Imaging gene expression in real-time using aptamers

by

Ilchung Shin

A dissertation submitted to the graduate faculty

In partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Biochemistry

Program of Study Committee:
Marit Nilsen-Hamilton, Major Professor
Alan Myers
Gustavo MacIntosh
Ted Huiatt
David Hannapel

Iowa State University

Ames, Iowa

2011

Table of Contents

ABSTRACT

Signal transduction pathways are usually activated by external stimuli and are transient. The downstream changes such as transcription of the activated genes are also transient. Real-time detection of promoter activity is useful for understanding changes in gene expression, especially during cell differentiation and in development. A simple and reliable method for viewing gene expression in real time is not yet available. Reporter proteins such as fluorescent proteins and luciferase allow for non-invasive detection of the products of gene expression in living cells. However, current reporter systems do not provide for real-time imaging of promoter activity in living cells. This is because of the long time period after transcription required for fluorescent protein synthesis and maturation. We have developed an RNA reporter system for imaging in real-time to detect changes in promoter activity as they occur. The RNA reporter uses strings of RNA aptamers that constitute IMAGEtags (Intracellular MultiAptamer GEnetic tags), which can be expressed from a promoter of choice. The tobramycin, neomycin and PDC RNA aptamers have been utilized for this system and expressed in yeast from the GAL1 promoter. The IMAGEtag RNA kinetics were quantified by RT-qPCR. In yeast precultured in raffinose containing media the GAL1 promoter responded faster than in yeast precultured in glucose containing media. IMAGEtag RNA has relatively short half-life (5.5 min) in yeast. For imaging, the yeast cells are incubated with their ligands that are labeled with fluorescent dyes. To increase signal to noise, ligands have been separately conjugated with the FRET (Förster resonance energy transfer) pairs, Cy3 and Cy5. With these constructs, the transcribed aptamers can be imaged after activation of the promoter by galactose. FRET was confirmed with three different approaches, which were sensitized emission, acceptor photobleaching and donor lifetime by FLIM (fluorescence lifetime imaging microscopy). Real-time transcription was measured by FLIM-FRET, which was detected by the decrease in donor lifetime resulting from ligand binding to IMAGEtags that were newly synthesized from the activated GAL1 promoter. The FRET signal was specific for transcribed IMAGEtags.

Chapter 1. INTRODUCTION

Dissertation organization

There are five chapters in this dissertation. Chapter 1 covers the general introduction of the research on imaging gene expression using aptamers to detect promoter activity. The introduction explains the purpose and significance of the research, questions to be addressed, and results of this study. Chapter 2 is a literature review of the mechanism and dynamics of transcriptional regulation as well as of the imaging methods that have previously been developed to detect new transcripts. Chapter 3 addresses the development of IMAGEtags (Intracellular MultiAptamer GEnetic tags) to detect promoter activity in real time in live yeast cells. Chapter 4 addresses conditions and requirements of IMAGEtags to support the Chapter 3 manuscript. *In vitro* analysis was performed before IMAGEtags were applied in the *in vivo* system. The general conclusions of this research are discussed in Chapter 5.

The manuscript is in preparation to submit to Nature Methods. The roles of the co-authors are as follows: Ilchung Shin initiated the studies, developed the yeast model system of IMAGEtags, worked mostly with tobramycin IMAGEtags, studied the real-time imaging of transcription by IMAGEtags, and wrote the first draft of the manuscript. Judhajeet Ray worked with PDC IMAGEtags and Muslum Ilgu tested neomycin IMAGEtags in yeast. Vinayak Gupta synthesized the fluorescent dye-labeled ligands and George Kraus mentored Gupta V for the synthesis. Ming Zhao and Leilei Peng from Optical Science, University of Arizona are collaborators who performed FLIM (fluorescence lifetime imaging microscopy). Marit Nilsen-Hamilton is the corresponding author, conceived the project, mentored Shin I, Ray J and Ilgu M, and supervised the work including the experimental designs and data analysis, and revised the manuscript. All coauthors contributed to editing the manuscript.

Significance of imaging gene expression at the RNA level

Transcription is the first and most highly regulated step of gene expression. It is composed of several steps. After the promoter is activated, the DNA template is melted at the transcription start site, the general transcription factors and RNA polymerase II (RNAPII) are recruited to the promoter region, and the preinitiation complex (PIC) is formed. Pre-mRNA synthesis is then initiated, and the initiation complex becomes an early elongation complex. In the early elongation complex, RNAPII pauses at the promoter proximal region in the metazoa, especially in cells from developmental stages of the organism. The mRNA is transcribed in the productive elongation phase, and capping and splicing are cotranscriptional events. When RNAPII reaches the termination signal, synthesis of the mRNA is terminated and polyadenine is added to the end of the mRNA. Finally, the pre-mRNA is matured in the nucleus and the mature mRNA is ready to export to the cytoplasm for protein expression.

Every step is regulated during transcription. The two rate-limiting steps are formation of the 1) PIC, and 2) the early elongation complex. To form the PIC, the general transcription factors (TFIIB, TFIID including TATA-binding protein, TFIIE, TFIIF, and TFIIH) and RNAPII are recruited sequentially to the promoter region. Other factors such as Mediator, various activators, histone-modifying enzymes, and ATP (adenosine triphosphate)-dependent chromatin remodeling complex are also recruited to the PIC. PIC assembly is regulated by interactions in the complex that involve structural and topological changes (Roeder 2005; Nechaev and Adelman 2011).

The transition from initiation to early elongation involves the C-terminal domain (CTD) of the RNAPII. The CTD contains a repeated heptad consensus sequence (YSPTSPS) that can be phosphorylated at several sites. The CTD is mostly not phosphorylated during initial promoter binding and this stabilizes the interaction between the CTD and PIC. The CTD of RNAPII is phosphorylated at Ser5 by the Cdk7 subunit of TFIIH, which destabilizes the interaction between RNAPII and the promoter, allowing promoter escape. In the elongation phase, Ser2 is

phosphorylated by P-TEFb (positive transcription elongation factor b) kinase, which results in highly processive transcription. Thus, phosphorylation and the specific site of CTD phosphorylation are both important for regulating the transcriptional elongation (Nechaev and Adelman 2011).

RNAPII pausing in the early elongation is regulated by several factors. After 25–50 nucleotides (nt) synthesis of mRNA, RNAPII pauses in the promoter proximal region in metazoan cells. RNAPII pausing is related to the presence of DSIF (DRB sensitivity-inducing factor) and NELF (negative elongation factor) in the early elongation complex. RNAPII is paused when DSIF is a heterodimer with NELF. P-TEFb is recruited and phosphorylates NELF, resulting in a transition to the productive elongation complex. The phosphorylated NELF is released from the complex, and RNAPII moves to productive transcriptional elongation. Factors such as NELF or P-TEFb, which interact with RNAPII, are important in determining the processivity of RNAPII.

During elongation, RNAPII must contend with chromosome structure because the DNA template is wrapped around histone octamers in the form of nucleosomes. The DNA template is released from the nucleosome during transcription as a result of changes in histone modification. Histone modifications change the net charge of the nucleosome, resulting in loosening of the inter- and intra-molecular DNA-histone interaction and dynamics of nucleosome structure. During elongation, RNAPII recruits histone acetyl transferases to acetylate histone. Acetylated histones are evicted in front of the elongation machinery and are redeposited onto the DNA behind RNAPII by histone chaperones. Newly deposited histones are methylated and deacetylated in turn by elongation factors, so that the chromatin structure is reset (Li, Carey et al. 2007). Dynamic changes of nucleosome structure during elongation are required in order that RNAPII moves through the DNA template.

Transcription is regulated in several steps of gene expression and also by epigenetic modifications. Imaging mRNA transcription in real time allows the correlation of intracellular events and visualization of the dynamics of live cells.

Several methods have been reported for visualizing intracellular mRNA. In one method oligonucleotide (ODN)-labeled fluorescent probes are hybridized with target mRNA to visualize intracellular mRNA by fluorescence. Because when a single fluorescent dye is used in the system the signal to noise ratio is too low to distinguish the bound and unbound probes, FRET (Förster resonance energy transfer) probes have been proposed. However, for visualizing mRNA dynamics, ODN probes have several problems. First, ODNs are difficult to deliver inside the cells. After delivery, the ODNs are mostly sequestered into the nucleus. To prevent ODNs from coming into the nucleus, they are modified with large molecules such as quantum dots and NeutrAvidin. These modifications affect the kinetics of modified ODN and hybridization to the target mRNA (Chen, Davydenko et al. 2010). Second, microinjection for delivery of ODNs cannot be performed on a large number of cells, so only a small number of cells can be examined with this method. In addition, microinjected ODNs are diluted along with cell division, so they cannot track mRNA dynamics for a long time (Zhang and Yu 2008). The last concern of ODNs is target accessibility. It is difficult to find a nonstructured target RNA that is not occupied by proteins (Bao, Rhee et al. 2009). Because of these problems of target accessibility and delivery to target cells the ODN probes are inadequate for imaging RNA in real time and, other than the initial reports, these methods have not been adopted for real-time imaging. Also, these methods cannot measure gene transcription activity but are intended to image the total amount of mRNA in the cell.

A method currently used to track transcription in real time involves an RNA tag that binds to a green fluorescent protein (GFP)-MS2 fusion protein. The transcribed RNA is monitored by an accumulation on the newly transcribed mRNA of the fluorescent signal from the GFP. The MS2 system is composed of two parts: 1) A DNA construct that expresses the aptamer-like MS2 RNA inserted into the 3' or 5' untranslated region (UTR) of a the encoded target RNA preceded by an inducible promoter, and 2) the MS2 protein fused with GFP and with nuclear localization signal that is expressed from a constitutively active promoter. With the MS2 system inserted in a 5' UTR, the

transcriptional initiation and dwell time at the transcription start site can be measured. By comparing the fluorescent detection time between a 5' UTR and a 3' UTR MS2 system, the transcriptional elongation rate can also be determined. The MS2 RNA elements are transcribed with the mRNA and bound by MS2-GFP, thus allowing the RNA to be visualized (Bertrand, Chartrand et al. 1998; Larson, Zenklusen et al. 2011). Although this system has enabled successful imaging of RNA transcription in living cells, MS2-GFP is constitutively expressed by the cells, so there is a high background signal in this system.

To decrease the background of the MS2-GFP system, a split GFP system was introduced. The N terminus of the GFP is linked with one RNA-binding domain, and the C terminus of the GFP is linked with the other RNA-binding domain. Two target RNAs are located close by on the target RNA transcript. After the target RNA is transcribed, two RNA-binding domains bind to two adjacent RNA elements. The complementary N and C terminal GFP domains are brought close together so that they can combine to form an intact GFP. The target RNA can be detected as complemented GFP (Bao, Rhee et al. 2009).

RNA-tags linked with GFP have drawbacks. First, the target mRNA is bound by large protein molecules that could alter the dynamics of RNA inside the cells. Second, the cells must be constructed to express the RNA with MS2 motifs and the MS2-GFP fusion proteins. To ensure that the MS2-GFP fusion proteins are not limiting for visualization they should be stably expressed in the cells. Third, oxygen is required to form mature GFP (Tsien 1998), and therefore this system cannot be used under anaerobic conditions.

Research aim and significance

To detect promoter activity, reporter proteins such as luciferase and fluorescent proteins are used for noninvasive detection of the products of gene expression in living cells. However, these reporter protein systems do not provide real-time imaging of promoter activity because it takes time

to synthesize mature proteins. In this dissertation, the development of an RNA reporter system for real-time imaging to detect changes in promoter activity is described. The RNA reporter consists of strings of RNA aptamers called IMAGEtags (Intracellular Multiaptamer GEnetic Tags) that can be expressed from a promoter of choice (Figure 1). The IMAGEtag system has been developed in a eukaryotic model yeast system in which they are expressed from an inducible GAL1 promoter. For imaging, the yeast cells are incubated with fluorescently labeled aptamer ligands. To increase signal to noise, FRET is used to obtain a signal. To this end, ligands are separately conjugated with the FRET pairs, Cy3 and Cy5. The transcribed RNA aptamers bind the fluorescently labeled ligands and are imaged by a FRET signal from bound ligands after activation of the promoter by the addition of galactose.

For an effective IMAGEtag system, there are certain requirements of the ligand and aptamer: 1) the ligands should not be toxic to target cells, 2) there should be no intracellular competitors of the aptamer-ligand interaction to decrease the signal, 3) the ligands should be able to enter and leave the cells easily, 4) the aptamers should be properly folded under an intracellular condition so they can bind with the ligands, 5) the aptamers should have high affinity for their ligands to provide a signal at low ligand concentrations, and 6) the aptamers should be short RNAs so as to minimally change the endogenous RNA and its spatial and temporal dynamics.

Here the IMAGEtags have been applied to transcriptional initiation. To measure transcriptional initiation, IMAGEtags were inserted directly after the promoter region. The FRET signal was measured from ligands bound by the transcribed IMAGEtags that were synthesized in response to activation of the GAL1 promoter by galactose.
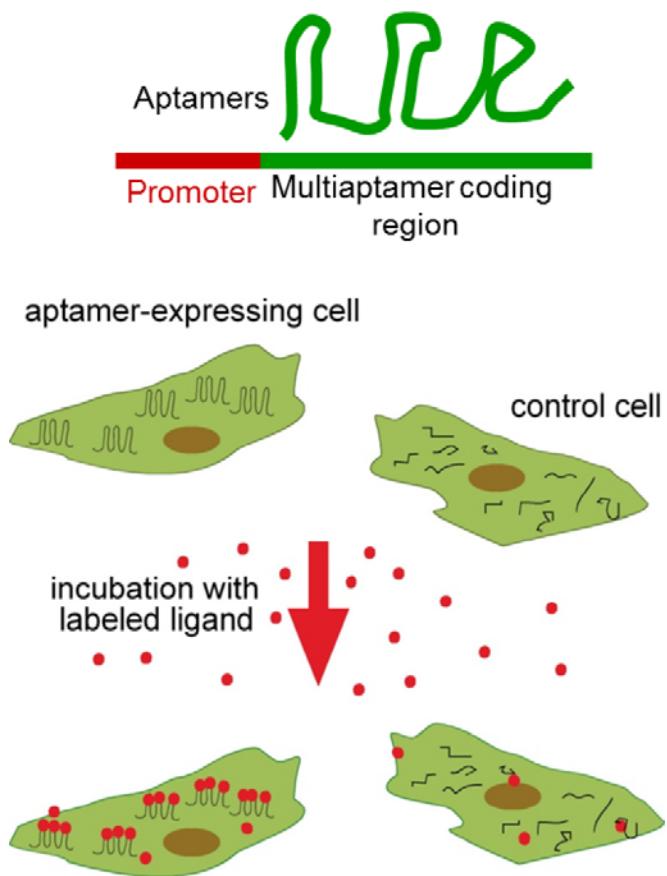
Figure1, Schematic diagram of IMAGEtags. IMAGEtags (tandemly repeated aptamers) are transcribed in the cells as parts of mRNA with an inducible promoter. IMAGEtags bind fluorescently labeled ligands and are detected by fluorescence compared to control cells. IMAGEtags visualize gene expression for promoter activity.

Chapter 2. LITERATURE REVIEW

Regulation of gene expression in eukaryotes

The regulation of gene expression from transcription to translation is a continuous process. Activation of a receptor in the membrane results in a signal that activates transcription factors (TFs). Activated TFs move to the nucleus and bind to their DNA regulatory sequences located upstream of the transcriptional start site (TSS). Other transcription factors are localized in the nucleus and activated by ligands, such as steroids, that can move across membranes. As TFs interact with other components of transcription machinery, they recruit chromatic remodeling factors to melt the DNA template at the TSS and assist in the recruitment of RNAPII to the TSS. Transcription of the protein-coding gene is catalyzed by RNAPII.

To protect the newly synthesized mRNA, the nascent RNA is modified by 5' capping that is added soon after initiation of transcription. RNAPII moves from 5' to 3' along the gene to elongate the pre-mRNA. Introns, which are long noncoding sequences, are removed by pre-mRNA splicing as cotranscriptional events. RNAPII stops transcription when it reaches the termination signal, and the newly synthesized RNA is cleaved and polyadenylated at the 3' end of the pre-mRNA. Each step in transcription is functionally and physically connected to the next as a continuous process. The transcriptional machinery plays a role in recruiting the machinery for capping and pre-mRNA splicing and promotes transcriptional elongation and efficient export of mRNA to the cytoplasm (Orphanides and Reinberg 2002).

Gene expression is also regulated epigenetically by cytosine DNA methylation, RNA-mediated gene regulation, and histone modification. DNA methylation at carbon-5 of cytosine often leads to suppression of gene expression. Regulatory regions in sense transcripts interact with regulatory RNAs in processes such as RNA interference and RNA editing. Natural antisense

transcripts have been shown to affect gene regulation through direct interaction with the sense transcripts or indirect interactions with methylated DNA or regions containing modified histone such as acetylated and methylated histone tails (Su, Xiong et al. 2010). The chromatin also controls access of the transcriptional machinery to genes to regulate gene transcription (Orphanides and Reinberg 2002).

Transcription and Regulation

Transcription is a process of synthesizing RNA from a DNA template. It is initiated either at common core promoter elements that are recognized by general transcription initiation factors or gene-specific DNA elements that are recognized by regulatory factors that activate general initiation factors. Transcription is composed of several major steps. RNAPII access to a promoter starts the transcription cycle. To access the promoter region, it is necessary to clear the nucleosomes that block access of RNAPII and general transcription factors to the promoter. A PIC forms on the core promoter. The DNA is unwound, and then the RNAPII initiates transcription. The transcription machinery escapes the core promoter and proceeds to the promoter-proximal pause region. The paused RNAPII is hyperphosphorylated on its CTD and then clears from the pausing region for either termination or productive elongation. If it is not terminated, the RNAPII productively elongates through the entire gene until it meets the cleavage site of the poly (A) signal. Transcription is terminated, and the RNAPII and transcription machinery are released and recycled to a new round of transcription. Each transcription step is regulated, and each can be the rate-limiting step.

By ChIP experiments, three types of RNAPII distribution have been observed on different genes. In the first type, the gene is without RNAPII, suggesting that it is in the off state and its association with the gene is limited by clearance of the nucleosome in the promoter region. Here, chromatin remodeling in the core promoter is believed to be the rate-limiting step for initiating transcription. In the second case, an even distribution of RNAPII is found across the gene, which is

interpreted to mean that recruitment of RNAPII and general transcription factors, namely the PIC

formation, is a rate-limiting step. The last type of distribution shows RNAPII enriched at the 5' end.

This suggests that another rate-limiting step occurs after recruitment of the RNAPII, such as at

promoter escape or during early elongation (Fuda, Ardehali et al. 2009).

RNA polymerase recruitment and initiation

Initiation of transcription requires a core promoter in DNA, several transcription factors, and

RNAPII. The core promoter elements, which function independently or synergistically, are the

TATA element located near the -30 or -25 position and a pyrimidine-rich initiator placed near the

transcription start site. General transcription factors (TFIID, TFIIA, TFIIB, TFIIE, TFIIF, and

TFIIH) as well as RNAPII are assembled at the promoter element sites to start transcription (Roeder

1996).

Transcriptional initiation is a complicated and dynamic process that involves several

transcription factors and RNAPII. Initiation requires that the transcription factors are added to the

PIC in sequential order. These series of events increase the size of the complex associated with the

DNA template. TATA-binding protein (TBP) subunit of TFIID binds first to the TATA box in the

minor groove to form a stable complex that distorts the DNA such that it bends at the TATA box.

When bound to the TATA box, the TBP produces a platform for subsequent association of other

transcription factors. TFIIA interacts with this complex through direct contact both with the TBP and

with upstream DNA sequences. The TBP-DNA interaction is also stabilized by binding to TFIIA.

TFIIB binds directly to TBP and DNA sequences both upstream and downstream of the TATA box

and stabilizes the TBP and TATA interaction. TFIIB recruits THIIF and RNAPII. The TFIIF and the

RNAPII complex bind directly to TFIIB. The complex of TBP and TAFs (TBP-associated factors)

interact with the CTD tail of RNAPII. TFIIF is recruited by RNAPII, binds to the PIC complex,

regulates promoter targeting by RNAPII and decreases nonspecific binding of RNAPII to DNA.

Binding of TFIIH with TFIIE completes the PIC assembly, stabilizes TFIIE binding, and stimulates

CTD phosphorylation. TFIIB, TFIIH, RNAPII and TFIIF subunits are involved in promoter opening,

unwinding of the DNA at the start site, and transcriptional initiation (Roeder 1996; Kostrewa, Zeller

et al. 2009). The DNA helicase subunit of TFIIH performs ATP-dependent functions in

transcriptional initiation such as formation of the open complex and promoter escape. The DNA is

unwound to form an open complex that allows single-stranded DNA to enter the RNAPII machinery

to initiate RNA synthesis (Ptashne and Gann 1997; Moreland, Tirode et al. 1999). The PIC is formed

by protein-protein and DNA-protein interactions. These dynamic interactions are involved with the

mechanism of transcriptional regulation.

Following formation of the PIC, transcription proceeds along one of at least two pathways.

The first pathway is that the initial transcribing complex undergoes abortive initiation, which results

in the polymerase repeatedly synthesizing and releasing short (2–3 nt) transcripts. More than one

short transcript is produced per template by abortive initiation, which means that they are not stably

attached in the transcription complex. With the addition of a fourth nucleotide there is a noticeable

reduction in abortive initiation. RNAPII machinery escapes from abortive initiation and is moved to

productive RNA synthesis after registering 11 nt (Holstege, Fiedler et al. 1997). Abortive initiation is

the rate-limiting step for transcription *in vitro*, and escape from abortive initiation is stimulated by the

general transcription factors (TFIIB, TFIIE, TFIIF, and TFIIH) (Nechaev and Adelman 2011). TFIIE

and TFIIH are necessary for promoter escape and synthesis of full RNAs. In their absence, the

transcription machinery produces abortive transcripts (Kugel and Goodrich 1998).

The second pathway for transcription is synthesis of the full-length RNA after promoter

escape. RNAPII moves towards promoter escape step by adding nucleoside triphosphates. As the

transcription is transformed from an open complex to an elongation complex, the first phosphodiester

bond is synthesized and the region of melted DNA moves away from the start site (Kugel and

Goodrich 1998). During the transition of the RNAPII from initiation to elongation, the basal

transcription factors TFIID, TFIIA, TFIIH, and TFIIE and Mediator (which has a role in assisting formation and/or function of the PIC) are left at the promoter site as a scaffold for the next round of initiation. RNAPII moves forward with elongation whereas TFIIB and TFIIF dissociate from the PIC during the RNAPII transition from initiation to elongation. TFIIE and TFIIH are required to melt the DNA template to allow RNAPII movement. The yeast kinase Ctk1, which has sequence homology with mammalian Cdk9, the catalytic subunit of P-TEFb, promotes dissociation of the general transcription factors in the transition from initiation to elongation. This is known because, deletion of Ctk1, results in the general transcription factors failing to release from RNAPII when RNAPII enters productive elongation (Ahn, Keogh et al. 2009). During promoter escape, the RNAPII dissociates from the initiation factors and makes other interactions with elongation factors to form the elongation complex.

Transcriptional activation of the GAL promoter

Transcriptional activation and regulation are based on the recruitment of transcription complexes to target genes. Transcriptional activators bind to specific DNA sequences and recruit co-activators, general transcription factors, and other transcriptional machinery to form a PIC complex on the promoter. In yeast, the expression of the *GAL* genes is induced by galactose and is initiated by the transcription activator Gal4p, which binds to an upstream region of the GAL1 promoter. Gal4p has two domains, the DNA-binding domain and the transcriptional activation domain. Free Gal4p can interact with and recruit transcription machinery and co-activators such as Mediator, histone modifying factors, and chromatin remodelers to the *GAL* genes. In the absence of galactose, Gal4p is regulated by binding with an inhibitor protein Gal80p that masks the transcriptional activation domain of Gal4p. In the presence of galactose, Gal80p is dissociated from Gal4p in the nucleus by an inducer protein, Gal3p, so that the activation domain of Gal4p becomes exposed. Gal3p sequesters Gal80p in cytoplasm and prevents it from entering the nucleus, so it mitigates the inhibitory effect of Gal80p on

Gal4p. Recruitment of co-activators to the *GAL* genes occurs sequentially. The SAGA (Spt-Ada-Gcn5-Acetyltransferase) complex is recruited by Gal4p, and then Mediator is assembled on the promoter by interacting with Gal4p. Finally, RNAPII and the general transcription factors are recruited to the GAL promoters. RNAPII and the general transcription factors are not recruited in the absence of the SAGA and Mediator (Jiang, Frey et al. 2009; Weake and Workman 2010).

The *GAL* gene is regulated by both negative and positive feedback loops. The galactose signal is transferred through a signaling cascade of *GAL* gene activation. First, the Gal2p, galactose permease, brings extracellular galactose into the cells. Intracellular galactose binds to Gal3p and activates Gal3p in the cytoplasm. Gal3p is exclusively located in cytoplasm, whereas Gal80p is a nucleocytoplasmic shuttle protein. The activated Gal3p binds with the Gal80p monomer in the cytoplasm, sequesters Gal80p in the cytoplasm and thereby depletes Gal80p in the nucleus. The transcriptional activator, Gal4p, located in the nucleus, is thereby released from inhibition by Gal80p and activates expression of the *GAL* genes including *GAL1*, *GAL2*, *GAL3,* and *GAL80*. Gal2p, Gal3p, and Gal80p are expressed at their basal levels in the absence of galactose. The increase of Gal2p and Gal3p concentration results in enhanced transcriptional activity. Thus, Gal2p and Gal3p constitute a positive feedback loop by increasing transcriptional activity, whereas the Gal80p effect is a negative feedback loop (Acar, Becskei et al. 2005). Gal80p is increased 2- to 3-fold in the presence of galactose. However, the Gal3p concentration in cells growing in a galactose-containing medium is 5-fold higher than the Gal80p concentration (Verma, Bhat et al. 2003).

Deletion of either one of feedback loops affects *GAL* gene activation. In yeast containing a knockout of Gal80p, the *GAL* switch is fully induced at 0.01% of galactose, whereas wild type yeasts are induced approximately 13% at the same concentration of galactose. By comparison, the *GAL* response is slow and reduced 5-fold in a Gal3p knockout. Therefore, the galactose response depends on a balance between the Gal3p- and Gal80p-mediated feedback loops, and the distribution of Gal80p

between nucleus and cytoplasm is a key component to controlling the *GAL* switch (Ramsey, Smith et al. 2006).

In addition to Gal3p, Gal1p also plays a role in the positive feedback loop. Gal1p is bifunctional, with the roles of galactokinase and a weak transcriptional inducer. Gal1p and Gal3p have 70% amino acid identity and 90% sequence similarity. Gal1p also binds with Gal80p and can substitute for Gal3p when it is present at a high concentration. After adding galactose, the expression level of Gal1p is increased 1,000-fold and it further enhances the Gal3p positive feedback loop (Thoden, Sellick et al. 2005; Sellick, Jowitt et al. 2009; Kundu and Peterson 2010). Thus, in the presence of galactose, several positive feedback loops work together to activate the *GAL* gene.

Transition to the early elongation complex and elongation

The transition from initiation to the early elongation complex is tightly regulated by the exchange of factors due to the phosphorylation of the CTD. The CTD is mostly not phosphorylated during promoter binding. The unphosphorylated CTD stabilizes the interaction between the PIC and factors such as Mediator. The CTD is phosphorylated at Ser5 by the Cdk7 subunit of TFIIH. Phosphorylated Ser5 destabilizes the interaction between the promoter and the general transcription factors, thereby facilitating promoter escape (Nechaev and Adelman 2011).

The transcribing complex undergoes structural changes during the transition from the initiation to the elongation complex. The N-terminal B finger of TFIIB approaches the RNA exit channel toward the active site of RNAPII and stabilizes the short, nearly synthesized RNA. The B finger collides with the nascent RNA beyond about the 5th nucleotide position. Therefore, TFIIB must be displaced from the transcribing complex as the nascent RNA is extended beyond 4 nt in length. After synthesis of at least 7 nt RNA, the upstream DNA is rewound and the RNA-DNA hybrid is stabilized at the active site of RNAPII. The growing RNA facilitates dissociation of the

general transcription factors and promoter escape (Kostrewa, Zeller et al. 2009; Liu, Bushnell et al. 2010).

The rate-limiting step of promoter escape, which coincides with transition to the elongation complex, occurs with the addition of the 9th nucleotide and translocation of the active site of the RNAPII at the 9th position. During promoter escape, the upstream transcript slippage is decreased (at +8/+9), and abortive transcription also ends (at +10/+11). The requirement of the ATP and TFIIH is also terminated (at +8/+10). Finally, the transcription bubble is broken, which results in the transition to the early elongation complex. The rate of the promoter escape is sequence-dependent. A complex containing weak RNA-DNA interaction completes the promoter escape more slowly (Pal and Luse 2003; Hieb, Baran et al. 2006).

Promoter-proximal pausing is an important regulating step of RNAPII transcription *in vivo* and functions as a checkpoint prior to the transcription complex moving to productive elongation. Following promoter escape, the RNA remains stably bound with the transcription machinery. However, the early elongation complex is less stable than the productive elongation complex. In the early elongation complex, the transcription machinery is more likely to backtrack and stall than in the productive elongation complex. This happens near the promoter region and is called transcriptional pausing. RNA pausing was first detected at the *Drosophila melanogaster* heat-shock gene, *Hsp70*, and has since been found in many genes in eukaryotes. After synthesis of 25–50 nt of RNA, the RNAPII is paused. RNAPII pausing is due to its interaction with the heterodimer of DSIF and NELF. DSIF bound to NELF inhibits transcriptional elongation, while DSIF alone promotes RNAPII elongation. NELF is a complex of four subunits, NELF-A, B, C/D, and E, and is conserved in higher eukaryotes but has not been found in *Caenorhabditis elegans, Saccharomyces cerevisiae,* or *Arabidopsis thaliana*. NELF is colocalized with the hypophosphorylated RNAPII and has been observed to be associated with the RNAPII-bound gene by chip-chip analysis. Deletion of NELF in

Drosophila cells by the RNAi decreases RNAPII occupancy at promoters (Saunders, Core et al. 2006; Nechaev and Adelman 2011).

Unlike NELF, DSIF remains with transcriptional machinery after its transition into the productive elongation. DISF has both negative and positive effects on transcription depending on the activity of the proteins with which it interacts. NELF binds with DSIF causing RNAPII pausing in early elongation, while DSIF binds with P-TEFb in productive elongation (Saunders, Core et al. 2006; Nechaev and Adelman 2011).

Pausing in transcriptional elongation depends on specific local sequences such as the downstream promoter element and the pause button (GAGA elements). During transcription along A+T-rich regions the DNA-RNA hybrid stability is lower and dislodges the RNA 3' end from the active site, which causes initial pausing and backtracking of the RNAPII complex. The RNAPII machinery is preferentially arrested at the G+C-rich sequence between +26 and +23 with respect to the TSS, which is followed by stretches of A+T in the *Drosophila* embryo (Hendrix, Hong et al. 2008).

TFIIS plays an important role in promoting efficient escape of paused RNAPII. During backtracking and arrest of the RNAPII complex, RNA transcript becomes misaligned at the RNAPII active site. TFIIS stimulates displacement of backtracked RNA, cleavage to create a new RNA 3'OH at the active site and enables resumption of transcriptional elongation (Adelman, Marr et al. 2005; Cheung and Cramer 2011).

P-TEFb, which is located on active genes, is a central factor for productive elongation. Depletion of P-TEFb results in failure of gene expression, and inhibition of P-TEFb reduces the RNAPII density on genes. P-TEFb is a heterodimer of the cyclin-dependent kinase 9 (Cdk9) and regulatory subunits, cyclinT in *Drosophila* and mostly cyclinT1 and less frequently cyclinT2a and cyclinT2b in HeLa cells (Peng, Zhu et al. 1998). It phosphorylates Ser2 of the CTD of RNAPII, providing for a platform of assembly for complexes that travel with the polymerase along the gene.

P-TEFb also phosphorylates DSIF and NELF. Phosphorylated NELF is released from DSIF, which

promotes productive elongation (Saunders, Core et al. 2006; Nechaev and Adelman 2011).

RNA polymerase II (RNAPII)

RNAPII synthesizes all messenger RNA in the eukaryotes. It consists of 12 subunits with a

highly conserved structure and size throughout eukaryotes. Rpb1 and Rpb2 form the central part of

the RNAPII in *S. cerevisiae* and are catalytic subunits that form the active site (Cramer, Bushnell et

al. 2000; Malagon, Kireeva et al. 2006). The two largest subunits, Rpb1 and Rpb2, associate with a

cleft between them that is bridged by the helix of Rpb1 and the carboxy terminal region of Rpb2 in

the opposite side. The canonical B-form of DNA lies in the Rpb1-Rpb2 cleft and continues in a

straight path to the active site (Cramer, Bushnell et al. 2000). Two Mg^{2+} ions are crucial for catalysis

by the RNAPII transcribing complex. The two ion binding sites are in the active region of Rpb1 and

Rpb2 (Wang, Bushnell et al. 2006).

Rpb1

Rpb1 is the largest subunit of the RNAPII and contains a CTD that consists of 52 repeats in

mammals and 26 repeats in yeasts of the consensus heptapeptide YSPTSPS. The CTD is responsible

for guiding RNA processing by recruiting the appropriate proteins complexes for different stages in

processing. Transcripts produced by RNA polymerase I (RNAPI) and RNA polymerase III

(RNAPIII) that lack CTDs cannot be processed into mRNA. Rpb1 is involved in the pre-mRNA

processing and transcription depending on the level and pattern of phosphorylation of the CTD

(Orphanides and Reinberg 2002; Proudfoot, Furger et al. 2002). Recruitment of different pre-mRNA

processing machineries is associated with distinct regions of the CTD (Orphanides and Reinberg

2002). Ser5 phosphorylation is generally increased when the RNAPII is at the promoter region,

whereas Ser2 phosphorylation is detected when the enzyme is approaching the 3' ends of genes. Ser7

phosphorylation is related to maturation of small nuclear RNAs (Koch, Jourquin et al. 2008). The

CTD phosphorylation pattern defines the stage of transcription and the events associated with that stage. Partial phosphorylation of the CTD motif during transcription initiation recruits the capping enzyme. Phosphorylation of Ser5 in the CTD motif occurs between transcription initiation and promoter clearance. Further phosphorylation of the CTD upon promoter escape recruits the pre-mRNA splicing machinery.

Rpb2

Rpb2 is the second largest subunit of the RNAPII. It is encoded by a single copy gene of the haploid genome and is essential for yeast growth. The Rpb2 protein sequence contains conserved motifs, a nucleotide-binding domain, and a metal-binding domain, which contribute to a portion of the catalytic site for RNA synthesis (Sweetser, Nonet et al. 1987). Mutation studies of amino acids in the active site have identified amino acids K979, K987, and R1020 of Rpb2 as important for transcription and essential for viability in yeast. Rpb2 R1020 interacts with i+1 NTP γ-phosphate. In the TFIIS deletion strain, the R1020K/Q mutation of Rpb2 causes defects in elongation and consequently cell growth is slowed. Mutation of the adjacent amino acid in Rpb2 active site (S1019F) results in very slow RNAPII elongation *in vitro*. It was concluded that S1019F may cause mislocation of R1020, which is required for catalysis (Domecq, Kireeva et al. 2010). In the rbp2-10 mutant yeast (P1018S in rbp2) RNAPII is found more frequently present in the arrest site compared to the wild type. The rpb2 (P1018S) demonstrates a reduced elongation *in vitro*. A reduced transcription reaction rate mimics the effect of a low substrate concentration, thus enzymes encounter an increasing risk of becoming arrested at the arrest sites of the DNA (Powell and Reines 1996).

To measure the elongation rate of RNAPII, the large (8kb) and nonessential gene, YLR454, was used under the control of the GAL1 promoter. The elongation rate and processivity were determined by RNAPII occupancy at different positions within the YLR454 coding region. In these studies it was found that rpb2-10 displays delayed dissociation of RNAPII from the gene compared to

the wild type. The elongation rate in rpb2-10 was reported as around 1kb/min, whereas the rate was measured as 2kb/min in the wild-type cells. In the rpb2-10 strain, the RNAPII did not successfully translocate to the 3' end of the YLR454 coding region. The reason for this is probably that the reduced elongation rate increases the dwell time of the RNAPII, resulting in an increased probability of premature dissociation from the template. Thus mutation of Rpb2 subunit (rpb2-10) decreased the RNAPII elongation rate and processivity *in vivo* (Mason and Struhl 2005).

Cotranscriptional events

Capping

Capping of the 5' end is the first cotranscriptional modification of the mRNA. Capping usually occurs after 20–30 nt of the pre-mRNA has been transcribed and when the 5' end of the nascent transcript becomes accessible. The RNAPII pauses after transcription of 20–30 nt of RNA, which allows time for 5' capping. The pre-mRNA must be capped for the RNAPII to further elongate it (Orphanides and Reinberg 2002). Three steps are involved in the addition of the cap to the 5' end of a pre-mRNA. First, an RNA 5' triphosphatase hydrolyzes the triphosphate of the first nucleotide, changing it to a diphosphate. Then a guanylyltransferase catalyzes coupling with GMP (guanosine monophosphate) using a GTP (guanosine triphosphate) substrate to form a 5'-5' triphosphate linkage to the first nucleotide of the pre-mRNA. Finally, a methytransferase methylates the N7 position of the GMP (Proudfoot, Furger et al. 2002).

Recruitment of capping enzymes is also part of the transition from the initiation to elongation complex of RNAPII. Phosphorylation of the CTD plays an important role in capping. The components of the capping enzymes, such as guanylyltransferase and methyltransferase, bind directly to the phosphorylated CTD at both Ser2 and Ser5, but do not bind unphosphorylated CTD. Guanylyltransferase is allosterically activated by Ser5 (but not Ser2)-phosphorylated CTD, which promotes formation of the enzyme-GMP complex (Ho and Shuman 1999). DSIF also recruits a

capping enzyme to the transcription machinery. Spt5, the larger subunit of DSIF, interacts with the triphosphatase and guanylyltransferase to stimulate capping. The capping enzyme overcomes NELF-mediated repression of RNAPII. Therefore, capping serves as a checkpoint during transcription to move productive elongation through promoter escape and from the paused early elongation complex (Mandal, Chu et al. 2004).

The cap structure also affects transcription. The cap binding complex (CBC) interacts with P-TEFb in HeLa cells and is necessary for optimal phosphorylation of Ser2 of CTD. When the CBC is depleted by siRNA treatment, the level of Ser2 phosphorylated CTD is decreased due to decreased occupancy of P-TEFb and transcriptional elongation is inhibited. Therefore, the CBC mediates pre-mRNA capping and RNAPII elongation by interacting with P-TEFb, which ensures optimal phosphorylation of Ser2 of the CTD (Lenasi, Peterlin et al. 2011).

Splicing

Mammalian genes consist mostly of noncoding sequences (introns), so mRNA splicing is required to produce functional mRNA. Splicing of the transcribed pre-mRNA is cotranscriptional. The pre-mRNA contains several consensus elements in *cis* for the splicing reaction. The 5' splice site is identified by the consensus sequence AG/*GURAGU*, and the 3' splice site by *YAG*/RNNN (R: purine and Y: pyrimidine) in which the splice position is marked by / in each case. The branch point, $CURA^{2'OH}Y$, lies about 100 nt upstream of the 3' splice site with a conserved adenosine that is followed by a pyrimidine-rich track. There are two steps in the splicing reaction. First, the 2' OH of the branch point adenosine attacks the 5' exon-intron border as a nucleophile, resulting in transesterification between the 2' OH and 5' intron border. A free 5' exon and a lariat containing the intron and 3' exon are formed after this first reaction. In the second step, the 3' OH group of free 5' exon attacks the 3' intron-exon border. As a result, the two exons are fused and a lariat-shaped intron is released by these two transesterifications (Moore and Sharp 1993).

The five small nuclear RNAs (U1, U2, U4, U5, and U6) and other proteins are assembled with introns to form the splicesome, which promotes splicing. Interactions of the 5' splice site and U1 snRNP (small nuclear ribonucleoprotein) initiates the splicesome assembly. U2 snRNP is recruited to the branch point with other branch point-binding proteins by the pyrimidine track and the interaction of the splicing factor U2AF at the 3' splice site. U2 snRNP is paired with the branch point sequence so it promotes the first nucleophilic attack by bulging out the adenosine. U4 snRNP does not directly interact with the pre-mRNA, but it recruits U5 and U6 to the splicesome. After the heterotrimer, U5/U4/U6, is formed, U5 binds the 5' splice site and U6 binds with U2. After U1 is released from the 5' splice site, U6 interacts with the 5' splice site. The U6/U2 portion of the complex forms the catalytic core of the splicesome and catalyzes transesterification. U5, which binds the exon at the 3' splice site, releases U4 and the first splicing reaction occurs. The U2/U5/U6 complex remains bound to the lariat-shaped intron while the second reaction takes place. U5 is positioned in the exon for the second nucleophilic attack (Lewin 2004).

The U1 snRNP is associated with the transcription machinery on all transcribing genes whether or not they contain introns. U1 snRNP goes with the polymerase during transcription to identify the splice sites and recruits the remaining components of the splicing machinery when it binds the 5' splicing site. The polymerase elongation kinetics are not affected by splicing events on the emerging pre-mRNA. The more introns that are on the pre-mRNA, the more splicing factors are recruited at the active transcription site. The polymerase activity is separated from the splicing complexes once transcription has terminated. The splicing complexes stay with the pre-mRNA transcript until splicing is completed (Brody, Neufeld et al. 2011).

Activity of the splicing reaction is also coupled with the CTD of RNAPII. CTD-truncated RNAPII is associated with severely inhibited splicing in mammalian cells because the CTD is required for targeting the splicing factors to the transcription sites (McCracken, Fong et al. 1997; Hirose and Ohkuma 2007). The phosphorylated CTD directly interacts with U2AF65, which

recognizes and binds to polypyrimidine-rich sequences found near the 3' end of the splice site. Interaction between the phospho-CTD and U2AF65 recruits PRP19C, which is an essential splicing factor in activation of the splicesome but does not tightly associate with snRNPs. The recruitment of PRP19C through interaction with the CTD-bound U2AF65 facilitates efficient formation of the splicesome on nascent transcripts. Thus, splicing is coupled with RNAPII transcription by CTD-mediated molecular interactions (David, Boyne et al. 2011).

Splicing of pre-mRNA is also affected by other cotranscriptional events, such as 5' capping. Deletion of the capping enzyme leads to the accumulation of unspliced pre-mRNA (Burckin, Nagel et al. 2005). Recruitment of the U1, U2, and U5 snRNPs and assembly of the splicesome are decreased in the CBC-deleted yeast. However, recruitment of Mud2p, the 3' splice site factor, is not affected by deletion of the CBC, and both spliced and unspliced mRNA are formed in the yeast CBC deletion strain. Therefore, the CBC is necessary but not sufficient for cotranscriptional splicing (Gornemann, Kotovic et al. 2005).

Termination and polyadenylation

Transcriptional termination occurs when RNAPII stops RNA synthesis and both RNAPII and the nascent RNA are released from the DNA template. Termination is coupled with process of cleavage and polyadenylation. This 3' end processing is divided into two steps: 1) Transcription of poly (A) site is followed by pausing of RNAPII transcription and cleavage of the nascent RNA; and 2) The upstream cleaved RNA is polyadenylated, whereas the downstream cleavage product is degraded. The poly (A) site is the key determinant of termination (Kuehner, Pearson et al. 2011).

The CTD provides a scaffold for recruitment of RNAPII termination factors. Truncation of the CTD severely inhibits RNA cleavage at the poly (A) site because cleavage and polyadenylation factors are bound to the CTD (McCracken, Fong et al. 1997). Ser2 phosphorylation of the CTD is important for 3' end processing. A polyadenylation factor such as Pcf11 and termination factor such

as Rtt103 are preferentially bound to the phospho-Ser2 CTD. The loss of Ser2 phosphorylation leads to a defect in 3' end processing. Thus, high phospho-Ser2 near the poly (A) site provides a platform to recruit polyadenylation and termination factors, resulting in proper mRNA termination (Hirose and Ohkuma 2007; Lunde, Reichow et al. 2010).

Polyadenylation

Polyadenylation occurs at the end of transcription and is coincident with transcriptional termination. Most mRNAs contain a uniform 3' end consisting of approximately 200 adenosine residues of polyadenylation. The common cleavage site in eukaryotes consists of one A-rich element and one or more U-rich regions, although the *cis*-element sequence differs among mammalian, yeast, and plant pre-mRNAs. The formation of poly (A) in pre-mRNA is directed by the polyadenylation machinery. Before polyadenylation, pre-mRNA is cleaved at between a highly conserved AAUAAA hexanucleotide sequence and a downstream sequence element (DSE), which is a U- or GU-rich motif in Metazoa. The AAUAAA sequence occurs 10 to 30 nucleotides upstream of the cleavage site. This sequence is essential for proper 3' formation. Cleavage and polyadenylation is predominately at a CA dinucleotide. The cleavage and polyadenylation specificity factor (CPSF) interacts with AAUAAA. The DSE acts as a platform for interaction with the cleavage stimulatory factor. Cleavage factor I (CFI) and cleavage factor II (CFII) are also involved in 3' end cleavage. CFI directly interacts with the pre-mRNA substrate, and CFII is the essential factor for cleavage. A poly (A) polymerase (PAP) is also required in the cleavage reaction with CPSF and catalyzes the poly (A) addition. The poly (A) binding protein binds to poly (A) and enhances processivity of the PAP (Proudfoot, Furger et al. 2002; Millevoi and Vagner 2009).

Saccharomyces cerevisiae uses different, but complex, signals to direct the mRNA 3' end formation. The yeast mRNA does not have AAUAAA sequence in the 3' untranslated region. Instead, it uses various sequences such as TTAAGAAC (*CYC1*), AATAATGA (*GAL7*), and

AAAAAAAA (*ADH1*) as signals for 3' end formation. Efficiency and positioning elements are required for the yeast 3' end formation signal, and flanking sequences also affect the precise selection of poly (A) sites but to a lesser degree than do the efficiency and positioning elements. The efficiency elements, which contain the sequences TATATA, TAG⋯TATGTA, TTTTTATA, or their derivatives, increase the activity of the positioning elements. The positioning sequences are composed of TTAAGAAC, AAGAA, and other A-rich sequences such as AATAAA directing the position of the poly (A) sequences. Polyadenylation begins 16 to 27 nt downstream of the positioning sequence, usually after pyrimidine residues that are followed by one or more adenosine residues (Guo and Sherman 1995).

Transcription and chromatin remodeling

The DNA template is wrapped around histone proteins in the form of nucleosomes that are a barrier to transcription. The accessibility of DNA is highly regulated by chromatin dynamics such as ATP-dependent nucleosome remodeling and histone modification during transcription. The nucleosome remodeling complex weakens histone-DNA contacts by using energy from ATP hydrolysis, which increases nucleosome movement and nucleosome disassembly and assembly (Cosgrove, Boeke et al. 2004). For efficient transcription, nucleosomes are disassembled in front of an RNA polymerase path. Therefore, transcriptional initiation and elongation is associated with a loss of histone density. For example, the histone density of the yeast *GAL* gene is decreased in an active gene locus at the promoter region and in the coding region (Li, Carey et al. 2007).

Histone modification, depending on its location in the histone octamer, changes chromatin dynamics. It alters the packing of chromatin, which changes internucleosomal contacts or electrostatic charge. For example, acetylated histones are displaced easily from DNA. In addition, it changes the chromatin structure, resulting in creating binding sites for protein domains that are

recognized by many transcription factors. Transcription factor binding to induce transcription is influenced by chromatin structure (Li, Carey et al. 2007).

Histone acetylation

The N-terminal tails of histones H3 and H4, which are highly evolutionarily conserved proteins, can be posttranslationally modified by acetylation. Acetylation neutralizes the positive charges of the lysine residues of the N termini of H3 and H4. Acetylation of histone limits chromatin folding *in vitro*. Each histone can be acetylated at multiple lysine residues. Once acetylated, the neutralized histone tail interacts more loosely with DNA, so it increases access of the transcription factor to the DNA. Acetylation and deacetylation of histones (except for H4K12) are linked to transcriptional activation or repression, respectively. In an active gene, acetylation of histones H3 and H4 occurs mainly at the 5' UTR and promoter regions. Histone acetylation is regulated by histone acetyltransferase complexes (HAT) and histone deacetylase (HDAC), which are associated with gene regulation (Workman and Kingston 1998). However, acetylation of some histone residues is involved in inactivating some genes; for example, Hos2, histone deacetylase in *S. cerevisiae*, deacetylates Lys12 of histone H4 (H4K12) in activating *GAL* genes and is preferentially associated with the coding region of genes that have high transcriptional activity (Wang, Kurdistani et al. 2002).

Histone acetylation is generally related to transcriptional activation. Lys9 of histone H3 (H3K9) and Lys14 of histone H3 (H3K14) are highly acetylated at transcriptional start sites of active genes, which is catalyzed by the Gcn5 HAT. Lys of histone H4 (H4K5, H4K8, and H4K16) is acetylated by the Esa1 HAT. Hyperacetylation of H4 peaks at the start sites of active genes. Histone acetyl transferases Gcn5 and Esa1, which are generally recruited by transcription activators, are enriched in the promoter regions and at the transcription start sites of active genes (Pokholok, Harbison et al. 2005).

Histone acetylation activity is also associated with elongation by RNAPII. Lys56 of histone H3 (H3K56), which is acetylated by the Rtt109 HAT, is associated with an actively transcribed gene and is associated with the elongating RNAPII in yeast and *Drosophila* (Schneider, Bajwa et al. 2006). Acetylation of H3K56 by Rtt109 and ASF1, HAT, is important for RNAPII transcription in heterochromatic structures in the coding region, which is known for repressive loci in yeast (Varv, Kristjuhan et al. 2010).

Histone methylation

Histone methylation on lysine residues can be associated with gene activation or repression, depending on the site of methylation. There are six lysine residues (H3K4, H3K9, H3K27, H3K36, H3K79, and H4K20) for methylation on histones H3 and H4. In general, methylation at H3K9, H3K27, and H4K30 correlates with transcriptional repression, whereas methylation at H3K4, H3K36, and H3K79 correlates with transcriptional activation. Also, the lysine residues of histones can be modified by one, two, or three methyl groups, and distinct functions in transcription are associated with different numbers of methyl groups. Methylation of histones is performed by different methyltransferases, such as Set complex and Dot1, at different positions of Lys residues (Martin and Zhang 2005).

Lys4 of histone 3 (H3K4) is methylated by the Set1 methyltransferase in yeast. Phosphorylation of Ser5 of the CTD by TFIIH recruits the Set1 complex to the 5' end of the gene, resulting in trimethylation of H3K4, which is concentrated at the 5' region of the gene and activates transcription. Dimethylation of H3K4 predominates downstream of the transcription start site, whereas monomethylation of H3K4 is distributed across the coding region (Dillon, Zhang et al. 2005; Saunders, Core et al. 2006). Different numbers of methyl groups on H3K4 play different roles in transcriptional activation.

The interaction of Set2 and the CTD of the RNAPII is important for elongation. The elongating RNAPII is hyperphosphorylated at the CTD, and *S. cerevisiae* Set2 interacts with the hyperphosphorylated RNAPII. Deletion of the Rpb1 interacting domain of Set2 resulted in the same effect as deletion of the entire Set2. Set2 methylates Lys36 of histone H3 (H3K36) results in an increase of dimethylation (H3K36me2) and trimethylation (H3K36me3). H3K36me2 is distributed near the promoter region in *Drosophlia*, whereas H3K36me3 is highly enriched at the 3' end of the active coding region in yeast and *Drosophila*. Methylated H3K36 recruits the Rpd3S HDAC complex in yeast and catalyzes elongation-associated histone deacetylation. Therefore, the major role of H3K36 methylation is to regulate the cycle of acetylation and deacetylation connected with transcriptional elongation (Pokholok, Harbison et al. 2005; Bell, Wirbelauer et al. 2007; Selth, Sigurdsson et al. 2010).

H3K79 interacts with Sir3 (silent information regulator 3) protein. The Dot1 histone methyltransferase modifies Lys79 of histone 3 (H3K79) and the interaction between H3K79 and Sir3 is blocked by methylation. Dot1 makes a complex with ENL (eleven nineteen leukemia)-associated proteins that is associated with P-TEFb, and knockdown of ENL decreases genome-wide dimethylation of H3K79 and transcriptional activity. In addition, trimethylation of H3K79 in yeast is found on the transcribed regions of genes. Thus, methylation of H3K79 is related with RNAPII elongation (Pokholok, Harbison et al. 2005; Nguyen and Zhang 2011).

Dynamics of transcription

Gene transcription involves several steps and is very dynamic. However, the real-time kinetics of transcription have not yet been fully investigated and studies to date are mostly based on a single *in vivo* imaging technique to observe the RNA production. For this purpose, the Singer group has developed a single molecule assay to observe the synthesis of RNA from an active gene by the RNAPII in live cells. To investigate RNAPII enzymatic activity, 24 repeats of the PP7 bacteriophage

binding site were inserted in the 5' UTR of a target gene under a cell cycle-regulated *POL1* promoter. In the same cells, PP7 was constitutively expressed as a fusion protein with GFP. When the 24 repeats of the PP7 binding site are transcribed, they are bound with PP7-GFP. The transcribing RNA can be imaged because of the concentration of the GFP on the repeats. The concentrated fluorescent signal continues with the transcription of *GLT1* gene until the nascent RNA is released and moves away from the transcription site. This method can measure both pre-mRNA synthesis and the dwell time of RNA at the transcription site. By inserting PP7 cassettes in the 3' and 5' UTRs of a long gene such as *MDN1* comparing the fluorescent signals, initiation, elongation, and termination events can be distinguished. From this it was determined that the dwell time and rate of initiation and elongation of MND1 transcription varies with the cell cycle. The dwell time in the G1 phase is longer than in the late S/G2 phase, so the rate of initiation and elongation in the G1 phase is longer than in the late S/G2 phase (Larson, Zenklusen et al. 2011).

Stochastic events are characteristic of many biological systems and processes. The noise of gene expression is a stochastic process in transcription and/or translation. Diverse phenotypes observed in isogenic cells result from expression noise in spite of the genetic equivalence. Negative feedback regulation can decrease the noise (Dong, Shao et al. 2010). For example, the transcriptional initiation of the *GAL* gene shows low stochasticity because of the negative feedback loop of Gal80p. In a yeast containing knockout of Gal80p, the coefficient of variation (c.v.) of fluorescence of GFP under the GAL1 promoter was larger than for the wild type. Thus, it was concluded that the fluctuation in *GAL* gene expression is suppressed in the presence of negative feedback (Ramsey, Smith et al. 2006).

In vivo imaging

Reporter proteins

β- galatosidase

Reporter proteins, such as β-galactosidase, luciferase and green fluorescent protein, have been widely used to monitor gene regulation when coupled with a defined promoter. β-galactosidase, encoded by bacterial *LacZ* gene, is stable and retains high activity through harsh chemical treatments such as tissue fixation. It can hydrolyze chromogen substrates including o-nitrophenol β-D-galactopyranoside (ONPG), 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal), and 3,4-cyclohexenoseculetin-β-D-galactopyranoside (S-gal), which produce yellow, blue, and black products, respectively. When expressed by an activate promoter, β-galactosidase can be used to easily detect gene regulation by monitoring the color change of its substrate. With a different pH optimum from the mammalian lysosomal β-galactosidase, it can be measured in the absence of interference by the endogenous enzyme. However, the β-galactosidase system cannot be applied to real-time detection because cell lysis or fixation is required for its quantitative measurement. By contrast, GFP and luciferase reporter proteins can be monitored in the absence of cell lysis. β-galactosidase is also less sensitive than luciferase and is therefore not useful for monitoring the activity of weakly or moderately active promoter (von Degenfeld, Wehrman et al. 2009; Ghim, Lee et al. 2010; Smale 2010).

Fluorescent proteins

Green fluorescent protein (GFP) from *Aequorea Victoria* has been used as a protein tag for *in vivo* labeling since 1994, after its cloning in 1992. GFP and its variants have been applied in various studies of the structure and function of living organisms. The chromophore group of GFP is formed by posttranscriptional cyclization of Ser65-Tyr66-Gly67. The chromophore is located in the center of

the GFP β-barrel, and the side chains of amino acids inside the barrel play important roles in fluorophore formation. Variation in the side chain sequence alters the excitation and emission spectra, resulting in different colors of fluorescent proteins. The barrels of these fluorescent proteins are stabilized by several noncovalent interactions, which give high stability to thermal or chemical modification; thus the chromophore is well protected from the surrounding environment. Fluorescence from fluorescent proteins appears after protein folding and chromophore maturation, which involves several covalent modifications (Chudakov, Matz et al. 2010).

Chromophore maturation is the rate-limiting step for developing fluorescence in GFP and related fluorescent proteins. Depending on the fluorescent protein, oxygen concentration, and temperature, the maturation takes from several minutes to hours, or even days. Oxygen is required to dehydrogenate the α,β bond at residue 66, which is the slowest step in maturation. Because of the oxygen requirement for its maturation, GFP cannot be used in anaerobes. The maturation time for most fluorescent proteins is ~40 min to 1–2 hours. For applications that involve studies of rapidly changing protein levels, rapidly maturing fluorescent proteins, such as yellow fluorescent proteins, are used. For example, Venus is the fastest maturing fluorescent protein and still takes several minutes for maturation *in vitro* at 37°C (Tsien 1998; Nagai, Ibata et al. 2002; Chudakov, Matz et al. 2010).

The main applications of fluorescent proteins are as fusion proteins to visualize protein expression, localization, translocation, interaction, and degradation in living systems in real time. Fluorescent proteins report promoter activity in a given genetic environment cloned under the control of a promoter of interest. Fluorescent proteins as reporters are useful because of their broad-host applicability and absence of cell toxicity or the need for addition of substrate. However, to be used to report promoter activity, the rates of synthesis, maturation and turnover of fluorescent proteins should be considered. The rate-limiting maturation time delays detection of the fluorescence signal (Chudakov, Matz et al. 2010). Fluorescent proteins also give a false signal after the gene has been

turned off because they are stable and continue to fluorescence (Ghim, Lee et al. 2010). For example, the half-life of GFP is more than 420 min in yeast (Natarajan, Subramanian et al. 1998). For temporal analysis of gene expression, a short half-life GFP was developed by inserting a short C-terminal peptide tag, which is susceptible to a housekeeping protease. By insertion of the C-terminal peptide tag, the half-life of GFP variants ranges from 40 min to 110 min in *E. coli*, while the half-life of the wild type GFP is one day in *E. coli* (Andersen, Sternberg et al. 1998). A short lived cyan fluorescent protein (CFP) that is destabilized by ubiquitin fusion at N-terminal was used as a transcriptional reporter in living yeast cells (Hackett, Esch et al. 2006). Fluorescence intensity of the short half-life CFP ($\tau_{1/2}$ = 12 min) showed a similar response to time-dependent profile of transcription in an oscillating system, whereas the fluorescence intensity of a long half-life fluorescent protein ($\tau_{1/2}$ = 76 min) did not show changes that paralleled the transcription profile. In addition to the short half-life, a fast maturing fluorescent protein increased the dynamic range or the reporter to allow accurate tracking of rapid changes in transcription (Wang, Errede et al. 2008). However, there is still a time gap between the change in transcription and fluorescence intensity of GFP-derived fluorescent proteins. For example, the maximum RNA level synthesized from an active FUS promoter induced by pheromone occurred after 15 min, while the maximum fluorescence intensity of short half-life CFP synthesized from the FUS promoter occurred after 150 min (Hackett, Esch et al. 2006). Although fluorescent proteins track rapid change of gene expression, they showed delayed response after stimuli due to time for protein synthesis.

Bioluminescence proteins

Luciferase is a bioluminescence reporter gene that is used as a noninvasive method of monitoring gene expression. It does not require posttranslational modification and is active immediately after translation. *In vivo,* bioluminescence can be detected by a key enzyme, luciferase that releases visible light by oxidizing its substrate, luciferin, which is provided exogenously. To

monitor and measure reporter activity, bioluminescence imaging is used for various host systems including bacteria, mammalian cells, virus, and transgenic animals. However, the light emitted from luciferase is extremely dim, so it requires an especially sensitive charge-coupled camera to detect bioluminescence imaging (Prescher and Contag 2010).

Firefly luciferase (*Pontinus pyralis*) and insect luciferase (*Pyrophorus plagiophthalamus*) emit photons at around 600 nm due to their oxidizing D-luciferin in the presence of molecular oxygen and ATP. Luciferase from the *Renilla* species oxidizes coelenterazine and releases primarily blue-green light. Renilla luciferase does not require cofactors except oxygen to catalyze the release of light. Bacterial luciferase also produces blue-green light and needs long-chain aldehydes and flavin mononucleotides, but this luciferase and required proteins are encoded in a single *lux* operon. For this reason, bacterial luciferase is limited to use in bacteria (Gross and Piwnica-Worms 2005; Prescher and Contag 2010). For application in different systems, the bacterial luciferase requires modifications such as codon optimization and simultaneous expression of all five genes (*luxCDABE*) of the *lux* operon. A modified bacterial luciferase cassette has been applied to mammalian cells in culture for whole animal bioluminescence imaging that does not require exogenous substrate addition (Close, Patterson et al. 2010).

Compared with GFP, luciferases show low backgrounds in cell culture and in animal models. It also takes less time to see the signal from firefly luciferase (0.17h in mammalian cells) than from modified bacterial luciferase (16h in mammalian cells) or GFP (22h in mammalian cells). In addition, the signal from firefly luciferse is stronger than from bacterial luciferase. On the contrary, bacterial luciferase does not depend on the addition of substrate with the result that a continuous luminescent signal can be tracked, whereas firefly luciferase requires addition of the exogenous substrate, luciferin (Close, Hahn et al. 2011).

RNA imaging systems

Gene expression is the process that results in production of mRNAs and proteins. It involves several steps, including transcription, RNA splicing, translation, and posttranslational modification. Gene expression is measured by quantification and localization of the mRNA and protein products. To analyze gene expression, mRNA levels can be observed by a polymerase chain reaction (PCR), microarray, and *in situ* hybridization. These techniques require lysed or fixed cells and cannot be used to detect the mRNA level in real time. For imaging transcription of RNA in real time in live cells, a noninvasive means of detecting gene expression in the response of promoter activity is needed. There are several requirements for optimal real-time imaging of transcription activity. The first requirement is for an RNA tag that produces a strong signal. One means of increasing the signal to noise is to have a repeated RNA signaling sequence in the tag. A second requirement is that the tag does not interfere with the sequence, structure, and function of mRNA. It is also desirable that the tag be small for efficient RNA synthesis (Kubota, Ikeda et al. 2010).

MS2 system

To tag mRNA, GFP fused proteins that bind to specific RNA elements have been developed for mRNA imaging (Figure 1F). The MS2 system for imaging gene expression in real time consists of two plasmids. In the first plasmid, the GFP sequence is fused with the coding sequence for the RNA phage capsid protein MS2 that is expressed from a constitutive promoter. In the second plasmid, 24 copies of the MS2-binding RNA motif are incorporated into an untranslated region of the target mRNA that is under the control of a regulatable promoter such as the galactose-inducible promoter in yeast. The target mRNA can be visualized by the interaction between the sequence MS2-hairpin RNA elements and GFP-fused MS2 protein. Cytoplasmic localization of *ASH1* mRNA was visualized by the MS2 system in living yeast (Bertrand, Chartrand et al. 1998). Other bacteriophage coat proteins such as PP7 are also used for similar constructions (Larson, Zenklusen et al. 2011). The

binding of multiple GFP-MS2 fusion proteins to a single RNA transcript produces high fluorescence compared with the background. Therefore, the MS2 system has been used to localize and track the movement of mRNA transcripts in several species, including *Xenopus* oocytes, fruit flies, *E. coli,* and mammalian cells (Bao, Rhee et al. 2009; Tyagi 2009).

The MS2 system has some limitations. For example, the fluorescent protein is always fluorescent after formation of the GFP regardless of its association with the RNA. Therefore, it is difficult to distinguish GFP-bound RNA from unbound GFP. Because MS2-GFP has a nuclear localization signal, most RNA-bound GFPs are in the cytoplasm and unbound GFPs are in the nucleus. However, the RNA transport system has to deal with two conflicting processes, which are the mechanism of mRNA transport from the nucleus and the entry into the nucleus of the MS2-GFP by way of its nuclear localization signal. These opposing signals affect the export of mRNA from the nucleus (Tyagi 2009). The binding of multiple proteins to the target mRNA could also change the intracellular dynamics of the mRNA, so the GFP-bound, MS2-modified mRNA does not closely represent the condition of the endogenous mRNA (Itzkovitz and van Oudenaarden 2011).

RNA-mediated reconstruction of GFP

To decrease the background contribution of the GFP not bound to MS2-modified RNA, a split-GFP approach was developed (Figure 1G), in which the GFP and related proteins are split into two nonfluorescent fragments. In this system assembly of the fluorescence protein depends on additional protein interactions. The eukaryotic initiation factor 4A (eIF4A) is a dumbbell-shaped protein with two globular domains. Selected aptamers for each domain have strong affinity for their respective domains but low affinity for the other domain. Each domain of eIF4A was fused with a fragment of enhanced GFP (eGFP) and expressed together in *Escherichia coli*. When aptamers were coexpressed with fusion proteins, each domain of eIF4A was bound to its respective aptamer and the

aligned eGFP fragments came closer, resulting in reconstruction of the eGFP fluorescence (Valencia-Burton, McCullough et al. 2007).

In another system, the RNA tags consist of two RNA-binding domains of the human Pumilio homology domain (PUM-HD), in which each domain is connected with a split fragment of eGFP. The split eGFP can be reconstructed by binding PUM-HD to the target mitochondrial RNA, which encodes the nicotinamide adenine dinucleotide dehydrogenase subunit 6. By targeting the split eGFP and PUM-HD to the mitochondrial matrix, it is possible to visualize the localization of target mitochondrial RNA (Ozawa, Natori et al. 2007).

Approaches to imaging RNAs that rely on reconstruction of the two halves of a fluorescent protein depend on an additional protein interaction that links the reconstructed fluorescent proteins to the RNA. Without this interaction, the fluorescent signal is at background levels. Monitoring the fluorescent signal allows spatial and temporal information regarding mRNA localization in single living cells. However, the split GFP system is not good for detecting dynamics of mRNAs in real time because reconstitution of eGFP takes 0.5–4 hours. The dissociation of the split fragment of the GFP is also very slow once the complete eGFP is formed. Thus, the remaining GFP gives a high background fluorescence after degradation of the target mRNA and makes it difficult to follow the mRNA levels in real time (Bao, Rhee et al. 2009).

Hybridization of endogenous mRNA with fluorescent probes

Approaches to imaging endogenous mRNA by using fluorescent antisense nucleic acids have been attempted that include linear oligonucleotide (ODN) probes, FRET reporters, and hairpin ODN probes (molecular beacons) (Figure 1A-E).

The fluorescently-labeled linear antisense ODN is the simplest mRNA probe that targets mRNA with fast hybridization kinetics that is compatible with detecting mRNA dynamics and rapid changes in mRNA expression (Okabe, Harada et al. 2010). However, the signal/noise is low because

it is difficult to distinguish the fluorescence of the bound, unbound and degraded probes. To visualize endogenous RNA in live cells, multiple probes targeting the same RNA are required to increase the signal (Bao, Rhee et al. 2009; Tyagi 2009). This and other methods that involve the extracellular application of labeled ODN probes are limited by 1) the rapid degradation of much of the probe because the cells take them up through the lysosomal system and 2) concentration of probe in regions of the nucleus while the RNA to be measured is mostly in the cytoplasm.

To decrease the background signal, FRET probes including molecular beacons, dual molecular beacons, and autoligation FRET probes were developed for imaging endogenous RNA. Molecular beacons have a stem-loop structure with antisense ODNs in the loop region and are labeled with a fluorophore on one end and a quencher on the other end. In the absence of target RNA, the molecular beacon adopts a structure in which the fluorophore is close to the quencher, showing low fluorescence intensity. After hybridization with the complementary sequence, the fluorophore is separated from the quencher and fluorescence increases. To get a signal from the molecular beacon, the complementary sequence must hybridize with target mRNA, which causes the stem to separate and thereby separates the fluorophore from its quencher. This characteristic enables the molecular beacon to function as a sensitive probe with a high signal to noise (Tyagi and Kramer 1996; Bao, Rhee et al. 2009). The molecular beacon is microinjected for efficient delivery. However, microinjection of beacons can damage living cells and cannot be applied in a large number of cells (Ozawa, Natori et al. 2007) and degradation is still a problem. The microinjected molecular beacons are also sequestered in the nucleus and can give a false signal if structurally distorted by interaction with cellular molecules.

To prevent sequestration in the nucleus, quantum dots and NeutrAvidin have been conjugated with molecular beacons, but these conjugations make it difficult for molecular beacon delivery and hybridization with complementary sequences (Chen, Davydenko et al. 2010). The molecular beacon also takes a longer time to hybridize with target mRNA in living cells than the linear ODN. The

hybridization kinetics depend on the composition of the ODN and its stem length. For example, a molecular beacon with a 4-base stem showed 100 times greater hybridization on-rate constant ($M^{-1}s^{-1}$) than a molecular beacon with a 6-base stem (Santangelo, Nitin et al. 2006). The longer time required for hybridization of the molecular beacon makes it difficult to explore the fast dynamics of mRNA synthesis (Okabe, Harada et al. 2010).

FRET probes have been tested that utilize two linear ODNs, which are fluorescently labeled at the 3' and 5' ends with donor and acceptor fluorophores, respectively, or that use two molecular beacons, that are labeled with donor and acceptor fluorophores, respectively. These FRET probes are targeted to adjacent regions on an mRNA such that the two fluorophores come close, and both probes are hybridized with target mRNA. This generates the FRET signal. However, the requirement for two adjacent complementary sequences on the target mRNA makes it difficult to implement this method because it is challenging to find a large unstructured stretch of mRNA that can be used for binding of two FRET probes (Bao, Rhee et al. 2009).

ODN probes have many limitations, such as delivery damaging the cells and being inefficient, degradation inside cells, impediments due to RNA structure, and bound proteins (Bao, Rhee et al. 2009; Tyagi 2009). To deliver ODN tags efficiently, streptolysin O, which reversibly permeabilizes the cell membrane, was used to deliver ODN probes into living cells. Delivery of ODN by streptolysin O was shown to be rapid, efficient, and applied in various cell types. An optimized amount of streptolysin O was not toxic to the cells (Faria, Spiller et al. 2001; Santangelo, Nix et al. 2004; Santangelo, Lifland et al. 2009). In addition, cell penetrating peptide-mediated transduction, which has the ability to traverse biological membranes efficiently, was used for ODN delivery to cytoplasm. TAT peptide (YGRKKRRQRRR, transacting activator of transcription of the human immunodeficiency virus), a cell-penetrating peptide, was covalently linked with ODNs and shown to enter the cytoplasm of a living cell with high efficiency providing for fast delivery kinetics (Nitin, Santangelo et al. 2004; Lifland, Zurla et al. 2010). Also, hairpin DNA-functionalized gold

nanoparticles are an efficient intracellular delivery without transfection or permeabilization agents (Jayagopal, Halfpenny et al. 2010). In another approach, ratiometric bimolecular beacons were made to mimic siRNA structures with 2nt overhangs at the 3' end. siRNA is efficiently exported from the nucleus by a nucleo-cytoplasmic transporter (Ohrt, Merkle et al. 2006; Chen, Davydenko et al. 2010).

To deliver tags into the cytoplasm, ODN probes have been modified with quantum dots to prevent sequestration in the nucleus (Chen, Behlke et al. 2007). Instead of using a big molecule, a polyethylene glycol (PEG) core was used to link with eight fluorescently labeled ODN probes. The PEG core provides several advantages when conjugated to ODNs that are utilized *in vivo*, including reduced toxicity, reduced nonspecific binding, and improved internalization (Lifland, Zurla et al. 2010).

To increase signal to noise, the intrinsic ability of gold colloids to quench fluorescence has been tested. In this approach, hairpin ODNs, with a fluorophore attached to the 3' end, were attached to gold particles by a thiol group at the 5'end. The gold nanoparticles quenched the fluorescence of the attached ODN until hybridization with the target RNA. Once hybridized, the fluorescent dyes moved further from the gold, so fluorescence was increased (Jayagopal, Halfpenny et al. 2010).

Aptamer tags

Aptamer tags have been used to create detectable signals such as changes in fluorescence after binding with ligands (Xu and Lu 2010). For example, the fluorescence of malachite green is increased more than 2,000-fold when it is bound to the malachite green aptamer (Babendure, Adams et al. 2003). This feature could allow it to be used as a tag to examine the formation, location, or degradation of RNA. However, malachite green produces free radicals upon irradiation. As this property eventually damages the RNA binding motif, the destruction decreases the fluorescent signal, causing photodestruction of the RNA (Stojanovic and Kolpashchikov 2004). Malachite green is also highly toxic to cells (Kraus, Jeon et al. 2008).

Another means of obtaining a signal from an RNA aptamer is the photoinduced electron transfer (PET) system. This system consists of two linked chemical moieties, a fluorophore and a quencher. When a fluorophore is excited, decay of the electron from an excited state to ground state is accompanied by release of energy via fluorescence. If a quencher is near the fluorophore, PET occurs from the quencher to the excited fluorophore, the excited electron decays by nonradiative relaxation, and thus the fluorescence signal is decreased. As the selected RNA aptamer binds to the quencher of a PET sensor, the increased quencher-fluorophore distance suppresses the PET, resulting in fluorescence enhancement. The cell-permeable PET sensor contains a dichlorofluorescein as the fluorophore and N-(p-methoxyphenyl)piperazine (MPP) as the quencher, showing weak fluorescence due to PET between the fluorophore and the quencher. When MPP binds the aptamers, the PET sensor has stronger fluorescence by reduced PET efficiency (Sparano and Koide 2005; Sparano and Koide 2007). However, we have found that MPP is highly insoluble in an aqueous environment and thus not likely to be useful for cell imaging.

Light-up RNA tag is another RNA tag with a fluorescent output. For this tag, the chemical structure of Hoechst 33258 was modified to suppress its binding to AT-rich dsDNA. An RNA aptamer was selected that recognized the modified Hoechst. The fluorescence intensity increased when the aptamers bound the modified Hoechst. The authors demonstrated the efficacy of the tag by showing that the fluorescent signal from Hoechst was enhanced by luciferase mRNA (synthesized by in vitro transcription) that was fused with five tandem light-up tags. There was no increase in the Hoechst fluorescence signal in the presence of control luciferase mRNA. The in vitro study demonstrated an enhanced fluorescence signal due to transcribed RNA aptamers binding with the modified Hoechst. Thus, the RNA aptamer light-up tag in the presence of the dye was proposed as a means of monitoring RNA transcription in vitro (Sando, Narita et al. 2008).

Recently, Spinach RNA, which binds to the fluorophore of GFP (4-hydroxybenzlidene imidazolinone: HBI), was used to image ribosomal RNA in a cell culture system. Excited HBI

dissipates its energy by molecular vibrations. However, HBI fluorescence is enhanced by suppression

of its molecular motion when in the context of maturated GFP or when bound with an aptamer that

recognizes it. HBI derivatives were found to fluoresce after binding with aptamers *in vitro* whereas

there was no fluorescence with HBI derivatives and control RNA. Each HBI derivative bound to

aptamers has a different spectrum of fluorescence, which allows them to be used like different colors

of fluorescent proteins. Spinach RNA was used to tag 5S rRNA and the cells imaged to localize the

rRNA in HEK293 T cells. The authors propose that spinach RNA might be used to track RNA

dynamics in living cells (Zimmer and Baffour-Awuah 2004; Paige, Wu et al. 2011).

The advantages of the aptamer tags over GFP tags include a shorter time period between

transcription and visualization of the fluorescence signal than for GFP maturation. The advantages of

the aptamer tags over ODN probes are better dye permeability and less toxic than fluorescently

labeled ODNs (Raj, van den Bogaard et al. 2008). However, if the aptamer ligand binds

nonspecifically to intracellular molecules then this could increase the background.

IMAGEtag (Intracellular MultiAptamer GEnetic tag)

The proposed IMAGEtag system is a RNA reporter for imaging to detect promoter activity

by FRET in real-time *in vivo*. It can detect transcriptional events faster than reporter proteins such as

GFP after stimulating promoter. For example, to detect promoter activity by GFP, there is ~90 lag

time for protein synthesis and maturation (Li, Wang et al. 2000; Hackett, Esch et al. 2006).

IMAGEtags are composed of strings of RNA aptamers that can be expressed from a promoter

of choice. For imaging, IMAGEtags are transcribed after the promoter has been activated, and the

transcribed IMAGEtags bind with fluorescently labeled ligands, resulting in a fluorescent signal. To

increase signal to noise, a FRET system is used. IMAGEtags can be transcribed by cells and their

fluorescently labeled ligands enter the cells by passive transport. Thus, there is no need to deliver

RNA reporters such as is necessary for ODN probes and, if the IMAGEtag expression constructs are

stably expressed then they are not lost by dilution with cell division that is a problem with ODN probes. In addition, IMAGEtag can be observed in a large number of cells in each experiment, whereas ODN probes are delivered by microinjection, which limits them to a small number of cells.

IMAGEtag RNA also has relatively short half-life in yeast, so the FRET signal from IMAGEtags is transient. In the MS2 system, fluorescent signal from GFP remains long after degradation of RNA and therefore it gives a false signal.

Another advantage of the IMAGEtag system is that the IMAGEtags are short RNAs. Therefore they pose less metabolic burden than do protein reporters. For example, to see a signal from the MS2 system, both MS2-GFP and MS2 RNA must be synthesized, causing additional metabolic burden on the cells. A metabolic burden due to the synthesis of expressed recombinant protein has been demonstrated in in *E. coli* in which effects were observed in carbon metabolism and cell growth (Heyland, Blank et al. 2011).

Aminoglycoside antibiotics and aptamers

Aminoglycoside antibiotics

Aminoglycosides are important antibiotics for treating gram-negative bacterial infections. Most aminoglycoside antibiotics are from *Streptomyces* and *Actimonyces* strains or are chemically modified derivatives from these strains. They consist of one or more amino sugars attached to an aminocyclitol moiety by a glycosidic bond. Aminoglycosides have anti-bacterial activity due to their inhibition of a prokaryotic protein synthesis. They are polar basic compounds that are positively charged at physiological pH. The compounds bind to various intracellular and membrane anionic molecules such as DNA, RNA, and phospholipids (Jin, Katritch et al. 2000).

The highly polar aminoglycosides do not easily cross the bacterial membrane. Being cationic, they bind to the outer negatively charged lipopolysaccharides and may enter the

periplasmatic compartment through an aqueous channel formed by porin-type proteins (Silva and Carvalho 2007). Aminoglycoside transport to the cytoplasm is believed to depend on electron transport, because the membrane electric potential (interior negative) is a driving force for aminoglycoside entry. The transport of aminoglycosides is blocked or inhibited by divalent ions such as Ca^{2+} and Mg^{2+}, hyperosmolarity, decrease of pH, and anaerobiosis (Brunton 2006).

Aminoglycosides selectively bind to the conserved sequence of the 16S ribosomal RNA (rRNA) A-site. Two adenine residues (A1492 and A1493 in *Escherichia coli*) of the A-site are involved in a drug-induced conformational change of the rRNA. Specificity for bacterial ribosomes over eukaryotic ribosomes is due to a sequence difference in the A-site at the position 1408 (*E. coli* numbering) with adenine in prokaryotes and guanine in eukaryotes. Eukaryotic ribosomes that contain a guanine at 1408 instead of adenine are resistant to aminoglycosides, and mutation A1408G in prokaryotes ribosomes also results in resistance to aminoglycosides. These binding affinities were determined by isothermal titration calorimetry using the *E. coli* 16S rRNA A-site model oligonucleotide and the mutant oligonucleotide (A1408G) to paromomycin. Paromomycin binds to 16S rRNA ~31-fold higher than to the A1408G variant by enthalpy-driven binding (Kaul, Barbieri et al. 2004).

Aminoglycosides also bind to the Rev Responsive Element (RRE) and Trans Activation Responsive Element (TAR) HIV RNA and compete with natural ligands *Rev* and *Tat* proteins, preventing replication of HIV in the cell culture. For this reason, they are potentially useful as antiviral agents (Jin, Katritch et al. 2000; Silva and Carvalho 2007).

Tobramycin

Tobramycin is a broad spectrum aminoglycoside antibiotic produced by *Streptomyces tenebrarius* and is used for treatment of a gram-negative infection. Chemical analysis of tobramycin is difficult because tobramycin does not contain a UV-absorbing chromophore. Therefore, liquid

chromatography combined with amperometry, mass spectrometry, pulsed electrochemical detection, or evaporative light scattering detection is used to detect tobramycin (Chopra, Vanderheyden et al. 2010). To detect tobramycin by UV absorption, it is necessary to introduce a chromophore group to the molecule (Guo, Wrisley et al. 2006).

Tobramycin is a polyamino compound containing five primary amino groups. In Figure 2, the pKa of each amino group is indicated. The 6' amino group has the highest pKa of the five amino groups as reported from pH-dependent nuclear magnetic resonance (NMR) studies. The reaction of tobramycin with the N-hydroxysuccinimidyl ester of the 2,4-dinitrophenylthioglycolic acid results in a mono-acylated product. Only the 6' amino group of tobramycin is involved in monoacylation in accord with the mass spectral fragmentation of the product (Singh, Pirio et al. 1984; Jin, Katritch et al. 2000).

Aptamer

Ligands

Aptamers are single-stranded DNAs or RNAs that originate from an *in vitro* selection method called SELEX (Systematic Evolution of Ligands by EXponential enrichment), showing high affinity and specificity for their ligands. Aptamers are predominantly unstructured in solution but fold upon associating with their ligands. Ligands of aptamers are diverse from small molecules such as aromatic ligands, amino acids, and oligosaccharides to larger molecules including peptides, proteins, and whole cells. Aptamers have generally been found to bind to their targets by an induced fit mechanism. Intermolecular hydrogen bonding and stacking interactions contribute to ligand binding. Aptamers frequently bind with high specificity. For example, the theophylline aptamer discriminates between theophylline and caffeine in which the former has hydrogen in place of the methyl group at N7. Steric hindrance due to the presence of the N7 methyl group interferes with binding of caffeine to the theophylline aptamer (Hermann and Patel 2000).

SELEX (Systematic Evolution of Ligands by EXponential enrichment)

In 1990, two research groups independently developed a method of selecting nucleic acids *in vitro* to bind their target molecules with high specificity. The Ellington and Szostack group selected RNA sequences bound to seven different organic dyes, whereas the Tuerk and Gold group selected two RNAs to bind to T4 DNA polymerase. SELEX can be divided into three steps. It starts with synthesizing a randomized nucleotide pool in which the random sequence is flanked by sequences at both ends for hybridizing PCR primers. To synthesize a random pool, an equimolar mixture of four bases is used in an automated synthesizer. The length of the random sequence generally is 20–60 bases, creating the potential for very large combinatorial libraries of which 10^{12}–10^{15} sequences can generally be screened. The second step in SELEX is selection, in which the pool of RNA or DNA molecules is applied for binding with the selected ligand and the bound sequences are separated from the unbound by a means such as affinity chromatography or nitrocellulose filtration. Negative selections are also performed against the affinity matrix. The last step is amplification of the selected sequences. The bound sequences are amplified and prepared for the next round by either PCR followed by single strand selection for DNA aptamers or reverse transcription followed by PCR and *in vitro* transcription for selecting RNA aptamers. Newly selected and amplified sequences are expected to be enriched for oligonucleotides with higher affinities for the target and are used for starting the next round. Multiple rounds (usually 8–15) of selection are carried out until a large portion of the pool can be demonstrated to bind the targets (Ellington and Szostak 1990; Tuerk and Gold 1990; Syed and Pervaiz 2010).

Whereas DNA is quite stable under most conditions, RNA can be more rapidly degraded or hydrolyzed. To increase the stability of an RNA aptamer for applications, several modifications can be made. Examples of modifications are substitutions in the ribose sugar residues, such as 2'-F, 2'-NH$_2$, 2'-O-methyl, 2'-O-methoxyl, and 2'-O-dimethylallyl; in the phosphate backbone, such as phosphorothioate and methyl phosphonate; or in the nitrogen base, including propenyl, 5-(N-

aminoalkyl) carbamoyluracil, methyl, trifluoromethyl, phenyl, and 2-thiopyrimidine. Incorporation of modified nucleotides during SELEX requires the use of mutant T7 polymerases that will utilize these NTPs by *in vitro* transcription. These modifications minimize degradation of RNA oligonucleotides by nucleases (Bunka, Platonova et al. 2010; Syed and Pervaiz 2010).

Application of aptamers

Aptamers are used in a broad range of applications in therapeutics, diagnostics, biosensors, cell tracking, and purification processes.

Aptamers are good candidates for therapeutics due to their high specificity, easy production and modification, and low immunogenicity. One aptamer is already being used as a therapeutic agent that has been approved by the U.S. Food and Drug Administration, the European Union and others are being tested in clinical trials. Pegaptanib (Macugen, marketed by Pfizer) is an RNA aptamer that binds to and antagonizes the action of vascular endothelial growth factor (VEGF) by blocking VEGF binding to the receptor. It is used for treatment of age-related macular degeneration (Cerchia and de Franciscis 2010; Keefe, Pai et al. 2010). The prostate specific membrane antigen (PSMA) is a biomarker of prostate cancer. The anti-PSMA aptamer has been demonstrated capable of drug delivery of an anti-cancer drug in studies in which the aptamer was conjugated with anti-cancer drugs, such as cisplatin, or siRNAs targeting the survival factors Plk-1 and Bcl-2 (Bunka, Platonova et al. 2010; Syed and Pervaiz 2010). Aptamers against the blood-clotting factor IXa RB006 and its complementary sequence antidote RB007 are being tested as an anticoagulation system in a phase II clinical trial (Bunka, Platonova et al. 2010).

The small size of aptamers is advantageous for accessing their targets *in vivo*, but this also results in rapid renal clearance. To extend retention time in the system, aptamers are conjugated with PEG to increase molecular weight (Syed and Pervaiz 2010).

Aptamers can also be selected against whole cells by using Cell-SELEX. A justification for this approach is that aptamers selected against the purified target might not recognize the same target in its native conformation in a cellular context. For example, a highly glycosylated protein such as full-length EGFRvIII target cannot be recognized by an RNA aptamer selected against histidine-tagged EGFRvIII ectodomain produced by a bacterial expression system (Liu, Kuan et al. 2009). Hence, aptamers are selected by Cell-SELEX to recognize complex targets in the cellular context. Aptamers selected against tumor cell surface markers by Cell-SELEX are more likely to be useful for diagnostics compared with those selected against the equivalent recombinant proteins (Cerchia and de Franciscis 2010; Syed and Pervaiz 2010). .

Aptamers often undergo conformational changes upon binding their ligands. These motions enable the integration of aptamers in biosensors that use optical, electrochemical, or mass-based modalities for detecting the conformational change.

Electrochemical detections have been used to develop aptamer-based biosensors. These are electrochemical impedance spectroscopy, potentiometry with ion-selective electrodes, electrogenerated chemiluminescence, cyclic voltammetry, and amperometry. Electroactive reporters such as methylene blue have been used. Here signals are detected by the electrodes' redox state of methylene blue. Methylene blue has been conjugated with DNA aptamers that recognize cocaine or platelet-derived growth factor (PDGF). Due to the conformational change in the aptamer upon binding ligand, the distance between methylene blue and the electrode is different in the absence or presence of target molecules, thus the redox state of methylene blue could be used to detect aptamer-ligand binding (Cho, Lee et al. 2009). For example, anti VEGF aptamer has been reported as a folding-based electrochemical DNA aptasensor to detect of VEGF directly in blood serum and whole blood. Methylene blue-modified and surface-immobilized anti-VEGF aptamer changes its conformation upon binding VEGF and brings methylene blue close to electrode, which results in an increase in the observed current (Zhao, Yang et al. 2011).

The ability to detect the conformational change of an aptamer upon binding ligand can also be built into the aptamer. For example, if an organic fluorophore is introduced into the binding region of aptamers, ligand binding can change the local environment of the fluorophore, resulting in changes in the fluorescence characteristics of the aptamer, such as its intensity and anisotropy. The quantum yield of base analogs, such as 2-aminopurine (adenine analog), 4-amino-6-methylpteridone (adenine analog), and 3-methylisoxanthopterine (guanine analog) is highly dependent on the local environment, especially the degree of base-stacking. The analogs are quenched when they are involved in base-paired stacking, but fluorescence is increased when they are part of a single-stranded region. For example, the fluorescence intensity DNA aptamers against thrombin, immunoglobulin E, and PDGF with incorporated base analogs increased up to 30-fold after binding with their target molecules (Cho, Lee et al. 2009).

Modifications of the 3' and 5' ends of aptamers with a fluorophore and quencher pair can also signal a conformational change in an aptamer. These modified aptamers, called aptamer beacons, can be monitored for ligand binding by fluorescence (Li, Fang et al. 2002). In addition, flexibility of single-stranded RNA structure was also applied an aptamer-based biosensor, called TRAP (targeted reversibly attenuated probe). TRAP is composed of three parts, inactive aptamer, intervening antisense and attenuator (cis-complementary regulatory nucleic acid). The intervening sequence is complementary to an mRNA or DNA (regulatory oligonuceotide).   In the absence of regulatory oligonucleotide, the aptamer is hybridized with attenuator and is inactive. When the regulatory sequence is present and hybridizes with the intervening sequence, the aptamer is separated from the attenuator and becomes active (Cong and Nilsen-Hamilton 2005). The ability of Lcn2 mRNA to activate a malachite green aptamer TRAP, which contains part of anti Lcn2 mRNA sequence, can be measured by the increase in fluorescence of malachite green after binding with the aptamer (Cong 2006).

Gold nanoparticles (GNPs) have also used in association with aptamers to create biosensors because the color changes depending on the aggregation state of the GNP. The color is purple in the aggregated state, while it is red in the disaggregated state. An aptamer against ATP was hybridized to a complementary ODN that in turn was attached to a GNP. The highly negatively charged phosphate backbones prevent aggregation of the GNP. Binding of ATP displaces the aptamers from the GNPs, and in turn the GNPs aggregate, changing the color from red to purple (Cho, Lee et al. 2009).

Tobramycin aptamer

Tobramycin aptamers were first selected by the Rando group. Two consensus sequences are the binding region of the aptamers and form a stem-loop structure (Wang and Rando 1995). The solution structures of tobramycin bound by two stem-loop RNA aptamers were solved by NMR (Figure 3). The binding pattern between tobramycin and the two RNA aptamers is similar. Tobramycin binds site-specifically to the RNA major groove and is encapsulated between the floor of the major groove and the loop-out cytosine for RNA aptamer I or guanine for RNA aptamer II, which is positioned as a flap in the binding pocket. The tobramycin binding site includes the hairpin loop of the RNA aptamer, which has a specific sequence and bends around the ligand. The specific sequence of the loop is UUARNU, where R has to be a purine and N can be any base. The RNA major groove is widened by a one base bulge in RNA aptamer I or one G·A and two G·U mismatches in RNA aptamer II. The non-sugar ring I and amino sugar ring III of tobramycin are involved in binding to the aptamers by intermolecular interactions such as hydrogen bonding. The non-sugar ring I interacts with the major groove of the RNA, and the amino sugar ring III is encapsulated by a flap, either cytosine in RNA aptamer I or guanine in RNA aptamer II. The amino sugar ring II is involved in the binding to a less extent than the other rings. The 6'-amino group from sugar ring II is directed toward the solvent (Jiang, Suri et al. 1997; Jiang and Patel 1998; Jin, Katritch et al. 2000). Hence, tobramycin that is modified on the 6'-amino group on sugar ring II can still bind with the tobramycin

RNA aptamer. These results suggested that, for the imaging described in this dissertation, tobramycin could be modified with fluorescent dyes at 6' amino group and the product would still bind the tobramycin aptamer.
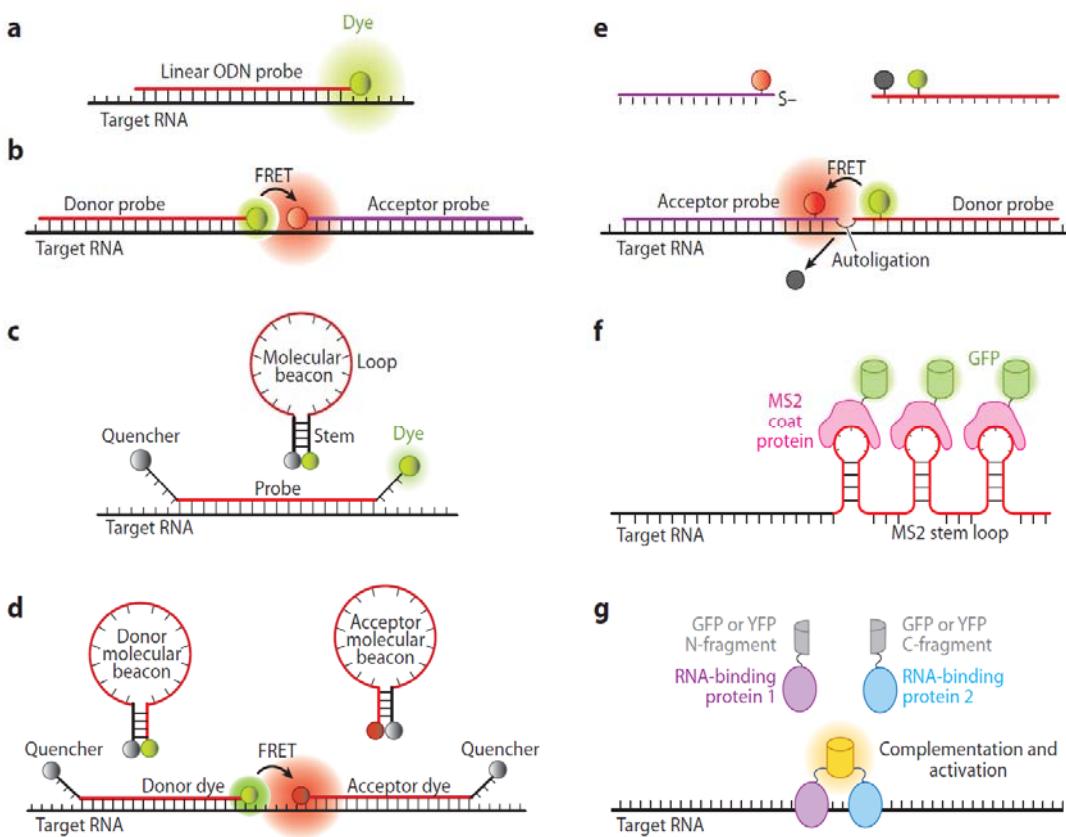
Figure 1, Illustration of fluorescent probes developed for live-cell RNA detection. Figures taken

from (Bao, Rhee et al. 2009). Republished with permission. (a) Linear fluorescence oligonucleotide

probes; (b) Linear FRET probes; (c) Molecular beacons; (d) Dual FRET molecular beacons; (e)

Autoligation FRET probes; (f) MS2-system; (g) split GFP probes.

a

pKa-1 = 7.4
pKa-3 = 6.2
pKa-2' = 7.6
pKa-6' = 8.6
pKa-3"= 7.4

b

c
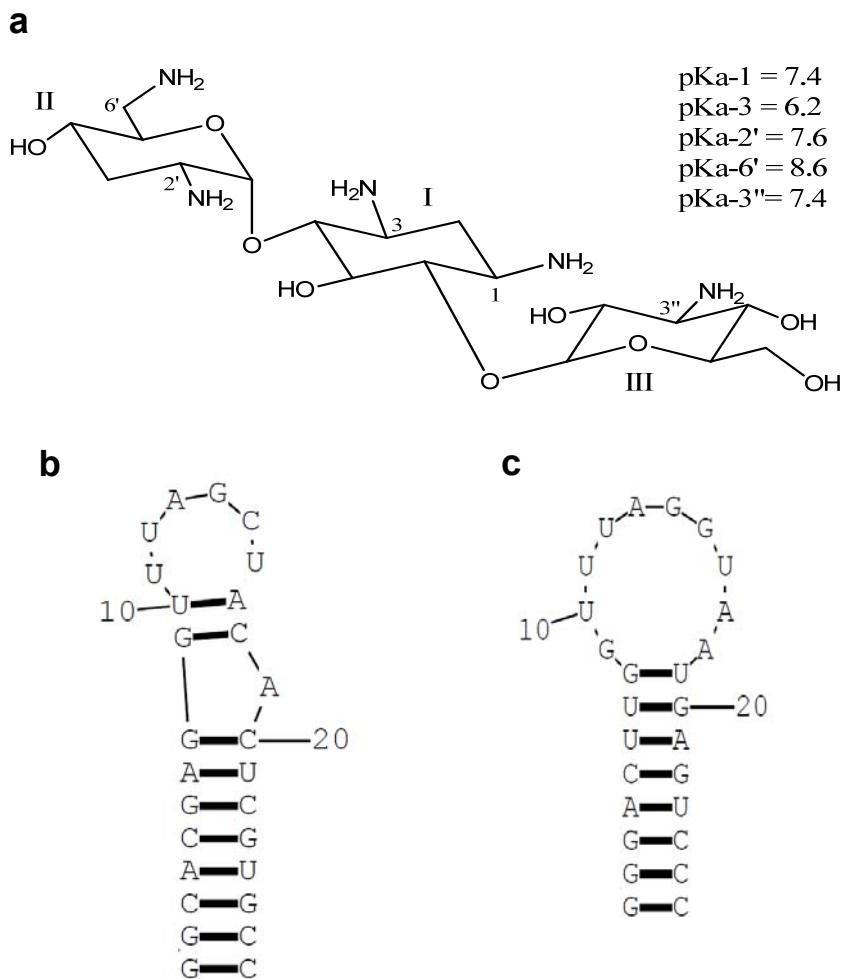
Figure 2, The chemical structure of tobramycin and secondary structure of two RNA aptamers.

(a) Chemical structure of tobramycin in which the atomic and ring numbering systems are denoted in Arabic and Roman numerals, respectively. The NMR-driven pKa values of the five amino groups are shown (Jin, Katritch et al. 2000). (b) Secondary structure of 27 nt tobramycin RNA aptamer (RNA aptamer I) (Jiang, Suri et al. 1997). (c) Secondary structure of 26 nt tobramycin RNA aptamer (RNA aptamer II) (Jiang and Patel 1998).
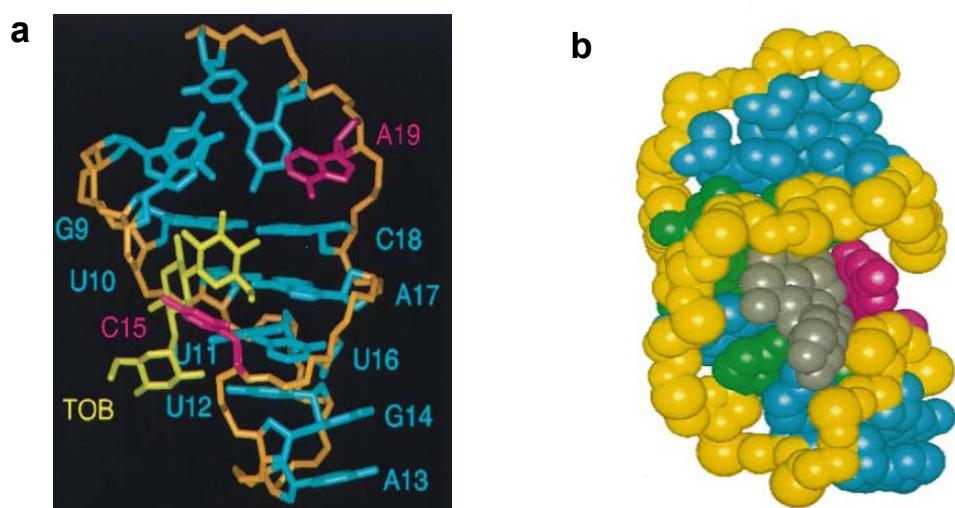
Figure 3, Solution NMR structure of the tobramycin-RNA complex. (a) Figures taken from (Jin, Katritch et al. 2000). A view looking into the major groove of the core of the complex of tobramycin and RNA aptamer I. (b) Figures taken from (Jiang and Patel 1998). A space-filling view of a refined structure of the complex of tobramycin and RNA aptamer II. The RNA backbone is in yellow, and the bound tobramycin is in grey.

References

Acar, M., A. Becskei, et al. (2005). "Enhancement of cellular memory by reducing stochastic transitions." Nature **435**(7039): 228-32.

Adelman, K., M. T. Marr, et al. (2005). "Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS." Mol Cell **17**(1): 103-12.

Ahn, S. H., M. C. Keogh, et al. (2009). "Ctk1 promotes dissociation of basal transcription factors from elongating RNA polymerase II." EMBO J **28**(3): 205-12.

Andersen, J. B., C. Sternberg, et al. (1998). "New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria." Appl Environ Microbiol **64**(6): 2240-6.

Babendure, J. R., S. R. Adams, et al. (2003). "Aptamers switch on fluorescence of triphenylmethane dyes." J Am Chem Soc **125**(48): 14716-7.

Bao, G., W. J. Rhee, et al. (2009). "Fluorescent probes for live-cell RNA detection." Annu Rev Biomed Eng **11**: 25-47.

Bell, O., C. Wirbelauer, et al. (2007). "Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in Drosophila." EMBO J **26**(24): 4974-84.

Bertrand, E., P. Chartrand, et al. (1998). "Localization of ASH1 mRNA particles in living yeast." Mol Cell **2**(4): 437-45.

Brody, Y., N. Neufeld, et al. (2011). "The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing." PLoS Biol **9**(1): e1000573.

Brunton, L. L., Lazo, J.S and Parker, K.L (2006). Goodman and Gilman's The Pharmacological Basis of Therapeutics, McGRAW-HILL

Bunka, D. H., O. Platonova, et al. (2010). "Development of aptamer therapeutics." Curr Opin Pharmacol **10**(5): 557-62.

Burckin, T., R. Nagel, et al. (2005). "Exploring functional relationships between components of the gene expression machinery." Nat Struct Mol Biol **12**(2): 175-82.

Cerchia, L. and V. de Franciscis (2010). "Targeting cancer cells with nucleic acid aptamers." Trends Biotechnol **28**(10): 517-25.

Chen, A. K., M. A. Behlke, et al. (2007). "Avoiding false-positive signals with nuclease-vulnerable molecular beacons in single living cells." Nucleic Acids Res **35**(16): e105.

Chen, A. K., O. Davydenko, et al. (2010). "Ratiometric bimolecular beacons for the sensitive detection of RNA in single living cells." Nucleic Acids Res **38**(14): e148.

Cheung, A. C. and P. Cramer (2011). "Structural basis of RNA polymerase II backtracking, arrest and reactivation." Nature **471**(7337): 249-53.

Cho, E. J., J. W. Lee, et al. (2009). "Applications of aptamers as sensors." Annu Rev Anal Chem (Palo Alto Calif) **2**: 241-64.

Chopra, S., G. Vanderheyden, et al. (2010). "Comparative study on the analytical performance of different detectors for the liquid chromatographic analysis of tobramycin." J Pharm Biomed Anal **53**(2): 151-7.

Chudakov, D. M., M. V. Matz, et al. (2010). "Fluorescent proteins and their applications in imaging living cells and tissues." Physiol Rev **90**(3): 1103-63.

Close, D. M., R. E. Hahn, et al. (2011). "Comparison of human optimized bacterial luciferase, firefly luciferase, and green fluorescent protein for continuous imaging of cell culture and animal models." J Biomed Opt **16**(4): 047003.

Close, D. M., S. S. Patterson, et al. (2010). "Autonomous bioluminescent expression of the bacterial luciferase gene cassette (lux) in a mammalian cell line." PLoS One **5**(8): e12441.

Cong, X. (2006). Development of aptamer based targeted reversibly attenuated probes. Biochemistry, Biophysics and Molecular Biology. Ames, Iowa State University. Doctor of Philosophy: 261.

Cong, X. and M. Nilsen-Hamilton (2005). "Allosteric aptamers: targeted reversibly attenuated probes." Biochemistry **44**(22): 7945-54.

Cosgrove, M. S., J. D. Boeke, et al. (2004). "Regulated nucleosome mobility and the histone code." Nat Struct Mol Biol **11**(11): 1037-43.

Cramer, P., D. A. Bushnell, et al. (2000). "Architecture of RNA polymerase II and implications for the transcription mechanism." Science **288**(5466): 640-9.

David, C. J., A. R. Boyne, et al. (2011). "The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex." Genes Dev **25**(9): 972-83.

Dillon, S. C., X. Zhang, et al. (2005). "The SET-domain protein superfamily: protein lysine methyltransferases." Genome Biol **6**(8): 227.

Domecq, C., M. Kireeva, et al. (2010). "Site-directed mutagenesis, purification and assay of Saccharomyces cerevisiae RNA polymerase II." Protein Expr Purif **69**(1): 83-90.

Dong, D., X. Shao, et al. (2010). "Gene expression variations are predictive for stochastic noise." Nucleic Acids Res **39**(2): 403-13.

Ellington, A. D. and J. W. Szostak (1990). "In vitro selection of RNA molecules that bind specific ligands." Nature **346**(6287): 818-22.

Faria, M., D. G. Spiller, et al. (2001). "Phosphoramidate oligonucleotides as potent antisense molecules in cells and in vivo." Nat Biotechnol **19**(1): 40-4.

Fuda, N. J., M. B. Ardehali, et al. (2009). "Defining mechanisms that regulate RNA polymerase II transcription in vivo." Nature **461**(7261): 186-92.

Ghim, C. M., S. K. Lee, et al. (2010). "The art of reporter proteins in science: past, present and future applications." BMB Rep **43**(7): 451-60.

Gornemann, J., K. M. Kotovic, et al. (2005). "Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex." Mol Cell **19**(1): 53-63.

Gross, S. and D. Piwnica-Worms (2005). "Spying on cancer: molecular imaging in vivo with genetically encoded reporters." Cancer Cell **7**(1): 5-15.

Guo, M. X., L. Wrisley, et al. (2006). "Measurement of tobramycin by reversed-phase high-performance liquid chromatography with mass spectrometry detection." Anal Chim Acta **571**(1): 12-6.

Guo, Z. and F. Sherman (1995). "3'-end-forming signals of yeast mRNA." Mol Cell Biol **15**(11): 5983-90.

Hackett, E. A., R. K. Esch, et al. (2006). "A family of destabilized cyan fluorescent proteins as transcriptional reporters in S. cerevisiae." Yeast **23**(5): 333-49.

Hendrix, D. A., J. W. Hong, et al. (2008). "Promoter elements associated with RNA Pol II stalling in the Drosophila embryo." Proc Natl Acad Sci U S A **105**(22): 7762-7.

Hermann, T. and D. J. Patel (2000). "Adaptive recognition by nucleic acid aptamers." Science **287**(5454): 820-5.

Heyland, J., L. M. Blank, et al. (2011). "Quantification of metabolic limitations during recombinant protein production in Escherichia coli." J Biotechnol **155**(2): 178-184.

Hieb, A. R., S. Baran, et al. (2006). "An 8 nt RNA triggers a rate-limiting shift of RNA polymerase II complexes into elongation." EMBO J **25**(13): 3100-9.

Hirose, Y. and Y. Ohkuma (2007). "Phosphorylation of the C-terminal domain of RNA polymerase II plays central roles in the integrated events of eucaryotic gene expression." J Biochem **141**(5): 601-8.

Ho, C. K. and S. Shuman (1999). "Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme." Mol Cell **3**(3): 405-11.

Holstege, F. C., U. Fiedler, et al. (1997). "Three transitions in the RNA polymerase II transcription complex during initiation." <u>EMBO J</u> **16**(24): 7468-80.

Itzkovitz, S. and A. van Oudenaarden (2011). "Validating transcripts with probes and imaging technology." <u>Nat Methods</u> **8**(4 Suppl): S12-9.

Jayagopal, A., K. C. Halfpenny, et al. (2010). "Hairpin DNA-functionalized gold colloids for the imaging of mRNA in live cells." <u>J Am Chem Soc</u> **132**(28): 9789-96.

Jiang, F., B. R. Frey, et al. (2009). "Gene activation by dissociation of an inhibitor from a transcriptional activation domain." <u>Mol Cell Biol</u> **29**(20): 5604-10.

Jiang, L. and D. J. Patel (1998). "Solution structure of the tobramycin-RNA aptamer complex." <u>Nat Struct Biol</u> **5**(9): 769-74.

Jiang, L., A. K. Suri, et al. (1997). "Saccharide-RNA recognition in an aminoglycoside antibiotic-RNA aptamer complex." <u>Chem Biol</u> **4**(1): 35-50.

Jin, E., V. Katritch, et al. (2000). "Aminoglycoside binding in the major groove of duplex RNA: the thermodynamic and electrostatic forces that govern recognition." <u>J Mol Biol</u> **298**(1): 95-110.

Kaul, M., C. M. Barbieri, et al. (2004). "Fluorescence-based approach for detecting and characterizing antibiotic-induced conformational changes in ribosomal RNA: comparing aminoglycoside binding to prokaryotic and eukaryotic ribosomal RNA sequences." <u>J Am Chem Soc</u> **126**(11): 3447-53.

Keefe, A. D., S. Pai, et al. (2010). "Aptamers as therapeutics." <u>Nat Rev Drug Discov</u> **9**(7): 537-50.

Koch, F., F. Jourquin, et al. (2008). "Genome-wide RNA polymerase II: not genes only!" <u>Trends Biochem Sci</u> **33**(6): 265-73.

Kostrewa, D., M. E. Zeller, et al. (2009). "RNA polymerase II-TFIIB structure and mechanism of transcription initiation." <u>Nature</u> **462**(7271): 323-30.

Kraus, G. A., I. Jeon, et al. (2008). "Fluorinated analogs of malachite green: synthesis and toxicity." <u>Molecules</u> **13**(4): 986-94.

Kubota, T., S. Ikeda, et al. (2010). "Sets of RNA repeated tags and hybridization-sensitive fluorescent probes for distinct images of RNA in a living cell." <u>PLoS One</u> **5**(9): e13003.

Kuehner, J. N., E. L. Pearson, et al. (2011). "Unravelling the means to an end: RNA polymerase II transcription termination." <u>Nat Rev Mol Cell Biol</u> **12**(5): 283-94.

Kugel, J. F. and J. A. Goodrich (1998). "Promoter escape limits the rate of RNA polymerase II transcription and is enhanced by TFIIE, TFIIH, and ATP on negatively supercoiled DNA." <u>Proc Natl Acad Sci U S A</u> **95**(16): 9232-7.

Kundu, S. and C. L. Peterson (2010). "Dominant role for signal transduction in the transcriptional memory of yeast GAL genes." <u>Mol Cell Biol</u> **30**(10): 2330-40.

Larson, D. R., D. Zenklusen, et al. (2011). "Real-time observation of transcription initiation and elongation on an endogenous yeast gene." <u>Science</u> **332**(6028): 475-8.

Lenasi, T., B. M. Peterlin, et al. (2011). "Cap-binding Protein Complex Links Pre-mRNA Capping to Transcription Elongation and Alternative Splicing through Positive Transcription Elongation Factor b (P-TEFb)." <u>J Biol Chem</u> **286**(26): 22758-68.

Lewin, B. (2004). <u>Genes VIII</u>, Pearson Education, Inc.

Li, B., M. Carey, et al. (2007). "The role of chromatin during transcription." <u>Cell</u> **128**(4): 707-19.

Li, J., S. Wang, et al. (2000). "Green fluorescent protein in Saccharomyces cerevisiae: real-time studies of the GAL1 promoter." <u>Biotechnol Bioeng</u> **70**(2): 187-96.

Li, J. J., X. Fang, et al. (2002). "Molecular aptamer beacons for real-time protein recognition." <u>Biochem Biophys Res Commun</u> **292**(1): 31-40.

Lifland, A. W., C. Zurla, et al. (2010). "Single Molecule Sensitive Multivalent Polyethylene Glycol Probes for RNA Imaging." <u>Bioconjug Chem</u> **21**(3): 483-488.

Liu, X., D. A. Bushnell, et al. (2010). "Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism." <u>Science</u> **327**(5962): 206-9.

Liu, Y., C. T. Kuan, et al. (2009). "Aptamers selected against the unglycosylated EGFRvIII ectodomain and delivered intracellularly reduce membrane-bound EGFRvIII and induce apoptosis." Biol Chem 390(2): 137-44.

Lunde, B. M., S. L. Reichow, et al. (2010). "Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain." Nat Struct Mol Biol 17(10): 1195-201.

Malagon, F., M. L. Kireeva, et al. (2006). "Mutations in the Saccharomyces cerevisiae RPB1 gene conferring hypersensitivity to 6-azauracil." Genetics 172(4): 2201-9.

Mandal, S. S., C. Chu, et al. (2004). "Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II." Proc Natl Acad Sci U S A 101(20): 7572-7.

Martin, C. and Y. Zhang (2005). "The diverse functions of histone lysine methylation." Nat Rev Mol Cell Biol 6(11): 838-49.

Mason, P. B. and K. Struhl (2005). "Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo." Mol Cell 17(6): 831-40.

McCracken, S., N. Fong, et al. (1997). "The C-terminal domain of RNA polymerase II couples mRNA processing to transcription." Nature 385(6614): 357-61.

Millevoi, S. and S. Vagner (2009). "Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation." Nucleic Acids Res 38(9): 2757-74.

Moore, M. J. and P. A. Sharp (1993). "Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing." Nature 365(6444): 364-8.

Moreland, R. J., F. Tirode, et al. (1999). "A role for the TFIIH XPB DNA helicase in promoter escape by RNA polymerase II." J Biol Chem 274(32): 22127-30.

Nagai, T., K. Ibata, et al. (2002). "A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications." Nat Biotechnol 20(1): 87-90.

Natarajan, A., S. Subramanian, et al. (1998). "Comparison of mutant forms of the green fluorescent protein as expression markers in Chinese hamster ovary (CHO) and Saccharomyces cerevisiae cells." J Biotechnol 62(1): 29-45.

Nechaev, S. and K. Adelman (2011). "Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation." Biochim Biophys Acta 1809(1): 34-45.

Nguyen, A. T. and Y. Zhang (2011). "The diverse functions of Dot1 and H3K79 methylation." Genes Dev 25(13): 1345-58.

Nitin, N., P. J. Santangelo, et al. (2004). "Peptide-linked molecular beacons for efficient delivery and rapid mRNA detection in living cells." Nucleic Acids Res 32(6): e58.

Ohrt, T., D. Merkle, et al. (2006). "In situ fluorescence analysis demonstrates active siRNA exclusion from the nucleus by Exportin 5." Nucleic Acids Res 34(5): 1369-80.

Okabe, K., Y. Harada, et al. (2010). "Real time monitoring of endogenous cytoplasmic mRNA using linear antisense 2'-O-methyl RNA probes in living cells." Nucleic Acids Res 39(4): e20.

Orphanides, G. and D. Reinberg (2002). "A unified theory of gene expression." Cell 108(4): 439-51.

Ozawa, T., Y. Natori, et al. (2007). "Imaging dynamics of endogenous mitochondrial RNA in single living cells." Nat Methods 4(5): 413-9.

Paige, J. S., K. Y. Wu, et al. (2011). "RNA mimics of green fluorescent protein." Science 333(6042): 642-6.

Pal, M. and D. S. Luse (2003). "The initiation-elongation transition: lateral mobility of RNA in RNA polymerase II complexes is greatly reduced at +8/+9 and absent by +23." Proc Natl Acad Sci U S A 100(10): 5700-5.

Peng, J., Y. Zhu, et al. (1998). "Identification of multiple cyclin subunits of human P-TEFb." Genes Dev 12(5): 755-62.

Pokholok, D. K., C. T. Harbison, et al. (2005). "Genome-wide map of nucleosome acetylation and methylation in yeast." Cell 122(4): 517-27.

Powell, W. and D. Reines (1996). "Mutations in the second largest subunit of RNA polymerase II cause 6-azauracil sensitivity in yeast and increased transcriptional arrest in vitro." J Biol Chem 271(12): 6866-73.

Prescher, J. A. and C. H. Contag (2010). "Guided by the light: visualizing biomolecular processes in living animals with bioluminescence." Curr Opin Chem Biol 14(1): 80-9.

Proudfoot, N. J., A. Furger, et al. (2002). "Integrating mRNA processing with transcription." Cell 108(4): 501-12.

Ptashne, M. and A. Gann (1997). "Transcriptional activation by recruitment." Nature 386(6625): 569-77.

Raj, A., P. van den Bogaard, et al. (2008). "Imaging individual mRNA molecules using multiple singly labeled probes." Nat Methods 5(10): 877-9.

Ramsey, S. A., J. J. Smith, et al. (2006). "Dual feedback loops in the GAL regulon suppress cellular heterogeneity in yeast." Nat Genet 38(9): 1082-7.

Roeder, R. G. (1996). "The role of general initiation factors in transcription by RNA polymerase II." Trends Biochem Sci 21(9): 327-35.

Roeder, R. G. (2005). "Transcriptional regulation and the role of diverse coactivators in animal cells." FEBS Lett 579(4): 909-15.

Sando, S., A. Narita, et al. (2008). "Transcription monitoring using fused RNA with a dye-binding light-up aptamer as a tag: a blue fluorescent RNA." Chem Commun (Camb)(33): 3858-60.

Santangelo, P., N. Nitin, et al. (2006). "Nanostructured probes for RNA detection in living cells." Ann Biomed Eng 34(1): 39-50.

Santangelo, P. J., A. W. Lifland, et al. (2009). "Single molecule-sensitive probes for imaging RNA in live cells." Nat Methods 6(5): 347-9.

Santangelo, P. J., B. Nix, et al. (2004). "Dual FRET molecular beacons for mRNA detection in living cells." Nucleic Acids Res 32(6): e57.

Saunders, A., L. J. Core, et al. (2006). "Breaking barriers to transcription elongation." Nat Rev Mol Cell Biol 7(8): 557-67.

Schneider, J., P. Bajwa, et al. (2006). "Rtt109 is required for proper H3K56 acetylation: a chromatin mark associated with the elongating RNA polymerase II." J Biol Chem 281(49): 37270-4.

Sellick, C. A., T. A. Jowitt, et al. (2009). "The effect of ligand binding on the galactokinase activity of yeast Gal1p and its ability to activate transcription." J Biol Chem 284(1): 229-36.

Selth, L. A., S. Sigurdsson, et al. (2010). "Transcript Elongation by RNA Polymerase II." Annu Rev Biochem 79: 271-93.

Silva, J. G. and I. Carvalho (2007). "New insights into aminoglycoside antibiotics and derivatives." Curr Med Chem 14(10): 1101-19.

Singh, P., M. Pirio, et al. (1984). "Controlled Coupling of Aminoglycoside Antibiotics to Proteins for Use in Homogeneous Enzyme Immunoassays." Canadian Journal of Chemistry-Revue Canadienne De Chimie 62(11): 2471-2477.

Smale, S. T. (2010). "Luciferase assay." Cold Spring Harb Protoc 2010(5): pdb prot5421.

Sparano, B. A. and K. Koide (2005). "A strategy for the development of small-molecule-based sensors that strongly fluoresce when bound to a specific RNA." J Am Chem Soc 127(43): 14954-5.

Sparano, B. A. and K. Koide (2007). "Fluorescent sensors for specific RNA: a general paradigm using chemistry and combinatorial biology." J Am Chem Soc 129(15): 4785-94.

Stojanovic, M. N. and D. M. Kolpashchikov (2004). "Modular aptameric sensors." J Am Chem Soc 126(30): 9266-70.

Su, W. Y., H. Xiong, et al. (2010). "Natural antisense transcripts regulate gene expression in an epigenetic manner." Biochem Biophys Res Commun **396**(2): 177-81.

Sweetser, D., M. Nonet, et al. (1987). "Prokaryotic and eukaryotic RNA polymerases have homologous core subunits." Proc Natl Acad Sci U S A **84**(5): 1192-6.

Syed, M. A. and S. Pervaiz (2010). "Advances in aptamers." Oligonucleotides **20**(5): 215-24.

Thoden, J. B., C. A. Sellick, et al. (2005). "Molecular structure of Saccharomyces cerevisiae Gal1p, a bifunctional galactokinase and transcriptional inducer." J Biol Chem **280**(44): 36905-11.

Tsien, R. Y. (1998). "The green fluorescent protein." Annu Rev Biochem **67**: 509-44.

Tuerk, C. and L. Gold (1990). "Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase." Science **249**(4968): 505-10.

Tyagi, S. (2009). "Imaging intracellular RNA distribution and dynamics in living cells." Nat Methods **6**(5): 331-8.

Tyagi, S. and F. R. Kramer (1996). "Molecular beacons: probes that fluoresce upon hybridization." Nat Biotechnol **14**(3): 303-8.

Valencia-Burton, M., R. M. McCullough, et al. (2007). "RNA visualization in live bacterial cells using fluorescent protein complementation." Nat Methods **4**(5): 421-7.

Varv, S., K. Kristjuhan, et al. (2010). "Acetylation of H3 K56 is required for RNA polymerase II transcript elongation through heterochromatin in yeast." Mol Cell Biol **30**(6): 1467-77.

Verma, M., P. J. Bhat, et al. (2003). "Quantitative analysis of GAL genetic switch of Saccharomyces cerevisiae reveals that nucleocytoplasmic shuttling of Gal80p results in a highly sensitive response to galactose." J Biol Chem **278**(49): 48764-9.

von Degenfeld, G., T. S. Wehrman, et al. (2009). "Imaging beta-galactosidase activity in vivo using sequential reporter-enzyme luminescence." Methods Mol Biol **574**: 249-59.

Wang, A., S. K. Kurdistani, et al. (2002). "Requirement of Hos2 histone deacetylase for gene activity in yeast." Science **298**(5597): 1412-4.

Wang, D., D. A. Bushnell, et al. (2006). "Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis." Cell **127**(5): 941-54.

Wang, X., B. Errede, et al. (2008). "Mathematical analysis and quantification of fluorescent proteins as transcriptional reporters." Biophys J **94**(6): 2017-26.

Wang, Y. and R. R. Rando (1995). "Specific binding of aminoglycoside antibiotics to RNA." Chem Biol **2**(5): 281-90.

Weake, V. M. and J. L. Workman (2010). "Inducible gene expression: diverse regulatory mechanisms." Nat Rev Genet **11**(6): 426-37.

Workman, J. L. and R. E. Kingston (1998). "Alteration of nucleosome structure as a mechanism of transcriptional regulation." Annu Rev Biochem **67**: 545-79.

Xu, W. and Y. Lu (2010). "Label-free fluorescent aptamer sensor based on regulation of malachite green fluorescence." Anal Chem **82**(2): 574-8.

Zhang, Y. and L. C. Yu (2008). "Microinjection as a tool of mechanical delivery." Curr Opin Biotechnol **19**(5): 506-10.

Zhao, S., W. Yang, et al. (2011). "A folding-based electrochemical aptasensor for detection of vascular endothelial growth factor in human whole blood." Biosens Bioelectron **26**(5): 2442-7.

Zimmer, M. and N. Y. A. Baffour-Awuah (2004). "Hula-twisting in green fluorescent protein." Chemical Physics **303**(1-2): 7-11.

Chapter 3. REAL-TIME IMAGING OF PROMOTER ACTIVITY TO MONITOR GENE EXPRESSION

Ilchung Shin, Judhajeet Ray, Muslum Ilgu, Vinayak Gupta, Ming Zhao, Leilei Peng,

George Kraus and Marit Nilsen-Hamilton

Abstract

A signaling pathway that is activated by an external stimulus is usually transient. For this reason, real-time detection of promoter activity is useful to determine the progress of changes in gene expression, especially during cell differentiation and development. We describe here an RNA reporter system for imaging changes in promoter activity in real-time. The RNA reporter consists of strings of RNA aptamers that constitute IMAGEtags (Intracellular MultiAptamer Genetic tags), which can be expressed from a promoter of choice. Tobramycin, neomycin and PDC RNA aptamers have been employed in this system and expressed in yeast from the GAL1 promoter. For imaging, the yeast cells are incubated with their ligands labeled with fluorescent dyes. To increase signal to noise, the ligands have been separately conjugated with the FRET pairs, Cy3 and Cy5. With these constructs, the transcribed aptamers can be imaged after activation of the promoter by galactose. Real-time transcription was measured by FLIM -FRET, which was detected by the decrease in donor lifetime resulting from ligand binding to IMAGEtags that were newly synthesized from the activated GAL1 promoter. The FRET signal was specific for transcribed IMAGEtags. Compared with other methods that are available to monitor promoter activity in real time, the IMAGEtag system is more sensitive and more readily applicable to a broad range of cell types.

Introduction

Changes in gene expression are central to most alterations in cellular functions, yet our current means of measuring transcriptional changes in real time are limited. Indirect measurements of reporter protein levels are mostly used as a monitor of transcriptional activity, which requires that the biological samples should be destroyed in the process of analysis and does not allow the monitoring of a single cell or cell population *in situ* and with time. Although the use of fluorescent proteins allows measurement in living cells, all protein reporter systems include a significant lag in time between transcriptional initiation and protein appearance.[1] Although the use of externally supplied oligodeoxynucleotide (ODN) probes has been reported for monitoring transcriptional activity, these methods are plagued with problems such as nuclear sequestration, false positives due to off-target interactions and nuclease degradation.[2-8] ODN probes are also diluted by cell division, so they cannot be used for long-term tracking *in vivo*. RNA reporter systems are the most likely to provide more timely detection of changes in transcription.

Two types of RNA reporter systems have been shown to function in living cells. One utilizes a repeated segment of a RNA element that binds to a protein/peptide that is linked as a fusion protein to a fluorescent protein such as GFP. The most established example of this approach is the use of the bacteriophage MS2 RNA binding domain and MS2-GFP fusion proteins.[9] Other bacteriophage coat proteins such as PP7 have also been used for similar constructions.[10] The detection sensitivity in these systems is limited by the background fluorescence of the constitutive expressed GFP fusion proteins, and thus they require sophisticated image analysis to separate the true signal from background noise. To resolve this problem, several RNA reporters bring together split GFPs linked with two different RNA binding proteins.[11,12] However, these methods for monitoring gene expression require that the host cells constitutively express one or more fluorescent proteins as well as the tagged reporter RNA, thus limiting their potential application to cells for which stable transfectants can be obtained with retention of critical cell behavior such as capability of

differentiating. The second type of RNA reporter provides a signal after interacting with a small

molecule as its ligand. A recent example of this type of reporter, referred to as Spinach,[13] creates a

signal by promoting the fluorescence of a GFP fluorophore mimic. A similar approach was proposed

for the malachite green aptamer, which selectively increases its ligand's fluorescence efficiency by

2,000-fold in the aptamer-ligand complex.[14] The Spinach tag provides a sufficient signal when linked

to the very abundant ribosomal RNA to provide an image of that RNA in the cell. However, we have

found that the background fluorescence in cells of this type of ligand is too high relative to the signal

obtained from a limited number of mRNAs to allow discrimination of the signal from the aptamer

over the background noise.

 Here we report a small RNA transcriptional reporter that we call Intracellular Multiaptamer

Genetic tags (IMAGEtags) by which promoter activity is visualized in living cells. The image is

obtained from a string of RNA aptamers (IMAGEtags) that recognize exogenously supplied

fluorescent ligands. The sensitivity of this is heightened by utilizing a FRET signal between nearby

ligands for recognizing the presence of the IMAGEtags. To demonstrate the applicability of this

method, we have visualized GAL1 promoter activity using IMAGEtags in living yeast cells. This

new RNA reporter system enables real-time monitoring of newly transcribed mRNA in response to

changes in promoter activity and is broadly applicable to many cell types.

Results

Basic Strategy

 Our strategy for measuring promoter activity as an RNA reporter in real-time *in vivo* is shown

in Figure 1. IMAGEtags, which are strings of repeated RNA aptamers, are inserted as a synthetic

coding region after the promoter of choice for plasmid-based expression or can be inserted into the

endogenous gene. Two forms of fluorescent ligands, that constitute a FRET pair, are exogenously provided and enter the cells by diffusion or passive transport. Binding of the transcribed IMAGEtags to the fluorescent ligands is observed by an increase in the FRET signal. In these studies, Cy3- and Cy5-linked aptamer ligands were used as the FRET pairs. The abilities of the aptamers and IMAGEtags used in this study to bind with their fluorescent ligands are shown in Figure 2 and Table 1 of Chapter 4. Similarly, it is demonstrated in Figure 1 of Chapter 4 that these fluorescent ligands are not toxic to target cells over the range of time used for imaging.

IMAGEtags are specifically detected by FRET

Three different aptamers that recognize tobramycin, neomycin or PDC[15] were used to construct IMAGEtags placed under the control of the GAL1 promoter. In all instances a FRET image was observed 15-60 min after inducing the promoter with galactose when the appropriate ligands were used to visualize the expressed IMAGEtags (Fig. 2). FRET signals, calculated as the ratio of fluorescence intensity of a FRET image (F_{FRET}) to that of a donor image (F_{donor}), were quantified from Z stacks of individual yeast cells.

Although at least two aptamers are required for FRET, the number of aptamers that can be included in the IMAGEtag appears to be flexible, with cells expressing IMAGEtags containing 5, 10 and 14 tobramycin aptamers, 24 neomycin aptamers and 6 PDC aptamers all giving FRET signals higher than the control RNA (0mer). However, the presence of a cognate aptamer-ligand pair is important for obtaining a FRET signal, with an incorrect pair (for example Cy3-PDC and Cy5-PDC and the tobramycin IMAGEtags shown in Fig. 2) giving no additional signal compared with the control despite the presence of similar levels of IMAGEtag or control RNA expression.

Validation of the IMAGEtag FRET signal by acceptor photobleaching and FRET lifetime measurements

Although FRET has the advantage of greater sensitivity due to the improved signal/noise, FRET observed as sensitized emission must be validated to ensure that it is not due to bleed-through of the donor signal or other potential artifacts. Therefore, the FRET signal from IMAGEtag expressing cells was quantified by acceptor depletion in which energy transfer between donor (Cy3) and acceptor (Cy5) is prevented by irreversibly photobleaching the acceptor. For true FRET, this treatment results in an increase in donor fluorescence intensity (fluorescence dequenching). FRET measurements by acceptor photobleaching were carried out with cells expressing tobramycin or PDC IMAGEtags or the control RNA (Fig. 3A). The donor fluorescence intensity was selectively increased after acceptor photobleaching in the IMAGEtag expressing cells and the calculated FRET efficiency was statistically higher in IMAGEtag expressing cells than control cells.

To confirm FRET from IMAGEtags, the measurement of donor lifetime was performed by fluorescence lifetime imaging microscopy (FLIM). FLIM-FRET measures the decrease of donor lifetime due to FRET and is independent of fluorophore concentration and spectral detection crosstalk.[16,17] The donor lifetime is reduced in the presence of a FRET acceptor due to the availability of an alternative pathway for excited-state deactivation by interaction with the acceptor.[18] Photobleaching the acceptor results in an increase in the donor lifetime and this change can be used to quantify the FRET efficiency.

The donor lifetime was measured before and after acceptor photobleaching in cells incubated with Cy3- and Cy5-tobramycin that were expressing tobramycin IMAGEtags or control RNA (Fig. 3B). Both donor lifetime and donor intensity were increased in yeast expressing tobramycin IMAGEtags after the acceptor was photobleached. By contrast, the increase of donor lifetime and donor intensity due to acceptor photobleaching was relatively small in yeast expressing control RNA.

These results confirm that the IMAGEtags can be detected by a FRET signal from the ligand pair recognized by aptamers in the RNA.

Real-time and continual detection of promoter activity with IMAGEtags

The IMAGEtag technology can provide a snapshot of the current activity level of the promoter that is being reported while the cells are still living. The image is equivalent to quantifying the RNA by RT-qPCR, with the advantage that the cells are still alive after the data is collected. To demonstrate this capability, yeast cells expressing PDC IMAGEtags (6mer PDC) were induced for different time periods and the FRET signal was measured in individual cells after each induction time (Fig. 4A). As expected from our current knowledge of stochasticity of gene expression[19] all cells did not respond identically. For most of the cells the FRET signal increased with incubation time. However, a few cells (20%) did not respond with an increased FRET signal along the time course. By comparison, the FRET signal from the control cells (0mer) changed very little over this time period (Fig. 4B). At the each time point, average of FRET signal from the PDC IMAGEtags was statistically higher than the average from control RNA ($p < 0.001$). In addition, the rate of the average increase in FRET signal paralleled the amount of IMAGEtag mRNA (Chapter 4, Figure 3B). Thus, the IMAGEtag system reflects the level of reporter RNA present in the cells at the time the IMAGEtag is taken. This is different from promoter reporters for which there is a significant delay in time before the protein is detected compared with the RNA (Chapter 4, Figure 4).

The IMAGEtag technology also provides an opportunity for continuous monitoring of changes in promoter activity in living cells. To demonstrate this capability, we monitored the decrease in donor lifetime due to FRET with time after adding galactose to cells expressing tobramycin IMAGEtags under control of the GAL1 promoter (Fig. 4C). The cells were preincubated in raffinose and induced by galactose, which are conditions that result in a rapid increase in promoter activity that is almost complete by 20 min (Chapter 4, Figure 3A). To continuously monitor FRET by

activating promoter in real time, yeast cells were preincubated in raffinose with the Cy3- and Cy5-tobramycin. Galactose was added to initiate transcription from the GAL1 promoter and the donor lifetime was monitored every 2 min by FLIM-FRET (Fig. 4C-D). Because the half-life of IMAGEtags is short, the level of IMAGEtag RNA measured by RT-qPCR should closely approximate the rate of their transcription (Chapter 4, Figure 5). The donor lifetime changed in parallel with the IMAGEtag RNA levels over the 60 min period of monitoring. The donor lifetime in the same cells that were not induced or cells that were induced but expressed the control RNA changed very little over the same time period (Fig. 4D,E). Thus, promoter activity can be observed continuously with IMAGEtags over a period of at least one hour.

Discussion

The IMAGEtag system described here is an RNA reporter system that enables continual or snapshot visualization of promoter activity in real time *in vivo*. This method has advantages over currently available mRNA reporter systems for which the MS2 system and Spinach RNAs are the most prominent. The MS2 system has been successfully used to visualize transcriptional events in yeast and mammalian cells in real time and to follow an mRNA as it traffics through the cell.[9,10] However, this system requires stable expression of the fluorescent fusion proteins that bind the reporter, which limits its application to many cells for which specialized cellular behaviors are frequently not retained in stably transfected cells. The requirement to express two proteins for MS2-based imaging also creates a metabolic burden for the cells that could have unintended impacts on cellular activities and might also affect the transcriptional response that is monitored.[20] Although technologies for imaging mRNA synthesis are likely to always have a footprint on the cell, it is important that this should be minimal. Aptamer tags that bind small molecules as opposed to proteins are likely to have a smaller impact on cell metabolism and behavior. This approach was taken with the Spinach RNA tag that has been successfully used to track ribosomal RNA through cells.[13]

However, the quantity of an individual mRNA is far smaller than an individual ribosomal RNA by thousands of fold even if the mRNA is overexpressed. Thus, for imaging mRNA expression in real time the challenge is to obtain a high signal to noise ratio from an RNA tag.

The IMAGEtag is an RNA tag that utilizes FRET for high signal/noise to provide high sensitivity. The method does not require the stable expression of proteins and the IMAGEtags are small so as to minimally impact on cellular metabolism and RNA processing and trafficking. As for the Spinach ligand, the IMAGEtag ligands are not toxic to the cells in the time-frame of our current experiments.

Here we have demonstrated that the IMAGEtags can be detected over the background noise by measuring the FRET output of the bound IMAGEtag ligands. The FRET signal was measured and validated in yeast by sensitized acceptor fluorescence, quenching of donor fluorescence and donor fluorescence lifetime. IMAGEtags can readily be attached to any RNA and used as reporters in transiently transfected cells. Thus, this system provides a generally applicable means of studying gene expression in real time in individual cells.

This new ability to image gene expression in individual cells by IMAGEtags will provide a means of expanding our knowledge of how individual cells in a population respond with altered gene expression. Even for isogenic yeast cells, cell signal transduction and gene expression upon stimulation vary between individual cells.[10,21] Using the IMAGEtags we have also been able to observe cell to cell variations in GAL1 promoter activity.

IMAGEtags also have the potential to be applied to investigating the interaction between cells during developmental events or in prokaryotic communities or for observing the response of selected cells in the host in response to pathogen invasion. These and many other important prokaryotic and eukaryotic communication events that involve changes in gene expression are accessible to imaging in real time with IMAGEtags.

Acknowledgment

References

1. T. Endoh, M. Mie, and E. Kobatake, *J Biotechnol* **133** (4), 413 (2008).
2. K. Okabe, Y. Harada, J. Zhang et al., *Nucleic Acids Res* **39** (4), e20 (2010).
3. B. Juskowiak, *Anal Bioanal Chem* **399** (9), 3157 (2011).
4. C. J. Yang, K. Martinez, H. Lin et al., *J Am Chem Soc* **128** (31), 9986 (2006).
5. S. Tyagi and F. R. Kramer, *Nat Biotechnol* **14** (3), 303 (1996).
6. A. K. Chen, O. Davydenko, M. A. Behlke et al., *Nucleic Acids Res* **38** (14), e148 (2010).
7. G. Bao, W. J. Rhee, and A. Tsourkas, *Annu Rev Biomed Eng* **11**, 25 (2009).
8. S. Tyagi and O. Alsmadi, *Biophys J* **87** (6), 4153 (2004).
9. E. Bertrand, P. Chartrand, M. Schaefer et al., *Mol Cell* **2** (4), 437 (1998).
10. D. R. Larson, D. Zenklusen, B. Wu et al., *Science* **332** (6028), 475 (2011).
11. M. Valencia-Burton, R. M. McCullough, C. R. Cantor et al., *Nat Methods* **4** (5), 421 (2007).
12. T. Yamada, H. Yoshimura, A. Inaguma et al., *Anal Chem* **83** (14), 5708 (2011).
13. J. S. Paige, K. Y. Wu, and S. R. Jaffrey, *Science* **333** (6042), 642 (2011).
14. J. R. Babendure, S. R. Adams, and R. Y. Tsien, *J Am Chem Soc* **125** (48), 14716 (2003).
15. G. A. Kraus, V. Gupta, M. Mokhtarian et al., *Bioorg Med Chem* **18** (17), 6316 (2010).
16. D. M. Grant, J. McGinty, E. J. McGhee et al., *Opt Express* **15** (24), 15656 (2007).
17. L. Wang, T. Chen, J. Qu et al., *J Fluoresc* **20** (1), 27 (2010).
18. L. Albertazzi, D. Arosio, L. Marchetti et al., *Photochem Photobiol* **85** (1), 287 (2009).
19. A. Raj and A. van Oudenaarden, *Annu Rev Biophys* **38**, 255 (2009).
20. J. Heyland, L. M. Blank, and A. Schmid, *J Biotechnol* **155** (2), 178 (2011).
21. S. Rajan, H. Djambazian, H. C. Dang et al., *BMC Genomics* **12**, 115 (2011).

Figure legends

Figure 1, Schematic diagram of IMAGEtags and the structures of ligands. Left: IMAGEtags
(tandemly repeated aptamers) are transcribed in the cells as an mRNA construct from a promoter of
choice. IMAGEtags bind fluorescently labeled ligands that are labeled with one of two members of a
FRET pair and the bound ligands are visualized by FRET. **Right:** The structures of representative
ligands used in this study.

Figure 2, Analysis of IMAGEtags expressed in yeast. FRET from ligand-bound IMAGEtags was
detected by sensitized emission. Images of fluorescent yeast cells expressing IMAGEtags from the
GAL1 promoter were obtained by confocal microscopy **(A-C)** and the DIC images show all cells in
the field **(B,C)**. **A) Tobramycin IMAGEtags with Cy3- and Cy5-tobramycin.** Cells expressing
control RNA or IMAGEtags containing 5, 10 or 14 tandem tobramycin aptamers were induced by
galactose for 1 h, and then incubated with 25 μM Cy3-tobramycin and 25 μM Cy5-tobramycin for 20
min. Scale bar, 10 μm. **B) PDC and tobramycin IMAGEtags with Cy3- and Cy5-PDC.** Cells
expressing control RNA or IMAGEtags of 6 tandem PDC aptamers or 5 tandem tobramycin aptamers
were induced by galactose for 30 min and incubated with 5 μM Cy3-PDC and 5 μM Cy5-PDC for 30
min. **C) Neomycin IMAGEtags with Cy3- and Cy5-paromomycin.** Cells expressing control RNA
or IMAGEtags of 24 tandem neomycin aptamers were induced by galactose for 15 min and incubated
with Cy3-paromomycin and Cy5-paromomycin for 15 min. **D) Quantification of FRET from PDC
IMAGEtags and tobramycin IMAGEtags in single cells**. Cells were induced by galactose for 30
min and incubated with 5 μM Cy3-PDC and 5 μM Cy5-PDC for 30 min (images shown in **B**). The
fluorescence intensities of the donor and FRET summed through the Z stack were measured from 9
different cells in each field. The mean fluorescence intensity was calculated as ΣPixel volume/(Pixel
count x number of slices). FRET was determined by the formula, $FRET = F_{FRET}/F_{donor}$. **E)
Quantification of RNA.** The amount of the IMAGEtag RNA in each population was determined by
quantitative RT-PCR and normalized to ACT1 mRNA.

Figure 3, Acceptor photobleaching. A) FRET efficiency: The average FRET efficiency from three independent experiments is shown with the standard deviation in error bars. $p < 0.0001$. FRET efficiencies were calculated as $1 - FD/F'D$, where FD and F'D are the donor intensities before and after photobleaching the acceptor, respectively. **B) FLIM-FRET:** Yeast cells were preincubated with 40 μM Cy3- and Cy5-tobramycin for 90 min. Cells that contain expression vectors for the control (0mer) or tobramycin IMAGEtags (5mer tob) driven by a GAL1 promoter, were induced by galactose for 30 min and incubated with 40 μM Cy3- and Cy5-tobramycin for 30 min. Donor intensities, FRET lifetimes and the differences (after-before photobleaching) are shown (color bar for lifetime: 10^{-3} μsec).

Figure 4, Time-dependent change in IMAGEtag RNA level after activation of the GAL1 promoter. Yeast cells transformed with a 2 micron plasmid for expression of PDC IMAGEtags (**A**), tobramycin IMAGEtags (**C,E**) or control RNA (**B,D**) from the GAL1 promoter were induced with galactose (**A-D**) or not induced (**E**). **FRET:** Yeast cells (n = 15 – 27) expressing PDC IMAGEtags (**A**) or control RNA (**B**) were preincubated for 90 min with 5 μM Cy3-PDC and 5 μM Cy5-PDC, induced with 2% galactose at time 0, and imaged at 10, 40 and 90 min. **FLIM-FRET:** Cells transformed with tobramycin IMAGEtag (**C,E**) or control RNA (**D**) expression vectors were preincubated with 40 μM each of Cy3-tobramycin and Cy5-tobramycin in 1% raffinose SD-uracil for 2 h. Galactose was added to 2% at time 0 (**C,D**) or the cells were left uninduced in 1% raffinose (**E**). FLIM-FRET images of the same individual cells were captured every 2 min. and the donor lifetimes were quantified.

Methods

Yeast and plasmids

Saccharomyces cerevisiae (S288C, Genotype: MATalpha ade2delta::hisG his3delta200 leu2delta0 met15delta0 trp1delta63 ura3delta0) were cultured in YPD medium. The yeast expression plasmid, pYES2, is a yeast 2 micron plasmid carrying a URA3 marker and a GAL1 promoter for galactose inducible gene expression in *S. cerevisiae*. The pYES2 was modified to express a reporter RNA consisting of a series of tandem aptamers (multiaptamers) of tobramycin, neomycin, or PDC. These IMAGEtag sequences were inserted after a GAL1 promoter and after a transcriptional start site. Plasmids expressing multiaptamers were transformed into S288C using the Lazy Bones yeast transformation method by selection on SD-uracil plates.(Amberg 2005)

FRET Image Acquisition and Analysis

For FRET analysis by sensitized emission, yeast cells were cultured in SD-uracil medium containing 2 % galactose for the induction period and incubated with the Cy3- and Cy5 modified ligands. Cells were washed once with Tris-based saline (TBS: 50mM Tris-HCl, 150mM NaCl and pH7.6), resuspended in TBS and a 30 μL volume placed on a poly L-lysine coated cover glass or a poly D-lysine coated glass bottom culture dish (MatTek). The cells were observed using a Nikon Eclipse 200 or Leica SP5X laser scanning confocal microscope with a 63X objective and immersion oil. Cells were excited by a 568 nm Argon/Krypton laser (Nikon) or a 550 nm white light laser (WLL, Leica) and FRET images were taken to measure sensitized emission using emission filters for FRET of 700-750nm or 660-710 nm, respectively.

For FRET analysis by acceptor photobleaching, yeast cells were grown in SD-uracil containing 2% galactose for a 30 min induction period and incubated together with a mixture of Cy3- and Cy5-modified ligands in SD-uracil containing 2% galactose. The cells were washed once with TBS and placed on a poly D-lysine coated glass bottom culture dish. Acceptor bleaching was performed with a Leica SP5X laser scanning confocal microscope using the FRET acceptor bleaching

wizard. Prebleach and postbleach images were taken serially with excitation of 550 nm of WLL with lower laser intensity and filters of 560-640 nm for collecting Cy3 emission and 660-710 nm for Cy5emission. The acceptor was bleached with high laser intensity at 650 nm of WLL. The fluorescence intensities of the donor prebleach (FD) and postbleach (FD') conditions were measured by using the LAS AF Lite program. The FRET efficiency was calculated using the formula FRET efficiency = 1 − FD/F'D.(Bhaumik 2006)

As a quantitative readout of FRET, the fluorescence lifetime of the donor, defined as the mean time between fluorophore excitation and photo emission, was measured.(Lleres, James et al. 2009) Yeast cells were preincubated with 40 μM each of Cy3- and Cy5-tobramycin in SD-uracil containing 1% raffinose for 90 min. The GAL1 promoter was then activated by adding galactose (final 2%) and incubating for 30 min. After centrifugation, yeast cells were placed on poly-D lysine coated glass bottom culture dish. Donor fluorescence lifetime was measured before and after acceptor photobleaching by frequency-sweeping Fourier spectroscopy.(Zhao and Peng 2010) The lifetime measurements from different cells were averaged over 5X5 pixel region at the center of the cells.

Quantitative Analysis of IMAGEtag Transcripts by RT-qPCR

Steady-state levels of mRNAs were analyzed by quantitative real-time RT-PCR (RT-qPCR). Yeast RNA was extracted by phenol chloroform using glass beads.(Del Aguila, Dutra et al. 2005) Briefly, cells were harvested in 0.6 mL of RNA extraction buffer (10 mM EDTA, 50 mM Tris-HCl pH 7.5, 0.1 M NaCl, 5% SDS [sodium dodecyl sulphate]) and 0.6 mL of phenol:chloroform:isoamyl alcohol (49.5:49.5:1). After 6 min incubation at room temperature, ~0.2 g of glass beads (0.45 mm diameter) was added and the cells were lysed by vigorous agitation for 2 min. Following centrifugation, the aqueous phase was collected; the RNA was precipitated by ethanol, dissolved in water, and then treated with RNase-free DNase I (1 U/μL) at 24°C for 15 min to remove traces of genomic DNA. One μg of RNA was transcribed to create cDNA using the SuperScript III reverse

transcriptase and oligodT as a primer in a final volume of 20 μL. After reverse transcription, 2 μL of a 1:2 diluted cDNA was used as a template for qPCR with SYBRgreen and Taq polymerase into a final volume of 15 μL in the Opticon (Bio-rad). The forward 4656 (TGCAGATATCCATCACACTGGC) and reverse 4657 (CCTAGACTTCAGGTTGTCTAACTCC) primers were located outside the IMAGEtag sequences. The IMAGEtag RNA expression levels were normalized to the ACT1 mRNA levels, which were determined in the same samples using the forward and reverse primers 4609 (ATTCTGAGGTTGCTGCTTT) and 4610 (GTCCCAGTTGGTGACAATAC), respectively. The PCR thermal cycling conditions were 5 min denaturation at 94°C, 40 cycles at 94°C for 15 sec, 60°C for 15 sec, and 72°C for 15 sec.

Time lapse FLIM

Yeast cells, precultured in 2% raffinose containing media, were incubated with SD-uracil containing 1% raffinose with 40 μM Cy3-tobramycin and 40 μM Cy5-tobramycin for 2h. After washing, the cells were placed on poly-D lysine coated glass bottom culture dish with SD-uracil containing 1% raffinose. The donor lifetime was measured immediately after addition of two percent galactose every 2 min by FLIM.

Figures

Figure 1
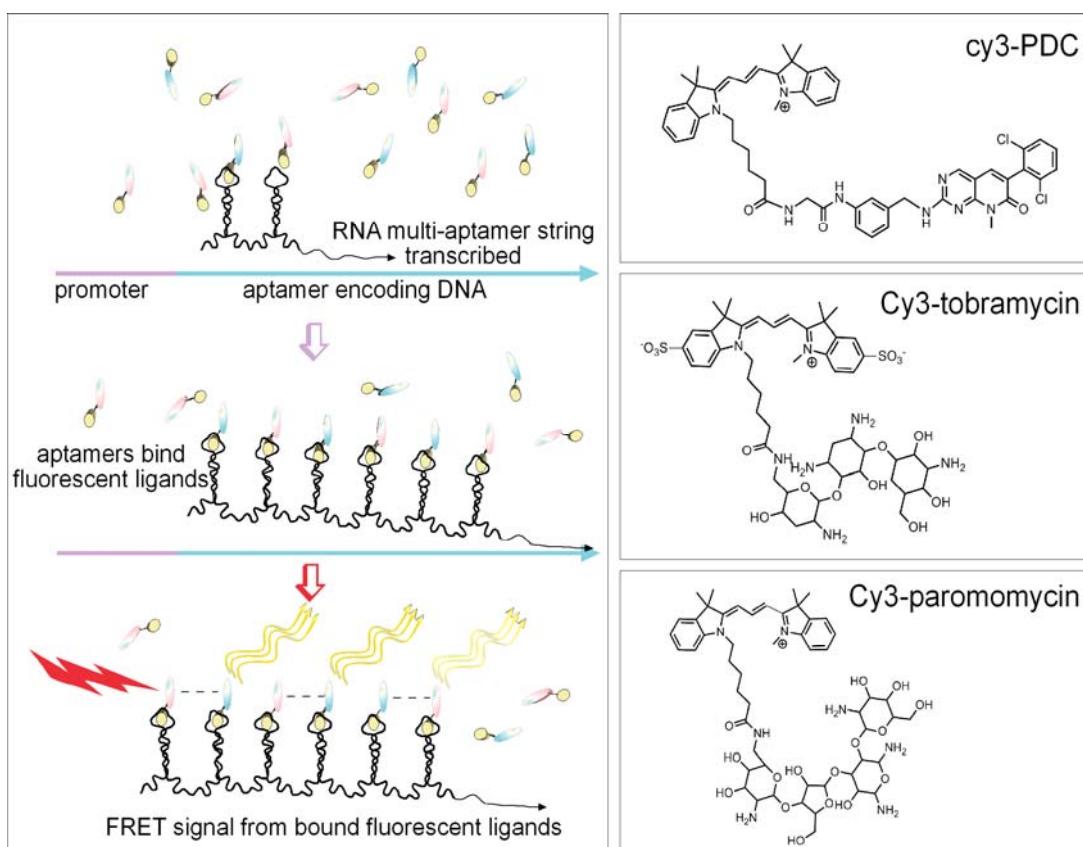
Figure 2
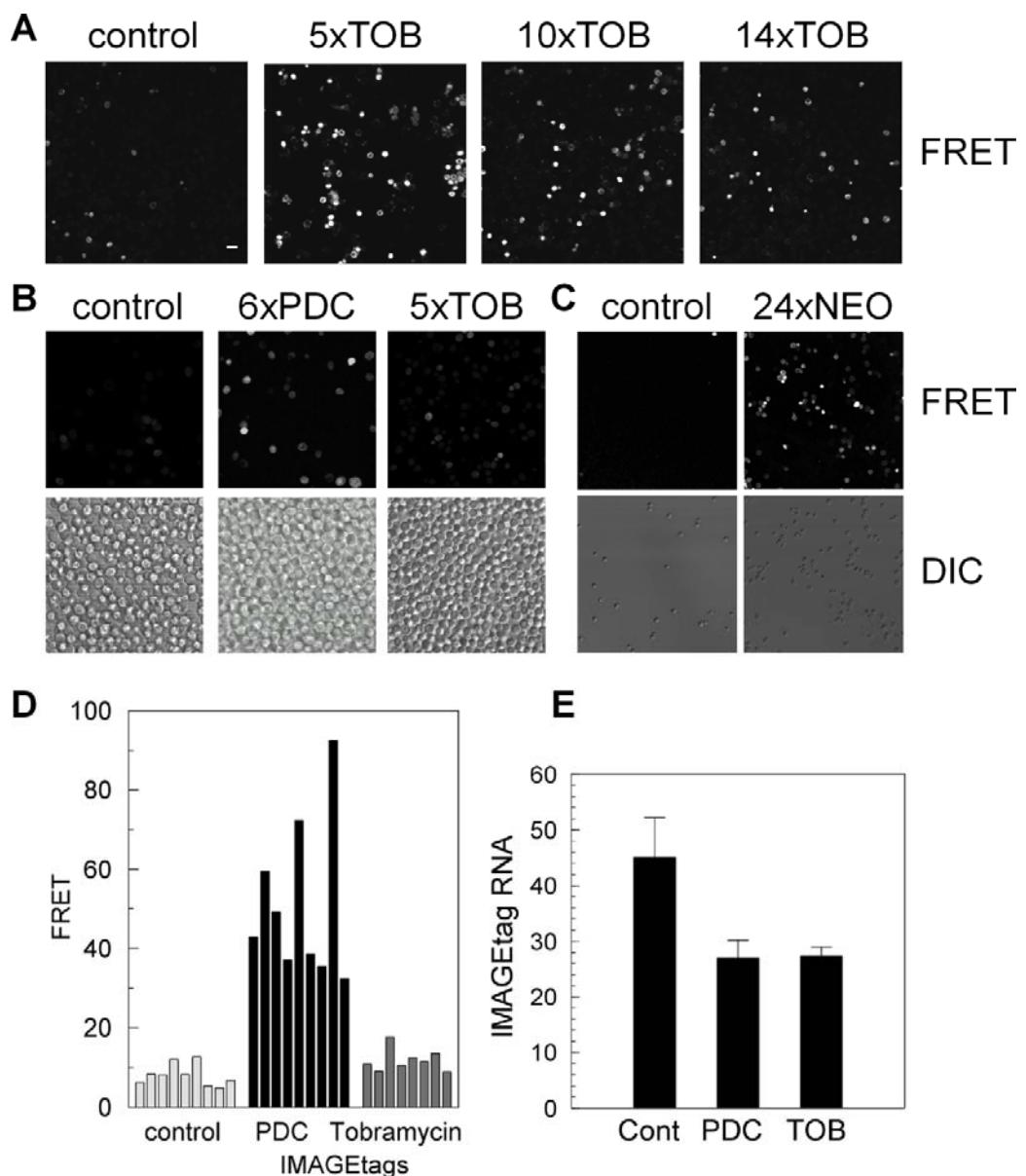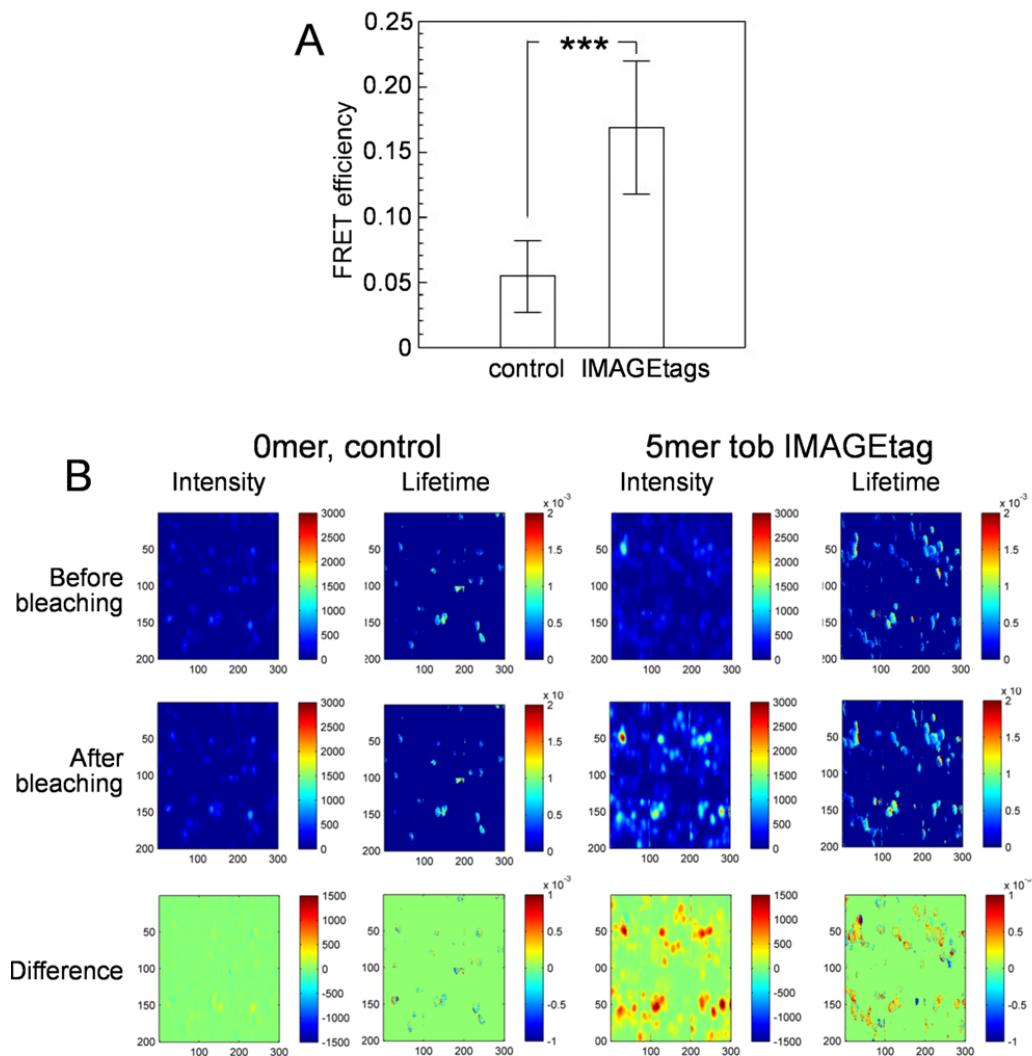
Figure 3

Figure 4
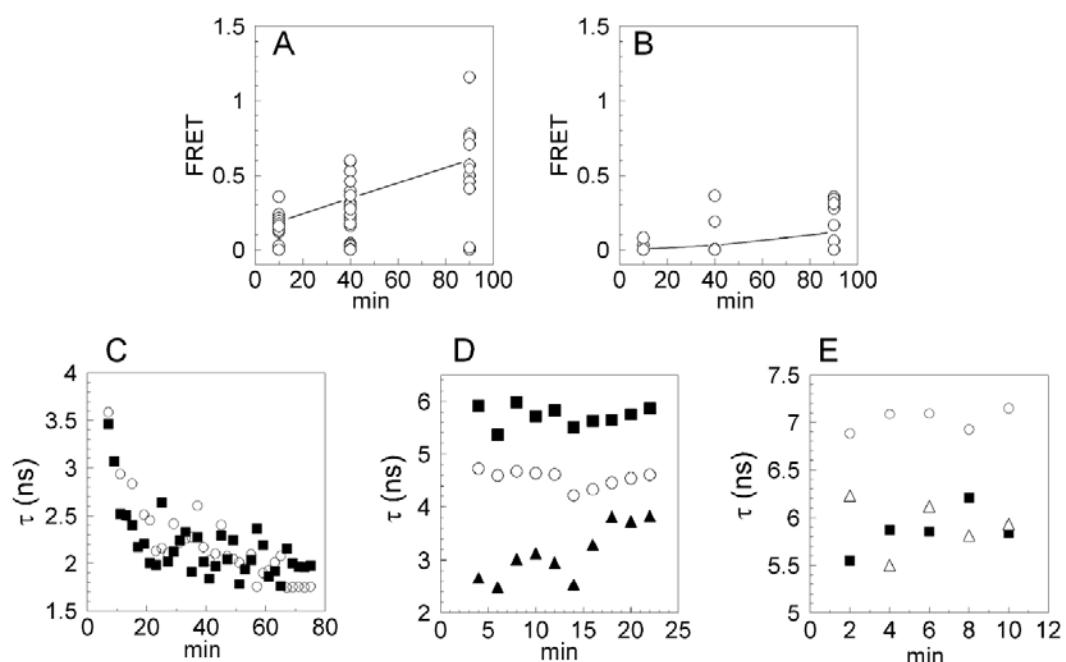
Chapter 4. STUDY OF TOBRAMYCIN IMAGETAGS AND LIGANDS

Supporting material for a paper to be submitted to *Nature Methods*

Introduction

The IMAGEtags system involves two parts: RNA IMAGEtags (multiaptamers), which are expressed inside the cell after promoter activation, and fluorescent labeled ligands, which are bound by the aptamers resulting in a fluorescence signal. For an effective IMAGEtag system, the properties of aptamer ligands should be that they 1) are able to enter and leave cells easily, 2) are not toxic to target cells, and 3) do not have endogenous competitors for aptamer binding. The aptamers in the IMAGEtags should 1) bind to ligands with high affinity, 2) fold properly inside the cells, and 3) be short RNAs so the endogenous RNA is not much altered.

Tandem multiple aptamer sequences were used for a higher fluorescent signal/noise for clear imaging of the IMAGEtag reporters. To increase fluorescent signal, MS2 system introduced multiple copies of the phage hairpin motif in the untranslated region of target mRNA. Binding of multiple copies of the MS2 coat protein-GFP to RNA motifs allows more fluorescent target mRNA than the background (Tyagi 2009).

The aptamers in the IMAGEtags should fold properly inside the cells for IMAGEtag functionality. This was tested in a study in which three tobramycin aptamers in series were expressed in bacteria and tested for their effect on bacterial growth. If the aptamers folded properly in the cells, they would be expected to bind the tobramycin and protect the cells from killing by tobramycin, which normally binds to the ribosomes and prevents translation. The results of the experiment were that the cell growth was not inhibited when treated with tobramcyin whereas the control cells died.

These studies showed that the three tobramycin aptamers were functional in bacteria (Werstuck and Green 1998).

The GAL1 promoter is inhibited in the presence of glucose, while it is activated in the presence of galactose. Others have observed activity of the GAL1 promoter in living cells in which GFP expression was under the control of the GAL1 promoter and its fluorescent signal was monitored. In the presence of a mixture of carbon sources, glucose and galactose, the appearance of the GFP reporter showed that the GAL1 promoter turned on after the depletion of glucose (Li, Wang et al. 2000). The response of the GAL1 promoter varied depending on the carbon sources in the previous culture. The *GAL10* mRNA reached a maximum level within 10 minutes if the yeast had been previously cultured in raffinose as the carbon source. Raffinose is neither an activator nor inhibitor of the GAL promoter. However, it took several hours to reach the maximum level of the *GAL10* mRNA when the yeast had previously been grown in glucose as a carbon source (Kundu and Peterson 2010).

The steady-state level of mRNA in a cell is balanced by its rates of production and decay. Control of mRNA decay can be an important regulatory step in gene expression. The decay rate of mRNA is affected by various stimuli and cellular signals. The half-lives of yeast mRNAs vary from short (~ 3 min) to long (more than 90 min). The transcripts encoding enzymes involved in energy metabolism such as glycolysis, gluconeogenesis and tricarboxylic acid cycles show a longer half-life. In contrast, transcripts encoding proteins with a regulatory role such as the mating pheromone signal cascade turn over rapidly (Wang, Liu et al. 2002).

We used tobramycin and the tobramycin aptamer as components of the IMAGEtag system. In order to detect IMAGEtags from bound fluorescent labeled ligands, we modified tobramycin with fluorescent dyes, adding either Cy3 or Cy5. As part of the process of establishing the conditions for measuring the binding activity of the IMAGEtags, we tested the binding affinity of a single tobramycin aptamer to fluorescently labeled ligands in a buffer that approximates the intracellular

conditions. Multiple tandem tobramycin aptamers that could constitute the IMAGEtags were also tested for binding with tobramycin in the same intracellular buffer to confirm that the aptamers were properly folded for ligand binding activity. We also measured the toxicity of Cy3- and Cy5-tobramycin in yeast as the model organism.

We measured the IMAGEtag RNA expressed from the GAL1 promoter in cells incubated in different carbon sources in preculture and in various carbon sources during the analysis period. To determine the length of time that we should expect to see the specific signal from IMAGEtags after the promoter shuts down again, we determined the decay rate of the IMAGEtags.

Materials and methods

Reagents

The tobramycin RNA aptamer was ordered from IDT (Coraville, Iowa). YPD, tobramycin, galactose, raffinose, phenol:chloroform:isoamyl alcohol (49.5:49.5:1) and glass beads (0.45 mm diameter) were purchased from Sigma (St. Louis, Missouri). Dextrose was purchased from Fisher Scientific (Pittsburgh, Pennsylvania). RNase-free DNase I, SuperScript III reverse transcriptase, SYBRgreen, and DH5α competent cells were purchased from Invitrogen (Carlsbad, California). Taq polymerase, *BsmAI,* and *BsaI* were purchased from New England Biolabs (Ipswich, Massachusetts). *HindIII*, *BamHI*, T4 polynucleotide kinase, T4 DNA ligase, and pSV-β-galactosidase control vector were purchased from Promega (Madison, Wisconsin). Thiolutin was purchased from Santa Cruz Biotechnology (Santa Cruz, California). All sequencing was done by the Iowa State University DNA facility. The oligonucleotides for use as primers or inserts were synthesized by the DNA facility. Cy3-tobramycin and Cy5-tobramycin were synthesized by Vinayak Gupta in the lab of Dr. George Kraus (Department of Chemistry, Iowa State University).

Cloning of tobramycin IMAGEtags

The method for cloning the IMAGEtags, which consist of repeated tandem aptamers, is described in Haarberg's thesis (Haarberg 2009). Briefly, two pairs of complementary oligonucleotides were used to generate inserts containing five repeats of the tobramycin aptamer sequences, flanked by restriction enzyme sites for *BsaI* and *BsmAI* for subsequent multiplication. Annealed oligonucleotides were phosphorylated by T4 polynucleotide kinase and ligated to a minimized pZErO-2 vector that contained the *E. coli* ori and a gene for kanamycin resistance. The ligation mixture was transformed into DH5α *E. coli* by heat shock. For cloning longer repetitive sequences, an initial tobramycin repetitive sequence was cut by *BsmAI* and the vector was opened by *BsaI*. A purified insert and vector were ligated together and transformed into DH5α *E. coli*. The number of inserted tobramycin aptamers was identified by sequencing and electrophoresis in agarose gel. Clones were obtained containing five, 10, and 14 tobramycin aptamers inserted into the vector in tandem. The tobramycin IMAGEtags were cut by restriction enzymes and pasted into a pYES2 yeast expression vector.

Yeast and plasmids

The cDNA of β-galactosidase was inserted in a pYES2 vector after restriction digestion by *HindIII* and *BamHI*. Plasmids expressing β-galactosidase and tobramycin IMAGEtags were transformed into *Saccharomyces cerevisiae* (S288C) using the Lazy Bones yeast transformation method (Amberg 2005). Transformants were isolated by selection on SD (synthetic dropout)-Uracil plates.

Toxicity determination

S. cerevisiae (S288C) was maintained in YPD (1% yeast extract, 1% peptone, 2% glucose) medium at 30°C overnight. The initial yeast cell density in YPD medium was approximately 0.02 OD_{600} units and the cells were dispensed into a 96-well plate to be tested in quadruplicate. Several concentrations of Cy3-tobramycin and Cy5-tobramycin were added to each well of the 96-well plate.

To test PDC derivatives, several concentrations of Cy3-PDC and Cy5-PDC were treated in each well of a 96-well plate in the presence of 2−5% DMSO (dimethyl sulfoxide). For all studies, the control wells contained a similar cell density of yeast in YPD medium with 2−5% DMSO. The plates were sealed with an oxygen permeable membrane and incubated on a plate shaker at 250 r.p.m at 30°C for 24 h. The turbidity was measured every hour by measuring the absorbance at 600 nm in a plate reader (Synergy2 plate reader Biotek, Winooski, Vermont). The growth curves of the treated and control cultures were compared and toxicity was observed as a decrease in growth rate and final density as determined by OD_{600}.

Isothermal titration calorimetry (ITC)

ITC experiments were carried out at 25°C with a stirring speed of 300 r.p.m in intracellular buffer (13.5 mM NaCl, 150 mM KCl, 0.22 mM Na_2HPO_4, 0.44 mM KH_2PO_4, 100 μM $MgSO_4$, 120 nM $CaCl_2$, 120 μM $MgCl_2$, 20 mM HEPES, pH7.3 at 24°C), which was formulated to approximate intracellular pH and cation concentrations based on literature reports for these values. The ITC experiments were performed using a VP-ITC instrument (Microcal, Piscataway, New Jersey). Thirty injections of 10 μL from a 300 μL syringe were made at a spacing time of 300 sec. Titrations were performed with 100 μM tobramycin in the syringe and 10 μM tobramycin RNA aptamer in the cell. Control experiments to determine the heat of the dilution were performed injecting tobramycin into a cell containing intracellular buffer. These values for the heat of dilution of tobramycin to buffer were subtracted from the corresponding values of heat from titrating tobramycin into the tobramycin aptamer. To measure the binding constants for tobramycin analogs, 50 μM tobramycin RNA aptamer in intracellular buffer was titrated into a cell containing intracellular buffer and 5 μM Cy3-tobramycin or Cy5-tobramycin, respectively. For control experiments, 50 μM tobramycin aptamer was injected into a cell containing intracellular buffer. To measure the binding affinity for the IMAGEtags, 150 μM tobramycin in intracellular buffer was titrated into a cell containing intracellular buffer and 2.4 μM *in vitro* transcribed 5mer of tobramycin aptamer. For the

control experiment, 150 μM tobramycin was injected into a cell containing intracellular buffer.

Before each titration, the RNA was heated to 95°C and incubated for 5 min in the intracellular buffer,

then cooled down slowly to room temperature for 30 min. Data from the ITC experiments were

analyzed using Origin 7.0 (Microcal) with N (number of binding sites per molecule in the sample

cell), Ka (association constant in M^{-1}), ΔH (enthalpy change in cal/mol), and ΔS (entropy change in

cal/mol/deg) as adjustable parameters.

Quantitative analysis of transcripts by RT-qPCR

      Steady-state levels of mRNAs were analyzed by quantitative real-time RT-PCR (RT-qPCR).

Yeast RNA was extracted by phenol chloroform using glass beads (Del Aguila, Dutra et al. 2005) and

described in Chapter 3. The IMAGEtag and ACT1 were measured by oligo 4656, 4657 and oligo

4609, 4610 respectively. The sequence is described in Chapter 3. The level of galactokinase mRNA,

the endogenous positive control of the GAL1 promoter, was determined by the forward 4654

(TTTGATATGCTTTGCGCCGTC) and reverse 4655 (AGTCCGACACAGAAGGATCAATT)

primers. The IMAGEtag and galactokinase RNA expression levels were normalized to the ACT1

mRNA level, which was determined in the same samples.

Beta-galactosidase assay

      Yeasts containing β-galactosidase under the control of the GAL1 promoter were inoculated

into either SD-uracil containing 2% glucose or SD-uracil containing 2% raffinose at 30°C overnight.

Cells from the glucose preculture were then diluted into either SD-uracil containing 2% galactose

(inducing condition) or SD-uracil containing 2% glucose (noninducing condition). The yeast from

the raffinose preculture was diluted in SD-uracil containing 2% galactose (inducing condition). The

5 mL of yeast culture was collected at the identified time points after adding galactose and

centrifuged at 1100 x *g* for 5 min. The yeasts were resuspended in Z buffer (60 mM Na_2HPO_4,

40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM 2-mercaptoethanol, pH 7.0), placed on ice, and

measured for the OD_{600} value. After lysis by 0.1% SDS and chloroform, the yeast lysates were

equilibrated for 15 min at 30°C. ONPG (2-nitrophenyl beta-D-galactopyranoside) was incubated with

the lysates at 30°C. The color of the reaction mixes changed to a range of yellows depending on the

amount of β-galactosidase activity in the sample. The reaction was stopped by adding sodium

carbonate. After centrifugation, the OD_{420} was detected to measure β-galactosidase activity and the

OD_{550} was also measured to detect particulate matter such as cell debris. The β-galactosidase activity

was calculated by the following equation (Reynolds, Lundblad et al. 2001).

Unit = 1000 x [(OD_{420})-(1.75 x OD_{550})] / [(Time) x (vol) x OD_{600}]

The time of reaction is in minutes, vol is volume of culture used in the assay (mL), OD_{600} is a

measure of the cell density at the start of the assay, OD_{420} is a combination of absorbance by o-

nitrophenol and light scattering by cell debris, and OD_{550} is light scattering by cell debris.

Half-life of mRNA

Yeast cells were precultured in SD-uracil containing 2% raffinose and then were transferred

into SD-uracil containing 2% galactose in which they were grown until reaching mid log phase.

Transcription was stopped by adding thiolutin (3 μg/ml), which is an inhibitor of RNAPI, RNAPII

and RNAPIII. Five mL of yeast was collected at each sampling point of the time course and RNA

was extracted as described above. The IMAGEtag, galactokinase and ACT1 RNA expression levels

were normalized to the 25S rRNA level, which was determined in the same samples using the

forward and reverse primers, 4669 (GGTAGGAGTACCCGCTGAAC) and 4670

(CCCAAAGTTGCCCTCTCCAA), respectively.

Results and discussion

Toxicity of fluorescently labelled ligands

Tobramycin is an aminoglycoside antibiotic that is used to treat gram-negative bacterial

infections. Known side effects of tobramycin treatment are nephrotoxicity and ototoxicity (Brunton

2006). PDC is a derivative of PD173955 that selectively inhibits the Bcr-Abl tyrosine kinase that

drives cell growth in chronic myelogenous leukemia (Kraus, Gupta et al. 2010). To be used as the

ligands for IMAGEtags, the fluorescent ligands should not be toxic to the target cells at the

concentrations required and over the period of incubation used for measuring promoter activity. To

test the toxicity of fluorescently labelled ligands, *S. cerevisiae* (S288C) were incubated with each of

the fluorescent ligands and the growth curves were measured at OD_{600}. The results in Figure 1 show

the amount of yeast growth in the presence of the fluorescent ligands compared to the growth in the

control condition after 4 h incubation with the ligands. Four hours are the maximum incubation time

with fluorescently labelled ligands that were used in these studies. These results show that the ligands

were not toxic to yeast at the concentrations used for measuring IMAGEtag presence in the cells and

over a 4 h incubation period.

Binding affinity between aptamers and ligands

The dissociation constant (Kd) of the tobramycin aptamer for tobramycin was previously

reported to be 9 to 12 nM. In that study, aptamer binding was determined by an indirect assay in

which increase of fluorescence intensity from PYT (6'-N-4-pyrenebutyryl tobramycin) was detected

under a condition in which it competed with tobramycin for binding at 20°C in the selection buffer

(140 mM NaCl, 5 mM KCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$, 20 mM Tris-acetate, pH 7.4), which had a

high sodium and low potassium concentration (Wang and Rando 1995). When the tobramycin

aptamers were analyzed for affinity to tobramycin by competition with CRT (5-carboxytetramethyl-

rhodamine tobramycin), the measured Kd tobramycin aptamers to tobramycin by fluorescence

depolarization was 0.77 nM at 20°C in the selection buffer (Wang, Killian et al. 1996). To see the

IMAGEtags by bound fluorescent ligands inside the cells, the aptamer in the IMAGEtag should bind

to the target under intracellular conditions. Therefore we used a buffer with concentrations of

sodium, potassium, calcium and magnesium that approximate reported intracellular concentrations for

these ions. The binding affinity and thermodynamic parameters of tobramycin with the tobramycin

aptamer were measured by ITC at 25°C in a buffer designed to replicate intracellular ionic conditions

(Figure 2A and Table 1). The result showed that the tobramcyin aptamer bound to its ligand,

tobramycin with a Kd of 160 nM. The difference of the Kd value from the published data may be due

to a difference in methodology, binding buffer and temperature. However the difference in method

may also lead to this result because the ligands used by Wang et al. (Wang and Rando 1995; Wang,

Killian et al. 1996) are not tobramycin, but are modified versions of tobramycin, which were assumed

to have the same affinity as tobramycin for the aptamer.

Aptamers are short molecules with many unpaired bases that results in a flexible structure.

Consequently, their conformations and interactions with ligands are very sensitive to environmental

conditions. For example, the ATP (adenine triphosphate) aptamer showed a Kd value 18 times lower

in a 2X SSC (saline-sodium citrate) buffer (30 mM sodium citrate, pH 7.4 300 mM NaCl, 5 mM

MgCl$_2$, 0,01% Tween) compared with the Kd in the selection buffer, which contained 20 mM Tris-

HCl, pH 7.4, 300 mM NaCl, 5 mM MgCl$_2$ and 0.01% Tween. It was concluded that the lower

binding affinity of ATP and aptamer in the SSC buffer is due to the interaction of the negative citrate

ions in SSC buffer with the Mg^{2+} ions that are essential for ATP and aptamer binding (Baaske,

Wienken et al. 2010). Similarly, the thrombin aptamer had an apparent binding affinity six times

lower in buffer containing human serum (1:1 mixture with selection buffer) than in the selection

buffer (20 mM Tris-HCl, pH7.4, 150 mM NaCl, 5 mM KCl, 1 mM CaCl$_2$, 1 mM MgCl$_2$, 0.01%

Tween20, 4% BSA). It is likely that the thrombin aptamer interacts with some of the protein

components of human blood (Baaske, Wienken et al. 2010). The function of the aptamer is affected

by temperature, buffer condition, purification method, and conformational state (Cho, Lee et al. 2009)

and Kd of thrombin aptamer has been reported in the range of 50 to 450 nM from studies between

which the methodology, presence of a washing step and temperature of incubation were variables (Hu

and Easley 2011).

In a tandem array of aptamers, such as describes IMAGEtags, the regions containing the aptamer sequences may not fold as they do in oligonucleotides containing single aptamer sequences. Therefore it was necessary to determine that the aptamers in the IMAGEtags bound to ligand. To test the binding, IMAGEtags were transcribed *in vitro* and their binding affinity was measured by ITC (Figure 2B). The data showed that the IMAGEtag bound with ligands. From the data it appears that there is not a single Kd value, which suggests that all the aptamer sequences in the series do not fold exactly as expected from the oligonucleotide structures. It was not possible to calculate a Kd value for these RNAs because the aptamers may display a range of Kds depending on their positions in the overall sequence. In addition, the proportion of active aptamers is not known for these multiple, tandem aptamer sequences and therefore the concentration of active RNA aptamers is not known for insertion into the calculation for Kd. When the secondary structure of five tandem tobramycin aptamers was predicted by MFold or RNA structure, the predicted structure did not present five properly folded tobramycin aptamers. However, this same result was found from Mfold analysis of four tandem malachite green aptamers. In this latter case, all four aptamers were shown to bind to malachite green with the Kd expected of a single aptamer. Also, all four malachite green aptamers were demonstrated to be properly folded and to bind malachite green by the RNaseI footprinting assay (Wang 2008). However, because this was not true for the tobramycin IMAGEtags, which gave a range of Kds for tobramycin, further optimization of this IMAGEtag will likely require optimization of the sequence for achieving the correctly folded structure reproducibly along the RNA.

To visualize the IMAGEtags binding their ligands when expressed in cells, tobramycin was conjugated with the fluorescent dyes, Cy3 and Cy5. It was important to identify a site of conjugation with the dyes that did not interfere with aptamer binding. In an NMR structure of tobramycin and its aptamer, the primary amine group linked with primary carbon was not involved in binding with the aptamer (Jiang, Suri et al. 1997). This amine is the most basic and active group in tobramycin and has been previously conjugated to affinity gel column for selection of the aptamer (Singh, Pirio et al.

1984; Wang and Rando 1995; Jin, Katritch et al. 2000). Therefore, this amino group was conjugated with Cy3 and Cy5 fluorescent dyes. The binding affinities for Cy3- and Cy5-tobramycin were measured by ITC in an intracellular-simulating buffer (Table 1). Cy3- and Cy5-tobramycin showed binding to the tobramycin aptamer with similar Kds, but the binding affinities were lower than for tobramycin.

Transcriptional regulation of the GAL1 promoter

Glucose is the primary carbon source for yeast growth. Given the option of two carbon sources, glucose and galactose, yeast metabolize glucose first and then use galactose as a carbon source after the glucose in the medium has been depleted (Li, Wang et al. 2000). To achieve this regulation, the GAL1 promoter is inhibited in the presence of glucose by the interaction of the *GAL4* and *GAL80* gene products. Gal4p is found bound upstream of the *GAL* gene, but activation of the promoter is prevented because of its binding with the inhibitory protein, Gal80p. To measure transcriptional activation of the GAL1 promoter, yeast containing 2 micron plasmids, in which IMAGEtags are under the control of the GAL1 promoter, were activated by adding galactose to the culture. Expression of the IMAGEtag RNA was first evaluated by RT-qPCR. It was found that the lag in response to the GAL1 promoter after activation varied with the carbon source in the culture prior to the addition of galactose (preculture). When yeast was precultured in SD-uracil containing 2% raffinose, the response time of GAL1 was faster than when they had previously been cultured in SD-uracil containing 2% glucose (Figure 3 A,B). In addition, the transcription level was higher with raffinose preculture than with glucose preculture. When galactose was used as the lone carbon source for induction, the response of GAL1 promoter was faster than with two carbon sources (Figure 3B,C). In previous studies in which RT-PCR was used to measure the mRNAs, it was shown that *GAL10* mRNA reached a peak level 40 min after adding galactose with raffinose preculture, whereas it reached the peak level 7 h after adding galactose with glucose preculture (Kundu and Peterson 2010).

In another study, GFP was used as the reporter of GAL1 promoter activity in *S. cerevesiae*. As well as the fluorescence intensity of GFP, the culture medium levels of glucose and galactose were determined. After the glucose in the medium was depleted, the level of galactose decreased and the fluorescence intensity from GFP increased after another 95 min, with this lag time being necessary for formation of the GFP chromophore (Li, Wang et al. 2000). The GAL1 promoter probably responded more slowly in cells treated with glucose preculture because glucose still remained inside the cells after its depletion in the medium. These studies confirm the conclusion that, yeast metabolizes galactose only after glucose is depleted both in the medium and in the cells.

Galactose metabolism (Leloir pathway) in *S. cerevisiae* requires four inducible enzyme activities that are galactose mutarotase, galactokinase, galactose-1-phosphate uridyl transferase, and uridine diphosphogalactose 4-epimerase. The galactokinase gene is regulated by the GAL1 promoter and this enzyme is the first to utilize galactose (Hopper, Broach et al. 1978; Thoden, Sellick et al. 2005). The galactokinase gene was used as a monitor of the rate of response of the chromosomal GAL1 promoter to compare its response to that of the extrachromosomal GAL1 promoter from which the IMAGEtags are expressed. The level of galactokinase was measured after adding galactose to cells previously incubated with glucose or raffinose (Figure 3). The time-courses of galactokinase mRNA and IMAGEtag RNA expression were similar but not identical with either glucose or raffinose preculture. Compared with the time course of IMAGEtag RNA expression, there was an initial lag of about 20 min in the time course of galactokinase mRNA increase. This lag is likely due to the need for the chromosomal GAL1 promoter to be remodeled to make an activator region available prior to initiation of transcription. These results demonstrate the need for the IMAGEtags to be integrated into the genome to accurately measure promoter activity of endogenous gene.

In summary, to activate the GAL1 promoter, the carbon source in the media is important. When galactose was the sole carbon source, the GAL1 promoter response was faster and more robust than when two carbon sources, raffinose and galactose, were present. In addition it was found that

the location of the promoter (in the genome or on the plasmid) can affect the response time of transcribing mRNA due to activation.

Reporter RNA vs. reporter protein for tracking GAL1 promoter activation

To compare the response time of the IMAGEtag reporter with a protein reporter, β-galactosidase was inserted after the GAL1 promoter in the 2 micron yeast plasmid and β-galactosidase activity was measured (Figure 4). Unlike for GFP, β-galactosidase does not require posttranslational modification. Therefore, it is in the active form immediately after translation so can be detected faster than a fluorescent protein that requires posttranslational modification. β-galactosidase activity was measured several hours after the addition of galactose, with the lag depending on the preculture conditions in the same way as demonstrated for RNA levels of galactokinase and the IMAGEtags (Figure 4). Whereas β-galactosidase activity was detectable after 2 h with raffinose preculture, it was measurable after 4 h with glucose preculture. However, the IMAGEtag mRNA could be measured several minutes later after adding galactose (Figure3). As for the IMAGEtag mRNA reporter, β-galactosidase activity in cells precultured in raffinose-containing media was faster and higher than for cells precultured in glucose-containing media (Figure 4).

In summary, RNAs like the IMAGEtags are good candidates as reporters to visualize gene expression under circumstances where gene regulation changes rapidly such as in early developmental stages and during inflammation. Consequently, RNA reporters will provide images closer to the time of promoter activation than will protein reporters. In order to be detected, RNA reporters need only to be transcribed and elongated. By contrast, protein reporters need transcriptional initiation, elongation and termination as well as translational initiation, elongation and termination and sometimes posttranslational modifications in order to be observed.

Half-life of IMAGEtag mRNA

One difficulty with protein reporters is that they are quite stable in most cells. Therefore, they do not disappear when the promoter turns off. As for proteins, the abundance of mRNA in a cell is a function of the rates at which it is produced and degraded. However, the half-lives of mRNAs are often shorter than for proteins. The half-lives of mRNAs in yeast vary from ~ 3 min to more than 90 min, whereas the range of protein half-lives is from ~2 min to more than 17 hour with a median of 43 min (Belle, Tanay et al. 2006). Proteins from the production category such as ribosomal proteins and other proteins involved protein biosynthesis show longer half-lives, while proteins from the regulation category such as cell cycle protein and proteins involved in transcriptional regulation have short half-lives (Belle, Tanay et al. 2006). For the exogenous protein, GFP, the half-life is greater than 7h in *S. cerevisiae*. Even after GFP was modified to adopt a shorter half-life, the half live was still ~30 min in *S. cerevisiae*, which is a long time compared with the time periods for transcriptional initiation and elongation (Mateus and Avery 2000).

Thiolutin from *Streptomyces luteoreticuli* was used to detect the mRNA half-life because it inhibits all yeast RNA polymerases (RNAPI, RNAPII and RNAPIII), especially at the level of initiation (Pelechano and Perez-Ortin 2008). Therefore to measure the half-life of IMAGEtag RNA, the amount of IMAGEtag RNA (under the control of the GAL1 promoter) was determined in cells that had been preincubated in galactose-containing medium then treated with 3 μg/ml thiolutin. The calculated half-life of IMAGEtags was 5.5 min, which is significantly shorter than half-life of the galactokinase mRNA (43 min) and the half-life of the ACT1 mRNA (29 min) (Figure 5). Previously reported values for the half-life of ACT1 are 50 min measured by Northern blot (Pelechano and Perez-Ortin 2008), 46 min from overall decay and 11 min from poly A decay measured in a temperature sensitive rpb1-1 mutant (Wang, Liu et al. 2002). The reported value for the half-life of galactokinase mRNA are 31 min from overall decay and 18 min from poly A decay measured in the temperature sensitive rpb1-1 mutant (Wang, Liu et al. 2002). Thus the half-life of the ACT1 mRNA

determined in this study is similar to average of reported ACT1 mRNA half-life, whereas the galactokinase mRNA half-life determined in this study is longer than previously reported. Although the reason for this discrepancy is unknown, the salient point is that the half-life of the IMAGEtags is in the range of the very short RNA half-lives demonstrated previously in yeast. With such a short half-life, IMAGEtags can give a more accurate real-time image of the transcriptional activity from the promoter by which they are driven.

References

Amberg, D. C., Burke, D. J., and Strathern, J. N. (2005). Methods in yeast genetics.

Baaske, P., C. J. Wienken, et al. (2010). "Optical thermophoresis for quantifying the buffer dependence of aptamer binding." Angew Chem Int Ed Engl **49**(12): 2238-41.

Belle, A., A. Tanay, et al. (2006). "Quantification of protein half-lives in the budding yeast proteome." Proc Natl Acad Sci U S A **103**(35): 13004-9.

Brunton, L. L., Lazo, J.S and Parker, K.L (2006). Goodman and Gilman's The Pharmacological Basis of Therapeutics, McGRAW-HILL

Cho, E. J., J. W. Lee, et al. (2009). "Applications of aptamers as sensors." Annu Rev Anal Chem (Palo Alto Calif) **2**: 241-64.

Del Aguila, E. M., M. B. Dutra, et al. (2005). "Comparing protocols for preparation of DNA-free total yeast RNA suitable for RT-PCR." BMC Mol Biol **6**(1): 9.

Haarberg, H. E. (2009). Theophylline IMAGEtags (intracellular multi aptamer aptamer genetic tags): the development and evaluation of an RNA reporter system based on the theophylline aptamer. Neuroscience. Ames, Iowa State University. Master: 77.

Hopper, J. E., J. R. Broach, et al. (1978). "Regulation of the galactose pathway in Saccharomyces cerevisiae: induction of uridyl transferase mRNA and dependency on GAL4 gene function." Proc Natl Acad Sci U S A **75**(6): 2878-82.

Hu, J. and C. J. Easley (2011). "A simple and rapid approach for measurement of dissociation constants of DNA aptamers against proteins and small molecules via automated microchip electrophoresis." Analyst **136**(17): 3461-8.

Jiang, L., A. K. Suri, et al. (1997). "Saccharide-RNA recognition in an aminoglycoside antibiotic-RNA aptamer complex." Chem Biol **4**(1): 35-50.

Jin, E., V. Katritch, et al. (2000). "Aminoglycoside binding in the major groove of duplex RNA: the thermodynamic and electrostatic forces that govern recognition." J Mol Biol **298**(1): 95-110.

Kraus, G. A., V. Gupta, et al. (2010). "New effective inhibitors of the Abelson kinase." Bioorg Med Chem **18**(17): 6316-21.

Kundu, S. and C. L. Peterson (2010). "Dominant role for signal transduction in the transcriptional memory of yeast GAL genes." Mol Cell Biol **30**(10): 2330-40.

Li, J., S. Wang, et al. (2000). "Green fluorescent protein in Saccharomyces cerevisiae: real-time studies of the GAL1 promoter." Biotechnol Bioeng **70**(2): 187-96.

Mateus, C. and S. V. Avery (2000). "Destabilized green fluorescent protein for monitoring dynamic changes in yeast gene expression with flow cytometry." Yeast **16**(14): 1313-23.

Pelechano, V. and J. E. Perez-Ortin (2008). "The transcriptional inhibitor thiolutin blocks mRNA degradation in yeast." Yeast **25**(2): 85-92.

Reynolds, A., V. Lundblad, et al. (2001). "Yeast vectors and assays for expression of cloned genes." Curr Protoc Mol Biol **Chapter 13**: Unit13 6.

Singh, P., M. Pirio, et al. (1984). "Controlled Coupling of Aminoglycoside Antibiotics to Proteins for Use in Homogeneous Enzyme Immunoassays." Canadian Journal of Chemistry-Revue Canadienne De Chimie **62**(11): 2471-2477.

Thoden, J. B., C. A. Sellick, et al. (2005). "Molecular structure of Saccharomyces cerevisiae Gal1p, a bifunctional galactokinase and transcriptional inducer." J Biol Chem **280**(44): 36905-11.

Tyagi, S. (2009). "Imaging intracellular RNA distribution and dynamics in living cells." Nat Methods **6**(5): 331-8.

Wang, T. (2008). Function and dynamics of aptamers: A case study on the malachite green aptamer. Molecular, Cellular and Developmental Biology. Ames, Iowa State University. Doctor of Philosophy: 157.

Wang, Y., J. Killian, et al. (1996). "RNA molecules that specifically and stoichiometrically bind aminoglycoside antibiotics with high affinities." <u>Biochemistry</u> **35**(38): 12338-46.

Wang, Y., C. L. Liu, et al. (2002). "Precision and functional specificity in mRNA decay." <u>Proc Natl Acad Sci U S A</u> **99**(9): 5860-5.

Wang, Y. and R. R. Rando (1995). "Specific binding of aminoglycoside antibiotics to RNA." <u>Chem Biol</u> **2**(5): 281-90.

Werstuck, G. and M. R. Green (1998). "Controlling gene expression in living cells through small molecule-RNA interactions." <u>Science</u> **282**(5387): 296-8.

Figure Legends

Figure 1, Toxicity of fluorescently modified ligands. A) Toxicity of Cy3- and Cy5-tobramycin.

Yeast growth was measured by OD_{600}. The yeast growth treated by these compounds was compared

to the growth of the control yeast culture and the ratio is shown. The control was incubated only in

YPD medium. **B)** The toxicity of Cy3- and Cy5-PDC for yeast. All samples, including the control

cultures, were incubated with YPD medium with 2% of DMSO (for up to 9 μM PDC conjugates) or

5% of DMSO (20 and 40 μM PDC conjugates).

Figure 2, Binding affinity between tobramycin and single or multiple tandem tobramycin

aptamers determined by ITC. A) The binding affinity of tobramycin with a single tobramycin

aptamer. Tobramycin was titrated into the cell containing the tobramycin aptamer to determine the

binding constants of aptamers in the presence of salt concentrations and pH similar to the conditions

inside eukaryotic cells. **B)** Tobramycin was titrated into an ITC cell expressing an IMAGEtag

consisting of five tandem tobramycin aptamers. **Top panel**: The raw ITC data with the blank

subtracted showed the heat release with time. **Bottom panel**: The integrated data from the top panel

showed the normalized heat release plotted with the molar ratio of tobramycin/aptamer.

Figure 3, Time-course RNA levels of tobramycin IMAGEtag and galactokinase in yeast. A)

IMAGEtags and galactokinase RNA levels were determined in GAL1-inducing media (2% galactose

and 1% raffinose) in cells that had been precultured in 2% raffinose-containing media. The RNA

levels were determined by RT-qPCR and normalized to the ACT1 mRNA. **B)** The levels of

IMAGEtag and galactokinase RNAs were measured in GAL1 inducing media (2% galactose and 1%

raffinose) in cells that had been precultured in 2% glucose-containing media. **C)** The levels of

IMAGEtag and galactokinase mRNAs were quantified in GAL1 inducing media (2% galactose) in

cells that had been precultured in 2% glucose-containing media.

Figure 4, Time-course of change in β-galactosidase activity. The time-course of β-galactosidase

activity was determined in GAL1 inducing media (2% galactose) and noninducing media (2%

glucose, marked as Glu [Glucose]). In 2% galactose and SD-uracil, yeast was precultured in either

2% glucose containing media (marked as Gal [Glucose]) or 2% raffinose containing media (marked

as Gal [Raffinose]). The sugar in brackets is that present during the preculture.

Figures 5, IMAGEtags have a short half-life. Yeast containing a 2 micron plasmid with

IMAGEtags driven by the GAL1 promoter was incubated with galactose until reaching mid-log phase

after preculture in 2% raffinose containing media. Thiolutin (3 μg/ml) was then added (set as zero

time) to block transcription. Total RNA was isolated at the indicated times and IMAGEtags RNA,

galactokinase mRNA and ACT1 mRNA were quantified by RT-qPCR and normalized to the 25S

rRNA values in the same samples. The half-lives were calculated as: 1) 5.5 min (regression

coefficient = 0.98) for IMAGEtag RNA, 2) 43 min (regression coefficient = 0.91) for galactokinase

mRNA, and 3) 29 min (regression coefficient = 0.91) for ACT1 mRNA.

Table

Table 1. ITC assay to determine the binding constants of aptamers under salt and pH conditions similar to those inside eukaryotic cells.

		N	Kd (nM)	ΔH (cal/mole)	ΔS (cal/mol/deg)
Tobramycin aptamer	Tobramycin	1.19 ± 0.0246	160 ± 31	-1.33E4 ± 227	-13.4
	Cy3-tobramycin	0.73 ± 0.049	690 ± 2470	-1.96E4 ± 1854	-37.5
	Cy5-tobramycin	0.50 ± 0.019	490 ± 5290	-3.75E4 ± 191	-96.8

The table summarizes the calculated stoichiometry (N), binding constants, enthalpy, and entropy for tobramycin aptamer to its ligands. The buffer contained: 20 mM HEPES, 13.5 mM NaCl, 150 mM KCl, 0.22 mM Na_2HPO_4, 0.44 mM KH_2PO_4, 100 uM $MgSO_4$, 120 uM $MgCl_2$, 120 nM $CaCl_2$, pH 7.3. Due to technical constraints, the temperature was maintained at 25°C.

Figures

Figure 1.
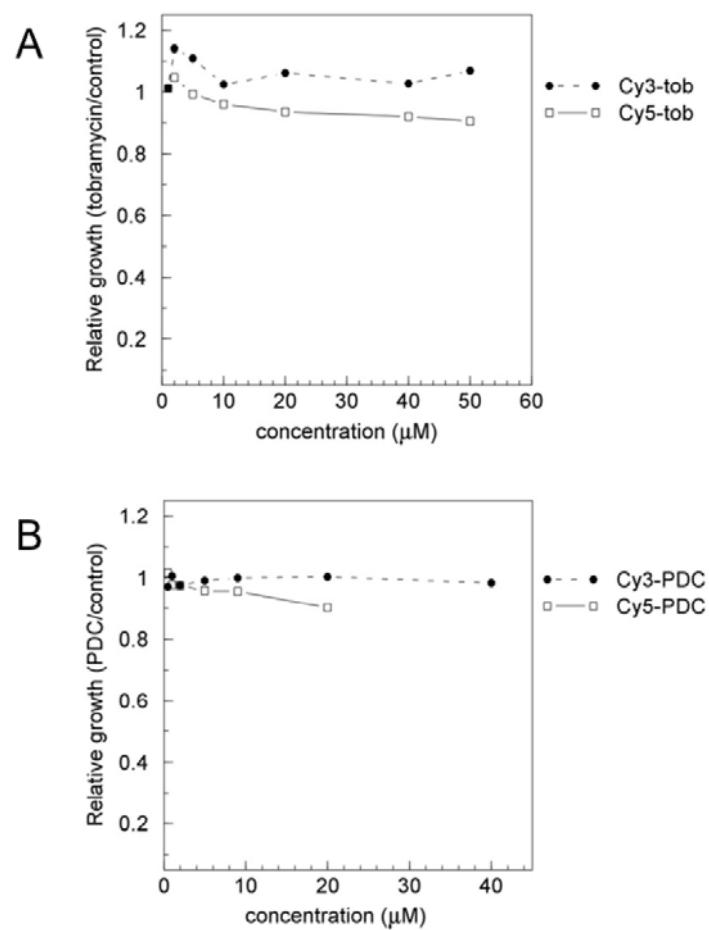
Figure 2.
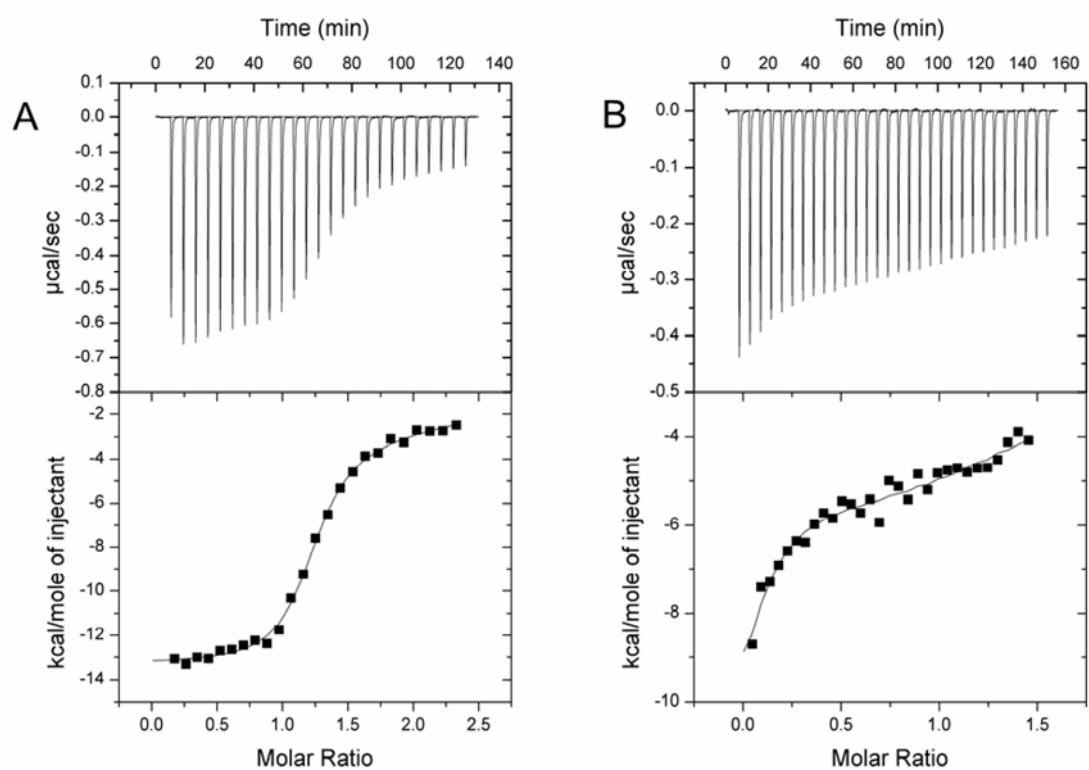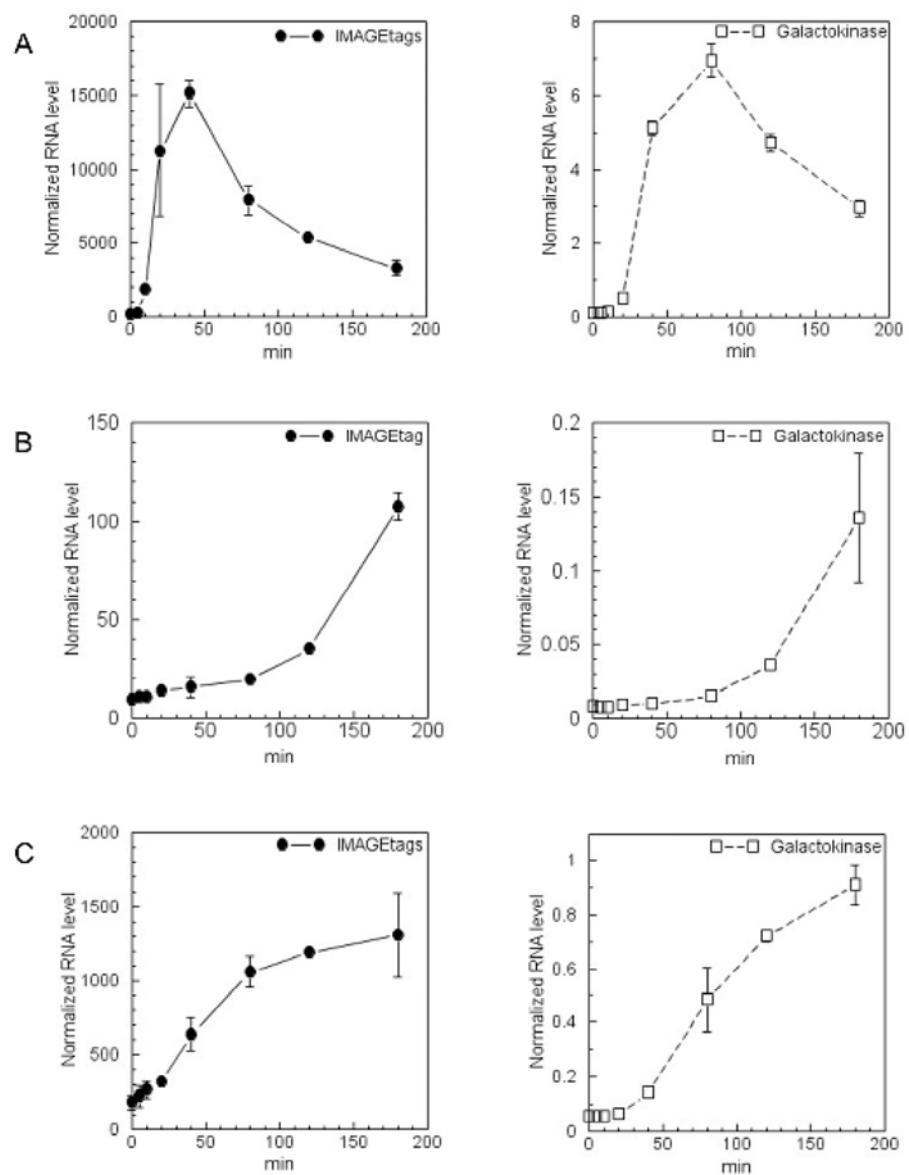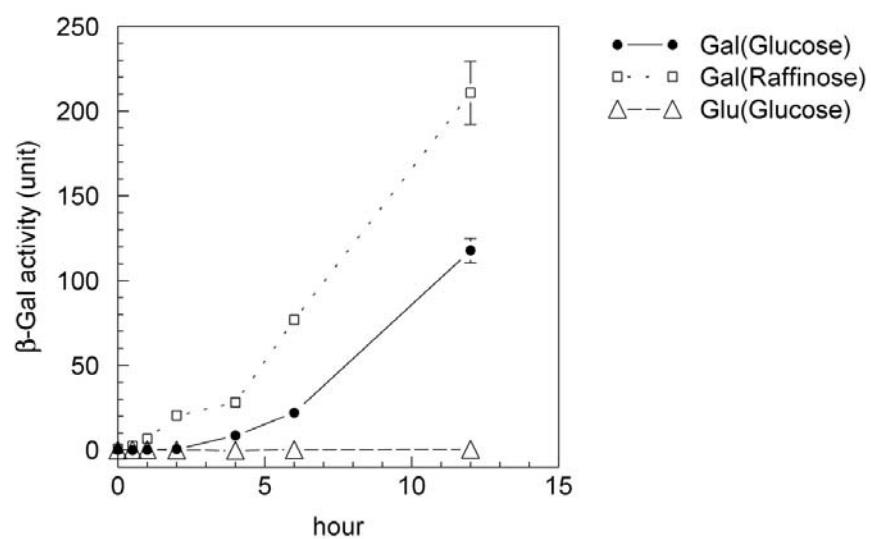
Figure 3.

Figure 4.

Figure 5.
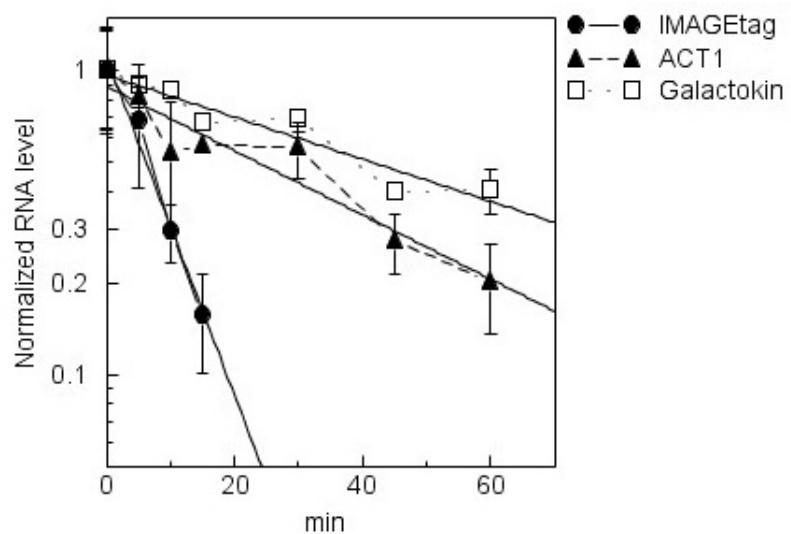
Chapter 5. GENERAL CONCLUSION

The design of IMAGEtags is practical

The IMAGEtag system is an RNA reporter system designed to detect promoter activity in real time *in vivo*. The system is composed of two parts. One part is the IMAGEtags, a string of tandem aptamers. In these studies they have been cloned into plasmids that result in their being transcribed in the format of an mRNA (Pol II promoter and polyadenylation signal) under the control of the GAL1 promoter. The other part is a set of two fluorescently labeled ligands, which give a fluorescence FRET signal when bound to the aptamers of the IMAGEtags. The promoter activity is measured at the RNA level by the binding of the fluorescently labeled ligands by the newly transcribed IMAGEtags. To increase the signal to noise of the image, a FRET system was used. FRET dyes, Cy3 and Cy5, were used to modify the ligands. FRET is only observed between two dyes when they are within about 10 nm of each other, which results in a decreased background signal from fluorescent dyes present in the cell, but not bound to the IMAGEtags.

IMAGEtag system is a practical concept and has flexibility of application and design. Here, IMAGEtag systems were tested with three different ligand and aptamer pairs, which were tobramycin and the tobramycin aptamer, paromomycin and the neomycin aptamer, and PDC and the PDC aptamer. In this study yeast was used as the model system for measuring the GAL1 promoter activity using these three IMAGEtag systems as a model, but the IMAGEtag system can also be applied to measuring the activities of other promoter activities in other cell types.

Advantages of IMAGEtags as the reporter system for real time measurements of gene expression

IMAGEtags are an RNA reporter system for detecting promoter activity. They report at the transcriptional level, so the detection time is faster than with reporter proteins because of the

additional steps involved in making mature reporter proteins. The time-course experiments of qRT-PCR of IMAGEtags and activity of β-galactosidase showed that the IMAGEtag RNAs were detected sooner than the β-galactosidase activity after activation of the GAL1 promoter. Li *et al.* used GFP to see GAL1 promoter activity in real time in *S. cerevisiae*. The fluorescence intensity of GFP was increased after 95 min of incubation with galactose. There was a delay after the addition of galactose until it started to be consumed by the yeast as the carbon source. There was an additional delay in time for maturation of GFP before it was detected (Li, Wang et al. 2000). In this study, neomycin IMAGEtags were detected after 15 min of induction with galactose. Tobramycin IMAGEtags were also used for real-time detection of transcription by the FRET that occurred as a result of binding of newly transcribed IMAGEtags to Cy3- and Cy5-tobramycin. The FRET signal could be detected 7 min after adding galactose. The time lag was necessitated by the manipulations required to establish the cells under the microscope for monitoring. To validate the FRET signal, the donor lifetime was measured and shown to decrease after galactose was added to the cells. Therefore, IMAGEtags are RNA reporters that can detect promoter activity in real-time.

Many methods that have been proposed for measuring intracellular RNA levels in real time have involved the use of complementary nucleic acid probes that were delivered to the cells from an external source. If these probes are not degraded in the cells, they will be diluted in concentration as the cells divide. By contrast, with the IMAGEtag system, cells are genetically modified by promoter-reporter constructs that can express IMAGEtags, which have been inserted as a synthetic coding region after a promoter. Thus, with IMAGEtags, the investigator can examine transcription in a large number of cells with no dilution due to cell division or progressive degradation of the probes. As well, the fluorescently labeled ligands are small enough to enter the cells by passive transport. Delivery of the probes to the cells is one of the major problems for ODN probes due to their large size and negative charge. For efficient delivery, microinjection of ODNs has been used but this causes mechanical cell damage. Microinjection can only be performed with in a small number of cells at a

time, so a limited number of cells can be scrutinized by microinjected ODNs. Microinjected ODNs are also diluted by cell division and subject to degradation, so they cannot be traced for a long time (Zhang and Yu 2008). Therefore, IMAGEtags can be measured in a large number of cells without delivery issues for short and long tracking times.

The IMAGEtag system also has the advantage that it makes a relatively small footprint on the cell that it is being used to monitor. The IMAGEtags are composed of short aptamers of 20–40 nt long with short spacers between them. For example, the total length of the 5-mer tobramycin tag is ~150 nt. This is a smaller size than RNAs that encode fluorescent reporter proteins. For example, GFP is composed of 238 amino acids that are encoded by 714 bases. Small and cell-permeant molecules are used to identify create the image for the IMAGEtag system. Thus, compared with the MS2 system that uses one or two fusion proteins (depending on if it uses GFP-MS2 or split GFP-MS2) to create the image, the IMAGEtag system is also energy-efficient for the cells. This latter advantage is likely to be important under conditions of limited energy supply where the requirement to synthesize one or more exogenous proteins is an extra metabolic burden for the cells. For example, in *E. coli*, the synthesis of recombinant proteins causes a metabolic burden, stressing the cells, forcing a change in carbon metabolism to accommodate the extra energy requirements and affecting cell growth (Heyland, Blank et al. 2011). Another disadvantage of the MS2 system is that the cells into which the MS2 RNA is transfected must already by stably transfected by expression vectors encoding the fusion proteins used for detection.

In summary, IMAGEtags are excellent reporters of promoter activity because they have a small footprint on the cell, can be used with any cell type that can be genetically modified, they are an energy-efficient reporter system, they will give a signal even in the absence of oxygen and they can be applied in transient transfection assays. Of all systems described so far for imaging gene expression in real time, the IMAGEtags are likely to have the least impact on cellular function and gene expression.

With IMAGEtag systems, a promoter activity can be measured in individual living cells. In population studies to measure mRNA levels such as qRT-PCR, it is not possible to determine cell-to-cell variation in the strength or timing of the promoter response or in the spatial dynamics of transcription in a mixed cell population. Gene expression is now understood to be a stochastic process with many factors contributing to intrinsic and extrinsic noise. Even for isogenic yeast cells, cell signal transmission and gene expression upon stimulation vary between individual cells (Rajan, Djambazian et al. 2011). These variations can be quantified with the IMAGEtag system.

Problems of current IMAGEtags and proposed future work

In this study tobramycin IMAGEtags were tested. Although they were shown to work, they still have drawbacks as an RNA reporter system. Tobramycin is a positively charged molecule at physiological pH (Jin, Katritch et al. 2000) that can interact nonspecifically with negatively charged molecules such as RNA and DNA. This nonspecific binding can result in a high background. Although the FRET system was applied in the IMAGEtags, Cy3- and Cy5-tobramycin showed a background FRET. To develop a better IMAGEtag system, the ligand should interact less, and preferably not at all, with cellular molecules and the aptamer should have a high affinity for its ligand. If the aptamer binds the fluorescently labeled ligands with higher affinity, a low concentration of ligands can be applied externally. Thus, studies with IMAGEtags made of higher affinity aptamers will have a lower background because, at the lower concentrations of ligands required for imaging, there will be less nonspecific interaction with other cellular molecules.

The tobramycin IMAGEtags were constructed as a string of a single aptamer. For FRET, tobramycin was modified by Cy3 and Cy5. However, Cy3- and Cy5-tobramycin both bind the same aptamer. The binding of Cy3-tobramycin and Cy5-tobramycin to the IMAGEtag is random and will not always result in adjacent FRET pairs. If the same dyes bind at adjacent sites on the IMAGEtags then homo FRET would be the result. Homo FRET affects hetero FRET by shifting the emission

spectra, leads to perturbation of the measured fluorescence decay (Grailhe, Merola et al. 2006) and changes the donor lifetime (Koushik and Vogel 2008). If two different aptamers are alternated in the IMAGEtags, the binding order of Cy3 and Cy5 can be predetermined and this should result in an increase of the Cy3-Cy5 FRET signal.

The GAL1 promoter is a well-known inducible promoter in *S. cerevisiae*. It is widely used as an inducible promoter because of the strict control of the GAL1 promoter by strong catabolite repression. Although the *GAL* system is regulated well by glucose repression and strong induction, it does not initiate rapid gene expression due to the regulatory mechanism of the galactose signal transduction pathway (Matsuyama, Yamanishi et al. 2011). For a faster promoter response, a different promoter such as *POL1* could be used to control IMAGEtag expression. The *POL1* promoter is a cell-cycle regulated promoter that is active during the late G1/S phase. This promoter was recently used for expressing the PP7 RNA in order to image gene expression (Larson, Zenklusen et al. 2011). The MDN1 promoter was also fused with PP7 RNA cassette and the long gene *MDN1* gene (15kb) was used to image the kinetics of RNAPII initiation, elongation and termination *in vivo* (Larson, Zenklusen et al. 2011). The use of IMAGEtags with different promoters can similarly be used to monitor gene expression.

To test the ability of IMAGEtags to image real-time gene expression, the IMAGEtag system was first tested in yeast as the model system. Yeast is a model system for eukaryotic cells and is easily stably transformed. Yeast also responds to pheromone by activating a transient transcription that could be used for further testing (Wang, Errede et al. 2008). However, we found that the cell wall of yeast created difficulty for permeation of the IMAGEtag ligands and this made it more difficult to work with yeast for testing IMAGEtag functionality. Now that the IMAGEtag system has been established it can be applied to measure promoter activity of mammalian cells and other eukaryotic model systems such as *Caenorhabditis elegans* and *Drosophila melanogaster*. This system for non-invasive tracking of gene expression in real-time will be especially useful for studying

gene expression that changes rapidly such as during cell differentiation and in development as well as under anoxic conditions when fluorescent protein reporters cannot mature.

References

Grailhe, R., F. Merola, et al. (2006). "Monitoring protein interactions in the living cell through the fluorescence decays of the cyan fluorescent protein." Chemphyschem **7**(7): 1442-54.

Heyland, J., L. M. Blank, et al. (2011). "Quantification of metabolic limitations during recombinant protein production in Escherichia coli." J Biotechnol **155**(2): 178-184.

Jin, E., V. Katritch, et al. (2000). "Aminoglycoside binding in the major groove of duplex RNA: the thermodynamic and electrostatic forces that govern recognition." J Mol Biol **298**(1): 95-110.

Koushik, S. V. and S. S. Vogel (2008). "Energy migration alters the fluorescence lifetime of Cerulean: implications for fluorescence lifetime imaging Forster resonance energy transfer measurements." J Biomed Opt **13**(3): 031204.

Larson, D. R., D. Zenklusen, et al. (2011). "Real-time observation of transcription initiation and elongation on an endogenous yeast gene." Science **332**(6028): 475-8.

Li, J., S. Wang, et al. (2000). "Green fluorescent protein in Saccharomyces cerevisiae: real-time studies of the GAL1 promoter." Biotechnol Bioeng **70**(2): 187-96.

Matsuyama, T., M. Yamanishi, et al. (2011). "Improvement of galactose induction system in Saccharomyces cerevisiae." J Biosci Bioeng **111**(2): 175-7.

Rajan, S., H. Djambazian, et al. (2011). "The living microarray: a high-throughput platform for measuring transcription dynamics in single cells." BMC Genomics **12**: 115.

Wang, X., B. Errede, et al. (2008). "Mathematical analysis and quantification of fluorescent proteins as transcriptional reporters." Biophys J **94**(6): 2017-26.

Zhang, Y. and L. C. Yu (2008). "Microinjection as a tool of mechanical delivery." Curr Opin Biotechnol **19**(5): 506-10.

APPENDIX. ABBREVIATIONS

(In alphabetical order)

ATP	Adenosine triphosphate
CBC	Cap binding complex
Cdk9	Cycline dependent kinase 9
CFI	Cleavage factor I
CFII	Cleavage factor II
CFP	Cyan fluorescent protein
CPSF	Cleavage and polyadenylation specificity factor
CTD	C-terminal domain
DSE	Downstream sequence element
DSIF	DRB sensitivity-inducing factor
eIF4A	Eukaryotic initiation factor 4A
eGFP	Enhanced green fluorescent protein
ENL	Eleven nineteen leukemia
FLIM	Fluorescence lifetime imaging microscopy
FRET	Förster resonance energy transfer
GFP	Green fluorescent protein
GMP	Guanosine monophosphate
GNP	Gold nanoparticle
GTP	Guanosine triphosphate
HAT	Histone acetyltransferase
HBI	4-Hydroxybenzliden imidazolinone
HDAC	Histone deacetylase
IMAGEtags	Intracellular multiaptamer genetic tags
ITC	Isothermal titration calorimetry

NADH	Nicotinamide adenine dinucleotide
NELF	Negative elongation factor
NMR	Nuclear magnetic resonance
ODN	Oligonucleotide
PAP	Poly (A) polymerase
PCR	Polymerase chain reaction
PDGF	Platelet-derived growth factor
PEG	Polyethylene glycol
PET	Photoinduced electron transfer
PIC	Preinitiation complex
PSMA	Prostate specific membrane antigen
P-TEFb	Positive transcription elongation factor b
PUM-HD	Human Pumilio homology domain
RNAPI	RNA polymerase I
RNAPII	RNA polymerase II
RNAPIII	RNA polymerase III
RRE	Rev responsive element
rRNA	Ribosomal RNA
SAGA	Spt-Ada-Gcn5-Acetyltransferase
SELEX	Systematic evolution of ligands by exponential enrichment
snRNP	Small nuclear ribonucleoprotein
TAF	TBP associated factor
TAR	Trans activation responsive element
TAT	Transacting activator of transcription
TBP	TATA-binding protein
TRAP	Targeted reversibly attenuated probe

TF   Transcription factor

TSS   Transcriptional start site

UTR   Untranslated region

VEGF   Vascular endothelial growth factor

ACKNOWLEDGEMENTS