Biochemistry 201: RNA Processing
Herschlag
Feb. 16, 2000

Lecture 2. The odyssey of mRNA: An overview

I. The simple view:

Figure 1 provides a sketch of what a mRNA has to go through before it can perform its most obvious biological function of coding for a protein.

The 5' End:
-Transcription (you heard about this in Roger Kornberg's lectures)

-As the RNA is transcribed it is "capped" on its 5' end. This only occurs with RNA polymerase II (pol II) transcripts. As suspected, the enzyme that does has been shown to associate with pol II -its CTD (C-terminal domain).

[There was prior evidence for this type of interaction for a viral polymerase and capping enzyme (Hagler & Shuman (1992) *Science* 255, 983-986: "A freeze-frame view of eukaryotic transcription during elongation and capping of nascent mRNA").]

The Middle:
-Nuclear pre-mRNAs from higher eukaryotes often contain more intron than exon sequence.

-The introns are removed in the nucleus by the spliceosome in a process that involves several steps (Alluded to in last lecture).

The 3' End:
-At the other end of the pre-mRNA, the 3' end is determined by cleavage of the growing RNA chain. The sequence UUAUUU about 15-25 nucleotides 5' of a U or UG rich sequence determines this site for cleavage. Following cleavage, a poly(A) tail ~200 in length is added onto the new 3' end. The enzymes involved in this process have been isolated and are being studied in order to understand how the cleavage site is chosen and how the length of the poly(A) tail is determined (Wahle & Keller, *Annual Rev. Biochem.* (1992) 61, 419-440: "The biochemistry of 3'-end cleavage and polyadenylation of messenger RNA").
-The polyadenylation factors are also associated with the CTD of Pol II. (Hirose & Manley (1998) *Nature* 395, 93-96: "RNA polymerase is an essential mRNA polyadenylation factor.")

Getting Out of the Nucleus:
The processed mRNA exits the nucleus via a nuclear pore.

In the Cytoplasm:

The mRNA can associate with ribosomes and be translated to give protein.

II. The Complexities of pre-mRNA Processing

-What's wrong with this picture: Figure 1 revisited!!

IIA. RNA as a ribbon?

-RNA, left alone, folds into multiple structures that are stable for long periods of time:

In Vitro Observations:

-Isolated tRNAs often need to be renatured in order to be charged by tRNA amino acyl synthetases: Adams, Lindahl & Fresco, Proc. Natl. Acad. Sci. USA (1967) 57, 1684-1691: "Conformational differences between the biologically active and inactive forms of a tRNA").

-RNAs often adopt multiple and inactive structures. For example, often in self-splicing reactions, a fraction of the group I intron will excise itself fast, but the remaining fraction will self-splice much slower. Sometimes the RNA can be renatured (e.g. Walstrum & Uhlenbeck, Biochemistry (1990) 29, 10573-10576).

-Misfolded RNAs can be extremely stable kinetically, with alternative conformations often remaining for days (e.g., Fedor & Uhlenbeck, Proc. Natl. Acad. Sci. USA (1990) 87, 1688-1672: "Substrate sequence effects on "hammerhead" RNA catlaytic efficiency").

This is not the behavior we are used to with proteins. What about RNA could be responsible? (Figure 2)

-20 amino acid side chains vs 4 RNA side chains: Diversity & uniqueness.

-In secondary structure, protein side chains point out, ready for interactions that can uniquely determine tertiary structure; RNA side chains point in in base pairs, masking their differences.

-Protein secondary structure is of marginal stability, so that if the wrong secondary structure forms it is quickly unformed; in contrast RNA secondary

structure is very stable. For example, a duplex with 10 base pairs can take a year to dissociate into single strands!!

-Proteins can help RNA fold and unfold. For example, proteins can denature tRNA. hnRNP's and related proteins can speed duplex formation and dissociation (Pontius & Berg, *Proc. Natl. Acad. Sci.* (1990) 87, 8403-8407; Tsuchihashi, Khosla & Herschlag, *Science* (1993) 262, 99-102; Herschlag, Khosla, Tsuchihashi & Karpel, *EMBO J.* 13, 2913-2924).

-The problems that RNA has in folding and misfolding must be dealt with by the cell.

In Vivo Observation:

As the RNA transcript is made, it is coated with proteins to form hnRNP particles (heterogeneous nuclear ribonucleoprotein). There are >20 different hnRNP proteins that seem to coat the RNA (Dreyfuss, Matunis, Pinol-Roma & Burd (1993) *Annu. Rev. Biochem.* 62, 289-321: "hnRNP proteins and the biogenesis of mRNA.")

This suggests that the hnRNP proteins play the biological role of "RNA chaperones"
(Herschlag (1995) J. Biol. Chem.270, 20871-20874: "RNA Chaperones and the RNA Folding Problem"). (Chaperones are proteins that help proteins fold correctly by preventing their misfolding into aggregates.)
(Clodi, Semrad & Schroeder (1999) EMBO J. 18, 3776: "Assaying RNA chaperone activity in vivo using a novel RNA folding trap")

Relating back to the RNA world, it is reasonable (but not established) that the hnRNP's were initally used as RNA chaperones and that later in evolution they were co-opted into additional functions:

-To help distinguish different mRNAs, and perhaps control their turnover. hnRNP proteins have been shown to exhibit some specificity in the RNA sequences that they prefer (e.g., Bennett, Pinol-Roma, Staknis, Dreyfuss & Reed, *Molec. Cell. Biol.* (1992) 12, 3165-3175: "Differential binding of heterogeneous nuclear ribonucleoproteins to mRNA precursors prior to spliceosome assembly in vitro").

-To help splicing or help control splicing. hnRNP protein A1 has been shown to influence splice site choice in vitro (Mayeda & Krainer, *Cell* (1992) 68, 365-375: "Regulation of alternative splicing by hnRNP A1 and splicing factor SF2").

-To help retain RNAs in the nucleus until they are fully processed and to aid in transport out of the nucleus (see below).

Now we have hnRNP, but pol II keeps going after the cleavage event, continuing on to make apparently non-functional transcripts that are hundreds or thousands of nucleotides long. These RNAs are rapidly degraded, presumably because they lack the 5' cap.

Why is so much RNA made, when so little is actually used?

-Only about 5% of total RNA makes it out of the nucleus into the cytoplasm!

-The lost RNA is from introns, the 3' region beyond the cleavage/polyadenylation site, and turnover of RNA without release.

-Why is the process so complex and so inefficient!?

1. For control

2. Because the evolutionary driving force is not large enough.
-This can arise if the organism doesn't care as much about energy efficiency as environmentally concerned individuals care about fuel efficiency in automobiles. Alternatively, the organism might care, but there may be no accessible selection pathway toward greater efficiency. (I.e., How facile are the evolutionary mechanisms to evolve greater efficiency? How easily can the molecules be rearranged through selection to do the job?)

IIB. Complexities of pre-mRNA Splicing

Guthrie (1996) The Harvey Lectures, Series 90, p59-80: "The spliceosome is a dynamic ribonucleoprotein machine"

-Compare with group II self-splicing: Same chemical pathway, but now >50 proteins & 5 RNP complexes!

-'Lariat' identified based on unusual chemical properties:
-Resistant to RNase & other nucleases
-Primer extension stop
-Anomolous migration by PAGE

-Discovery of protein factors:
Yeast genetic screens, synthetic lethals, etc.
Purification from mammalian system

Neubaur, King, Rappsilber, Calvio, Watson, Ajuh, Sleeman, Lamond & Mann (1998) Nature Genetics 20, 46: "Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome"

-Why is pre-mRNA splicing so complex? (And for that matter, all of RNA processing)

1. It has a more difficult problem than in self-splicing, as many different exon/intron junctions must recognized. Especially in mammalian splicing, the consensus sequences are quite weak. It has recently been suggested that proofreading is involved to ensure choice of the correct exons (Guthrie, *Science* (1991) 253, 157-163: "Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein"). Need to break a process up into distinct steps to allow proofreading.

The general principle of proofreading is that an input of energy is required to allow organization and control -there's no such thing as a free lunch. (It is also worth reading about and understanding the thermodynamic and kinetic requirements for proofreading as this comes up in all areas of biochemistry. [Thopmson (1988) *TIBS* 13, 91-93: "EFTu provides an internal kinetic standard for translational accuracy"; Yarus (1992) *TIBS* 17, 130-133: "Proofreading, NTPases and translation: constraints on accurate biochemistry"; the classic in the field, though incorrect in the detailed mechanism suggested, is Hopfield (1974) *Proc. Natl. Acad. Sci. USA* 71, 4135-4139: "Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity"; Burgess & Guthrie (1993) TIBS 18, 381-384: "Beat the clock: Paradigms for NTPases in the maintenance of biological fidelity."]

2. A complex pathway can allow multiple options for control, allowing alternative splicing. Examples of alternative splicing abound. The classic example is sex determination in Drosophila. The basic principles here, at the very crude level they are known, appear to be analgous to those in trascriptional control: The use of protein factors, whose expression or activity is controlled, to either recruit a splicing factor to a weak splicing site or to block a strong splicing site.

Mattox, Ryner & Baker, *J. Biol. Chem.* 267, 19023-19026: "Autoregulation and mulifunctionality among trans-acting factors that regulate alternative pre-mRNA processing"

Valcarcel, Singh, Zamore & Green, *Nature* (1993) 362,171-175, "The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA".

Tian & Maniatis, *Science* (1992) "Positive control of pre-mRNA splicing in vitro"

Berget (1995) *J. Biol. Chem.* 270, 2411-2414: "Exon recognition in vertebrate splicing"

Hertel, Lynch & Maniatis (1997) *Curr. Opin. Cell Biol.* 9, 350-357: "Common themes in the function of transcription and splicing enhancers"

Lopez (1998) *Annu. Rev. Genet.* 32, 279-305: "Alternative splicing of pre-mRNA: Developmental consequences and mechanisms of regulation"

Grabowski (1998) *Cell* 92, 709-712: "Splicing regulation in neurons: Tinkering with cell-specific control"

-Some of the details:

Complex assembly pathway (native gel analysis)

Kinetics of cross-linking "Sontheimer & Steitz (1993) Science 262, 1989: "The U5 and U6 small nuclear RNAs as active site components of the spliceosome"

Mutually exclusive pairings (see below)

Six base pairing rearrangements in the 1st step (Figure 3a&b).

Several steps require DEAD box proteins (Figure 4&5).

-What are DEAD box proteins and what do they do?

-Share sequence/ structural homology with DNA helicases
-RNA-dependent ATPases
-Some can unwind RNA duplexes
-Involved in every interesting biological process that utilizes RNA:
    -Splicing
    -mRNA transport
    -Ribosome biogenesis
    -Translation initiation
    -RNA localization
    -RNA turnover
-But in no case is the molecular process understood (Figure 6)

Staley & Guthrie (1998) Cell 92, 315: "Mechanical devices of the spliceosome: Motors, clocks, springs, and things"

Jankowsky, Gross, Shuman & Pyle (2000) Nature 403, 447: "The DexH protein NPH-II is a processive and directional motor for unwinding RNA"

U1/U6 base pairing switch (Figure 7):
(Staley & Guthrie (1999) Molecular Cell 3, 55: "An RNA switch at the 5' splice site requires ATP and the DEAD box protein prp28")

Background:
-U1 initially recognizes 5'-splice site, but replaced by U6 prior to first chemical step. This replacement requires ATP.
Experiment:
-Hyperstabilize U1•5'-splice site base pairing
Reduced splicing
-Compensated by increased 5'-splice site•U6 base pairing
(Why?)
-Screened for mutants that exacerbate the effect from hyperstabilized U1•5'-splice site base pairing: prp28 mutant: a DEAD box protein (Figure 8)
-Further biochemical characterization:
Form stalled complex with mutant prp28 & then 'chase' with wt prp28 & cold pre-mRNA: Splicing completed.

U2 base pairing switch with U4/U6
(Madhani & Guthrie (1992) Cell 71, 803: "A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome"

Background (Figure 9):
-U4/U6 form complex, but after spliceosomal assembly, U4 is no longer tightly associated with the complex.
-Inspection of sequences in U2 and U6 led to the hypothesis that there are specific base pairing interactions between these two snRNPs.

Experiments (Figure 10):
-The residues in U2 and U6 were changed to disrupt and restore these potential base pairs. (Rescue of activity is a particularly strong result.)
-Rescue was observed for several of the proposed base pairs; a couple of positions did not give rescue, leading to the suggestion that the base identity and not just the presence of a base pair is critical at these positions.

General: Power of 'genetics with RNA:
•Easy to stabilize as energetic effects local, compared with proteins.
•'Second site suppression' very powerful as base pairing rules simple and local.

IIC. RNA Transport

-Subset of hnRNP and SR proteins shuttle between nucleus & cytoplasm.
(Caceres, Scrfeaton & Krainer (1998) *Genes & Develop.* 12, 55-66: "A specific subset of SR proteins shuttles continuously between the nucleus and cytoplasm.")
-Thus transport is presumably of (remodeled) protein/RNA complex.
-The 5'"cap" may serve as an export signal
(Hamm & Mattaj, Cell (1990) 63, 109-118: "Monomethylated cap structures facilitate RNA export from the nucleus").
-However, very little is known about how RNA is transported.

-hnRNP's and snRNP's appear to be rather specifically localized within the nucleus, and there is evidence that hnRNP's are associated with the nuclear matrix.
(Subcellular organelles should not be considered "bags of enzymes".)

-In vivo hybridization experiments have suggested that a viral RNA exits the nucleus on "tracks" that lead to a nearby nuclear pore. Such localized transport into the cytoplasm might help in establishing organization within the cytoplasm. (Lawrence, Singer & Marselle, *Cell* (1989) 57, 493-502: "Highly localized tracks of specific transcripts within interphase nuclei visualized by in situ hybridization")

IID. In the Cytoplasm

-Must find translation initiation factors and ribosomes (next Lectures)

-Specific RNAs can be localized to certain regions within the cytoplasm. This is thought to play a role in establishing gradients that are necessary for development (Gottlieb, *Curr. Opin Cell Biol.* (1990) 2, 1080-1086: "Messenger RNA transport and localization"; Melton, *Science* (1991) 252, 234-241: "Pattern formation during animal development"; Mowry & Melton, *Science* (1992) 255, 991-994: "Vegetal mRNA localization directed by a 340-nt RNA sequence element in Xenopus Oocytes"; Sundell & Singer, *Science* (1991) 253, 1275-1277: "Requirement of microfilaments in sorting of actin mRNA")

-mRNA's in the cytoplasm vary in their lifetime, with the average half-life being ~30 min. Some sequences important in determining turnover rates have been. The turnover is controled in many cases, with factors both increasing and decreasing the turnover. (e.g., Malter & Hong, *J. Biol. Chem.* (1991) 266, 3167-3171: "A redox switch and phosphorylation are involved in the post-translational up-regulation of the adenosine-uridine binding factor by phorbol ester and ionophore"; Ratnasabapathy, Hwang & Williams, *J. Biol. Chem.* (1990) 265, 14040-14055: "The 3'-untranslated region of apolipoprotein II mRNA contains two independent domains that bind distinct cytosolic factors")

III. Other Fun Facts About RNA: More Modifications to the Central Dogma

-RNA editing

(Bass "RNA editing: New uses for old players in the RNA world" , "The RNA World (1993) Cold Spring Harbor Press, pp 383-418; Simpson "RNA editing -An evolutionary perspective", in 'The RNA World' (1999) Cold Spring Harbor Press, pp585-608.)

-RNA modifications

-Nonsense mediated decay

-Nuclear amino acylation of tRNA

-RNAi

(Sharp (1999) Genes & Devel. 13, 139-141: "RNAi and douible strand RNA")

-Frameshifting in translation

-Protein splicing

(Perler et al. (1992) *PNAS* 89, 5577-5581: "Intervening sequences in an Archaea DNA polymerase gene"; Gimble & Thorner (1992) *Nature* 357, 301-306: " Homing of a DNA endonuclease gene by meiotic gene conversion in Sacchoromyces cerevisiae".)

Are the exceptions the rule?

IV Lessons & Future Prospects

1. Must understand origins/ paradigm: EVOLUTION!

In order to understand modern day biology, we need to understand how it got to be this way:

One thing we'll see is that the biological answers to problems may be different than those one would come up with by "design".

2. Must understand molecular properties: these define what molecules can and can't do and thus, what the capabilities and limitations of biological systems.

3. (1) & (2) define all that is basic to biology.

Conclusion: An appreciation for the molecular nature of RNA has led to insights into a possible biological role of the hnRNP proteins that otherwise would have remained obscure. However, remember that these observations cannot prove the biological role of hnRNP's. This underscores the necessity of approaching a problem from multiple standpoints.

Questions:
-What is the level of our understanding of RNA processing?
-What approaches will be required to understand how pre-mRNA processing and splicing occur?
-Is it critical to understand these underlying mechanisms to understand cellular control and regulation?

Future questions:
-What is the interplay of different aspects of RNA processing between steps and RNAs and with other cellular components? Does this interplay 'create' properties of systems of molecules that are not simply predicted from the behavior of the individual components?
-What, if any, is the resolution of RNAs in terms of intra-cellular and intra-organellar localization? Are there subsets of RNAs processed via different pathways, different RNA binding proteins, or in different places within the cell?

How Can We Understand the Complexity (and Underlying Principles) of RNA Processing?

-Look at <u>all</u> of the RNAs
Hypothesis-driven vs hypothesis-limited

- Pictures are appealing, but misleading by necessity of their static nature.

Whereas moving pictures (movies) can capture the workings of a 'machine', they are not sufficient to provide appreciation for the stochastic (probabilistic) nature of events on the molecular level.

-Can single molecule detection in vivo save the day?

-Femino, Fay, Fogarty & Singer (1998) Science 280, 585-590: "Visualization of single RNA transcripts in situ".

-Arn & Macdonald (1998) Cell 95, 151-154: Motors driving mRNA localization: New insights from in vivo imaging".

-What haven't we thought of yet?