MOLECULAR EVOLUTION OF INTRONS AND OTHER RNA ELEMENTS

Organizers: Donal Hickey, Marlene Belfort and Alan Lambowitz

February 2-8, 1991

Plenary Sessions Page

Poster Sessions

Molecular Evolution of Introns and Other RNA Elements

Evolutionary Origin of Introns

I 001 HOW BIG IS THE UNIVERSE OF EXONS? Walter Gilbert, Harvard University, Department of Biological Laboratories, 16 Divinity Avenue, Cambridge, Massachusetts 02138.

We try to estimate how many different kinds of exons were used to make up the primordial proteins. We do this by thinking of each gene as a random drawing of a set of exons from some underlying universe of possible exons. The frequency at which the same exon appears in different genes gives us a way of estimating how big that universe is. Our best current estimate is that a few thousand exons were sufficient to construct all proteins.

I 002 THREE-DIMENSIONAL ARCHITECTURE AND EVOLUTION OF GROUP I INTRONS
François Michel[1], Luc Jaeger[1,2] and Eric Westhof[2], (1) Centre de Génétique Moléculaire du C.N.R.S., 91190 Gif-sur-Yvette, France and (2) Institut de Biologie Moléculaire et Cellulaire du C.N.R.S., 15 rue René Descartes, 67084 Strasbourg-Cedex, France.

Alignment of 87 sequences of group I introns reveals several instances of covariation between distant sites which can be accounted for neither by chance events during the divergence of present-day members of group I, nor by the well-established secondary structure of these molecules, and are therefore best explained as reflecting conserved tertiary interactions. With the help of stereo-chemical modelling, we have shown that these newly uncovered, putative tertiary contacts can all be incorporated into a single, coherent three-dimensional picture of the core of group I introns (ref. 1). The validity of our model is being assessed in a number of ways. First, we are checking that grafting the various peripheral parts and pseudo-knots that characterize individual subgroups of group I introns on the common core structure is straightforward and does not lead to topological impossibilities. Second, we are verifying that nucleotide combinations favored in evolution are the ones that perform best in *in vitro* tests. More importantly, we are taking advantage of our geometrical interpretations of patterns of covariation to devise novel, isosteric nucleotide combinations, which we predict should be compatible with function. Conceivable extensions of this approach include the replacement of entire subdomains and, eventually, the deliberate engineering of new functional versions of the ribozyme.

ref. 1: F. Michel & E. Westhof, J. Mol. Biol., in the press.

I 003 **EUBACTERIAL ORIGIN OF GROUP I AND GROUP II INTRONS,** **Jeffrey D. Palmer, Maria G. Kuhsel and James R. Manhart*, Department of Biology, Indiana University, Bloomington, IN 47405 and *Department of Biology, Texas A&M University, College Station, TX 77843.**

We are studying the origin and evolution of chloroplast introns. We have found that certain chloroplast introns were acquired within the chloroplast lineage after its endosymbiotic origin, whereas others were present in the endosymbiotic, bacterial ancestors of chloroplasts. The former category includes group II introns located in tRNA-Ala and tRNA-Ile genes of land plant chloroplasts, which were acquired specifically in the green algal ancestors of land plants[1]. The latter category includes both a group I intron in the gene for tRNA-Leu(UAA) and a group II intron in the gene for tRNA-Val(UAC). Both of these introns are present in chloroplasts of all major groups of algae and in diverse eubacterial lineages, including not only cyanobacteria[2], the immediate progenitors of chloroplasts, but also purple bacteria and basal thermophilic bacteria such as Thermotoga. In all cases, the introns are present in the same position within the tRNA-Leu and tRNA-Val genes of chloroplasts and eubacteria, and in some cases they are highly similar in primary sequence and secondary structure. These results suggest that at least some group I and group II introns were present in the common ancestor of all eubacteria. Efforts to detect these introns in archaebacterial and eukaryotic nuclear tRNA genes will be presented. The implications of these findings for the ongoing debate regarding the origin of eukaryotic mRNA introns - early or late - will be discussed.

[1]Manhart, J.R. and Palmer, J.D. 1990. The gain of two chloroplast tRNA introns marks the green algal ancestors of land plants. Nature 345:268-270.
[2]Kuhsel, M.G., Strickland, R., and Palmer, J.D. 1990. An ancient group I intron shared by eubacteria and chloroplasts. Science, in press.

I 004 MITOCHONDRIAL INTRONS : EARLY FOUNDERS OR LATE INVADERS.

P. Slonimski Centre de Génétique Moléculaire, Laboratoire propre du CNRS, associé à l'Université P. et M. Curie, 91190 GIF-sur-YVETTE (France)

The majority of mitochondrial genomes are devoid of introns. I have asked the question whether mitochondrial genes with no introns (eg. from metazoa) display singular properties surrounding the homologous positions corresponding to intron insertion positions in mitochondrial genes from lower eucaryotes. To this end I have developed a new method for measuring evolutionary divergence based on distance and variance matrices. The method is robust and applicable to any set of homologous sequences. Analysis of some 50 insertion positions corresponding to more than 80 introns demonstrates that introns are inserted in the evolutionary most conserved positions of mitochondrial genes coding for proteins. Monte-Carlo simulations demonstrate the remarkable significance of this constraint. Interestingly, this constraint is stronger at the protein than at the nucleic acid level and holds true for any comparison of homologous genes from bacteria, plants, fungi and animals. Furthermore, introns are inserted in, or in the vicinity of codons specifying amino acids essential for the electron transfer catalysed by the cytochrome reductase,(bc1) and cytochrome oxydase complexes. The degree of evolutionary conservation of individual intron insertion positions will be compared with the known intron properties (protein coding capacity, mobility, self-splicing, secondary structure, etc.) Two alternatives of the origin of mitochondrial introns (a late evolutionary invasion of mtDNA in the fungi-algae phylum versus intron persistance from the early RNA world tagging the early catalysis of mitochondrial functions) will be discussed and compared with the results of a similar analysis bearing on nuclear introns in metazoa.

Molecular Evolution of Introns and Other RNA Elements

Self-Splicing Introns

I 005 CONVERSION OF A "PERFECT" RIBOZYME FROM TETRAHYMENA PRE-rRNA TO AN EVEN BETTER ENZYME BY SITE-DIRECTED MUTAGENESIS D. Herschlag, B. Young and T. R. Cech, Howard Hughes Medical Institute, Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO, 80309-0215

The L - 21 ScaI ribozyme, derived from the self-splicing group I intron of *Tetrahymena* pre-rRNA, catalyzes a site-specific endonuclease reaction with RNA substrates that is analogous to the first step of self-splicing: $G_2CCCUCUA_5 + G = G_2CCCUCU + GA_5$. Pre-steady state and steady state kinetic experiments have revealed that binding of the oligonucleotide substrate is rate limiting with subsaturating oligonucleotide substrate (S) and saturating guanosine (G).[1] Thus, the ribozyme can be described as having achieved evolutionary "perfection," as defined by Albery and Knowles,[2] in that every time S binds it is converted to product so that a change in the catalytic machinery that sped the chemical step would not increase the observed rate of reaction. However, this aspect of "perfection" refers only to the situation with subsaturating substrate (i.e. k_{cat}/K_m). With saturating S the ribozyme is a very slow enzyme, with a turnover number of only 0.1 min^{-1} (pH 7, 50 °C, 10 mM $MgCl_2$), because of the slow release of the product $G_2CCCUCU$. Thus, even though the ribozyme operates at the rate of binding with very low concentrations of S, when more S is present the ribozyme is tied up in the "E·P" complex , leaving little free ribozyme to catalyze the reaction of free substrate. The slow release of $G_2CCCUCU$, an analog of the 5' exon, results from binding that is ~10^4-fold stronger than that predicted for a simple duplex with the 5' exon binding site, suggesting that tertiary interactions contribute to binding.[1,3] Even though tight binding makes the ribozyme a bad enzyme for multiple turnover, it makes good biological sense: after the first step of self-splicing the 5' exon is no longer covalently attached; tight binding presumably prevents loss of the 5' exon thereby ensuring efficient splicing. Mutations in a non-conserved region adjacent to the 5' exon binding site weaken binding of $G_2CCCUCU$ and thereby increase the turnover number by up to ~100-fold. With subsaturating S, binding remains rate-limiting for the mutant ribozymes. Thus, the mutant ribozymes remain "perfect," while being faster enzymes for multiple turnover, with saturating S. While ribo-oligonucleotides are bound by the mutant ribozymes up to 10^3-fold worse than wild type, the binding of deoxy-oligonucleotides, which are bound by the wild type ribozyme much weaker than the corresponding ribo-oligonucleotides, is essentially unaffected by the mutations. This strongly suggests that tertiary interactions with the ribozyme that specifically stabilize binding of the ribo-substrate are disrupted by these mutations. It is suggested that the mutated region is not directly involved in interactions with 2'-hydroxyl groups, but rather, is important for positioning the ribozyme/S duplex so that it can form the tertiary contacts. Results with the mutant ribozymes also have implications for ribozyme specificity, both in terms of discrimination against alternate substrates and in terms of the choice of the cleavage site.

1. Herschlag, D., & Cech, T. R. (1990) *Biochemistry* In press.
2. Albery, W. J., Knowles, J. R. (1976) *Biochemistry 15*, 5631.
3. Pyle, A. M., McSwiggen, J. A., & Cech, T. R. (1990) *Proc. Natl. Acad. Sci. USA* In press.
DH is a Lucille P. Markey Scholar in Biomedical Science.

I 006 GROUP II INTRON SPLICING: MINI-INTRONS, MUTANT INTRONS AND REVERSE GENETICS, P. S. Perlman, J. L. Koch, S. D. Dib-Hajj, S. M. Belcher and S. C. Boulanger, Department of Biochemistry, UT Southwestern Medical Center, Dallas, Texas 75235 and Department of Molecular Genetics, Ohio State University, Columbus, Ohio 43210.

Our previous studies have identified a key role for the 34 nt long domain 5 in *in vitro* 5' exon release reactions carried out by the group II introns aI5γ, bI1 and aI1 of yeast mtDNA. Bimolecular reactions involving domain 5 transcripts (lacking other intron domains) and various substrate RNAs lacking domain 5 indicate that there must be a binding site for domain 5. Precursor RNAs containing aI5γ singly deleted for intron domains 2, 3, 4 or 6 were found to be highly reactive *in vitro*. We have analyzed *in vitro* reactions of RNAs transcribed from derivatives of the original aI5γ clone containing combinations of deletions of domains 2, 3, 4, and 6. RNA lacking domains 2, 4 and 6 (Δ2,4,6) carries out the 5'-exon release reaction under our standard 0.5 M salt conditions and regains the splicing reaction when the salt concentration is raised. Δ2,3,4,6 RNA is inactive under those standard conditions but regains the 5'-exon release reaction in the presence of higher salt levels. Retention of activities by these mini-intron constructions indicates that the reaction center of this catalytic RNA is made up of domains 1 and 5. These findings also show that the binding site for domain 5 is probably located within domain 1 and greatly diminish the possibility that it is created by an interaction between domain 1 and one or more of domains 2, 3, 4 and 6.

Various point and partial deletion mutants affecting chiefly domain 5 and the C1 substructure of domain 1 of two yeast group II introns were constructed and characterized *in vitro*. We have devised *trans* assays for d1, d3 and d5 function and those assays were a part of the characterization of each mutant intron form. Complementation experiments using intron domains from different introns yielded positive results although, especially for domain 5, homologous reactions were more efficient. This suggests that there may be some intron-specific accomodations of these generally conserved domains. Most point mutations tested have only modest effects on self-splicing even though highly conserved positions are changed. These studies of mutant RNAs set the stage for analysis of selected mutations *in vivo*. Mitochondrial transformation experiments using pMIT vectors developed in R. A. Butow's laboratory are in progress.

I 007 Self-Splicing Group I Introns in Bacteria. David A. Shub; Department of Biological Sciences; State University of New York, Albany, Albany, NY 12222.

The existence of ribozymes has stimulated speculation that they may have been the earliest genetic units. In these hypotheses, self-splicing introns are presumed to have been part of the primordial gene structure.

The weight of evidence from contemporary biology, however, has not given much support to the view that introns are ancient. Even though mitochondria and chloroplasts, which are clearly descended from eubacteria, contain many group I and group II introns, introns were never found in eubacterial genes. The only exceptions are group I introns in bacteriophage, which may have been acquired by horizontal gene transfer. In addition, many introns in organelles contain genes that facilitate intron transfer. Thus, the self-splicing introns may represent a relatively recent class of transposons, and introns in eukaryotic organelles may have been acquired from the nucleus after establishment of endosymbiosis.

We have searched directly for group I introns in eubacterial chromosomal genes. The ubiquity of introns in chloroplast genomes is striking. In spite of the fact that these genomes have undergone extensive gene loss, retaining only a few percent of the DNA of their bacterial ancestors, all plant chloroplasts have group I introns. In fact, such introns also exist in cyanobacteria. We have identified an intron, whose sequence is remarkably conserved, located in precisely the same position of the same gene in a variety of cyanobacteria and chloroplasts. (1) I will discuss the implications of this finding for theories of the origin and function of group I introns.

1. Ming-Qun Xu, Scott D. Kathe, Heidi Goodrich-Blair, Sandra A. Nierzwicki-Bauer, and David A. Shub. *Science* (1990), in press.

I 008 *IN VITRO* SELECTION OF RNA MOLECULES WITH NOVEL FUNCTIONS Jack W. Szostak, Andrew D. Ellington, Rachel Green, Jennifer A. Doudna. Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114

We have developed a simple method of isolating RNA molecules with ligand binding or certain types of catalytic activity from complex mixtures of random sequence molecules, or pools of mutagenized sequences. This method relies on the repeated application of *in vitro* selection for binding or catalysis followed by *in vitro* amplification of the selected sequences. We originally applied this method to the isolation of RNA molecules that specifically bound certain organic dyes; this approach has now been extended to the isolation of RNA molecules that specifically recognize and bind metabolic cofactors such as NAD.

We have also used *in vitro* selection and amplification as a tool for the genetic analysis of the structure of the group I introns. After randomizing discrete regions of the sequence of the intron, we select for the variants that retain catalytic activity. The sequences of these variants are aligned so that covarying residues may be detected. This 'artificial phylogeny' has advantages over natural phylogenetic data in that there are few extraneous sequence changes, and the activity of all variants can be readily measured. In this way we have obtained additional evidence in favor of the base-triple interactions postulated to exist between base-pairs 2 and 3 of stem P4.

The same methods have been used to select for ribozymes with improved stability. We have developed small derivatives of the *sunY* ribozyme by deleting the P7.1 and P7.2 stems. While these derivatives retain catalytic function, they are active only at very high salt and Mg^{++} concentrations. This ribozyme is small enough that it can be transcribed from a single synthetic deoxyoligomer 171 nucleotides in length. This DNA template was synthesized with the coding region doped with 5% incorrect bases to generate a heavily mutagenized pool of RNA transcripts. The initial pool showed extremely low activity, but repeated rounds of *in vitro* selection have led to the isolation of RNA molecules as active as the wild type undeleted parental ribozyme. These small stable ribozymes should be useful for structural analysis, and as a starting point for the development of an RNA replicase.

Molecular Evolution of Introns and Other RNA Elements

Intron Mobility and Evolution

I 009 INTRON MOBILITY IN PHAGE T4: CIS AND TRANS REQUIREMENTS OF INTRON HOMING,

Marlene Belfort, Deborah Bell-Pedersen, Jonathan Clyman and Susan Quirk,
Wadsworth Center for Labs and Research, New York State Department of Health, Empire State Plaza, Albany, New York, 12201.

Two of the three group I introns in phage T4, residing in the td and sunY genes, are mobile (1). That is, they are able to insert themselves with high efficiency into their respective intronless (In^-) alleles. This homing event is dependent on an intron-encoded endonuclease, which makes a double-strand cut in the In^- homologue. We are studying the cis and trans requirements of the homing process. First, the recognition and cleavage sites of the unusual phage endonucleases, whose cut sites are remote from the intron insertion site (2, 3), are being defined. Thus, although distant cleavage site sequences can influence enzyme activity, the endonuclease binding domain, which spans the intron insertion site, is both necessary and sufficient for cleavage. Second, the role of intron core DNA sequences in intron homing has been probed. Consistent with a double-strand break repair mechanism for intron inheritance, mobility is independent of any specific intron sequence provided the endonuclease is supplied in trans. Third, to define the cellular requirements for intron mobility, we are monitoring intron transfer in mutant host backgrounds. In testing various rec, lex, fis, him and hup alleles, a strong dependency on RecA function was demonstrated. Finally, by exploiting a plasmid-to-plasmid intron transfer assay, we have shown that phage factors are required for intron mobility. These phage recombination functions are currently being identified.

Our demonstration that endonuclease-dependent intron transfer is independent of intron sequences that encode the catalytic core, supports the argument that the endonucleases are the primary agents of mobility. The endonuclease coding sequences are therefore likely to be the primordial mobile elements, having made their way into the genetically silent self-splicing introns and thereby imparted mobile properties on these introns more recently in evolution.

(1) Quirk, S., et al., Cell, 56 (1989) 455-465.
(2) Bell-Pedersen, D., et al., Nucleic Acids Res., 18 (1990) 3763-3770.
(3) Chu, et al., Proc. Natl. Acad. Sci., 87 (1990) 3574-3578.

I 010 EVOLUTIONARY ENGINEERING OF RNA ENZYMES

Gerald F. Joyce, Department of Chemistry and Department of Molecular Biology
Research Institute of Scripps Clinic, 10666 North Torrey Pines Road, La Jolla, California 92037

RNA, by virtue of its template properties and catalytic activity, is able to serve both a genotypic and phenotypic function. We have exploited this dual role to develop an RNA-based evolving system that operates entirely *in vitro* [1]. A population of mutant ribozymes is subjected to repeated rounds of selective amplification. Only those individuals that perform a chosen catalytic task are amplified, so that over time the population "adapts" to the task at hand. We can selectively amplify 10^{10} - 10^{12} ribozymes in one hour, providing a very powerful search method. By gradually changing the nature of the catalytic task we are attempting to guide the evolving population toward the expression of novel catalytic behaviors.

Working with the *Tetrahymena* ribozyme, a molecule that catalyzes sequence-specific phosphoester transfer reactions involving RNA substrates, we used *in vitro* evolution techniques to obtain a mutant form of the enzyme that best cleaves a single-stranded DNA substrate [2]. We are currently taking two approaches in an attempt to engineer RNA enzymes to meet our specifications. First, we are selecting ribozymes that best cleave a target RNA or DNA substrate under conditions that resemble those of the cellular environment. The goal is to develop a family of "antisense ribozymes" that could be used to inactivate a target nucleic acid *in vivo*. Second, we are investigating a series of substrates of the form: d(CCCTCT$A_3TA_3TA_3TA_3$), where T refers to an analogue of thymidylate and the ribozyme is selected for its ability to cleave the phosphodiester bond following the sequence CCCTCT. In this way we are seeking to expand the catalytic repertoire of RNA to include a broader range of phosphoester chemistry.

1 Joyce, G.F. (1989) *Gene* **82,** 83-87.
2 Robertson, D.L. & Joyce, G.F. (1990) *Nature* **344,** 467-468.

I 011 MITOCHONDRIAL INTRONS AND RETROID ELEMENTS, A.M. Lambowitz, Departments of Molecular Genetics and Biochemistry, The Ohio State University, Columbus, OH 43210.

I will update two lines of research: (1) Involvement of mt tyrosyl-tRNA synthetase (tyrRS) in RNA splicing. *Neurospora* mt tyrRS, which is encoded by nuclear gene *cyt-18*, functions in splicing the mt large rRNA intron and other group I mitochondrial introns. We have expressed functional *cyt-18* protein in *E. coli*, using the T7 expression system. The expressed protein was purified to near homogeneity and shown to have splicing and tyrRS activities virtually identical to those of authentic *cyt-18* protein from mitochondria. These findings indicate that the *cyt-18* protein is by itself sufficient to splice the mt large rRNA intron *in vitro*. We also developed an *E. coli in vivo* system in which splicing of the mt large rRNA intron is required for the expression of β-galactosidase and is dependent on the expression of the *Neurospora cyt-18* protein. *In vitro* mutagenesis using the *E. coli* expression systems indicates that splicing activity of the *cyt-18* protein is dependent on an idiosyncratic N-terminal domain not present in bacterial or yeast mt tyrRS's, which do not function in splicing. This domain apparently acts together with other regions of the *cyt-18* protein, including the putative tRNA-binding domain at the C-terminus, to promote splicing. Competition of splicing and tyrRS activities of the *cyt-18* protein with mutant introns containing large deletions localizes the *cyt-18* protein binding site to a region of the intron core. We present a model of the *cyt-18* protein binding site, a key feature of which is the similarity between the highly conserved P7 stem in the intron and the variable loop region of *Neurospora* mt $tRNA^{tyr}$. Together, our results suggest that splicing activity of the *Neurospora* mt tyrRS may have been acquired after the divergence of *Neurospora* and yeast and may be based on a fortuitous resemblence between a conserved region of group I introns and the variable loop of *Neurospora* mt $tRNA^{tyr}$. The adaptation of the tyrRS and other cellular proteins to promote splicing of group I and group II introns may be a relatively recent evolutionary development, possibly reflecting the recent dispersal of the introns themselves. (2) Reverse transcription of mt plasmids. The Mauriceville and Varkud mt plasmids are closely-related, closed-circular DNAs (3.6 and 3.7 kb, respectively) that have characteristics of mtDNA introns and retroid elements. The plasmids use a novel mechanism of reverse transcription in which (-) strand DNA initiates directly at the 3' end of the plasmid transcript via recognition of a 3' terminal tRNA-like structure, similar to those at the 3' ends of plant RNA viruses. We identified the plasmid RT as an 81 kDa protein encoded by the 710 amino acid ORF of the plasmid and have expressed active RT in *E. coli*. Recent studies have focused on biochemical analysis of the plasmid RT and the mechanism of priming of (-) strand DNA synthesis. Autonomous intron-related elements, similar to the Mauriceville/Varkud plasmids, may have played a role in horizontal transfer of introns between organisms. The characteristics of the plasmid reverse transcription mechanism suggest that it may have arisen early in evolution, at the time of transition from an RNA to DNA world, and may be ancestral to the mechanism used by retroviruses and related elements.

Introns in Nuclear Genes

I 012 **BIOCHEMISTRY OF MAMMALIAN NUCLEAR PRE-mRNA SPLICING AND SPLICE SITE SELECTION**, Adrian R. Krainer, Akila Mayeda, David Helfman, Greg C. Conway and Diane Kozak. Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, New York 11724.

SF2 is one of several protein components necessary for 5' splice site cleavage and lariat formation during pre-mRNA splicing in HeLa cell extracts. Purified SF2 from HeLa cells is a 33 kDa protein that sometimes resolves into two or three closely spaced bands on SDS-PAGE. The purified 33 kDa polypeptides are active, in that they are sufficient to complement an extract containing the remaining splicing factors, to splice several different pre-mRNAs. However, different pre-mRNAs require different amounts of SF2 for optimal splicing. SF2 is necessary for the assembly or stabilization of the earliest specific pre-spliceosome complex, although in the absence of other components, it can bind RNA in a non-specific manner. The RNA-binding properties of SF2 in the presence of other purified components are currently under study. SF2 copurifies with an activity that promotes the annealing of complementary RNAs. Thus, SF2 may promote specific RNA-RNA interactions between snRNAs and pre-mRNA, between complementary snRNA regions, and/or involving intramolecular pre-mRNA helices. However, other purified proteins with RNA annealing activity cannot substitute for SF2 in the splicing reaction. In addition to its essential role in constitutive splicing, SF2 can determine the choice of 5' splice site when multiple sites are available, as shown with mutant and artificial pre-mRNAs, as well as with pre-mRNAs derived from the alternatively spliced rat β-tropomyosin pre-mRNA. In general, high concentrations of purified SF2 promote utilization of proximal 5' splice sites, and have no effect on 3' splice site selection. The basis for the polarity of this effect is under study. These effects of SF2 on splice site selection may reflect the cellular mechanisms that prevent exon skipping and ensure the accuracy of splicing. In addition, alterations in the concentration or activity of SF2, and of other general splicing factors, may serve to regulate alternative splicing in vivo. We have determined the overall aminoacid composition of each of the 33 kDa polypeptides that copurify with SF2 activity. We have also determined the aminoacid sequence of the N-terminus of the smallest polypeptide, and of several internal fragments of the remaining polypeptides. Based on these sequences, we are isolating the corresponding cDNAs. A full length cDNA for the smallest polypeptide has already been obtained and used to overexpress soluble protein in E. coli. The sequence of this protein is not homologous to entries in the current databases, and it is highly acidic. In contrast, the other polypeptides are highly arginine-rich. Experiments are in progress to determine whether both the acidic and basic proteins are necessary for SF2 activity. We are also characterizing an activity that inhibits the effect of SF2 on 5' splice site selection, but does not affect its splicing complementation activity. This activity may act independently of SF2 to promote the utilization of distal 5' splice sites, or it may bind to SF2 to inhibit its activity. Such an activity can potentially participate in regulating the selection of alternative 5' splice sites in conjunction with SF2. We are currently purifying this protein(s) on the basis of the inhibition assay.

I 013 U1 snRNP AND SPLICE SITE CHOICE IN S. CEREVISIAE, Michael Rosbash, Nadja Abovich, Hildur V. Colot, Valerie Goguel and Xiaoling Liao, Department of Biology, Brandeis University, Waltham, MA 02254

We are interested in the mechanism of nuclear pre-mRNA splicing in the yeast, Saccharomyces cerevisiae. Our focus is on splice site recognition and splice site partner formation, i.e., the early events of spliceosome assembly that lead to apposition of the 5' splice site and branchpoint regions such that splicing reactions can take place. In vitro experiments indicate that both of these regions are involved very early in splicing complex formation. Surprisingly, they interact prior to the involvement of any known branchpoint binding factor. A set of in vivo experiments give some indication that polarity and/or proximity are important features of 3' splice site choice in yeast as has been reported for mammalian cells. Interestingly, the polarity is much less apparent in vitro, suggesting that some aspect of the system is absent or defective in vitro. The data give some indication that the rules (and molecules?) that contribute to splice site choice decisions and that suppress exon skipping in mammalian cells may be components of the general splicing machinery, conserved between yeast and mammals. Both in vivo and in vitro, U1 snRNP figures prominently. It is involved in the earliest in vitro splicing complexes that can be detected. Also, there are indications that it is involved in 3' splice site choice as well as direct evidence confirming its role in 5' splice site choice.

I 014 ALTERNATIVE SPLICING OF MAMMALIAN PRE-mRNA, Manley, J.L., Ge, H., Harper, J., and Wu, J. Dept. of Biological Sciences, Columbia University, N.Y., N.Y., 10027

We have purified a protein of 32-33kD that is able to modulate selection of 5' splice sites during in vitro splicing of pre-mRNAs containing alternative 5' splice sites. High levels of the protein, called alternative splicing factor (ASF), leads to preferential utilization of the 5' splice site proximal to the 3' splice site, apparently independent of sequence. Lower concentrations can, in some cases, promote utilization of distal sites. The function of ASF appears to involve binding to the pre mRNA and influencing the binding of other proteins. Changes in the concentration or activity of the protein may play a role in cell-specific alternative splicing. Finally, protein sequencing and cDNA cloning suggests that ASF is a member of the protein family containing the conserved ~80 amino acid RNA recognition motif, with a somewhat diverged RNP consensus octamer.

Molecular Evolution of Introns and Other RNA Elements

Retroelements

I 015 TRANSPOSITION IN YEAST: MECHANISM AND HOST FACTORS, Jef D. Boeke, Lelita Braiterman, Karen Chapman, Daniel Eichinger and Hua Xu, Department of Molecular Biology & Genetics, Johns Hopkins University School of Medicine, 725 N. Wolfe St. Baltimore, MD 21205

The yeast *Saccharomyces cerevisiae* contains about 30 copies of a transposon, Ty1, per haploid genome. Its transposition has been studied in vivo and in vitro, and has been shown to occur by a mechanism that resembles the replication process of retroviruses. This element has two open reading frames, *TYA* and *TYB*, that encode analogs of retroviral *gag* and *pol*. As in retroviruses, the *TYB* product is made via a translational frameshift near the end of *TYA*. Ty1 transcripts and proteins assemble into a Ty1-VLP (virus-like particle) in which reverse transcription takes place. The DNA-containing form of the VLP can then participate in an in vitro reaction in which the DNA within the VLP can be inserted into a phage λ target DNA molecule. This simple reaction requires only a divalent cation and VLPs. The VLPs will also utilize exogenously added substrates resembling the termini of Ty1 DNA. These have been manipulated so as to provide information about the DNA requirements for transposition. Recently, we have developed a PCR assay for in vitro transposition.

Using sensitive in vivo assays for transposition, we have identified a number of host factors that play a role in Ty1 transposition. One of these is a tRNA whose abundance regulates the frequency of translational frameshifting from *TYA* into *TYB*, and hence, transposition frequency. Another host factor was identified in the form of a mutation that blocks transposition. Unexpectedly, intron lariats derived from a variety of pre-mRNAs accumulate to high levels in this mutant. We have isolated the gene corresponding to this mutation and will present evidence suggesting that the gene encodes the intron lariat debranching enzyme. A model for how this enzyme may play a role in retrotransposition will be proposed.

I 016 msDNA AND BACTERIAL REVERSE TRANSCRIPTASE, Masayori Inouye and Sumiko Inouye, Department of Biochemistry, University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Medical School at Rutgers, Piscataway, NJ 08854-5635.

A peculiar type of satellite DNA called msDNA exists in myxobacteria and some natural isolates of Escherichia coli. In this molecule, single-stranded DNA is branched out from an internal guanosine residue of an RNA molecule by a unique 2',5'-phosphodiester linkage. Reverse transcriptase (RT) is required for the synthesis of msDNA. The discovery of a retroelement in bacteria raises many intriguing questions including the origin of RT, the function and the biosynthesis of msDNA, and the extensive diversity found in msDNA and RT among different bacterial strains.

On the basis of msDNA analysis of natural isolates of myxobacteria, and the codon usage in the RT gene, it is proposed that myxobacterial RT genes are the oldest and their precursor was the ancestor of all present-day retroelements. In contrast, only a minor population of wild E. coli strains contains the msDNA-synthesizing system called "retron". Furthermore, these retrons are extensively different from each other with regard to their msDNA and RT structures, indicating that E. coli retrons were more recently acquired from some foreign sources. We will describe new retrons from clinical strains of E. coli. In one case, a retron was found to be a part of a 12.7-kb foreign DNA fragment flanked by 29-bp direct repeats and integrated into the gene for selenocystyl tRNA (*selC*) at 82 min on the E. coli chromosome. Except for the 2.4-kb retron region, the integrated DNA fragment showed remarkable homology to most of the bacteriophage P4 genome. We will explore the possibility that the retron is part of a mobile genetic element excised from the E. coli genome as an infectious virion.

I 017 HUMAN LINE-1 (L1Hs) ELEMENTS. Maxine F. Singer, Laboratory of Biochemistry, National Cancer Institute, Bethesda, MD 20892

The consensus L1Hs element is 6.2 kbp long and has all the features of a nonLTR (or polyA type or classII) retrotransposon (1). L1Hs is known to be transposed in the human genome. The process is assumed to involve a full length RNA intermediate and reverse transcription catalyzed by a reverse transcriptase encoded by the 3' ORF (ORF2). One or more of the several thousand full-length L1s in the human genome is presumed to be competent for independent transposition. However, most L1Hs elements must be incompetent because they are truncated, or carry sequence variations that render them transcriptionally silent, or cannot be translated into functional proteins.

Full length, polyA$^+$ L1Hs RNA and the product of the 5'ORF1 have been detected in the cytoplasms of teratocarcinoma (NTera2D1 and 2102EP) and choriocarcinoma (JEG3) cells. Likewise, expression of a reporter gene (lacZ) fused in frame near the 5' end of ORF1 and under control of the 900bp long L1Hs 5' UTR is more efficient in teratocarcinoma cells than in other cell types tested. The 5' UTR appears to contain all the regulatory sequences required for high level and cell-type specific transcription (see abstract by Swergold and Singer). Comparison of the polypeptide predicted by ORF1 (338 codons) with sequences in the gene bank indicated no striking homology to known proteins. However, a 'leucine-zipper' occurs suggesting an ability to associate as a coiled-coil with a second polypeptide. A region of the 1273 codon long ORF2 predicts a polypeptide with homology to reverse transcriptase.

The work to be described was carried out by T. Fanning, V. Krek, D. Liebold, G. Swergold and R.E. Thayer.

(1) J.D. Boeke and V.G. Corces. 1989. Annu. Rev. Biochem. 43 403-434.

I 018 MAIZE TRANSPOSABLE ELEMENTS AND RETROTRANSPOSONS ARE SPLICED FROM PRE-mRNA., Susan Wessler, Rita Varagona and Michael Purugganan; Department of Botany, University of Georgia, Athens, GA 30602.

Several transposable and retro-element induced mutations in maize are characterized by a leaky phenoytpe. To understand the molecular basis for this expression, our lab has focused on the mutant alleles of the *waxy* (*wx*) gene which encodes an enzyme involved in amylose biosynthesis. The *wx* gene is expressed only in the endosperm and pollen where the mutant phenoytpe is viable and easily visualized. Analysis of the leaky *wx-m9* allele, which contains 4.3 kb *Ds* element in exon 10, indicated that the residual *wx* activity resulted from the splicing of the element from pre-mRNA. Subsequent studies revealed that three different classes of *Ds* elements, associated with three different *wx* mutations, are spliced from pre-mRNA in a similar manner. In all cases, splicing utilizes multiple 5' splice sites nested in one *Ds* terminus and multiple 3' splice sites either near the other terminus or in adjacent *wx* sequences. Most, but not all of the *Ds* element is processed from pre-mRNA. Whereas Group I and II introns are thought to be evolving into mobile DNA, the *Ds* elements may be mobile DNA evolving into introns.

The splicing of DNA insertions appears to be a feature common to many maize elements. We have determined that four retrotransposon-induced *wx* mutations are leaky due to the processing of element sequences from either introns or exons. Two of these four mutations display a tissue specific phenotype, leaky in the endosperm and null in pollen. Finally, in addition to the splicing events which result in functional gene products, all four retro-induced mutations encode alternatively spliced *wx* transcripts that indicate exon skipping around the retrotransposon.

Insertion Elements

I 019 THE ANALYSIS OF INTRON-ENCODED PROTEINS OF THE YEAST MITOCHONDRIAL GENOME, Ronald A. Butow, Catherine Wernette, Paul Q. Anziano Roland Saldahna, and Philp S. Perlman, Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75235.

Intron 4α (aI4α) of the coxI gene of *Saccharomyces cerevisiae* is a mobile group I intron that encodes a site-specific DNA endonuclease, I-*Sce* II, required for aI4α mobility. That intron product is unique among mobile group I introns in that it is also a latent maturase required for RNA splicing. To begin dissecting those functions, we have first purified I-*Sce* II to near homogeneity and partially characterized the protein. The active endonuclease is a dimer of identical 31 kD subunits derived by proteolytic processing of a precursor protein, p56, encoded by a fusion of coxI exons 1-4 to the aI4α ORF. I-*Sce* II, which catalyzes a 4 bp staggered ds cut near the intron insertion site, is most active at ~30°C and at neutral pH. From site-directed and randomized oligonucleotide mutagenesis, we have identified positions over a 18 bp range spanning the cleavage site that affect the cleavage reaction. Further, we have identified cleavage sites in human genomic cosmid clones, and the analysis of four of these shows conservation of 6 out of 18 positions spanning the cleavage site. I-*Sce* II cuts human DNA infrequently (~once per 10^5 bp) and does not cleave λ DNA at all. These studies suggest a complex cleavage/recognition site specificity. As a first approach to study maturase functions, we have developed a mitochondrial transformation system using expression plasmids (pMITs) designed for transient expression and reverse genetic analysis of intron-encoded proteins. As a test system we have introduced a novel maturase gene consisting of an in-frame fusion of the coding regions of the maturase genes encoded by the group II introns 1 (aI1) and 2 (aI2) of the coxI gene. This fusion maturase gene will correct in *trans* chain terminating mutations in the maturase gene of aI1, and by recombination, *cis* mutations that block the splicing of aI1 or aI2. Deletion analysis indicates that the C-terminal portion of the fusion maturase, which includes a putative zinc finger, is required for maturase activity. Supported by Grants GM35510 and GM41426 from the NIH.

I 020 RETROTRANSPOSONS SPECIFIC TO THE 28S RIBOSOMAL RNA GENES OF INSECTS, Thomas H. Eickbush, John L. Jakubczak, William D. Burke and Dongmei Luan, Department of Biology, University of Rochester, Rochester, NY 14627.

R1 and R2 are non-LTR containing retrotransposable elements each found at a precise location in the 28S ribosomal RNA genes. The elements insert independently of each other at locations that are 74 base pairs (bp) apart. The presence of either R1 or R2 within a rDNA unit appears to inactivate that unit. Few copies of R1 and R2 are located outside the rDNA sites and they appear to be nonfunctional. The fraction of the total number of rDNA units that are occupied by R1 and R2 elements can vary between populations of a species from a few percent to well over 50%. The high insertion specificity of the R2 retrotransposable elements can be explained by an encoded endonuclease activity specific to the DNA sequence of the 28S gene insertion site. The enzyme makes a 4 bp staggered, 5' overhang cut, consistent with this product being an intermediate in the integration process. The recognition sequence for the enzyme is approximately 30 bp in length and extends almost entirely 5' of the cleavage site. Additional properties and activities of the enzyme will be discussed.

The DNA sequence of the 28S gene region surrounding the R1 and R2 insertion sites shows very little variation within eukaryotes. We have taken advantage of the remarkable insertion specificities of these two elements for such a highly conserved location in the genome to determine their distribution in a wide variety of organisms. Over 90% of the insect species tested from nine orders appear to contain these rDNA insertions. Surprisingly, single species were found to harbor multiple families of R1 and R2 elements within their rDNA loci. For example, insect species where found that contain at least six different families of R1 and/or R2 elements. These families of R1 and R2 elements within the same species are clearly distinct elements, since the level of amino acid divergence for their encoded endonuclease is as great as the levels seen between elements from widely different species. The possibility of the horizontal transfer of these elements will be discussed. Finally in our analysis of other animal taxa, R1 and R2 elements were not found outside of Arthropod species. However, we occasionally did encounter other insertion elements located near the R1 and R2 sites which presumably are also retrotransposable elements.

I 021 GROUP II AND GROUP III INTRONS AND TWINTRONS OF *EUGLENA* CHLOROPLAST OPERONS, Richard B. Hallick[1,2], Donald W. Copertino[2], Robert G. Drager[2], Ling Hong[1], Kristin P. Nelson[2] and Jennifer K. Stevenson[2], Depts. of [1]Biochemistry and [2]Molecular and Cellular Biology, University of Arizona, Tucson, AZ 85721

Euglena gracilis is a unicellular photosynthetic protist that is evolutionarily quite distant from plants and green algae. The large single copy DNA of the *Euglena* chloroplast genome (125 kbp of 145 kbp) contains >100 introns, more than 50 each of group II and group III introns. This is perhaps the largest collection of introns in any known organelle genome. The group II introns are small, of average length 475 nt (range 277-629 nt), and primarily localized in photosynthesis related genes. They often lack domains typical of group II introns. The properties of the recently described group III introns are: (i) small, uniform size, range 93-111 nt; (ii) degenerate versions of group II 5'- and 3'-intron boundary sequences; (iii) no conserved secondary structure; and (iv) location primarily in genes for proteins involved in translation and transcription.

In characterizing the organization and RNA maturation pathway of *Euglena gracilis* chloroplast intron-containing operons, we have recently determined some highly novel RNA splicing events. These include the following: **(1)** A population of intercistronic introns. The intercistronic regions within chloroplast ribosomal protein operons between *rps4-rps11*, *rpl23-rpl2*, and *rpl14-rpl5* contain group III introns. In two cases, the introns lie between the ribosome binding site and the initiator codon. **(2)** A group II intron internal to a second group II intron. The *psbF* locus of *Euglena* chloroplast DNA, encoding a component of photosystem II reaction center, has an unusual 1042 nt group II intron that is formed from the insertion of one group II intron into structural domain V of a second group II intron. We have determined that a 618 nt internal intron is first excised from the 1042 nt intron of *psbF* pre-mRNA, resulting in a partially spliced pre-mRNA containing a 424 nt group II intron that is spliced in domain V. The 424 nt intron is then removed to yield the mature *psbF* mRNA. Therefore, the 1042 nt intron of psbF is a group II intron within another group II intron. We use the term "twintron" to define this new type of genetic element. **(3)** A group II intron internal to a group III intron. This has been designated a mixed twintron. Twintrons are not confined to introns of the same type. The 409 nt intron of the gene for ribosomal protein S3 was found to be a 311 nt group II intron inserted into a 98 nt group III intron. As with the psbF twintron, sequential splicing of the mixed twintron occurs. Finally, **(4)** another new class of organelle intron with a size expected for group III twintrons has also been discovered.These double-length group III introns (range 198-213 nt) are the size expected for twintrons, but may be spliced by a novel mechanism. The excision of one of these introns has been examined by cDNA cloning and sequencing. A splicing pathway was found in which a 125 nt group III intron of the 5' portion of the intron was excised resulting in a partially spliced pre-mRNA containing an 83 nt group III intron. This intron structure resembles eukaryotic alternatively spliced introns with multiple 3' splice sites. A possible mechanism for twintron formation by intron transposition and the implications of twintrons and other novel introns to gene evolution will be discussed (Supported by Grants from the NIH).

I 022 RETROVIRAL DNA INTEGRATION, R. Katz,[1] E. Khan,[1] J. Mack,[2] G. Merkel, J. Kulkosky,[1] M. Katzman,[3] J. Leis,[4] and A. M. Skalka,[1] [1]Fox Chase Cancer Center, Institute for Cancer Research, Philadelphia, PA 19111; [2]NCI-Frederick Cancer Research and Development Center, Frederick, MD 21702; [3]The Milton S. Hershey Medical Center, Hershey, PA 17033; [4]Case Western Reserve University, Cleveland, OH 44106

In some respects retroviruses and certain other retroelements can be viewed as DNA transposable elements that use RNA intermediates. As with true DNA transposable elements, retroviral DNAs contain *cis*-acting terminal sequences and encode *trans*-acting proteins, which are required for joining to host DNA. Our studies with avian and human retrovirus (HIV), together with studies of others on HIV and murine retroviruses, have shown that only one viral protein, IN, is required for integration. Our *in vitro* analyses of the key features of the retroviral integration reaction, and functional domains of IN, highlight additional similarities between retroviral integration and recombination reactions catalyzed by eukaryotic and prokaryotic transposable elements. Such studies suggest specific evolutionary relationships and provide a productive framework for obtaining a more detailed understanding of these reactions.

Biological Diversity of RNA-Mediated Phenomena

I 023 EVOLUTION OF RNA LIGANDS: SITES RECOGNIZED BY PROTEINS, Larry Gold and Craig Tuerk, Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80309-0347

We have studied several proteins that bind to RNA and, as a result of that binding, regulate translation. In general the binding sites on target messages must overlap, directly or indirectly, the translation initiation region. Well understood proteins (and their respective target RNAs) include the R17 coat protein, the bacteriophage T4 regA, ssb, and DNA polymerase proteins, and the bacteriophage T4 regB "Shine and Dalgarno" nuclease. The target sites for these proteins illustrate, collectively, most if not all of the sequence and structural motifs available to small RNA molecules.

Our interests now include RNA sequences not found in nature. Solutions to a sequence/structure problem can be studied independently of a specific (and usually idiosynchratic) biological solution. From such work one derives new respect for those biological solutions, coupled with a deeper understanding of the narrow experimentation in "sequence space" that took place during evolutionary history. One sees as well new homologues and chemalogues (or "aptameres"), widening the appreciation for the flexibility of nucleic acids in "filling" a binding pocket in a target.

I 024 A FUNCTIONAL PSEUDOKNOT IN RIBOSOMAL RNA. Harry F. Noller and Ted Powers. Sinsheimer Laboratories, UCSC, Santa Cruz, CA, 95064.

Transfer RNA protects several sets of bases in ribosomal RNA from chemical probes. Protection of specific sets of bases is characteristic for binding to the A, P and E sites on the two ribosomal subunits. A site binding to the 30S subunit protects residues in the 1400/1500 regions (the decoding site) and also in the 530 loop, which is believed to be remote from the decoding site. Protection of the latter bases is likely to be indirect. Mutation of G530, one of the A site-protected bases, confers a dominant lethal phenotype, and the physiological effects are consistant with blockage of ribosomal A-site function. Woese and Gutell have proposed a set of tertiary base-base interactions between the 530 hairpin loop and the bulge loop around position 510, which would form a type of pseudoknot structure. We have tested their model by site-directed mutagenesis, and show that disruption of pairing interferes with 16S rRNA function. These effects are suppressed by restoration of pairing, providing direct evidence for the tertiary interaction. Certain mispairs confer unexpected phenotypes that begin to shed light on the functional role of this universally conserved feature of ribosomes.

Molecular Evolution of Introns and Other RNA Elements

Exon Shuffling and Alternative Splicing

I 025 THE INTERPLAY BETWEEN *TRANS*-SPLICING AND *CIS*-SPLICING IN *C. ELEGANS,* Tom Blumenthal, Richard Conrad, Ruey Fen Liou and Jeffrey Thomas, Department of Biology, Indiana University, Bloomington, IN 47405

In nematodes both conventional splicing (intron removal) and *trans*-splicing occur on the same transcripts. In *trans*-splicing a short 5' untranslated leader is donated to the transcripts of some genes by a 100 nucleotide RNA molecule (SL RNA) which exists in the cell as a snRNP. *Trans*-splicing and *cis*-splicing are mechanistically quite similar, and we have been studying the interplay between these two related reactions using transgenic worms. Since the two processes use the same 5' and 3' splice sites and both occur by similar branched intermediates, it is reasonable to presume that the standard snRNPs catalyze both. The exception appears to be U1 which recognizes the 5' splice site in *cis*-splicing but may be replaced by the SL snRNP, which contains its own 5' splice site, in *trans*-splicing. The key difference between *trans*-splicing and *cis*-splicing is that the former involves a recipient molecule containing a 3' splice site with no 5' splice site upstream. We have shown that this feature determines which transcripts get *trans*-spliced. It is possible to convert a conventional gene into a *trans*-spliced gene by placing an intron without a 5' splice site into its 5' untranslated region. We have also demonstrated the converse: a *trans*-spliced gene can be converted to a *cis*-spliced gene by placing a 5' splice site upstream of what is normally its 3' *trans*-splice site. Clearly, the presence of the 5' splice site is the key difference, since a *trans*-splice site will act as a *cis*-splice site if given the opportunity and a *cis*-splice site will act as a *trans*-splice site if denied access to a 5' splice site.

The importance of the highly conserved pyrimidines in the UUUCAG *trans*-splice site consensus of the *rol-6* gene have also been investigated. There exists a cryptic *trans*-splice site 20 bases upstream of the normal *trans*-splice site. If the normal site is inactivated, *trans*-splicing occurs at the upstream site. Point mutations in the U's alter the site of *trans*-splicing, demonstrating their importance in 3' splice site definition.

C. elegans contains a second spliced leader (SL2), donated by a different SL snRNP. The transcripts of three genes have been reported to receive specifically SL2. A comparison of the sequences of these genes in the region of the SL2-accepting *trans*-splice site reveals no obvious sequences held in common. We have therefore initiated experiments designed to determine what sequences on the SL2 recipients and on SL2 RNA result in SL2-specific *trans*-splicing. Replacement of the 20 bases upstream of the *rol-6 trans*-splice site (an SL1 acceptor) with 20 bases from the same region of an SL2 accepting gene (*gpd-3*) does not result in a change of specificity: the construct is *trans*-spliced to SL1.

I 026 SOME WIDELY DISTRIBUTED SEQUENCE MOTIFS FOUND IN DIVERSE ANIMAL PROTEINS: SPREADING AND SURVIVAL PROBLEMS, Russell F. Doolittle, Center for Molecular Genetics, University of California, San Diego, La Jolla, CA 92093-0634

A variety of otherwise unrelated proteins are known to contain a relatively small set of readily recognized sequence motifs. These sequences are sufficiently similar and frequent that it is virtually impossible that they could have been concocted independently. Rather, it appears that the genetic material for these motifs has been generously spread around the genome, especially during animal evolution, by the phenomenon usually referred to as "exon shuffling." As a first step in addressing how these units spread and survive, we have compiled an inventory of some of the better-defined units, including the EGF-domain, the β_2-glycoprotein unit, kringles, and the three motifs found in fibronectin. There are several features of these particular sequence patterns that warrant our attention. First, their distribution appears to be confined to animal extracellular proteins (or the extracellular portion of transmembrane proteins). Second, combinations of these patterns are often found in the same proteins or interacting proteins. Many of the patterns correspond to exons, and it has been assumed that the surrounding intron plays a role in promoting their spread. Why this should be so is not immediately obvious, although in some instances, the inconsequential vulnerability of introns to invasion by transposable or repetitive elements may provide increased opportunities for homologous crossing over. Beyond that, the intrinsic property of the expressed protein sequences to fold independently and without influencing the folding of the sequences around them may be the key. All the same, the fact that these motifs occur together so often implies interactive, as opposed to independent, structural compatabilities.

I 027 REGULATED MUTUALLY EXCLUSIVE EXON SPLICING. Bernardo Nadal-Ginard, Christopher W.J Smith, James G. Patton and Mary Mullen. Howard Hughes Medical Institute. Department of Cardiology, Children's Hospital. Departments of Pediatrics, Cell and Molecular Physiology, Harvard Medical School. Boston, MA. U.S.A.

Alternative pre-mRNA splicing has emerged in recent years as a widespread device for regulating gene expression and generating protein diversity. In addition, it has provided new insights into some fundamental aspects of splicing. This mode of regulation is particularly prevalent in muscle cells, where genes such as Troponin T are able to generate up to 64 different isoforms from a single transcriptional unit. The α-tropomyosin gene has proven particularly useful for the analysis of the mechanisms involved in this type of regulation. Exon 2 and 3 of this gene are a mutually exclusive pair that are sterically inhibited from splicing together by the presence of an abnormally positioned upstream branch point. Exon 3 is incorporated constitutively as the default choice in all tissues except in smooth muscle cells where it is replaced by exon 2.

Analysis of a large variety of constructs containing this region of the α-TM gene in different cell types and in vitro cell-free system demonstrated that the default exon selection is determined through competition between the two exons rather than down-regulation of exon 2. Exon 3 out-competes exon 2. This behaviour is determined by the nature of the polypyrimidine tract in front of exon 3 as well as by the sequence of the branch point. Thus, the relative quality of the pyrimidine tracts is the primary determinant of default exon 3 selection, with the relative branch point sequences also playing a role. This role of the pyrimidine tracts is mediated by trans-acting factors. Using U.V. cross-linking a 55 kd protein has been identified, partially sequenced and cloned. This protein binds with high affinity to the branch point/polypyrimidine tract. The binding of this protein to a large number of different intron sequences demonstrates a perfect correlation between affinity for this factor and splice site strength as determined by in vitro and in vivo cis-competition assays. These results suggest that the binding of this factor pays an important role in determining splice site selection and, at least in part, determines the default splicing pattern of α-TM. The demonstration that this protein factor is an essential splicing factor that binds to the pre-mRNA and to U1 snRNP in an ATP independent manner strongly suggest that exon selection occurs early in the spliceosome assembly pathway.

How is the regulated pattern determined? In vivo expression studies demonstrate that the switch to the regulated selection of exon 2 involved inhibition of the splice site elements of exon 3 and the sequences involved in this down-regulation have been identified.

I 028 TISSUE-SPECIFIC SPLICING OF THE DROSOPHILA P ELEMENT THIRD INTRON Christian W. Siebel, Patrick S. McCaw, Lucille D. Fresco and Donald C. Rio, Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142 and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

The germline restriction of pre-mRNA splicing of the *Drosophila* P transposable element third intron (IVS3) is responsible for the tissue-specificity of transposition. IVS3 is removed only in the germline, yielding an mRNA that encodes the 87kD transposase protein. By contrast, in somatic cells, IVS3 is retained in the mature mRNA allowing synthesis of a truncated 66kD protein that functions as a negative regulator of transposition. We have used a variety of biochemical assays in an attempt to identify factors involved in regulating this tissue-specific splicing event. IVS3 pre-mRNA was accurately spliced in vitro in heterologous human HeLa cell extracts but is not processed in active splicing extracts derived from Drosophila somatic cells. This observation allowed us to develop a biochemical complementation assay in which we asked what effect addition of Drosophila somatic extract had on splicing of IVS3 pre-mRNA in HeLa extract. This assay revealed that components in the Drosophila somatic extract specifically inhibited splicing of IVS3 but not splicing of several other introns tested. Ultraviolet photochemical crosslinking, native gel electrophoresis and binding competition experiments led to the detection of specific RNA binding proteins that bound preferentially to IVS3 pre-mRNA yet did not bind efficiently to other introns that were spliced in the presence of Drosophila somatic extract. The 97kD protein bound to the IVS3 pre-mRNA in 5' exon sequences contained within a genetically-defined cis-acting regulatory region immediately adjacent to the IVS3 5' splice site. Mutations in the 5' exon sequence that inactivate a pseudo-5' splice site (see below) result in activation of accurate IVS3 splicing in HeLa cell extracts as though 5' splice site competition or occlusion is occurring in the presence of excess U1 snRNP.

We have also detected another RNA binding protein of 65kD that binds to the IVS3 5' exon, perhaps by recognition of a sequence resembling a 5' splice site. Indeed, this protein bound only to RNAs containing a 5' splice site-like sequences and could be involved in specifying or facilitating 5' splice site recognition by U1 snRNP. The presence of a 5' splice site-like sequence in the 5' exon as well as a total of six such "pseudo-5' splice sites" surrounding the accurate 5' splice site suggests that competition among these sites for binding of U1 snRNP or other 5' splice site recognition factors could play a role in the inhibition of IVS3 splicing in somatic cells. The binding of U1 snRNP to the accurate IVS3 5' splice site and the pseudo-5' splice sites as well as the possible influence of the 97kD and 65kD RNA binding proteins on U1 snRNP-pre mRNA interactions are being investigated. We have also shown that Drosophila splicing extracts contain biochemical activities functionally analogous to mammalian splicing factors U2AF and SF2/ASF. These activities are being purified and characterized. Results using 2'-o-methyl anti-sense oligonucleotide U1 snRNP-depleted extracts will also be presented.

Other Catalytic RNAs

I 029 **THE HAMMERHEAD RIBOZYME DOMAIN: THE USE OF EXPERIMENTAL DATA AS CONSTRAINTS FOR A 3-D MODEL**, Robert Cedergren, Daniel Gautheret, François Major*, Éric Fillion, Marcel Turcotte* and Guy Lapalme*, Département de biochimie and *d'informatique et de recherche opérationnelle, Université de Montréal, Montréal, Québec H3C 3J7, Canada.

We have exploited the technique of deoxyribonucleotide substitution in RNA (mixed RNA/DNA polymers) to evaluate the role of 2'-OH groups in the hammerhead domain of catalytic RNA and have found that nucleotide hydroxyls of one G and one A of the single-stranded consensus region influences catalytic activity of the ribozyme. The purported role of these 2'-OH's in the binding of Mg2+ has been confirmed by the fact that the catalytic ability of some deoxyribonucleotide-containing ribozyme analogs having reduced activity are stimulated in the presence of higher Mg2+ concentration. This data can be interpreted in terms of distance constraints: the Mg2+ ion must be in close contact with the scissile phosphate and the influential 2-OH's. The approximate distance between the hydroxyls and the scissile phosphate can be determined using the diameter of a solvated cation.

RNA structure prediction can be defined as the combinatorial problem of finding the spatial arrangement of all atoms consistent with known constraints. The heavy computational demands of this definition are greatly reduced using a constraint satisfaction problem algorithm augmented with help functions in a symbolic programming environment. In our protocol, atomic level 3-D structures generated in this environment are refined by energy minimization. The precision of models produced by this algorithm (MC-SYM) was evaluated by comparing predicted RNA loop and pseudoknot structures derived from structural information (excluding crystal structure data) with known or consensus structures. The root mean square deviation among predicted structures and between them and the "known" structures is on the order of 1.3 Å.

Entry data for the hammerhead 3-D structure determination by MC-SYM include the secondary structure, distance constraints inferred from the activity of deoxyribonucleotide-substituted analogs and knowledge of the mechanism of the reaction. From the 10^9 different structures that could be generated, MC-SYM finds 7 which satisfy all constraints. One of these solutions has a potential energy value of -1,500 Kcal/mol compared to -991 Kcal/mol for a model which was obtained solely by energy minimization. The modeling procedure can easily be applied to other catalytic RNA molecules in the hopes that the comparison of these structures will give some insight to the evolutionary relationships among the variety of RNA catalytic processes. (Supported by MRC of CANADA and FCAR of Québec)

I 030 CLEAVAGE AND LIGATION MEDIATED BY A NEUROSPORA MITOCHONDRIAL PLASMID RNA. Barry J. Saville and Richard A. Collins, Department of Botany, University of Toronto, 25 Willcocks St., Toronto, Ontario, Canada M5S 3B2.

Neurospora VSRNA is a mitochondrial single-stranded RNA that combines certain features of catalytic RNAs, Group I introns and retroelements. VSRNA synthesized in vitro performs a magnesium-dependent, site-specific, self-cleavage reaction which yields 2',3' cyclic phosphate and 5' hydroxyl termini typical of the products of many catalytic RNAs. However, VSRNA shows no substantial sequence similarity to these other RNAs, suggesting that the nucleotides and/or the RNA structure involved in cleavage is different. The in vitro cleavage site is at the same position as the 5' end of a prominent class of linear VSRNAs found in vivo, suggesting that this reaction is involved in the natural processing pathway of VSRNA.

Cleavage can be reconstituted in trans using separate, non-overlapping VSRNA fragments. One fragment contains the cleavage site; the other apparently contains the sequences that facilitate site-specific cleavage. Also, monomer VSRNA formed by self-cleavage of a multimeric transcript can perform RNA-mediated ligation reactions to produce circular monomer and linear dimer RNAs indistinguishable from those isolated from mitochondria. These reactions require magnesium and can occur in vitro in the absence of protein.

I 031 The Molecular Biology of Hepatitis Delta Virus RNA.

Michael M. C. Lai[1] and Huey-Nan Wu[2], [1]Howard Hughes Medical Institute and Department of Microbiology, University of Southern California School of Medicine, Los Angeles, CA 90033 and [2]Institute of Molecular Biology, Academia Sinica, Nankang, Taipei, Taiwan 11529.

Hepatitis delta virus (HDV) contains a 1.7 kb circular RNA genome, which has a high degree of self-complementarity and can undergo self-catalyzed cleavage and ligation. These properties suggest that HDV RNA is closely related to viroid or virusoid RNAs. However, the antigenomic sense HDV RNA possesses an open reading frame capable of encoding a protein, hepatitis delta antigen (HDAg). Thus, HDV RNA is distinct from other ribozyme RNAs, which do not have a coding capacity. We have investigated the ribozyme activity of HDV RNA. By constructing nested set deletion mutant RNAs, we have demonstrated that the minimum RNA sequences required for the self-cleavage of HDV RNA resides within a stretch of sequence of no more than 87 nucleotides. By performing site-specific mutagenesis, we have determined the RNA conformation involved in the ribozyme activity. This RNA conformation has also been confirmed by probing with various ribonucleases and diethylpyrocarbonate (DEPC). The HDV ribozyme does not contain the consensus sequence and conformation characteristic of the virusoid "hammerhead" RNAs. These results indicate that HDV RNA represents a new class of ribozyme. The HDV ribozyme can also act *in trans*.

We have also shown that HDAg interacts with the ribozyme domain of HDV RNA. The HDAg is a phosphoprotein localized exclusively in nuclei or nucleoli of the infected cells. It binds specifically to the HDV RNA. One of the binding sites is in the 3'-side of the ribozyme cleavage site. Two arginine-rich, nucleic acid-dinding motifs in the middle domain of HDAg are responsible for this binding. The protein-RNA binding likely plays a significant role in HDV RNA replication and its ribozyme activity.

I 032 VIROIDS AND SATELLITE RNAs: DID ROLLING CIRCLE RNA REPLICATION EVOLVE IN THE RNA WORLD?, Diener T.O., Center for Agricultural Biotechnology and Department of Botany, University of Maryland, College Park, MD 20742 and Agricultural Research Service, U.S.Department of Agriculture, Beltsville, MD 20705. Viroids, the smallest known agents of infectious disease, are small (246-375 nt.), covalently closed circular, single-stranded RNAs that, in contrast to viruses, are not encapsidated. They cause a number of important plant diseases. Despite their severely limited genetic information, viroids replicate autonomously in susceptible cells. Viroids do not function as mRNAs; they are replicated by a host enzyme, probably RNA polymerase II. Viroidlike satellite RNAs resemble viroids in size and molecular structure, but are found within the capsids of specific helper viruses, on which they depend for their own replication. Both types of RNA replicate by transcription, from the circular templates, of opposite polarity oligomers of as many as 7 tandem repeats, hence requiring a precise mechanism for the excision of monomers and their ligation to form circular progeny molecules. One viroid and known viroidlike satellite RNAs are self-cleaving ribozymes, whereas other viroids require a host factor. Phylogenetic analyses of their nucleotide sequences suggest that these RNAs are of monophyletic origin. Viroids and viroidlike satellite RNAs may represent "living fossils" of precellular evolution in a hypothetical RNA world -- widely believed to have existed before the advent of DNA or protein. In primitive, template-directed replicating systems, rolling circle replication of circular templates involving oligomeric intermediates would have been advantageous in overcoming the presumed error-proneness of such systems, because no initiation or termination signals are required and because redundancy of the genetic information would be automatic. Also, mutation in the cleavage cassette would result in doubling of the genome, suggesting one mechanism by which larger genomes could have evolved. Indeed, remnants of primary structure periodicity are recognizable in all but one viroid.

Molecular Evolution of Introns and Other RNA Elements

Evolutionary Models of Introns and RNA Elements

I 033 INTRONS: HISTORY AND PHILOSOPHY, W. Ford Doolittle, Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia, Canada, B3H 4H7

There are two schools of thought concerning the origins of introns -- the "introns early" (or exon theory of genes) school, which would have it that modern splicesomal introns are relics of precellular gene assembly processes, and the "introns late" school, which sees splicesomal introns as degenerate members of transposable element families which may still be active in eukaryotic genomes. Possibly, neither extreme view is correct, introns being of mixed origin. So the real question(s) is (are): were genes first put together from smaller parts (exons), and do modern genes show in their structure any evidence of this? Various kinds of data bear, with different degrees of logical force, on these questions. For instance, the fact that many mitochondrial (and phage T4) introns are also highly evolved DNA homing devices has nothing whatever to do with the case, while evidence that the eukaryotic versions of genes present also in eubacteria have introns which clearly define structural modules at the the level of protein is potentially compelling support for the antiquity of at least those introns -- but such structural correlations are not alas always clear. Most recently, good arguments have been based on the presence, in genes which "came in" with the endosymbiotic ancestors of organelles, of introns in positions identical to those in which their always-nuclear homologs show interruptions. These and other components of the debate will addressed and evaluated, in an historical context.

I 034 EVOLUTIONARY DYNAMICS OF INTRON GAIN AND LOSS, Donal A. Hickey, Dept. of Biology, Univ. of Ottawa, Ottawa, Canada, K1N 6N5.

Both the origin and subsequent evolution of introns constitutes a major unsolved problem in molecular evolution. There is a large and highly-developed body of population genetics theory available on the evolutionary dynamics of classical mendelian genes. Despite the existence of this theoretical framework, and despite the importance of the "intron problem", there has been relatively little work done on the application of quantitative population genetics theory to the evolution of introns. I will outline the major features of one possible approach to the quantification of the evolutionary forces which affect the structure and distribution of introns. The model is based on a consideration of Darwinian natural selection acting at both the cellular and sub-cellular (molecular) levels. The balance between selective forces acting at different levels of organization will be discussed. Because of the transposition capability of introns, their evolution is in some ways similar to the evolution of other transposable elements. There are, however, special evolutionary constraints on introns that do not apply to other transposons. An understanding of these selective forces can help explain some aspects of intron function, such as self-splicing. Quantitative evolutionary models can also be used to make predictive statements about the patterns of distribution of introns among various eukaryotic and prokaryotic genomes. Finally, the possible evolutionary relationships between different classes of introns will be discussed in light of this model.

I 035 GENOMIC TAGS: FROM RNA VIRUSES TO RETROELEMENTS TO TELOMERES
Nancy Maizels and Alan M. Weiner, Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, CT 06510.

We suggest that some of the diverse styles of genomic replication evident in contemporary organisms may be molecular fossils of ancient replication mechanisms. A variety of evidence supports our earlier suggestion that 3'terminal tRNA-like structures tagged ancient genomes for replication in the RNA world [Proc. Nat. Acad. Sci. USA 84:7383-7387 (1987)]. Recent work on Neurospora mitochondrial retroplasmids and ciliate telomerases now enables us to trace the evolution of 3' terminal tRNA-like genomic tags from the RNA world to the contemporary DNA world. We will show how tRNA-like structures evolved from genomic tags on RNA genomes in the RNA world, to genomic tags on transcripts of retroplasmids in the early DNA world, to tRNA primers for retroviral reverse transcription in the middle DNA world, and most recently to an internal RNA template for the telomere terminal transferase (telomerase) that replenishes chromosome ends in the contemporary DNA world. Thus many viruses (and perhaps extrachromosomal elements?) can be regarded as contemporary examples of the replication strategies employed by more ancient genomes. Modern duplex DNA chromosomes in fact represent a highly evolved version of the retrovirus replication strategy, in which the RNA transcript that primes the initiation of replication is a degenerate form of the retroviral tRNA primer, and the telomerase that completes replication is a specialized form of reverse transcriptase.

In addition to providing a unifying view of the evolution of replication strategies from the RNA to DNA world, genomic tags also provide a plausible model for the origin of encoded protein synthesis. A central problem in envisioning the evolution of modern protein synthesis has been that none of the individual components -- ribosomal RNAs and proteins, initiation and elongation factors, tRNAs and tRNA synthetases -- appears to be useful alone, yet the molecular apparatus for translation must have evolved stepwise. We therefore proposed {ref} that tRNA was the first component to evolve, originating as 3' terminal tRNA-like structures that tagged genomes for replication by serving as initiation sites for the replicase and functioning as primitive telomeres. An enzyme capable of aminoacylating 3' terminal tRNA-like structures may then have evolved in the RNA world to facilitate RNA replication. Thus the first tRNA synthetase would have been made of RNA, and these aminoacylating activities would have been perfected for use in replication before being coopted for protein synthesis. This model implies that RNase P evolved to convert genomic into functional RNA molecules by removing the 3' terminal tag, that the CCA-adding activity was originally a telomerase for RNA genomes, and that the structural diversity of modern tRNA synthetases reflects divergent pathways for descent from ancestral RNA enzymes.

I 036 BEFORE THE RNA WORLD, Leslie E. Orgel, The Salk Institute for Biological Studies, P.O. Box 85800, San Diego, CA. 92186-5800.

If there was an RNA world it must have evolved in an environment containing adequate concentrations of nucleotides. Could they have been synthesized abiotically, or must they have been the products of the metabolism of some more primitive organized system with a different genetic material? Relevant experimental evidence will be discussed.

Late Abstract

THE GENETIC MOBILITY OF GROUP I INTRONS AT THE DNA LEVEL,

Bernard Dujon, Arnaud Perrin, Ilana Stroke, Ettore Luzi, Laurence Colleaux, Anne Plessis and Agnès Thierry, Unité de Génétique moléculaire des levures, Institut Pasteur, 25 Rue du Dr. Roux F-75724 Paris-CEDEX 15, FRANCE.

A number of group I introns from various phylogenetic groups are now known as mobile genetic elements that propagate at predetermined positions of genes (a phenomenon called *homing*) as a consequence of the action of the endonucleases they encode. All known properties of intron homing indicate that it is a DNA mediated mechanism that involves two critical steps: -1- the site specific recognition and cleavage of the intron-less gene by the intron encoded endonuclease and -2- the subsequent repair of the double strand break of the intron-less copies of the gene, using as template the uncleaved intron-plus copies of the same gene. Our present level of understanding of these two steps markedly differ. Step 1 can be successfully completed *in vitro* with purified elements allowing precise protein-DNA interactions to be studied in details. Step 2, on the contrary, requires "cellular factors" that have not yet been characterized.

Group I intron-encoded endonucleases belong to at least two different and heterogeneous families of site specific enzymes, differing in their recognition sites and their mode of cleavage. Most intron-encoded endonucleases of eucaryotic origin contain a common dodecapeptide motif and produce 4 bp staggered cuts with 3' OH overhangs, suggesting that they act on the minor groove of DNA. We have shown that the endonuclease I-*Sce* I, encoded by the mobile intron Sc LSU•1 of yeast, recognizes an 18 bp long sequence, forming the *homing site* of the intron, and catalyzes double strand cleavage *in vitro* in an Mg^{++} dependent reaction. Because the recognition sequence is partly degenerated, the exact specificity of the enzyme has been directly assayed on genomic DNA of various organisms and proved that I-*Sce* I is, so far, the only endonuclease able to cleave an entire eucaryotic genome such as yeast at a single site. Secondary sites found in some genomes are now under study in an attempt to better characterize the rules of site-recognition for this new class of endonucleases. Mutants of the I-*Sce* I endonuclease have been isolated from artificial genetic constructs. They demonstrate the functional importance of the dodecapeptide motifs and of the N-terminal domain of the protein.

In an attempt to characterize the second step of the homing process, we have expressed the I-*Sce* I endonuclease from appropriately engineered artificial yeast nuclear vectors and found that its activity initiates a double strand break repair mechanism in the nucleus with the same intermediates as those observed when using the nuclear coded HO endonuclease. In particular, the I-*Sce* I generated double strand break can be observed for more than 30 minutes without significant degradation, suggesting that the ends are protected. Experiments are now in progress to try to visualize intermediates of the reaction in the normal situation in mitochondria.

Molecular Evolution of Introns and Other RNA Elements

Poster Session I

I 100 PROPERTIES OF DOWNSTREAM PROMOTERS FOUND IN DROSOPHILA RETROELEMENTS,

Irina R. Arkhipova[1,2] and Yurii V. Ilyin[1], [1]Engelhardt Institute of Molecular Biology, Moscow 117984, USSR and [2]Department of Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, Scotland, UK

Drosophila melanogaster retrotransposons are RNA polymerase II transcription units, transcription of which is initiated and terminated within the long terminal repeats (LTRs). In an earlier study (Arkhipova et al., Cell 44, 555) we have shown that transcription of retrotransposons mdg1, mdg3 and mdg4(gypsy) is initiated at a unique site at or near the common sequence TCAGTY in the absence of discernable TATA motifs. Analysis of the promoter activity for 5'- and 3'-deletion mutants of mdg1 LTR shows that the sequences responsible for correct and precise initiation of RNA synthesis at basal levels lie within the transcribed region. A similar type of RNA polymerase II promoters has been described in LINE-like Drosophila retroelements (Mizrokhi, Georgieva and Ilyin, Cell 54, 685). A comparison of promoter sequences between retroelements and Drosophila developmentally regulated genes which are known to have transcriptionally important downstream regions reveals two types of a common downstream sequence element. Experiments on DNA-protein interactions indicate the presence of a protein factor binding to the downstream consensus sequence. The existence of such a promoter organization in retrotransposons as well as LINEs indicates that the two classes of retroelements may be more similar than previously suggested.

I 101 SEQUENCES INHIBITORY TO SPLICING ARE PRESENT IN THE INTRON OF ROUS SARCOMA VIRUS RNA, Karen L. Beemon and Mark T. McNally, Department of Biology, Johns Hopkins University, Baltimore, MD 21218

The retroviral life cycle requires incomplete splicing of the primary viral RNA transcript to generate both spliced and unspliced mRNAs. In the case of Rous sarcoma virus (RSV), no viral regulatory proteins appear to be involved in controlling the balance of spliced and unspliced RNAs. However, deletion of sequences from the intron of the RSV pre-mRNA resulted in an increase in levels of spliced RNAs. These sequences, termed a negative regulator of splicing (NRS), were inserted into introns of heterologous genes, where they caused an accumulation of unspliced RNA in the nucleus. Maximal splicing inhibition was observed with a 300-nt fragment (RSV nts 707-1006), located 300 nts downstream of the RSV 5' splice site. Within this fragment, two non-contiguous domains were both found to be essential for NRS activity. The NRS element was active only in the sense orientation. While it inhibited splicing from several different intron positions, it did not have any effect at an exon position >400 nts upstream of the 5' splice site, nor did it affect the splicing of a downstream intron. The presence of a polypyrimidine tract in an essential region of the NRS element suggests that it may inhibit splicing by binding splicing factors or by competing with 3' splice sites.

I 102 REARRANGEMENTS IN MITOCHONDRIAL DNA AND INTRON MOBILITY IN Podospora anserina, Léon BELCOUR, Annie SAINSARD-CHANET, Carole SELLEM, Odile BEGEL, Michèle ROSSIGNOL and Marguerite PICARD* Centre de Génétique Moléculaire du CNRS, 91198 Gif-sur-Yvette, FRANCE. * Université Paris XI, 91405 Orsay, FRANCE.

Amplification of intron alpha (cox1-i1, a self-splicing group II intron) as circular multimeric DNA molecules and disorganization of the residual mt-DNA are known to be associated with the syndrom of senescence in the fungus Podospora anserina.

1°-In accordance with the idea that intron alpha plays a major role in the process of senescence, a protein encoded by this intron was detected by immunological methods in some senecent cultures and only in senescent cultures. 2°-Highly efficient transposition (homing) for a class I (cytb-i3) and a class II (cox1-i4) intron, associated with the co-conversion of flanking markers over several kbp have been evidenced. 3°-Mycelia carrying two particular nuclear genes have the following characteristics: a) a premature cessation of growth; b) a systematic site specific deletion of mt-DNA whose one of the boundaries is exactly localized at the 5' end of the intron alpha; c) the accumulation of an RNA species corresponding to the first cox1 exon (the exon 5' to intron alpha).

These data and previous studies of mutants selected as escaping senescence suggest that the cleavage at the 5' end of intron alpha, most probably correlated with its mobility, is a crucial step in several mt-DNA rearrangements in Podospora (senescence and premature death). Furthermore, this work allowed the identification of two nuclear genes involved in the control of the stability of the mitochondrial genome.

I 103 FUNCTIONAL STUDIES OF DOMAINS OF GROUP II INTRONS OF YEAST MTDNA, S. D. Dib-Hajj, S. C. Boulanger, S. M. Belcher, J. L. Koch and P. S. Perlman, Department of Biochemistry, UT Southwestern Medical Center, Dallas, Texas 75235 and Department of Molecular Genetics, Ohio State University, Columbus, Ohio 43210.

Point and full or partial domain deletion mutants of group II introns of yeast mtDNA were constructed and characterized *in vitro*. Ten point mutations of d5 of aI5γ were analyzed most of which have little effect on self-splicing even though they disrupt pairing at sites in conserved helical regions of d5 . A highly conserved G-U basepair located near the base of d5 was saturated for point mutations. All changes of the G residue greatly slowed self-splicing while all changes of the U residue had slight effects on the splicing rate. This shows that a G residue is important at that site but that its being paired is not essential. A double mutant having A-U instead of G-U at that position as well as a second mutation in d5, which alone has little effect on splicing, blocks splicing almost completely. We have also studied the function of the C1 substructure of domain 1, shown by others for aI5γ to play a minor role in 5' splice site selection. Point mutations of bases proposed to pair with the 5' end of the intron *block* splicing in aI1 but not in aI5γ. Various deletions of the C1 substructure of aI5γ and aI1 block splicing; these include partial deletions that leave the presumed functially important portion of C1 intact.

I 104 RETROIDS OF SOME MARINE INVERTEBRATES. Roy J. Britten, Mark S. Springer, David S. Kossack and Eric H. Davidson. California Institute of Technology, 101 Dahlia Ave, Corona del Mar CA 92625.

Retroid copies are interspersed in the DNA of all sea urchins yet tested including species with lineages that diverged 200 million years ago. About 5kb long, the retroids include a gag gene, two copies of an RNA binding site, protease, reverse transcriptase (RT), RNAse H and integrase gene regions. It is not known if they include an LTR or if they are retroviruses or retrotransposons. Their close relatives, based on RT amino acid sequence similarity (shown in parentheses), are yeast Ty3 (37%), Drosophila gypsy (27%), cauliflower mosaic virus (24%). The sea urchin retroids occur in many subfamilies showing about 70% RT amino acid sequence similarity to each other. The predominance of synonymous substitutions in such comparisons demonstrates sequence dependent selection and thus active examples of the retroids exist that depend on the RT gene for their replication and survival. Subfamilies may include 10 to 50 copies in the DNA of one species that are quite similar to each other (1% to 5% DNA sequence divergence). Members of the same subfamilies occur in related sea urchin species that had a last common ancestor 30 to 50 million years ago.

I 105 IDENTIFICATION OF A POTENTIAL RNA INTERMEDIATE FOR TRANSPOSITION OF THE LINE-LIKE ELEMENT I FACTOR IN DROSOPHILA, Alain Bucheton, Marie-Christine Chaboissier, Alain Pélisson and Isabelle Busseau, Centre de Génétique Moléculaire, CNRS, 91198 Gif-sur-Yvette, Cédex France

LINEs are transposable elements found in many eukaryotic species including plants, protists, insects and mammals. Their transposition is usually difficult to study, even in humans where some diseases have been shown to result from LINE insertion mutations, because most copies are defective and their frequency of transposition is very low.

The I factor is related to LINEs and controls the IR system of hybrid dysgenesis in Drosophila melanogaster. It transposes at high frequency in the germ-line of the female progeny of crosses between females of the reactive class of strains (which contain functional I factors).

We have introduced into an I element a DNA fragment containing an intron. This marked I element has been introduced into the genome of dysgenic females where its transposition resulted in precise removal of the intron, providing the first direct evidence for retrotransposition of LINEs. Northern blot and S1 mapping experiments have shown that a full-length RNA of the I factor is synthesized in the ovaries of dysgenic females. It has all characteristics to be the RNA intermediate required for transposition.

I 106 CAN THE RATE OF SLICING INFLUENCE THE RATIO OF TWO ALTERNATIVELY CHOSEN 3' SPLICE SITES? Daniela Castanotto and John J. Rossi, Department of Molecular Genetics, City of Hope, Duarte, CA 91010

We have constructed sequence variants of the *Saccharomyces cerevisiae* ACT intron to follow alternative splicing in a cis competitive context. In our construct we have inverted the UACUAAG and UACUAAC signals creating a new 3' splice site 40 nucleotides upstream of the normal one [UACUAAC...UACUA**AG**.....AG]. We show that in vitro there is an equal utilization of the two 3' splice sites, in contrast to the in vivo reaction, where the utilization of the first 3' splice site is 10 times greater than the second or normal AG acceptor. It is our belief that splicing is occurring on nascent transcripts. Thus, it follows that the first synthesized AG is used preferentially in vivo. We are presently testing whether the rate of splicing can influence the recognition of the two 3' splice sites and trying to obtain in vivo a similar ratio of AG usage previously observed in vitro. The rule "first come first serve," however, does not apply to branch point selection. In fact, in our second construct, where we have created two canonical branch points separated by seven nucleotides, the second (or downstream) UACUAAC is still favored over the first one. When we combined the above mentioned mutation with an AG to AC mutation at the 3' splice site [UACUAAC...UACUAAC.....AC], the ratio of branch point usage was found to be 20:80 in favor of the downstream UACUAAC. This preference might be due to interactions with 3' splice site recognition factors or to the overall structure of the intron. Experiments are in progress to distinguish between these two alternatives and to test the possible involvement of sequences surrounding the 3' splice site as being important in yeast branch point recognition.

I 107 ISOLATION OF A MUTANT DEFECTIVE IN Ty1 TRANSPOSITION AND INTRON DEBRANCHING IN *SACCHAROMYCES CEREVISIAE*
Karen B. Chapman and Jef D. Boeke, Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205.

The yeast transposon Ty1 transposes through an RNA intermediate, by a mechanism similar to that of retroviral replication. Using a genetic screen aimed at identifying cellular factors involved in Ty transposition, we have identified a mutation in a host gene that reduces transposition frequency. The mutant, *dbr1-1*, is also defective in the process of intron turnover. In *dbr1* mutant cells, excised introns are remarkably stable and accumulate in the form of a lariat which is missing the RNA sequences 3' to the branchpoint. The *DBR1* gene has been isolated and is shown to encode debranching enzyme, an RNA processing activity that hydrolyses the 2'-5' phosphodiester linkage at the branchpoint of the excised intron lariats. Disruption of the *DBR1* locus has no apparent deleterious phenotype. The absence of debranching activity in vivo results in the accumulation of excised intron lariats, while splicing of pre-mRNAs is apparently unaffected.

Transposition of Ty1 in *Δdbr1* mutant cells is reduced 9-fold, but not completely abolished. Debranching activity appears to be required for maximal transposition efficiency in vivo, but is not an essential function. A model for the role of debranching enzyme in the mechanism of Ty1 transposition will be presented.

I 108 GROUP II TWINTRON: AN INTRON-WITHIN-AN-INTRON IN A CHLOROPLAST CYTOCHROME B559 GENE, Donald W. Copertino[2], and Richard B. Hallick[1,2], Departments of Biochemistry[1] and Molecular and Cellular Biology[2], University of Arizona, Tucson, AZ, 85721

The *psbF* gene of chloroplast DNAs encodes the ß-subunit of cytochrome b559 of the photosystem II reaction center. The *psbF* locus of Euglena gracilis chloroplast DNA has an unusual 1042 nt group II intron that appears to be formed from the insertion of one group II intron into structural domain V of a second group II intron[1]. Using both direct primer extension cDNA sequencing and cDNA cloning and sequencing, we have determined that a 618 nt internal intron is first excised from the 1042 nt intron of *psbF* pre-mRNA, resulting in a partially spliced pre-mRNA containing a 424 nt group II intron that is spliced in domain V. The 424 nt intron is then removed to yield the mature *psbF* mRNA. Therefore, the 1042 nt intron of psbF is a group II intron within another group II intron. We use the term "twintron" to define this new type of genetic element. Intermediates in the splicing pathway were detected by Northern hybridization. Splicing of both the internal and external introns occurs via lariat intermediates. Twintron splicing was found to proceed by a sequential pathway, the internal intron being removed prior to the excision of the external intron.

[1] Michel, F., Umesono, K., and Ozeki, H. *Gene* 82: 5-30.

I 109 SEARCH FOR RELATED GROUP I INTRONS IN THE Chlamydomonas CHLOROPLAST GENOME, Marie-José Côté, Claude Lemieux and Monique Turmel, Département de biochimie, Faculté des sciences et de génie, Université Laval, Québec, QC, Canada, G1K 7P4.

The chloroplast genomes of the interfertile algae Chlamydomonas eugametos and C. moewusii feature numerous polymorphic loci, some of which result from differences in intron composition. The 14 group I introns that have been associated so far with these polymorphic regions reside in the rRNA and psbA genes. They have been classified into three subgroups: subgroups IA and IB have been well documented in other intron systems, while subgroup IC appears to be specific to the Chlamydomonas chloroplast. It is notable that the introns belonging to the most abundant subgroup (IA) share substantial sequence conservation within their core structure. This may indicate that, at least during a certain period of their evolutionary history, Chlamydomonas chloroplast genomes contained introns with the ability to move from one locus to another. We recently attempted to identify the C. eugametos and C. moewusii genes that share similar group IA introns by PCR amplification of intron sequences, using oligodeoxynucleotide primers complementary to conserved P and S elements. The PCR products were resolved by agarose gel electrophoresis and individual products were used as probes to determine their genomic locations. This approach allowed us to map group IA introns in two additional chloroplast genes, psbC and psaB, and most surprisingly, in the mitochondrial gene coding for the large subunit rRNA. (Supported by NSERC Canada).

I 110 GENE PRODUCTS REQUIRED IN tRNA SPLICING, M.R. Culbertson, D. Ursic, and F. Webb, Laboratories of Genetics and Molecular Biology, University of Wisconsin, Madison, Wi 53706.

We have previously identified mutations in two genes, *SEN1* and *SEN2*, that affect the activity of an endonuclease that cleaves introns from yeast pre-tRNAs. C. Ho and J. Abelson have shown that *SEN2* is a likely candidate for an endonuclease structural subunit. *SEN1* is probably not a structural subunit, but is nonetheless required for enzyme activity. We have concentrated on establishing the role of *SEN1* in tRNA splicing by first characterizing the *SEN1* gene and its product. The *SEN1* gene codes for a 230 kd protein that contains a consensus NTP biding site and shares amino acid sequence similarity with the yeast *UPF1* gene, which is involved in mRNA turnover. The C-terminal 140 kd of *SEN1* is essential for viability, whereas the N-terminal 90 kd is dispensable. The temperature-sensitive mutation *sen1-1* resides in the essential C-terminal region. Using immunofluorescence, we have shown that the *SEN1* protein localizes to the nucleus. By comparing the steady-state levels of *SEN1* mRNA and protein in strains carrying multiple copies of the *SEN1* gene, we have shown that expression is post-transcriptionally controlled. The N-terminal portion of the protein that is not required for viability is likely to be involved in control.

I 111 DNA ENDONUCLEASE OR RNA MATURASE FUNCTIONS OF GROUP I INTRON ENCODED PROTEINS OF YEAST MITOCHONDRIA, A. Delahodde, V. Goguel*, D. Hatat, C. Jacq, Laboratoire de Génétique Moléculaire, Ecole Normale Supérieure, CNRS URA 1302, 46 rue d'Ulm 75230 Paris Cedex 05, *Department of Biology, Brandeis University.

The eight putative group I intron-encoded proteins of yeast mitochondria belong to the same family of P1-P2 containing proteins. Previous genetic analyses have shown that three of them are required for the RNA intron splicing process (RNA maturases). Recent biochemical studies showed us that at least four out of the five remaining intron-encoded proteins might possess a specific DNA endonuclease activity. In two cases, this DNA endonuclease activity has already been shown to promote the DNA intron mobility. Considering our recent observations and the published works from different laboratories, it appears to exist a correlation between the function of the intron-encoded proteins and the ribozyme properties of the intron, suggesting that these two elements have not evolved independently. To further study this point we have focused our activity on the closely related aI4/bI4 introns which code for either a DNA endonuclease (paI4) or an RNA maturase (pbI4). Universal genetic code equivalent versions of the two intron ORFs were made. This allowed us to translate these two proteins either in E.coli, where their recombinational properties were studied, or in yeast cytoplasm where their RNA maturase properties were analyzed after they were targeted to mitochondria. Structure-function studies of the paI4 and pbI4 proteins were carried out by taking advantage of the engineered forms of the genes. In vivo hyrids of the two genes were also constructed in yeast nucleus to characterize new RNA maturase like proteins. The numerous sequence arrangements observed confirm the functional proximity of the two intron-encoded proteins and suggest that the RNA maturase activity could have derived from the DNA endonuclease activity; the two activities being mutually exclusive. An evolutionary pathway can be inferred from these observations.

I 112 RNA-MEDIATED RECOMBINATION, Leslie K. Derr, M. Joan Curcio, David J. Garfinkel and Jeffrey N. Strathern, Laboratory of Eukaryotic Gene Expression, NCI-FCRDC, Frederick, MD 21702

The existence of processed pseudogenes suggests that RNAs can act as intermediates in recombination. Additionally, it has been proposed that RNA recombination intermediates participate in gene conversion without crossing over. We have developed an extremely sensitive assay system in yeast *Saccharomyces cerevisiae* that has allowed us to detect RNA-mediated recombination and distinguish these events from events not involving an RNA intermediate. The assay is based on a plasmid-borne *his3* gene with an artificial intron inserted in an antisense orientation relative to the *HIS3* promoter (thereby inactivating *his3*). Transcripts can initiate at both the normal *HIS3* promoter and at a *GAL1* promoter placed at the 3' end of *his3*. Transcripts initiating at GAL1 can be spliced and are then equivalent to an antisense copy of a functional *HIS3* gene. The formation of His^+ prototrophs was shown to depend on galactose induction and the correct orientation of the intron for splicing from GAL-initiated transcripts. The resulting *HIS3* genes had the intron precisely removed. These data indicate that an RNA intermediate is central to the formation of *HIS3* prototrophs. *HIS3* prototrophs result from two distinct recombinational events. About 70% of the *HIS3* recombinants were due to gene conversion of plasmid *his3* sequences, consistent with RNA-mediated gene conversion. The remainder had *HIS3* inserted at novel positions in the genome as functional analogues of pseudogenes. Initial characterization of these processes suggests that the formation of *HIS3* prototrophs requires the SPT3 gene product and growth at low (20°) temperature; both of which are hallmarks of yeast retrotransposon Ty transposition.

I 113 A STRUCTURAL ELEMENT FROM A YEAST PRE-mRNA INTRON REQUIRED FOR ACCURATE 3' SPLICE SITE SELECTION SHOWS A STRIKING HOMOLOGY TO THE GAMMA DELTA TRANSPOSON FROM *ESCHERICHIA COLI*, **James O. Deshler and John J. Rossi, Department of Molecular Genetics, City of Hope, Duarte, CA 91010**

The 3' splice site of the *Kluveromyces lactis* ACT gene is highly unusual in that the branch point is located relatively far from the 3' splice site (122 nucleotides), and a PyAG is located between the branch point and 3' splice site. We designed a genetic screen to isolate randomly induced mutations between the UACUAAC and 3' splice site of this gene which activated the normally silent AG as a 3' splice site. The mutations isolated disrupted an extensive hairpin structure required for splicing to the distal AG.

Hypothesizing that this structural element of the intron might have been acquired from a transposition event, we searched GENBANK for homologies to this region of the *K. lactis* ACT intron. A portion of the hairpin was found to be highly homologous to the right terminal inverted repeat of gamma delta (Tn1000). It is interesting to note that *E. coli* can transfer its F factor (which contains a copy of Tn1000) to the yeast *Saccharomyces cerevisiae*. While we cannot say that Tn1000 was the origin of this intron element, it is possible that the *K. lactis* ACT gene acquired an insertion element between the branch point and 3' splice site which folded in such a way as to allow proper splicing to occur.

I 114 GROUP II INTRONS OF THE *EUGLENA GRACILIS* CHLOROPLAST *atp*A, *atp*B, *atp*F, *atp*H and *atp*I GENES, Robert G. Drager[1], Ling Hong[2], Donald W. Copertino[1]and Richard B. Hallick[1,2], Departments of [1]Molecular and Cellular Biology and [2]Biochemistry, University of Arizona, Tucson, AZ 85721

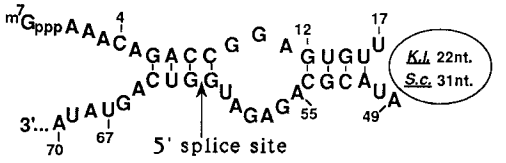
The *Euglena gracilis* chloroplast genes encoding chloroplast atpase subunits $CF_1\alpha$ $CF_1\beta$, CF_0I, CF_0III and CF_0IV; *atp*A, *atp*B, *atp*F, *atp*H and *atp*I, respectively, have been located on the chloroplast genome. *atp*I, *atp*H, *atp*F, and *atp*A are located between the large ribosomal protein operon similar to the *E. coli* S-10, spec and alpha operons and the *trn*P, *trn*S gene loci *atp*B is located between the *rbc*L and *pet*D genes. Genomic and cDNA sequence analysis have shown that these five genes are interrupted by eleven group II and eight group III introns. The group II introns are representative of the 52 group II introns which comprise approximately 25% of the single copy region of the *Euglena gracilis* chloroplast genome. This is the largest known number of group II introns from a single source. These group II introns range in size from 277 to 1605 nucleotides, generally have typical 5' and 3' boundary sequences, retain at least domain V and VI secondary structures and are excised as lariats. Intron II of *atp*F, a 361 nucleotide group II intron, is in the exact position of the only intron interrupting these genes in the chloroplast genomes of higher plants.

I 115 CHARACTERIZATION AND PURIFICATION OF I-*Ppo*, THE ENDONUCLEASE THAT MEDIATES INTRON MOBILITY IN THE rDNA OF PHYSARUM POLYCEPHALUM, Eldora L. Ellison, Donna E. Muscarella and Volker M. Vogt, Department of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853

Mobile group I introns are dependent upon the activity of a novel class of intron-encoded enzymes. We have characterized several aspects of the endonuclease that mediates homing of intron 3 in the rDNA of *Physarum polycephalum*. This site-specific endonuclease, termed I-*Ppo*, can be synthesized *in vitro* from cloned intron DNA. Translation can be initiated at either of two sites, generating functionally equivalent polypeptides of 138 or 160 amino acids in length. We have used I-*Ppo* synthesized *in vitro* to map the exact site where the endonuclease cleaves the rDNA. Cleavage generates a 4 nucleotide overhang with 3' protruding ends. To define the recognition sequence for the endonuclease we have generated a series of deletions surrounding the cloned homing site. A 16 bp sequence is sufficient for cleavage *in vitro*. I-*Ppo* can also be expressed in *E. coli* using a bacteriophage lambda-T7 RNA polymerase induction system. I-*Ppo* synthesized in *E. coli* cleaves plasmid DNA containing the homing site *in vivo*. We are currently purifying I-Ppo from crude extracts of these cells, using ion exchange chromatography and other methods.

I 116 STRUCTURAL BASIS OF THE AUTOGENOUS REGULATION OF SPLICING OF THE YEAST RIBOSOMAL PROTEIN L32 TRANSCRIPT, Francis J. Eng and Jonathan R. Warner, Depts. of Biochemistry and Cell Biology, Albert Einstein College of Medicine, Bronx, NY 10461

We have shown that overexpression of ribosomal protein L32 leads to the inhibition of splicing of the transcript of its own gene, RPL32. [PNAS,83,5854,1986] We have now determined that two sequence elements within the 5' exon and 5' splice site of the transcript are involved in this regulation of splicing. By a combination of deletion analysis, site directed mutagenesis, and sequence comparison with the closely related yeast, *K. lactis*, whose homologous gene is similarly regulated, we propose that the structure shown is responsible for preventing the 5' splice site from interacting with the splicing apparatus. Compensatory mutations in both stems confirm the validity of this structure. (Supported by an ACS Grant)



I 117 SIMILARITIES BETWEEN msDNA SYNTHESIS AND THE SPLICING OF TYPE II AND PRE-mRNA INTRONS, D. Frendewey, Cold Spring Harbor Lab, Cold Spring Harbor, NY 11724

Multi-copy single stranded DNA is an unusual DNA-RNA hybrid molecule consisting of a short single stranded DNA covalently linked to an RNA molecule (msdRNA) by a 2′–5′ phosphodiester bond between the 2′–OH of an internal guanosine near the 5′ end of the RNA and the 5′ terminal phosphate of the DNA. This structure is intriguingly similar to the branched lariat RNAs produced during the splicing of type II and pre-mRNA introns. The branch point in type II introns is located at an adenosine that is bulged from a base-paired stem. Both the bulge and stem structures are conserved and essential for branch formation. The branch point adenosine of pre-mRNA introns is also predicted to be bulged from a short helix formed by base pairing in *trans* between the branch point region and a conserved sequence in the U2 snRNA. In msDNA synthesis reverse transcription initiates with the formation of a 2′–5′ phosphodiester bond between the first deoxynucleotide of the msDNA and the branch point G in the msdRNA, which is proposed to lie at the end of a helix formed by the base pairing of inverted repeats in the msdRNA precursor. However, an inspection of the published msdRNA sequences indicates that base pairing can be extended for at least two bases beyond the branch point if the branch point G is bulged out of the helix. In addition, there is a loose primary sequence conservation around the branch point that resembles the branch point consensus sequences proposed for type II and pre-mRNA introns. Mutational analyses (Hsu et al., 1989, *J. Biol. Chem.* **264**, 6214-6219) support the hypothetical bulge structure and suggest that primary sequence near the branch site is important for efficient msDNA synthesis. It has also recently been reported (Xiong and Eickbush, 1990, *EMBO J.***9**, 3353-3362) that the reverse transcriptases associated with msDNAs are closely related to those encoded by ORFs found in some type II introns. Other similarities and their implications will be presented.

I 118 A GENE FAMILY OF GROUP I INTRONS FOUND IN THE NUCLEAR rRNA GENES OF GREEN ALGAE AND OTHER LOWER EUKARYOTES, Paul A. Fuerst+, Lee W. Wilcox*, Louise A. Lewis* and Gary L. Floyd*, Departments of Molecular Genetics(+) and Plant Biology(*), The Ohio State University, Columbus, OH 43210

Group I introns have been found in the nuclear small subunit rRNA gene of several chlorophycean green algae. Phylogenetic analyses of their sequences suggest that most are members of an evolutionary family. They are characterized by: 1) relatively small size (approximately 400 bases); 2) lack of open reading frames; and 3) similarity of primary sequence in both the core structure and in other parts of the intron. The small subunit rRNA gene of Asteromonas gracilis was found to contain two related introns; single introns in Dunaliella salina and in Characium saccatum are found at identical locations to the 5' and 3' intron in Asteromonas, respectively. Some of these algal introns have been found to self-splice. Introns which occur at identical locations in the 16S-like rRNA gene are more similar to one another in primary sequence than are those which occur at different sites. Sequence similarity of introns is strongly correlated with the phylogenetic relationship of the organisms in which they are found. The family includes the group I intron found in the nuclear 16S-like rRNA from the fungus Pneumocystis, and the set of group I introns in the nuclear 23S-like rRNA gene of Tetrahymena, each of which share significant sequence similarity to the algal introns. Outside the core structure, this family of introns has little sequence similarity with other group I introns. While mobility within the nucleus has clearly occurred, evidence for interspecific transfer of these introns is lacking.

I 119 RNA POLYMERASE III PROMOTER ELEMENTS IN THE INTRON OF THE FISSION YEAST U6 RNA GENE, Marc Gillespie and David Frendewey, Cold Spring Harbor Lab, P.O. Box 100, Cold Spring Harbor, NY 11724

The gene for the U6 small nuclear RNA (snRNA) in the fission yeast *Schizosaccharomyces pombe* is interrupted by an intron whose structure is similar to those found in messenger RNA precursors. We have investigated the structures of the U6 genes from five *Schizosaccharomyces* strains by PCR amplification of a fragment of the U6 coding sequence from the genomic DNA of each strain. The sequences of these fragments revealed that each had an intron of approximately 50 base pairs in precisely the same position. In addition to the splice sites and putative branch point regions, a sequence immediately upstream of the branch point consensus was found to be conserved in all of the *Schizosaccharomyces* U6 genes. This sequence matches the consensus for the B box of eukaryotic tRNA promoters. We have also identified a possible A box element in the first exon of the *S. pombe* U6 gene. This sequence is almost identical to the A box in the gene for the *S. pombe* 7SL RNA and is conserved in the *Saccharomyces cerevisiae* U6 gene. These results suggested that synthesis of U6 RNA in fission yeast might involve the use of internal promoter elements similar to those found in other genes transcribed by RNA polymerase III (Pol III). To test this hypothesis we have introduced a number of deletions and point mutations into the *S. pombe* U6 gene and assayed the ability of these mutant genes to serve as templates for Pol III transcription in an *S. cerevisiae* whole cell extract and in vivo in *S. pombe*. The in vitro results indicate that (i) 50 base pairs of 5′ flank is sufficient for efficient transcription, (ii) replacement of the TATA box causes a shift in the initiation site, (iii) removal of the entire 5′ flank reduces the efficiency but not the accuracy of transcription, (iv) deletion of the start site and half of the A box abolishes transcription and (v) deletion of the intron with the 5′ flank intact has no effect. In vivo experiments are in progress.

I 120 SPLICE SITE SELECTION IN THE YEAST *S. CEREVISIAE*, Valerie Goguel, Xiaoling Liao and Michael Rosbash, Department of Biology, Brandeis University, Waltham, MA 02254

We are interested in splice site selection during yeast nuclear pre-mRNA splicing. For that purpose, we duplicated either the 5' splice site or the 3' splice site, including the branch point sequence. We did not observe any preferential use among the duplicated 5' splice sites, except when one of these sites was weakened by the introduction of a mutation in the consensus sequence. If the strain carries a U1 snRNA with a mutation which allows base pairing with both the wild type and the mutant sites, the mutant site is no longer recognized as a weak site, indicating the role of U1 snRNP in 5' splice site selection. 3' splice site selection has been examined both in vivo and in vitro, and the results are different. In vivo experiments with the 3' splice site duplication show that if the two sites are separated by more than one hundred nucleotides, the proximal site is preferentially used, suggesting that polarity and/or proximity may be important features in this process. This pattern in lost in vitro where both sites are equally used, suggesting that something directing this selection is absent or defective in vitro. Studies with some yeast strains carrying mutations in the U1 snRNA gene have shown an increase in the use of the distal 3' site. Therefore, U1 snRNP is not only required for 5' splice site recognition and selection but is also clearly involved in 3' splice site selection.

I 121 INTRONS IN HMU BACTERIOPHAGE: INTRON STRUCTURE AND OPEN READING FRAME FUNCTION, Heidi Goodrich-Blair and David A. Shub, Department of Biological Sciences and Center for Molecular Genetics, State University of New York at Albany, Albany, NY 12222
We have described a self-splicing group I intron in the hydroxymethyl uracil (HMU) phage SPO1, that infects the gram positive *Bacillus subtilis* (1). The related phages SP82G, 2c and ϕe (which also contain HMU in their DNAs) also have an intron. The introns of SPO1 and SP82G have been sequenced on both strands directly from PCR-generated single stranded DNA. The two introns are very similar; they interrupt the DNA polymerase gene in identical positions, can be folded into secondary structures typical of group IA (with core secondary structures that are almost identical), and have open reading frames (ORFs) looped out of the structure in the same region (L8). The nucleotide sequences of the two ORFs are highly divergent, but a FASTA alignment of the expected translation products revealed 43.6% identity between them.
The intron ORFs of SPO1 and SP82G are preceded by a consensus Shine-Dalgarno sequence and may therefore be expressed in vivo. Preliminary experiments indicate that the ORFs encode DNA binding proteins, with specificity for the sequence of the ligated exons. This suggests that the ORFs may be involved in intron mobility.

1. Goodrich-Blair, H. et al., *Cell* **63** 417-424 (1990).

I 122 BINDING OF *NEUROSPORA* MT TYROSYL-tRNA SYNTHETASE TO A GROUP I INTRON RNA, Q. Guo, R. Akins, and A. Lambowitz, Departments of Molecular Genetics and Biochemistry, The Ohio State University, Columbus, OH 43210. The *Neurospora* mt large rRNA intron is a group IA intron of 2.3 kb, which requires the *cyt-18* gene product, mt tyrosyl-tRNA synthetase (mt tyrRS) for splicing *in vivo* and *in vitro*. We constructed a 388 nt mini-intron that lacks sequences extraneous to the splice sites and conserved group I intron secondary structure. This mini-intron is still spliced in a protein-dependent manner, but further modifications that affect different regions of the core structure abolish splicing, suggesting that the conserved RNA structure is still required for catalysis and cannot be substituted by the protein. UV-crosslinking experiments confirmed that the *cyt-18* protein binds directly to the intron. Experiments in which tyrRS or splicing activity of the *cyt-18* protein were competed by introns containing large deletions localized the protein binding site to a region of the intron core between P3 and P9. We present a model of the *cyt-18* protein binding site, a key feature of which is the similarity between the highly conserved P7 stem in the intron and the variable loop region of *Neurospora* mt $tRNA^{tyr}$. The possibility that the binding site includes the conserved P7 stem is consistent with the finding that the *cyt-18* protein is required for splicing a number of different introns *in vivo*. The similarity between the possible binding site and the variable loop region of the $tRNA^{tyr}$ accounts for the observations that *Neurospora* mt $tRNA^{tyr}$ is a much better inhibitor of splicing than is *E. coli* $tRNA^{tyr}$, which has a different variable loop, and that *E. coli* tyrRS does not splice the *Neurospora* mt large rRNA intron. Our findings suggest that the ability of the *Neurospora* tyrRS to function in splicing may be based in part on a fortuitous resemblance between a highly conserved region of group I introns and an idiosyncratic feature of *Neurospora* mt $tRNA^{tyr}$.

I 123 DE NOVO SYNTHESIS OF AN INTRON BY THE MAIZE TRANSPOSABLE ELEMENT, DS., L. Curtis Hannah, Lynwood Ingham, Michael Giroux, Donald McCarty, and John Baier, Department of Vegetable Crops, University of Florida, Gainesville, Florida, 32611
Since the discovery of intervening sequences or introns within genes, much effort has been directed towards identifying their origin(s) and their role(s) in the host organism. Recent insight into a possible origin of introns has come from the discovery that transposable elements can sometimes alter RNA processing. Newly-created transcripts are produced by use of a splice junction within the element itself or a site created upon insertion into the gene. Here, we report two novel splice reactions involving the well-characterized maize element, dissociation or Ds. In one transcript, *Ds* located within an exon of the shrunken-2 (Sh2) gene alters splicing of an adjacent, upstream intron such that splicing now involves the use of a 3' acceptor site normally silent in the wild type gene. This results in the inclusion in the mature RNA of sequences usually spliced from the processed transcript. Coupled to its ability to act from a distance to alter splicing, this *Ds* element with its associated duplication of host sequences also can behave like a conventional intron. A 5' donor site arising from one copy of the duplicated host sequence is sometimes spliced to the downstream 3' acceptor site. This removes the *Ds* element plus one copy of the host duplication and results in a transcript of wild type sequence at this junction. This latter situation then represents a direct demonstration of the creation of an intron within a modern eukaryotic gene.

I 124 TRYPANOSOMATID RNAB RNA HAS A UNIQUE STRUCTURE AND IS ASSOCIATED WITH THE NUCLEOLAR PROTEIN, FIBRILLARIN; Toinette Hartshorne and Nina Agabian; Intercampus Program in Molecular Parasitology, School of Pharmacy, UCSF, Laurel Heights, 94143.

Four major small nuclear U RNAs (U snRNAs) were previously identified in *T. brucei* by immunoprecipitation of U snRNPs with anti-TMG antibodies and by sequence analysis of their RNA components. Three of these are counterparts to the U2, U4 and U6 snRNAs involved in *cis*-splicing of pre-mRNAs, and are required in trypanosomatids for *trans*-splicing of the spliced leader (SL) RNA to all acceptor RNAs. The fourth trypanosomatid U snRNA, RNAB, does not share homology with other known U RNAs and is found uniquely in trypanosomatids by Southern and Northern analysis of nucleic acids isolated from various organisms. We determined the secondary structure of RNAB by comparative analysis of *T. brucei* sequences and those deduced from the distantly related trypansomes, *T. cruzi* and *L. collosoma*; the resultant structure also bears no resemblance to known U snRNAs. However, a single stranded region of RNAB RNA contains the sequence, UGAUGA, which is found in the nucleolar U3, U8 and U13 snRNAs believed to have roles in processing of pre-rRNA. A region containing this sequence in HeLa U3 RNA is known to bind the 34 kd fibrillarin protein reactive to anti-scleroderma patient autoantibodies. We have detected a 33 kd fibrillarin species in *T. brucei* whole cell and nuclear extracts by Western analysis using anti-fibrillarin monoclonal antibodies. Trypanosomatid fibrillarin and RNAB are immunoprecipitable with anti-scleroderma patient seras, indicating that RNAB is associated in a snRNP particle with fibrillarin. We are currently examining the subcellular location of RNAB and the putative role of RNAB in rRNA processing.

I 125 FUNCTIONAL ANALYSIS OF THE PROMOTER OF THE RAT L1 ELEMENT
B.E.Hayward, E.Pascale, and A.V.Furano, Laboratory of Biochemical Pharmacology, NIDDK, NIH, Bethesda, MD, 20892.

The rat L1(LINE, long interspersed repeated DNA) element, close homologues of which are found in all mammalian species thus far analyzed, is approximately 6.5kbp long and comprises about 10% of the rat genome (100,000 copies). L1 elements contain two highly conserved open reading frames, the larger of which is believed to encode a reverse transcriptase. The amplification and dispersion of these elements through the genome is believed to involve an RNA intermediate. The termini of the element, although sharing structural features differ markedly at the sequence level between species. The 3' terminus contains a G-rich homopurine stretch capable of assuming non-B DNA structures .

The 5' terminus of the element is a 707 bp tract of CpG-rich DNA. When placed upstream of a reporter gene (CAT) this sequence functions as a strong promoter in non-rat cell lines (COS [monkey] and 293 [human]) but as a weak one in a rat cell line (P113). Deletion analysis of the promoter sequence shows that it contains both stimulatory and inhibitory regions.

Using gel shift competition assays and exonuclease III protection assays we show here that several proteins bind to specific sequences in the promoter. Work is underway to delimit the binding sequence of each of these proteins and to assay its contribution to promoter function.

I 126 STRUCTURAL AND EVOLUTIONARY CHARACTERIZATION OF FOUR NUCLEAR GROUP-I INTRONS, Steinar Johansen, Terje Johansen and Finn Haugli, Department of Cell Biology, Institute of Medical Biology, University of Tromsø, PO Box 977, N-9001 Tromsø, Norway.

The ribosomal DNA in the myxomycetes Didymium iridis and Physarum polycephalum are linear extrachromosomal molecules of about 20 kb and 60 kb, respectively. The LSUrRNA genes in both organisms contains two immobile Group-Ib introns. We have determined the sequence of these four nuclear encoded Group-I introns and proposed complete secondary structure models. Pairwise sequence comparisons of the catalytic core region show that the introns from these two species are evolutionary related, and probably have been present in a common ancestor. Furthermore, we have identified several differences in the intron structure between an american (M3B) and a japanese (PPO-1) strain of Physarum, indicating short-time sequence evolution of Group-I introns. To further investigate the structural and functional relationships, work is in progress to analyse in vitro self-splicing properties of these introns.

I 127 REVERSE TRANSCRIPTION OF *NEUROSPORA* MITOCHONDRIAL PLASMIDS, J. Kennell, H. Wang, M. Kuiper, J. Sabourin, and A.M. Lambowitz. Departments of Molecular Genetics and Biochemistry, The Ohio State University, Columbus, OH 43210. The Mauriceville and Varkud mitochondrial plasmids are closely related, closed-circular DNAs of 3.6 and 3.7 kb, respectively, that have characteristics of mtDNA introns and retroid elements. The plasmids use a novel mechanism of reverse transcription in which (-) strand DNA initiates directly at the 3' end of the plasmid transcript via recognition of a 3' tRNA-like structure. The plasmid reverse transcriptase (RT) was partially purified by chromatographic procedures and shown to copurify with the protein encoded by the 710 amino acid ORF of the plasmid. The column-purified RT utilizes *in vitro* transcript substrates; it shows specificity for substrates having the 3' end of the plasmid transcript, and synthesizes full length cDNA copies of these transcripts beginning directly at their 3' ends. The column-purified RT has no detectable RNase H activity, and the initial product of (-) strand synthesis is a full length RNA/DNA hybrid. The (-) strand DNA synthesized using purified RT begins with an extra DNA of approx. 20 nt ("20 mer") that is presumably derived from a primer. An activity in RNP preparations specifically cleaves the "20 mer" primer, apparently generating the correct 5' end of (-) strand DNA. PCR cloning shows that the "20 mers" are complementary to various regions of the plasmid RNA. Using RNP preparations purified free of "20 mer" cleavage activity, we show that (-) strand DNA synthesized from endogenous plasmid RNA also begins with extra DNA of 1 to 75 nts. PCR cloning shows that most of these putative primers are complementary to the 5' end or internal regions of the plasmid RNA, and a few are complementary to mt RNAs. Together, these findings suggest that priming of (-) strand synthesis occurs primarily by the RT switching template strands. The synthesis of full length (-) strand DNAs appears to depend on the ability of the RT to recognize the 3' end of the plasmid transcript and a cleavage activity that specifically removes the primer to generate the correct 5' end of (-) strand DNA.

I 128 INTRONS SHARED BY BACTERIA AND CHLOROPLASTS, Maria G. Kuhsel and Jeffrey D. Palmer, Department of Biology, Indiana University, Bloomington, IN 47405

The presence of introns within chloroplast genes is interpreted controversially. They are mostly regarded as eucaryotic features, since to date no introns have been found in eubacteria. However, the discovery of a group I intron within the chloroplast tRNA-Leu (UAA) gene of *Cyanophora paradoxa* at the same position as in land plant chloroplasts suggested the possibility of the presence of this intron within the progenitors of chloroplasts, the cyanobacteria (1). We report the presence of group I introns within the tRNA-Leu (UAA) genes and of group II introns within the tRNA-Val (UAC) genes from chloroplasts of all major groups of algae and from diverse eubacterial lineages including cyanobacteria (2), purple bacteria and ancient thermophilic bacteria. These are the first reports of introns within eubacterial genes. In all cases, the bacterial introns are present in identical gene positions as in chloroplasts and are therefore most likely homologous. The presence of both the tRNA-Leu intron and the tRNA-Val intron within two species of *Thermotoga*, possibly the most ancient eubacterium so far described, implies that these introns are at least as ancient as the eubacteria and have been stably maintained during eubacterial and chloroplast evolution. We are currently looking for the presence of tRNA-Leu and tRNA-Val introns within archaebacteria and protists. If these introns turn out to be homologous within eubacteria, archaebacteria and protists, this would be strong evidence for their presence from the beginning of gene evolution.

1. J.-L. Evrard, M. Kuntz, N. A. Strauss, J.-H. Weil, Gene 71, 115 (1988).
2. M. G. Kuhsel, J. D. Palmer, Science, in press

I 129 CONSERVATION OF 7SL RNA FOLDING PATTERN DURING EVOLUTION OF MOUSE B1 REPEATS. Damian Labuda, Daniel Sinnett, Chantal Richer and George Striker[1], Service de Génétique Médicale, Hôpital Ste-Justine, Département de Pédiatrie, Université de Montréal, Montréal, Québec, Canada H3T 1C5; [1]Max-Planck-Institut für biophysikalische Chemie, D-3400 Göttingen, Germany

The numerous B1 elements found in mouse DNA were dispersed by reverse transcription of their RNA transcripts and subsequent reinsertion in various genomic locations. B1 repeats are derived from 7SL RNA by internal deletions and a partial duplication, retaining about 82% of shared sequence positions. They were classified into six families by Quentin (1989), each with a typical consensus sequence. We observed that nucleotide substitutions accumulate in the individual family members at a three to four times faster rate than consensus sequences themselves. The substitutions in individual family members are distributed randomly among the sequence positions, although at CpG-dinucleotides they occur with the ten-fold greater rate. Predicted minimal energy folding of B1 RNAs, derived from consensus sequences of the families, reveals remarkable conservation of secondary structure domains of the Alu-fragment of 7SL-RNA. We confirmed this structure by enzymatic approach using B1 RNA transcribed *in vitro* from recently retroposed B1 sequence. These data are consistent with a model of B1 retroposition in which conserved master sequences, corresponding to family consensus, give rise to multiple copies of B1 pseudogenes. Slower mutation rate in masters is related to the conservation of secondary folding of their RNA. This may either indicate a function of B1 RNAs or be the result of a specific adaptation of B1 elements for retroposition responsible for their dispersal within the genome. Comparison of rodent B1 with primate Alu repeats reveal important similarities among 7SL-like retroposons that extend beyond sequence level and may well also apply to short retroposons in other mammalian systems. Supported by The Cancer Research Society, Inc.

I 130 YEAST mtDNA GROUPII INTRONS AS ACTIVE DONORS DISPLAY DIFFERENTIAL EFFECTS IN TRANSPOSITION, J. Lazowska, B. Meunier, G-L.Tian and C.Macadre
Centre de Génétique Moléculaire, CNRS,F-91198 Gif-sur-Yvette,France
Unidirectional gene conversion in yeast mitochondria is initiated by endonucleolytic cleavage of the intron-less mtDNA by specific proteins encoded by the groupI intron. We have shown that the two introns ai1 and ai2(groupII) of the S. cerevisiae COXI gene, are very active as donors in gene conversion.Two strains with only one intron(ai1 or ai2) were crossed to intronless strains and revealed that these introns transpose by themselves and do not require any other intron. Experiments with donor strains carrying splicing defects demonstrate that mutations in ai1, whether trans- or cis-acting, abolish the transposition. In contrast, in the groupI intron ai4, a cis-acting mutation is known not to abolish gene conversion. Interestingly, the ai2 intron is very efficiently transposed to the mtDNA of S. douglasii while the ai1 is not transposed at all. Differential effects are also observed in S. cerevisiae when introns ai1 and ai2 are the sole introns present and are confronted in a trans-configuration: transposition of ai2 is almost complete and that of ai1 is greatly reduced, while both introns are efficiently transposed in a cis-configuration or in the absence of one another. In addition,we did not observe a DNA-endonuclease activity,that cleaves recipient DNA at the site of intron insertion in the ai1 strain.Thus, the transposition of groupI and groupII introns may occur via different mechanisms. In the latter case, we favor the hypothesis of a RNA intron containing intermediate which would be reverse-spliced into recipient RNA and then reverse-transcribed into intron containing DNA.

I 131 COMPARISON OF GEL MOBILITY FOR TELOMERIC DNA, HYBRID DNA-RNA AND RNA OLIGONUCLEOTIDES. LEE Bok Luel, LEE kwang Moon, LEE Suk Kyung, LEE So Young and SUH Suk Soo, College of Pharmacy, Kyung-Sung Univ., Pusan National Univ., 110 Dae yeon Dong Namgu, Pusan, Korea 608-736

To investigate the effects of deoxyguanosine and thymidine residue on telomeric structure, we have synthesized the following DNA, hybrid DNA-RNA and RNA oligonucleotides.

DNA oligonucleotides : $d(T_2G_4)_4$, $d(T_2AG_3)_4$

DNA-RNA oligonucleotides : $h(T_2rG_4)_4$, $h(U_2dG_4)_4$

RNA oligonucleotides : $r(U_2G_4)_4$

We have analyzed for conformational changes as a function of temperature by non-denaturing polyacrylamide gel electrophoresis.

We show that two telomeric DNA oligonucleotides have formed unusual intramolecular hairpin structures under physiological conditions. $d(T_2G_4)_4$has a more stable structure compared to $d(T_2AG_3)_4$by UV absorbance thermal analyses. Hybrid $(U_2dG_4)_4$had the same mobility as $d(T_2G_4)_4$.

But hybrid (T_2rG_4) migrated to $r(U_2G_4)$. $r(U_2G_4)$ showed more slow gel mobilty compared to $d(T_2G_4)_4$ in non-denaturing gel.

I 132 SEQUENCE VARIATION OF INTRONS REVEALS FINE-GRAINED SUBDIVISION IN NATURAL POPULATIONS OF POCKET GOPHERS, Enrique P. Lessa, Museum of Vertebrate Zoology, University of California, Berkeley, CA 94720

An approach for surveying allelic variation of nuclear DNA sequences in natural populations is presented. Conserved sequences of exons are used to design oligonucleotide primers on both sides of intros, whose sequence variation is used for population genetic studies. Double-stranded copies of the segments of interest are amplified by the polymerase chain reaction (PCR), and allelic variants are identified on parallel denaturing gradient gels. The alleles are subjected to asymmetric amplifications via PCR and sequenced directly. This approach provides efficient ways of estimating the two major types of information needed for population genetics using DNA sequences, namely the frequencies and distributions of alleles, and their phylogenetic relationships. The application of this approach to the first intron of the adult beta-globin of pocket gophers (Thomomys bottae) revealed significant genetic variation and population subdivision.

I 133 PROTEIN-CATALYZED SPLICING OF A GROUP I INTRON: THE EFFECT OF CBP2 ON COB INTRON 5, Alfred S. Lewin, Department of Immunology and Medical Microbiology, University of Florida College of Medicine, Gainesville, FL 32610

The fifth intron of the *COB* gene (bI5) of yeast mitochondria belongs to group IA and splices autocatalytically. In vivo, splicing of this intron requires the product of the nuclear gene *CBP2*, and this protein also enhances the splicing of this intron in a cell-free reaction. I cloned the *CBP2* gene in E. coli in the context of a promoter from bacteriophage T7. When this construct was present in the same cell as a plasmid containing lacZ disrupted with bI5, expression of b-galactosidase was dependent on expression of the complete *CBP2* reading frame. The cbp2 protein purified from bacterial cells catalyzed the splicing of bI5 in vitro, under conditions in which self-splicing was minimal. In an attempt to define the site of interaction between the intron and the protein, the sites of cross-links between these molecules are being mapped.

I 134 DIFFERENTIAL INTRON LOSS AND ENDOSYMBIOTIC TRANSFER OF CHLOROPLAST GAPDH GENES TO THE NUCLEUS.

Marie-Françoise Liaud, De Xing Zhang and Rüdiger Cerff, Institut für Genetik, Biozentrum, Technische Universität Braunschweig, D-3300 Braunschweig, FRG.

Chloroplast glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is composed of two different subunits, GAPA and GAPB, which are encoded in the nucleus by two related genes of eubacterial origin (Brinkmann et al. 1989, Plant Mol. Biol. 13, 81-94). In the present work the genes encoding chloroplast GAPA and GAPB from pea have been cloned and sequenced (Proc. Natl. Acad. Sci. USA, in the press). The gene for GAPB is split by 8 introns. Two introns interrupt the region encoding the transit peptide and 6 are found within the region encoding the mature subunit, four of which are in identical or similar positions relative to genes for cytosolic GAPDH of eukaryotic organisms. As opposed to this, the gene encoding pea GAPA has only two introns in the region encoding the mature subunit. These findings strongly support the "intron early" hypothesis and suggest that the GAPA gene lost most of its introns relatively recently during the streamlining period of the chloroplast genome and as a result of its "retarded" transfer into the nucleus.

I 135 CLONING AND CHARACTERIZATION OF AN AUTOCATALYTIC GROUP I INTRON WITHIN THE PNEUMOCYSTIS CARINII 16S-LIKE RIBOSOMAL RNA GENE, P.A.Liberator and J.W.Anderson, Dept. of Biochemical Parasitology, Merck Sharp & Dohme Research Laboratiories, Rahway, N.J. 07065

Steroid-induced immunosuppression of rodents induces the expansion of a latent Pneumocystis carinii (P.c.) infection in the lung. The 16S-like ribosomal RNA (rRNA) genes from P.c. prepared from both immunosuppressed rats and mice have been cloned using the polymerase chain reaction (PCR). Limited nucleotide sequence of the coding region indicates that the two genes are not identical to one another. Both genes feature what appear to be intervening sequences (IVS) that are nearly 400nt in length. The two IVS' are similarly positioned very close to the 3'-end of the gene and appear to be present in each member of the multi-gene family. By Northern blot analysis, the rat-derived P.c. IVS probe is not found associated with mature rRNA. This same probe does not hybridize with mouse-derived P.c. genomic DNA. Similarly, the mouse P.c. IVS probe does not hybridize with rat P.c. genomic DNA. This indicates that the P.carinii prepared from rat and mice are genetically distinct organisms and that the IVS' may be used to distinguish between them. Detailed sequence analysis of the two IVS' supports the lack of cross hybridization observation. Both IVS' do however possess the four Group I IVS consensus sequences which are required for accurate and efficient autocatalytic splicing in vitro. The rat-derived P.c. IVS along with minimal flanking sequence has been cloned into a transcription vector and synthetic RNA has been produced. Apparent excision of the IVS from the in vitro generated primary transcript occurs, i) in an orientation-dependent fashion, ii) in a reaction mixture composed only of Mg^{++}, NaCl and guanosine without any exogenous protein, and iii) by a mechanism which generates a radioactive product when 32P-GTP is included as a reaction substrate. Each of these is a hallmark feature of Group I IVS autocatalytic excision.

I 136 THE COMPOSITE NATURE OF A GENETIC ELEMENT CONTAINING THE GENE FOR REVERSE TRANSCRIPTASE IN E. COLI. Dongbin Lim and Werner K. Maas, New York University Medical Center, 550 First Ave., New York, NY 10016.
Recently, it has been shown that certain strains of E. coli contain a genetic element, called retron encoding a reverse transcriptase (RT). In order to understand the nature of retrons in E. coli, we compared the chromosomal regions bordering the E. coli B RT with the corresponding regions of other RT-producing E. coli strains. We show that: (1) The E. coli B RT is present in DNA fragment (about 30 kb long) which is inserted into min 19 of the E. coli K12 chromosome; (2) Four other strains have a similar inserted piece; (3) The inserted piece contains a small unique region within a larger homologous region. The unique region (about 2 kb) is involved in the synthesis of msDNAs. We define the unique region as the retron, and we present data suggesting that retrons may have been inserted into the homologous region by a novel transposition mechanism. We also discuss the possible function of retrons in E. coli.

I 137 CHLOROPLAST INTRONS: MOLECULAR EVOLUTION AND PHYLOGENETIC STUDIES John M. Logsdon, Jr., Kenneth H. Wolfe, Stephen R. Downie, and Jeffrey D. Palmer, Department of Biology, Indiana University, Bloomington, IN 47405

Chloroplast (cp) genomes from different plant and algal species contain variable numbers of introns. In studies focused primarily on land plants, we have surveyed approximately 300 species for the presence or absence of introns in cp genes. We have shown the loss of two highly conserved cp introns from the genes *rpl2* and *rpl16*. The *rpl2* intron has been found to be lost in at least five land plant lineages. Sequencing shows that the intron has been precisely removed, probably via homologous recombination of a reverse-transcribed RNA intermediate. The *rpl16* intron loss, which has occurred in at least two plant lineages, as well as intron losses from the rice cp *rpoC1* and *clpP* genes (1) also appear to have occurred by the same mechanism. We are currently investigating possible intron losses from from the cp *trnI*, *rps16*, and *clpP* genes. Intron losses have been found to be useful phylogenetic markers. In addition, we are studying sequence evolution in chloroplast introns. The *rpl2* intron, for example, shows an extraordinarily high level of sequence and length conservation. Whether this is due to functional constraint or simply to this gene's location in the conserved inverted repeat of the cp genome is being evaluated. A molecular phylogeny of known cp intron sequences will be presented in order to reconstruct the history of intron duplication and spread within the chloroplast genome.

(1) Hiratasuka *et al.* (1989) Mol. Gen. Genet. **217**:185-194

I 138 AN ESSENTIAL REVERSE TRANSCRIPTASE-LIKE PROTEIN IS REQUIRED FOR TELOMERE REPLICATION IN YEAST, Victoria Lundblad and Elizabeth H. Blackburn, Department of Microbiology and Immunology, University of California, San Francisco, CA 94143.
Telomerase is a ribonucleoprotein enzyme with essential RNA and protein components. The RNA moieties of the Tetrahymena and Euplotes telomerases each contain a telomere-complementary sequence which serves as the template for addition of telomeric repeats, establishing telomerase as a specialized reverse transcriptase that uses an internal RNA template. The protein component(s) of telomerase, however, have not yet been identified. The *EST1* gene has previously been shown to be required for telomere length maintenance and long term viability in *S. cerevisiae* (1). Yeast cells lacking a functional *EST1* gene show progressively shorter telomeres, increased frequencies of chromosome loss and a senescence phenotype. Inspection of the sequence of the predicted EST1 protein has revealed significant similarity with DNA and RNA polymerases that use an exogenous RNA template, including reverse transcriptases (2). We propose that *EST1* encodes an RNA-dependent polymerase in yeast, and that, on the basis of the phenotype of *est1* mutants, EST1 is a component of an essential yeast telomerase. This sequence similarity is supported by the behavior of site-directed mutations in the conserved regions of the EST1 protein, in that mutations in highly conserved residues confer a mutant phenotype. By analogy with the ciliate enzymes, this data also suggests that the EST1 protein is part of a complex with a yeast telomerase RNA. We are currently testing whether immunoprecipitation of the EST1 protein co-precipitates this putative telomerase RNA, as well as looking for ciliate homologues of the *EST1* gene.

(1) Lundblad, V. and Szostak, J.W. (1989). Cell *57*, 633-643.
(2) Lundblad, V. and Blackburn, E.H. (1990). Cell *60*, 529-530.

Poster Session II

I 200 INTRONS, TRANSCRIBED SPACERS AND 28S rRNA EXPANSION SEGMENTS : SIMILARITIES AND DIFFERENCES, B. Edward H. Maden, Paul Ajuh and Paul Heeney, Department of Biochemistry, University of Liverpool, P.O. Box 147, Liverpool L69 3BX, England.

Parallels have been drawn between the transcribed spacers of eukaryotic ribosomal transcription units and: (i) introns in protein-coding genes, (ii) expansion segments in eukaryotic 28S rRNA. Here we explore these relationships using examples from genes in vertebrates. (i) Transcribed spacers resemble introns in three major respects: they are transcribed into RNA primary transcripts; they are removed in generation of the mature RNA products; their nucleotide sequences are highly labile in evolution. They differ from introns in two major respects: their removal in RNA processing is not accompanied by splicing; in vertebrates they are highly GC rich with abundance of the dinucleotide CpG, whereas introns show the general vertebrate characteristic of CpG deficiency. (ii) Transcribed spacers resemble rRNA expansion segments in the following respects: both types of sequence are GC rich in vertebrates and AT rich in many invertebrates (i.e. they co-evolve with regard to base-composition); also, insertions and deletions occur during their evolution. They differ from rRNA expansion segments in that they are removed during rRNA processing and they evolve at least tenfold more rapidly than expansion segments. These respective evolutionary rates will be documented with new data from Xenopus.

I 201 CLEAVAGE PATTERN AND MINIMAL RECOGNITION SITE OF THE HOMING ENDONUCLEASE ENCODED BY THE FIFTH INTRON IN THE CHLOROPLAST LARGE SUBUNIT rRNA GENE OF Chlamydomonas eugametos, Philip Marshall and Claude Lemieux, Département de biochimie, Faculté des sciences et de génie, Université Laval, Québec, QC, Canada, G1K 7P4.

A distinctive property of the fifth intron in the chloroplast large subunit rRNA gene of Chlamydomonas eugametos (CeLSU•5) is that it undergoes nearly quantitative, duplicative transposition during interspecific crosses between C. eugametos and C. moewusii. This phenomenon of intron homing is most certainly mediated by the double-stranded DNA endonuclease (I-CeuI) encoded by CeLSU•5. The activity of I-CeuI against the C. moewusii homing site has been recently demonstrated by expressing the CeLSU•5 ORF in E. coli, using the vectors pKK233-2 and pTRC-99a (1). In this study, the I-CeuI activity proved to be highly toxic to E. coli and as a consequence, was detected at very low levels. To eliminate this problem and characterize the cleavage pattern and recognition site of the I-CeuI endonuclease, we have expressed the CeLSU•5 ORF in E. coli under the control of a bacteriophage T7 promoter in a tightly regulated M13 system and developed an in vitro system to assay partially purified I-CeuI activity. This allowed us to determine that I-CeuI has a recognition sequence of less than 26 bp centered around the homing site and makes a staggered cut 1 bp downstream of this site, yielding 4-base (CTAA), 3' hydroxyl overhangs. (Supported by NSERC Canada).

1. Gauthier, A., Turmel, M. and C. Lemieux (1990) Curr. Genet., in press.

I 202 **EVOLUTIONARY CONSERVATION OF NEURAL BC1 RNA IN RODENTS,** **John A. Martignetti, Henri Tiedge and Jürgen Brosius, FRCN, Mount Sinai School of Medicine, New York, N.Y. 10029**

BC1 RNA is a developmentally regulated RNA polymerase III transcript which is expressed predominantly in the nervous system. In the subset of neurons where it is expressed it is found in the somatic and dendritic compartments. BC1 is the product of a single-copy gene. This gene was probably generated via tRNA retroposition. Eventually, BC1 RNA itself became the founder of the transcriptionally incompetent ID repetitive elements (SINEs).

We have identified and characterized the BC1 genes of rat, mouse, hamster and guinea pig. The coding regions and certain putative regulatory elements are shown to be conserved in contrast to the flanking regions which are subject to neutral drift. This indicates that evolutionary pressure exists to maintain this rodent gene. Together with the RNA's distinct cellular and subcellular localization, evolutionary conservation supports the notion that BC1 RNA has been recruited for a function in the nervous system.

I 203 CHARACTERIZATION OF cDNAs FROM L1 RIBONUCLEOPROTEIN PARTICLES, Sandra L. Martin, Dan Branciforte, Rosa Maria Mendoza, Hilary K. Srere, and Audrey L. Wright, Department of Cellular and Structural Biology, Program in Molecular Biology, University of Colorado Health Sciences Center, Denver, CO 80262

Mammalian genomes contain thousands of copies of LINE1, or L1, DNA. Most copies of L1 are truncated pseudogenes, but a few are thought to be full-length, functional transposable elements. Long elements in the mouse are characterized by the presence of two open reading frames which encode basic proteins that we think are necessary, and perhaps sufficient, for transposition via an RNA intermediate. Our working model for L1 transposition begins with specific transcription leading to a full-length, sense strand, poly-adenylated RNA. This RNA is then transferred to the cytoplasm and translated into protein(s). L1 proteins then associate with L1 RNA to form an RNP complex that carries the RNA back into the nucleus, converts it to cDNA and integrates the new copy of L1 into the genome. We have recently described a complex of RNA and protein found in mouse embryonal carcinoma cells (F9) which may be an intermediate in the transposition of L1. RNA from these RNA-protein complexes has been converted into cDNA and cloned into lambda-ZAPII. The properties of these cDNA clones and their relationship to genomic copies of L1 will be described.

I 204 MOUSE U14 snRNA IS ENCODED IN THREE DIFFERENT INTRONS OF THE MOUSE COGNATE hsc70 HEAT SHOCK GENE, E. Stuart Maxwell and Joyce Liu, Department of Biochemistry, North Carolina State University, Raleigh. NC 27695.

U14 snRNA is an evolutionarily conserved nucleolar snRNA that plays a role in the processing of pre-ribosomal RNA. A single genomic fragment of mouse DNA containing the U14 snRNA genes has been isolated from a mouse genomic library and sequenced. Analysis of the cloned DNA has revealed that three highly-homologous U14 snRNA coding regions are positioned on the coding strand of the mouse cognate hsc70 heat shock gene within introns 5, 6, and 8. Comparative analysis with the previously reported rat and human hsc70 gene sequences revealed a similar positioning of three U14 snRNA sequences within introns 5, 6, and 8 of the respective rat and human hsc70 genes. The U14 sequences contained in all three introns of all three organisms are highly homologous to each other and well conserved with respect to the diverging intron sequences flanking each U14 coding region. Comparison of the isolated mouse U14 snRNA transcript with the determined mouse U14 DNA coding sequences contained in the three mouse hsc70 introns indicated that the coding region of intron 5 is utilized for U14 snRNA synthesis in normally growing mouse ascites cells. Analysis of the mouse, rat, and human U14 snRNA coding sequences with their upstream and downstream flanking regions did not reveal the presence of any obvious and conserved RNA polymerase I, II, or III promoter sites. This indicates that a new RNA polymerase promoter site may be responsible for U14 snRNA transcription. Alternatively, as suggested by secondary structural analysis of the U14 snRNA coding region and flanking intron sequences, U14 snRNA may be generated by intron processing of the hsc70 pre-mRNA transcript.

I 205 EVOLUTION OF RETROPOSONS BY ACQUISITION OR DELETION OF RETROVIRUS-LIKE GENES, Marcella A. McClure, Department of Ecology & Evolutionary Biology, University of California, Irvine, CA 92717, USA

The retroid family consists of all genetic elements that encode a potential reverse transcriptase (RT). Members of this family include a diversity of eukaryotic genetic elements (viruses, transposable elements, organelle introns and plasmids) and the retrons of prokaryotes. Based on RT sequence similarity, the retroposon group is defined as the eukaryotic long interspersed nuclear elements (LINEs), the transposable elements of 1) Drosophila melanogaster (I & F factor), 2) Trypanosoma brucei (ingi factor), 3) Zea mays (Cin4), 4) Bombyx mori (R2Bm) and members of the group II introns and plasmids of yeast mitochondria. The data presented here elucidate the extent of the relationship between these representative retroposons and other retroid family members. Protein-sequence alignment data demonstrate that subsets of the retroposons contain different assortments of retroviral-like genes. Sequence similarity can be detected between the retroviral capsid protein sequences and several retroposon sequences. The relationship among the retroposon capsid-like sequences is congruent with the RT sequence phylogeny. In contrast the similarity between ribonuclease H (RH) sequences varies in different sub-branchs of the retroposon lineage. These data suggest that homologous recombination and/or independent gene assortment played a role in the evolution of the retroposons. Models are proposed for the descent of two hypothetical ancestors to the present day retroposons by either acquisition or deletion of retrovirus-like genes.

I 206 RETROVIRAL-LIKE ELEMENTS AS MEDIATORS OF REGULATORY EVOLUTION: AN EVOLUTIONARY MODEL, McDonald J.F., Csink, A.K. and Cuticchia, A.J., Department of Genetics, University of Georgia, Athens,GA. 30602

The insertion of retroviral-like elements (RLEs) within non-coding regions of chromosomal genes is frequently associated with the acquisition of novel patterns of gene expression. Although there is a growing body of evidence suggesting RLE insertions have been important in the evolution of new regulatory networks, the fact that RLE insertion variants are typically less fit than their wild type progenitors seems to be inconsistent with such a significant evolutionary role. We propose a model, supported by computer simulations, where naturally occurring suppressor genes shield RLE insertion variants from natural selection. Such a mechanism would permit the maintenance of hidden pools of RLE-mediated regulatory variation within natural populations for extended periods of time, thereby increasing the probability that these novel regulatory variants will become fixed within the species by random or selective processes. Evidence that natural populations of Drosophila carry suppressor genes is also presented.

I 207 ANALYSIS OF THE PROMOTER OF THE I FACTOR - THE RETRO-TRANSPOSON CONTROLLING I-R HYBRID DYSGENESIS IN *DROSOPHILA MELANOGASTER*, Carol Mclean and David Finnegan. Institute of Cell and Molecular Biology, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JR, SCOTLAND.

LINE-like retrotransposons are found in diverse eukaryotic organism but the conditions under which transposition takes place are only known for the I factor. The I factor controls the I-R system of hybrid dysgenesis in *Drosophila melanogaster*. I-R hybrid dysgenesis occurs when male flies containing complete I factors (Inducer strains) are crossed with females lacking them (Reactive strains). In the germline of the resultant female progeny (SF females) the I factor transposes at a high frequency.

Transposition of LINE-like elements is thought to occur via reverse transcription of a full length RNA intermediate. We have undertaken to study those sequences involved in transcription initiation of the RNA intermediate. The promoter of the I factor has been assayed using the reporter gene chloramphenicol acetyl transferase (CAT), in tissue culture and in transformed flies. All promoter sequences lie within the transcription unit. Mutational analysis has revealed sequences which are essential, stimulate or suppress promoter activity. Some of these sequences are conserved in LINE-like elements of other species.

I 208 HUMAN TELOMERASE RECOGNITION OF A CHROMOSOME TRUNCATION SITE: EVIDENCE FOR TWO SITE PRIMER RECOGNITION, Gregg B. Morin and Joan A. Steitz, Department of Molecular Biophysics and Biochemistry, Yale University, 333 Cedar St., New Haven, CT 06510.

One mechanism for stabilization of truncated chromosomes is the addition of a new telomere to replace the deleted chromosome terminus. The sequence (see below) of a truncation/healing event that causes an inherited human α-thalassaemia has been described (Wilkie et al., 1990, *Nature*, *346*, 868). Chromosome healing could be accomplished by the action of the RNP, telomere terminal transferase (telomerase). We investigated whether the region upstream of the telomere addition site could function as a primer for human telomerase *in vitro*. The -12 to -25 region is very G rich, but the -1 to -11 region is not. Oligonucleotides spanning both domains or only the G rich domain function as primers. Oligonucleotides corresponding to the regions from -1 to -11 and -2 to -11 function as primers, but the region from -3 to -11 does not. Addition of 3 G's to the 5' end of the -3 to -11 primer restores activity. These data, and studies of other oligonucleotides, suggest a two site primer recognition model in which the primer 3' end base pairs to a template in the telomerase RNA and the 5' region interacts with a second binding site recognizing G richness. Therefore, we suggest that telomerase was responsible for the healing of the initial chromosome break in the case of this α-thalassaemia, allowing the stable inheritance of this defective chromosome. Furthermore, the sequence requirements for telomerase recognition are minimal. Telomerase mediated chromosome healing may be involved in the evolution of chromosome size and number.

```
truncated chromosome:  AGCAAGGAGG TGGGTGGGGC GAGCAGAGTT (AGGGTT)n
                                 |          |          |     ↓
                               -21        -11        -1   telomeric repeats
```

I 209 **GROUP II LARIAT IVS ACTS AS NUCLEOTIDYL-TRANSFERASE , PHOSPHO-TRANSFERASE AND AS AN POLY-N POLYMERASE IN VITRO**, Manfred W. Müller, Paul H. Stocker & Rudolf J. Schweyen,
Institut für Mikrobiologie und Genetik, Universität Wien, Althanstrasse 14, A-1090 Wien, Austria

Group II lariat IVS bl1 catalyzes the rearrangement of 3' exon sequences from heterologous 5'E-3'E substrate RNAs with multiple turnover by several cycles of reverse- and forward self-splicing reactions. We show that the IVS accepts substrate RNAs having substituted the 3' exon sequences by either G, A, U or C mononucleotides or by a single phosphate group. Any of these 3'E substitutes (3'E*) is ligated to the 3' hydroxyl of the lariat IVS by transesterification. This nucleotidylation and phosphorylation of the IVS is reversible by the 5'E sequences (5'E*). Thus, lariat IVS acts as a potential phospho- or nucleotidyl-transferase with multiple turnover. The transfer reactions can be generalyzed as follows:

IVS_{OH} + 5'E-3'E* <-> IVS-3'E* + $5'E_{OH}$.

In a side reaction the IVS can substract the 3' terminal nucleotide from the 5' exon. By a series of transesterificaction reactions, lariat IVS converts 5'E*pN substrate RNA (were N represents the 3' terminal nucleotide of the 5' exon, either G,A,U or C) into polyribonucleic acid. Co-existence of both versions of nucleotidyl transfer reactions enables lariat IVS to act as sequential and random RNA polymerase. The overall reaction can be formulated as follows:

IVS_{OH} + 5'E*pN_{OH} <--> IVS_{OH} + 5'E(-1) + ...$pN_4pN_3pN_2pN_{1OH}$.

I 210 SPLICE SIGNALS, INTRON STRUCTURE AND SPLICEOSOMAL snRNAs IN THE CILIATE TETRAHYMENA, Henrik Nielsen, Henrik Ørum and Jan Engberg, Dept. of Biochemistry B, Panum Inst., Univ. of Copenhagen, Denmark.

We have investigated the possibility, that pre-mRNA splicing in Tetrahymena is different from that of yeast and mammals by analysis of intron structure and snRNA components. Tetrahymena contains the full complement of spliceosomal snRNAs, that is U1, U2, U4, U5 and U6. These snRNAs are found in abundancies intermediary to those found in S. cer. and mammals as would be expected from the proportion of intron containing genes in these organisms. A specific intron, namely the single intron in the 1-3 copy ribosomal protein gene rpS25, was sequenced from different Tetrahymena species. The introns were very AT-rich and bounded by the canonical GT and AG dinucleotides. A striking conservation of long stretches of sequence was found at the 5´- end of the intron, whereas the 3´- end was highly divergent. The sequence 5´- ATGAA - 3´was found in similar positions in closely related introns and subsequently in most (including non-rpS25) introns at their 3´- end and is therefore a candidate for a branch site sequence. Altogether, Tetrahymena introns are more similar to mammalian introns than to S. cer. introns as judged from the low stringency of splice signals and the structures of individual snRNAs. Differences include the high AT-content, the sequence of the putative branch signal and the low copy number (2-4) of snRNA genes.

I 211 DISTRIBUTION OF INTRONS IN LAMPREY FIBRINOGEN GENES, Yang Pan and Rusell F. Doolittle, Center for Molecular Genetics, University of California, San Diego, La Jolla, CA 92093

Fibrinogen is a complex and highly differentiated protein found in all vertebrate species. It is encoded by three homologous genes: α, β and γ. The earliest diverging creatures with authentic fibrinogen are the cyclostomes (lamprey and hagfish). We have cloned and characterized these three genes from lamprey so that the cyclostome and mammalian gene structures can be compared. Our primary goals were to see if exons correlate with structural features, and to pinpoint when introns have been gained or lost. With the new information from the lmaprey, the history of the fibrinogen gene system can be divided into discrete time periods: the early preduplication period, the intermediate period before the cyclostomes diverged and, finally, the 450 million-year period since the last common ancestor of lampreys and mammals. Thus, one of our models begins with 16 different exons being assembled into a prototypic chain that could be bundled into trimers or hexamers. Of the fifteen starting introns, only three are now common to all three genes in both groups of organisms. In the simplest model, most of the intron loss occurred after the two duplications leading to the three subunits, but before the divergence of the lampreys and mammals. The notion of random loss is consistent with the early conclusions of Crabtree et al (J. Mol. Biol. 185, 1, 1985).

I 212 STUDIES ON THE PROTEINS ENCODED BY THE I-FACTOR, THE RETROPOSON RESPONSIBLE FOR I-R HYBRID DYSGENESIS IN *Drosophila melanogaster*.
Trevor Paterson and David J Finnegan, Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh, Scotland. EH9 3JR

The nucleotide sequence of the I-factor reveals a typical retroposon LINE-like structure, containing two long ORFs within the 5.4kb element. The first of these, ORF1, may perform similar structural and regulatory functions to the retroviral gag proteins and resembles the gag proteins by possessing a cysteine/histidine motif. The second, larger ORF potentially encodes a protein with clear homologies to the enzymes reverse transcriptase and RNAseH. The product of ORF1 has been expressed using the recombinant baculovirus system and a bacterial expression system. The 52kDa ORF1 protein product binds strongly to both single and double stranded DNA and RNA. The specificity of this binding is under investigation, as are the protein sequences involved in binding activity, in order to elucidate the role of ORF1 in transposition or its regulation. Expression of the larger, second open reading frame product is also being pursued.

I 213 INTERMOLECULAR RECOMBINATION OF CODING SEQUENCES AND ALTERNATIVE SPLICING IN THE TISSUE SPECIFIC EXPRESSION OF C-MYB PROTO-ONCOGENE, Bernard, V. Perbal*✚, Michel Vellard*, Johann Soret*, Evani Viégas-Péquiniot❖, Bernard Dutrillaux❖ Francis. Galibert✛ and Alain Sureau*, Institut Curie, Bat 110, Centre Universitaire, 91405 Orsay France*, Institut Curie, Paris❖, Hopital St Louis Paris✛, and Université P. et M. Curie Paris, France✚
In chicken, the c-myb mRNA species expressed in hematopoietic cells of the B and T lineages contain different 5' termini. We have characterized a new 5'-proximal c-myb exon (ET) whose sequences are specifically represented in c-myb mRNA species from thymic cells. We have established that ET and the bulk of 15 c-myb exons are localized on distinct chromosomes in chicken, suggesting that intermolecular recombination is involved in the tissue specific expression of c-myb. We also identified human sequences homologous to ET and found that ET and the other c-myb exons are split between different human chromosomes. We therefore speculated that an intermolecular recombination mecanism might also be involved in a tissue specific expression of the human c-myb proto-oncogene. We also report the existence of several alternate 5' ends for the human c-myb mRNA species expressed in normal hematopoietic tissues of different origins.

I 214 RETROTRANSPOSON INSERTIONS WITHIN AN EXON OF THE DROSOPHILA vermilion GENE ARE SPLICED FROM THE PRECURSOR RNA, Anne-Marie Pret, Robert A. Fridell, and Lillie L. Searles, Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599
Intergenic suppressor mutations reverse the phenotype caused by insertion of a 7.4 kilobase 412 retrotransposon into the 5' transcribed, but untranslated region of the Drosophila vermilion gene. These suppressor mutations partially restore the level of vermilion transcript accumulation. Analysis of transcripts reveals that most of the 412 sequences are eliminated by splicing at sites within the element. The insertion of a 3.6 kilobase truncated B104/roo retrotransposon into one end of the 412 element also reverses the vermilion mutant phenotype. This reversion mutant allele produces transcripts from which both transposable elements have been removed from the precursor RNA by splicing at new donor sites within the secondary insertion and at the same acceptor site within the 412 element. Both suppressor and reversion mutations may increase the level of vermilion mRNA by affecting one of the steps in the splicing pathway.

I 215 RECOGNITION ELEMENTS OF THE td INTRON ENCODED ENDONUCLEASE, Susan M. Quirk, Deborah Bell-Pedersen, Mary Bryk and Marlene Belfort, Wadsworth Center for Laboratories and Research, New York State Department of Health, Empire State Plaza, Albany, New York 12201.

The intron in the td gene of bacteriophage T4 encodes an endonuclease, I-TevI, which promotes the transfer of the intron to an intronless version of the cognate gene. I-TevI cleaves 23 nt 5' to the intron insertion site (IS). The intron is then acquired by a recombination event in which the intron-containing gene is postulated to provide the template for repair synthesis. Current studies are directed towards defining the recognition sequences for I-TevI and its mode of interaction with DNA target sequences. DNase I footprinting identifies a protected region including 9 to 11 nt 5' to the IS in exon I and 16 to 18 nt 3' to the IS in exon II. The distant cleavage site (CS) is not protected. The requirement for sequences spanning the IS for cleavage was shown by the failure of I-TevI to direct cleavage in constructs in which 21 nt surrounding the IS were deleted or in which 10 nt were inserted into the IS. Furthermore, duplexes containing 15 bp 5' and 16 bp 3' to the IS (HS31) and 12 bp to each side of the IS (HS24) were specifically bound by I-TevI in gel retardation assays, indicating that they contain I-TevI recognition sequences. HS31 cloned into pBSM13 in both orientations directed cleavage by I-TevI into vector sequences either 23 nt or 22 nt 5' to the IS. In other constructs in which sequences 5' to the CS were deleted, cleavage occurred at the normal distance from the IS (23 nt). In all of these constructs cleavage occurred at reduced efficiency. The data suggest that while the I-TevI IS/binding domain is both necessary and sufficient for cleavage, upstream sequences influence cleavage site selection as well as cleavage efficiency.

I 216 ORIGIN OF A NOVEL INTRON IN THE GENE FOR A TWO-DOMAIN GLOBIN, Austen F. Riggs, Yasushi Naito, Thomas L. Vandergon, and Claire K. Riggs, Department of Zoology, University of Texas, Austin, TX 78712-1064

Analysis of the gene for a clam globin with two heme-binding domains in one polypeptide shows that a novel intron separates the DNA encoding the domains. The close correspondence of parts of the intron with the 3' non-coding region of the cDNA for the globin strongly suggests that the intron evolved from the 3' non-coding region of an ancestral single-domain gene. Many introns may have a similar origin to the extent that exons evolved from independent genes as has been suggested. This hypothesis explains the finding that splice junctions map to protein surfaces. They do so because NH_2- and COOH- termini are on the surface.

Supported by NIH grant GM 35847, Welch Foundation Grant F-0213 and NSF grant DMB-8502857.

I 217 **A FUNCTIONAL DELETION OF THE P2 ELEMENT OF THE T4 *td* INTRON DERIVED BY GENETIC SELECTION,** Jill L. Salvo and Marlene Belfort, Wadsworth Center for Laboratories and Research, New York State Department of Health, Empire State Plaza, Albany, NY, 12201

Through a number of genetic, comparative and functional studies, the required core elements and secondary structure of the self-splicing group I introns have been defined. It is also clear however, that additional elements, although less well conserved, are necessary to maintain activity, and it is likely that the subgroups of group I introns utilize these elements in different ways to accomplish similar functions. Evaluation of the role of such accessory elements in splicing is facilitated by the three group IA introns of bacteriophage T4 which exhibit a high degree of both structural and sequence homology (Shub *et al., PNAS,* 85, 1151–1155, 1988). Direct comparison of these three introns emphasizes two differences – placement of the intron ORF, and the variable occurrence of P2. In studies concentrating on the *td* intron, we have previously demonstrated that the removal of extraneous ORF sequence results in a fully functional intron (Salvo *et al., JMB,* 211, 537–549, 1990). Results presented here show that the P2 element can also be successfully removed from the intron, but only under specific circumstances. Deletions which result in a wild type P1 with a shortened J1/3 are phenotypically Thy^- and incapable of ligating the exons. Instead, an upstream site is cleaved at low efficiency. Subjecting such a non-functional P2 deletion to genetic selection for the Thy^+ phenotype produced a further deletion of J1/3 and a truncated P1, with concomitant restoration of splicing function, including precise exon ligation. According to a current model, the role of P2 is to present the 5' splice site to the catalytic core for cleavage by stacking with P1 (Michel and Westhof, *JMB,* in press). It is likely that in the absence of P2, the shortened J1/3 restricts the flexibility of P1, aligning it with the core for appropriate 5' splice site cleavage. According to the Michel–Westhof model, correct orientation of the P1 helix under such conditions would require that it be shortened. These results indicate that successful deletion of P2 is subject to constraints based on its apparent role in facilitating the association of P1 with the catalytic core.

I 218 ALU RNA SECONDARY STRUCTURE CONSISTS OF TWO INDEPENDENT 7SL RNA-LIKE FOLDING UNITS. Daniel Sinnett, Chantal Richer, Jean-Marc Deragon and Damian Labuda. Service de Génétique médicale, Hôpital Sainte-Justine, Département de Pédiatrie, Université de Montréal, Montréal, Canada, H3T 1C5.

Alu repeats are ubiquitous sequence elements that occupy more than 5% of the human genome. Their amplification by retroposition, i.e. by genomic reintegration of reverse transcribed RNA copies, implies an important role for Alu RNA. We report enzymatic studies of the secondary structure of Alu RNAs transcribed *in vitro* from two recently retroposed Alu sequences. These experiments show that the dimeric organization of an Alu element is reflected in its RNA. Alu subunits fold independently conserving secondary structure motifs of their progenitor 7SL RNA molecule. Energy minimization analysis indicates that this folding pattern is also characteristic of different Alu and Alu-like sequences. This 7SL-like folding in Alu RNA may be important for specific interactions with proteins whereby the putative Alu ribonucleoprotein could serve as an active form of Alu RNA in retroposition. Although a possibility of cellular function of Alu RNA cannot be excluded, comparison of different mamalian systems suggests that its conserved secondary structure may represent an adaptation leading to efficient retroposition of Alu elements within the genome. Supported by The Cancer Research Society, Inc.

I 219 INTERCISTRONIC INTRONS IN *EUGLENA GRACILIS* CHLOROPLAST RIBOSOMAL PROTEIN OPERONS, Jennifer K. Stevenson[1], Robert G. Drager[1], Donald W. Copertino[1], Kristin P. Jenkins[1], David A. Christopher[1] and Richard B. Hallick[1,2], Department of Molecular and Cellular Biology[1], Department of Biochemistry[2], University of Arizona, Tucson, AZ 85721

An unusual ribosomal protein locus is found in the genome of *Euglena gracilis* chloroplast. It encodes the genes for ribosomal proteins S4 and S11 (*rps4* and *rps11*). The coding region of the *rps11* gene is interrupted by two introns of 107 bp and 100 bp. We have characterized the major transcript from this region by RNA blot analysis, primer extension sequencing, and cDNA cloning and sequencing. The transcript is detected only as a dicistronic message, the introns found in *rps11* have been removed, and an additional 95 nt are absent from the intercistronic region between *rps4* and *rps11*. From further examination of genomic sequence in this region, it is apparent that an intercistronic intron exists between the *rps4* and *rps11* genes. We examined other RNA transcripts from regions of the genome which could potentially contain intercistronic introns and found two more examples. These are located in a large ribosomal protein operon between the genes for the ribosomal proteins L23 and L2, and L14 and L5. The introns found in these ribosomal protein loci belong to a class of introns unique to *Euglena gracilis* chloroplast. They were first described by Christopher and Hallick (1989) and have been termed group III introns. There are at least 50 group III introns in the *Euglena gracilis* chloroplast genome and all but 5 of them are found in genes encoding protein components of the transcriptional and translational apparatus. The distribution of group III introns and the unusual location of intercistronic group III introns may reflect some aspect of gene expression, or provide some insight into the mechanism of their splicing.

[1]Chrisotpher D.A. and Hallick R.B. (1989) *Nucleic Acids Res.* 17:7591-7608

I 220 THE HUMAN LINE-1 (L1Hs) ELEMENT: IDENTIFICATION AND ANALYSIS OF IT'S INTERNAL PROMOTER, Gary D. Swergold, and Maxine F. Singer, Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, Md. 20892

L1Hs elements belong to the non-LTR (or poly A) class of transposable elements. L1Hs contains two long open reading frames; the 3' ORF predicts a polypeptide with regions similar to those found in RNA-dependent nucleic acid polymerases. Full length L1 RNA is observed in human teratocarcinoma cells. Several cDNAs, each with an individual sequence, have been cloned, suggesting that many genomic L1Hs units are transcribed in a cell-specific manner although they do not appear to share similar flanking sequences. These and other data suggest that L1Hs transposition proceeds via a full-length RNA intermediate which is synthesized from an internal promoter.

To identify the L1Hs promoter we constructed elements in which the bacterial *lacZ* gene replaced the L1Hs ORFs and was fused in frame, 14 codons after the presumed ORF1 initiator codon. The L1Hs 5'UTR promoted expression of ß-galactosidase in a variety of transiently transfected cell types. Full-length RNA was detected in the transfected cells; most of the transcripts initiated at or near the beginning of the L1Hs segment. Deletion analysis revealed that the sequences most critical for expression are located within the first 100 bases of L1Hs and that other sequences within the first 668 base pairs also contribute to RNA accumulation. Several regions within the first 100 base pairs are required for maximum expression. Expression from the L1Hs 5' UTR is high in human and mouse teratocarcinoma cells and low in other cell types.

I 221 A SELF-SPLICING GROUP I INTRON FROM CHLOROPLASTS OF CHLAMYDOMONAS REINHARDTII. Andrew J. Thompson and David L. Herrin, Botany Department, University of Texas at Austin, Austin, TX 78713.

Ten group I introns are known to occur among the chloroplast genomes of the genus *Chlamydomonas*. The 23S (large subunit) ribosomal RNA gene from the chloroplast genome of *Chlamydomonas reinhartii* contains a 888bp intron classified as a group Ic intron. This intron (CrLSU) self-splices *in vitro* and the conditions for this reaction have been optimized. The reaction mechanism is typical of other group I introns. Cyclization of the excised intron occurs at three major sites which have been mapped. The intron contains a 491bp open reading frame starting in L6 and extending to a stop codon in J6/7, in the catalytic core of the conserved group I secondary structure. A 368bp region of the open reading frame in L6 is not required for in vitro self-splicing and we are currently investigating, using a particle gun to achieve transformation of chloroplast DNA via homologous recombination, whether a functional ORF is required for *in vivo* splicing.

I 222 DISTRIBUTION OF GROUP I INTRONS IN THE CHLOROPLAST LARGE SUBUNIT RIBOSOMAL RNA GENE OF Chlamydomonas: POSSIBLE IMPLICATION OF THE REVERSE SELF-SPLICING REACTION IN THE GENERATION OF NOVEL INTRON INSERTION SITES, Monique Turmel, Marie-José Côté, Jean-Patrick Mercier, Christian Otis and Claude Lemieux, Département de biochimie, Faculté des sciences et de génie, Université Laval, Québec, QC, Canada, G1K 7P4.

Although it has not been directly demonstrated that group I introns can move from one genomic site to another, several circumstantial lines of evidence support this phenomenon. Two mechanisms have been proposed to explain this transposition. One of them occurs at the DNA level and involves intron-encoded endonucleases (1), while the other occurs at the RNA level and involves the insertion of group I introns into foreign RNAs by reversal of the self-splicing reaction, followed by reverse transcription of the recombined RNA and integration into the genomic DNA by homologous recombination (2). Our recent sequence analysis of the chloroplast large subunit (LSU) rRNA genes from 17 species of Chlamydomonas strongly suggests that the reverse self-splicing reaction has played a major role in the transposition of group I introns during the evolution of chloroplast genomes. A total of 39 group I introns representing 12 distinct insertion sites within domains II, IV and V of the LSU rRNA secondary structure were identified in the LSU rRNA genes of the algae examined. Their distribution is highly variable and shows no correlation with the phylogeny inferred from the coding LSU rRNA gene sequences. (Supported by NSERC Canada).

1. Dujon, B., (1989) Gene 82:91-113
2. Woodson, S.A. and Cech, T.R. (1989) Cell 57:335-345

I 223 **STREPTOMYCIN INHIBITS SPLICING OF GROUP I INTRONS**
Uwe von Ahsen and Renée Schroeder, Institut für Mikrobiologie und Genetik, University of Vienna, Austria

Streptomycin is an aminoglycoside antibiotic, that interacts with the 16S rRNA from *E. coli* (Moazed & Noller, Nature 327, 389, 1987). Here we report that streptomycin also interferes with self-splicing group I introns by competing with the substrate guanosine for the G-binding site (guanosine-binding site). Guanosine binds to the intron RNA via hydrogen bonds using the guanidino, the 6-keto and the ribose hydroxyl groups (Bass & Cech, Biochemistry 25, 4473, 1985). Two hydrogen bond interactions have been assigned between the guanidino group of the substrate guanosine and the G-264 residue in the conserved P7 stem in the *Tetrahymena* rRNA intron (Michel et al., Nature 342, 391, 1989). Streptomycin contains two guanidino groups that could potentially interfere with guanosine binding. We performed splicing reactions on the preRNA of the T4 phage *td* gene intron with increasing amounts of streptomycin and found that splicing is totally inhibited at a concentration of 10 mM streptomycin and 1 uM GTP (the K_M concentration). Increasing the GTP concentration restores splicing, suggesting a competitive inhibition mechanism. The inhibitory effect extends to other group I introns, but does not affect group II introns. In order to define the streptomycin-binding site, we are currently performing a mutational analysis on the td intron.

I 224 BIOCHEMICAL CHARACTERIZATION OF I-SceII, A YEAST MITOCHONDRIAL INTRON-ENCODED ENDONUCLEASE Catherine Wernette and Ronald A. Butow, Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75235

We have purified to near homogeneity, a site-specific ds DNA endonuclease, I-*Sce* II, encoded by intron aI4α of the yeast mitochondrial coxI gene, which is required for aI4α mobility. We can detect I-*Sce* II activity in mitochondrial or whole cell extracts in strains lacking a functional bI4 maturase, but not in wild-type strains. Typical yields of I-*Sce* II are 3-5% with specific activities of about 500,000 u/mg protein (one unit of activity cleaves 50 ng of DNA substrate/hr at 30 °C). In the last step of purification we use a synthetic DNA affinity column composed of multimers of an 18 basepair sequence containing the recognition sequence. I-*Sce* II is derived by proteolytic processing of a primary translation product, p56. Experiments to define the actual site of processing are in progress. We have previously determined that the recognition site of I-*Sce* II is no larger than 18 basepairs. We have analyzed a large set of mutations of our standard substrate and find that substitutions are tolerated at several positions. Other experiments indicate that certain combinations of multiple substitutions are tolerated. A kinetic analysis of the the activity of purified I-*Sce* II in the presence of natural and mutant substrates will be presented. CW was supported by a US Public Health Service Postdoctoral Fellowship.

I 225 ISOLATING RETROTRANSPOSONS BY PCR OF REVERSE TRANSCRIPTASE AND PHYLOGENETIC SCREENING, Holly A. Wichman and Ronald A. Van Den Bussche, Department of Biological Sciences, University of Idaho, Moscow ID 83843.
The phylogenetic screening procedure was designed to identify mammalian transposable elements, but it will also detect other rapidly evolving repetitive sequences. In an attempt to refine this procedure to be specific for retrotransposons, we took advantage of the conserved amino acid domains of reverse transcriptase. Degenerate primers with 5-prime cloning sites were designed to amplify an approximately 400 bp fragment of reverse transcriptase. The upstream primer is 14 bp long with an additional 8 bp overhang, and has 312-fold degeneracy. The downstream primer is 12 bp long with an 8 bp overhang, and has 8 fold degeneracy. These primers have been used for PCR reactions on genomic DNA from a number of taxa including human, flying lemur, deer mouse, horse, zebra, rhino, chicken and trout, and several elements with marked amino acid similarity to reverse transcriptase have been identified. Because these primers are unusually short and degenerate, they produce many nonspecific products, and a secondary screening procedure is desirable to aid in identification of target sequences. We have used the phylogenetic screening procedure on 87 clones from PCR of the deer mouse. 23 clones were found to be repetitive in the deer mouse genome, while only 16 of these were also repetitive in the house mouse. Of the remaining 7 clones, only one was of the predicted size. Sequence analysis of this clone revealed a high degree of similarity to reverse transcriptase. This clone is thought to represent a retrotransposon from the deer mouse, but is distinct from the previously identified *mys* element. This work was supported by NIH grant GM38727 and a grant from the Idaho State Board of Education to HAW and a Sloan Postdoctoral Fellowship to RVDB.

I 226 HETEROGENEITY IN THE STRUCTURE AND EXPRESSION OF MEMBERS OF THE RTVL-H FAMILY OF PRIMATE RETROTRANSPOSON-LIKE SEQUENCES, David A. Wilkinson, Nancy L. Goodchild and Dixie L. Mager, Terry Fox Laboratory, B. C. Cancer Agency and Dept. of Medical Genetics, University of British Columbia, Vancouver, British Columbia, V5Z 1L3
RTVL-H elements constitute a large family of retrotransposon-like sequences. They are present in as many as 1000 copies in the genomes of humans, apes and old world monkeys. The prototypic element defined in humans is 5.8 kb in size and contains ~450 bp LTRs. RTVL-H elements are expressed in a number of human cell lines as a 5.6 kb unit-length message and a 3.7 kb spliced derivative, as well as a few other RNAs apparently transcribed from elements containing deletions. Sequence analysis of numerous cDNA and genomic isolates has allowed us to designate three sub-types of RTVL-H elements based on sequence differences in the U3 region of the LTR. Typically, this region of the LTR contains regulatory sequences governing transcription from the LTR promoter. We present evidence that elements representing these U3 sub-types have different patterns of expression among various human cell lines, as well as differences in copy number among several primate species. In addition to the LTR variants, another sub-type of RTVL-H element has been recently defined by a stretch of sequence within the pol-region, which is deleted in the majority of RTVL-H elements, that preserves the homology and colinearity of this region to that of the murine leukemia virus pol gene. The expression of elements of this sub-type is either undetectable or a very minor component of the total RTVL-H RNA expressed in various cell lines. The identification of RTVL-H sub-types that have preferentially expanded in one species, or that more closely conform to the predicted genomic structure of a retrotransposon may be of value in isolating RTVL-H sequences capable of encoding the requisite proteins for transposition.

I 227 INTRON STRUCTURE IN FILARIAL NEMATODES, Steven A. Williams and Daniel J. Freedman, Department of Biological Sciences, Smith College, Northampton,MA 01063 and Program in Molecular and Cellular Biology, University of Massachusetts, Amherst,MA 01003.
Two basic models have been proposed to explain the nearly universal presence of intervening sequences within the protein-coding regions of eukaryotic genes. Doolittle (Am.Nat.1987.130:915-928) has summarized them succinctly as the "introns early" scenario and the "introns late" model. In order to synthesize these views into a unified theory regarding the origin and function of introns, more data from a variety of organisms is necessary. A comprehensive study of the free-living nematode, Caenorhabditis elegans, has defined two classes of introns with respect to the distribution of information-containing sites within regions required for splicing (Fields, Nuc. Acids Res.1990 18:1509-1512). Here we present a comparative study of intron sequences from genes coding for major surface and structural proteins from filarial nematodes including Brugia malayi, Brugia pahangi, Dirofilaria immitis, and Onchocerca volvulus. This class of nematodes includes organisms causing widespread human disease such as African River Blindness and elephantiasis. These intron sequences will be compared to intron sequences from two non-filarial nematodes: Ascaris lumbricoides and Caenorhabditis elegans.

I 228 ALTERNATIVE SPLICING PATTERNS CHARACTERIZE RETROVIRAL-LIKE INSERTIONS IN THE MAIZE WAXY GENE, Michael Purugganan, Marguerite Varagona and Susan Wessler, Botany Department, University of Georgia, Athens, Georgia 30602
The maize waxy gene encodes an enzyme required for amylose biosynthesis in the endosperm and pollen. Four spontaneous mutant waxy alleles result in leaky expression in the endosperm and, in two cases, in the pollen. Genomic cloning of these alleles - wxM, wxStonor, wxG and wxB5 - reveal the presence of retroviral-like insertions within the waxy transcription unit. The wxStonor element was inserted within a splice acceptor site, causing splice site duplication at the insertion point. Both wxG and wxB5, which are related elements, are inserted within introns. These insertions create large introns which are nonetheless spliced out in the mature transcript. Moreover, cDNA sequencing has revealed that the presence of these elements create a variety of alternative splicing patterns, utilizing multiple donor and acceptor sites that result in exon skipping. The disruption of splice site recognition is long-range, involving splice sites up to 5 exons away. The possible role of retroviral-like elements in creating alternative splicing patterns is presented.

Late Abstracts

EXON SHUFFLING AND THE UNDERLYING MOTIFS OF PROTEIN EVOLUTION, Robert L. Dorit, Lloyd Schoenbach and Walter Gilbert, Department of Cellular and Developmental Biology, Harvard University, Cambridge, MA. 02138.
This study updates and expands our previous analysis of the frequency of exon shuffling in modern eukaryotic proteins (Science, in press). We examine all protein-encoding DNA sequences of known intron-exon structure available in the Genbank and EMBL databases. When homologous proteins or multiple related members of a gene family are present in the databases, we retain only single examples of each protein. This distilled database is translated into a collection of individual exons; we then carry out pairwise amino-acid sequence comparisons using all exons in the collection. Using statistically significant sequence similarity as our criterion, we identify all cases where putatively homologous exons are present in non-homologous proteins. We consider such exon pairs to be examples of exon shuffling. The eukaryotic protein sequence motifs involved in shuffling events are then used to screen all available prokaryotic proteins. We also use the frequency of exon shuffling to estimate the size of the original exon universe from which the ancestors of all modern proteins were likely assembled.

POTENTIAL REGULATORY ELEMENTS IN NUCLEAR GENES, WHICH ENCODE SUBUNITS OF RESPIRATORY CHAIN COMPLEXES, MAY RESIDE IN THEIR INTRONS

John D. Phillips and Bernard L. Trumpower, Department of Biochemistry, Dartmouth Medical School, Hanover, NH 03756

The mitochondrial bc_1 complex of *S. cerevisiae* is composed of 8 nuclear and one mitochondrial encoded subunits. The 11 subunits of cytochrome c oxidase are similarly encoded by both nuclear and mitochondrial genomes. Cloning and sequencing of the nuclear gene for the smallest subunit of the bc_1 complex, *QCR9*, revealed the presence of a 213 bp intron, which separates the initial methionine from the rest of the protein. The intron contains a region that has a high degree of homology to an intron in *COX4*, a nuclear gene that encodes a subunit of cytochrome c oxidase. In this homologus region there are a minimum of 15 out of 18 bases which are identical. The cDNA for *QCR9* is being used to test the essentiality of the intron, and the homologus sequence is being examined for activity in regulating expression of these electron transfer complexes.

COX 4 intron

5' Sp Seq		Con Sp Seq	3' Sp Seq
AATGTATGT..142bp..	TTAACTCGTTCCCTGCTTTTCTATTTGG..118bp..	TACTAAC..41bp..	AAATAGAT
ATGGTATGT..150bp..	TGAATATGTTCCATGCGTTTCCATTCAG...2bp...	TACTAAC..14bp..	GAGTAGTC

QCR 9 intron