11-13-2003

Making and breaking the message

David A. Mangus
University of Massachusetts Medical School

Ambro van Hoof

Meeting report

Making and breaking the message

David A Mangus* and Ambro van Hoof†

Addresses: *Department of Molecular Genetics and Microbiology, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655, USA. †Department of Microbiology and Molecular Genetics, University of Texas Health Sciences Center, 6431 Fannin JFB 1.700, Houston, TX 77030, USA.

Correspondence: David A Mangus. E-mail: david.mangus@umassmed.edu

Published: 21 October 2003

Genome ***Biology*** 2003, **4:**346

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2003/4/11/346

A report on the Cold Spring Harbor Laboratory meeting 'Eukaryotic mRNA Processing', Cold Spring Harbor, USA, 20-24 August 2003.

Amid the Cold Spring Harbor Laboratory celebrations marking the 50th anniversary of the elucidation of the structure of DNA, researchers gathered to discuss recent advances in understanding the mechanisms regulating gene expression. The formation of a functional mRNA from the initial pre-mRNA synthesized by RNA polymerase II requires the addition of a cap and poly(A) tail as well as removal of introns. These individual reactions have traditionally been studied in isolation, leading to a perception that they are unconnected. It has, however, become clear that the events that transform a pre-mRNA into an mRNA are often carried out by large ribonucleoprotein (RNP) complexes and that these reactions are coupled and interdependent. Interestingly, these events and the links between them provide quality-control checkpoints to ensure that the message encoded in the genome is accurately communicated to the translation apparatus. These underlying concepts framed the presentations we describe below.

Origins and mechanisms of pre-mRNA splicing

The removal of introns from a pre-mRNA is catalyzed by a large set of RNAs and proteins collectively termed the spliceosome. Exciting advances have been made in the visualization of the spliceosome complexes within the last year. Cryoelectron micrographs presented by Melissa Jurica from the laboratory of Melissa Moore (Brandeis University, Waltham, USA), Cindy Will from Reinhard Lührmann's group (Max Planck Institute for Biophysical Chemistry, Göttingen, Germany), and Melanie Ohi from the laboratory of Thomas Walz (Harvard Medical School, Boston, USA) all indicate that the active spliceosome (C complex) has three domains of density surrounding a central cavity. The spliceosome structures are roughly 270 × 250 × 240 Å in size, with resolutions ranging between 20 Å and 40 Å. The current challenges are to achieve greater resolution and to locate important proteins and RNAs within the structure. We expect that the generation of high-resolution structures of the spliceosome will be as reinvigorating for research into the biochemistry of splicing as the determination of the structure of the ribosome has been for the study of translation.

Because the mechanism of pre-mRNA splicing by the spliceosome is very similar to that observed for the group II self-splicing introns, it has been widely postulated that the catalytic core of the spliceosome would be formed by the small nuclear RNAs (snRNAs). James Manley's group at Columbia University (New York, USA) previously reported that a mixture of three RNA molecules derived from the spliceosome and intron (U2 and U6 snRNAs and branchpoint RNA) has catalytic activity that produces an RNA species called 'RNA X'. This catalytic activity had only limited similarity to splicing, however. At the conference, Saba Valadkhan from the Manley laboratory reported that a different product (RNA Y) is formed when the 5′ splice site is tethered to the U6 RNA. By several criteria, RNA Y resembles the product of the first splicing step, in which the branchpoint is linked to the 5′ end of the intron through a 2′-5′ phosphodiester bond. It therefore appears that the spliceosome is indeed a ribozyme.

Answers to the question of how introns arose during evolution have remained elusive. One long-standing hypothesis is that introns arose very early and facilitated the evolution of proteins by exon shuffling. Support for this hypothesis has previously been found in the fact that introns are more

frequently found between codons than in the two positions that interrupt codons. Richard Padgett (Cleveland Clinic Foundation, USA) reported that this distribution is true only for the major U2-dependent class of introns, and that a minor class of U12-dependent introns is less frequently found between codons. He suggested that these introns were initially more prevalent, but have been unilaterally converted to U2-type introns, and that the current distribution of introns should not be used to address the introns-early versus introns-late conundrum.

Interdependency of different processes

Several presentations addressed the links between the RNA processing events that lead to maturation of mRNA. One of the most dramatic, by Susan Janicki from the laboratory of David Spector (Cold Spring Harbor Laboratory, New York, USA), included movies visualizing gene expression in living mammalian cells. Using a reporter gene that allowed simultaneous localization of the gene, the RNA and the encoded protein, as well as various tagged proteins, the group can observe the different steps in gene expression. Upon activation of the gene, RNA polymerase II, splicing factors and polyadenylation apparatus are all recruited to the site of transcription and the encoded mRNA accumulates. Simultaneously, heterochromatin proteins leave the transgene locus and the chromatin becomes less compact.

As splicing of pre-mRNA can occur co-transcriptionally, it is possible that the rate of transcription could influence the selection of 5′ and/or 3′ splice sites. Nicole Robson from the group of Mariano Garcia-Blanco (Duke University, Durham, USA) reported on experiments designed to test this hypothesis in which they determined the effects of RNA pausing on the alternative splicing of the pre-mRNA encoding the fibroblast growth-factor receptor FGFR2. The team observed that the regulation of alternative splicing of this pre-mRNA required active transcription and that the presence of an RNA polymerase II pause site promoted inclusion of nearby exons. Similarly, Alberto Kornblihtt (University of Buenos Aires, Argentina) described how his group used slow alleles of RNA polymerase to demonstrate effects on multiple modes of alternative splicing. Together, these studies indicate an important role for RNA polymerase II in the control of splicing.

Numerous studies have reported 'unusual' non-transcriptional effects of promoter changes on gene expression. How is it possible that promoter sequences can influence downstream events in gene expression? Emanuel Rosonina from the laboratory of Benjamin Blencowe (University of Toronto, Canada) described a correlation that the group observed between the strength of activation and the efficiency of splicing and 3′-end cleavage, using a reporter gene that can be activated by a variety of *trans*-acting factors. From this observation, it was inferred that transcriptional activators can influence recruitment of factors required for pre-mRNA processing.

DNA damage leads to a strong but transient block in polyadenylation. James Manley's laboratory had previously shown that a BRAC1/BARD1 heterodimer interacts with the polyadenylation factor CstF and represses 3′-end formation. At this meeting, Frida Kleiman from the Manley lab reported their further observations that the BRAC1/BARD1 factors ubiquitinate the elongating, phosphorylated form of RNA polymerase, RNAP IIO, and that this activity is dependent on the presence of CstF. Collectively, their data suggest that the targeted degradation of RNA polymerase during DNA repair reflects functional interactions between the machineries for transcription, 3′-end processing and DNA repair.

In yeast, where the number of introns is relatively low, DNA microarrays have been generated that can distinguish each unspliced pre-mRNA from its mature mRNA. Todd Burckin from Grant Hartzog's lab (University of California Santa Cruz, USA) reported on a collaboration between the Hartzog lab and Manuel Ares' group (also at University of California Santa Cruz) in which these 'splicing-sensitive' arrays were used to understand better the effects of mRNA-processing events on splicing by assaying the effects of 70 different yeast mutants defective in mRNA synthesis and processing. Their analysis revealed that defects in mRNA capping or export affect the splicing of many genes. In contrast, mutations in the carboxy-terminal domain of RNA polymerase II affect only some genes, whereas mutants in the cleavage and polyadenylation machinery have relatively little effect.

Maturation of the 3′ ends of most pre-mRNAs requires a site-specific cleavage followed by addition of the poly(A) tail. The enzyme that synthesizes the tail, poly(A) polymerase, was initially characterized and cloned in the early 1990s, but the identity of the factor that mediates the cleavage reaction has remained elusive. In a poster presentation from Jim Manley's lab, Kevin Ryan presented preliminary data suggesting that the protein CPSF-73 (in metazoans, or Ysh1p in yeast) is the long-sought endonuclease. This protein is similar to members of the β-CASP family, which participate in nucleic acid metabolism. More significantly, this factor is not present in the cytoplasmic polyadenylation apparatus, which does not have a cleavage activity, and a mutation in the predicted active site of Ysh1p is lethal. Although it remains to be shown definitively that CPSF-73 has endonuclease activity, this is a promising development in an area that has frustrated researchers for a quarter of a century.

Alexandre Akoulitchev (University of Oxford, UK) reported that termination of transcription of the human β-globin mRNA is dependent on cleavage of the nascent RNA. This cleavage is distinct from the cleavage and polyadenylation reaction and does not appear to require *trans*-acting factors. Instead, cleavage occurs *in cis* by a ribozyme present in the

globin transcript. The ribozyme sequence is highly conserved in the globin genes of primates, but it is unclear what the function of this ribozyme is and whether similar ribozymes are present in other genes.

Quality-control mechanisms in gene expression and mRNA decay

One of the main factors that degrades aberrant mRNAs in the nucleus is the exosome. David Tollervey (University of Edinburgh, UK) reported the discovery of a new exosome subunit, Rrp47p/Lrp1p. He also described gradient-centrifugation and western-blotting experiments that show that the Rrp6 protein (another subunit of the nuclear exosome) is present in three complexes that are distinct from the exosome. Many functions have been ascribed to the nuclear exosome because of their disruption in an *rrp6Δ* mutant. Analysis of an *rrp47* mutant strain revealed that many RNA-processing reactions indeed appear to be carried out by the exosome. The previously reported role for the exosome in degrading aberrant mRNAs is not affected in the *rrp47* mutant, however, and may therefore reflect functions of the non-exosomal Rrp6p-containing complexes.

Proper processing of an mRNA in the nucleus leads to its preferential export to the cytoplasm. How does the export machinery differentiate mature from pre-mature mRNAs? This question was addressed by Anita Corbett (Emory University, Atlanta, USA). In yeast, the hnRNP protein Nab2p is thought to function in poly(A) tail synthesis. Corbett's group has identified two components of the nuclear pore, Mlp1p and Mlp2p, which interact with Nab2p. When Mlp binding to Nab2p was inhibited, polyadenylated mRNA was retained in the nucleus. More surprisingly, deletion of the nonessential *MLP* genes results in increased transport of polyadenylated mRNA. This enhanced export may be due to increased export of improperly processed mRNAs, a notion supported by genetic interactions observed between nuclear mRNA decay factors and the Mlp proteins. Taken together, these data suggest that the Mlp proteins provide a quality-control checkpoint to export properly processed mRNAs.

Most eukaryotes preferentially degrade aberrant mRNAs that carry a premature stop codon in a process termed nonsense-mediated decay (NMD). Premature stop codons in mammalian mRNAs are recognized when a stop codon is encountered more than 55 nucleotides upstream of the junction of the last two exons. The exon-exon junction is recognized by the presence of an exon-junction complex (EJC) that is deposited by the spliceosome. David Gatfield from the laboratory of Elisa Izaurralde (European Molecular Biology Laboratory (EMBL) Heidelberg, Germany) reported the surprising result that NMD in *Drosophila* more closely resembles NMD in yeast, in that a poorly defined element downstream of the premature stop codon is required instead of the EJC. Importantly, however, flies do have homologs of most EJC components, suggesting that the EJC may have a role distinct from that of NMD. Chia Chan from the laboratory of Gideon Dreyfuss (University of Pennsylvania, Philadelphia, USA) and Thomas Tange from Melissa Moore's group (Brandeis University) independently reported the discovery of several additional components of the EJC. One protein identified by both groups is the RNA helicase eIF4AIII. Nahum Sonenberg's group at McGill University, Montreal, Canada, has previously shown that eIF4AIII is homologous to, but functionally distinct from, the translation initiation factors eIF4AI and eIF4AII. In their published work, Sonenberg's group showed that eIF4AIII inhibits translation in reticulocyte lysates, although at concentrations significantly higher than are physiological. One possible explanation is that eIF4AIII represses translation on freshly spliced mRNAs. An alternative explanation is suggested by a previous report from Moore's lab that spliced mRNAs yield more protein per mRNA than an identical mRNA that never contained introns. This suggests that eIF4AIII may be important for the stimulation of translation by the EJC. A unifying hypothesis is that the EJC functions to signal to the translation apparatus that the message has been properly processed.

Assembly of spliceosome and RISC complexes

Most, if not all, reactions involved in gene expression are carried out by large multiprotein complexes. Several presentations addressed how these complexes are assembled from their individual components. In one example, Jean Beggs (University of Edinburgh) reported that the nuclear Lsm complex (Lsm2p-Lsm8p) in yeast is directly involved in U4/U6 assembly. The Lsm/U6 complex can be formed from Lsm proteins translated *in vitro* and U6 snRNA transcribed *in vitro*. This U6/Lsm complex was more efficient than naked U6 snRNA in assembling with *in vitro* transcribed U4 snRNA. U4/U6 assembly was further enhanced by the Prp24 protein, which has previously been implicated in this process. The U4/U6 complex has to be reassembled after every splicing reaction. Mutant extracts lacking one of the Lsm proteins could carry out splicing but not multiple rounds of splicing, providing further support for a role for Lsm proteins in U4/U6 reassembly.

The RNA interference (RNAi) pathway has become a favorite tool for functional genomics, as it allows for inactivation of specific gene expression in a wide range of model organisms, but the mechanism(s) of RNAi remain largely uncharacterized. A key step in RNAi appears to be the assembly of the RISC complex, which contains small interfering RNAs (siRNAs) of 21-23 nucleotides together with multiple proteins. Erik Sontheimer (Northwestern University, Evanston, USA) described the results of a native gel electrophoresis analysis looking at siRNA-containing complexes. This study identified three complexes (R1, R2 and R3), of which the R3 complex is most like the previously described RISC complex,

but much larger in size. The R1 and R2 complexes appear to be precursors that can be chased into the R3 form. Qinghua Liu from Xiaodong Wang's group (University of Texas Southwestern Medical Center, Dallas, USA), Tingting Du from the laboratory of Phillip Zamore (University of Massachusetts Medical School, Worcester, USA), and Sontheimer all reported biochemical analyses of fly mutants defective in RISC complex assembly. Liu reported that purified Dicer activity (the activity that generates the siRNAs) from *Drosophila* cells contains the protein Dicer2 as well as the R2D2 protein, a homolog of the *Caenorhabditis elegans* RDE-4 protein that is required for RNAi. R2D2 does not affect Dicer activity *in vitro*, but rather stimulates the incorporation of the siRNA product generated by Dicer into the RISC complex. Du reported that the *Drosophila* protein Armitage, an RNA helicase whose *Arabidopsis* homolog SDE3 is required for RNAi, is also required for RISC complex assembly on siRNAs. These studies provide further support for the notion that the RNAi pathway is conserved between plants and metazoans. A better understanding of the mechanisms of RNAi should further increase its utility as a genetic technique.

The importance of understanding mRNA processing has increased in the genomics era. As the number of sequenced eukaryotic genomes gets larger, the task now is to identify and define the messages contained within them. The human genome is currently predicted to have roughly 30,000 protein-coding genes, but 40-60% of human genes are likely to be subject to alternative splicing, suggesting that the genome encodes many more products. The physiological relevance of alternative splicing has been clearly demonstrated in a variety of ways. Far less is known about how other mRNA-processing events and alternative translation initiation might impact on the proteome. The further identification and characterization of alternative mRNA processing events, as well as the factors controlling them, will help clarify the capacity of the genome. For example, our understanding of alternative splicing will be advanced significantly with the advent and commercial availability of microarrays sensitive to alternative splice forms. Such tools should generate important insights into mRNA processing mechanisms and the diseases caused when those mechanisms go awry. We look forward to seeing advances in these and other areas at the next 'Eukaryotic mRNA Processing' meeting in 2005.