Case for support. Single-molecule dynamics of human transcription regulation

We will apply cutting-edge super-resolution imaging techniques to dissect the mechanism of human transcription regulation at single-molecule resolution. This study will provide fundamentally new insights into the mechanism of human gene regulation, not attainable by any other method. This is an interdisciplinary project at the interface of physics, biology, and chemistry.

2.1 Background: Transcription regulation is at the heart of all cell state decisions. Indeed, most forms of cancer are linked to mutations in transcription activators(*1*), and forced expression of only four transcription activators reprograms somatic cells into pluripotent stem cells(*2*). Thus, if we had a better understanding of transcription regulation, we could find better ways to treat diseases and to engineer tissues. Transcription of messenger RNA (mRNA) in the human cell is a dynamic, multi-step process carried out by RNA Polymerase II (Pol II (*3*)). To initiate transcription, Pol II, together with five General Transcription Factors (GTFs) and other proteins, forms a so-called preinitiation complex (PIC) with the promoter (Fig. 1). Then, Pol II melts the promoter, initiates transcription, escapes the promoter, and begins elongation, often interrupted by pauses(*4*). After the first Pol II molecule clears the promoter, the next Pol II molecule can *re*-initiate transcription.

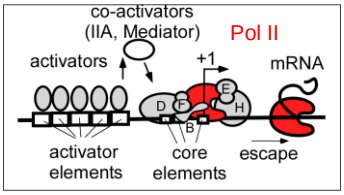

Fig. 1. Minimal set of factors required for transcription by Pol II (red).

Pol II and the five GTFs (named IIB, IID, IIF, IIE, and IIH) are sufficient to initiate transcription from strong promoters *in vitro* ('basal transcription'). GTFs have been assigned the following functions (*5*): (i) IID binds several core promoter elements; (ii) IIB binds its own core element, serves as an adapter between IID and Pol II, and positions Pol II at the transcription start (+1). (iii) IIF, together with IIB, positions Pol II at +1. (iv) IIH and IIE stimulate promoter melting and escape.

Most native promoters require additional proteins, called activators, to initiate in a regulated manner ('activated transcription' (*6*)). Activators recognize their own DNA elements (usually located upstream from the core promoter) and stimulate transcription by contacting GTFs and chromatin-modifying complexes. Specificity of activation is thought to be achieved by cooperative binding of activators in different combinations.

Although most proteins involved in the PIC are known, the *dynamics* of the PIC assembly are poorly understood. Understanding PIC dynamics is essential, because this is the step at which the cell commits to turn on a gene, and numerous regulation pathways converge on the PIC (*6*). First, it remains unclear in what order, and at what rates GTFs and Pol II assemble on the promoter. Early studies suggested that the GTFs bind in a specific order: IID, IIB, Pol II+F, IIE+H (the 'ordered assembly' model (*7*)). However, later studies suggested that Pol II binds to the promoter pre-assembled with some GTFs (the 'holoenzyme' model (*8*)).

Second, it remains unclear what happens to the PIC after Pol II escapes the promoter. Early studies suggested that some of the GTFs remain at the promoter, and then rapidly recruit the next Pol II molecule (the 're-initiation scaffold' model (*9*,*10*)). An alternative view suggests that the PIC falls apart after escape, and re-assembles for re-initiation (*11*). This 'dynamic PIC' view recently received support from studies *in vivo* (*12*) and from our single-molecule work (*13*).

Third, it remains unclear how activators stimulate transcription. Some activators interact with GTFs directly (e.g. human activator Sp1 binds IID (*14*)), providing basis for the classical 'recruitment' model of activation (*15*). However, kinetic studies generally fail to pinpoint the step(s) of PIC assembly that activators 'speed up' (for Sp1 see (*16*) and references therein). Recent *in vivo* studies found that activators bind their DNA elements transiently (dwell times of seconds -- the so-called 'hit-and-run' mode of activation (*17, 27*)). The picture is further complicated by the fact that activators act co-operatively and require co-activators (e.g. Sp1 requires co-activator IIA).

The lack of understanding of PIC dynamics could be due to inherent limitations of the biochemical approach itself. Indeed, biochemical methods generally have poor time resolution and are biased towards stable interactions that withstand purification. More importantly, biochemistry requires millions of molecules to get a signal. Thus, unless the molecules in an ensemble are homogenous and synchronous -- which is hard to achieve in the multi-component, multi-step process like transcription -- the signal gets averaged across molecules.

In contrast to biochemical methods, single-molecule methods do not require synchronization, can catch transient (sub-second) interactions and, thus, are ideal to study transcriptional dynamics (*18*). Recently, we have developed a single-molecule technology to monitor transcription initiation

by human Pol II (*13*). We have discovered that Pol II initiates from single DNA templates in a stochastic manner, suggesting indirectly that the PIC disassembles after every initiation event (consistent with the 'dynamic PIC' view). However, our study focused on RNA production, and has not directly probed the dynamics of GTFs and Pol II molecules at the promoter.

We are now in a unique and timely position to determine the mechanism of PIC assembly in initiation and re-initiation, and to determine the mechanism of PIC regulation, at the single-molecule level. Insights that we will gain may fundamentally change the way we view gene regulation, because transcription is inherently a 'single-molecule,' stochastic process from the viewpoint of a single cell (because a gene in a diploid cell is represented by only two copies (*19*,*20*)). Thus, in a broad sense, this study will be a step towards understanding whether fluctuations in transcription initiation can propagate to stochastic cell-state decisions (e.g. induction of pluripotency (*2*), transformation (*21*), and differentiation (*22*)).

2.2. Specific Objectives.

1. To determine the mechanism of PIC assembly in initiation and re-initiation (months 1-36).
2. To determine the mechanism of PIC regulation by the model activator Sp1 (months 30-48).

Part 3. Programme and Methodology.

3.1. The mechanism of PIC assembly in initiation and re-initiation (months 1-36).

In Objective 1, we will use real-time single-molecule imaging to answer three questions:

-- What is the mechanism of PIC assembly in *de novo* initiation?
-- What are the fates of GTFs after Pol II escapes the promoter?
-- What is the mechanism of PIC assembly during *re*-initiation?

3.1.1. Scope of Objective 1. To tackle the mechanism of PIC assembly, we will use a well-defined system comprised of purified human IIB, IID, IIF, IIE, IIH, and Pol II -- the factors necessary and sufficient for basal transcription. In addition to the single-molecule aspect, the use of the defined system will set this study apart from previous studies which were performed in practically fractionated nuclear extracts.

3.1.2. Experimental approach. To visualize PIC assembly at single-molecule resolution, we will use fluorescence co-localization microscopy (Fig. 2 (*23*,*13*)). We will end-tether linear DNA molecules to a glass surface in a flow cell. Each DNA will be labeled with a single fluorescent dye (e.g. 'red'), and contain a single promoter directing transcription towards the untethered DNA end. We will supply a solution of GTFs and Pol II, some of which are labeled with dyes of different colours (e.g. 'green' IID and 'blue' Pol II). We will hit the surface with lasers in total internal reflection configuration (TIR). In TIR, the electromagnetic field of the laser light exists as a thin (<150 nm), non-propagating evanescent wave next to the surface (*24*). Thus, only dyes within 150 nm from the surface will get excited. The emitted red, blue and green fluorescence will be imaged with a microscope. In an image, a tethered DNA will appear as a red spot, and IID and Pol II bound to the DNA will appear as green and blue spots co-localizing with the DNA. In contrast, unbound protein molecules will be mostly invisible, because they are, on average, out of the evanescent wave and diffuse rapidly. Thus, from the order of appearance of IID and Pol II spots we will deduce the order of assembly of the PIC. To detect Pol II escape (and to confirm PIC activity), we will supply a labeled RNA hybridization probe to the flow cell. The probe will also appear as a spot once the nascent RNA is produced. Because the promoter is directing Pol II towards the untethered DNA end, the elongating Pol II and the probe will reach the DNA end, and run off the DNA together, 'recycling' the DNA for re-initiation. We have previously shown that the outlined approach works as described (see below).

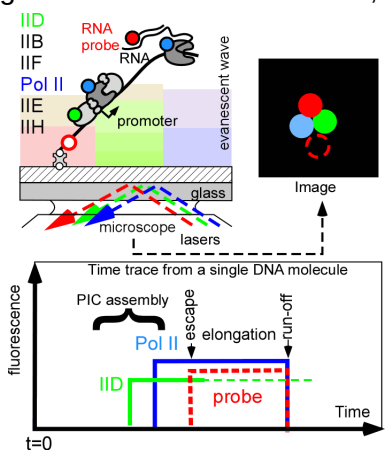

Fig. 2. Co-localization approach to visualize assembly and transcription by the PIC at single-molecule resolution. (**Top**) Schematic (**Bottom**). An idealized 3-colour time trace.

We chose the TIR approach, because: (i) it permits single-molecule imaging at high (<20 nM) protein concentrations required for Pol II activity; (ii) tethering of DNA allows tracking full transcription cycles; (iii) data collection can be multiplexed.

3.1.3. Instrument development (months 1-6)

The experimental approach proposed above cannot be cost--effectively implemented on a turnkey system. Therefore, we will

develop a custom three-colour TIR instrument (Fig. 3). Three colours is the minimum required to determine the order of PIC assembly, and to correlate PIC formation to RNA synthesis. The setup will have two features essential for the success of the project:

Active 3D stabilization with nanometer precision of the sample with respect to the objective lens (std. dev$_{xyz}$~1-5 nm) will minimize spatial drift and, thus, enable co-localization of molecules within the dimensions of the PIC (~20 nm (*3*)). Drift is an important factor during long incubations at elevated temperatures (>30 min at 30ºC) required for PIC assembly. I have previously developed a simple, robust method for active stabilization (*13*).

Multiplexed data acquisition. The Pol II system is complex (6+ protein factors), and transcription events *in vitro* are rare (0.05…4 events per DNA in 30 min (*25,13*)). To get sufficient statistics, we will have to image >1000 DNA molecules per experiment, which will require wide (~100μm) fields of view, 3 high-sensitivity cameras (one camera per 'colour') and high-power lasers (>200 mW).

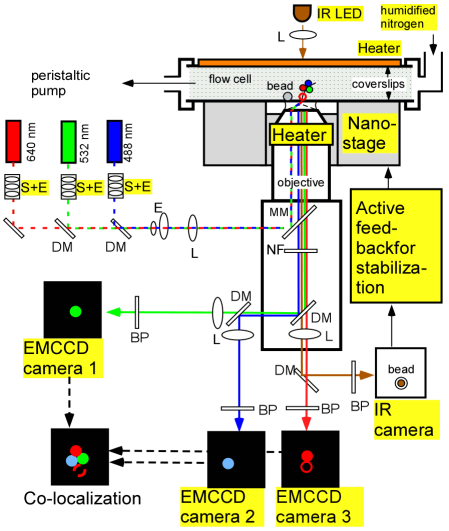

Fig. 3. Three-colour TIR fluorescence imaging instrument. Abbreviations: IR LED, infrared LED; MM, multichroic mirror; NF , notch filter; DM, dichroic mirror; BP, bandpass filter; S, beam shaper; E, beam expander; L, lens. Modules for stabilization and multiplexing are in yellow. Stabilization is achieved by IR bead tracking.

We will collaborate with the group of Dr. A. Hudson (Leicester, see letter of collaboration) on building the setup. No caveats are expected, because I have previously built and used a two-colour version of the setup (*26, 27*, *13*). To calibrate the new instrument, we will visualize transcription by unlabeled GTFs and Pol II which I have purified.

3.1.4. Preliminary data. I have already shown that the single-molecule co-localization approach detects promoter-specific transcription by human Pol II. In a representative experiment (Fig. 4), I immobilized DNA templates labeled with a 'red' fluorophore on a chemically passivized glass surface, and incubated the templates with GTFs and Pol II. Here, the promoter was directing transcription *towards* the surface. Thus, Pol II transcribed towards the surface and remained stalled there as a ternary Pol II-DNA-RNA complex (*13*). After the incubation, I supplied a green RNA probe, imaged and co-localized DNA spots and probe spots, and found robust, non-random co-localization (colour-coded yellow in Fig. 4). Controls unequivocally showed that the DNA-probe co-localization was due to promoter-specific transcription by Pol II. This experiment showed that GTFs and Pol II were active on surface-immobilized DNA (see 3.1.6), and that the co-localization approach to visualize the assembly and activity of the PIC will work.

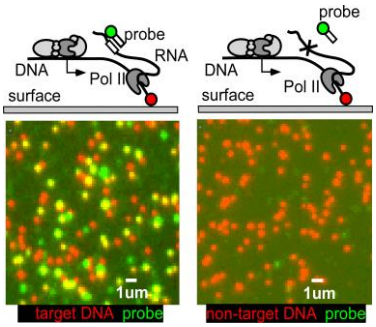

Fig. 4. Detection of transcription initiation by human Pol II. (**Top**). Schematics of experiments (see the text). (**Bottom**). Representative 16x16 μm fields of view showing DNA spots before transcription (red), and RNA probe spots after transcription (green). Co-localization is colour-coded in yellow. (**Left**). Experiment. (**Right**) Negative control with DNA template lacking the probe target.

In this proposal, we will direct transcription *away* from the glass, and monitor PIC assembly and RNA production *in real time*. To show the feasibility of this idea, my collaborator Z. Zhang (Dr. Tjian lab) and I performed proof-of-concept experiments with RNA polymerase of the T7 phage (RNAP). T7 RNAP is a simple, single-subunit enzyme that can initiate transcription, escape, elongate, and terminate without additional factors ((*28*) and references therein). Thus, we used fluorescently labeled RNAP ('red') to perform real-time two-colour transcription experiments using red RNAP and a green RNA probe (Fig. 5A). Because the promoter now directed Pol II towards the free DNA end, Pol II could 'run-off' at the end of each transcription cycle. We observed striking, repetitive cycles of coordinated binding and disappearance of RNAP and probe spots at the DNA (Fig. 5B). Each cycle consisted of binding of RNAP, followed (after a ~1 sec delay) by binding of the probe, followed (after another ~1 sec delay) by simultaneous departure of RNAP and the probe (run-off). Controls unequivocally showed that these events were initiation, escape, elongation, termination, and re-initiation by single RNAP molecules (*26*). We will use a three-colour version of this approach, combined with labeled GTFs and Pol II, to determine the dynamics of the PIC in initiation, escape, re-initiation (Objective 1), and activation (Objective 2).

3.1.5. Purification and labeling of GTFs and Pol II (months 1-24). We will express recombinant factors (IIB, IIF, and IIE) in *E. coli*, and purify them using established protocols. Native factors -- IID, IIH and Pol II -- will be purified from HeLa nuclear extracts ((*13*) and references therein). To help us jump-start the work, Dr. R. Tjian has kindly offered hybridoma supernatants with antibodies against IID, IIH, and Pol II (see letter of collaboration).

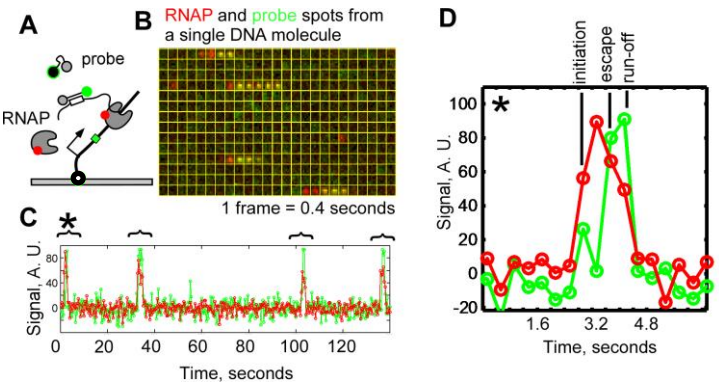

Fig. 5. Real-time visualization of promoter binding, escape, elongation, and termination T7 RNAP. (**A**).Schematic. (**B**). Two-colour movie montage from a single DNA molecule. RNAP is red, RNA probe is green, and co-localization is colour-coded yellow. (**C**). 2D representation of (B). Transcription events are shown by brackets. (**D**). The first event in C.

We will covalently and site-specifically label GTFs and Pol II with the dyes Alexa488 ('blue'), Cy3/tetra-methylrhodamine (TMR, 'green'), and Atto633/Cy5 ('red'), quantify the efficiency of labeling by absorbance- and mass-spectroscopy, and confirm activity of all factors (unlabeled and labeled) with radioactivity-based primer-extension transcription assays. I have the experience with GTF purification and labeling, and with activity assays, and will provide initial guidance to the staff. I already have labeled IIF, and my collaborators already have labeled Pol II.

To label IIB, we will use an established strategy based on the genetically encoded ybbR tag. ybbR is an 11-aminoacid peptide which reacts with CoA-activated ligands at a single serine residue (D**S**LEFIASKLA). The reaction is catalyzed by recombinant phosphopantetheinyl transferase (*29*). The strategy has been used by the Cramer lab to tag yeast IIB (*30*).

I have already labeled IIF (a heterodimer of IIFα and IIFβ) by introducing a cysteine (S311C) into a hydrophilic linker in IIFα (*31*). I expressed IIFα–S311C and IIFβ separately, labeled IIFβ–S311C using maleimide chemistry, and reconstituted the full, ~100% functional labeled IIF.

To label IIE (a heterodimer of IIEα and IIEβ), we will express IIEα and IIEβ separately, label the unique cysteine in IIEβ (C208) using maleimide chemistry, and reconstitute the full labeled IIE.

To label Pol II (a 12-subunit complex), my collaborators X. Darzacq and Z. Zhang (Dr. R. Tjian lab, see letter of collaboration) have prepared a stable human U2OS cell line expressing a fusion of the Pol II RPB1 subunit with the HaloTag. HaloTag is a 33 Kda polypeptide that reacts with haloalkane ligands (HaloLigand (*32*)). Halo-Pol II was affinity-purified and labeled *in vitro* using a HaloLigand-TMR conjugate. The native untagged Pol II was pre-emptively eliminated by propagating the cells in the presence of the toxin α-amanitin, and using an α-amanitin-resistant RPB1 mutant to encode the HaloTag fusion (*33*).

To label IID (a 13-14-subunit complex), we will collaborate with the groups of Dr. S. Cowley and I. Eperon (Leicester, see letters of collaboration). We will use gene-editing tools (TALEN (*34*), CRISPR (*35*)) or knock-ins to establish HeLa cell lines that express a core subunit of IID fused with the HaloTag (e.g. TAF4). To suppress expression of the native, untagged TAF4, we will express a small hairpin RNA against the 3'-Untranslated Region of the TAF4 mRNA. We will then affinity-purify the tagged IID, and label it *in vitro*.

To label IIH (a 10-subunit complex), we will collaborate with the groups of Dr. S. Cowley and I. Eperon to pursue the HaloTag strategy as described for IID. As targets, we will use the XPD subunit of core IIH (XPD was tagged with GFP by other groups, without affecting function (*36*)).

3.1.6. Design of DNA templates, detection of nascent RNA, and passivation of imaging surfaces.

We will use linear DNA templates containing the synthetic Supercore promoter -- the strongest known Pol II promoter (*37*) -- which will maximize the throughput of single-molecule data collection. To minimize non-specific binding of GTFs to DNA, we will use templates of a minimal length allowing surface immobilization, PIC assembly and RNA detection (254 bp, Fig 6).

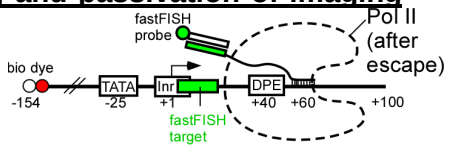

Fig. 6. DNA template for imaging PIC assembly and activity. Pol II and the DNA fragment are roughly to scale. Bio(tin) is for surface attachment. Based on Pol II footprint (*3*), RNA will be detected by fastFISH probe when Pol II is at ~+60.

To detect nascent RNA, we will employ a fluorescently labeled probe complementary to the 5' end of the transcript (Fig. 6). We will achieve near-real-time RNA detection (uncertainty <0.3 seconds) by using special probes that hybridize very rapidly at physiological temperatures (~10^7 $M^{-1}s^{-1}$, up to 1000-fold faster

than current state of the art). The method for sub-second RNA detection is based on using unstructured, complex RNA and probe sequences (we call the method 'fastFISH' *(26))*.

TIR imaging requires that the 45-polypeptide PIC assembles on a tethered promoter within 150 nm from the imaging surface. Repeated interactions of the promoter with the surface greatly increase the chances of PIC inactivation. To preserve the activity of the PIC, we have previously developed a glass passivation protocol that permits Pol II transcription from immobilized DNA at levels matching activity in solution. The method is based on hydrophobic silanes. To ensure the reproducibility of surface passivation, we will quality-control every batch of glass using surface-based biochemical transcription assays *(13)*.

3.1.7. Single-molecule measurements (months 6-36).

To dissect the dynamics of PIC assembly with our three-color TIR setup, we will assign colour 1 to one GTF, colour 2 to Pol II, and colour 3 to the fastFISH probe for RNA detection. Using different labeled GTFs as colour 1, we will reconstruct the dynamics of each GTF with respect to Pol II recruitment and escape. We will begin (months 6-24) by dissecting the dynamics of GTFs that are easy to label (or already labeled) -- IIB, IIE, IIF and Pol II -- which will establish the system, catch caveats, and generate the first publication on the dynamics of these factors in the PIC. We will then (months 24-36) dissect the dynamics of IID and IIH (which will require more time to label). Dissecting dynamics of IID and IIH will provide new insights into promoter recognition, melting, and escape mechanisms, which will have an enormous impact in the field.

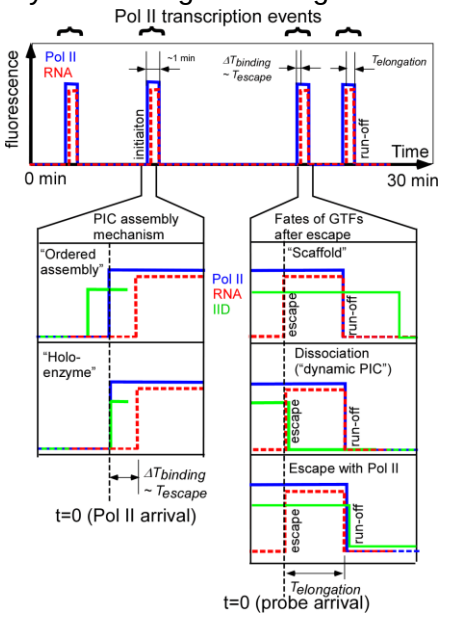

Fig. 7. Elucidating the mechanism of PIC assembly and the fate of labeled GTF (e.g. IID) after Pol II escape from three-colour time-traces (**Top**). Identification of Pol II transcription events. (**Left**). IID entry into the PIC. (**Right**). Fate of IID after escape.

In a typical experiment, we will immobilize labeled DNA templates on the passivized surface, with the promoter pointing away from the surface. We will localize each DNA to within ~10 nm (*38, 13*), and then irreversibly 'bleach' the dye on the DNA (bleaching will let us re-use the dye colour as a label on a different molecule). We will then incubate the DNA with Pol II, GTFs and the probe, and acquire synchronized movies for colors 1, 2, and 3. Then, we will use our home-written software (*13*) to identify DNA loci that repeatedly co-localized with the labeled GTF, Pol II, and the probe during the incubation (~30 min), and generate three-colour time traces (see Fig. 7 for IID as labeled GTF). Analysis of the time traces will determine (i) in what order with respect to Pol II, and at what rate, the labeled GTF entered the PIC, (ii) what was the fate of the GTF after escape, and (iii) what was the role of the GTF in re-initiation. To catch fast, cooperative interactions, as well as transient, non-productive interactions, we will use fast acquisition rates (~30 Hz).

We will develop software for automated analysis of time traces in collaboration with the group of Dr. A. Hudson (*39*). First, we will identify Pol II transcription events (Fig. 7), comprised of binding of Pol II to DNA, followed by binding of the probe (after a delay due to escape and early elongation, T_{escape}), followed by Pol II and probe co-occupancy (while Pol II elongates towards the free DNA end, $T_{elongation}$), followed by simultaneous departure of Pol II and the probe from the DNA. Based on previous studies (*40*,*13*), we expect transcription events to be rare and rapid (1-4 events in 30 min, each lasting <1 min; ~10-fold slower than T7 RNAP events in Fig. 5B-D). To ensure correct interpretation of traces, we will do controls with a null promoter mutant, with an incomplete set of NTPs, and with DNA templates lacking the probe target.

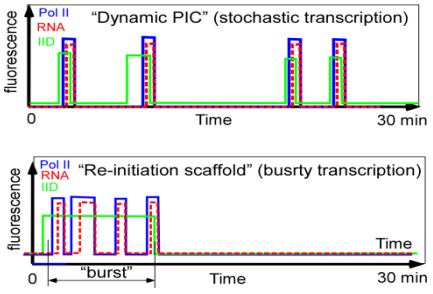

Fig. 8. Elucidating the mechanism of re-initiation from three-colour time traces.

To determine the order of PIC assembly (Fig. 7, left), we will ask if, statistically, the labeled GTF arrived before Pol II (supporting the ordered assembly model) or simultaneously with Pol II (supporting the holoenzyme model). To determine the fate of the labeled GTF after Pol II escape (Fig 7, right), we will measure its dwell time on the DNA after probe arrival. Long, stochastic dwell times after escape (minutes) will indicate that the GTF remained at the promoter after escape (supporting the

're-initiation scaffold' model). Short, stochastic dwell times after escape (seconds) will indicate that the GTF dissociated after escape (supporting the 'dynamic PIC' model). Finally, a deterministic departure of the GTF together with Pol II and the probe would indicate that the GTF remained bound to Pol II after escape. The use of home-purified oxygen scavenging enzymes and the use of nitrogen atmosphere for imaging (Fig. 3) will ensure that photobleaching of protein labels is not a significant factor in dwell time measurements.

To determine the role of the labeled GTF in *re*-initiation, we will ask whether *every,* or only *the first* transcription event from a given DNA required binding of the GTF (Fig. 8). The former scenario would suggest that the PIC re-assembled at every transcription round (the 'dynamic PIC' model), while the latter scenario would suggest that the GTF remained at the promoter, and supported a 'burst' of transcription (the 're-initiation scaffold' model).

Analysis of the promoter-interaction dynamics of the five GTFs and Pol II will answer long-standing questions on the mechanism PIC assembly and disassembly, and will result in the first data-driven, quantitative model of Pol II transcription initiation, which will identify the rate-limiting steps in initiation that are likely to be regulated (see Objective 2).

3.1.8. Caveats. Labeling. If labeling of recombinant IIB and IIE compromises their function, we will try alternative tagging sites (N- and C-termini and unstructured linkers), and try different labeling strategies (e.g. intein (*41*), and KKCK-tag (*42*)). If labeling of endogenous IIH and IID with the HaloTag compromises HeLa cells, we will try the smaller ybbr-tag (see for IIB above). To ensure successful labeling of IID and IIH, we will pursue *parallel* efforts to tag *several* subunits of IID (TAF4, TAF5, TAF6) and IIH (XPB, XPC, XPG). We note that the success of Objective 1 will *not* be compromised if a particular factor, e.g. IID, proves difficult to label. Even a minimal subset of labeled factors (e.g. IIB, IIF, IIE, and Pol II, of which two are already labeled) will provide a wealth of new insights into the PIC dynamics, and will lead to at least one high-impact publication for Objective 1, and set the benchmarks for Objective 2.

Non-specific binding of GTFs to DNA. Although promoter specificity of initiation is mainly achieved through IID, other factors can interact with DNA non-specifically. We will exclude such events based on the absence of correlation with Pol II binding, probe binding and run-off (Fig. 7). In addition, we found that post-translational modifications improve behavior of GTFs. For instance, IIF is hyper-phosphorylated *in vivo* (*43*), and, in our hands, phosphorylation of IIF by the kinase CK2 *in vitro* (*44*) eliminates DNA binding by IIF, without affecting its function in transcription.

Effects of imaging surface. If we find that Pol II activity from immobilized DNA is reduced compared to activity in solution, we will try to: (i) screen for a different siliconizing agent; (ii) orient the promoter perpendicular to the surface using bottom-up DNA bio-nanostructures *(49)*; (iii) lift the promoter DNA away from glass with ~0.3 pN of force using magnetic tweezers (AR has years of experience with magnetic tweezers, and the method is easy to implement with objective-type TIR).

3.2. The mechanism of PIC regulation by the model activator Sp1 (months 30-48).

In Objective 2, we will use the model human activator Sp1 to understand, at the single-molecule level, how the dynamic interplay between activators, co-activators, GTFs and Pol II leads to stimulated transcription. Because many activators have the organization of Sp1 (*45*), the insights that we will attain are likely to be general. Sp1 recognizes repeats of GC-rich 10-nt DNA elements ('GC-boxes') located upstream from the target promoter (Fig. 9 (*46*)). Sp1 interacts with IID (47), requires several GC-boxes and the co-activator IIA to achieve physiological levels of activation (*47*), and forms unstable oligomers with GC-box-repeats. In Objective 2, we will answer three questions:

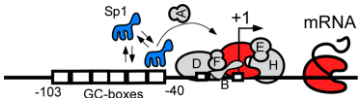

Fig. 9. Factors required for activation by Sp1 *in vitro.*

-- Which step of transcription does Sp1 target?
-- What is the role of the co-activator IIA in activation?
-- Does Sp1 (or do Sp1 oligomers) act by a 'hit-and-run' mechanism, or by 'PIC recruitment?'

3.2.1. Scope of Objective 2. We will use a defined *in vitro* system comprised of purified human Sp1, co-activator IIA, IIB, IID, IIF, IIE, IIH, and Pol II. We have previously showed that this minimal set of proteins supports 3- to 11-fold activation by Sp1 in ensemble and single-molecule assays (*13*). We will use the toolbox of labeled GTFs and Pol II developed in Objective 1, and, in addition, will prepare labeled activator Sp1 and co-activator IIA.

3.2.2. Labeling (months 24-36). To label Sp1, we will pursue the HaloTag- based strategies described for IIB, Pol II, IID and IIH. Because Sp1 is modified post-translationally, we will overexpress and purify tagged Sp1 fusions from HeLa cells. We do not expect the tagging to

interfere with Sp1 function, because GFP fusions of Sp1 have been previously shown to be functional (*48*). In parallel, we will purify and label the b168C mutant of Sp1 (lacks the Gln-rich transactivation domain), and the $Zn_{1,2,3}$ mutant of Sp1 (defective in DNA binding), to be used as negative controls (*47*). To label recombinant co-activator IIA (comprised of IIA$\alpha\beta$ and IIAγ) we will use strategies outlined for IIB, IIE, and IIF.

3.2.3. DNA templates. To create a model promoter responsive to Sp1 activation, we will insert six GC-boxes immediately upstream from Supercore promoter containing a mutant Inr element (*mInr*). We previously showed that the *mInr* mutation in Supercore is essential to achieve activation by Sp1 (*13*).

3.2.4. To determine which step of transcription Sp1 activates, we will perform three-colour single-molecule transcription experiments exactly as for Objective 1, but in the presence of *unlabeled* Sp1 and co-activator IIA. We will visualize the single-molecule dynamics of GTFs and co-activator IIA (one factor at a time labeled, colour 1), Pol II (colour 2), and RNA probe (colour 3), compare the results with those obtained in Objective 1, and ask whether Sp1 (i) increased the rate of recruitment of a GTF (e.g., IID) or Pol II; (ii) changed the order of PIC assembly (e.g. from 'ordered' to 'holoenzyme'); (iii) increased the rate of Pol II escape, or elongation; (iv) increased the fraction of productive GTF-DNA interactions; (v) stabilized a GTF at the promoter after escape.

3.2.5. To determine if Sp1 acts by 'hit-and-run' or 'PIC recruitment', and to determine the role of IIA in activation, we will use fluorescently tagged Sp1 to correlate single-molecule Sp1-DNA interactions with instantaneous transcription outcomes. Specifically, we will determine how many Sp1 molecules, and for how long, were present on the DNA during a transcription activation event. The experiments will be performed with labeled Sp1 (colour 1), one labeled GTF at a time (colour 2, including co-activator IIA), and the labeled probe (colour 3). These experiments will reconcile the results of biochemical studies (postulating that activators bring the PIC to the promoter) with the results of *in vivo* studies (postulating rapid exchange of activators). In addition, these experiments will validate if temporal co-occurrence of multiple activator molecules at promoters is required for physiological levels of activation -- a model proposed decades ago (15), but only directly testable at single-molecule resolution. Finally, we will quantify the dynamics of entry of the co-activator IIA into the PIC, and establish the correlation between IIA dynamics and Sp1-dependent activation. These experiments will directly address the generally unknown molecular mechanism of co-activator-dependent transcription stimulation.

3.3.References:
1.Vogelstein,Kinzler,Nat. Med. 10,789–99,2004. **2.**Yamanaka,Nature 460, 49–52, 2009. **3.**Cheung,Cramer,Cell 149,1431–37 ,2012). **4.**Zhou,Li,et al. Annl Rev Biochem 81,119–4 , 2012). **5.**Thomas,Chiang,Crit Rev Biochem Mol Biol 41,105–78 ,2006). **6.**Kadonaga,Cell 116, 247–57, 2004). **7.**Orphanides,et al.,Genes Dev 10, 2657–83 ,1996). **8.**Koleske,Young,Trends in Biochem. Sci 20,113–6 ,1995.**9.**Yudkovsky,Ranish et al. Nature 408,225–9 ,2000. **10**.Zawel,Kumar et al. Genes Dev 9, 1479–90 ,1995). **11**.Kadonaga,JBC 265, 2624–31 , 1990. **12**.Larson, et al., Science 332, 475–8 , 2011. **13**.Revyakin, Zhang, Coleman et al., Genes Dev 26, 1691–702 , 2012). **14**. Pugh, Tjian, Cell 61, 1187–97 , 1990. **15**.Ptashne, Nature 335, 683–9 , 1988. **16**.Narayan, Wilson, Oncogene 19, 4729–35, 2000.**17**.Voss et al., Cell 146, 544–54 , 2011. **18**.Bai, Santangelo, et al. Ann Rev Biophys 35, 343–60 , 2006. **19**.Raj, Van Oudenaarden, Ann Rev Biophys 38, 255–70 , 2009. **20**.Lionnet, Singer, EMBO Rep. 13, 313–21 , 2012. **21**.Gupta et al., Cell 146, 633–44 , 2011. **22**. Blood 81, 2844–53 , 1993. **23**.Hoskins et al., Science 331, 1289–95 , 2011. **24**.Axelrod, JCB 89, 141–5 , 1981. **25**.Dignam et al., NAR 11, 1475–89 , 1983. **26**.Zhang, Revyakin, et al., Submitted to Nat. Met , 2013. **27**.Chen, Li et al., Submitted to Cell , 2013. **28**.Tahirov et al., Nature 420, 43–50 , 2002. **29**.Yin et al., PNAS 102, 15815–20 , 2005. **30**.Treutlein et al., Mol Cell 46, 136–46 , 2012. **31**.Wang, Burton, JBC 270, 27035–44 , 1995. **32**.Los, Wood, Met Mol Biol 356, 195–208 , 2007. **33**. Darzacq et al., Nat Struct & Mol Biol 14, 796–806 , 2007. **34**.Bedell et al., Nature , 2012. **35**. Mali et al., Science , 2013. **36**.Giglia-Mari et al., PLoS 7, 15 , 2009. **37**.Juven-Gershon, et al., Nat Methods 3, 917–22 , 2006. **38**.Yildiz et al., Science 300, 2061–5 , 2003. **39**.Kelly et al., PLoS 7, 29703 , 2012. **40**.Hawley, Roeder, JBC 262, 3452–61 , 1987. **41**.Mekler et al., Cell 108, 599–614 , 2002. **42**.Eckert. et al., PloS 7, 50161 , 2012. **43**.Burton, et al., Mol and Cel Biol 8, 1602–13 , 1988. **44**.A. Újvári, Pal, Luse, JBC 286, 23160–7 , 2011. **45**.Black et al., JCPhys 188, 143–60 , 2001. **46**.Dynan, Tjian, Cell 35, 79–87 , 1983. **47**.Su, Jackson, Tjian, et al. Genes & Dev. 5, 820–6 , 1991. **48**.Pascal, Tjian, Genes Dev 5, 1646–56 , 1991. **49**.Ito, Kuwahara, BBA 380, 28–39 , 2009. **49.**Gietl et al. NAR 2012 40, 14:1-10.