Utilizing functional cell-free extracts to dissect ribonucleoprotein complex biology at single-molecule resolution

Elizabeth Duran | Andreas Schmidt | Robb Welty | Ameya P. Jalihal | Sethuramasundaram Pitchiaya | Nils G. Walter

Abstract
Cellular machineries that drive and regulate gene expression often rely on the coordinated assembly and interaction of a multitude of proteins and RNA together called ribonucleoprotein complexes (RNPs). As such, it is challenging to fully reconstitute these cellular machines recombinantly and gain mechanistic understanding of how they operate and are regulated within the complex environment that is the cell. One strategy for overcoming this challenge is to perform single molecule fluorescence microscopy studies within crude or recombinantly supplemented cell extracts. This strategy enables elucidation of the interaction and kinetic behavior of specific fluorescently labeled biomolecules within RNPs under conditions that approximate native cellular environments. In this review, we describe single molecule fluorescence microscopy approaches that dissect RNP-driven processes within cellular extracts, highlighting general strategies used in these methods. We further survey biological advances in the areas of pre-mRNA splicing and transcription regulation that have been facilitated through this approach. Finally, we conclude with a summary of practical considerations for the implementation of the featured approaches to facilitate their broader future implementation in dissecting the mechanisms of RNP-driven cellular processes.

This article is categorized under:
- RNA Structure and Dynamics > RNA Structure, Dynamics and Chemistry
- RNA Interactions with Proteins and Other Molecules > RNA-Protein Complexes
- RNA Structure and Dynamics > Influence of RNA Structure in Biological Systems

KEYWORDS
- cell extract, conformational change, macromolecular crowding, RNA:protein complex, single molecule fluorescence microscopy
1 INTRODUCTION

The transfer of genome-encoded information into functional proteins and the complex regulatory mechanisms that control gene expression across all domains of life rely on the activity of ribonucleoprotein complexes (RNPs) with RNA-binding proteins (RBPs) at their core. For example, prokaryotic clustered regularly interspersed short palindromic repeat (CRISPR) based adaptive immunity mechanisms and eukaryotic gene regulatory mechanisms such as RNA splicing and RNA interference are all RNP-driven processes. In the CRISPR-mediated silencing mechanisms of bacteria and archaea, RNPs composed of a guide RNA and a CRISPR-associated endonuclease can target specific DNA cleavage sites, thereby blocking gene expression at the DNA level (Adli, 2018; Doudna & Charpentier, 2014). In eukaryotes, precursor mRNA (pre-mRNA) splicing relies on the formation of large RNP complexes that assemble into a sequence of stable spliceosome assemblies to mediate the identification of splice sites, excision of introns, and ligation of exons leading to mature mRNA synthesis (Wilkinson et al., 2020). In the RNA silencing mechanisms of eukaryotes, RNPs known as RNA-induced silencing complexes (RISC), formed through the association of microRNAs (miRNAs) or small interfering (si)RNAs with Argonaute (AGO) proteins, target RNA transcripts for translation regulation and/or degradation (O’Brien et al., 2018). In addition, PIWI-interacting (pi)RNAs form RNP complexes with PIWI proteins of the Argonaute family and maintain germline genomic integrity by repressing transposon activity (Ozata et al., 2019).

Decades of detailed genetic, biochemical, and structural studies have largely elucidated the composition and molecular events carried out by these and other gene regulatory mechanisms. Yet what remains less clear is how RNP-mediated gene regulatory and gene expression mechanisms are influenced by the complex cellular environment.

Short-lived, often non-specific interactions between RBPs and potential targets dominate molecular interactions in the complex, crowded intracellular environment. Furthermore, RNPs rely on a network of specifically interacting regulatory components. RNP activity will thus be sensitive to small changes in their concentration and local cellular properties, especially those resulting from processes such as macromolecular phase separation (An et al., 2021; Rhine et al., 2022). All these factors may significantly perturb RNP-mediated mechanisms in cells. In eukaryotic gene silencing mechanisms, for example, it remains unclear how the activity of miRNAs or RISC-associated RBPs is modulated by competitive binding to untranslated regions (UTRs) of non-target RNAs (S. Kim et al., 2021). The accessibility of RBP target sites is also influenced by the conformational dynamics inherent to RNAs and RBPs. In pre-mRNA splicing for example, local RNA secondary structural perturbations can either enhance or inhibit the interactions with splicing factors, leading to direct effects on the splicing efficiency or identity of spliced products, with associated downstream effects (Warf & Berglund, 2010). Thus, one important aspect of studying native biochemical behavior of molecular RNP complexes is designing assays that approximate the complex environment of the cell and allow for characterization of the role that RNA conformational dynamics plays on their mechanisms.

Recent advances in single molecule fluorescence microscopy techniques such as fluorescence co-localization and single-molecule Förster resonance energy transfer (smFRET) have made it possible to follow coupled and non-equilibrium processes in gene expression (Roy et al., 2008; Serebrov & Moore, 2016). Studies with these techniques often follow two experimental designs: in vitro experiments with fully reconstituted, recombinant systems or experiments using live cells. Yet, each of these approaches impose significant constraints on studying complex RNP-driven mechanisms. On the one hand, in vitro experiments using purified proteins and mixtures of purified biomolecules offer a large degree of control and reproducibility. This allows one to address very targeted aspects of assembly dynamics and molecular rates. Yet, the large number of components and complex assemblies of much of the cellular expression machinery often precludes them from purification of single components and their functional reconstitution. Single molecule microscopy studies in cells, on the other hand, offer an alternative to fully reconstituted systems. These studies can elucidate the timing and localization of particular components within larger RNP assemblies, as in the case of tracking miRNA assembly within RNA-induced silencing complexes (Pitchiaya et al., 2012). However, due to the immense complexity of cells, biochemical RNP interactions cannot be well controlled in composition or environmental conditions, making it difficult to perturb these systems in a way that leads to direct insights about the molecular interactions underlying biological functions. Imaging within cells also raises significant measurement challenges such as limited fluorophore longevity, making it very difficult to address targeted mechanistic questions about RNP assembly and dynamics.

An alternative approach for approximating complex cellular environments that has proven effective for characterization of complex RNP biology is the reconstitution of the desired RNP system within cell extracts. Cell extracts have enabled seminal discoveries in RNA biology, including the first mechanistic insights of RNA silencing (Wightman et al., 1993) and various mechanistic aspects of RNA splicing (Padgett et al., 1983; Ruskin et al., 1984). In general, cellular extracts are prepared by culturing the targeted cell type up to the desired scale, lysing the cells, and subjecting the
lysate to repeated centrifugal cycles to isolate the soluble fraction. Combining the use of reconstituted RNPs within cell extracts with biophysical techniques that allow imaging and tracking of RNP binding interactions can provide a way to discern RNP mechanisms in approximated native conditions, especially in cases where not all components can be made recombinantly.

Designing in vitro experiments that combine single molecule approaches with cell extracts therefore is a practical alternative to studying the complex reaction pathways of cellular machineries in close-to-native cellular environments with a high degree of control. In contrast to single molecule experiments performed using highly purified biomolecules, in these experiments the reaction trajectories of individually labeled biomolecules within cell extracts are followed. This approach has been applied extensively to large RNP systems that cannot (yet) be reconstituted from purified components, such as yeast and human spliceosomes. The strategy overcomes the substantial technical challenges of detailed biochemical RNP characterization in whole cells, while maintaining some important aspects of the native cellular environment such as physiologically relevant compositions that lead to RNA processing. This is a powerful strategy for mapping the molecular mechanisms of low-abundance or highly transient interactions within biomolecular pathways as has been shown for splicing biochemistry over many years and can be applied to many other RNP-processes.

In this review, we survey single molecule strategies for investigating genomic processes in crude or partially reconstituted cell extracts. We start by providing a historical context for the use of cell extracts to approximate RNP assembly and behavior within a cellular-like environment. We subsequently describe the single molecule fluorescence microscopy approaches that have enabled mechanistic understanding of RNP biology in cellular extracts and highlight some recent insights that these approaches have generated. Finally, we conclude with a summary of practical considerations and experimental designs to follow biomolecule trajectories in cell extracts in order to facilitate future discoveries using this approach.

2 | STUDYING MULTI-COMPONENT RNP SYSTEMS IN VITRO AND IN LIVE CELLS

2.1 | Historical perspective and the need to shift toward studying RNP biochemistry in cellular environments

Historically, in vitro purification has been pivotal in the characterization and evaluation of biomolecular systems. While originally restricted to only the most robust enzymes (Payen, 1838), the sophistication of this approach has matured for over 150 years. Experiments elucidating the biochemistry of RNPs can be traced back to work in the 1950s on the ribosome (Schachman et al., 1952). Piece-by-piece interrogation of ribosome assembly provided the basis for our understanding of the chemical foundations of the world’s most conserved RNP (Moore, 1988). Since those early days, there have been great strides in RNP component structure determination (Schluenzen et al., 2000; Wimberly et al., 2000), mapping of component interactions (Korostelev et al., 2008; Noller et al., 2017), descriptions of intramolecular motion (Dashti et al., 2014), and assays of enzymatic activity (Monro, 1967). These biochemical methods provided a methodical approach to understanding the capabilities and properties of multi-component RNP molecular machines like the ribosome.

While these approaches provide a clear window into the nature of RNPs, they are insufficient to explain the scope of interactions in the cellular context. By necessity, in vitro experiments simplify the composition and physicochemical conditions of reactions relative to the complex intracellular milieu. As a result, in many cases they only partially recapitulate in vivo behaviors (Perez & Gonzalez, 2011; Wu et al., 2015). An alternative is to incorporate the extracted contents of cells that may play a necessary biological role within the system being studied in an otherwise in vitro experiment.

Over time, cell extracts have been used to probe a diverse set of biological questions. One strategy is to use them when the biological process under study is too complex, in terms of biochemical components or knowing their exact concentration for in vitro work. A classic example is the study of effects of sperm nuclei stimulating multiple facets of the mitotic cycle when exposed to extracts from ova in Xenopus laevis (Lohka & Masui, 1983). More recently, extracts from Drosophila embryos have been used to study the regulation of protein translation through development by repressive granules (Avilés-Pagán et al., 2021). Another example highlighting the specificity that cell extracts allow is the study of patient-specific misregulation resulting in neurodegeneration by studying amyloid formation in extracts from patient-derived tissue (Sanders et al., 2014). In this instance, patient cell extract was used to collect and form
experimentally tractable aggregates for insights into neurodegeneration. Recently, cell extracts have additionally been used for biophysical characterization of protein stability. By mimicking the complexity of the cellular environment, the authors were able to approximate the effect of cellular components showing that the stability of cold shock protein B monotonically increased with extract concentration (Welte & Kovermann, 2020).

The utility of using cell extracts to help investigate RNP biology may become apparent considering these and other examples, yet advances are still not trivial. For instance, mRNAs can interact dynamically with >1500 RNA-binding proteins throughout their lifetimes with varying affinities. These interaction partners decorate mRNAs in complex temporal orders (Serebrov & Moore, 2016; Singh et al., 2015). Consequently, studying specific RNP processes in the backdrop of this complexity remains an ambitious endeavor.

Over the past 10 years, live-cell single molecule fluorescence methods have made great strides toward the goal of understanding close-to-native molecular behavior with developments and increasingly widespread adoption of live-cell single particle tracking, molecular tethering (Biswas et al., 2019; Sonneveld et al., 2020) and multi-color colocalization (Boersma et al., 2019; Cialek et al., 2020; Schmidt et al., 2020). Innovations in intracellular site-specific labeling of both RNAs and proteins (Schmidt et al., 2020), genetic manipulation methods, and constant improvements in imaging technologies suggest that live-cell single-molecule methods are on their way to become increasingly widely adopted, heralding a new paradigm of intracellular biochemistry (Walter, 2019).

Despite recent advances, intracellular live-cell studies are still challenging, particularly for the study of large multi-component systems. Even with the maturation of CRISPR-based tools for gene editing, creating cell lines with mutations and deletions on a large scale, especially in essential proteins, still remains out of reach. Further, many important RNP complexes such as the ribosome are expressed at high copy numbers, making them challenging to study at single molecule resolution due to optical crowding during microscopic imaging. Moreover, molecular conformational dynamics remain difficult to study in live cells using smFRET due to rapid photobleaching. Given these challenges, combining single molecule fluorescence methods to image and track RNP biology within cell extracts provides a more tractable approach to dissecting RNP mechanisms in a cell-like environment that offers a greater degree of control.

2.2 | Cell extracts: From crude to partially reconstituted

Whole cell extracts range from crude to partially reconstituted, the latter of which involve modifying crude lysates by the addition of purified components to reconstitute the RNP system of interest (Figure 1). The absence of biochemical

---

**Figure 1** Cell extract versus purified RNP preparations for single molecule characterization. Prior to single molecule microscopy imaging, cell extract is prepared by culturing a cell line of choice to a high density and subsequently subjecting the cells to multiple rounds of centrifugation. (1) The isolated soluble fraction makes up crude cell extract. (2) In order to follow RNP interactions with other biomolecules, crude extracts can be supplemented with purified biomolecules thereby creating partially reconstituted cell extracts. (3) If desired and/or possible, crude extract can be further purified using traditional biomolecule purification strategies (e.g., affinity tags, ion exchange, gel filtration, etc.) to fully isolate and image the RNP of choice in the absence of complex cellular extract components.
manipulations to overexpress or isolate components within cell extracts allows for them to potentially retain endog-
ous RNP concentrations, or at least relative levels, as well as the biomolecular complexity and macromolecular
crowding that approximates the cellular environment (Ellis, 2001). Crude extract preparations containing genomic
alterations such as deletions or additions of specific components are typically achieved by genetically manipulating
parental cells (Folco & Reed, 2014; Ilagan & Jurica, 2014). For example, in order to characterize the role of small
nuclear RNAs in the spliceosome, a number of yeast strains have been created containing genetically engineered cDNA
complementary regions that lead to RNase H mediated depletion of the targeted endogenous RNA (Fabrizio et al.,
1989; McPheeters et al., 1989; O’Keefe et al., 1996; Pan & Prives, 1988; Stark & Rader, 2014). Crude extracts can also
be fractionated, thereafter referred to as fractionated extracts, to enrich for components from subcellular compart-
ments, such as nuclear versus cytosolic extracts (Holden & Horton, 2009). For example, the recently developed direct
analysis of ribosome targeting technology (DART) leverages the use of fractionated whole cell extracts in combination
with next generation sequencing to determine the role of 5’ UTR elements on translation (Niederer et al., 2022).

While crude and fractionated cell extracts have proven useful in providing mechanistic understanding of various
RNA processes, the protocols to prepare them can be complex and arduous, often leading to functional variation among
batches. Moreover, there is limited ability to control biomolecular concentration of specific components or ensure their
functionality. Partially reconstituted cell extracts serve as an alternative to control cell extracts that gets around some of
these limitations (Kelwick et al., 2020; Laohakunakorn et al., 2020; Perez et al., 2016). These systems are created by care-
fully mixing a small number of purified components with crude or fractionated cell extracts. Controlled biomolecular
additions are enabled by transcription or translation (Laohakunakorn et al., 2020). Small molecule and buffer exchanges
are also possible through dialysis and depletion of specific components, achieved by immunodepletion or degron tagging
(Aggarwal & Ha, 2014; H.-W. Lee et al., 2013; Y. K. Lee et al., 2017; Shetty et al., 2019; Srinivasan et al., 2017).

Extracts containing RNP complexes of interest depleted of a specific component offer a useful strategy to character-
ize the function of that missing component. Such an approach using Xenopus cell extract partially reconstituted with
CMG (Cdc45–MCM2–7–GINS) helicase complex has been employed to uncover novel DNA replication and repair
mechanisms. For example, DNA replication of templates containing DNA-protein cross-links (DPC) in the presence of
CMG-reconstituted extracts immunodepleted of the accessory helicase RTEL1 or the DNA-dependent protease SPRTN
revealed the role of these proteins in a DPC repair mechanism in which the replication fork triggers DPC proteolysis
(Larsen et al., 2019; Sparks et al., 2019). In another example, auxin-inducible degron tags were used for targeted deple-
tion of the kinetochore motor protein, Kip3, from cell extracts to characterize its role in kinetochore attachment to
microtubules through both single molecule microscopy and ensemble experiments (Torvi et al., 2022).

Like all experimental model systems, both crude and reconstituted cell extracts are associated with a number of lim-
itations, such as limited control of component concentration, that may complicate the study of certain biological prob-
lems (Table 1). Another important consideration is the possibility of batch-to-batch composition variability incurred
during cell extract preparation. Despite precisely optimized protocols, subtle changes in the cell culture growth condi-
tions, as well as in the lysis and post-lysis clearing steps, could lead to the differential loss or inactivation of components
between preparations. As a result, it is critical to implement quality control measures for each preparation such as activity
assays, western blots and mass spectrometry analysis to ensure that the components of interest are present and
active in the extract prior to using it in the desired experiments. Similarly, it is advisable to implement such quality con-
trol measures over longer periods of storage and, possibly, to re-optimize the experimental conditions (i.e., temperature,
time, buffer composition, etc.). Despite these limitations, cell extracts offer an important set of capabilities that can be
powerful when combined with single molecule techniques, as exemplified by the discoveries that this approach has
enabled as detailed below. To facilitate future experimental designs, Table 1 summarizes the scope and limitations of
single-molecule experiments in cell extracts, relative to both in vitro reconstituted and live-cell systems.

3 | COMBINING SINGLE MOLECULE FLUORESCENCE MICROSCOPY AND
PARTIALLY RECONSTITUTED CELL EXTRACTS

3.1 | Fluorescence microscopy tools for single molecule imaging in cellular extracts

Single molecule experiments with cellular extracts require the same tools as classical in vitro single molecule experi-
ments. While there are potential differences in sample viscosity, turbidity, autofluorescence, and non-specific interac-
tions associated with cellular extracts, the basic measurement tenets are the same. The techniques discussed in this
review all use total internal reflection (TIRF) microscopy. While TIRF systems vary in terms of laser lines, microscopes, and cameras, these technical variations are outside of the scope of this review and have been reviewed extensively elsewhere (Niederauer et al., 2018; Oheim, 2016; Oheim et al., 2019). More specifically, here we focus on techniques with surface-immobilized biomolecules as these have been well established for use in the characterization of RNP complex assemblies within cellular extracts. Confocal techniques for solution measurements are, thus far, underutilized for experiments in cellular extracts (Currie et al., 2015; Mateu-Regué et al., 2021).

Surface immobilization techniques can facilitate observation of individual biological complexes for extended periods of time (Roy et al., 2008). This begins with passivating an optical surface (e.g., microscope slide) while adding specific attachment points for molecules of interest. Next, the cellular extracts and fluorescently labeled biomolecules are introduced to the surface. The surface is then illuminated with laser light and the behavior of individual labeled molecules can be observed. The observable behaviors include, but are not limited to conformational changes via FRET, co-localization, and molecular interactions with antibody surface captured interaction partners (Figure 2).

Over the last two decades, a number of single molecule modalities have been established to investigate multi-domain complexes under near-native conditions in various contexts. One approach is to immobilize exogenous macromolecules (typically protein or oligonucleotides) to surfaces that nucleate biological complex formation. This is a powerful tool that can be used to enrich a surface with numerous complexes. When used in conjunction with labeled endogenous components (e.g., SNAP tag fluorescent labeled protein), such as in the colocalization single-molecule spectroscopy (CoSMoS) technique (Crawford et al., 2008), one can measure how different species colocalize within biomolecular complexes.

While the classic biotin/avidin surface attachment chemistry is still popular for simpler systems, immunoprecipitation has been recently used to great effect for pulling down biological machinery directly from bulk lysate. SiMPull and SiMPLex (Aggarwal & Ha, 2014; Fareh et al., 2016; X. Wang et al., 2018) methodologies utilize biotinylated antibodies to pulldown a biomolecule of choice from the lysate. These immunoprecipitation-based approaches make it possible to study endogenous proteins of interest without the need to add exogenous proteins.

As exemplified by the biological contributions highlighted in Sections 3.2–3.4, the common principles underlying all these methods are:

<table>
<thead>
<tr>
<th>Experimental consideration</th>
<th>In vitro reconstitution</th>
<th>In live cells (in cellulo)</th>
<th>In cell extract</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature of measurement</td>
<td>Can be independently varied</td>
<td>Confined to physiological temperatures</td>
<td>Similar to in vitro, can be independently varied within optimized range</td>
</tr>
<tr>
<td>Concentration of component</td>
<td>Can be independently varied</td>
<td>Can be altered by manipulating gene expression characteristics</td>
<td>Can be extended beyond in cellulo by doping in purified components</td>
</tr>
<tr>
<td>Purity of component</td>
<td>Highly pure, but involves time-consuming purification steps</td>
<td>Complex mixture</td>
<td>Similar to in vitro, but affords local “pull-down” to enrich for specific components (Aggarwal &amp; Ha, 2014; H.-W. Lee et al., 2013; Y. K. Lee et al., 2017; Srinivasan et al., 2017)</td>
</tr>
<tr>
<td>Macromolecular Crowding</td>
<td>Can be simulated</td>
<td>Has intrinsic crowding</td>
<td>Can be simulated</td>
</tr>
<tr>
<td>Equilibrium measurements</td>
<td>Long-term equilibration is possible</td>
<td>Non-equilibrium</td>
<td>Similar to in vitro</td>
</tr>
<tr>
<td>Experimentally tractable (long term imaging)</td>
<td>Oxygen scavenging systems can enhance imaging times</td>
<td>Short term imaging only</td>
<td>Similar to in vitro</td>
</tr>
<tr>
<td>Small molecule testing and/or exchange</td>
<td>Yes. Highly controllable</td>
<td>Possible, but limited within the constraints of maintaining cell viability</td>
<td>Similar to in vitro, but retaining much of the complexity of the cell</td>
</tr>
</tbody>
</table>
3.1.1 | Enrichment of target molecules

In the case of reconstitution experiments, proteins or complexes of interest are purified based on affinity principles after heterologous expression in bacteria and/or eukaryotic systems of choice. RNAs may be synthesized by chemical synthesis or in vitro transcription. Alternatively, RNAs or proteins may be overexpressed in transcription/translation-competent extracts.

3.1.2 | Surface proximal immobilization

These strategies use various high affinity non-covalent molecular interactions, such as biotin-streptavidin binding or antibody-antigen interactions to tether a molecule or complex of interest to a surface; typically, PEG (polyethylene...
glycol) passivated glass or quartz. This has two-fold utility: (1) it allows performing single-molecule total internal reflection fluorescence (TIRF) microscopy to selectively excite target molecules proximal to the surface, and (2) allows individual, tethered molecules or complexes to be followed over time (seconds-to-tenth of minutes), enabling the detection of multiple interaction events and/or extraction of kinetic information.

### 3.2 Insights from the spliceosome

One RNP complex that has been extensively studied using single molecule fluorescence experiments in whole cell extract is the spliceosome (Crawford et al., 2008; Hoskins, Gelles, & Moore, 2011; Ray et al., 2018). Unlike molecular machines that adopt an enzymatically active complex which remains compositionally intact through multiple rounds of catalysis on a given substrate, spliceosomes assemble stepwise onto targeted pre-mRNA splice sites forming precatalytic complexes. These are then remodeled into catalytically active spliceosomes that further transition through several compositionally distinct states on the catalytic path in a process involving dozens of transiently acting factors. The large number of components (both RNA and proteins) that assemble and transition through each stage of the splice spliceting cycle precludes reconstitution of this RNP machine from purified materials. The size, dynamics, and poorly understood regulatory pathways of the spliceosome machinery also pose enormous technical challenges for using in vivo experiments to extract molecular level understanding of spliceosomal transitions. Thus, in vitro single molecule fluorescence microscopy experiments using partially reconstituted cell extracts have been implemented to provide insights about spliceosome assembly and RNA conformational dynamics at various points during the pre-mRNA splicing reaction.

A number of single molecule colocalization experiments have been implemented to investigate multiple aspects of spliceosome assembly. Hoskins et al. used yeast cell extract with genetically engineered SNAP-tags on various splicing components to follow the kinetics of spliceosome assembly on a model pre-mRNA substrate with multi-wavelength CoSMoS (Hoskins, Friedman, et al., 2011). Real-time kinetic analysis of the single molecule trajectories revealed the dynamic and reversible nature of spliceosome subcomplex associations during step-wise assembly. In a later study, yeast cell extracts were used in single molecule CoSMoS and CoSMoS-FRET experiments to follow how pre-mRNA dynamics were coupled to spliceosome assembly and activation (Crawford et al., 2013). The data suggest a model in which the spliceosome machinery coordinates the physical proximity of the splice sites to ensure that splicing chemistry takes place only once the catalytic complex is fully assembled.

Single molecule co-localization studies using cell extracts from mammalian cell lines have also advanced our understanding of spliceosome composition and mechanisms of human spliceosomes. For example, Cherny et al. quantified the photobleaching events exhibited by immobilized spliceosomal complexes pulled down from HEK293 cell extracts to determine the number of polypyrimidine tract-binding proteins within single spliceosome complexes (Cherny et al., 2010). Given that human pre-mRNAs are often alternatively spliced, others have used similar approaches to characterize alternative splice site selection mechanisms. For example, Hodson et al. followed the colocalization of U1 snRNPs with pre-mRNA substrate containing alternative splice sites to show that, although multiple U1 snRNPs initially bind the pre-mRNA, only a single U1 is bound after the ATP-dependent transition to complex A during spliceosomal assembly (Hodson et al., 2012). More recently, single molecule CoSMoS experiments with HEK293 cell extracts revealed insights about cross-intron and cross-exon interactions in human pre-spliceosome assembly (Braun et al., 2018). In these experiments, Braun et al. fluorescently labeled spliceosome subcomplexes in order to follow their assembly with model pre-mRNA substrates (Figure 3a). They observed that pre-mRNA dynamics of the subcomplex that recognizes the 5’ splice site (U1) was not affected by the presence of the subcomplex that recognizes and binds 3’ splice sites (U2). In contrast, they found that U2 binding is strongly accelerated when U1 is bound at the 5’ splice site either upstream (cross-intron) or downstream (cross-exon) revealing an alternative kinetic pathway for U2 recruitment during pre-spliceosome assembly that may be key to exon definition in the multi-intron containing pre-mRNAs that predominate within human genomes (Figure 3b).

In addition to providing insights about spliceosome complex assembly, smFRET experiments have revealed insights about the conformational dynamics of substrate pre-mRNAs captured bound within splicing complexes (Abelson et al., 2010; Krishnan et al., 2013). Alongside these measurements, computational methodologies such as single molecule clustering analysis (SiMCAn) have enabled computational sorting and identification of kinetic signatures associated with various steps of the splicing cycle (Blanco et al., 2015). Characterization of the kinetics of such RNA rearrangements and how those are correlated to spliceosomal transitions and interacting protein partners that mediate...
those transitions continue to expand our understanding of the kinetic proofreading mechanisms that underlie the tightly controlled regulation of spliceosomal transitions and splice site selection (Beier et al., 2019; Semlow et al., 2016).

3.3 Revealing dynamic aspects of eukaryotic transcriptional activation

Similar to splicing, the transcription of RNA is a highly dynamic process requiring the temporally controlled coordination of many RNPs to tune transcriptional fate. In eukaryotes, the transcription of protein-coding genes is carried out by RNA polymerase II (RNAP II), while transcription of ribosomal and transfer RNAs are carried out by RNAPs I and III, respectively. A large number of transcription factors interact with each RNAP to control transcription initiation, elongation, and termination.

While the identities and biochemical properties of the essential eukaryotic transcriptional factors are known, the dynamics of their interactions remain largely unknown. For example, it is known that the process of turning on eukaryotic gene expression by RNAP II is preempted by a highly coordinated assembly of a minimum of six general transcription factors (GTFs: TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH) with Mediator and RNAP II into a pre-initiation complex (PIC) at the promoter just upstream of the coding region (Hahn, 2004; Schier & Taatjes, 2020). It has also been shown that transcriptional activators and co-activators increase the transcription of particular genes through interactions with DNA elements such as enhancer or upstream activator sequences (UAS) found upstream of the transcription start site (Morse, 2022; Petrenko & Struhl, 2021). Yet many questions remain regarding how these transcriptional activators are recruited to transcription complexes, the effect of activators and co-activators on PIC assembly, and how they facilitate transitions between PIC assembly and productive elongation to stimulate transcription.

Single molecule studies of eukaryotic transcription initiation have enabled important mechanistic insights on PIC assembly. Using fully reconstituted PIC from purified GTFs and RNAP II, single molecule experiments have revealed previously averaged out aspects of the assembly dynamics. For instance, Zhang et al. used immobilized transcription templates and fluorescently labeled GTFs, show that TFIIB interreacts transiently with the promoter and undergoes a
transient-to-stable transition only in the presence of RNAP II (Zhang et al., 2016). Others have used optical or magnetic tweezers with reconstituted transcription reactions to show the role of TFIIH in ATP-dependent expansion of the transcription bubble prior to RNAP II-driven promoter escape (Fazal et al., 2015; Tomko et al., 2017). More recently, PIC assembly dynamics has been followed in cell extracts using CoSMoS in order to determine the impact of the rich repertoire of transcription factors present in cell lysates on transcription initiation (Rosen et al., 2020). Using fluorescently labeled DNA template containing UAS ± promoter, and cell lysates with fluorescently labeled RNAP II, TFIIF, TFIIE, or TFIIH, Baek et al. observed that RNAP II associated with UAS with the same kinetic “on” rate whether the promoter was present or not (Baek et al., 2021). Similarly, they observed that the general transcription factors TFIIE and TFIIF also associated with the UAS independently of the promoter, but the TFIIH did not. Together their findings support a model in which some GTFs and RNAP II can be recruited to UAS, potentially through activator interactions, in order to enhance PIC assembly and thereby increase transcription of the downstream gene (Figure 4). In a complementary study, Nguyen and colleagues employed live-cell single particle tracking experiments to follow the nuclear dynamics of 10 PIC components in budding yeast (Nguyen et al., 2021). Their findings support a highly dynamic model in which the Mediator and TFIID complexes stimulate the “trapping” of other GTFs in radially limited zones in order to promote PIC assembly. Together these findings show how single molecule studies using cell extracts can complement and bridge gaps in the mechanistic understanding of dynamic RNP processes recapitulated from fully reconstituted systems versus within the rich cellular milieu of in vivo conditions.

3.4 Functional, reconstituted cell-free systems to interrogate gene silencing mechanisms

Cell extracts have also been the cornerstone of miRNA research, enabling mechanistic interrogation of miRNA biogenesis (Betancur & Tomari, 2012; Förstemann et al., 2007; Miyoshi et al., 2010; O’Brien et al., 2018), miRISC assembly (Azuma-Mukai et al., 2008; Iwasaki et al., 2010; Kawamata et al., 2009; Kawamata & Tomari, 2011; Yoda et al., 2010) and miRISC-mediated gene regulation (Fukao et al., 2014; Mathonnet et al., 2007; Pillai et al., 2005; Wakiyama et al., 2006; Wakiyama et al., 2007; Wakiyama et al., 2018; B. Wang et al., 2006; B. Wang et al., 2007) at the ensemble level. miRNAs are evolutionarily conserved small (21–24 nucleotides) non-coding RNAs which guide miRNA-induced silencing complexes (miRISC) to fine tune gene expression through translational repression and/or mRNA degradation (Jonas & Izaurralde, 2015). These complexes are composed of Argonaute (Ago), miRNA, and mRNA-resident miRNA response elements (MREs). miRNAs typically bind MREs via incomplete complementarity, requiring only a short (7–8 nucleotide) “seed” (perfect) match. Consequently, one miRNA can regulate many mRNAs (Selbach et al., 2008) and, conversely, a single mRNA can be regulated by many miRNAs (Uhlmann et al., 2012), resulting in a complex gene regulatory network that controls 70% of all mammalian mRNAs. Although the basic mechanism of miRNA-mediated gene
silencing is widely accepted, it is still unclear how the ~200,000 miRNA molecules of the cell find, bind, and regulate their targets, which are interspersed among a pool of ~7,000,000 mRNA molecules (Milo et al., 2010). Moreover, mRNAs typically contain multiple MREs, be it for miRNAs of similar or distinct sequences (Agarwal et al., 2015), and whether all these MREs are simultaneously occupied, that is, stoichiometric aspects, for optimal regulation is unclear.

In vitro single-molecule microscopy of purified Ago proteins has revealed lateral (1-dimensional) diffusion and augmented hybridization upon seed recognition as major mechanisms of MRE recognition by minimal si/miRISC on sparsely populated, tethered individual mRNAs (Salomon et al., 2015). Recently, smFRET measurements utilizing site specifically-labeled human Ago2 (hAgo2) immunopurified from mammalian lysates have revealed mechanistic details of guide and target RNA loading by hAgo2 (Willkomm et al., 2022). Rather than finding a distinct conformational state for the ternary RNP consisting of hAgo2, guide, and target RNA, the authors observed various distinct and conformationally dynamic states required for regulatory activity. Experimental research on small DNA-guided prokaryotic Ago analogs (Cui et al., 2019) and the theoretical treatment of protein-nucleic acids interactions (Cui & Joo, 2019; Halford & Marko, 2004) have suggested that short intersegmental transfers, that is, 3-dimensional hopping, in addition to 1-dimensional scanning, may make target search more efficient.

While these efforts are significant, they do not recapitulate (1) the complexity of miRISC, a large (>1 MDa) RNP with numerous proteins that act in concert to mediate gene regulation and (2) the crowded confines of the cell. In fact, supplementing purified Ago with TNRC6B, a miRISC scaffold protein for tertiary regulators, leads to phase separation of Ago-TNRC6B-mRNA complexes in vitro and this phenomenon is correlated with accelerated mRNA deadenylation (Sheu-Gruttadauria & MacRae, 2018).

Most RISC-active cell extracts have been derived from Drosophila or Caenorhabditis elegans embryos and human cell culture, requiring the meticulous optimization of cell lysis (Fukao et al., 2014; Mathonnet et al., 2007; Wakiyama et al., 2006; Wakiyama et al., 2007; Wakiyama et al., 2018). While commercial cell-free extracts recapitulate gene silencing, they are incompetent in miRNA biogenesis and miRISC assembly (B. Wang et al., 2006; B. Wang et al., 2007; B. Wang et al., 2008), often involving non-physiological molecular manipulation, such as pre-annealing of small RNAs with mRNAs prior to extract addition. Cell-free transcription/translation extracts have the potential to overcome these limitations by expressing key miRISC factors such as AGO, TNRC6A and DICER at near physiological levels to recapitulate all facets of the miRNA pathway. The future use of such functional, reconstituted cell extracts for in vitro single-molecule imaging of miRISC, in combination with intracellular single-miRNA imaging (Pitchiaya et al., 2012; Pitchiaya et al., 2017; Pitchiaya et al., 2019) may resolve the long-standing questions pertaining to target search and miRNA-binding stoichiometry, critical aspects to consider for the pharmacokinetics of potential anti-miRNA drugs.

4 | CONCLUSIONS AND FUTURE DIRECTIONS

Many of the biological insights into pre-mRNA splicing, transcription and gene regulation discussed here were uncovered by combining the use of cell extracts and single-molecule fluorescence microscopy. These examples demonstrate the feasibility and underexplored potential of this approach, yet so far target only few of the many complex RNP genomic maintenance systems spanning the central dogma of molecular biology. We hope that the works highlighted here serve to inspire and encourage many more researchers to consider exploring cell extracts as a way to bridge recombinant in vitro and in cellulo experimental inquiries. To further facilitate the use of this approach, here we additionally describe some universal practical considerations for applying specifically single molecule fluorescence microscopy to cell extracts.

4.1 | Component concentration considerations

Investigation of any biological sample of interest employing single-molecule fluorescence microscopy requires the mitigation of several practical challenges. One of these challenges is exceeding the necessary affinity threshold between interacting partners. In vitro single molecule spectroscopy is performed at highly diluted concentration of the fluorescently labeled species to facilitate the observation to single molecule events (Holzmeister et al., 2014). This means that to study the biological interactions of two or more partners, a low dissociation constant ($K_D$) of those interactions is required.
Reconstituted cell-free systems support the introduction and/or removal of biomolecules, as previously described (Section 2.2). Careful addition or removal of components with minimal change to the near-cellular composition of the extract is a useful strategy for characterizing the role of specific biomolecules in a given RNP mechanism. Conversely, an alternative to adding recombinantly expressed components that is emerging as a powerful additional tool to study RNP systems is to express proteins of interest in translation-competent extracts (Lo Gullo et al., 2019). However, this is associated with some degree of variability in protein expression across those pathways that are highly sensitive to experimental conditions (Hunter et al., 2018).

4.2 | Fluorescent labeling of proteins

Since all single-molecule fluorescence microscopy techniques depend on a fluorescence readout, another challenge that must be met is the incorporation of one or more chromophore-labeled biomolecule. In keeping with the scope of this article, here we limit this discussion to strategies particularly suitable for cell extracts. A more exhaustive description of labeling strategies for single molecule experiments was recently published elsewhere (Ha et al., 2022).

In general, fusion of a fluorescent protein via standard molecular cloning methods represents a straightforward method to site-specifically introduce a fluorophore into the protein of interest (POI). For some fluorescent proteins such as GFP or GFP-derivatives, size of the fluorescent protein and poor photophysical properties often render it unsuitable for single molecule observation (Gust et al., 2014; Thorn, 2017). As a result, care must be taken to ensure that the fluorescent reporter chosen for single molecule measurements meets the suitable criteria of size, brightness, and lifetime required for the chosen experimental design.

A flexible strategy for introducing chromophores suitable for single molecule applications is to fuse genetically encoded self-labeling protein tags (Halo-, Snap-, and CLIP-tags) to a POI. Cell extracts derived from cell lines containing these tagged genes can then be incubated with an organic dye of desired emission wavelength for site-specific attachment to the POI (Blackstock & Chen, 2014; Cole, 2013; Liss et al., 2016). Furthermore, the covalent organic dye is buried in the active site of the self-labeling protein tag and is thereby protected from the environment with respect to photophysical perturbation, resulting in a non-environmentally affected, and potentially enhanced, fluorescence readout (Banaz et al., 2019; Erdmann et al., 2019). A potential concern of such chimeric systems is the significant size of the protein tag (~20–30 kDa) that may potentially perturb aspects of the system of interest, especially in multi-component RNP systems (Banaz et al., 2019). Additionally, precise distance measurement and/or structural dynamics may be complicated due to the size of the tag and potential dye orientation effects. These limitations make this system mainly suitable for colocalization and super resolution microscopy studies, as well as stoichiometry measurements (Banaz et al., 2019).

For applications in which a recombinantly expressed and fluorescently labeled protein is added to the cell extract, an alternative organic fluorophore-to-protein attachment strategy is to hijack the reactivity of amino acid side chains. Initial studies relied on the reactivity of primary amino moieties, for example, lysine side chain within proteins and N-hydroxysuccinimidyl (NHS) ester functionalized fluorescent probes (Nanda & Lorsch, 2014). The high abundance of lysine residues often found on protein surfaces makes site-specifically labeling very challenging (Carugo, 2008). To improve the site-specificity, the less abundant cysteine amino acid may be utilized using thiol-maleimide chemistry (Y. Kim et al., 2008). This labeling strategy comes at the expense of needing to mutate other cysteines to achieve specificity within POIs where they are abundantly present. In cases where this need is untenable, an alternative is to site-specifically incorporate an unnatural amino acid into the POI that can subsequently be bioorthogonally labeled (Chin, 2017; Lang & Chin, 2014; Ryu & Schultz, 2006). Initially, this strategy was only applicable for bacterial proteins or eukaryotic proteins recombinantly expressed in bacteria. More recently, it has been expanded to POIs in mammalian cell lines (Gust et al., 2018). While this strategy circumvents the limitation of site-specifically labeling proteins with high lysine and cysteine abundance, it comes with additional challenges of its own, briefly summarized below.

The limitations of site-specifically labeling of proteins using incorporation of non-natural amino acids can be summarized as four challenges: (i) expression of functionalized protein containing a non-natural amino acid; (ii) potential functional and structural integrity change of the POI upon functionalization; (iii) change in the functional integrity of the POI due to attachment of the fluorescence dye; and (iv) difference in fluorescent dye properties caused by the changing physicochemical microenvironment of the dye upon labeling. Despite these hurdles, this technique allows for very precise control over the labeling site and for incorporation of multiple dyes, a unique feature that can be used to address structural dynamics (T. C. Lee et al., 2016; K. J. Lee et al., 2019; K. Wang et al., 2012). To overcome photophysical perturbation of the dye, a long and flexible linker (~20 Å) was developed that increases a dye’s mobility and...
reduces hydrophobic and electrostatic interactions between the POI and the dye (Sánchez-Rico et al., 2017). Together with progress in modeling of fluorophores at specific attachment sites and as well as recent emergence of biochemical workflows to identify suitable labeling sites, these challenges are becoming more easily surmountable (Kalinin et al., 2012; Nagy et al., 2015; Schmidt et al., 2018). One might therefore predict that cell extracts will further expand as a tool that allows the dissection of RNP biology at the single molecule level.

AUTHOR CONTRIBUTIONS
Elizabeth Duran: Conceptualization (equal); funding acquisition (supporting); supervision (lead); writing – original draft (lead). Andreas Schmidt: Conceptualization (equal); funding acquisition (supporting); writing – original draft (equal). Robb Welty: Conceptualization (equal); writing – original draft (equal); writing – review and editing (equal). Ameya Jalihal: Conceptualization (equal); writing – original draft (equal); writing – review and editing (equal). Sethuramasundaram Pitchiaya: Conceptualization (equal); writing – original draft (equal); writing – review and editing (equal). Nils Walter: Conceptualization (lead); funding acquisition (lead); project administration (equal); writing – review and editing (supporting).

FUNDING INFORMATION
This work was supported by NIH R35 grant GM131922 to Nils G. Walter, an NIH IRACDA fellowship on grant K12 GM111725 and NIH K99 MOSAIC grant GM144735 to Elizabeth Duran, and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project-ID 449562930 to Andreas Schmidt.

CONFLICT OF INTEREST STATEMENT
The authors have declared no conflicts of interest for this article.

DATA AVAILABILITY STATEMENT
Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID
Elizabeth Duran https://orcid.org/0000-0002-3988-5091
Robb Welty https://orcid.org/0000-0002-8362-5297
Ameya P. Jalihal https://orcid.org/0000-0003-4569-8636
Sethuramasundaram Pitchiaya https://orcid.org/0000-0003-2529-5186
Nils G. Walter https://orcid.org/0000-0002-7301-1275

RELATED WIREs ARTICLES
Advances in single-molecule fluorescent nanosensors
Following the messenger: Recent innovations in live cell single molecule fluorescence imaging
Recent advancement of light-based single-molecule approaches for studying biomolecules

REFERENCES


---