

# *Analysis of single-molecule FRET trajectories of transcription complexes based on Hidden-Markov Modelling*

A Kapanidis and J Hein (for Brian Chuang) Jan29, 2009

Transcription is a crucial step in gene expression, orchestrated by RNA polymerase (RNAP), a molecular machine that transfers genetic information from DNA to RNA<sup>1,2</sup>. Bacterial transcription provides a tractable model system which provides mechanistic insights on its more complex eukaryotic counterpart. Bacterial transcription is initiated after an RNAP holoenzyme (core RNAP bound to a  $\sigma$  initiation factor) melts the double-stranded DNA (dsDNA) around the transcription start to form a transcription bubble in the RNAP-promoter DNA open complex (RP<sub>o</sub>). Subsequently, RNAP performs cycles of RNA synthesis and dissociation (abortive initiation) and at a certain point, escapes from the promoter and enters elongation. RNAP has been studied extensively using genetic, biochemical and structural methods. Recent X-ray structures<sup>3,4</sup> vastly improved our understanding of transcription, leading to mechanistic proposals, and experiments that tested these proposals and further examined RNAP function. However, crystal structures cannot *directly* capture the dynamics of transcription. The transient nature, dynamics and heterogeneity of many intermediates hinders trapping and examination; it is often unclear whether a structure corresponds to an “on-pathway” intermediate or not. Analysis of intermediates is vital for understanding transcription, since it can reveal the order of steps during transcription and identify steps regulated by transcription factors and sequence elements (e.g., activators, repressors, DNA-bending proteins, promoter sequences). Analysis of intermediates also reports on conformational changes in DNA, RNAP,  $\sigma$  factors and activators during transcription.

Real-time studies of transcription and direct probing of intermediates has become possible thanks to single-molecule methods. In contrast to ensemble methods that report on the *mean* properties of large populations, single-molecule methods can study asynchronous reactions; uncover transient intermediates; and observe full time-trajectories of pathways.

In work with Shimon Weiss (UCLA), the Kapanidis group developed single-molecule fluorescence resonance energy transfer (FRET) and alternating-laser excitation (ALEX) tools to probe RNAP structure and interactions<sup>5-8</sup>, and with Richard Ebright (Rutgers), they addressed the fate of  $\sigma^{70}$  initiation factor in elongation<sup>9</sup> and the DNA-scrunching-based mechanism of abortive initiation<sup>10</sup>. They showed that the single-stranded (ssDNA) of the transcription bubble in RP<sub>o</sub> is surprisingly dynamic, helping RNAP to sample from multiple start sites on the same DNA (L Hwang and A Kapanidis, unpublished). They also *directly* observed the real-time kinetics of abortive initiation by studying active, surface-immobilised transcription complexes using total-internal reflection fluorescence (TIRF) microscopy and ALEX<sup>11</sup>, and identified the rate-limiting step of the process. However, these studies were hampered by low temporal resolution (400 ms; cf. with ~50 ms per nucleotide addition in elongation), which did not permit clear observation of individual abortive-RNA synthesis events or single base-pair (bp) steps; the time traces were short (due to fast photobleaching) and did not report on conformational changes either before or after abortive initiation. As a result, extension of the assay was needed to study initial transcription in any detail. At Oxford, the Kapanidis group improved the temporal resolution of the TIRF assay: by monitoring FRET (using DNA $\rightarrow$ DNA FRET) within initial transcribing complexes, we observed multiple abortive-RNA synthesis events at 100-ms temporal resolution. The Kapanidis group recently improved our resolution by another >5-fold (reaching ~4 ms exposures for fields of >50 molecules) by building an objective-type TIRF with an improved detection path. The aim is to monitor the entire initial transcription at high temporal (1-5 ms) and spatial (1-bp) resolution, enabling analysis akin to the one of elongation complexes.

The Kapanidis group used our improved measuring capabilities to study transcription initiation. To study the kinetics of abortive initiation as well as conformational changes occurring before and after initial RNA synthesis, we developed a real-time FRET assay that reports on the expansion and compaction of the DNA bubble. The new assay allows a clearer view (near single base-pair resolution) of DNA-scrunching and initial transcription, providing >100 RNA synthesis and release events for each initial transcribing complex. The assay allowed the first direct observations of “un-scrunching”, the process of RNA release and backward translocation that returns RNAP to the state of open complex. The assay also reports on conformational changes in the open complex. Detailed analysis of the intermediates populated and their dynamics are in progress.

To interpret the FRET changes observed in a systematic and user-independent fashion, we need to determine the number of FRET states populated by transcription initiation complexes, the observed transitions between states, and the dwell times in each state for a statistically significant population of single molecules. Such an approach, based in Hidden Markov Modelling, is available for time-binned data obtained from imaging single-molecule fluorescence on EMCCD cameras<sup>12</sup>. HMM approaches have also been used to examine time-series with low throughput but high temporal resolution, such as single-molecule measurements with record photon-arrival times for each photon<sup>13</sup>; such methods may be useful to look at faster transitions at a later stage.

Extraction of the underlying states and the transitions between them will allow us to build quantitative predictive theoretical models that improve our understanding and control of transcription initiation; such modelling has been

performed for transcription elongation and termination, which are better characterised process that have been described by various thermodynamic<sup>14</sup>, kinetic<sup>15</sup> and stochastic<sup>16</sup> theoretical models (based on biochemical and single-molecule observations).

### **Workplan:**

1. Training in data acquisition and analysis
  - a. Record TIRF movies of transcribing complexes
  - b. Image analysis and processing to obtain time-series from individual molecules
  - c. Data interpretation: learn to identify signal changes due to photophysics, photobleaching, conformational changes.
2. HMM Applications
  - c. Application of HMM to simple two-state model systems (blinking of fluorescent immobilised DNA)
  - d. Application of HMM to complex systems with 3-5 states (initial transcribing complexes that synthesize up to 4-nucleotide RNA)
  - e. Application of HMM to complex systems with 6-8 states (initial transcribing complexes that synthesize up to 7-nucleotide RNA)
3. (Depending on time and scope) Extension of HMM to data with high temporal resolution (using photon-arrival time detection).

### **References (key papers are highlighted)**

- 1 Landick, R. (1999) Shifting RNA polymerase into overdrive. *Science* 284 (5414), 598-599
- 2 Young, B.A. et al. (2002) Views of transcription initiation. *Cell* 109 (4), 417-420
- 3 Murakami, K.S. and Darst, S.A. (2003) Bacterial RNA polymerases: the whole story. *Curr Opin Struct Biol* 13 (1), 31-39
- 4 Vassilyev, D.G. et al. (2007) Structural basis for substrate loading in bacterial RNA polymerase. *Nature* 448 (7150), 163-168
- 5 Kapanidis, A.N. et al. (2005) Alternating-laser excitation of single molecules. *Acc Chem Res* 38 (7), 523-533
- 6 Kapanidis, A.N. et al. (2004) Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc Natl Acad Sci U S A* 101 (24), 8936-8941
- 7 Lee, N.K. et al. (2007) Three-color alternating-laser excitation of single molecules: monitoring multiple interactions and distances. *Biophys J* 92 (1), 303-312
- 8 Lee, N.K. et al. (2005) Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. *Biophys J* 88 (4), 2939-2953
- 9 Kapanidis, A.N. et al. (2005) Retention of Transcription Initiation Factor sigma(70) in Transcription Elongation: Single-Molecule Analysis. *Mol Cell* 20 (3), 347-356
- 10 Kapanidis, A.N. et al. (2006) Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* 314 (5802), 1144-1147
- 11 Margeat, E. et al. (2006) Direct observation of abortive initiation and promoter escape within single immobilized transcription complexes. *Biophys J* 90 (4), 1419-1431
- 12 McKinney, S.A. et al. (2006) Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys J* 91 (5), 1941-1951
- 13 Andrec, M. et al. (2003) Direct determination of kinetic rates from single-molecule photon arrival trajectories using hidden Markov models. *Journal of Physical Chemistry A* 107 (38), 7454-7464
- 14 Yager, T.D. and von Hippel, P.H. (1991) A thermodynamic analysis of RNA transcript elongation and termination in Escherichia coli. *Biochemistry* 30 (4), 1097-1118
- 15 Bai, L. et al. (2004) Sequence-dependent kinetic model for transcription elongation by RNA polymerase. *J Mol Biol* 344 (2), 335-349
- 16 Voliotis, M. et al. (2008) Fluctuations, pauses, and backtracking in DNA transcription. *Biophys J* 94 (2), 334-348

Comments. Our data series are very similar to the ones shown in Fig8 of the McKinney paper. We would definitely like to have transition density plots such as the ones in Fig4 of that paper. What distinguishes our data from the data of the Taekjip Ha lab (senior author in the McKinney work) is that we use brighter fluorophores and therefore collect more photons per unit time. This increases our temporal resolution, or rather for the same exposure time compared to studies with less bright fluorophores, it results in narrower FRET distributions for each FRET state (note that the photon-counting noise is Poissonian). In terms of states, we expect our system to cycle between 6 to 8 states (spanning the ~0.2-0.8 FRET range); 6 of the states should be resolvable on the basis of FRET. We also have some control of the dwell in many of the states since we can slow down the RNA synthesis phase by reducing the concentration of the nucleotides. We also can limit the number of states to 3-5 states (with 3 being resolvable). Other controls (single state, two states) are also available.