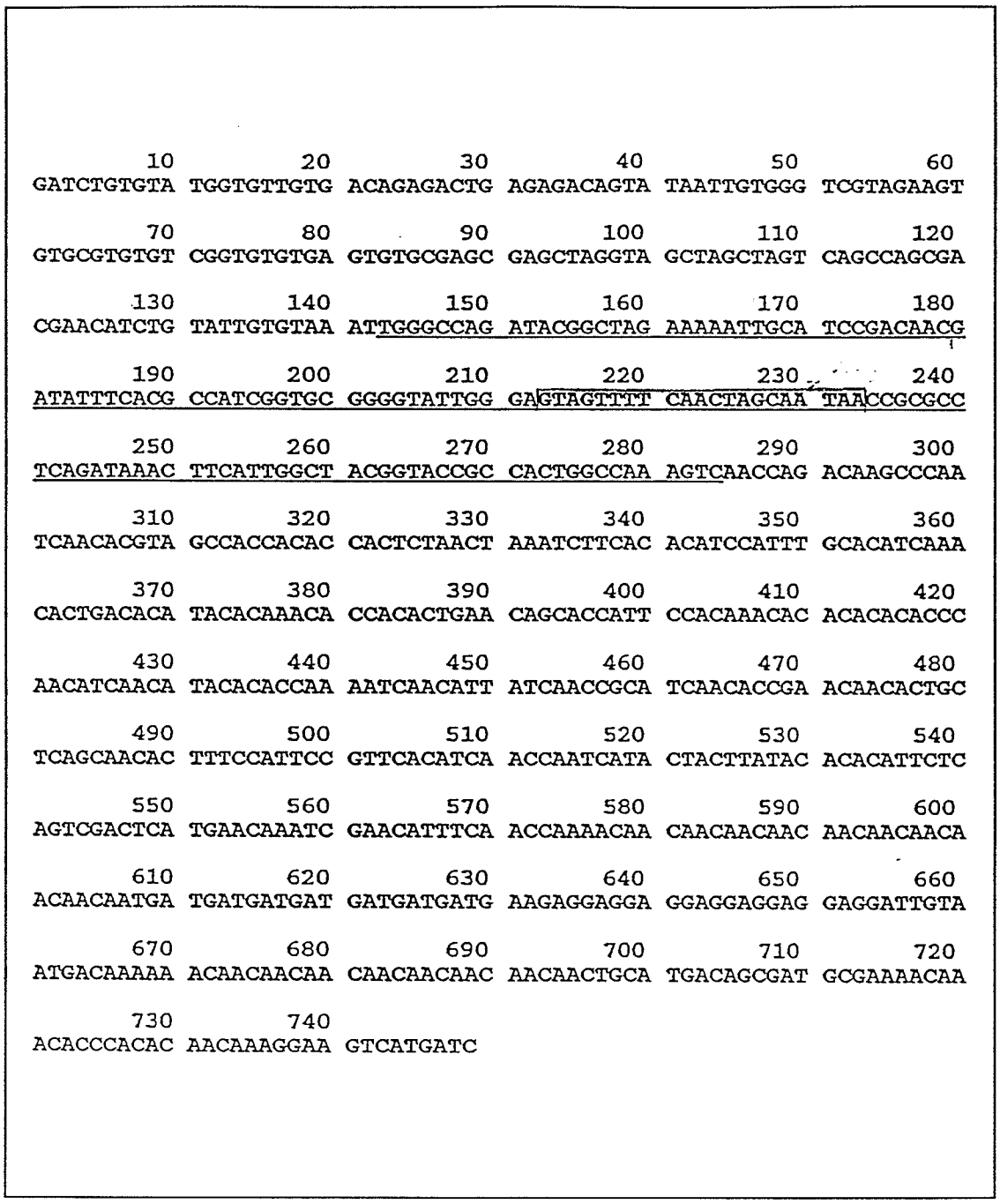


Figure 25: Nucleotide Sequence of U4c3 0.7 Kb *Sau* 3A Fragment

Underlined regions are sequences complementary to the pU4 snRNA. The region is complementary to nucleotides 142 to 1 of the pU4 snRNA. Boxed regions correspond to the sequence of the screening oligonucleotide.

pU4snRNA

```
        10          20          30          40          50          60
GACUUUGGCC AGUGGCGGUA CCGUAGCCAA UGAAGUUUAU CUGAGGCGCG GUUAUUGCUA

        70          80          90         100         110         120
GUUGAAAACU ACUCCCAAUA CCCCGCACCG AUGGCGUGAA AUAUCGUUGU CGGAUGCAAU

       130         140
UUUUCUAGCC GUAUCUGGCC CA
```

Figure 26: Nucleotide Sequence of Putative *S. mansoni* U4 snRNA

Underlined sequence represents the putative Sm site.

Secondary Structure Analysis of pU4 snRNA

As the U4 snRNA is most commonly found complexed with the U6 snRNA the secondary structure and possible folding patterns of the pU4 snRNA were determined in conjunction with the pU6 snRNA (discussed below).

Cloning and Characterisation of the U6 snRNA

Evaluation of U6-probe for Screening a Library

To assess the utility of the U6-Probe for screening the *S. mansoni* genomic library a set of Northern and Southern analyses were performed. As indicated in Figure #12 the probe clearly recognized a single band of 104 nucleotides in the Northern blot of *S. mansoni* A(-) RNA. This indicated that the U6 probe recognized a single non-polyadenylylated transcript of 104 nucleotides in length. The size of the RNA fragment recognized was within the range of U6 snRNA's already described[35]. Figure #27 indicates the U6-probe oligonucleotide used in a Southern blot of genomic *S. mansoni* DNA digested to completion with the restriction endonucleases; *Sal* 1, *Eco* R1 and *Sau* 3A. The probe recognized multiple bands with varying intensities.

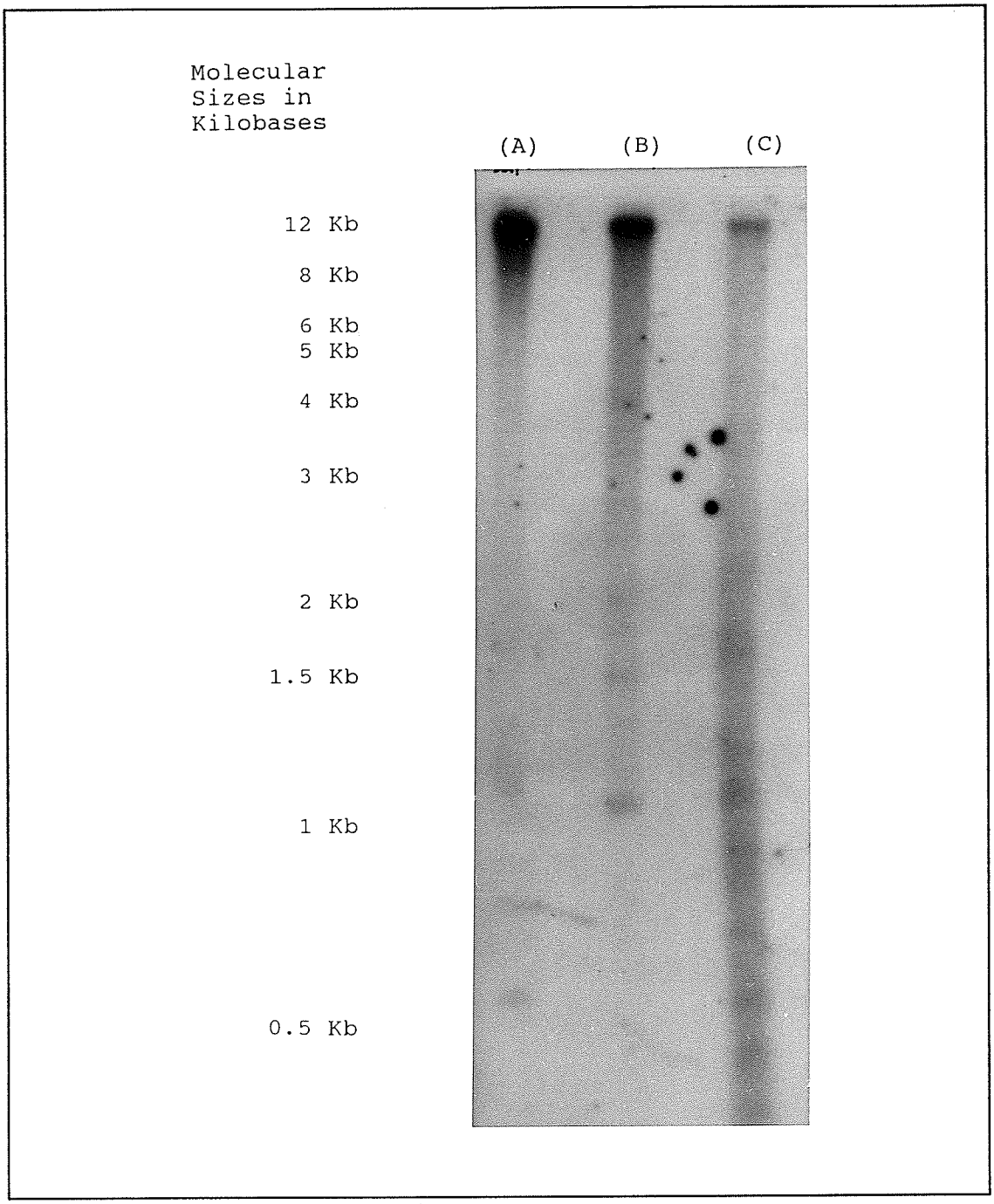


Figure 27: Southern analysis of U6 Probe

2 μgrams of DNA was digested to completion with the restriction enzymes; Lane A: *Sal* 1, Lane B: *Eco* R1, Lane C: *Sau* 3A. The DNA was resolved on an agarose gel and transferred. The blot was then probed with the indicated oligonucleotide probe. Numbers at left are Molecular weight markers.

Similar to the data from other snRNA's, these data suggested that the sequences recognised existed in multicopy, and that these sequences may exist in some sort of repeat motif. These data also suggested that the U6-probe recognised the U6 snRNA, as well as genomic copies of either the U6 snRNA gene or U6 snRNA psuedogenes and was therefore useful for screening the genomic library.

Screening a Genomic Library for U6 snRNA clones

With the utility of the probe established by the Northern and Southern analyses screening of the constructed genomic library was completed. Approximately 300 000 plaque forming units (PFU's) of recombinant lambda phage were screened using the U6-Probe. Six positive clones were identified and purified. They were labelled as U6c1 to U6c6. These clones were amplified, grown in large scale, and recombinant phage DNA extracted and purified as previously described.

Mapping of U6 Genomic Clones

The insert of the U6c3 clone was characterised by agarose gel electrophoresis. The insert was 13 KB in length. Restriction enzyme mapping of this insert was carried out as previously described. The map of U6c4 is presented in figure #28. This map was used to identify regions suitable for subcloning and sequencing. A single 1.4 Kb *Hinc* II/*Not* I fragment was targeted for subcloning. This fragment was isolated and subcloned into the sequencing vector pBluescribe SK+. Recombinant plasmids were isolated as already described.

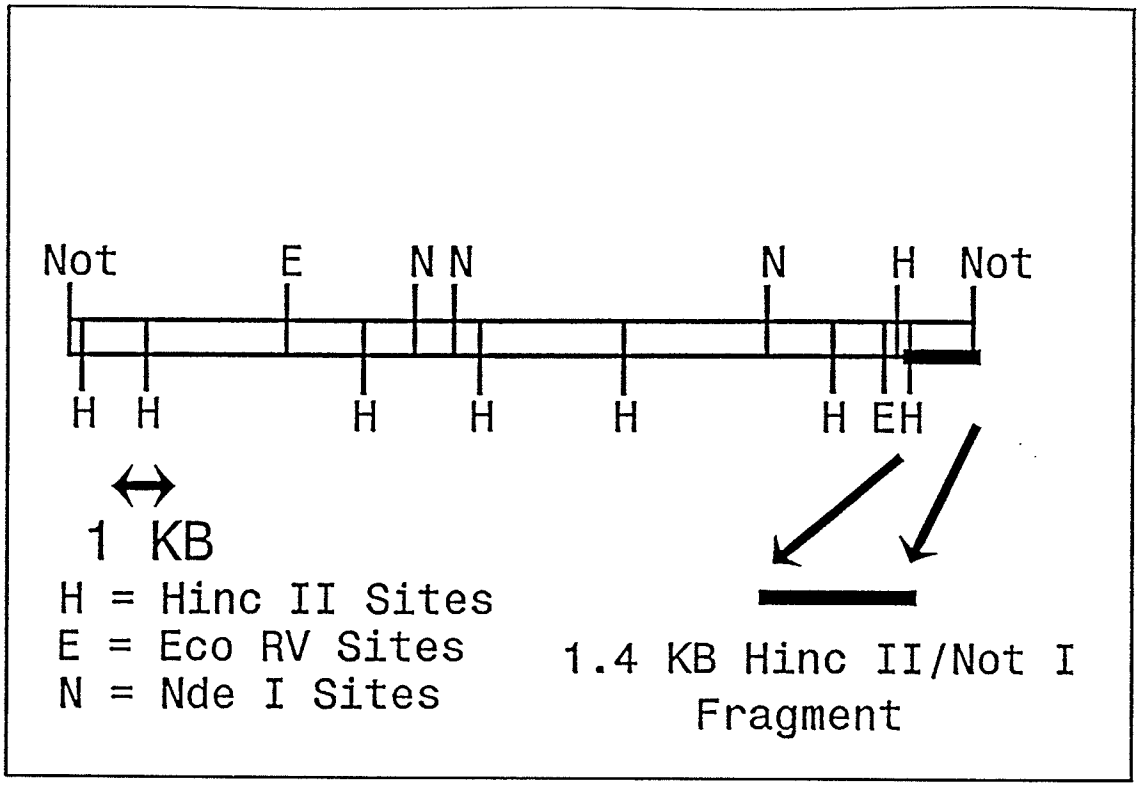


Figure 28: Restriction Map of U6c3 Lambda Phage Insert (13 Kb)

The restriction map of U6c3 is indicated above, the letters indicate the specific restriction enzyme sites, while the dark box indicates the region of the insert recognised by the U6-Probe. Regions targeted for subcloning are indicated by the arrows (not to scale).

Sequence Analysis and Identification of a Putative U6 snRNA

Clones resulting from the ligation of the 1.4 Kb *Hinc* II/*Not* I insert into pBluescribe were sequenced as describe above using the T3 or T7 oligonucleotides as primers (see figure #29). The sequence of this insert was used to identify regions corresponding to the U6-Probe oligonucleotide. Knowing the predicted size of the U6 snRNA from the northern analysis data, and knowing the location recognised by the U6-probe, the sequence of a putative U6 snRNA was determined.

Primer Extension Analysis of U6-probe

In an attempt to generate some preliminary data it was decided to perform primer extension sequencing using the U6-probe oligonucleotide as a primer. The primer extension analysis using the U6-Probe was expected to produce sequence for the first 77-78 nucleotides of the U6 snRNA at the 5′ end, based upon the binding position of the U6-probe. However due to numerous enzyme "crashes" readable sequence was only generated for the first 22 nucleotides at the 5′ end of the U6 molecule (indicated in figure #18). Again this extension sequence data validated the U6 subclone (figure #29), and confirmed the sequence of the 5′ end of the pU6 snRNA (figure #30).

```
        10          20          30          40          50          60
ATCGTGTAGC TCACTCGTGG TGTATCAGGG AGGATGGAGA GTGAGTGAAC TCTGTGGCAT

        70          80          90         100         110         120
CAGCATCACA CTGAGTGGTT GTGGTGACAG GGAGTGTGGA TTTGGGAGAT GAGAATGGAG

       130         140         150         160         170         180
TGTGGAAAAT ATGGAACCCT TCACGATTTT CCGTGTCATC CTTACCACCA GGGGCCATGC

       190         200         210         220         230         240
TAATCTTCTC TGTATCGTTC CAATTTTAGT ATATGTACCG CCGAAGCAGT ACGGGTTTCG

       250         260         270         280         290         300
CACATCACTA TCTCACTATC AGTCCATACA TGTGGTTTGT CTCGCACACA CACTCATGTG

       310         320         330         340         350         360
TACATCTCAC ATTGTTTTTC TCTACATCAG NTCCAGTCAG CATACACACA CATCAATTTG

       370         380         390         400         410         420
CATGACTCCC ANCACAAAAT ACAACAACCC CCTGCACCAT CATCAACACA CAATTTCAGA

       430         440         450         460         470         480
ATCTACAAAT CAATCACTGC CCTCACCCAC TTCTATTGCA GTAGGGTNGC CACCCACATC

       490         500         510         520         530         540
TTTTTTTAGC CCCAAGCCGT TCGACGTCCA ATCAANCATA CAAATTANCA ATCATATAAC

       550         560         570         580         590         600
ACACAGGTTC ATTGGATCAA CGGTGTGGGC CATAACCANA AAATTCATAG CAGTTTCGCT

       610         620         630         640         650         660
CAGAGAGGGT AGTGAACTAA TGCATANATA TACCATCATT CACTCCTACC ACTAGAGGGA

       670         680         690         700         710         720
TTCTCAGGTA TGTACCGTTT CATCCACTTT AACCGGCGGC AGATGTTTTA AATGAGTCCC

       730         740         750         760         770         780
AATAGGCNGA ATTCGAAGTG GGCAGCCGTC AATTTTACCT CTGGGGTAGA TAATGGTTTC

       790         800         810         820         830         840
ATTTTTTTAG GGGGGTTGTT GGGACTACGC ATAACCCTTG AAAGATTATC CAATCCGCCT

       850         860         870         880         890         900
TGGGCCAAGC CAAGACAACT TAAAGATATT GGCGGAATAT CAAAGCGCGA TCCATCGAGT

       910         920         930         940         950         960
CCATTCATGC AGNCCGAGAG ATTTTTTTAA CTATAAACCG CTGATGGAAG CGTTTATGCG

       970         980         990        1000        1010        1020
GGAAGAGGTA AAGCCCTTCC CGAGTAACAC AAATCAACAG CATAAATAAC CCCGCTCTTA

      1030        1040        1050        1060        1070        1080
CACATTCCAC CNCTGAGAGA GGGCATCACA TTAAACCACA CCTATGGTGT ATGCATTTAT
```

Figure 29A: Nucleotide Sequence of U6c3 1.4 Kb *Hinc* II/*Not* I Fragment

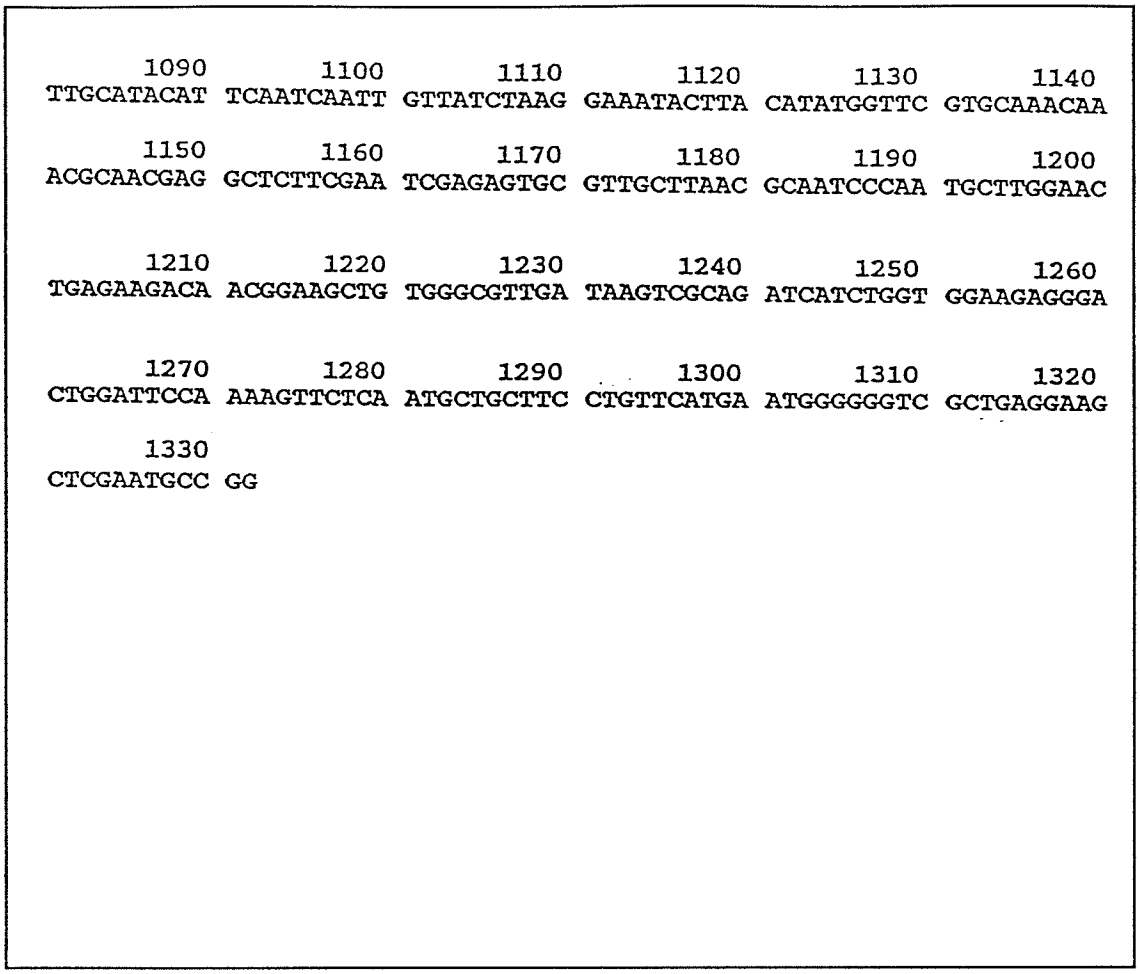


Figure 29: Nucleotide Sequence of U6c3 1.4 Kb *Hinc* II/*Not* I fragment

Underlined regions are sequence complementary to the pU6 snRNA. The region is complementary to nucleotides 104 to 1 of the pU6 snRNA. Boxed regions correspond to the sequence of the screening oligonucleotide, while regions in italics (208-229) are complementary to the primer extension sequence data.

Secondary Structure Analysis of pU6 snRNA

The secondary structure and possible folding patterns of the pU4/pU6 snRNA complex was determined by comparison to previously published U6-snRNA folding patterns, as well as computer assisted modelling based on the method of Zucker[90]. As can be seen in figure #30 the pU4 snRNA could be folded with a number of stem-loop structures. The overall shape of this folded pU4 is strongly suggestive of the folding pattern of U4-snRNA's from other species. In addition to this structural analysis the sequence of the U4-snRNA was compared to published sequences of the U4-snRNA's from other species. As also can be seen in figure #30 the pU6 snRNA could also be folded with a number of stem-loop structures. The overall shape of this folded pU6 is strongly suggestive of the folding pattern of U6 snRNA's from other species. In addition to this structural analysis the sequence of the U4 and U6 snRNA's were compared to published sequences of the U4 and U6 snRNA's from other species (figure #31 and #32).

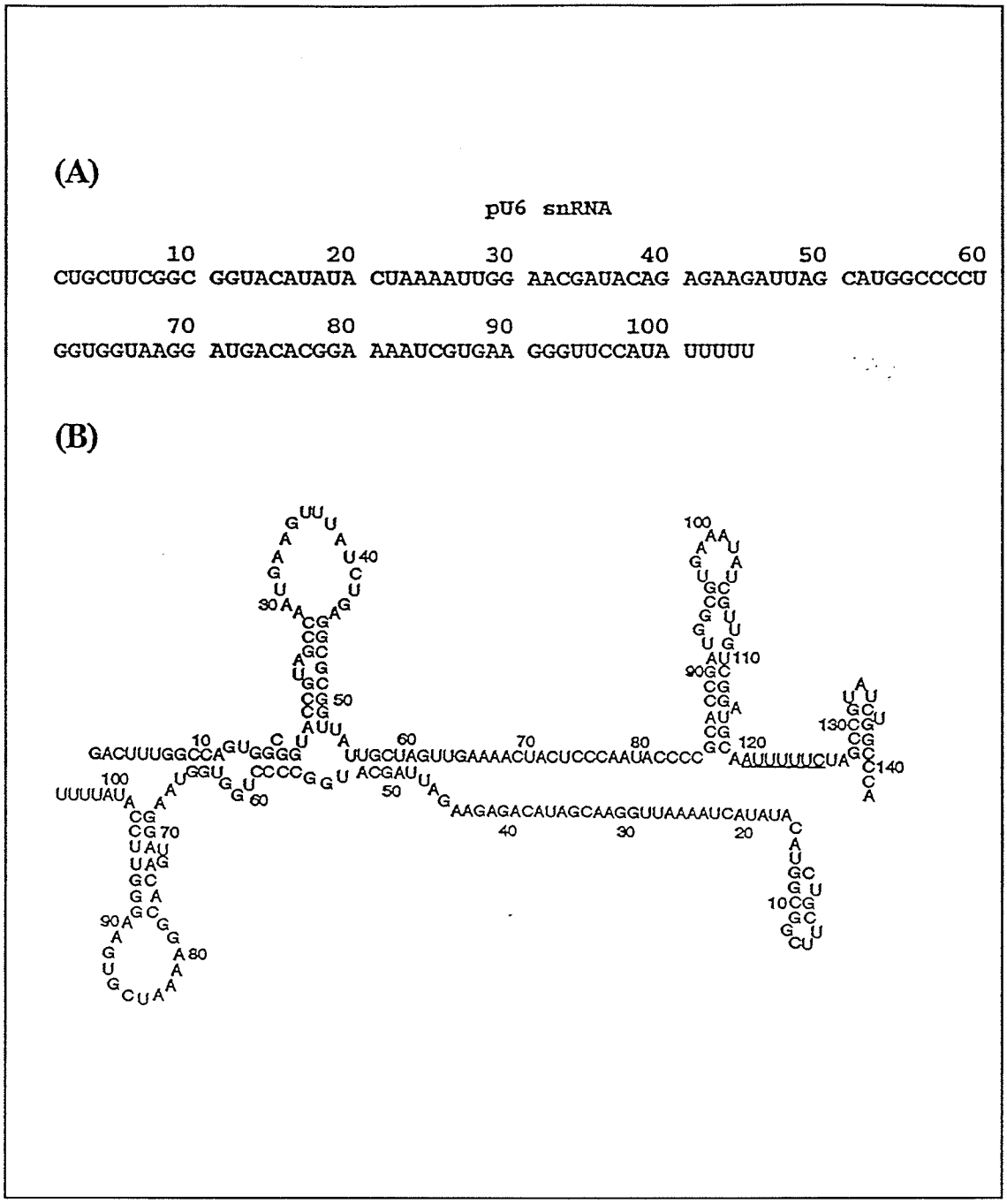


Figure 30: Nucleotide Sequence of pU6 snRNA and Potential Secondary Structure of pU4 and pU6 snRNA's

(A) Shows the sequence for the pU6 snRNA. Boxed sequence is confirmed by primer extension analysis, while underlined sequence is the putative Sm site. (B) Shows the predicted secondary structure of the pU4/6 snRNA's.

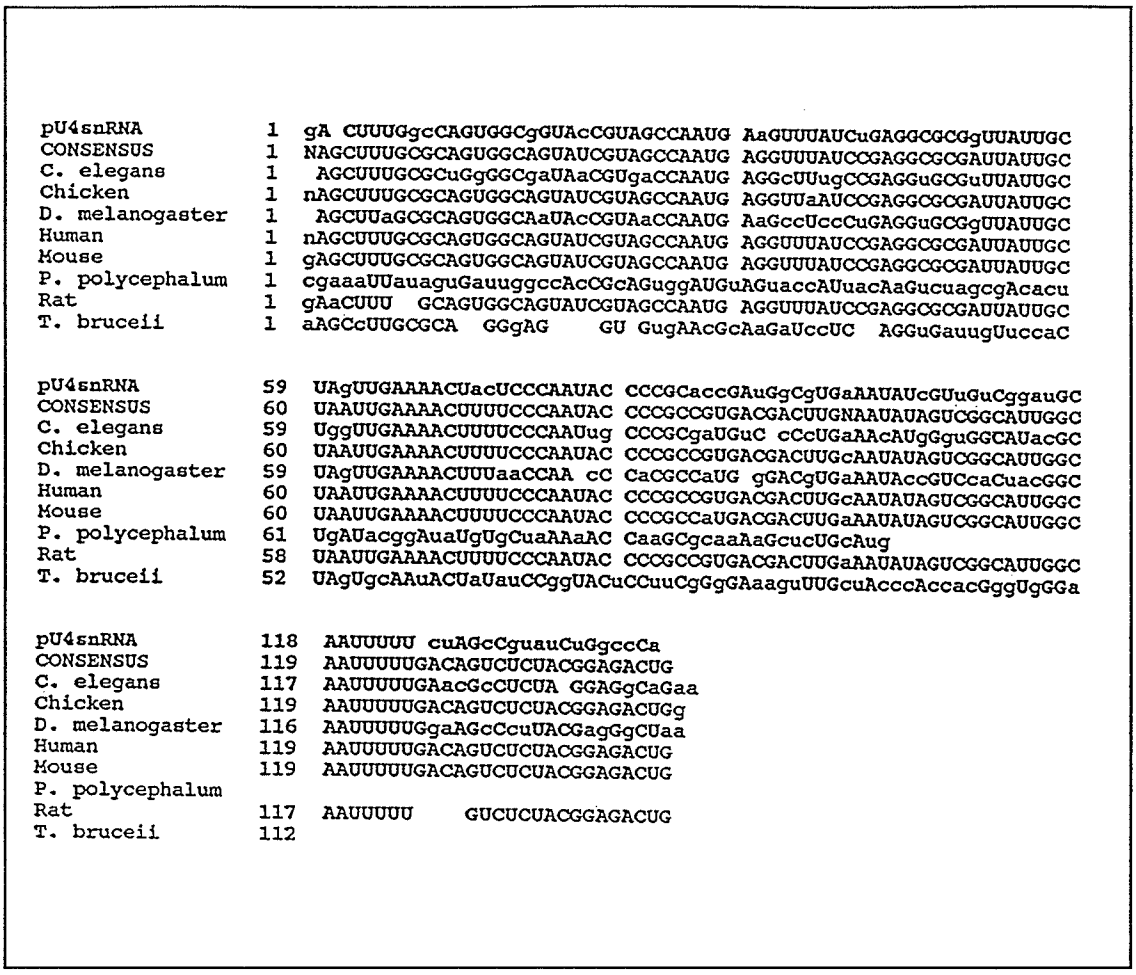


Figure 31: Sequence Alignments of pU4 snRNA and Published U4 snRNA's

A comparison of the pU4 snRNA (Top) with previously published U4 snRNA's. Sequences in agreement with the pU4 are in upper case, while sequences in disagreement are in lower case.

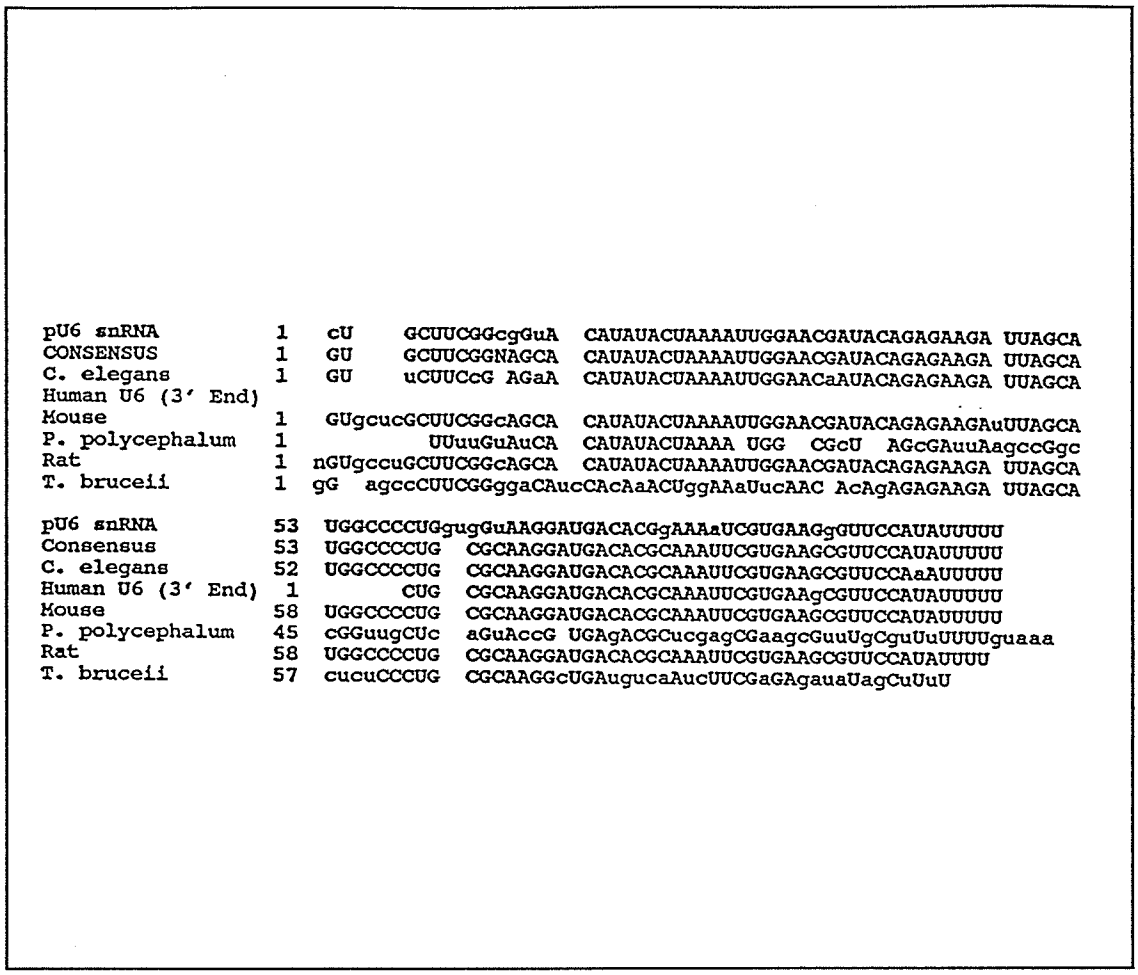


Figure 32: Sequence Alignments of pU6 snRNA and Published U6 snRNA's

A comparison of the pU6 snRNA (Top) with previously published U6 snRNA's. Sequences in agreement with the pU6 are in upper case, while sequences in disagreement are in lower case.

Discussion

Identification of U1, U2, U4 and U6 snRNA's in S. mansoni

The cloning and characterisation of the U1, U2, U4 and U6 snRNA's were carried out to help understand the mechanism by which certain transcripts are selected for trans-splicing, as well as to further understand spliceosome formation and function in *S. mansoni*. Based upon the current ideas for the roles of these snRNA's in splice site selection, an understanding of the snRNA's from *S. mansoni* may provide some understanding of the splice site selection process in a trans-splicing reaction. Oligonucleotide probes were constructed based upon published sequences of snRNA's from other species and upon conserved areas of the molecules believed to be important in their function. To insure that these oligonucleotides did indeed recognise potential snRNA transcripts, as well as potential snRNA genes, a series of Northern and Southern analysis were performed. These results clearly indicated two things; one, that a single non-polyadenylylated A(-) transcript was recognised by each specific oligonucleotide probe, and the size of the recognised fragments was similar to the size of the snRNA's from other species, secondly, it was obvious that multiple genomic copies of this transcript were detected in restricted *S. mansoni* DNA. This provides strongly suggestive evidence that these probes recognised the *S. mansoni* analogs of the U1, U2, U4 and U6 snRNA's. The data also provided some information about the possible genomic organisation of these snRNA genes. The recognition of multiple bands in the Southern analysis suggested that these snRNA genes exist in the *S. mansoni* genome in multiple

copies. This is in agreement with findings on the snRNA's from other species. For example, the snRNA's in *Xenopus* and *Psammechinus* have been shown to be encoded by multiple copies of snRNA genes in multicopy repeat arrays arranged in a head to head or head to tail array. However, the human genome contains multiple copies of its snRNA's genes randomly scattered through the genome[35]. The banding pattern of the restricted *S. mansoni* DNA fragments recognized by the oligonucleotide probes (the Southern blot data) also provided some suggestive evidence about the potential genomic organisation of these snRNA genes. Certain bands recognized by the U2 probe for example, appear to be more intense than other fragments within the same digest. This suggested the possibility that the sequence elements recognised may exist in some sort of repetitive motif. For example, the *Sau* 3A digest shows a very intense band at about 1 Kb in length (a larger number of *Sau* 3A restriction fragments of that specific size being recognized) while the other bands in the *Eco* R1 and *Sma* 1 digests are less intense and in a "smearing" pattern (a fewer number of different sized fragments being recognized). If the sequence recognised by the probe was present in the genome as some sort of multi-copy repeat unit, and if there was a *Sau* 3A site within this repeat unit, then this binding pattern would be predicted on a Southern blot of genomic DNA. Other snRNA genes have been shown to exist in multiple copies in other eukaryotic organisms, and in a number of cases these multiple copies have been shown to be in repeat unit arrays[35]. This possible organization was noted especially for the U2, U4, and U6 snRNA's. The identification of two partial pU2 snRNA sequences in the 1.0 Kb Xho-1 fragment was suggestive of this type of genomic organisation. These two partial

sequences seemed to be copies of two pU2 snRNA genes encoded in a head-to-tail array on the same DNA strand. This led to the attempt to PCR amplify a complete "repeat" unit using primers that would recognise the snRNA and adjacent regions. Final results however suggested an alternative genomic arrangement. PCR amplification was attempted using single primers to investigate this possibility. The presence of the resulting products from a single primer suggested that the pU2 snRNA gene lay not in a head to tail array, but in a head to head array, encoded on opposing strands. This is the only possible explanation (barring experimental error) of how one could obtain PCR product with a single primer. This finding did not agree with the sequence data from the *Xho* I fragment (which had two pU2 snRNA genes in a head-to-tail array on the same DNA strand). One possible explanation is that the *Xho* I fragment did not accurately reflect the most prevalent genomic organisation in this organism. A number of other organisms that encode genes in repeat arrays have been shown to also encode the same genes outside of these repeat arrays (so-called "orphon" genes)[5]. Another explanation is that this fragment may encode a U2 snRNA and some sort of U2 snRNA pseudogene, or even two U2 snRNA psuedogenes. In any case the PCR data suggests that at least some portion of the U2 snRNA or U2 pseudogenes are organised in a head to head array; it is possible that some of the U2 snRNA genes (or pseudogenes) do not follow this configuration. As there is little data about the transcription of *S. mansoni* genes it is uncertain as to which type of organisation would be preferentially used by the actively transcribed genes, if any.

One possibility that was considered during the cloning of these snRNA's was that the oligonucleotide probes used may recognise non-transcribed snRNA genes (ie. snRNA pseudogenes). This potential problem was to be addressed after the sequencing of any snRNA's was complete. There is no guarantee that any clones identified are actually transcribed to produce functional snRNA's. Further evidence would have to be presented to clearly indicate that an RNA transcript would have sequence identical to the putative snRNA genes already identified. This work would involve generation of primers complementary to the 3′ ends of any snRNA's sequenced. These primers would be used in primer-extension sequencing as already described. The preliminary primer extension data were able to confirm the sequence of the 5′ end of the pU2 and pU6 snRNA's. Full length primer extension data may be able to be generated by 3′ primers and would generate sequence directly from RNA transcripts. This would not be proof that these putative snRNA genes are expressed, but it would give suggestive evidence that these (or similar) snRNA genes may be expressed. The main focus of this project was to generate the sequence of the *S. mansoni* snRNA's, this goal would be obtained by full length primer extension data. To definitively prove that a certain gene is expressed a much more detailed investigation of *S. mansoni* transcription factors would have to be undertaken.

Subcloning of the U1, U2, U4 and U6 snRNA's

A clone for each snRNA was characterised by restriction endonuclease mapping to identify restriction fragments suitable for subcloning into a sequencing vector. This subcloning was performed for the U2, U4 and U6 snRNA's, however the U1 snRNA

could not be readily subcloned. Any positive colonies recovered during the subcloning process were recombinants containing both *S. mansoni* and Lambda bacteriophage DNA. The most likely explanation for this finding was that some sort of recombination event occurred sometime during subcloning or library construction. The rearrangement most likely occurred during the subcloning process for a number of reasons. The first reason is that restriction digests of positive lambda phage clones were always able to liberate an intact *S. mansoni* insert suggesting that no rearrangement of *S. mansoni* and phage DNA was occurring during the phage growth, or phage DNA isolation. Further evidence that this arrangement occurred during subcloning lies in the fact that the *E. coli* host strain used (DH5α) during the subcloning process is not deficient in the same recombination mechanisms that the *E. coli* SRB strain (used for phage growth and screening) is. It is not unreasonable to assume that DNA crossover or some other sort of mechanism may have accounted for these events during the subcloning process. These problems have been observed in the cloning of other eukaryotic genes. The presence of repeated sequence elements (such as AT rich regions) or possible sequence homologies between *S. mansoni* and lambda phage could be responsible for these recombination events. Recombination events due to repeated or related sequence elements or some other mechanism has led to the identification of "unclonable elements" in some eukaryotic organisms[91]. However, further investigation into these rearrangements in our cloning system was not pursued.

Identification and Sequence Analysis of pU2, pU4 and pU6 snRNA's

The subcloned fragments of genomic DNA were sequenced and the putative snRNA's identified based upon; the size of the *S. mansoni* snRNA's (northern data), the identification of the site that the snRNA probe would recognise on a consensus snRNA, and the primer extension data for the 5′ end of the U2 and U6 snRNA's. Repeated attempts at obtaining primer extension sequence data from the U1 and U4(2) probe oligonucleotide were unsuccessful. This may be due to incomplete recognition of the U4 snRNA by the degenerate U4(2)-probe and the inability of the reverse transcriptase enzyme to generate extension product if the primer is improperly annealled. The failure of the U1 probe to be extended is probably due to the design of the U1 probe. The sequence used to generate the U1 probe is at the very 5′ end of the U1 consensus molecule, and therefore would not be able to generate primer extension product. In any case it was possible to identify putative snRNA's for the U2, U4 and U6 snRNA's. The sequence of the pU2 snRNA showed a degree of homology to published U2 snRNA's ranging from 64-71 percent. In addition, the probable secondary structure of the pU2 snRNA is highly conserved compared to other U2 snRNA's. Among the regions of high conservation were a number of regions identified as being important in the intact spliceosome. These areas of high conservation include the majority of the 5′ end of the molecule (containing stem-loops I and II) which contain sequences thought to be important in recognising the branch point acceptor region. The pU2 snRNA contained sequence perfectly complementary to the highly conserved branch point acceptor consensus sequence. This region that has been shown to be necessary for the recognition

of the intron to be removed and formation of the intact spliceosome in a number of organisms. Another region of the pU2 snRNA that was shown to be conserved was the sequence surrounding the Sm site (necessary for the attachment of the "core" proteins). The identification of these points are all suggestive evidence that we have identified the U2 analog in *S. mansoni*, and this putative U2 may function similarly to the U2 from other species. Areas of divergence were also noted, which mainly include unpaired "loop" regions of stem-loop structures, and regions near the 3′ end of the molecule. One area of divergence that was notable, however, was the apparent Sm site of the pU2 snRNA. This site differs from the highly conserved Sm site of other eukaryotic organisms by having an Sm sequence of PyAUUUUUCPu, rather than the highly conserved consensus of $PuAU_{3-4}NUGPu$. What role this modified Sm site may play in the formation of a spliceosome is uncertain, as is the effect of areas of sequence divergence of the pU2 snRNA. Sequence analysis of the pU4 snRNA showed sequence homology of between 72-81% to published U4 snRNA's with considerable homology shown throughout the snRNA. The secondary structure of the pU4 snRNA was also highly conserved compared to published structural data. There were areas of sequence divergence, with the main areas being regions paired with the U6 snRNA as well as unpaired "loop" sections. Again the pU4 snRNA had a modified Sm consensus sequence of PuAUUUUUCPy compared to the highly conserved consensus of $PuAU_{3-4}NAGPu$. Sequence analysis of the pU6 snRNA showed sequence homology of between 80-91% to published U6 snRNA's with homology shown throughout the pU6 snRNA. This high degree of homology is not surprising considering the postulated role of the U6 snRNA

as a catalytic element. Regions believed to be important in the U6 snRNA's function were all present in the pU6 snRNA including regions important in interacting with the splice sites, such as the poly Uridine region at the 3′ end. The secondary structure of the pU6 snRNA was also highly conserved compared to published structural data. There were no great areas of sequence divergence noted. But some of the more notable were regions involved in base pairing with the pU4 snRNA (which in most cases were complemented by comparable changes in the pU4 snRNA), as well as potential "loop" structures.

Potential Mechanisms for snRNA Control of Selection of Trans-Splicing

As was discussed above it is apparent that there is a relatively high degree of homology between the snRNA's from *S. mansoni* and the snRNA's of other organisms that carry out either cis splicing, trans-splicing or both. This is not surprising as it has been shown that formation of the spliceosome (either a cis or trans spliceosome) is a highly conserved process among almost all eukaryotes that carry out mRNA processing reactions. The high degree of conservation was found in organisms ranging from the simplest yeasts to higher order eukaryotes such as mammals. It is not surprising therefore, that trematodes (ie *S. mansoni*) would therefore share a great number of similarities in their RNA-processing machinery with evolutionarily related organisms such as the nematodes, or even share certain characteristics with the more advanced eukaryotes. Therefore a relatively high degree of conservation of parts of the splicing machinery is to be expected, this degree of conservation helps to suggest that indeed the

true analogs of the *S. mansoni* snRNA's have been identified. This degree of conservation is especially necessary if the organism in question carries out "normal" cis-splicing as well as the more rare trans-splicing.

Divergent sequence elements were noted in these snRNA's compared to the consensus sequences generated. These differences may be simply due to the evolutionary distance between *S. mansoni* and the organisms used to generate the consensus, or these changes may have an effect on the control and regulation used to select a transcript to be trans-spliced. An examination of the specific roles played by the putative snRNA's studied above may provide more of an answer as to their possible involvement in targeting a specific transcript to be either cis or trans spliced.

It would have been perhaps most interesting to generate the sequence of an *S. mansoni* U1 snRNA as its primary role in formation of a spliceosome appears to be identification of the 5′ splice site on a transcript to be spliced. This may have provided insight into a mechanism to select a trans-spliced target transcript. It could be speculated that the sequence of the Sm site on the *S. mansoni* U1 snRNA may suggest a mechanism for selection of a cis-spliced product. It is not unreasonable to assume that the U1 snRNA may have a similar consensus sequence to the modified SM consensus found on the pU2 and pU6 snRNA's. If the U1 Sm site differed significantly from the SM consensus found on the SL-RNA then this sequence may indicate a different or modified set of "core" or accessory proteins that bind to each particle. A "U1-core" could be responsible for recognising a cis-spliced transcript, while the "core" proteins that form around the SL-RNA might select a transcript for trans-splicing. This proposed mechanism is supported

by an examination of the sequence of the SL RNA's of different schistosome species. The consensus Sm site of different schistosome species contains a cytosine residue in the middle of the poly-uridine sequence. This cytosine may help play a role in the selection of a trans-splice site. This selection may be mediated by the altered splice leader Sm site directly, or more likely be mediated by an altered set of "core" or even SL-specific or U1-specific accessory proteins. This is pure speculation however, and the mechanism for selection may lie within the other snRNA's that form the intact spliceosome or with some as yet unknown mechanism.

The U2 snRNA is thought to be involved in the proper alignment of the two splice sites in a target transcript. The regulation of such splice site alignment may allow for some manner of control or regulation of which transcripts are targeted for trans-splicing. The differences between consensus U2 snRNA's and the S. mansoni pU2 snRNA lie mainly in exposed "loop" structures as well as in the slightly longer 3′ end of this molecule. If the pU2 snRNA is responsible for the determination of whether a transcript is cis or trans-spliced, then these areas may play a role in such determination. If such a mechanism of control exists, it is unknown whether this recognition of a specific trans-spliced target would be mediated through the snRNA directly, or through some sort of accessory proteins.

Of the snRNA's that have been examined perhaps the least likely involved in selection of cis and trans spliced transcripts would be the U4/U6 snRNA's. The role of the U4 snRNA is believed to be that of a negative regulator for the activity of the U6 snRNA, as well as acting as the framework around which the U6 snRNA enters the

spliceosome. It is therefore unlikely that this snRNA plays much of a role in the selection process, as the function of the U4 snRNA would not be different in a either a cis or trans-spliceosome. There were not many significant differences between the pU4 snRNA and other snRNA's, but there were some differences noted. What effect these differences may have is still unknown. Of any snRNA the U6 snRNA is least likely to involved in selection of a trans-spliced transcript. This snRNA is believed to be the catalytic subunit necessary for both cis and-trans splicing and is highly conserved in all eukaryotes. This high degree of conservation was also found in the pU6 snRNA from *S. mansoni*, and it is therefore unlikely that the *S. mansoni* pU6 snRNA would play any role in splice site selection.

The changes noted in the putative *S. mansoni* snRNA's are mainly due to sequence differences with snRNA consensus sequences. This sort of sequence divergence has been noted in other organisms, even those that do not carry out trans-splicing reactions. Thus it is impossible to judge from this data what the effects or roles that this sequence divergence may play. The high degree of secondary structure homology suggests that these snRNA's behave very similarly to their counterparts in other organisms. One potentially important difference in the *S. mansoni* snRNA's is their altered Sm binding site. This alteration from the highly conserved consensus Sm site was observed in both putative snRNA's expected to carry this site (pU2 and pU4 snRNA's). Additionally this modified Sm site was noted in a cloned *S. mansoni* U5 snRNA (unpublished data). It is likely that these changes may be conserved among the *S.*

mansoni snRNA's, and may have an effect on the "core" structural proteins that bind to this site. It is unknown whether or not these potentially modified proteins would have any effect on the regulation of trans-splicing.

Conclusion and Further Directions

The goal of this work was to gain an understanding of the control and regulation of trans-splicing in a preliminary manner by first characterising the components involved in the catalysis of the splicing reaction. The most important components of either cis or trans-splicing are the small nuclear ribonucleotide particles (snRNP's) that form the spliceosomal complex. One of the most important components of these snRNP's are the snRNA's that are the active portion of these multimeric molecules. The cloning and characterisation of three putative snRNA's (U2,U4 and U6) from *S. mansoni* was carried out to see if there were any obvious differences between these snRNA's and the snRNA's of other organisms that may allow for the control and regulation of what mRNA in *S. mansoni* are trans-spliced. An examination of both the sequence and the potential secondary structure of these snRNA's revealed a degree of conservation with other snRNA's, as well as a number of differences. The degree of conservation suggests that these snRNA's do function similarly to their counterparts in other organisms. The differences noted in these snRNA's may or may not play any role in the selection of trans-splice sites. If these changes do play a role in splice site selection it is probable that any selection would be mediated by some sort of accessory protein intermediate that would perhaps bind to these altered regions and be able to differentiate between a cis and

trans-spliced substrate. To truly judge if these sequence changes do play a role in selection, it would be necessary to have some sort of *In Vitro* splicing system for *S. mansoni* to test what effects these sequence differences may have. The development of such a system is paramount to test any hypothesis that may arise from any sequence comparison gained from this work. With the basic understanding of an *S. mansoni* splicing system one could continue to further characterise other aspects of this splicing pathway such as the large amount of accessory proteins that are not only involved in splicing activity, but may also play a role in the trans-splicing selection process. While the work presented in this thesis did not directly answer whether or not the snRNA's of *S. mansoni* play a role in the control of selection of trans-splicing it provides the basis upon which other work could be done to answer this question. It would be clearly impossible to make any sort of hypothesis about the control and regulation of the splicing process in *S. mansoni*, without the basic understanding of the components of this splicing system.

This work was partially funded by a grant from the Manitoba Medical Services Foundation Inc. and by a Graduate Fellowship from the Department of Medical Microbiology, University of Manitoba.

Appendix

A) Media Used

i) NZY Medium and Plates

in 900 ml add:

5 grams NaCl

2 grams $MgSO_4*7H_2O$

5 grams Yeast Extract

10 grams NZ Amine (Casein hydrolysate)

Adjust pH to 7.5 with 5M NaOH, bring up volume to 1 litre and autoclave

If necessary add 100mg of Ampicillin (100 ug/ml)

For NZY Plates 15 g/l of Agar is added

For NZY Top-Agar 7.5 g/l of Agarose is Added

ii) NZYCM Medium

in 900 ml add:

5 grams NaCl

2 grams $MgSO_4*7H_2O$

1 gram Casamino acids

5 grams Yeast Extract

10 grams NZ Amine (Casein hydrolysate)

Adjust pH to 7.0 with 5M NaOH, bring volume up to 1 litre and autoclave

iii) SOC Medium

in 900 ml add:

0.5 grams NaCl

5 grams Yeast Extract

20 grams BactoTryptone

10 ml of 0.2 M KCl

Adjust pH to 7.0 with 5M NaOH, bring up volume to 1 litre and autoclave

Add 5 ml of sterile 2M $MgCl_2$

Add 10 ml of sterile 2M Glucose

iv) Luria-Bertani Medium (LB) and Plates

in 900 ml add:

5 grams NaCl

10 grams Yeast Extract

10 grams BactoTryptone

Adjust pH to 7.0 with 5M NaOH, bring up volume to 1 litre and autoclave

If necessary add 100mg of Ampicillin (100 ug/ml)

For LB Plates 15 g/l of Agar is added

For LB Top-Agar 7.5 g/l of Agarose is Added

v) Brain Heart Infusion Medium

in 1 litre add:

38 grams BHI Powder and Autoclave

If necessary add 100mg of Ampicillin (100ug/ml)

B) Solutions Used

i) 10X Tris-Borate EDTA Buffer

to 900 ml H_2O add:

121.1 grams Tris Base

61.8 grams Boric Acid

3.72 grams Na_2EDTA

Adjust volume to 1 liter and filter sterilize

final pH should be 8.3

ii) SM Phage Dilution Buffer

to 900 ml H_2O add:

5.8 grams NaCl

2.0 grams $MgSO_4*7H_2O$

50 ml of 1M Tris-HCl, pH 7.5

5 ml of 2% Gelatin

Adjust volume to 1 liter and autoclave

iii) 10X Northern Dyes

to 9 ml DEPC treated H_2O add:

25 milligrams of bromophenol blue (0.25%)

25 milligrams of xylene cyanol (0.25%)

1.5 grams Ficoll (Type 400 Pharmacia)

bring up to 10 ml final volume

iv) 10X SSC

to 900 ml H_2O add:

175.3 grams NaCl

88.2 grams of Sodium Citrate

Adjust pH to 7.0 with 5M NaOH, bring up volume to 1 litre and autoclave

v) 10X Ficoll Dyes

to 9 ml H_2O add:

25 milligrams of bromophenol blue (0.25%)

25 milligrams of xylene cyanol (0.25%)

1.5 grams Ficoll (Type 400 Pharmacia)

bring up to 10 ml final volume

vi) Low Hybridisation Solution

for 10 ml add the following:

0.1 grams Bovine Serum Albumin

1.5 ml of H_20

5 ml of 1M $NaPO_4$ - Made as follows

134 grams Na_2HPO_4-$7H_2O$

4 ml H_3PO_4(Phosphoric Acid)

H_2O to 1 Liter

20 uL of 0.5 M EDTA (pH 8.0)

3.5 ml of 20% SDS

vii) 10 mM Tris 1 mM EDTA (T10E1)

to 988 ml H_2O add:

10 ml of 1M Tris*Cl (pH 8.0)

2 ml of 0.5 M EDTA (pH 8.0)

vii) FSB

to 400 ml H_2O add:

0.49 grams KOAc

6.05 grams RbCl

4.45 grams $MnCl_2*H_2O$

0.40 grams $HaCoCl_3$

0.74 grams $CaCl_2*H_2O$

50 ml of glycerol

Adjust pH to 6.4 with dilute HCl bring up volume to

1 litre and filter sterilize

ix) Boiled Preparation Buffer (BPB)

to 262.5 ml sterile H_2O add:

200 ml sterile 20% Sucrose (8% final)

2.5 ml 10% Triton X-100 (0.5% final)

25 ml sterile 1M EDTA pH 8.0 (50 mM final)

10 ml sterile 1 M Tris-HCl pH 8.0 (10 mM final)

x) Plasmid Preparation Solution I

to 93 ml H_2O add:

2.5 ml sterile 2M Glucose (50 mM final)

5 ml sterile 2M Tris-HCl ph 8.0 (25 mM final)

2.0 ml 0.5 M EDTA pH 8.0 (10 mM final)

xi) Plasmid Preparation Solution II

to 95 ml H_2O add:

0.8 grams NaOH (0.2 M final)

5 ml 20% SDS (1% final)

xii) Plasmid Preparation Solution III

to 28.5 ml H_2O add:

60 ml 5 M KOAc

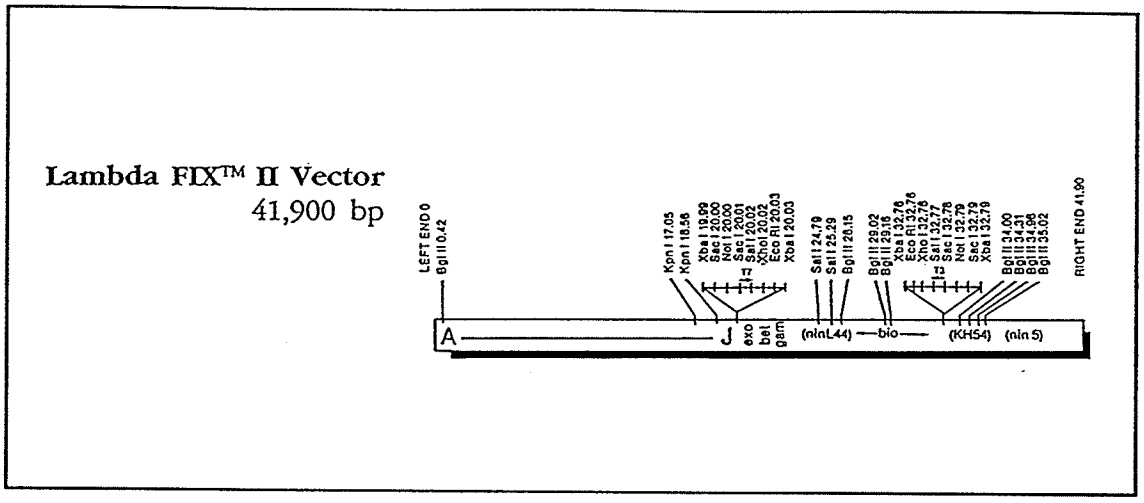
11.5 ml glacial Acetic Acid

C) Description of Bacterial Strains and Cloning Vectors

i) Bacterial Strains

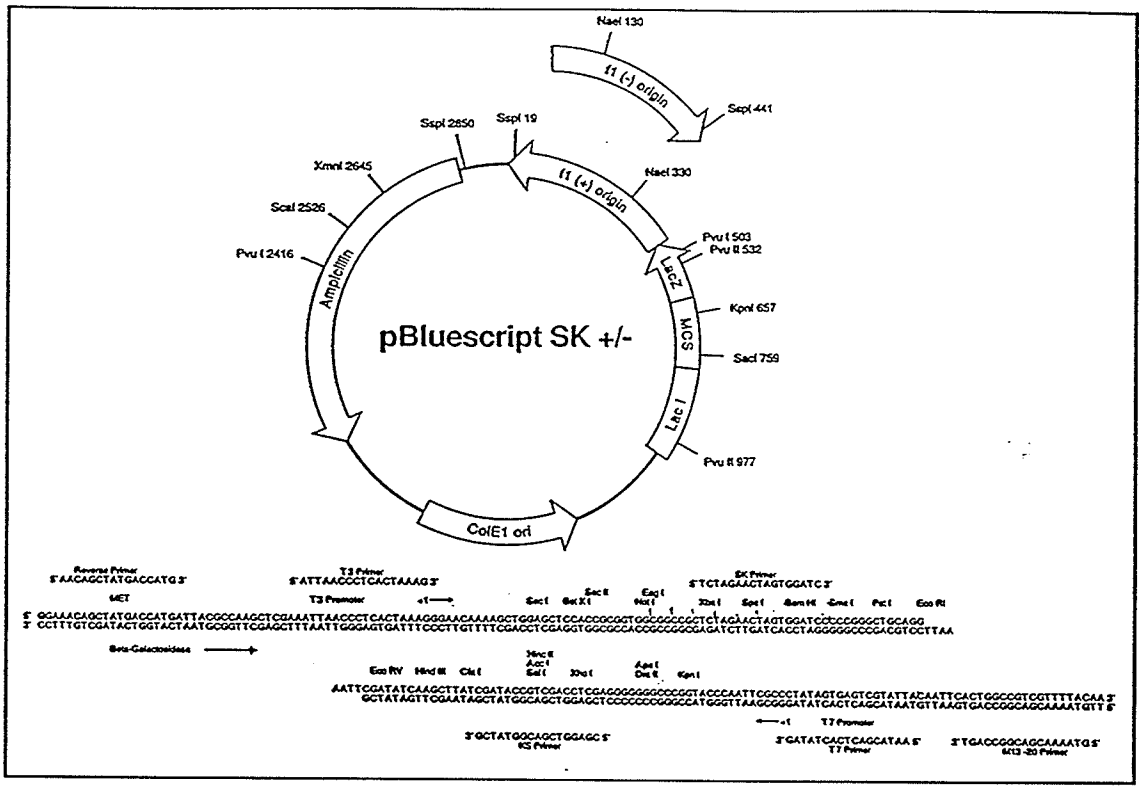
Name	Phenotype	Source
E. coli DH5-α^{TM}	F$^-$, endA1, hsdR17(r_k-, m_k+), supE44, thi-1,λ^- recA1, gyrA96, relA1, Δ(argF-laczya)U169, ϕ80dlacZΔM15	Gibco/BRL Burlington,Ont.
E. coli SRB	sbcC, recJ, uvrC::Tn5(kan^r), supE44, lac, gyrA96 rel A1, thi-1, end A1, mcrA, D(mcrC$^-$ hsdRMS$^-$mrr)171, [F' proAB, lacI^{q}ZΔM15]	Stratagene La Jolla, CA

ii) Description of cloning vectors



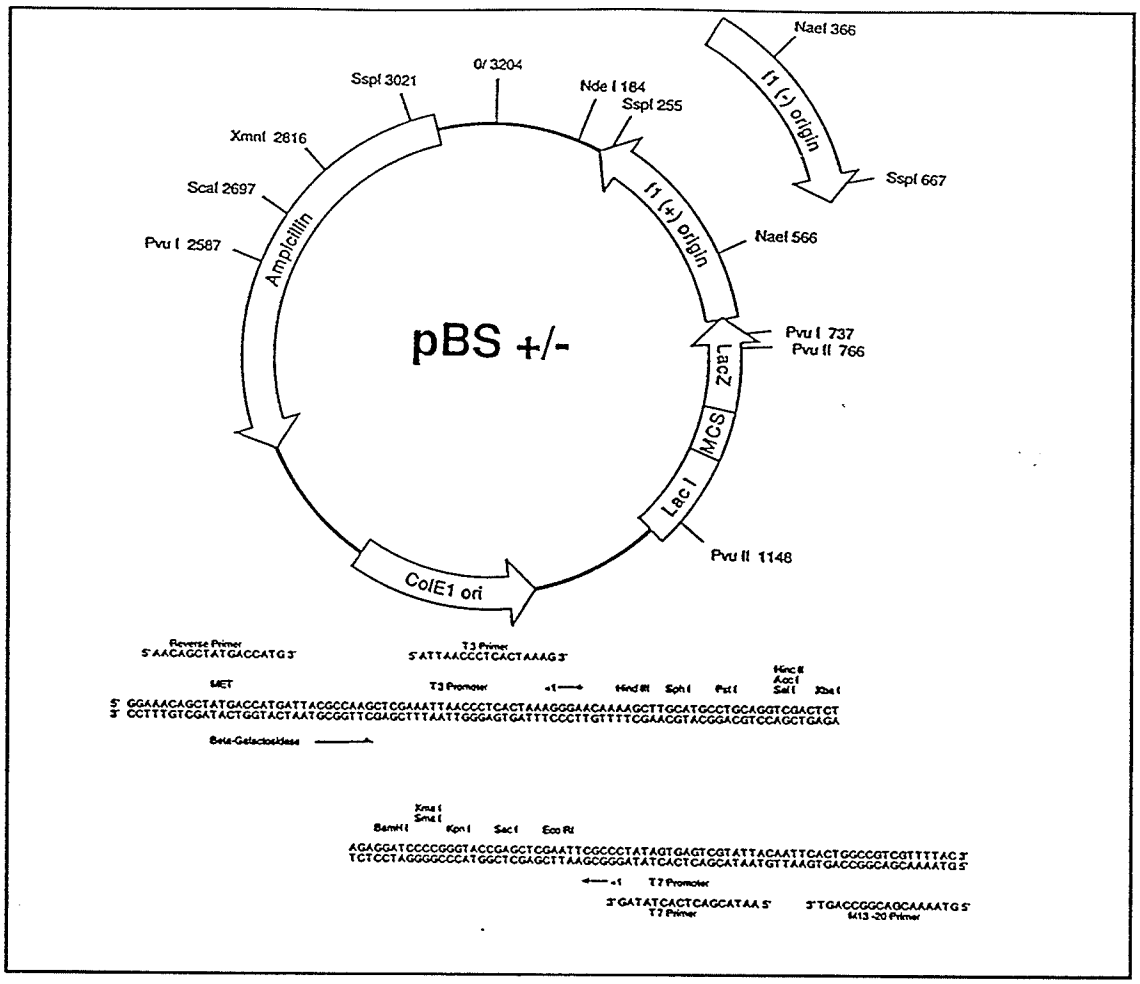
a) Lambda FIX™ II Bacteriophage Vector

Source: Stratagene, La Jolla, CA



b) pBluescript SK(+/-)© Phagemid

Source: Stratagene, La Jolla, CA



c) pBluescribe© Plasmid

Source: Stratagene, La Jolla, CA

D) Abbreviations Used

RNA	Ribonucleic Acid
mRNA	Messenger RNA
premRNA	Unprocessed mRNA
VSG	Variable Surface Glycoproteins
n.t.	Nucleotide
SL	Spliced Leader
SL-RNA	Spliced Leader RNA
gRNA	Guide RNA
HMG CoA	3-hydroxy-3-methylglutaryl coenzyme A reductase
snRNP	Small Nuclear Ribonucleotide Particles
snRNA	Small Nuclear RNA
SL-RNP	Spliced Leader Ribonucleotide Particle
ATP	Adenosine Triphosphate
UV	Ultra-violet Light
mg	Milligram
μg	Microgram
ng	Nanogram
l	Litre
ml	Millilitre
μl	Microlitre
mM	Millimolar

ODU	Optical Density Unit
PFU	Plaque Forming Unit
rpm	Revolutions per minute
mCi	Millicurie
°C	Degrees Celsius
mAmps	MilliAmps
PCR	Polymeraise Chain Reaction
Tm	Annealing Temperature of DNA
MW	Molecular Weight

References

1. Green, M.R. "Biochemical Mechanisms of Constitutive and Regulated Pre-mRNA Splicing." Ann. Rev. Cell Biol. (1991) 7:559-599.

2. Birnsteil M.L. (Ed.) "Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles." New York: Springer-Verlag (1988).

3. Sharp P.A. "Trans-splicing: Variation on a Familiar Theme." Cell (1987) 50: 147-148.

4. Boothroyd J. C. Cross G. A. M. "Transcripts Coding for the Variable Surface Glycoproteins of *Trypanosoma brucei* have a Short Identicle Exon at Their 5′ End." Gene (1982) 20:281-289.

5. Nelson R. G. Parsons M. Barr, P. J. Stuart K. Selkirk M. Agabian N. "Sequences Homologous to Variant Antigen mRNA Spliced Leader are located in tandem repeats and variable orphons in *Trypanosoma brucei*." Cell (1983) 34:901-909.

6. Parsons M. Nelson R. G. Watkins K. P. Agabian N. "Trypanosome mRNA's Share a Common 5' Splice Leader Sequence." Cell (1984) 38:309-316.

7. Ralph D. "Physical Identification of Branched Intron Side Products of Splicing in *Trypanosome brucei*." EMBO J. (1988) 7:2539-2544.

8. Perry K. Agabian N. "mRNA Processing in the Trypanosomatidae." Experientia (1991) 47:118-128.

9. Johnson P.J. "Inactivation of Transcription by UV Irradiation of *T. brucei* Provides Evidence for a Multicistronic Transcription Unit, Including a VSG Gene." Cell (1987) 51:273-281.

10. Imboden M.A. "Transcription of the Intergenic Regions of the Tubulin Gene Cluster of *Trypanosoma brucei*: Evidence for a Polycistronic Transcription Unit in a Eukaryote." Nuc. Acids Res. (1987) 15:7357-7368.

11. Grondal "Characterisation of the RNA Polymerases of *T. brucei*: Trypanosomal mRNA's are composed of Transcripts Derived from both Polymerase II and III." EMBO J. (1989) 8:3383-3389.

12. Cattaneo R. "Different Types of Messenger RNA Editing." Ann. Rev. Genet. (1991) 25:71-88.

13. Krause M. Hirsh D. "A Trans-Spliced Leader Sequence on Actin mRNA in *C. elegans*." Cell (1987) 49:753-761.

14. Bektesh S. van Doren K. V. Hirsh D. "Presence of the *Caenhorhabditis elegans* spliced leader on different RNA's in Different Genera of Nematodes." Genes-Dev. (1988) 2:1277-1283.

15. Bektesh S.L. Hirsh D. "*C. elegans* mRNA's Acquire a Spliced Leader through a Trans-splicing System." Nuc. Acid Res. (1988) 16:5692-5698.

16. Takacs A. M. Denker J. A. Perrine K. G. Maroney P. A. Nilsen T.W. "A 22 Nucleotide Spliced Leader Sequence in the Human Parasitic Nematode *Brugei malayi* is identicle to the Trans-spliced Leader exon in *C. elegans*." PNAS (1988) 85:7932-7936.

17. Maroney P.A. Hannon G. J. Nilsen T. W. "Transcription and Cap Trimethylation of a Nematode Spliced Leader RNA in a Cell-Fee System." PNAS (1991) 87:709-713.

18. Bektesh S. van Doren K. V. Hirsh D. "Presence of the *Caenhorhabditis elegans* Spliced Leader on Different mRNA's and in Different Genera of Nematodes." Genes-Dev. (1988) 2:1277-1283.

19. Takacs A. M. Denker J. A. Perrine K. G. Maroney P. A. Nilsen T.W. "A 22-nucleotide Spliced Leader Sequence in the Human Parasite *Brugia malayi* is Identicle to the Trans-spliced Leader Exon in *C. elegans*." PNAS (1988) 85:7932-7936.

20. Huang X.Y. Hirsh D. "A Second Trans-spliced RNA Leader Sequence in the Nematode *C. elegans*." PNAS (1989) 86:8640-8644.

21. Splieth et. al. "Operons in C. elegans: Polycistronic mRNA Precursers are processed by Trans-Splicing of SL2 to downstream coding Regions." Cell (1993) 73:521-532.

22. Rajkovic. A. Davis R. E. Simonsen J. N. Rottman F. M. "A Spliced Leader is Present on a Subset of mRNA from the Human Parasite *S. mansoni*." PNAS (1990) 87:8879-8883.

23. Unpublished Data.

24. Guthrie C. "Catalytic RNA and RNA Splicing." Am. Zool. (1989) 29:557-567.

25. Lerner M.R. Steitz J. A. "Antibodies to Small Nuclear RNA's complexed with Proteins are Produced by Patients with Systemic Lupus Erethematous." PNAS (1979) 76: 5495.

26. Anderson J. A. "Assembly and Intracellular Transport of snRNP Particles." BioEssays (1991) 13:57-64.

27. Kunkel G. R. Maser R. L. Cavet J. P. Pederson T. "U6 Small Nuclear RNA is Transcribed by RNA Polymeraise III" PNAS. (1983) 83: 8575.

28. Guthrie C. "Catalytic RNA and RNA Splicing." Am. Zool. (1989) 29:387-419.

29. Ruby SW. "An Early Hierarchic role of U1 Small Nuclear Ribonuclearprotein in Spliceosome Assembly." Science (1988) 242:1028-1035.

30. Lerner M. R. "Are snRNP's Involved in Splicing." Nature (1980) 283:220-224.

31. Zhuang Y. Weiner A. M. "A Compensatory Base Change In U1 snRNA Supresses a 5′ Splice Site Mutation." Cell (1986) 46:827-835.

32. Seraphin B. "Identification of Functional U1 snRNP-PremRNA Complexes Committed to Spliceosome Assembly and Splicing." Cell (1989) 59:349-358.

33. Goguel V. "Splice Site Choice and Splicing Efficiency are Positively Influenced by Pre-mRNA Intramolecular Base Pairing in Yeast." Cell (1993) 72:893-901.

34. Nandabalan, K. Price L. Roeder G. S. "Mutations in U1 snRNA Bypass the Requirement for a Cell Type-Specific RNA Splicing Factor" Cell (1993) 73, 407-415.

35. Zieve G.W. Sauterer R. A. "Cell Biology of the snRNP Particles." Criticle Reviews in Biochemistry and Molecular Biology (1990) 25:1-46.

36. Stutz. F. Liao X. C. Rosbash M. "U1 Small Nuclear Ribonucleoprotein Particle-Protein Interactions Are Revealed in *S. cerevisiae* by In Vivo Competition Assays." Mol. and Cell. Biology (1993) 13:2126-2133.

37. Parker R. A. Siliciano P. G. Guthrie C. "Recognition of the TACTAAC Box during mRNA Splicing in Yeast involves Base Pairing to the U2-Like snRNA." Cell (1987) 49:220-239.

38. Wu J. Manley J. L. "Mammalian pre-mRNA Branch Site Selection by the U2 snRNA Involves Base Pairing." Genes Dev. (1989) 7:1553-1561.

39. Ruskin B. Zamore P. D. Green M. R. "A Factor, U2AF, is Required for U2 snRNP Binding and Splicing Complex Assembly." Cell (1988) 52:207-219.

40. Ruby S.W. "An early Hierarchic role of U1 snRNP in Spliceosome Assembly." Science (1988) 242:1028-1035.

41. Reich C. I. VanHoy R. W. Porter L. P. Wise J. A. "Mutations at the 3′ Splice Site Can be Suppressed by compensatory changes in U1 snRNA in Fission Yeast." Cell (1992) 69:1159-1169.

42. Seraphin L. Kandels-Lewis S. "3′ Splice Site Recognition in S. cerevisiae Does Not Require Base Pairing with U1 snRNA." Cell (1993) 73:803-812.

43. Zieve G.W. Sauterer R. A. "Cell Biology of the snRNP Particles." Criticle Reviews in Biochemistry and Molecular Biology (1990) 25:1-46.

44. Luhrmann R. "Structure of Spliceosomal snRNP's and their Role in pre-mRNA Splicing." Biochem. Biophys. Acta (1990) 1087:265-292.

45. Zavavelli M. "Efficient Association of U2 snRNP's with premRNA Requires an essential U2 snRNA Structural Element." Genes Dev. (1987) 5:2521-2533.

46. Lamond A.I. Konarska M. M. Grabowski P. J. Sharp P. A. "Spliceosome assembly involves binding and release of U4 small nuclear ribonucleoprotein." PNAS (1988) 85:411-415.

47. Bindereif K. Green M. R. "An ordered pathway of snRNP binding during mammalian pre-mRNA splicing complex assembly." EMBO J. (1987) 6:2415-2424.

48. Newman A. J. Norman C. "U5 snRNA Interacts with Exon Sequences at 5′ and 3′ Splice Sites." Cell (1992) 68:743-754.

49. Konforti B. B. Koziolkiewicz M. J. Konarska M.M. "Disruption of Base Pairing between the 5′ Splice Site and the 5′ end of the U1 snRNA Is Required for Spliceosome Assembly." Cell (1993) 75:863-873.

50. Sontheime E. J. "The U5 and U6 snRNA's as Active Site Components of the Spliceosome." Science (1993) 262:1989-1996.

51. Cortes J. J. "Mutations in Conserved Loop of Human U5 snRNA Generates the use of Novel 5′ Splice Sites In Vivo." EMBO J. (1993) 12:5181-5189.

52. Brow E. "Splicing a spliceosomal RNA" Nature (1989) 337:14-15.

53. Vankan P. "Roles of U4 and U6 snRNA's in the Assembly of Splicing Complexes." EMBO Journal (1992) 11:335-343.

54. Lamond A.I. Konarska M. M. Grabowski P. J. Sharp P. A. "Spliceosome assembly involves binding and release of U4 small nuclear ribonucleoprotein. PNAS (1988) 85,411-415.

55. Guthrie C. Paterson B. "Spliceosomal snRNA's", Annual Review Genetics (1988) 22, 387-419.

56. Leser C. F. "Mutations in U6 snRNA That Alter Splice Site Specificity: Implications For the Active Site." Science (1993) 262:1982-1988.

57. Yu Y. T. Maroney P. A. Nilsen T. W. "Functional Reconstitution of U6 snRNA in Nematode Cis and Trans-splicing: U6 Can serve as Both a Branch Acceptor and a 5′

Exon." Cell (1993) 75:1049-1059.

58. Sawa H. Shimura Y. "Association of U6 snRNA with the 5′-splice site region of pre-mRNA in the spliceosome." Genes & Development (1992) 6:244-254.

59. Tani T. Ohshima Y. "The Gene for the U6 Small Nuclear RNA in Fission Yeast has an Intron." Nature (1989) 337:87-90.

60. Tani T. Ohshima Y. "mRNA-type Introns in U6 Small Nuclear RNA Genes: Implications for the Catalysis in Pre-mRNA Splicing." Genes Dev. (1991) 5:1022-1031.

61. Takahashi Y. "An mRNA-Type Intron Is Present in the Rhodotorula hasegawae U2 Small Nuclear RNA Gene." Mol. and Cell. Biology (1993) 13:5613-5619.

62. Wu. J. Manley J. L. "Base pairing between U2 and U6 snRNAs is necessary for splicing of a mammalian pre-mRNA." Nature (1991) 352:818-821.

63. Madhani H. D. Guthrie C. " A Novel Base-Pairing Interaction between U2 and U6 snRNA's Suggests a Mechanism for the Catalytic Activation of the Spliceosome." Cell (1992) 71:803-817.

64. Terns M. P. Lund E. Dahlberg J. E. "3′-End-Dependent Formation of U6 Small Nuclear Ribonucleoprotein Particles in *Xenopus laevis* Oocyte Nuclei." Mol. and Cel. Biol. (1992) 12:3032-3040.

65. Tazi J. Forne T. Jeanteur P. Cathala G. Brunel C. "Mammalian U6 Small Nuclear RNA Undergoes 3′ End Modifications within the Spliceosome." Mol. and Cel. Biology (1993) 13:1641-1650.

66. Zieve G.W. Sauterer R. A. "Cell Biology of the snRNP Particles." Criticle Reviews in Biochemistry and Molecular Biology (1990) 25:1-46.

67. Vankan P. McGuigan C. Mattaj I. W. "Roles of U4 and U6 snRNA's in the Assembly of Splicing Complexes." EMBO Journal (1992) 11: 335-343.

68. Wolff T. Bindereif A. "Reconstituted Mammalian U4/U6 snRNP Complements Splicing: A Mutational Analysis." EMBO Journal (1992) 11: 345-359.

69. Agabian N. "Trans-splicing of Nuclear pre-mRNA's" Cell (1990) 61:1157-1160.

70. Bruzik J.P. van Doren K. Hirsh D. Steitz J. A. "Trans-splicing Involves a Novel Form of Small Nuclear Ribonucleoprotein Particles." Nature (1988) 335:556-559.

71. Thomas J.D. "The *C. elegans* Trans-Spliced Leader RNA Is Bound to Sm and has a trimethylguanosine Cap." Cell (1988) 54:533-539.

72. Van Doren K. Hirsh D. "Trans-spliced leader RNA exist as small nuclear ribonucleoprotein particles in *C. elegans*." Nature (1988) 335:556-559.

73. Michaeli S. "Isolation of Distinct Small Ribonucleoprotein Particles Containing the Spliced Leader and U2 RNA's of *Trypanosoma brucei*." Journal Biological Chemistry 265:10582-10588.

74. Ullu E. Tshudi C. "Trans splicing in Trypanosomes Requires Methylation of the 5′ end of the SL RNA." PNAS (1991) 88:10074-10078.

75. Mottram J. "Isolation and Sequence of Four Small Nuclear U RNA Genes of *Trypanosoma brucei* subsp. *brucei*: Identification of the U2, U4 and U6 RNA Analogs." Mol. and Cel. Biology (1989) 9:1212-1223.

76. Tschudi C. Ullu E. "Destruction of U2, U4 or U6 Small Nuclear RNA Blocks Trans-splicing in Trypanosome Cells." Cell (1990) 61:459-466.

77. Watkins K.P. "Identification of a Small RNA That Interacts with the 5′ Splice Site of the *Trypanosoma brucei* Spliced Leader RNA In Vivo." Cell (1994) 76:171-182.

78. Hannon G. J. Maroney P. A. Nilsen T. W. "U Small Nuclear Ribonucleoprotein Requirements for Nematode cis and Trans-Splicing in vitro." Journal Biological Chemistry (1991) 266:22792-22795.

79. Bruzik J.P. Steitz J. A. "Spliced Leader RNA Sequences Can Substitute for the Essential 5′ End of U1 RNA during Splicing in a Mammalian In Vitro System." Cell (1990) 62:889-899.

80. Hannon G.J. "Interaction of U6 snRNA with a Sequence Required for Function of the Nematode SL RNA in trans-splicing." Science (1992) 258:1775-1780.

81. Bruzik J. P. Maniatis T. "Spliced leader RNA's from lower eukaryotes are trans-spliced in mammalian cells." Nature (1992) 360:692-695.

82. Ullu E. "Temporal Order of RNA-Processing Reactions in Trypanosomes: Rapid trans-splicing Precedes Polyadenylation of Newly Synthesised Tubulin Transcripts." Mol. and Cel. Biology (1993) 13:720-725.

83. Conrad R. Thomas J. Spieth J. Blumenthal T. "Insertion of Part of an Intron into the 5′ Untranslated Region of a Caenorhabditis elegans Gene Converts It into a trans-Spliced Gene." Mol. and Cel. Biology (1991) 11:1921-1926.

84. Conrad R. "Conversion of a trans-spliced *C. elegans* gene into a Conventional Gene by Introduction of a Splice Donor Site." EMBO J. (1993) 12:1249-1255.

85. Speith L. "Operons in C. elegans: Polycistronic mRNA Precursers Are Processed by Trans-Splicing of SL2 to Downstream Coding Regions." Cell (1993) 73:521-532.

86. Unpublished Results.

87. Rajkovic. A. Davis R. E. Simonsen J. N. Rottman F. M. "A Spliced Leader is Present on a Subset of mRNA from the Human Parasite *S. mansoni*." PNAS (1990) 87:8879-8883.

88. Sambrook J. Maniatis T. Fritch E. S. "Molecular Cloning: A Laboratory Manual 2nd Ed." Publisher: Cold Spring Harbor Laboratory Press, New York (1989).

89. Southern E. "Detection of Specific Sequences Among DNA Fragments Separated by Gel Electrophoreses" J. Mol. Biol. (1975) 98:503.

90. Zuker M. "On Finding All Suboptimal Foldings of an RNA Molecule" Science (1989) 244:48-52.

91. Wyman G. "Host Strains that Alleviate Underrepresentation of Specific Sequences" (1987) Methods Enzymol. 152:173-179.