Dynamics of transcription and mRNA export
Xavier Darzacq1,2, Robert H Singer1 and Yaron Shav-Tal1

Understanding the different molecular mechanisms responsible for gene expression has been a central interest of molecular biologists for several decades. Transcription, the initial step of gene expression, consists of converting the genetic code into a dynamic messenger RNA that will specify a required cellular function following translocation to the cytoplasm and translation. We now possess an in-depth understanding of the mechanism and regulations of transcription. By contrast, an understanding of the dynamics of an individual gene’s expression in real time is just beginning to emerge following recent technological developments.

Addresses
1 Departments of Anatomy and Structural Biology and Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA
2 UMR 8541 CNRS, Ecole Normale Supérieure, 75230 Paris Cedex 05, France

Corresponding author: Singer, Robert H (rhsinger@aecom.yu.edu)

Introduction
Molecular events governing gene expression are among the best-studied and best-understood mechanisms in biology. In many cases, our understanding of these processes has reached the atomic level, especially with the unraveling of the atomic order within the nuclear pore complex [2]. This high-resolution picture of the transcription and export machineries contrasts with our lack of knowledge regarding the time scales of these reactions within live cells. While mechanistic approaches shed light on how these events are regulated, the consequences of this regulation for cell fate can only be understood by integrating these mechanisms into kinetic pathways, allowing an understanding of the inherent plasticity within living cells. In this review, we focus on the recent technological and conceptual advances that provide a better understanding of the kinetic rules that govern the molecular processes of mRNA transcription and export.

Dynamics of transcription initiation: few factors for many genes
Gene promoters can be viewed as static binding elements on which transcription factors assemble. It is the combinatorial variety of transcription factors in a cell that will presumably modulate the transcriptional activity of a specific gene. Recent approaches in which immunoprecipitation of chromatin using antibodies against transcription factors is followed by analysis of the bound sequences by DNA microarrays (‘ChIP-chip’ methods) have identified factor-specific yeast and mammalian promoter sequences [3–5]. A complex network of interactions was revealed by a genome-wide ChIP-chip analysis of S. cerevisiae that identified 106 transcription factors bound to 2343 promoter sequences [6]. In a recent study, 142 transcription factors interacting with 3420 target genes were linked to the expression of the genes they regulate, uncovering the underestimated dynamics of the ‘transcription factor network’ [7]. One of the findings was that out of the 70 transcription factors that bind to transcription units during the cell cycle, 56 were also regulated by the cell cycle. Synthesis and binding of transcription factors could occur in the same phase of the cycle (as was the case for 13/56 genes) or also bind to genes expressed in subsequent phases (24/56 genes), providing a trigger to move to the next phase. The lag between transcription factor binding and gene expression, although highly reproducible, varied greatly from gene to gene and did not seem to depend on specific transcription factors but rather on the target gene [7]. Although these findings correlating the binding of transcription factors to their expression can provide crucial information on the wiring of the transcriptional network, it is still unclear how the time lag separating transcription factor binding from gene expression is controlled at the molecular level. So far, the complexity of the mammalian genome has not allowed genome-wide ChIP-CHIP analysis [3,8]. However, new approaches where chromatin immunoprecipitation is analyzed by tandem cloning, known as serial analysis of chromosome occupancy (SACO), allow the genome-wide quantitative interrogation of the binding of specific transcription factors [9]. One way to extend this analysis would then be to investigate the correlation between transcription factor binding and polymerase loading of a gene. Technologies to address these questions now exist. For example, a time-course ChIP approach has shown that the binding of the estrogen receptor to its promoter is cyclical, defining permissive and non-permissive states for gene activation [10–12]. These biochemical approaches can only address a population of cells at a fixed time determined by chemical fixation procedures, and often require cell...
transcriptional process will modulate the output. Expression, the many control points downstream in the
action of transcription elongation factors. Genes controlled at the promoter escape stage, for example heat
shock genes in Drosophila, can react rapidly to a stimulus since processive elongation-competent polymerases are
available [25,26]. Paused polymerases have been detected in the proximal regions of genes including c-Myc [27]. Recently, the activity of TFIIH on c-Myc transcription was investigated using cell lines deficient in one of TFIIH's components, XPB. While c-Myc protein levels were only moderately affected in the population, cell-to-cell variance in protein level increased by a factor of three. Although the function of TFIIH is not completely understood, it plays an important role in the initial stages of transcription that occur before promoter clearance. Mathematical modeling showed that TFIIH might function as an integrator of promoter firing events, therefore establishing stable expression levels of c-Myc from a stochastic input [28*].

Dynamics of elongation
Single bacterial polymerases tethered to a fixed substrate and used to drag a DNA-conjugated bead enabled the
determination of elongation speeds for single polymerases [29,30], and demonstrated that RNA polymerases have an intrinsic propensity to pause. The conservation between bacterial core polymerases and mammalian RNA polymerase II makes it likely that pausing is a property of the mammalian enzyme as well. It is difficult to address the synthesis speed of an mRNA as the kinetics will be determined by the limiting step in the transcriptional process (initiation, promoter escape or elongation). The human dystrophin gene (2.3 Mb) provides a model where initiation and promoter escape rates are minimal relative to elongation time. The measured transcription rates for this gene were 1.7–2.5 kb/min [31]. Other approaches using FISH to detect nascent pre-mRNAs at their locus of transcription following activation have found elongation speeds of 1.1–1.4 kb/min for the Drosophila UBX mRNA [32] and mammalian β-actin mRNA [33]. Similar rates were observed using nuclear run-on assays on the HSP70 gene following heat shock in insect cells [25].

The development of genetically encoded fluorescent markers such as GFP (green fluorescent protein), together with sophisticated imaging, have allowed kinetic measurements of many biochemical processes in live cells, including gene expression [34]. rRNA transcription is confined to the nucleolus, providing a system for studying a homogenous population of co-regulated genes. The kinetics of pol I transcription in live mammalian cells were calculated using FRAP (fluorescence recovery after photobleaching) and iFRAP (inverse FRAP) experiments. The kinetics of recruitment of this enzyme to the rDNA genes were correlated with the time it takes to transcribe one gene. This approach allowed the dissociation of the different steps of transcription (polymerase

Dynamics of post-initiation polymerases
Following assembly of factors on the promoter, the polymerase initiates transcription and then elongates the nascent pre-mRNA. The switch from the post-initiation stage to elongation (termed promoter clearance and promoter escape, respectively) is believed to be a limiting step leading to high levels of aborted short RNA molecules [23,24]. The instability of the polymerase on its DNA substrate results in aborted promoter clearance and it is unclear how different transcription factors can influence its processivity. Promoter usage could provide a positive regulation loop selecting for a conformation more favorable to polymerase progression. A more active control operates at the level of promoter escape, where the polymerase enters a paused state under the control of factors that inhibit elongation. Release from pausing is, at least partly, triggered by a combination of maturation events including capping of the nascent pre-mRNA and

association with the promoter and elongation), and the experimental measurements were compared to simulations based on mathematical kinetic models and provided an in vivo kinetic model for Pol I transcription [35,36**].

Pol II transcription was addressed in live cells by complementing a non-functional endogenous copy of the large subunit with a GFP-tagged functional version [37]. Pol II recruitment to nuclear transcription units comprised two detectable components, a fast component presenting a half life in the order of several seconds that was attributed to Pol II diffusion and binding to the promoter, and a slow component showing a half life of 14 to 20 min, resulting from engaged polymerases. The average transcription unit length in mammalian cells (~14 kb) and the previously reported polymerase speeds (1.1–2.5 kb/min) would predict an average elongation time in the order of 6–13 min, significantly shorter than that observed [38**]. This discrepancy could be explained if the polymerases were engaged on the DNA longer than the elongation process, raising the possibility that promoter escape, termination or pausing could add a rate-limiting step within the transcription unit. Ultimately, kinetic analyses at the single gene level will resolve these issues.

To follow the activity of specific genes in single living cells, cell lines have been generated where a particular gene is integrated as multiple tandem repeats in the genome (Figure 1) and can be visualized either using the lac operator repeat system (lacO/LacI) [39–41] or by the recruitment of fluorescent chimeric transcription factors to higher levels than the nucleoplasmic background [15–18]. Using this technology, a GFP-RNA-pol-II subunit was used to monitor the transcriptional activity of an MMTV tandem gene array: it was found that its activation was transient even in the constant presence of dexamethasone, showing a maximal activity 30 min post-activation [17]. Observing the loading of the RNA polymerases on the tandem gene array over time revealed variability among individual cells. The lacO/LacI system combined with a live-cell approach to RNA visualization [42,43] has provided information on the dynamics of chromatin at the locus of transcriptional activity and on the temporal resolution of gene activation (Figure 2) [44**]. Applying such techniques to the study of transcription kinetics within a gene will undoubtedly lead to a more precise understanding of this process.

Dynamics of mRNA within the nuclear environment

The mechanisms that regulate RNA flow from the transcription site are not well understood. It is now clear that mRNAs traverse the nucleoplasm by a diffusion-based mechanism. Measurements of nuclear RNAs labeled either with fluorescent probes [45–47] or with specific GFP-fused protein tags [48**,49] have yielded a range of diffusion coefficients. Single-particle tracking of messenger ribonucleoprotein particles (mRNPs) in the nuclei of living cells allowed the detection of random movements [48**]. This motion was energy-independent and was not directed, although correlated motion implied the presence of intra-nuclear structures hindering the movements; mRNPs could be seen ‘bouncing off’ nucleoli, suggesting the presence of chromatin domains inaccessible to mRNPs [50,51]. Saturation of the export machinery by RNA overexpression revealed a reticular network throughout the nucleus space [52], corroborating previous studies showing mRNA movement within inter-chromatin channels [46]. A simplistic view of chromatin domains separated from inter-chromatin channels has been complicated by the fact that large macromolecule complexes can move through condensed chromatin regions [53*]. Chromatin regions were completely accessible to fluorescent dextrins with sizes of 3–10 kDa, but 70-kDa dextrins could not penetrate highly condensed areas. The radius of gyration of these inert molecules was calculated and found to be 4, 6 or 10 nm, values which correspond to either spherical proteins with molecular weights of approximately 400, 1400 and 6600 kDa or to ellipsoid proteins of 55, 180 and 850 kDa, respectively. Therefore, large macromolecule complexes such as RNA polymerases could gain access to most chromatin areas, whereas larger complexes in the megadalton range, such as RNP complexes (~50 nm diameter [54]), would appear to be excluded.

Gene positioning and mRNA export

An efficient and rapid path of mRNA export could result if active genes were associated with nuclear pores. Although studies have indicated that the nuclear periphery is actually involved in gene silencing, a screen operated on yeast has shown that classes of highly active genes, for instance genes involved in glycolysis or protein biosynthesis, are preferentially associated with nucleoporins, karyopherins and nuclear-pore-associated proteins [55**]. Moreover, transcriptional induction of the GAL genes caused their relocation towards the nuclear envelope, suggesting a mechanism for ‘express shipping’ of highly required transcripts. In another study, an active gene region was tethered to the nuclear pore by export proteins, including the transportin Cse1 [56]. Live imaging of GFP-tagged lac repressor protein (LacI) fused to the C-terminus of Cse1 showed not only the preferential relocation of a lacO locus to the nuclear periphery, but the occasional penetration of these chromosomal domains into the cytoplasm. The proximity of genes to the pore might be different in mammalian cells where the nucleus is much larger and chromosome organization more complex. Inactive genes in mammalian cells have been found to be associated with heterochromatin and the nuclear periphery, whereas active genes were sometimes interiorly positioned [57]. Yet the nuclear distribution of a specific mRNA was not influenced by the relative position
Tandem gene arrays for the analysis of nuclear factor dynamics in living cells. The establishment of integrated tandem gene arrays in the genome provides a platform on which a signal from a fluorescently labeled DNA- or RNA-binding protein can be amplified, detected and analyzed kinetically. (a) The locus of integration occurs randomly in the genome and (b) many copies of the gene are inserted in tandem at this locus. (c) Different types of gene arrays have been used. (1) lac operator repeats (lacO) are bound by the lac repressor protein (LacI), which can either be tagged with a fluorescent protein to mark the locus of integration or can be fused to a protein of interest, thus tethering it to the chromatin. (2) Tandem arrays of promoters can serve for the analysis of transcription factor (TF) binding dynamics. (3) The production of mRNA can be followed using gene arrays that transcribe mRNAs containing MS2 repeats, which are bound by the YFP-MS2 protein. The kinetics of RNA pol II or RNA processing factors can also be analyzed on such arrays. (d) Photobleaching methods such as FRAP are used to measure transcription-factor binding to gene arrays. (e) The onset of transcription can be measured upon cellular stimulation (red line shows a classical activation with saturation; orange line shows activation with a negative feedback).
of the transcription site in comparison to the nuclear envelope, implying that transcriptional proximity to nuclear pores might have more subtle or indirect effects [48**]. Genes distally separated by large DNA regions on the same chromosome had a high probability of sharing the same nuclear transcription space, demonstrating a high degree of spatial linkage within chromosome territories [58*], and other studies have shown that genes adopt preferential locations during differentiation [59–61]. An analysis of the travels of endogenous mRNAs with respect to their distance from the nuclear envelope may reveal whether gene positioning can influence mRNA export in mammalian cells.

**Nucleo-cytoplasmic transit of mRNA**

Distinct RNA transport mechanisms are responsible for the export of different species of nuclear RNAs (mRNAs, spliceosomal U snRNAs and tRNAs). This implies the existence of mechanisms that are able to distinguish between RNA species while in transit. For example, mRNAs and spliceosomal U snRNAs are both produced by RNA polymerase II transcription, yet are preferentially exported by different pathways. Interestingly, RNA length was found to be a parameter measured by the export machinery [62,63]. If shortened to <120 nt, mRNAs were exported by the U snRNA pathway, whereas U1 snRNA with a 300 nt insertion behaved like an mRNA. It remains to be defined whether the commitment to a certain pathway occurs during the formation of an mRNP at the transcription site or during its travels to the nuclear pore [64]. If the mRNP is indeed remodeled during transit, it would be of interest to see whether this occurs in a specific nuclear compartment.

The time necessary for transport of an mRNA molecule from its transcription site to the cytoplasm is still an open question due to the lack of suitable living cell systems for measuring the rates of transport of endogenous RNAs. The study of mRNA export in purified nuclei from *Xenopus* oocytes has yielded export times ranging from 10 min to 1 h. These biochemical studies depend on viral RNA systems or on microinjection of mRNAs that have been transcribed in vitro. The rabbit β-globin gene expressed under the control of a doxycycline-controlled inducible promoter allowed the kinetic analysis of mRNP assembly and transport following induction [65]. This approach yielded nuclear residency half-lives of 2.5–4.4 min, which obeyed first-order kinetics. Kinetic modeling of these data showed that the nuclear pool of mRNA built up slowly, whereas cytoplasmic accumulation was

### Table: CFP–Lac repressor, CFP–peroxisomes, YFP fusion protein, Merge

<table>
<thead>
<tr>
<th></th>
<th>CFP–lac repressor</th>
<th>CFP–peroxisomes</th>
<th>YFP fusion protein</th>
<th>Merge</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS2 (mRNA)</td>
<td><img src="image" alt="a" /></td>
<td><img src="image" alt="b" /></td>
<td><img src="image" alt="c" /></td>
<td><img src="image" alt="g" /></td>
</tr>
<tr>
<td>RNA pol II</td>
<td><img src="image" alt="d" /></td>
<td><img src="image" alt="e" /></td>
<td><img src="image" alt="f" /></td>
<td><img src="image" alt="h" /></td>
</tr>
<tr>
<td>SF2/ASF</td>
<td><img src="image" alt="c" /></td>
<td><img src="image" alt="i" /></td>
<td><img src="image" alt="i" /></td>
<td><img src="image" alt="i" /></td>
</tr>
</tbody>
</table>

Visualizing the binding of nuclear factors to tandem gene arrays. Human cells containing a tandem gene array (200 copies) were used for real-time detection of the recruitment of different factors to the site of active transcription. The tetracycline-inducible gene module gives rise to an intron-containing mRNA coding for a peroxisome-targeted CFP (cyan fluorescent protein) and containing 24 MS2 repeats in the 3'UTR. Each gene module is flanked by 256 repeats of the lacO binding site. (a,b,c) After induction of transcription, the gene locus is detected by the binding of the CFP–LacI protein to the lacO repeats and the peroxisome-targeted CFP protein product is observed in the cytoplasm. (d) The production of the mRNA at the gene locus is visualized by the binding of the yellow fluorescent protein (YFP)–MS2 protein to the MS2 stem-loops in the mRNA. (e) The recruitment of the transcription machinery to this active site of transcription is seen by the enrichment of YFP–RNA pol II at the locus. (f) The pre-mRNA splicing factor YFP–SF2/ASF is also enriched at the locus. (g,h,i) Merge. (Adapted from Janicki et al. [44**]).
exponential. Differences in the nuclear dwell-times of the mRNAs were not observed if actin polymerization was inhibited or if cells were plated at low or high densities. The availability of new methods to label endogenous mRNAs [66] and the development of live cell systems for the study of mRNA transport, combined with kinetic modeling approaches, have been successfully applied to nuclear protein transport [67-69], and will provide new insights into the process of mRNA export.

Conclusions

mRNA transcription and export constitute a chain of molecular events offering many points of control. Although there are many combinations of promoter–transcription-factor associations, the downstream cellular responses cannot be explained by promoter firing alone. To understand the subtleties of these transcriptional pathways, single cell approaches are necessary. Recent developments in our ability to probe single cells in real time have yielded new information on the dynamics of gene expression. These studies will ultimately take us to the complex task of unraveling the dynamics of transcription within the live organism.

Acknowledgements

We thank Olivier Bensuade and Claire Dugast-Darzacq for critical reading of the manuscript. This work was supported by grants from the NIH EB 2060, DOE ER63056 and P20 EB4930 to RHS.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

• of special interest
**•• of outstanding interest


This work reveals the complexity of the transcriptional regulatory network and provides a genome-wide view of gene regulation in yeast.


The transcription network [4] is integrated with available expression data. The transcription factor expression patterns associated with different cellular treatments or with different phases of the cell cycle are used to define ‘active transcription networks’. This work shows the high dynamism and plasticity of this network.


Transcriptional variation observed at the single-cell level does not lead to protein variations in cell cultures. It has been proposed that the stochastic initiation of transcription onset are balanced by the regulations that apply to post-initiation events in the chain of gene expression [21,22]. This paper illustrates this theory by giving an example where post-initiation transcriptional regulation of the c-Myc maintains constant c-Myc protein content.


This study breaks new ground in its approaches to study transcription in live cells. The recruitment of the enzyme to the gene is correlated with its activity using mathematical kinetic modeling (see also [35]).


38. Kimura H, Sugaya K, Cook PR: The transcription cycle of RNA polymerase II in living cells. J Cell Biol 2002, 159:777–782. The nuclear association of the RNA pol II with chromatin is addressed by FRAP and iFRAP and the results are discussed in light of the transcriptional activity of this enzyme. This work lays the groundwork for the kinetic analysis of pol II in live cells. (See also [57] for the establishment of the cellular system.)


48. Shav-Tal Y, Darzacq X, Shenoy SM, Fusco D, Janicki SM, Spector DL, Singer RH: Dynamics of single mRNPs in nuclei of living cells. Science 2004, 304:1797–1800. Single mRNPs transcribed from a specific gene are tracked in the nucleoplasm of living mammalian cells and are found to move by a diffusion-based mechanism. Diffusion coefficients for these mRNPs are then measured by different techniques. This study shows that energy is not required for mRNP movement per se, but is essential for the integrity of the chromatin structure.


53. Verschure PJ, van der Kraan I, Manders EM, Hoogstraten D, Houtsmtuller AB, van Driel R: Condensed chromatin domains in the mammalian nucleus are accessible to large macromolecules. EMBO Rep 2003, 4:861–866. Using fluorescent dextran of various molecular sizes, this work shows that condensed chromatin domains are accessible to large macromolecules, implying that the nuclear machinery can freely enter highly ordered regions of the genome.


the same transcription factories, indicating that genes are recruited to active domains of transcription.


