The flow of gene expression

Tom Misteli

Gene expression is a highly interconnected multistep process. A recent meeting in Iguazu Falls, Argentina, highlighted the need to uncover both the molecular details of each single step as well as the mechanisms of coordination among processes in order to fully understand the expression of genes.

© 2004 Nature Publishing Group http://www.nature.com/natstructmolbiol

Gene expression is often metaphorically described as a flow of information from DNA to protein via transcription, RNA processing, RNA export and translation. This picture of a steady stream of events leading to gene expression has, however, recently been questioned by observations indicating complex functional interrelationships among the various steps[1]. It thus seemed fitting to hold the 2nd Symposium on Gene Expression and RNA Processing from 29 November to 3 December 2003 in Iguazu Falls, Argentina, where the peaceful flow of the Iguazu River is spectacularly interrupted by one of the world's most awe-inspiring waterfalls. This international meeting was superbly organized by Alberto Kornblihtt (University of Buenos Aires, Buenos Aires) with generous financial support from the International Center for Genetic Engineering and Biotechnology (ICGEB, Trieste, Italy).

Transcription basics and complexities

The last two decades have led to a detailed molecular description of transcription factors and complexes. Despite the tremendous progress, fundamental questions remain, maybe none more tantalizing than the issue of how transcription complexes achieve specificity to act on particular genes at particular times. The meeting opened with a keynote address by Robert Tjian (University of California Berkeley, Berkeley, California, USA), who provided two answers to this puzzle. First, cryo-electron microscopy demonstrates that coactivator complexes can take on distinct structures depending on what activator is bound to them. Second, distinct combinations of components in the basal transcription machinery can generate specificity. This principle is impressively illustrated by TAF105, which is expressed in a highly restricted pattern in ovarian granulosa cells and seems to be sufficient to convey tissue specificity. A different paradigm for generating specificity was described by Joaquin Espinosa (Emerson laboratory, The Salk Institute, San Diego) who reported that both the composition and kinetics of assembly of the transcriptional machinery on p53 target genes are distinct depending on the type of DNA damage, further suggesting that the formation of distinct transcription initiation complexes is a critical means of generating specificity in gene regulation[2].

An important mechanism of controlling gene expression is histone modification. Andreas Ladurner (European Molecular Biology Laboratory, Heidelberg, Germany) presented functional data demonstrating that the bromodomains of mammalian TAFII250, the large subunit of the general transcription factor TFIID, and the yeast protein Bdf1 show specificity *in vitro* for binding to acetylated histones and that this ability correlates with a role in the maintenance of euchromatin *in vivo*[3]. The physiological relevance of histone modification became clear in a presentation by Mauro Giacca (ICGEB). He reported that activation of the HIV-1 LTR promoter correlates with acetylation of histones at two discrete nucleosomal regions. This acetylation response parallels the recruitment of different cellular acetyltransferases to the promoter and thus probably serves a major regulatory function[4].

More often than not, transcriptional regulation is controlled by signaling pathways. Diego de Mendoza (University of Rosario, Rosario, Argentina) described an elegant signaling cascade responsible for maintaining membrane fluidity in *Bacillus subtilis* under low temperature conditions. The Des pathway includes a membrane-associated kinase (DesK) that senses the higher state of order in the lipid bilayer and activates a downstream transactivator (DesR) that in turn increases transcription of the *des* gene, an acyl lipid desaturase that counteracts the effect of the lowered temperature.

In mammalian cells, signaling pathways that control gene expression are often involved in cell proliferation and are frequently defective in disease. The group of José Bocco (University of Córdoba, Córdoba, Argentina) has identified Kruppel/Sp1-like transcription factor 6 (KLF6) as a target for viral and cellular oncoproteins and provided evidence that KLF6 might lead to the inhibition of cell proliferation, thus implicating KLF6 as a potential tumor suppressor. Another disease-relevant signaling pathway is the insulin regulatory pathway. Eduardo Canepa (University of Buenos Aires) reported that transcription factor HNF3β activates the 5-aminolevulinate synthase (ALAS) promoter by binding at sites that overlap with an insulin response element. He proposes that this activation is blocked by insulin through Akt-mediated phosphorylation of HNF3β, thus providing another example of a powerful transcriptional regulatory mechanism acting though signaling pathways.

Good and bad pre-mRNA splicing

Although the basic mechanistic principles of the pre-mRNA splicing process are now well established, several exciting new directions in the study of RNA splicing have

The author is at the National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.
e-mail: mistelit@mail.nih.gov

The waterfalls at Iguazu, Argentina. Image courtesy of Juan Valcarcel.

© 2004 Nature Publishing Group http://www.nature.com/natstructmolbiol

recently emerged. One is the realization that splicing defects are highly relevant for numerous disease mechanisms. A second is that alternative splicing is the major mechanism of generating protein diversity from a limited set of genes. Both of these clinically relevant aspects of pre-mRNA splicing are now being intensely investigated. Juan Valcarcel (ICREA-CRG, Barcelona, Spain) has designed microarrays to detect alternative splicing and as a first application showed correlation between the choice of alternatively spliced variants of several RNAs and tumor progression of Hodgkin's lymphoma. Adrian Krainer (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA) described the mechanisms responsible for the different splicing patterns of transcripts from the nearly identical *SMN1* and *SMN2* genes, which are involved in the neurodegenerative disease spinal muscular atrophy. A single nucleotide difference between the two genes in their penultimate exon accounts for the splicing differences. Analysis of the *cis* elements and the *trans*-acting factors that recognize them suggests that the sequence difference gives rise to an exon splicing enhancer in *SMN1* but not in *SMN2*, thus providing a molecular explanation for the reduced activity of the SMN2 protein as compared with that of SMN1 (ref. 5).

A substantial fraction of disease-causing mutations disrupts splicing regulatory elements, resulting in aberrant splicing. There are considerable difficulties in diagnosing mutations of this type because they are context-specific and often look like benign polymorphisms. Francisco Baralle (ICGEB) and his collaborators Marco Baralle and Franco Pagani (ICGEB) discussed a series of examples of mutations of this type located in intronic and exonic regions of several disease genes including neurofibromatosis 1 and CFTR[6,7]. In particular he showed that a previously defined silencing element in the human EDA exon of fibronectin was really an RNA structural determinant specific to the human gene context. This result suggests that secondary RNA structure is an important and underappreciated effector of splicing in disease.

Apart from disease, alternative splicing events are also critically involved in differentiation pathways. To begin to address these complex pathways of regulation, Anabella Srebrow (University of Buenos Aires) reported on the development of a system to study how parts of the cellular microenvironment, such as neighboring cells, growth factors and extracellular matrix, influence alternative splicing of fibronectin in the physiological context of a mammary epithelium cell culture system. She demonstrated that soluble factors can deliver critical signals for splicing control between different cell types.

Understanding aberrant splicing in disease processes requires knowledge of the complex interactions of proteins in the spliceosomes during normal splicing events. New insights into the mechanism of one of the most basic splicing factors, hnRNPA1, were provided by Benoit Chabot (University of Sherbrooke, Quebec, Canada) who suggested that hnRNPA1 may exert both negative and positive roles in splicing by simultaneously repressing looped-out splice sites and facilitating the pairing of splice sites. Finally, Héctor Torres (Institute for Genetic Engineering and Molecular Biology (Buenos Aires) made the important point that components common to the *trans*- and *cis*-splicing machineries are evolutionarily conserved. He described TcSR, a novel serine-arginine-rich SR protein, and a corresponding SR protein-specific kinase, TcSRPK, in the protozoa parasite *Trypanosoma cruzi*, with substantial homology to other metazoan members of the family.

Communication is everything

Cellular processes rarely occur in isolation and are frequently coupled in interconnected process networks. In the case of gene expression, research in the last decade has established a tight link between transcription and RNA processing events by physical interaction of some of the key factors. David Bentley (University of Colorado Health Sciences Center, Denver) has extended this work and presented a model of an 'mRNA factory' comprising RNA polymerase and a dynamic collection of mRNA processing and transport factors that travel along the gene with the transcription machinery during elongation. This factory complex permits communication between the transcription, processing and transport components. A direct functional consequence of coupling transcription and RNA processing was described by Alberto Kornblihtt (University of Buenos Aires). His laboratory demonstrated that a mutant RNA polymerase II with slowed elongation rate affects alternative splicing of several genes including the gene encoding ultrabithorax *in vivo*, resulting in a typical mutant phenotype[8].

A highly unexpected and intriguing link between transcription and RNA processing was described by Nick Proudfoot (University of Oxford, Oxford). In a collaboration with Jane Mellor (University of Oxford), genetic analysis has been done in yeast to uncover a role for chromatin remodeling enzymes in transcription termination[9,10]. Further support for tight coupling of termination and processing came from Proudfoot's work in collaboration with Alexandre Akoulitchev (University of Oxford). They established that termination of transcription for human

β-globin relies on a cotranscriptional cleavage (CoTC) sequence located over the site of RNA polymerase II termination. Biochemical analysis suggests that a ribozyme autocatalytic cleavage activity encoded by the cotranscriptional element is involved in termination.

Linkage and sharing of processing factors do not seem to be limited to the early nuclear steps of gene expression, but to extend to the late cytoplasmic steps of gene expression. Javier Caceres (Medical Research Council, Edinburgh) provided evidence for an association of SR protein splicing factors with translating ribosomes. This association is functionally relevant because it correlates with stimulation of translation both *in vivo* and *in vitro*. These results strongly suggest that shuttling SR proteins have multiple roles in the post-transcriptional expression of eukaryotic genes by coupling splicing and translation.

A highly interconnected and unusual pathway involving tight coordination of transcription, unusual RNA processing and signaling was described by Peter Walter (University of California San Francisco, San Francisco). The unfolded protein response pathway senses accumulation of misfolded proteins in the endoplasmic reticulum. Surprisingly, one component of the pathway, the RNA endonuclease IRE1, is activated by binding of ATP to the protein's kinase domain. The kinase in this pathway is therefore used as a conformational regulator of the endonuclease rather than as a catalytic enzyme[11].

© 2004 Nature Publishing Group http://www.nature.com/natstructmolbiol

Life after pre-mRNA splicing

Nonsense-mediated RNA decay and RNA export have long been viewed as largely unregulated default steps after RNA splicing and processing. This view has now been revised and several talks focused on the tight coupling between processing, nonsense-mediated mRNA decay (NMD) and export. Chris Smith (University of Cambridge, Cambridge) demonstrated using the example of the polypyrimidine tract–binding protein (PTB) that alternative splicing can lead to an mRNA that is removed by nonsense-mediated decay. Bioinformatics analysis shows that this behavior is not limited to the gene that encodes PTB, suggesting that alternative splicing not only generates protein isoform diversity, but can also acts to quantitatively control gene expression via NMD[12,13]. Similarly, Lynne Maquat (University of Rochester, Rochester, New York, USA) demonstrated that NMD in mammalian cells is dependent on splicing and targets newly synthesized mRNA for 5′→3′ and 3′→5′ decay as a consequence of a 'pioneer' round of translation[14,15].

Coupling between export and transcription was described by Michael Rosbash (Brandeis University, Waltham, Massachusetts, USA), who showed that the some mRNA export factors and putative TREX (transcription-export) complex components are cotranscriptionally recruited to transcription sites. Evidence for coupling of splicing and export comes from observations by Joan Steitz's group (Yale University, New Haven, Connecticut, USA), which has identified the three shuttling SR proteins 9G8, SRp20 and SF2/ASF as interactors with the export machinery[16]. Steitz suggests that these SR proteins act as RNA export factors via a phosphorylation cycle in which SR proteins become dephosphorylated during splicing. The removal of the phosphate group then enhances their binding to the mRNA export factor TAP.

Small RNAs strut their stuff

Small, often noncoding, RNAs have been found to be critical components of various gene expression steps. Joan Steitz (Yale University) discussed the biogenesis of small nucleolar ribonucleoprotein particles (snoRNPs), which act as guides to direct modification of ribosomal RNA. Her laboratory has developed a coupled splicing-snoRNA assembly assay to demonstrate that splicing is essential for proper snoRNP assembly. Using this system, they have demonstrated that the snoRNP components join the forming splicing complex at a distinct stage by physically interacting with components of the spliceosome[17].

Although the role of the small nuclear RNAs (snRNAs) and snoRNAs is well established, small RNAs of unknown functions are being discovered in unexpected places. Olivier Bensaude (École Normale Supérieure, Paris) described the association of the 7SK snRNA with Cdk9, cyclin T and the MAQ1/HEXIM1 protein to form the positive transcription elongation factor b (P-TEFb)[18]. Alexandre Akoulitchev (University of Oxford) reported on the association of U1 snRNA with the TFIIH component cdk7–cyclin H[19]. These two examples of small RNAs associating with cyclin-dependent kinases (CDKs) now provoke the question whether the presence of small RNAs is a general feature of CDK–cyclin complexes.

A class of small RNAs that has gained particular attention recently is that of the micro RNAs (miRNAs), small RNAs of 22–24 nucleotides implicated in various aspects of development and differentiation. James Dahlberg (University of Wisconsin, Madison, Wisconsin, USA) described the identification of exportin 5 as the export receptor for the hairpin-containing precursors of miRNA and showed that Exp5 binds its cargo RNAs directly, in a RanGTP-dependent manner[20]. These are some of the first insights into the molecular mechanisms of miRNA export.

Putting things into context

A key question, and another poorly appreciated level of complexity in gene expression, is the issue of how gene expression processes are spatially organized within the cell nucleus. Mauro Giacca (ICGEB) described how several of the cellular factors that control HIV-1 transcription reside in subnuclear PML bodies. Using fluorescence resonance energy transfer, both cyclin T1 and Cdk9 were found to physically interact with the PML protein in these bodies, and upon transcriptional activation PML itself was detected at the HIV-1 promoter. These data suggest that this protein regulates promoter activation by modulating the local availability of essential cofactors of the transcription machinery[21].

To study the spatial organization of gene expression events and to link localization to function, new approaches and tools must be developed. Angus Lamond (University of Dundee, Dundee, Scotland), in close collaboration with Matthias Mann's laboratory (University of Odense, Odense, Denmark), has expanded previous proteomic analysis of nucleoli and the collaborators have now cataloged >650 nucleolar proteins. Using stable isotope labeling with amino acids in cell culture (SILAC), they can accurately quantify dynamic changes in the relative levels of nucleolar proteins at multiple time points, thus providing a powerful tool to unravel the interactions and activities that take place in nuclear organelles.

Another largely unexplored area of gene expression is how the spatial organization of genomes affects gene expression. Tom Misteli (National Cancer Institute, Bethesda, Maryland, USA) described semi-high-throughput methods to map genes and chromosomes in three-dimensional space. Using these methods, his group has obtained evidence suggesting that genomes are organized differentially in different tissues and that the tissue-specific spatial proximity of chromosomes is related to the tissue-specific frequency of chromosomal translocations observed in cancer tissues. Taken together, these cell biological studies are first steps in linking the knowledge of molecular mechanism of gene expression with their *in vivo* organization.

Seeing the forest and the trees

The current studies of gene expression are characterized by an important dichotomy. On one hand, there is an urgent need to continue to uncover the fundamental principles, basic molecular mechanisms and even some of the key players of each step in gene expression. This is often done using highly reductionist approaches with simplified *in vitro* systems. On the other hand, the increasing awareness of the interconnectedness of gene expression steps requires more holistic, often *in vivo*, approaches that can probe the functional interrelationships among the various processes. The presentations, and especially the lively discussions, at the Iguazu meeting made it blatantly clear that both aspects of gene expression must be explored and used in combination to advance our understanding of this most fundamental biological process.

ACKNOWLEDGMENTS
I thank D. Bentley for refreshing my memory and for constructive comments on the manuscript. T.M. is a fellow of the Keith R. Porter Endowment for Cell Biology.

1. Maniatis, T. & Reed, R. *Nature* **416**, 499–506 (2002).
2. Espinosa, J.M., Verdun, R.E. & Emerson, B.M. *Mol. Cell* **12**, 1015–1027 (2003).
3. Ladurner, A.G., Inouye, C., Jain, R. & Tjian, R. *Mol. Cell* **11**, 365–376 (2003).
4. Lusic, M., Marcello, A., Cereseto, A. & Giacca, M. *EMBO J.* **22**, 6550–6561 (2003).
5. Cartegni, L. & Krainer, A.R. *Nat. Genet.* **30**, 377–384 (2002).
6. Pagani, F. *et al. Hum. Mol. Genet.* **12**, 1111–1120 (2003).
7. Baralle, M. *et al. J. Med. Genet.* **40**, 220–222 (2003).
8. de la Mata, M. *et al. Mol. Cell* **12**, 525–532 (2003).
9. Morillon, A., O'Sullivan, J., Azad, A., Proudfoot, N. & Mellor, J. *Science* **300**, 492–495 (2003).
10. Alen, C. *et al. Mol. Cell* **10**, 1441–1452 (2002).
11. Papa, F.R., Zhang, C., Shokat, K. & Walter, P. *Science* **302**, 1533–1537 (2003).
12. Wollerton, M.C., Gooding, C., Wagner, E.J., Garcia-Blanco, M.A. & Smith, C.W. *Mol. Cell* **13**, 91–100 (2004).
13. Lewis, B.P., Green, R.E. & Brenner, S.E. *Proc. Natl. Acad. Sci. USA* **100**, 189–192 (2003).
14. Ishigaki, Y., Li, X., Serin, G. & Maquat, L.E. *Cell* **106**, 607–617 (2001).
15. Lejeune, F., Li, X. & Maquat, L.E. *Mol. Cell* **12**, 675–687 (2003).
16. Huang, Y., Gattoni, R., Stevenin, J. & Steitz, J.A. *Mol. Cell* **11**, 837–843 (2003).
17. Hirose, T., Shu, M.D. & Steitz, J.A. *Mol. Cell* **12**, 113–123 (2003).
18. Nguyen, V.T., Kiss, T., Michels, A.A. & Bensaude, O. *Nature* **414**, 322–325 (2001).
19. Kwek, K.Y. *et al. Nat. Struct. Biol.* **9**, 800–805 (2002).
20. Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E. & Kutay, U. *Science* **303**, 95–98 (2004).
21. Marcello, A. *et al. EMBO J.* **22**, 2156–2166 (2003).

© 2004 Nature Publishing Group http://www.nature.com/natstructmolbiol