

# Nucleus and gene expression Multiprotein complexes, mechanistic connections and nuclear organization

Editorial overview

Gideon Dreyfuss\* and Kevin Struhl†

## Addresses

\*Howard Hughes Medical Institute, University of Pennsylvania School of Medicine, 328 Clinical Research Building, 415 Curie Boulevard, Philadelphia, PA 19104-6148, USA

†Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston MA 02115-5370, USA; e-mail: kevin@hms.harvard.edu

**Current Opinion in Cell Biology** 1999, 11:303–306

<http://biomednet.com/elecreff/0955067401100303>

© Elsevier Science Ltd ISSN 0955-0674

## Abbreviations

<b>HMG protein</b>	high mobility group protein
<b>NPC</b>	nuclear pore complex
<b>RAG</b>	recombinase activating gene
<b>RBD</b>	RNA-binding domain
<b>RNP</b>	ribonucleoprotein
<b>SR protein</b>	serine-arginine protein
<b>TAF</b>	TATA-box binding protein associated factor

One of the defining features of eukaryotic organisms is the nucleus, the cellular compartment that houses the genetic material. Consequently, all molecular processes involving DNA — for example replication, recombination and transcription — occur in a distinct subcellular location. In addition, the nucleus is the site of action for other steps in gene expression, such as RNA processing, that occur co-transcriptionally or post-transcriptionally and do not involve DNA. Indeed, there is increasing evidence that transcription and RNA processing are mechanistically coupled, a feature that is permitted by the location of both processes within a common cellular compartment. The compartmentalization of these biological events clearly indicates that specific proteins and RNAs must be imported into and exported from the nucleus. Such nuclear-cytoplasmic trafficking occurs through structures termed nuclear pore complexes (NPCs) and is mediated by a family of soluble transporters. As all of these processes are regulated by developmental and environmental signals, the nucleus has a central role in orchestrating the enormous diversity and complexity of eukaryotic organisms.

This issue of *Current Opinion in Cell Biology* contains 16 reviews on a variety of biological processes that occur in the nucleus. Despite the differences among these processes, there are several common themes. First, nuclear DNA is highly compacted by histones and other proteins into a discrete chromatin structure. This compaction occurs at several levels and it is essential in order for the full complement of genomic DNA to fit in the nucleus. A fundamental principle that has become clear over the past few years is that

chromatin does not play a passive and repressive role in transcription and other DNA-dependent events. Instead, static and dynamic aspects of chromatin structure are integral to these processes and are mediated in part by the specific recruitment of chromatin-modifying activities that create active or inactive regions of chromatin. Second, extremely large multiprotein complexes that function as molecular machines carry out most of the biochemical processes. For example, the RNA polymerase II transcription machinery and the mRNA splicing apparatus are both comparable in size and subunit complexity to ribosomes. Aside from the challenges in identifying all the components of these molecular machines, it will be important to determine how each of these components function in precise biochemical terms and how they contribute to the exquisite regulation and high fidelity of gene expression. Third, complex biological processes that have been historically considered as separate problems (e.g. transcription and RNA processing) are mechanistically coupled, presumably to facilitate a seamless passage through the various steps involved in gene expression. Fourth, the nucleus is not simply a liquid environment contained within a membranous envelope that separates the genome from other cellular components but rather contains distinct functional substructures that have important biological roles. Fifth, communication between the nucleus and cytoplasm — a process that involves transport of macromolecules between these two compartments — is often a key step in gene regulation.

## Chromatin structure

In nuclei, DNA is compacted at several levels, leading to a hierarchy of structures starting with nucleosomes, proceeding through 10 nm fibers, 30 nm fibers, and chromosome loops, and ending in fully condensed metaphase chromosomes. Belmont *et al.* (pp 307–311) discuss recent results in living cells that identify higher order folding (above the level of the 30 nm fiber) within interphase chromosomes. Structural models for these chromatin structures are discussed, as are the implications for genetic effects such as X-chromosome inactivation, dosage compensation, locus control regions, and boundary elements.

Most of our current knowledge about chromatin structure is at the level of nucleosomes. Although the structure of the histone octamer has been solved at atomic resolution, the location of the linker histone H1, which stabilizes the nucleosome, has been less clear. Thomas (pp 312–317) reviews recent studies leading to the unexpected result that histone H1 is asymmetrically disposed near the dyad axis, suggesting a polar arrangement along nucleosome

arrays. These results have implications for how transcription factors and H1 might compete for DNA in the linker regions between nucleosomes. The physiological role of histone H1 remains unclear, as yeast and *Tetrahymena* H1-like proteins are not required for cell growth, although they confer selective effects on gene expression.

### Maintenance and modification of the eukaryotic genome

Cells have evolved a variety of complex biochemical processes to ensure the faithful transmission of genetic information. Bryan and Cech (pp 318–324) focus on one such process, the replication of chromosome ends, which is mediated by telomerase, an enzyme composed of an essential RNA molecule and a protein that contains motifs common to reverse transcriptases. In addition, other telomeric proteins have been identified, including Ku, which has a dual function in protecting chromosome ends and in mediating DNA repair. Telomeric shortening occurs as cultured cells are continuously passaged and this shortening has been implicated in cellular senescence, particularly because overexpression of telomerase extends the life span of cultured cells. The role of telomerase in the aging of organisms (as opposed to cells) remains to be clarified.

Genomic stability is a hallmark of healthy organisms and the loss of such stability is associated with cancer and other diseases. Nevertheless, developmentally programmed and site-specific genomic rearrangement is the key mechanism for creating the enormous diversity of antibodies and T-cell receptors that are necessary for a functional immune system. As reviewed by Oettinger (pp 325–329) there is increasing knowledge about the RAG (recombination activating gene) proteins, which mediate the DNA recognition and cleavage events of V(D)J recombination. The biochemical mechanism is related to that of bacterial transposases but it has a number of novel features that are essential for the diversity and specificity of the recombination events. Much less is known about how V(D)J recombination events occur in the proper temporal order and under the correct developmental restrictions, although it is presumed that localized changes in chromatin structure are responsible.

### Transcriptional regulation

Berk (pp 330–335) reviews the advances and controversies about a central issue in gene expression, namely how activators stimulate transcriptional initiation by RNA polymerase II (pol II). Although the basic pol II machinery contains approximately 60 proteins — a number of which can interact with activators *in vitro* — multiprotein 'co-activator' complexes are also required for activation. There is increasing evidence that the pol II holoenzyme, which consists of the core enzyme, general transcription factors, and the Srb-containing mediator, associates as a single entity at promoters. A large number of poorly defined proteins and factors has now 'coalesced' into a smaller number of multiprotein complexes with known components, and

genetic and biochemical analyses have identified functionally distinct subcomplexes. Recent studies, however, have indicated an additional level of complexity in that individual proteins — for example TAFs (TATA-box binding protein associated factors) — can be components of completely different multiprotein complexes, such as TFIID and the SAGA/PCAF histone acetylase complexes. Although Berk argues that the TAF subunits of TFIID are generally required for transcription *in vivo*, this issue remains controversial.

A fundamental principle that has emerged over the past few years is that transcriptional regulation and chromatin structure are mechanistically intertwined. As reviewed by Berger (pp 336–341), there are a number of distinct histone acetylase and deacetylase complexes that have gene-specific effects on transcription. DNA-binding activators and repressors can recruit such histone acetylases and deacetylases to specific promoters whereupon they generate local disruptions in chromatin structure. Although recruitment of Gcn5 histone acetylase or Rpd3 histone deacetylase complexes causes a highly localized change in chromatin structure, it is likely that recruitment of other chromatin modifying complexes will be responsible for the creation of large regions of altered chromatin structure that are associated with long-range effects on gene activity. Unexpectedly, some histone acetylase complexes can also acetylate transcriptional activators (e.g. p53) or architectural DNA-binding proteins (e.g. HMG1), thus providing an additional level of complexity for gene transcription.

The switch from transcriptional initiation to transcriptional elongation is associated with the phosphorylation of the carboxy-terminal tail of pol II. The review of Reines, Conaway and Conaway (pp 342–346) indicates that transcriptional elongation is a complex process. A number of elongation factors have been characterized and these can function by directly suppressing the transient pausing or arrest of pol II at specific sites in the RNA or by specifically facilitating the passage of pol II through chromatin templates. The HIV Tat protein regulates transcription by interacting with sequences in the mRNA transcript and stimulating elongation in a manner that appears to require P-TEFb, a protein kinase that phosphorylates the carboxy-terminal domain (CTD) of pol II. There is also considerable evidence linking transcriptional elongation to DNA repair, somatic hypermutation, homologous recombination, and maintenance of genome stability, although the molecular mechanisms remain to be elucidated.

### Connection between transcriptional initiation and pre-mRNA processing

The fact that pol II transcripts undergo a series of common processing events, most of which take place on the nascent transcripts, has long suggested a connection between transcriptional initiation and mRNA processing. As discussed by Bentley (pp 347–351), the critical connection occurs when the CTD of pol II is phosphorylated and the enzyme

shifts from the initiation to elongation mode. The phosphorylated pol II tail interacts directly with factors that mediate mRNA capping, 3'-end processing and perhaps splicing, thereby recruiting the various RNA-processing machineries to the transcription apparatus and the RNA transcript. Another possible connection is that cleavage polyadenylation specificity factor (CPSF) contains certain TAFs as subunits. These results, and ultrastructural studies indicating colocalization of newly synthesized RNA, Pol II and a variety of mRNA processing factors, have strongly argued for the idea of an 'mRNA factory' in which synthesis and processing of mRNA are integrated. Consequently, events occurring during transcriptional initiation can have regulatory consequences at the level of mRNA processing and hence affect the final products of gene expression.

In addition to splicing, most pre-mRNAs are also processed at their 3'-ends by a coupled process of cleavage and polyadenylation. This process depends on specific signals in the pre-mRNA and is mediated by a large complex of proteins, most of which have been identified and characterized over the last few years. Minvielle-Sebastia and Keller (pp 352–357) review current knowledge in this field and highlight the emerging theme of coupling of 3'-end processing to the upstream processes of splicing and transcription.

### **Pre-mRNA splicing: components and nuclear organization**

Specific sequences in pre-mRNA define the introns and exons. These sequences are recognized by a large array of *trans*-acting factors — comprised of snRNPs (small nuclear ribonucleoproteins) and protein factors — which serve in the formation of the reaction centers, the spliceosomes, on each spliced intron. Many of the protein factors have a common structural motif and they constitute a protein family termed the SR proteins. These nuclear factors have one or more RNA-binding domains and a domain rich in arginine-serine dipeptides (hence the term SR proteins). The SR domain serves largely in protein-protein interactions that are critical for splicing and in the transport and localization of these proteins. The SR proteins are the subject of the review by Tacke and Manley (pp 358–362) who underscore the RNA-binding specificity of SR proteins and discuss their roles in both constitutive and alternative splicing.

Throughout their lifetime in the nucleus, pol II transcripts are associated with heterogeneous nuclear ribonucleoproteins (hnRNPs) proteins, a group of abundant RNA-binding proteins. Over the years, these proteins have turned out to have a surprisingly wide range of functions and be involved in almost every aspect of pre-mRNA processing as well as in mRNA transport and in translation. Krecic and Swanson (pp 363–371) outline recent developments on hnRNP proteins and present a thoughtful framework with which to consider their myriad activities. Clearly, pre-mRNAs and mRNAs must be thought of in the context of RNP complexes in which both the RNA and the proteins are critical to the fate and function of the transcription product.

Interestingly, the relative amounts of the RNA-binding protein cousins, the SR proteins and hnRNP proteins, can profoundly affect alternative splicing patterns. Therefore the tissue-specific expression of these two groups of proteins and the modulation of the distribution of proteins between the nucleus and the cytoplasm are important and largely unexplored determinants of gene expression.

In much the same way as for pre-mRNAs, small nuclear RNAs (snRNAs) also exist as RNP complexes. SnRNPs are comprised of uridine-rich snRNAs and a specific collection of RNA-binding proteins, the snRNP proteins. The snRNPs, in concert with protein splicing factors, interact with the hnRNP complexes and carry out the pre-mRNA processing reactions. But these reactions — most commonly investigated by biochemical approaches *in vitro* — take place in the nucleus and to think about them without considering the organization of these components in the nucleus is to look at only one half of the story. Sleeman and Lamond (pp 372–377) bring the organization of the splicing factors in the nucleus into the picture and discuss its relationship to their function. Although most of the pre-mRNA processing reactions occur on the chromatin-associated nascent transcripts, in most cells the splicing factors themselves are not only found to be associated with these transcripts but are also particularly concentrated in several distinct nuclear structures. These highly concentrated localized accumulations are dynamic compartments whose functions are still largely unknown yet they have captured the interest of many molecular and cell biologists. The study of these structures, an area of research rife with controversy, will undoubtedly continue to be exciting and it promises to yield important insights into how the post-transcriptional portion of the pathway of gene expression operates in the cell.

### **The nucleolus, pre-rRNA processing and ribosome biogenesis**

Anybody who has ever looked at the nucleus even with the simplest microscope, as Schleiden and Schwann did in the 1830s, cannot help but notice the most prominent morphological entity in it, the nucleolus. The nucleolus is the site of ribosomal RNA transcription and processing and of ribosome assembly. Pre-rRNAs are processed by a complex series of reactions that involve removal of intervening sequences and extensive modifications (e.g. ribose methylation and pseudouridylation) of specific regions of rRNAs. As is the case for mRNA biogenesis, these reactions are mediated by protein factors and numerous small nucleolar RNAs (snoRNAs) which, in the form of RNP complexes (snoRNPs), base pair with rRNA sequence. A plethora of fascinating recent observations, reviewed by Weinstein and Steitz (pp 378–384), reveal how snoRNAs are produced from precursor molecules, how they are targeted to the nucleolus, and how they guide the critical modifications of the rRNA. Along the way, the unexpected discovery was made that snoRNPs can also act to modify spliceosomal snRNAs such as U6. This raises the possibility that the many snoRNPs

may modify yet other RNAs in the cell and it should also make us consider seriously the possibility that the nucleolus may be a site of more than just ribosome biogenesis.

To serve in ribosome biogenesis the ribosomal proteins, which like all proteins are synthesized in the cytoplasm, must be imported into the nucleus and further localized to the nucleolus. Scheer and Hock (pp 385–390) summarize current knowledge on the structure of the nucleolus and pull together many recent observations on the localization of proteins and snoRNPs to this nuclear body. Here again, tantalizing hints that the nucleolus may have functions in addition to ribosome biogenesis emerge. At the very least though, the nucleolus is where ribosomal subunits are manufactured and from here they need to be exported to the cytoplasm. These are large macromolecular assemblies and the specific factors and requirements involved in their export are not yet known.

### **Trafficking of macromolecules between the nucleus and the cytoplasm**

All macromolecular trafficking takes place through gateways that perforate the nuclear envelope called nuclear pore complexes (NPCs). The NPCs, which are likely to be all of the same composition and probably not individually specialized for the transport of specific cargoes, are very large and complex structures (30 times the size of a ribosome). Much progress has been made in identifying the protein components of the NPC in divergent organisms from yeast to man. Though the cataloging of this immense inventory still has a long way to go before it is complete (for metazoans), several themes regarding the characteristics of the individual protein components (called

nucleoporins) have already become apparent. Aebi and co-authors (pp 391–401) describe recent progress in elucidating the structure of the NPC and the arrangement of specific nucleoporins within it. Some intriguing relationships between the form of the NPC and its function in transport have begun to reveal themselves although how this organized tunnel actually works remains to be learned.

The transport process, either import into the nucleus or export out of the nucleus, is initiated by the interaction of soluble receptors with specific targeting signals on the cargo. It is these soluble receptors that then dock at the mouth of the NPC and deliver the transport cargo. One of the most important developments over the past few years has been the discovery of multiple signals and multiple pathways for both nuclear import and export. Another key issue relates to the regulation of directionality of the transport process and compartment identity. This is now believed to be determined largely by the small GTPase Ran, which, with auxiliary factors, maintains Ran in a GTP bound form in the nucleus and a GDP bound form in the cytoplasm. Adam's review (pp 402–406) highlights these developments as well as the many critical questions that remain to be answered. Among them is the issue of the source of energy for the translocation process and the molecular mechanism of the movement of the cargo in the NPC. Importantly, for large RNA molecules such as mRNA — which is translocated as a RNP — it appears that the actual signals for export are on the proteins, not the mRNA itself. And these proteins, mostly hnRNP proteins and SR factors, have already shepherded the transcript from its birthplace, once again emphasizing the integration and co-operation of the machineries that function in gene expression.