Probing the Spatio-Temporal Organizations and Dynamics of Gene Expression and DNA Replication in the Mammalian Cell Nucleus

Probing the Spatio-Temporal Organizations and Dynamics of Gene Expression and DNA Replication in the Mammalian Cell Nucleus

A dissertation presented

by

Ziqing Zhao

to

The Committee on Higher Degrees in Biophysics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biophysics

Harvard University

Cambridge, Massachusetts

April 2015

Dissertation Advisor: Professor X. Sunney Xie Ziqing Zhao

Probing the Spatio-Temporal Organizations and Dynamics of Gene Expression and DNA Replication in the Mammalian Cell Nucleus

Abstract

The nucleus is an organelle of central importance to the mammalian cell. However, our understanding of the organizations and dynamics of many nuclear structures and processes remains inadequate, largely due to the difficulty in probing them *in situ*, with single-molecule sensitivity as well as ultra-high resolutions in space and time. In this dissertation, we develop approaches to interrogate, through imaging and modeling, the spatio-temporal organizations and dynamics of two key nuclear processes: gene expression and DNA replication. We first describe a novel fluorescence imaging technique, named reflected light-sheet (RLS) microscopy, that is capable of detecting single molecules with superior signal-to-background ratio inside the mammalian nucleus. By selectively illuminating only a thin section of the nucleus using a light-sheet reflected off a miniature mirror, RLS microscopy combines the capabilities of 3D optical sectioning, fast imaging speed, and applicability to single, normal-sized adherent cells. As demonstration, we apply RLS microscopy to directly monitor the DNA binding dynamics and spatio-temporal colocalization of single mammalian transcription factor molecules in live cells. By measuring their diffusion constants, DNA-bound fraction, as well as *in vivo* residence times, we resolve three distinct modes of their interaction with genomic DNA.

Furthermore, we take advantage of the prowess of RLS illumination for super-resolution microscopy (SRM), attaining resolution improvements critical for resolving nuclear structures

with high molecular density. Using RLS-SRM, we map the distribution of RNA polymerase II (RNAP II), the main workhorse of mammalian transcription, which has been proposed to heterogeneously cluster into spatially discrete foci termed "transcription factories". Leveraging on the photophysics of rhodamine-based dyes, we also develop an image analysis algorithm capable of accurately counting the copy number of RNAP II molecules in these foci. We found that majority of the foci originate from single RNAP II molecules, which exhibit no significant clustering within the length scale of the reported diameters of "transcription factories", arguing against the prevalent existence of such "factories" as previously believed. We also super-resolve in the mammalian nucleus individual DNA replication domains (RDs), and quantitatively characterize their physical morphology and propagation on a global scale. Our results support a spatio-temporal model for RD dynamics across different stages of S-phase, in which the progression of replicons along chromosomes as well as the nuclear lamina constrains the distribution of DNA synthesis sites and drives the spreading of RDs in specific spatial patterns.

Lastly, to better understand the catalytic mechanism of DNA replication at the molecular level, we simulate the dynamics of DNA polymerase, whose catalytic action is accompanied by a large nucleotide-induced movement of its finger domain, using a Langevin-type Gaussian Network Model. Our model captures the induced conformational dynamics of the polymerase upon substrate binding, and reveals its close coupling to the advancement towards transition-state along the reaction coordinate. These results demonstrate the precise role of conformational dynamics in achieving catalysis of the polymerization reaction, and indicate that the mechanism for lowering the reaction barrier through conformational motion is encoded in the structural topology of DNA polymerase. Overall, the strategies developed in this dissertation pave the way for quantitative mapping and characterization of nuclear processes at unprecedented levels of detail, both in space and in time.

Table of Contents

List of Figures and Tables

Acknowledgements

The writing of acknowledgements is always an occasion that calls for retrospection as well as introspection. As I look back on my graduate school career and try to come up with a fitting but pithy description for it, the saying that strangely comes to my mind is Julius Caesar's famous words after the Battle of Zela: "*veni, vidi, vici*", or "I came, I saw, I conquered". I certainly dare not proclaim that I have conquered anything in the past six years; instead, I would rather consider my journey as an odyssey of learning, and hence my modified version of the tricolon: I tried, I persevered, I learned. Although research is all about trying (and most of the time, failing), it is the perseverance in the face of such repeated failures that ultimately leads us to the sweet fruits, *i.e.* the things we end up learning (about science, about scientific research, and about being a scientist), that I will take away and cherish for the rest of my life.

That being said, many individuals have helped along the way to make my learning curve less steep (during those segments of ascent), or less precipitous (whenever there is a fall)! First and foremost, I would like to acknowledge my gratitude to my advisor, Prof. Sunney Xie, for his constant support and motivation, and for sharing with us his contagious passion for new science. I have learned from him that the art of this profession consists in learning to ask fundamental questions, and to tackle them with a fresh toolkit, both conceptually as well as technically. I am also grateful to him for giving me the opportunity to hone my teaching skills (through the course Chemistry 163) as well as to be involved in the development of a textbook *de novo*, thus allowing me to personally experience how quality teaching and quality research can go hand in hand. I am also fortunate to have, as members of my Dissertation Advisory Committee, Profs. Xiaowei Zhuang, Daniel Needleman, and Adam Cohen, as well as Prof. Ronald Walsworth as a member of my Dissertation Defense Committee. Their critical comments

and continuous encouragements have been crucial in helping me polish my ideas and consolidate my findings. Special thanks are due to Prof. Xiaowei Zhuang, from whom I learned so much about super-resolution microscopy during the 2nd STORM Workshop her group organized. I would also like to thank Prof. Yujie Sun, who co-supervised the DNA replication work described in Chapter 5, and whose hospitality to me during my brief research stint at Peking University I will never forget.

It is, however, to those people with whom I worked on a daily basis that I owe the greatest debt. Dr. Wei Min first introduced to me the light-sheet microscopic techniques, and Drs. Rahul Roy, Christof Gebhardt, and David Suter worked closely with me on the work presented in Chapters 2 to 4. During our collaborative effort in building the RLS and RLS-SRM microscopes, Rahul and Christof mentored me on everything ranging from optical instrumentation to biochemical sample preparation, and impressed me with, respectively, an Indian aptitude for computational analysis and the meticulousness and precision that exemplify German engineering. I also had wonderful collaborations with Qian Peter Su (both during his visit to Harvard in 2012 and during my visit to Peking University in 2014) on the work presented in Chapter 5, as well as with Prof. Hao Ge on the modeling work described in Chapter 6. The expertise, resourcefulness, and collegiality of these individuals made the collaborations highly enriching and enjoyable for me.

I am also grateful to other members of the Xie Group, past and present, including Dr. Yan Gu, Julie Chang, Dr. Sabin Mulepati, Dr. Asaf Tal, Wenting Cai, Dr. Gary Holtom, Dr. Erik Bröströmer, Dr. Yuichi Taniguchi, Srinjan Basu, Dr. Dan Fu, Shasha Chong, Chongyi Chen, Dr. Dong Xing, Zi Hertz He, Dr. Xiaohui Ni, Dr. Fa-Ke Frank Lu, Sijia Lu, Dr. Peter Sims, and Dr. Steve Mao for their stimulating discussions, generous sharing, and kind help along the way, as well as Alec Chapman and Xu Zhang, my fellow graduate school classmates, for their company.

Deserving special mention are Dr. Lin Song, Larry Valles, Dr. Patricia Purcell, Tracey Schaal, and Teri Howard, the "unsung heroes and heroines" of the Xie Group, whose support made my work go much smoother. I also appreciate the assistance of many members of the Zhuang group, in particular Graham Dempsey, Bryan Harada, Dr. Ke Xu, Jiang He, Dr. Steven Wang, Kok Hao Chen, and Dr. Ruobo Zhou, who allowed me to use their equipment and lent me materials. Lastly, I would like to acknowledge the help of Profs. Xiang-Dong Fu, Jie Xiao and Long Cai, who provided valuable advice and suggestions for the work described in Chapter 4.

And how could I forget the Biophysics Program, without a doubt *the best* graduate program at Harvard (and perhaps in the whole world)! All of us in the Program are truly blessed to have Jim and Michele, whose caring support and dedication have been a key contributor to my graduate school experience. I am deeply grateful to them for making the Biophysics Program such a wonderful home, intellectual and otherwise, for me for the past six years.

Last but definitely not the least, I am forever indebted to my parents, Drs. Yuanwei Zhao and Xiaojian Liu, who not only nurtured me every step of the way to where I am now, but also inculcated in me a broad interest in learning, scientific and humanistic alike. As to my wife, Chen, who shared with me the joys and frustrations of graduate school and who had a steadfast faith in my capability, I could only say: 取次花叢懶回顧，半緣修道半緣君.

Citations to Published Works

Parts of the work described in Chapters 2 - 4 have been published, with modification and re-organization, as the following papers:

Gebhardt, J.C.M.*, Suter, D.M.*, Roy, R., **Zhao, Z.W.**, Chapman, A.R., Basu, S., Maniatis, T., Xie, X.S. (2013). Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nat. Methods* **10**, 421-426;

Zhao, Z.W., Gebhardt, J.C.M., Suter, D.M., Xie, X.S. (2013). Reply to "Convergence of chromatin binding estimates in live cells". *Nat. Methods* **10**, 692 (correspondence);

Zhao, Z.W.*, Roy, R.*, Gebhardt, J.C.M.*, Suter, D.M.*, Chapman, A.R., Xie, X.S. (2014). Spatial organization of RNA polymerase II inside a mammalian cell nucleus revealed by reflected light-sheet super-resolution microscopy. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 681-686.

(Contributions: J.C.M.G. conceived of and built the RLS microscope prototype; R.R., Z.W.Z. and J.C.M.G. built the RLS-SRM setup; J.C.M.G., R.R. and Z.W.Z. performed characterization of the reflected light-sheet; R.R., J.C.M.G., D.M.S. and Z.W.Z. designed the experiments; D.M.S., J.C.M.G. and R.R. cloned the constructs and established the cell lines; J.C.M.G. and D.M.S. performed live-cell imaging and data analysis; Z.W.Z., R.R. and J.C.M.G. performed super-resolution imaging and data analysis; A.R.C. contributed to analysis software; S.B. contributed to cloning; X.S.X. supervised the projects; Z.W.Z., J.C.M.G., R.R., D.M.S., X.S.X. and T.M. wrote the manuscripts.)

The work described in Chapter 5 is being prepared as the following paper:

Su, Q.P.*, **Zhao, Z.W.***, Chen, M., Xie, X.S., Sun, Y. Spatio-temporal dynamics of mammalian DNA replication domains at sub-diffraction-limit resolution (manuscript in preparation).

(Contributions: Y.S. conceived of the project; Y.S., Q.P.S. and Z.W.Z. designed the experiments; Q.P.S., Z.W.Z. and M.C. performed labeling and imaging experiments; Q.P.S. and Z.W.Z. performed data analysis; Y.S. and X.S.X. supervised the project; Z.W.Z., Q.P.S. and Y.S. wrote the manuscript.)

The work described in Chapter 6 has been submitted, with modification, as the following paper:

Zhao, Z.W., Xie, X.S., Ge, H. Nonequilibrium relaxation of conformational dynamics facilitates catalytic reaction in an elastic network model of T7 DNA polymerase. *J. Phys. Chem. B* (in review).

(Contributions: H.G. conceived of the project; H.G. and Z.W.Z. designed the study; Z.W.Z. and H.G. performed the modeling and analysis; H.G. and X.S.X. supervised the project; Z.W.Z., H.G. and X.S.X. wrote the manuscript.)

(* denotes equal contribution)

To my parents, Drs. Yuanwei Zhao and Xiaojian Liu,

Who made me who I am today;

To my wife, Chen,

Whose companionship sustained me during the six-year odyssey;

And to the greatest Chinese poet of all time, who taught me:

In carefully deriving the Laws of Nature one should take pleasure,

Why shackle this body of yours with floating fame?

細推物理須行樂，何用浮名絆此身。

--- 杜甫《曲江二首・其一》

Chapter 1

Introduction: Probing the mammalian nucleus in space and time

1.1. The mammalian cell nucleus: A functional overview

The nucleus is a membrane-bound organelle of central importance to the eukaryotic cell, in which a variety of processes critical to the functioning, maintenance, and survival of the cell take place. To begin with, the nucleus is the place where the information encoded in the cell's genome is stored. As such, one of its primary functions is the expression of that information, through the processes of transcription and translation, into functional proteins that execute most of the cellular biochemistry. All aspects of the gene expression process, including which gene gets expressed at what time and in what quantity, are extremely tightly regulated. Such regulation is first and foremost carried out through the binding and unbinding of a variety of transcription factors (both general and specific) to the promoter, as well as their interactions with the components of the transcriptional machinery, chief among them RNA polymerases, during transcription initiation. Other forms of regulation that take place inside the nucleus include modulating chromatin accessibility through nucleosome remodeling, DNA methylation

and histone modifications, DNA looping, as well as the coordinated actions of a wide range of molecular players such as activators, repressors, enhancers, insulators, micro RNAs, and non-coding RNAs.

Equally important is the process of DNA replication, which maintains the integrity of the genomic information stored in the cell by ensuring that exactly the same information is passed down to the daughter cells during cell division. Mediated by DNA polymerases, this process takes place during the S-phase of the cell cycle. Similar to gene expression, DNA replication is also a highly regulated process (especially at the G1/S checkpoint), so as to ensure that each and every nucleotide in the genome gets replicated once and only once during the lifetime of the cell. Moreover, given the stringent requirement for fidelity, the newly synthesized DNA is meticulously proofread, and any damage or error made during replication is detected and repaired in a swift and precise fashion during G2-phase of the cell cycle.

Besides these two main functions, the wide variety of processes taking place in the cell nucleus range from telomere maintenance at chromosomal ends to post-transcriptional modifications of mRNA transcripts via capping, polyadenylation, and splicing. Lastly, the exchanges between the nucleus and cytoplasm are mediated by the nuclear envelope, a double-membrane structure studded with numerous nuclear pore complexes that serve as conduits for nucleo-cytoplasmic transport. Lining the inner side of the nuclear envelope is the nuclear lamina, which provides mechanical support for the nucleus as well as anchors for chromatin.

1.2. "Geography" and "history" in nuclear biology: The missing links

While we now have a rather comprehensive view of the biochemical mechanisms of the abovementioned processes as well as of the molecular machineries responsible for them, our

understanding of how they are organized spatially inside the nucleus remains inadequate. Such question is particularly pertinent in light of the fact that all of these processes and machineries co-exist in the extremely crowded nuclear space, thus suggesting that some degree of compartmentalization of nuclear functions is essential. Indeed, some of these processes are known to be carried out in certain specialized, intra-nuclear domains or "bodies" (**Figure 1.1**). For example, the transcription of ribosomal RNAs and the assembly of ribosomal components have long been known to take place inside the nucleoli, and Cajal bodies have been found to be the site of biogenesis of small nuclear ribonucleoproteins (snRNPs) involved in diverse processes such as pre-mRNA splicing and histone processing (Spector, 1993; Spector, 2001).

Despite these discoveries, the "geographical" distributions of many other key processes in the mammalian nucleus remain to be elucidated. For example, the spatial organization of transcription in the nucleus has been subject to substantial controversy over the years. Based on the observations that nascent mRNA transcripts are not homogeneously distributed in the nucleus but rather localize to discrete foci, the hypothesis of "transcription factories" that carry out the coordinated transcription of multiple genes was conceived (Jackson et al., 1993; Wansink et al., 1993). However, large variations in experimental estimates of the physical parameters of these "factories" have led to suspicions about their actual existence. Such controversy is further compounded by findings showing that genes whose transcription is to be repressed could be actively recruited to reposition to the vicinity of the nuclear lamina (Reddy et al., 2008).

The spatial distribution of another closely related process, pre-mRNA splicing, is also not well understood. Many spliceosomal snRNPs and other splicing factors are found to be primarily localized to nuclear speckles or interchromatin granule clusters (Spector and Lamond, 2011), although the morphology of these structures remain poorly defined. The observation that

active transcription sites are often found near the periphery of, rather than within these speckles has led to proposals of their role as a storage of splicing factors that can be supplied to nearby transcription sites (Xing et al., 1993; Xing et al., 1995), although such view was challenged by the recent finding that as much as 80% of splicing in fact takes place co-transcriptionally (Girard et al., 2012). Such coupling (or lack thereof) carries functional significance given that alternative splicing events have been found to be spatially uncoupled from transcription in two systems, while constitutively spliced introns tend to be spliced co-transcriptionally (Vargas et al., 2011). However, to date the extent to which splicing takes place inside nuclear speckles is not known.

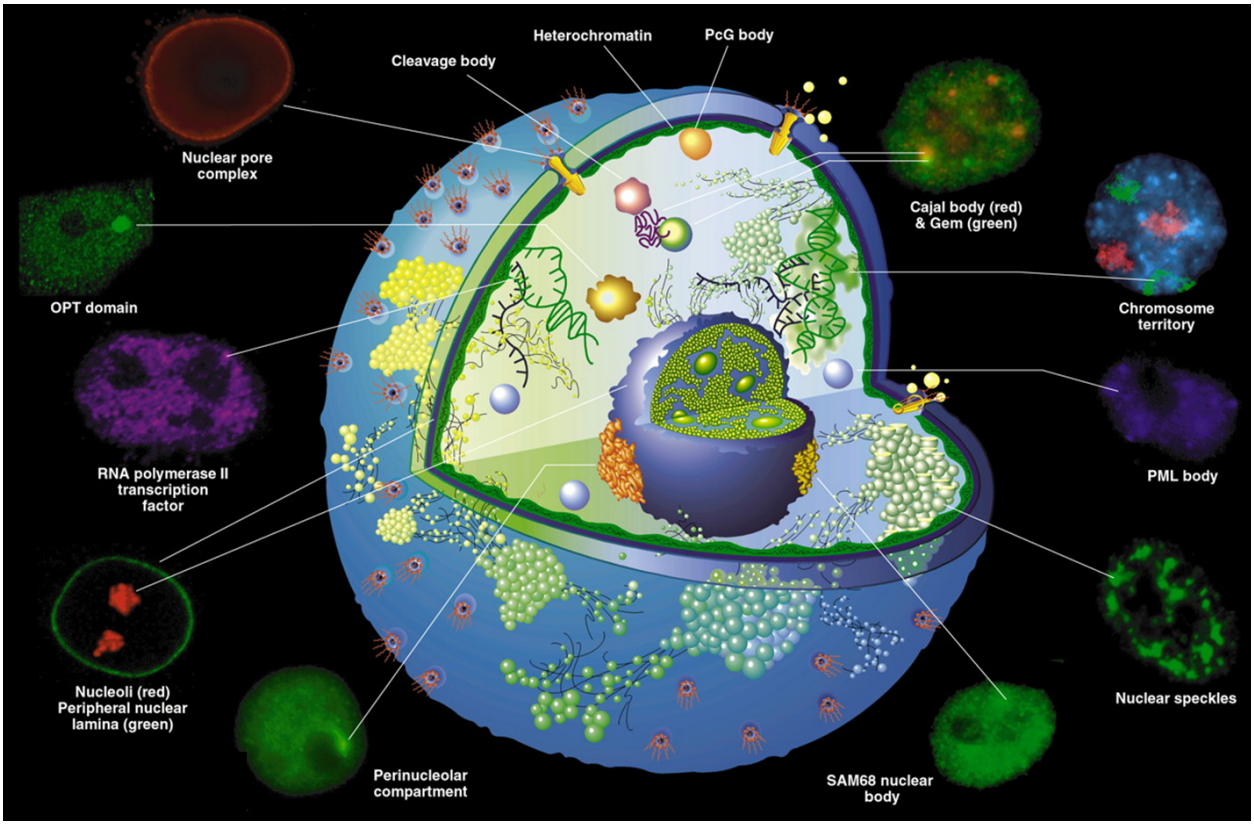

Figure 1.1. Compartmentalization of mammalian nuclear functions. A selected sample of key processes known to take place inside specialized nuclear domains are indicated. (Figure adapted from Spector, 2001).

Even in cases where the localizations of certain nuclear processes are known (either in Cartesian space or sequence space), their temporal dynamics, which are critical for the generation and maintenance of "an ever-changing but overall stable nuclear architecture" (Misteli, 2001), remain inadequately understood. If the spatial distributions of these processes can be likened to the "geography" of the nucleus, then these dynamical events, which are often asynchronous, constitute the "historical" details. For example, mammalian transcription factors have been shown to interact with DNA *in vivo* in a highly dynamic manner (McNally et al., 2000). While we now know the genomic binding sites of a large number of these factors, thanks to the advent of high-throughput sequencing and related technologies (MacQuarrie et al., 2011), our understanding of how they interact with DNA in time and whether there is any intra-nuclear spatial patterning in such interactions remain much more limited.

Another example is the organization of the genome in the three-dimensional (3D) nuclear space. Evidence from a variety of eukaryotic systems indicate that interphase chromosomes tend to occupy loosely defined and mutually exclusive "chromosome territories" in the nucleus (Cremer and Cremer, 2001). Such territorial compartmentalization could have significant bearing on gene expression, since the location of a gene within these "territories" can influence its access to the transcription and splicing machineries. While it is generally believed that these "territories" are dynamic structures, it is unclear how they change their size, shape, and position as a function of time in a live cell, and how are such dynamics correlated with specific functional consequences for gene expression. Moreover, our knowledge of the spatio-temporal organization of the mammalian genome during replication is even more limited, and till now we do not yet understand very well how DNA replication is coordinated across the genome, how are replication domains distributed in space, or how do these domains spread in time.

1.3. The need for high-resolution, sensitive, and dynamic imaging of the nucleus

The abovementioned inadequacies are to a large extent the consequence of limitations in our current capability to probe the mammalian cell nucleus in space and time. Firstly, much of our knowledge on these nuclear processes are derived from static biochemical studies, which are inadequate for revealing dynamic events taking place in a living cell. Even those studies that were performed *in vivo* using techniques such as fluorescence recovery after photobleaching (FRAP) are, given their ensemble nature, inevitably prone to masking many of the intrinsic heterogeneities among the individual molecules that participate in these processes. After all, many of the nuclear processes such as gene expression or DNA replication involve only one (or two) DNA molecule(s) and a small number of enzyme or regulatory molecules per event (Li and Xie, 2011), and are thus intrinsically stochastic in nature (Raj and van Oudenaarden, 2008).

Therefore, to illuminate these "geographical" and "historical" aspects of mammalian nuclear biology requires us to visualize these structures and processes *in situ*, with single-molecule sensitivity as well as ultra-high spatial and temporal resolutions. While single-molecule imaging in live bacterial cells can now be accomplished routinely (Biteen and Moerner, 2010; Xie et al., 2008), and has provided a wealth of valuable insights into a diverse range of bacterial processes (Cai et al., 2006; Choi et al., 2008; Chong et al., 2014; Elf et al., 2007; Garner et al., 2011; Golding et al., 2005; Hammar et al., 2012; Taniguchi et al., 2010; Wang et al., 2011; Yu et al., 2006), detecting single fluorescent biomolecules is still challenging in the nucleus of the much larger mammalian cell. This difficulty stems primarily from two inter-related features of the mammalian cell nucleus. Firstly, the mammalian nucleus is much thicker (usually 10 - 20 μm in diameter) than bacterial cells (~ 1 μm in diameter) or the cytoplasm of cultured mammalian cells (which significantly flattens upon adhering to the sample dish or

coverslip), and is notorious for its high auto-fluorescence background. As such, the fluorescence signal of a single molecule is easily overwhelmed by the out-of-focus background under a wide-field epi-fluorescence microscope. Secondly, many key molecular species in the mammalian nucleus are present at high copy numbers (**Table 1.1**), further contributing to the background and diminishing detectability for individual molecules.

Along the same line, the structural details of many of these nuclear domains lies at a spatial scale at least an order of magnitude below the diffraction limit of conventional optical microscopes, a problem that is further compounded by their high molecular density. To put things in perspective, a copy number of 100,000 per mammalian nucleus translates to a molecular density of ~ 110 molecules/μm^3, which gives a mean intermolecular distance of ~ 200 μm, assuming that the molecular species is homogenously distributed throughout the entire nuclear space. Given that many species either are excluded from certain parts of the nucleus or heterogeneously aggregate into clustered structures, the local molecular density will be much higher, leading to intermolecular distances far below the diffraction limit, which is ~ 250 nm laterally and ~ 500 nm axially. As such, resolving these structures has remained challenging until the recent advent of super-resolution microscopy (SRM) (Hell, 2007; Huang et al., 2009). Among the various approaches, stochastic optical reconstruction microscopy (STORM) and photoactivated localization microscopy (PALM) take advantage of the ability to detect single fluorescent probes attached to biomolecules while they transiently reside in their fluorescent "on" state, and subsequently reconstruct an image using the location map of their centroid positions (Betzig et al., 2006; Hess et al., 2006; Rust et al., 2006). Therefore, the high fluorescence background in the mammalian nucleus often hampers the localization of molecules with sufficient precision to reconstruct the super-resolution image.

Table 1.1. Copy numbers of some of the key molecular species in the mammalian cell nucleus.

Molecular species		Copy number	Cell type	Reference
Polymerases	RNA polymerase III	3,000 - 7,500	Hela	Jackson et al., 2000
	RNA polymerase I	15,000 - 30,000	Hela	Jackson et al., 1998; Jackson et al., 2000
	DNA polymerase α	20,000 - 60,000	-	Uri, 2013
	RNA polymerase II	60,000 - 75,000	Hela	Jackson et al., 1998; Jackson et al., 2000
General transcription factors	TFIIB	170,000	Hela	Wiśniewski et al., 2012
	TFIIE1	170,000	Hela	Wiśniewski et al., 2012
	TFIIE2	350,000	Hela	Wiśniewski et al., 2012
	TFIIF1	530,000	Hela	Wiśniewski et al., 2012
	TFIIF2	960,000	Hela	Wiśniewski et al., 2012
Specific transcription factors	Myc	60,000	MCF-7	Nieddu et al., 2005
	GR	100,000	Hela	van Steensel et al., 1995
	NF-κB (induced)	120,000	Jurkat	Hottiger et al., 1998
	STAT2 (induced)	150,000	Jurkat	Hottiger et al., 1998
	p53	160,000	MCF-7	Ma et al., 2005
	AP-2α	200,000	HepG2	Egener et al., 2005
	ER	260,000	MCF-7	Reese and Katzenellenbogen, 1992
	TCF-1	3,900,000	Jurkat	Van de Wetering et al., 1996
RNAs	miR-21 (microRNA)	12,000	Hela	Lim et al., 2003
	U6 snRNA (spliceosomal)	100,000	Hela	Lim et al., 2003
	Pre-rRNA (ribosomal)	112,000	Hela	Wolf and Schlessinger, 1977
Chromatin	Linker histone H1	33,000,000	Hela	Wiśniewski et al., 2012
	Core histones	160,000,000	Hela	Wiśniewski et al., 2012
	Genomic DNA	6,500,000,000 (base pairs)	Blood	Lander et al., 2001; Venter et al., 2001

All numbers are given in copies per nucleus.

1.4. Strategies for imaging the mammalian nucleus and their limitations

Among the many fluorescence imaging techniques, the most commonly used one has been the wide-field epi-fluorescence microscope (**Figure 1.2 A(i)**), which uses the same lens for both the excitation of fluorophores in the sample and the collection of fluorescence signal. While being easy to use, it unfortunately excites all fluorophores in the sample along the optical path, both above and below the focal plane, thus giving rise to high out-of-focus background. As such, it is now generally accepted that distinguishing single molecules present at high concentrations inside the mammalian nucleus requires a reduction in excitation volume beyond that afforded by the conventional wide-field epi-fluorescence microscope.

Over the years, a variety of strategies have been devised to provide such necessary reduction. Total internal reflection fluorescence microscopy illuminates only a thin section (a few hundred nanometers) close to the sample surface (**Figure 1.2 A(ii)**), and is thus well suited for visualizing single molecules in the cell membrane (Sako et al., 2000). However, selective excitation inside the nucleus, which is a few micrometers deep into the cell, cannot be achieved. Confocal or two-photon fluorescence microscopy (**Figure 1.2 A(iii)** and **(iv)**), on the other hand, can provide sufficient penetration depth while suppressing out-of-focus background. However, as they rely on scanning a point source across the sample, they do not permit acquisition speeds comparable to wide-field illumination schemes. While spinning-disk confocal microscopy can illuminate a large number of pixels simultaneously (Masters and So, 2008), it suffers from high photobleaching and phototoxicity since the entire sample, rather than just the in-focus plane, is illuminated, although the fluorescence signal is collected from only one plane at a time. Similarly, while multifocal multiphoton microscopy can illuminate multiple spots in parallel (Bewersdorf et al., 1998), it is also prone to significant photobleaching and phototoxicity due to

the requirement of very high laser power (to be split among multiple focal spots), and is less robust to the effects of scattering than ordinary two-photon microscopy.

Overall, none of the above techniques has proven satisfactory in their combined prowess of high sensitivity, 3D optical sectioning deep inside the cell, low photobleaching rate, and fast imaging speed, thus making them non-ideal for single-molecule and super-resolution imaging

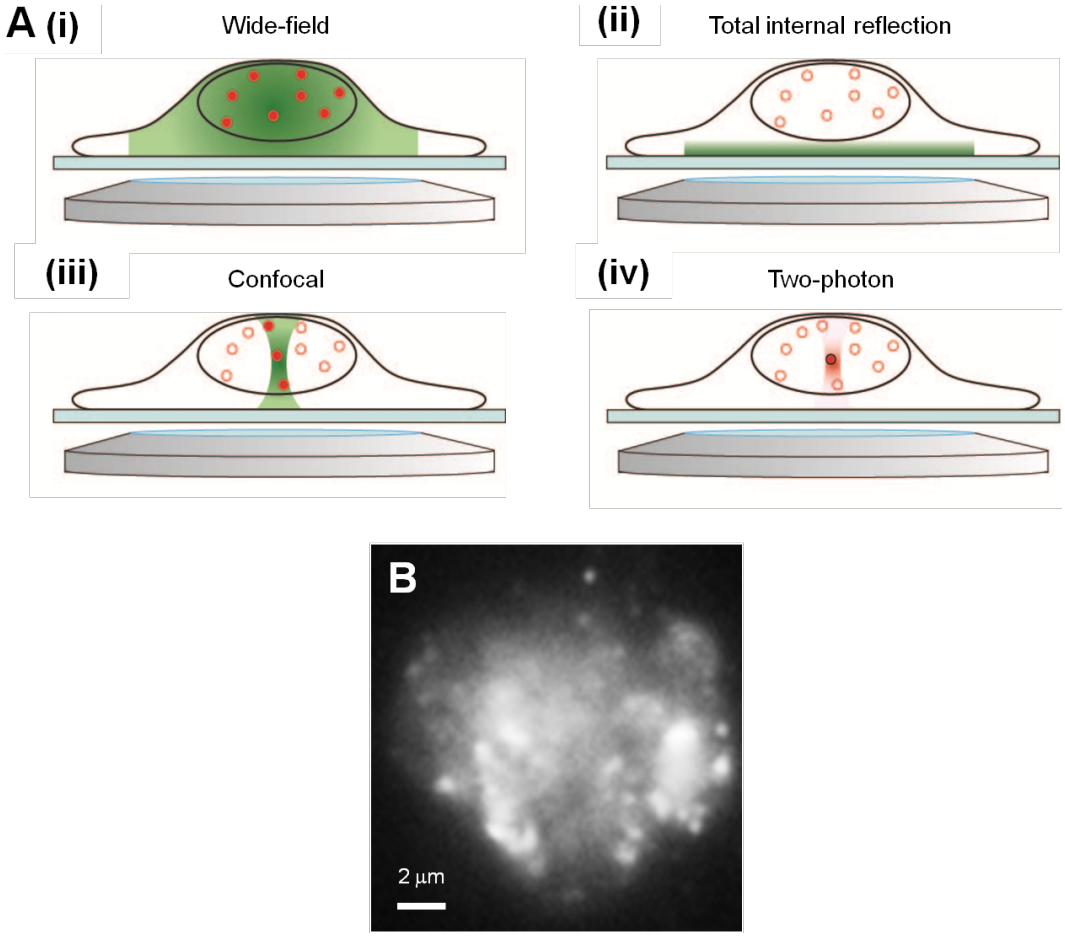

Figure 1.2. Limitations of conventional fluorescence microscopy for imaging the mammalian cell nucleus. (A) Schematic comparison of four major fluorescence imaging techniques: **(i)** wide-field epi-fluorescence; **(ii)** total internal reflection fluorescence; **(iii)** confocal; and **(iv)** two-photon fluorescence microscopy. **(B)** An epi-fluorescence image of a Hela cell nucleus stained with a Alexa 647-labeled antibody against RNA polymerase II, demonstrating the difficulty in discerning single molecules and thus achieving sub-diffraction-limit resolution.

inside the mammalian nucleus, especially for highly dense nuclear structures (**Figure 1.2 B**). It is only until the invention of light-sheet fluorescence microscopy did a generally satisfactory approach that achieves significant reduction in excitation volume and yet allows low photobleaching and fast image acquisition begin to emerge (Huisken et al., 2004), and this group of techniques will be the focus of Chapter 2.

1.5. Outline of the dissertation

This dissertation develops novel approaches, both experimental and theoretical, to probe the organization and dynamics of two key nuclear processes, gene expression and DNA replication, across a range of spatial and temporal scales. We first introduce in Chapter 2 a variant strategy of the light-sheet illumination scheme that adopts a parallel configuration for the two objectives, as opposed to the orthogonal configuration in conventional implementations. This technique, which we name reflected light-sheet (RLS) microscopy, achieves 3D optical sectioning and single-molecule sensitivity deep inside the mammalian nucleus, while at the same time allows individual, normal-sized adherent cells to be imaged, as compared to other light-sheet methods that could only be applied to embryos or cellular spheroids much larger in size. Taking advantage of the superior background suppression, we also combine RLS illumination with SRM, thus enabling us to resolve biomolecular species that are present at high molecular density in the mammalian nucleus. We demonstrate the capability of RLS microscopy for live-cell single-molecule imaging in Chapter 3 by tracking the DNA binding of two key mammalian transcription factors, glucocorticoid receptor and estrogen receptor-α, in real time. Our approach not only resolves their different modes of interaction with DNA, but also allows direct monitoring of the dynamics of single protein molecules *in vivo* without the need for additional calibrations or corrections that are commonly associated with FRAP and other techniques.

In Chapters 4 and 5, we apply RLS-SRM to probe the organizations of RNA polymerase II (RNAP II)-mediated transcription and DNA replication in the mammalian nucleus at sub-diffraction-limit resolution. Along the way, we further develop a density-based clustering algorithm for analyzing super-resolution images that is capable of accurate molecular counting with single-copy accuracy. Our global quantification of RNAP II clustering argues against the prevalent existence of "transcription factories" as previously proposed, and suggests that the model of concerted transcription of mammalian genes in these "factories" needs to be reconsidered. We also quantitatively characterize the physical morphology and distribution of DNA replication domains (RDs), which are distinct regions of the genome that tend to be replicated in a temporally specific manner, and probe their spatio-temporal dynamics across different stages of the S-phase. Our observation of a reversal in the radial direction of RD propagation from early to late S-phase suggests a model in which the progression of replicons along chromosomes helps constrain the distribution of DNA synthesis sites and drives RDs to spread in specific spatial patterns.

Lastly, in Chapter 6, we switch gear a little bit to focus on spatio-temporal dynamics at a different scale, *i.e.* the conformational dynamics within individual DNA polymerase molecules, whose catalytic action is accompanied by a large nucleotide-induced movement of its finger domain. Given the difficulty in experimentally following such dynamics with full atomic detail, we choose instead to simulate the catalytic process of the polymerase using a simplified Langevin-type Gaussian Network Model, which illustrates the precise role of conformational motion in achieving transition-state stabilization. Our results provide semi-quantitatively support for the theoretical framework of enzymatic catalysis based on a multisurface free energy description, and suggest that such approach, albeit coarse-grained, possesses the

potential to furnish significant mechanistic insights into the functioning of a variety of enzymatic systems.

In summary, the capabilities and insights provided by these studies allow us to quantitatively re-examine a diverse array of molecular processes that center on the mammalian genome, as well as the principles that govern their organizations and dynamics. It is hoped that the work presented in this dissertation will contribute to the growing efforts that regard the genome not only as a one-dimensional linear sequence but more as a 3D physical structure in space and time (Misteli, 2007), thereby shedding critical light on the spatio-temporal control of mammalian nuclear functions.

References

Betzig, E., Patterson, G.H., Sougrat, R., Lindwasser, O.W., Olenych, S., Bonifacino, J.S., Davidson, M.W., Lippincott-Schwartz, J., and Hess, H.F. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642-1645.

Bewersdorf, J., Pick, R., and Hell, S.W. (1998). Multifocal multiphoton microscopy. *Opt. Lett.* **23**, 655-657.

Biteen, J.S., and Moerner, W.E. (2010). Single-molecule and superresolution imaging in live bacteria cells. *Cold Spring Harb. Perspect. Biol.* **2**, a000448.

Cai, L., Friedman, N., and Xie, X.S. (2006). Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358-362.

Choi, P.J., Cai, L., Frieda, K., and Xie, X.S. (2008). A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* **322**, 442-446.

Chong, S., Chen, C., Ge, H., and Xie, X.S. (2014). Mechanism of transcriptional bursting in bacteria. *Cell* **158**, 314-326.

Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**, 292-301.

Egener, T., Roulet, E., Zehnder, M., Bucher, P., and Mermod, N. (2005). Proof of concept for microarray-based detection of DNA-binding oncogenes in cell extracts. *Nucleic Acids Res.* **33**, e79.

Elf, J., Li, G.-W., and Xie, X.S. (2007). Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191-1194.

Garner, E.C., Bernard, R., Wang, W., Zhuang, X., Rudner, D.Z., and Mitchison, T. (2011). Coupled, circumferential motions of the cell wall synthesis machinery and MreB filaments in *B. subtilis*. *Science* **333**, 222-225.

Girard, C., Will, C.L., Peng, J., Makarov, E.M., Kastner, B., Lemm, I., Urlaub, H., Hartmuth, K., and Lührmann, R. (2012). Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nat. Commun.* **3**, 994.

Golding, I., Paulsson, J., Zawilski, S.M., and Cox, E.C. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025-1036.

Hammar, P., Leroy, P., Mahmutovic, A., Marklund, E.G., Berg, O.G., and Elf, J. (2012). The lac repressor displays facilitated diffusion in living cells. *Science* **336**, 1595-1598.

Hell, S.W. (2007). Far-field optical nanoscopy. *Science* **316**, 1153-1158.

Hess, S.T., Girirajan, T.P., and Mason, M.D. (2006). Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.* **91**, 4258-4272.

Hottiger, M.O., Felzien, L.K., and Nabel, G.J. (1998). Modulation of cytokine-induced HIV gene expression by competitive binding of transcription factors to the coactivator p300. *EMBO J.* **17**, 3124-3134..

Huang, B., Bates, M., and Zhuang, X. (2009). Super-resolution fluorescence microscopy. *Annu. Rev. Biochem.* **78**, 993-1016.

Huisken, J., Swoger, J., Del Bene, F., Wittbrodt, J., and Stelzer, E.H. (2004). Optical sectioning deep inside live embryos by selective plane illumination microscopy. *Science* **305**, 1007-1009.

Jackson, D.A., Hassan, A.B., Errington, R.J., and Cook, P.R. (1993). Visualization of focal sites of transcription within human nuclei. *EMBO J.* **12**, 1059-1065.

Jackson, D.A., Iborra, F.J., Manders, E.M.M., and Cook, P.R. (1998). Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Mol. Biol. Cell* **9**, 1523-1536.

Jackson, D.A., Pombo, A., and Iborra, F.J. (2000). The balance sheet for transcription: an analysis of nuclear RNA metabolism in mammalian cells. *FASEB J.* **14**, 242-254.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.

Li, G.-W., and Xie, X.S. (2011). Central dogma at the single-molecule level in living cells. *Nature* **475**, 308-315.

Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. (2003). The microRNAs of *Caenorhabditis elegans*. *Genes Dev.* **17**, 991-1008.

Ma, L., Wagner, J., Rice, J.J., Hu, W., Levine, A.J., and Stolovitzky, G.A. (2005). A plausible model for the digital response of p53 to DNA damage. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 14266-14271.

MacQuarrie, K.L., Fong, A.P., Morse, R.H., and Tapscott, S.J. (2011). Genome-wide transcription factor binding: beyond direct target regulation. *Trends Genet.* **27**, 141-148.

Masters, B.R., and So, P.T.C. (2008). Handbook of Biomedical Nonlinear Optical Microscopy (New York, Oxford University Press).

McNally, J.G., Muller, W.G., Walker, D., Wolford, R., and Hager, G.L. (2000). The glucocorticoid receptor: Rapid exchange with regulatory sites in living cells. *Science* **287**, 1262-1265.

Misteli, T. (2001). Protein dynamics: Implications for nuclear architecture and gene expression. *Science* **291**, 843-847.

Misteli, T. (2007). Beyond the sequence: Cellular organization of genome function. *Cell* **128**, 787-800.

Nieddu, E., Melchiori, A., Pescarolo, M.P., Bagnasco, L., Biasotti, B., Licheri, B., Malacarne, D., Tortolina, L., Castagnino, N., Pasa, S., *et al.* (2005). Sequence specific peptidomimetic molecules inhibitors of a protein-protein interaction at the helix 1 level of c-Myc. *FASEB J.* **19**, 632-634.

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell* **135**, 216-226.

Reddy, K.L., Zullo, J.M., Bertolino, E., and Singh, H. (2008). Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* **452**, 243-247.

Reese, J.C., and Katzenellenbogen, B.S. (1992). Examination of the DNA-binding ability of estrogen receptor in whole cells: Implications for hormone-independent transactivation and the actions of antiestrogens. *Mol. Cell. Biol.* **12**, 4531-4538.

Rust, M.J., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793-795.

Sako, Y., Minoghchi, S., and Yanagida, T. (2000). Single-molecule imaging of EGFR signalling on the surface of living cells. *Nat. Cell Biol.* **2**, 168-172.

Spector, D.L. (1993). Macromolecular domains within the cell nucleus. *Annu. Rev. Cell Biol.* **9**, 265-315.

Spector, D.L. (2001). Nuclear domains. *J. Cell Sci.* **114**, 2891-2893.

Spector, D.L., and Lamond, A.I. (2011). Nuclear speckles. *Cold Spring Harb. Perspect. Biol.* **3**, a000646.

Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533-538.

Uri, M. (2013). http://bionumbers.hms.harvard.edu//bionumber.aspx?id=108716&ver=2. Accessed on March 30, 2015.

Van de Wetering, M., Castrop, J., Korinek, V., and Clevers, H. (1996). Extensive alternative splicing and dual promoter usage generate Tcf-1 protein isoforms with differential transcription control properties. *Mol. Cell. Biol.* **16**, 745-752.

van Steensel, B., Brink, M., van der Meulen, K., van Binnendijk, E.P., Wansink, D.G., de Jong, L., de Kloet, E.R., and van Driel, R. (1995). Localization of the glucocorticoid receptor in discrete clusters in the cell nucleus. *J. Cell Sci.* **108**, 3003-3011.

Vargas, D.Y., Shah, K., Batish, M., Levandoski, M., Sinha, S., Marras, S.A.E., Schedl, P., and Tyagi, S. (2011). Single-molecule imaging of transcriptionally coupled and uncoupled splicing. *Cell* **147**, 1054-1065.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., *et al.* (2001). The sequence of the human genome. *Science* **291**, 1304-1351.

Wang, W., Li, G.-W., Chen, C., Xie, X.S., and Zhuang, X. (2011). Chromosome organization by a nucleoid-associated protein in live bacteria. *Science* **333**, 1445-1449.

Wansink, D.G., Schul, W., van der Kraan, I., van Steensel, B., van Driel, R., and de Jong, L. (1993). Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. *J. Cell Biol.* **122**, 283-293.

Wiśniewski, J.R., Ostasiewicz, P., Duś, K., Zielińska, D.F., Gnad, F., and Mann, M. (2012). Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. *Mol. Syst. Biol.* **8**, 611.

Wolf, S.F., and Schlessinger, D. (1977). Nuclear metabolism of ribosomal RNA in growing, methionine-limited, and ethionine-treated HeLa cells. *Biochemistry* **16**, 2783-2791.

Xie, X.S., Choi, P.J., Li, G.W., Lee, N.K., and Lia, G. (2008). Single-molecule approach to molecular biology in living bacterial cells. *Annu. Rev. Biophys.* **37**, 417-444.

Xing, Y., Johnson, C., Dobner, P., and Lawrence, J. (1993). Higher level organization of individual gene transcription and RNA splicing. *Science* **259**, 1326-1330.

Xing, Y., Johnson, C.V., Moen, P.T., McNeil, J.A., and Lawrence, J. (1995). Nonrandom gene organization: structural arrangements of specific pre-mRNA transcription and splicing with SC-35 domains. *J. Cell Biol.* **131**, 1635-1647.

Yu, J., Xiao, J., Ren, X., Lao, K., and Xie, X.S. (2006). Probing gene expression in live cells, one protein molecule at a time. *Science* **311**, 1600-1603.

Chapter 2

Development of reflected light-sheet (RLS) microscopy and reflected light-sheet super-resolution microscopy (RLS-SRM) [1]

2.1. Abstract

The light-sheet microscopy approach offers an effective means to reduce excitation volume and suppress out-of-focus fluorescence background. However, conventional implementations of the technique suffer from either diminished illumination area in the focal plane or steric hindrance arising from the orthogonal placement of two high numerical aperture objectives in close proximity, thus restricting their applicability to only very large cells or cellular clusters. To overcome these hurdles, we develop reflected light-sheet (RLS) microscopy that selectively illuminates a thin section of a mammalian cell nucleus using a light-sheet reflected off a small adjustable mirror close to the sample surface. By adopting an anti-parallel configuration for the illumination and detection objectives instead of the conventional orthogonal geometry, this

[1] The work described in this chapter was performed in collaboration with Drs. J. Christof M. Gebhardt, Rahul Roy, and David M. Suter.

scheme allows optical sectioning throughout the nucleus of single normal-sized adherent cells. RLS microscopy achieves signal-to-background ratio superior to previous implementations, and enables imaging of single fluorescent protein molecules with up to 100 Hz temporal resolution. In addition, we take advantage of such background suppression and combine RLS illumination with super-resolution microscopy (SRM). The resulting technique, RLS-SRM, attains resolution improvements critical for resolving mammalian nuclear structures with high molecular density. Both RLS and RLS-SRM could be generally applied to provide real-time tracking and super-resolution mapping of key biomolecular species deep inside mammalian cells.

2.2. Introduction

2.2.1. The concept of light-sheet microscopy and its various incarnations

Even though light-sheet microscopy, also called selective plane illumination microcopy (SPIM), is a relatively recent technological invention, the concept of using a light-sheet for optical imaging dates back to the work of Siedentopf and Zsigmondy in 1903, who used a slit ultramicroscope with an illuminating light-sheet to look at gold particles in glasses (Siedentopf and Zsigmondy, 1903). Unfortunately, the idea was not well received at the time, and it would take another 90 years before it re-surfaced as a tool for imaging macroscopic biological samples (Voie et al., 1993). Many similar implementations of this principle soon followed, such as 3D light scanning macrography for surface scanning of small specimens (Huber et al., 2001) and thin light-sheet microscopy for imaging aquatic microbes (Fuchs et al., 2002). However, the use of a light-sheet for fluorescence microscopy was not realized until 2004, when the first SPIM was invented to study embryogenesis in Medaka and *Drosophila* (Huisken et al., 2004). It was only then did light-sheet microscopy finally become a mature technique with its own firmly

grounded niche in the realm of optical bioimaging (Mertz, 2011; Pampaloni et al., 2007; Stelzer, 2015).

The fundamental principle behind light-sheet microscopy is a simple one: a light-sheet of relatively constant thickness is used to selectively illuminate a thin layer of a biological sample around the focal plane of the detection objective (Greger et al., 2007; Huisken and Stainier, 2009; Reynaud et al., 2008; Ritter et al., 2008). As such, it reduces the illuminated volume and restricts sample excitation to the focal plane. In this sense, the light-sheet microscope is a wide-field version of the confocal microscope. The illumination light-sheet used normally comes in from the side in a direction perpendicular to the detection optical axis (**Figure 2.1 A**). As such, the illumination light is decoupled, both in shape and positioning, from the detection optics, and does not have to share the same physical constraints such as numerical aperture (NA) and working distance (Huisken and Stainier, 2009). The sample is placed at the intersection of the illumination and detection optical axes, whose fluorescence is then collected and imaged onto a camera. The sample could be scanned along the detection axis to generate a complete 3D image.

Over the years, various implementations of the light-sheet concept have been invented to suit the needs of different applications (**Figure 2.1 B**). In the original SPIM design (**Figure 2.1 B(i)**), the light-sheet comes in horizontally from the side, and the sample, usually embedded in a soft gel such as agarose, is immersed from the top into a water-chamber with transparent sides (Huisken et al., 2004). The advantage of such sample orientation lies in its ability to rotate the sample about the vertical axis, so that multiple-view reconstruction could be performed. However, such configuration is susceptible to motion-induced vibrations (Huisken and Stainier, 2009). To overcome this caveat, an alternative arrangement called objective-coupled planar illumination (OCPI) was devised, in which the sample is held stationary and the illumination

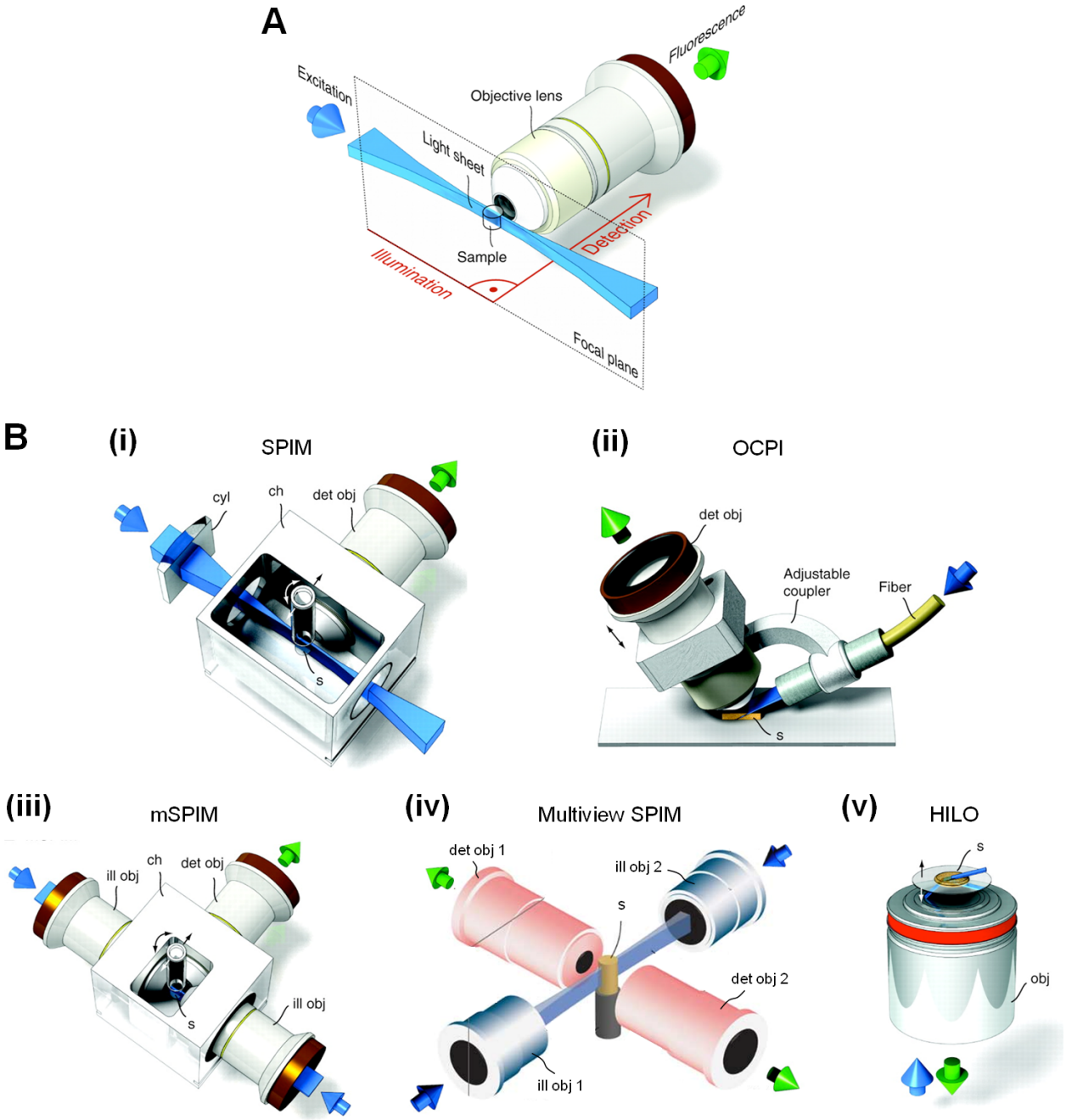

Figure 2.1. Principle and various implementations of light-sheet microscopy. (A) Scheme showing the selective illumination of a thin section of the sample around the focal plane of the detection objective using a light-sheet. **(B)** Some common implementations of the light-sheet concept: **(i)** SPIM; **(ii)** OCPI; **(iii)** mSPIM; **(iv)** multiview SPIM; and **(v)** HILO. Cyl: cylindrical lens; ch: sample chamber; det obj: detection objective; ill obj: illumination objective; obj: objective; s: sample (Figure adapted from Huisken and Stainier, 2009 and Tomer et al., 2012).

and detection optics (which are physically coupled together in an orthogonal geometry and placed at 45° with respect to the sample surface) are moved instead (**Figure 2.1 B(ii)**). In this way, vibrations in the sample are reduced, and the light-sheet can remain in the focal plane as the coupled optics are moved up and down (Holekamp et al., 2008; Turaga and Holy, 2008).

Another issue with the original SPIM design is the potential spherical aberrations induced by refractive index mismatch, since an air objective is used to generate the illumination light-sheet outside the sample chamber whereas the sample is immersed in an aqueous medium. This problem can be circumvented by using instead water-dipping objectives for both illumination and detection, as demonstrated in multi-directional SPIM (mSPIM) (Huisken and Stainier, 2007). In mSPIM, the sample chamber holds three confocally aligned water-dipping objectives, two for illumination and one for detection (**Figure 2.1 B(iii)**). Such an arrangement reduces the shadows and stripes along the direction of illumination, which are common to many SPIM setups due to absorption in the specimen (Huisken and Stainier, 2009). The use of two light-sheets also compensates for the attenuation of illumination light due to sample scattering. Recently, multiview SPIM that incorporates a total of four objectives was also developed (**Figure 2.1 B(iv)**), allowing panoramic imaging with even higher speed (Krzic et al., 2012; Schmid et al., 2013; Tomer et al., 2012). However, one caveat with these designs is that fitting three or four objectives into the sample chamber leaves little space for placing and maneuvering the specimen, thus imposing a stringent limit on the size of the sample that can be handled.

In contrast to mSPIM, a single-objective light-sheet-based technique named highly inclined and laminated optical sheet (HILO) microscopy has been invented for imaging samples much smaller in size (Tokunaga et al., 2008). Employing illumination conditions between total internal reflection (TIR) and epi-illumination, it uses an inclined light beam propagating near the edge of

the objective to image a sample at an angle just below the critical angle needed to achieve TIR. As such, the incident beam strikes the coverslip-specimen interface in a highly tilted fashion, and becomes a laminated light-sheet on the specimen side (**Figure 2.1 B(v)**). The same objective is used for the detection of fluorescence, thus resolving the crowding issue in mSPIM. However, the field of view that can be achieved by HILO microscope is limited, since reduction of light-sheet thickness is accompanied by a decrease in the illuminated area in the focal plane. In addition, the inclined nature of the illuminating laser beam still leads to out-of-focus excitation.

More recently, light-sheet microscopy has been combined with two-photon excitation to provide greater imaging depth (Truong et al., 2011), integrated with structured illumination microscopy (SIM) to further enhance its contrast (Keller et al., 2010), or coupled with fluorescence correlation spectroscopy to measure protein diffusion and interaction (Capoulade et al., 2011). Efforts have also been made to modify the physical characteristics of the light-sheet itself. For example, an illumination scheme based on a Bessel beam, instead of the traditional Gaussian beam, has been combined with two-photon excitation to achieve thinner light-sheet (< 0.5 μm) with near-isotropic resolution in 3D (Planchon et al., 2011), while the use of an Airy beam considerably expands the field of view (Vettenburg et al., 2014). Bessel beam plane illumination has also been combined with SIM to enable sub-diffraction-limit imaging in thick samples (Gao et al., 2012), although single-molecule detection was not reported for this method.

2.2.2. Limitations of conventional light-sheet microscope designs for imaging single normal-sized mammalian cells

Since their invention, microscopists have used this group of techniques for a variety of applications ranging from imaging whole embryos and whole mouse brains (Ahrens et al., 2013;

Dodt et al., 2007; Huisken et al., 2004; Keller et al., 2008), to *in vivo* heart imaging (Arnaout et al., 2007; Mickoleit et al., 2014; Scherz et al., 2008), to cell lineage tracing in *C. elegans* (Wu et al., 2011), and to laser microsurgery (Engelbrecht et al., 2007). On a finer spatial scale, light-sheet-based methods have enabled imaging of single quantum dots inside developing zebrafish (Friedrich et al., 2009), monitoring of single diffusing mRNA particles in real time in large salivary gland cell nuclei (Ritter et al., 2010), as well as super-resolution imaging of histones in cellular spheroids (Cella Zanacchi et al., 2011). The versatile power of light-sheet fluorescence microscopy has led *Nature Methods* to dub it "Method of the Year 2014".

While powerful and widely applicable, these techniques are not without limitations. Firstly, due to the steric hindrance that arises from placing two high NA objectives in close proximity, the light-sheet in conventional SPIM setups can only be positioned at distances > 10 μm above the sample surface. As such, the utility of these techniques was limited to only very large cells (a few hundred micrometers in size) or cellular clusters (Cella Zanacchi et al., 2011; Ritter et al., 2010; Verveer et al., 2007). Single normal-sized mammalian cells (such as Hela cells) that constitute the staple of model systems used in molecular and cell biology are, however, not accessible to these techniques. In addition, the full width at half maximum (FWHM) thickness of the light-sheet in many setups is > 2 μm, thereby reducing their optical sectioning power. While ultrathin light-sheets have been achieved in Bessel beam plane illumination (Planchon et al., 2011), its use of an upright orthogonal geometry between the two objectives (similar to that in OCPI) still prevents it from accessing single adherent cells in culture.

To circumvent these restrictions, we devised an illumination scheme that combines selective plane illumination with a vertical arrangement of illumination and detection objectives. In this new geometry, a disposable mirror reflects the light-sheet into a horizontal

plane close to the sample surface, thus allowing horizontal optical sectioning of single adherent mammalian cells. Such arrangement also allows us to use high NA objectives for both fluorescence detection and illumination, resulting in high photon collection efficiency as well as very thin light-sheet (< 1 μm). With this setup, we achieved single fluorescent protein imaging in live mammalian cell nucleus with superior signal-to-background ratio (SBR) and millisecond time resolution. We name our technique reflected light-sheet (RLS) microscopy. In addition, we integrated this illumination scheme with super-resolution capability using the STORM/PALM approach, resulting in a technique capable of imaging highly dense nuclear structures with sub-diffraction-limit resolution.

2.3. Results

2.3.1. Design and implementation of the reflected light-sheet microscope

We overcame the steric constraint issue in conventional SPIM setups by replacing the condenser of an inverted microscope with a vertically mounted, high NA water immersion objective. The thin vertical light-sheet generated is subsequently reflected by 90° using an adjustable miniature mirror and projected horizontally into the nucleus of the cell to be imaged (**Figures 2.2 A** and **2.3**), thus allowing sub-micrometer optical sectioning.

The optical scheme of the RLS setup is shown in **Figure 2.2 B** (see Section 2.5 for details). Briefly, the various lasers used are first collimated and aligned, and a cylindrical lens telescope and another cylindrical lens create an expanded and collimated laser beam that overfills the back aperture of the illumination objective. This objective focuses the elliptical beam incident from the top to form a diffraction-limited light-sheet, which is then reflected by the mirror. The waist of the light-sheet is positioned at the center of the field of view for optimal contrast and

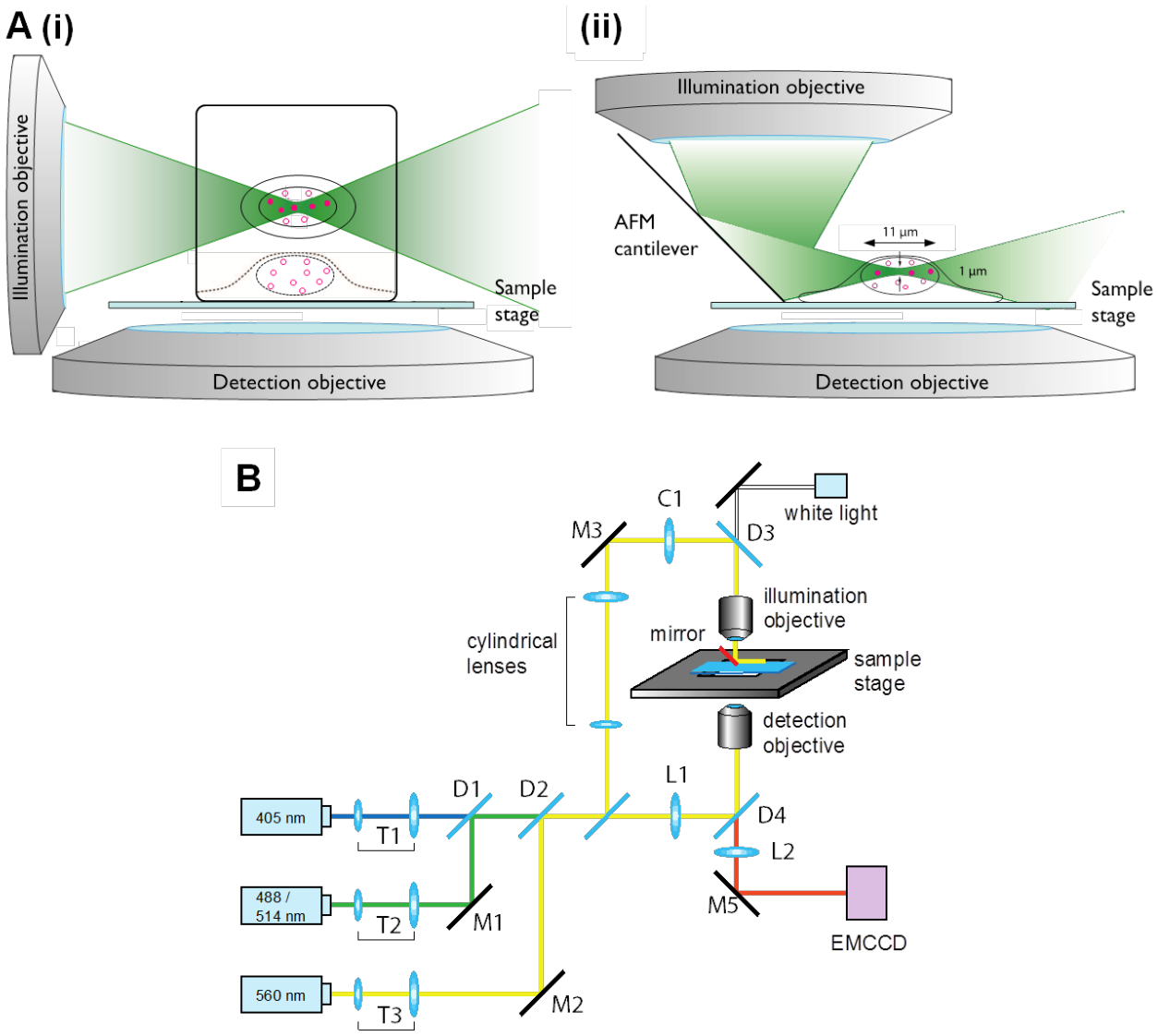

Figure 2.2. Design and scheme of RLS microscopy. (A) Schematic comparison (not drawn to scale) between **(i)** conventional light-sheet microscopy and **(ii)** RLS microscopy, in which a vertical light-sheet (green) formed by an illumination objective is reflected by 90° off an AFM cantilever next to the cell to be imaged. Fluorescence is detected by a second high NA objective placed anti-parallel to the illumination objective, and 3D optical sectioning is achieved by vertical displacement of the sample stage. Solid circles denote excited molecules while hollow circles denote unexcited molecules, and the black box in **(i)** denotes the sample chamber. **(B)** Optical scheme of the RLS setup, showing a combination of lasers of different wavelengths for multi-color imaging. C: cylindrical lens; D: dichroic beam splitter; L: lens; M: mirror; T: telescope.

sensitivity. Alternatively, the laser beams are focused into a separate path for wide-field illumination. Fluorescence emission is collected by a second high NA detection objective and focused onto an electron-multiplying CCD camera, thus allowing high sensitivity and temporal resolution. Vertical scanning is achieved by mounting the sample on a *xyz* piezo-stage and moving the stage during image acquisition while keeping the objectives fixed.

As reflector of the laser beam, we used a disposable tip-less atomic force microscopy (AFM) cantilever coated with an aluminum layer (see Section 2.5), which possesses high reflectivity across a broad spectrum in the visible range. The cantilever is mounted on a home-built *xyz*-positioning device clamped onto the illumination objective (**Figure 2.3**), granting it mechanical

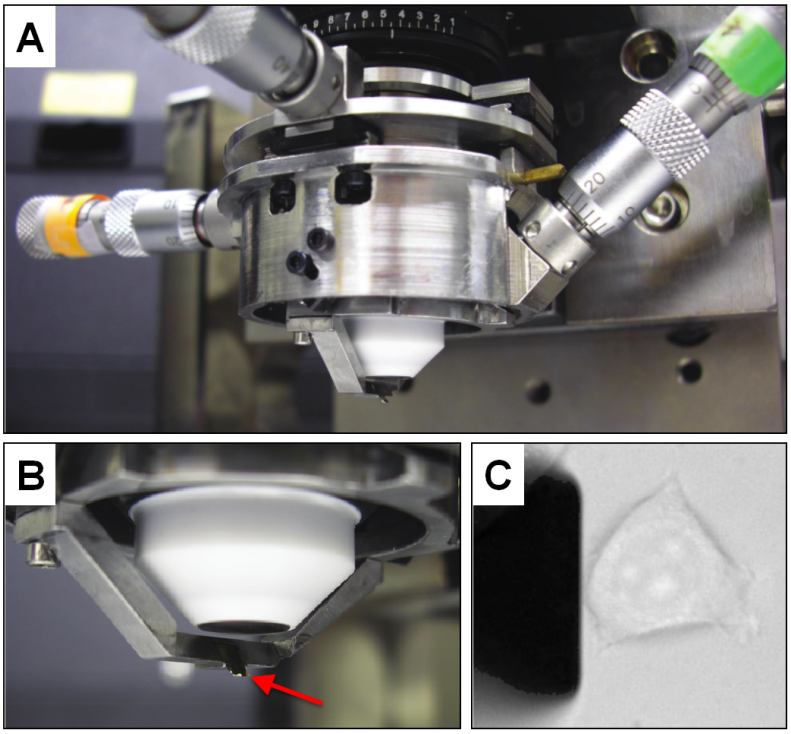

Figure 2.3. Implementation of RLS microscopy. **(A)** Photograph of the home-built *xyz*-positioning device for mounting the AFM cantilever. **(B)** Close-up view of the metallic bridge that holds the cantilever (red arrow) directly below the illumination objective. **(C)** Bright-field image of the AFM cantilever (dark wedge on the left) positioned next to a MCF-7 cell (right).

stability. Due to the upright geometry between the objectives, standard glass-bottom dishes can be used to both grow and image mammalian cells, thus simplifying experimental procedures.

2.3.2. Characterization of the reflected light-sheet

We used the signal from small fluorescent beads to visualize the spatial profile of the reflected light-sheet (**Figure 2.4**, and see Section 2.5). As expected, the reflection does not alter the shape of the laser beam, although the light-sheet displays a small deviation (2°) from the horizontal plane, which corresponds to a 2% deviation from the 45° cantilever angle. Different AFM cantilevers gave very similar performance (data not shown).

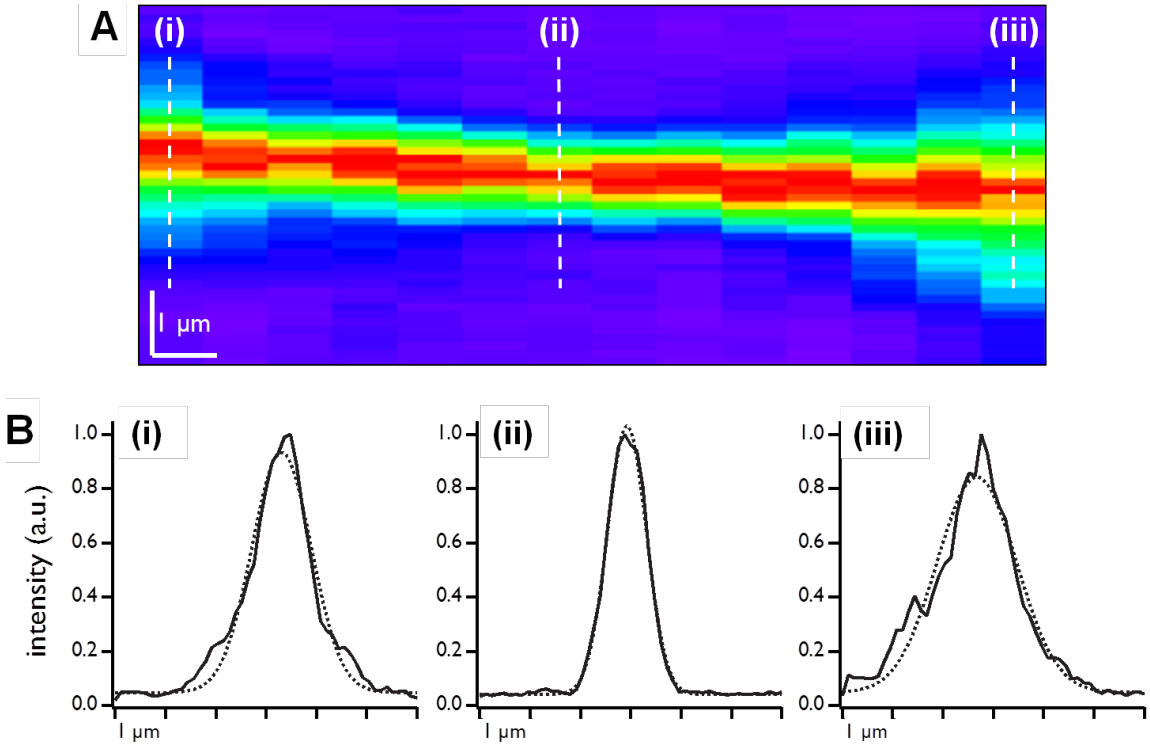

Figure 2.4. Visualization of the reflected light-sheet profile. **(A)** Intensity map of a fluorescent bead scanned vertically through the light-sheet at 4 mm aperture size after reflection by the AFM cantilever. **(B)** Intensity profiles of the bead (solid lines) and Gaussian fits (dashed lines) at positions **(i) – (iii)** along the light-sheet as indicated in **(A)**.

In addition, we quantitatively characterized the dimensions of the light-sheet in the vicinity of the focus. By changing the dimensions of the incident beam with a spherical aperture in front of the illumination objective, the Rayleigh length over which the light-sheet maintains a relatively constant thickness can be controlled. At an aperture size of 4 mm, the FWHM thickness at the focus and the Rayleigh length of the light-sheet was found to be ~ 0.9 µm and ~ 11 µm, respectively (**Figure 2.5 A** and **B)**. Light-sheet with this profile will be used for all subsequent experiments. Due to the shape of the light-sheet, a small gap (~ 1 - 2 µm) between the surface of the sample dish and the light-sheet cannot be illuminated (**Figure 2.5 C**).

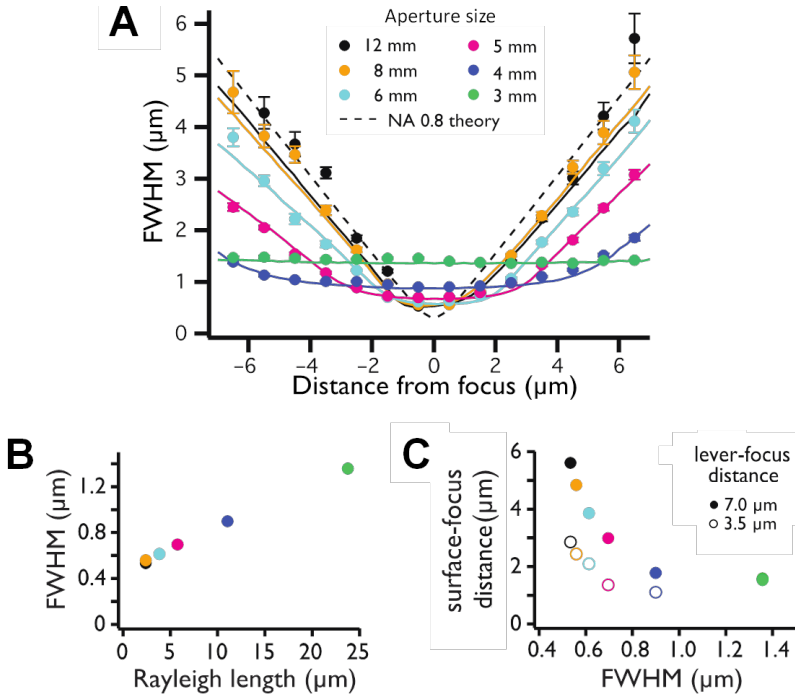

Figure 2.5. Physical characterization of the reflected light-sheet. (A) FWHM thickness of the light-sheet before (solid lines) and after (symbols) reflection as a function of distance from the focus, shown at different aperture sizes. The dashed line indicates profile expected for a 0.8 NA objective. Error bars denote ± s.d. (n = 3). **(B)** FWHM thickness of the light-sheet at the focus as a function of its Rayleigh length. **(C)** Minimal distance between surface of the sample dish and focus of the light-sheet as a function of its FWHM thickness, shown at two different distances between the cantilever edge and the focus of the light-sheet.

2.3.3. Comparing RLS and HILO illuminations

To demonstrate the superior capability of RLS microscopy in single-molecule detection, we compared our technique with HILO illumination using MCF-7 cells expressing histone H4 fused to the photoactivatable fluorescent protein mEos2. For HILO measurements, we chose a small illumination area with a diameter of ~ 10 μm to keep the light-sheet thickness small (~ 5 μm) (Tokunaga et al., 2008). Upon activating a subset of mEos2 molecules with a 405 nm laser in HILO illumination mode, the fluorescence from the molecules was excited with a 560 nm laser and alternated every 50 ms between RLS and HILO modes. Even at low density of mEos2 activation, RLS illumination already resulted in improved SBR as compared to that of HILO microscopy (**Figure 2.6 A**). At high activation densities, the SBR achieved by RLS microscopy

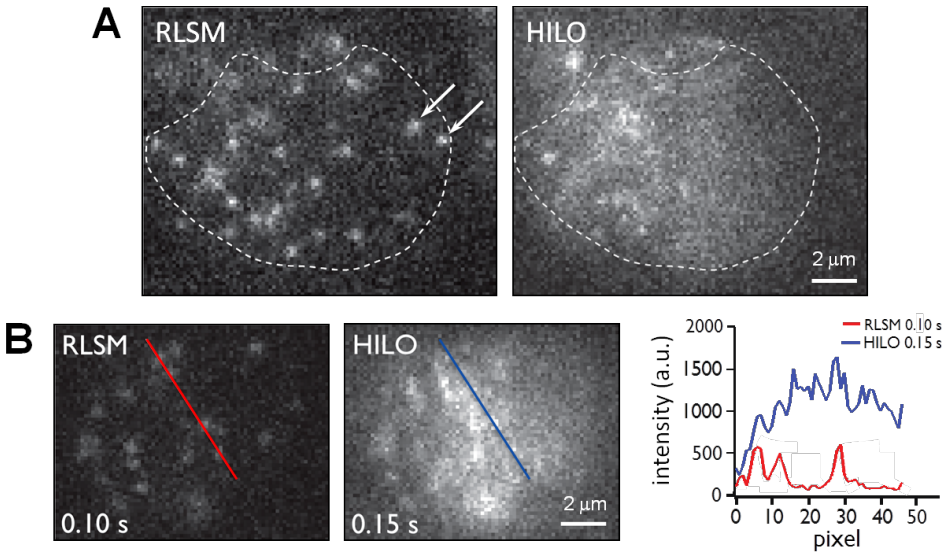

Figure 2.6. Comparison between RLS and HILO illuminations at different densities of fluorescent proteins. (A) RLS (left) and HILO (right) images of a MCF-7 cell expressing mEos2-H4, imaged at low density of mEos2 activation. Arrows in the left panel indicate mEos2 molecules detected with RLS but overlooked by HILO microscopy. Dashed line outlines the nuclear boundary. **(B)** RLS (left) and HILO (middle) images of a MCF-7 cell expressing mEos2-H4, imaged at high density of mEos2 activation. The right panel shows fluorescence intensity across the line for RLS illumination at 0.10 s (red) and for HILO illumination at 0.15 s (blue).

was 5.3 ± 0.4 fold (± s.e.m., n = 267 molecules from 3 cells) higher than that of HILO illumination (**Figure 2.6 B**). We also confirmed the superior SBR and field of view of RLS microscopy throughout the nucleus at different z-positions (**Figure 2.7 A** and **B**).

Moreover, RLS microscopy allows detection of single molecules throughout the cross-section of the nucleus, while for HILO the illuminated area is restricted to a central part of the cross-section. As a result, RLS microscopy can detect fluorescent protein molecules that are missed by HILO microscopy, as shown in **Figure 2.6 A**. At high activation density (**Figure 2.6 B**), the assignment of the molecules that are clearly detected by RLS (as indicated by the three

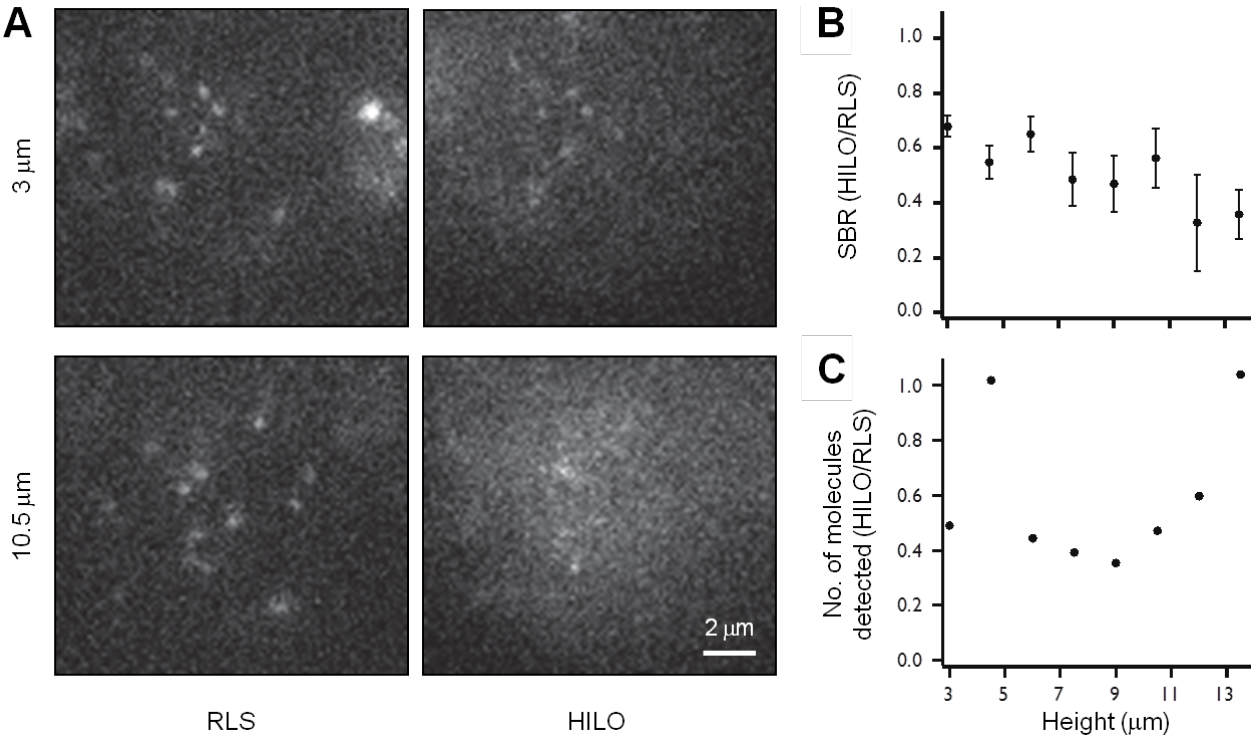

Figure 2.7. Comparison between RLS and HILO illuminations at different z-positions. (A) A MCF-7 cell expressing mEos2-H4 alternately imaged with RLS and HILO illuminations at two different z-positions. The ratios of SBR for single molecules **(B)** and of the number of molecules detected **(C)** between HILO and RLS microscopy are plotted for the cell shown in **(A)**, as a function of z-position. Error bars denote ± s.e.m. (n = 8 cells).

peaks in the red intensity profile) becomes ambiguous in the HILO image, due to the high and inhomogeneous background. Lastly, RLS has shown to be able to consistently detect 2 – 3 times more molecules than HILO at different *z*-positions of the nucleus (**Figure 2.7 C**).

2.3.4. Combining RLS with super-resolution microscopy: RLS-SRM

In order to take full advantage of the 3D optical sectioning capability and single-molecule sensitivity of RLS, we further combined it with the SRM technique STORM/PALM (Betzig et al., 2006; Hess et al., 2006; Rust et al., 2006). The resulting technique is termed RLS-SRM (**Figure 2.8**). We found that gold has better reflectivity for the redder laser lines (such as 647 nm) used in RLS-SRM, and therefore used gold-coated cantilevers as reflector of the light-sheet. To enhance the impact of optical sectioning, cells were fixed immediately after attaching to the coverslip of the imaging dish and before their nuclei had flattened out (see Section 4.5), thereby allowing the

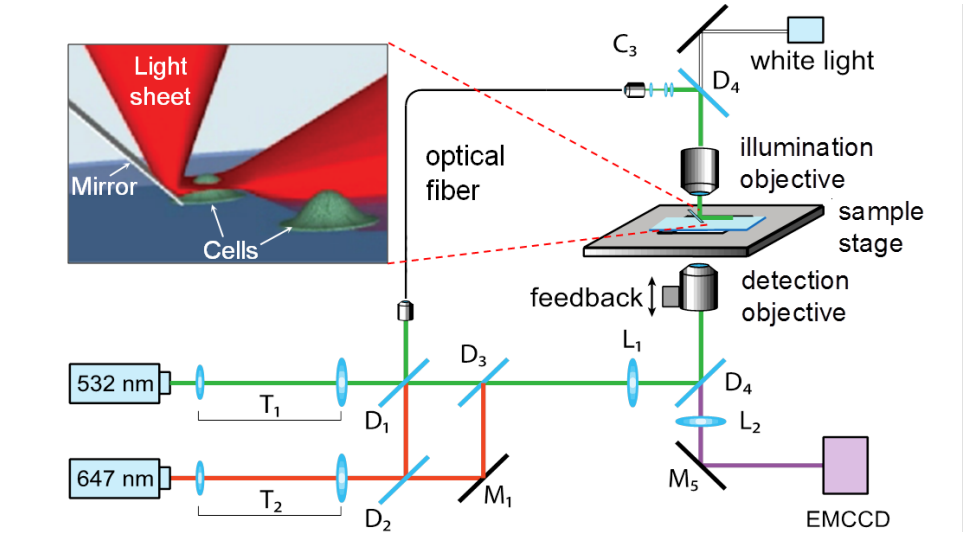

Figure 2.8. Optical scheme of RLS-SRM. Two lasers of different wavelengths are combined to achieve multi-color imaging. Inset shows the placement of the miniature mirror next to the cell to be imaged. C: cylindrical lens; D: dichroic beam splitter; L: lens; M: mirror; T: telescope.

thin light-sheet to selectively illuminate less than 10% of the nuclear volume, leading to further reduced fluorescence background. As a consequence of such enhanced ability in detecting single molecules (**Figure 2.9**), RLS-SRM accomplishes super-resolution imaging inside the mammalian cell nucleus with superior SBR compared to previous studies.

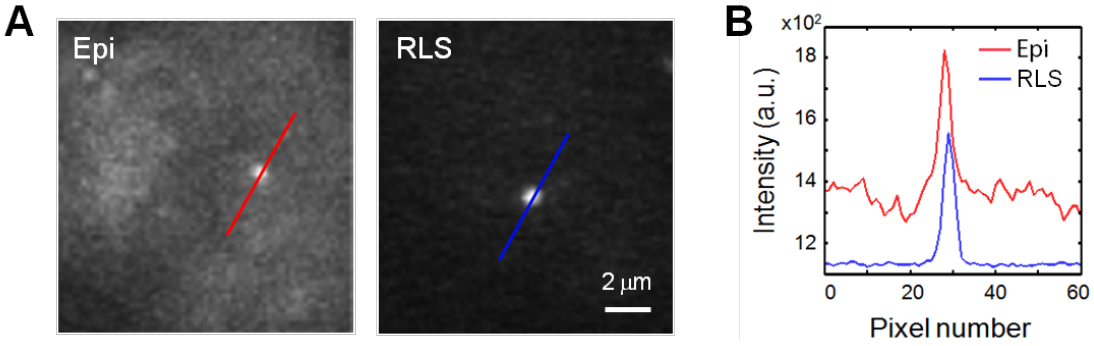

Figure 2.9. Comparison between SRM using RLS and epi-illuminations. (A) Images of a single TMR dye molecule inside a Hela cell nucleus taken under epi (left) and RLS (right) illuminations. **(B)** Comparison of fluorescence intensities across the lines shown in **(A)**.

In order to more quantitatively demonstrate the superior capability of RLS-SRM in performing super-resolution imaging, we imaged two Hela cells in the same Cy5-labeled sample using both epi- and RLS illuminations. The datasets collected, each containing 6,000 frames, were subject to the same procedure for image analysis. RLS illumination detects ~ 20-fold more localizations than epi-illumination, convincingly illustrating its superior sensitivity (**Table 2.1**). In addition, in the process of grouping localizations into spatial clusters (see Sections 4.3.2 and 4.5 for details) in order to pinpoint individual dye molecules in the nucleus, RLS-SRM not only identifies ~ 8-fold more clusters, but each cluster identified also contains ~ 3-fold more localizations than those identified by epi-SRM within a similar area, thus leading to higher localization precision.

Table 2.1. Demonstration of enhanced capability of RLS-SRM in performing super-resolution imaging.

		Epi only	RLS only	Epi + RLS [1]
No. of localizations detected		1537	33658	36470
Mean intensity of localizations (a.u.)		21000	8400	9800
No. of spatial clusters identified		137	1055	1224
Mean no. of localizations per spatial cluster		11	32	30
Dimensions of spatial clusters	σ_x **(nm)**	22.2	22.1	22.3
	σ_y **(nm)**	21.8	23.1	22.9

The dimensions of the spatial clusters identified are measured by the standard deviations in the lateral (x and y) spread of the localization positions. [1] denotes analysis in which the raw data from both epi and RLS imaging are pooled together as one dataset.

2.4. Discussion and conclusion

As a consequence of the decoupling between illumination and detection optics, the advantages of the light-sheet imaging configuration in our design are multifarious. Firstly, the intrinsic optical sectioning capability minimizes out-of-focus background, thereby enhancing contrast and reducing phototoxicity to the sample, which is especially advantageous for live-cell imaging. Concomitantly, the drastically reduced photobleaching rate also prolongs the period over which the sample could be imaged, as previously demonstrated (Reynaud et al., 2008). Lastly, since the image is acquired in a parallel, layer-by-layer fashion, ultrafast acquisition speed could be achieved without compromising the field of view. Overall, such configuration combines the resolution and contrast of a laser-scanning microscope with the speed and sensitivity of a video microscope, while avoiding their respective shortcomings.

The vertical orientation of the illumination and detection objectives unique to our RLS setup further introduces several advantages compared to the orthogonal geometry normally

employed in SPIM instruments. Firstly, any commercial inverted microscope may be converted to a RLS setup by adjusting the laser illumination beam path, replacing the condenser with a water-dipping objective and the connected mirror, and changing the sample stage to a piezo stage. Secondly, since both objectives can be chosen with high numerical aperture, very thin excitation light-sheet (down to ~ 0.5 μm) as well as high efficiency of fluorescence signal collection can be achieved, resulting in high SBR superior to that of epi- and HILO illuminations. Thirdly, the reflecting mirror allows positioning of the horizontal light-sheet close to the cover-glass surface, leaving only a small gap of ~ 2 μm that cannot be illuminated. This gap is small enough to enable sectioning of most part of the mammalian cell nucleus. Finally, there is no need for special observation chamber in RLS microscopy, as commercially available glass-bottom culture dishes can be used for both cell culture and imaging, further simplifying experimental procedures (Capoulade et al., 2011; Wu et al., 2011). Such configuration also eliminates any potential aberration to the light-sheet profile as a result of introducing a glass surface (through the side of the sample chamber) into the illumination light path (see, for example, **Figure 2.1 B(i)**), although the fact that the sample is open to air does impose limits on the duration of live-cell observation, in cases where long-term imaging is required.

While the spatial resolution of RLS alone is diffraction-limited, when coupled to super-resolution microscopy, RLS-SRM can achieve resolution improvements that are particularly important for imaging dense nuclear structures. Given that the light-sheet used has a FWHM thickness of ~ 0.9 μm and a Rayleigh length of ~ 11 μm, and that an unflattened mammalian cell nucleus has a diameter of 10 - 15 μm, the dimension of the light-sheet is perfectly suited to image almost the entire cross-section of the nucleus while providing excellent optical sectioning

by illuminating < 10% of the nuclear volume. As such, single-molecule localization in these dense structures, which is difficult with epi-illumination due to the substantial background, can be achieved with much higher precision. Such improvements in resolution as well as precision will prove critical for resolving, with single-copy accuracy, mammalian nuclear structures that are present at very high molecular density, as will be shown in Chapters 4 and 5. Overall, with their capability and versatility, both RLS and RLS-SRM could be generally applied to probe key biomolecular processes deep inside mammalian cells, either at super-resolution level or in real time.

2.5. Materials and methods

Optical setup of RLS microscope

The reflected light-sheet microscope is integrated into an inverted microscope (IX71, Olympus). Illumination lasers (405 nm, 50 mW, Electra-40, Laserglow; 488/514 nm, 1000 mW, Innova300, Coherent; 560 nm, 1000 mW, VFL-P-1000-560, MPB Communications; actual intensity was set to 3 mW in the sample plane) are collimated and co-linearly combined via dichroic beam splitters. Shutters (LS3M2, Uniblitz) are used to control the active laser times. A telescope of two cylindrical lenses (f = 40 mm, LJ1402L1-A and f = 400 mm, LJ1363L1-A, both Thorlabs) creates an expanded and collimated line that overfills the back aperture of the vertical illumination objective (LUMPLFLN 40x, water, NA 0.8, Olympus) and is focused to a diffraction-limited light-sheet. A third cylindrical lens (f = 150 mm, LJ1629L1-A, Thorlabs) is used to control the lateral extension of the light-sheet. A spherical iris in front of the illumination objective allows reduction of the illumination line dimensions and thus the FWHM of the light-sheet. A tip-less AFM cantilever (HYDRA2R-100N-TL-10, Nanoscience) is mounted to the illumination objective

via a custom-designed holder. The disposable AFM cantilever is custom-coated with a 1 nm Ti layer followed by 40 nm Al layer by thermal evaporation. A manual *xy*-stage (ST1XY-S, Thorlabs) and a *z*-stage (423, Newport) allow simultaneous positioning of illumination objective and cantilever holder with respect to the detection objective (either UPlanApo 100x, NA 1.35, oil; or UPlanSApo 100x, NA 1.4, oil, both Olympus). Fluorescence (filter sets for mEos2: dichroic Di01-R561, filters Brightline 617/73 and Edgebasic long wave pass 561, Semrock; YPet: dichroic FF495-Di03, Semrock and filter HQ545/30, Chroma; EGFP/TagRFP-T: dichroic Di01-R488/561 and filter FF01-523/610, Semrock) was focused onto a back-illuminated EMCCD camera (iXon+, DU-897E-CSO-BV, Andor). Bright field illumination was achieved using the microscope light source, coupled into the illumination objective using a dichroic (FF593-Di03, Semrock). The sample dish (Delta-T, Bioptechs) is mounted onto a custom designed manual *xy*-stage, in which a thermal control unit (Delta-T, Bioptechs) is integrated to control the sample temperature. An objective heater (FAB6318x, 1000 Oaks Optical) aids in thermal control. A *xyz*-piezo stage (Nano-Bio3200, Mad City Labs), controlled by custom-written LabVIEW software, is used for precision positioning and vertical scanning of the sample. The microscope, shutters, and EMCCD camera are controlled by MetaMorph software.

Design of the cantilever holder

The cantilever holder consists of four stainless steel cylinders. The first unit is stably mounted onto the objective, while the other units are successively connected via two miniature linear guides per plane (MR3MNSS1V0N15L-2.5-2.5, Precision Alliance), enabling movement of the fourth cylinder in three dimensions with respect to the objective. A micrometer drive (DM11-5, Newport) counteracted by a small spring is used for precise position control in each plane. The fourth cylinder holds a metallic bridge with a small groove that accommodates the AFM

cantilever. High vacuum grease (Dow Corning) is used to reversibly affix the cantilever and ensures stable mounting after ~ 10 min of settling time.

Determination of the light-sheet profile

Before reflection, the laser beam profile was characterized by imaging the intensity cross-section at various distances from the focus onto the EMCCD camera, and determining the FWHM from the Gaussian fits to each intensity distribution. To characterize the beam profile after reflection, fluorescent beads (TetraSpeck microspheres, 100 nm diameter, Invitrogen) attached to fixed Hela cells (for elevation above the surface of the sample dish) were scanned across the beam at various distances from the focus using the *xyz*-piezo stage. Fluorescence emitted by the beads was projected onto the EMCCD camera, and beads were kept in focus by moving the detection objective with a *z*-piezo stage (PIFOC, Physik Instrumente). The FWHM thickness of the laser beam was determined from Gaussian fits to the resulting intensity distributions.

Data acquisition

Data acquisition with RLS is straightforward and comparable to other single-molecule assays. Exchange of the AFM cantilever and positioning with respect to the light-sheet focus can be achieved within 20 min. Once in place, a cantilever typically can be used for up to 5 days of imaging, with only minor position adjustments for each sample. Positioning of the cantilever next to a cell is precisely controlled by the *xyz* piezo-stage and achieved within 1 min without perforating the cell membrane. Cells were imaged in OptiMEM medium. During imaging, the power of all illumination lasers was kept below 3 mW, which corresponds to ~ 5 kW/cm^2 in the focal plane of the illumination objective, comparable to the laser power density used in other live-cell single-molecule microscopy experiments (Jones et al., 2011).

Comparison of RLS and HILO illuminations

The position, peak intensity and background level of single molecules were determined by two-dimensional Gaussian fitting of the original uncorrected image. The SBR was defined as signal divided by the background level above the camera dark offset (Moerner and Fromm, 2003). For both RLS and HILO illuminations, the SBR for all molecules within a *z*-section of a cell nucleus were combined into a histogram, whose median value was used to calculate the ratio between RLS and HILO for this *z*-section.

Combining RLS with SRM

The RLS-SRM microscope was custom-built using the body frame of a Nikon TE300 inverted microscope. The microscope sample stage was replaced with a custom-made invar piece to increase the overall stability of the setup and reduce mechanical vibrations. The condenser was replaced with a water immersion objective (LUMPLFLN 40X, NA 0.8, Olympus) mounted vertically on a home-built holder that acts as a micro-positioning translational stage. Laser beams (647 nm line from Innova-300 or 532 nm line from Verdi-5W, both Coherent) were expanded, collimated, combined and delivered through an optical fiber onto a lens assembly that generates an elliptical beam. This beam was positioned at the back aperture of the vertical illumination objective with one axis overfilling the back aperture and the orthogonal axis being focused. The resulting light-sheet created by the objective was reflected by 90° by a gold-coated tip-less AFM cantilever (Applied NanoStructures). Separate laser beams were also sent through the microscope back-port to allow alternative epi-illumination. The sample dish was placed on a 3D piezo-driven stage (Physik Instrumente) coupled to a manual *xy* stage, allowing positional control in all three axes with nanometer precision. Vertical drift of the sample stage was corrected for by controlling the *z*-position of the detection objective with a piezo-driven mount

subjected to a feedback signal coming from the reflection of an infrared laser diode off the coverslip of the sample dish (Autofocus Trac, MotionX Corp.). The fluorescence emission was collected by an oil objective (UPlanSApo 100X, NA 1.4, Olympus) placed co-planar to the light-sheet and imaged onto an EMCCD camera (iXon DU-897, Andor). Astigmatism introduced into the imaging path using a cylindrical lens was used to recover super-resolved *z*-positions (Huang et al., 2008). The entire setup was encased in a home-built aluminum chamber and room temperature was controlled to within ± 0.1 °C to shield it from environmental influences.

References

Ahrens, M.B., Orger, M.B., Robson, D.N., Li, J.M., and Keller, P.J. (2013). Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nat. Methods* **10**, 413-420.

Arnaout, R., Ferrer, T., Huisken, J., Spitzer, K., Stainier, D.Y.R., Tristani-Firouzi, M., and Chi, N.C. (2007). Zebrafish model for human long QT syndrome. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11316-11321.

Betzig, E., Patterson, G.H., Sougrat, R., Lindwasser, O.W., Olenych, S., Bonifacino, J.S., Davidson, M.W., Lippincott-Schwartz, J., and Hess, H.F. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642-1645.

Capoulade, J., Wachsmuth, M., Hufnagel, L., and Knop, M. (2011). Quantitative fluorescence imaging of protein diffusion and interaction in living cells. *Nat. Biotechnol.* **29**, 835-839.

Cella Zanacchi, F., Lavagnino, Z., Perrone Donnorso, M., Del Bue, A., Furia, L., Faretta, M., and Diaspro, A. (2011). Live-cell 3D super-resolution imaging in thick biological samples. *Nat. Methods* **8**, 1047-1049.

Dodt, H.-U., Leischner, U., Schierloh, A., Jahrling, N., Mauch, C.P., Deininger, K., Deussing, J.M., Eder, M., Zieglgansberger, W., and Becker, K. (2007). Ultramicroscopy: Three-dimensional visualization of neuronal networks in the whole mouse brain. *Nat. Methods* **4**, 331-336.

Engelbrecht, C.J., Greger, K., Reynaud, E.G., Kržic, U., Colombelli, J., and Stelzer, E.H. (2007). Three-dimensional laser microsurgery in light-sheet based microscopy (SPIM). *Opt. Express* **15**, 6420-6430.

Friedrich, M., Nozadze, R., Gan, Q., Zelman-Femiak, M., Ermolayev, V., Wagner, T.U., and Harms, G.S. (2009). Detection of single quantum dots in model organisms with sheet illumination microscopy. *Biochem. Biophys. Res. Comm.* **390**, 722-727.

Fuchs, E., Jaffe, J., Long, R., and Azam, F. (2002). Thin laser light sheet microscope for microbial oceanography. *Opt. Express* **10**, 145-154.

Gao, L., Shao, L., Higgins, C.D., Poulton, J.S., Peifer, M., Davidson, M.W., Wu, X., Goldstein, B., and Betzig, E. (2012). Noninvasive imaging beyond the diffraction limit of 3D dynamics in thickly fluorescent specimens. *Cell* **151**, 1370-1385.

Greger, K., Swoger, J., and Stelzer, E.H.K. (2007). Basic building units and properties of a fluorescence single plane illumination microscope. *Rev. Sci. Instrum.* **78**, 023909.

Hess, S.T., Girirajan, T.P., and Mason, M.D. (2006). Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.* **91**, 4258-4272.

Holekamp, T.F., Turaga, D., and Holy, T.E. (2008). Fast three-dimensional fluorescence imaging of activity in neural populations by objective-coupled planar illumination microscopy. *Neuron* **57**, 661-672.

Huang, B., Wang, W., Bates, M., and Zhuang, X. (2008). Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* **319**, 810-813.

Huber, D., Keller, M., and Robert, D. (2001). 3D light scanning macrography. *J. Microsc.* **203**, 208-213.

Huisken, J., and Stainier, D.Y.R. (2007). Even fluorescence excitation by multidirectional selective plane illumination microscopy (mSPIM). *Opt. Lett.* **32**, 2608-2610.

Huisken, J., and Stainier, D.Y.R. (2009). Selective plane illumination microscopy techniques in developmental biology. *Development* **136**, 1963-1975.

Huisken, J., Swoger, J., Del Bene, F., Wittbrodt, J., and Stelzer, E.H. (2004). Optical sectioning deep inside live embryos by selective plane illumination microscopy. *Science* **305**, 1007-1009.

Jones, S.A., Shim, S.H., He, J., and Zhuang, X. (2011). Fast, three-dimensional super-resolution imaging of live cells. *Nat. Methods* **8**, 499-508.

Keller, P.J., Schmidt, A.D., Santella, A., Khairy, K., Bao, Z., Wittbrodt, J., and Stelzer, E.H.K. (2010). Fast, high-contrast imaging of animal development with scanned light sheet-based structured-illumination microscopy. *Nat. Methods* **7**, 637-642.

Keller, P.J., Schmidt, A.D., Wittbrodt, J., and Stelzer, E.H.K. (2008). Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. *Science* **322**, 1065-1069.

Krzic, U., Gunther, S., Saunders, T.E., Streichan, S.J., and Hufnagel, L. (2012). Multiview light-sheet microscope for rapid *in toto* imaging. *Nat. Methods* **9**, 730-733.

Mertz, J. (2011). Optical sectioning microscopy with planar or structured illumination. *Nat. Methods* **8**, 811-819.

Mickoleit, M., Schmid, B., Weber, M., Fahrbach, F.O., Hombach, S., Reischauer, S., and Huisken, J. (2014). High-resolution reconstruction of the beating zebrafish heart. *Nat. Methods* **11**, 919-922.

Moerner, W.E., and Fromm, D.P. (2003). Methods of single-molecule fluorescence spectroscopy and microscopy. *Rev. Sci. Instrum.* **74**, 3597.

Pampaloni, F., Reynaud, E.G., and Stelzer, E.H.K. (2007). The third dimension bridges the gap between cell culture and live tissue. *Nat. Rev. Mol. Cell Biol.* **8**, 839-845.

Planchon, T.A., Gao, L., Milkie, D.E., Davidson, M.W., Galbraith, J.A., Galbraith, C.G., and Betzig, E. (2011). Rapid three-dimensional isotropic imaging of living cells using Bessel beam plane illumination. *Nat. Methods* **8**, 417-423.

Reynaud, E.G., Krzic, U., Greger, K., and Stelzer, E.H.K. (2008). Light sheet-based fluorescence microscopy: more dimensions, more photons, and less photodamage. *HFSP J.* **2**, 266-275.

Ritter, J.G., Veith, R., Siebrasse, J.-P., and Kubitscheck, U. (2008). High-contrast single-particle tracking by selective focal plane illumination microscopy. *Opt. Express* **16**, 7142-7152.

Ritter, J.G., Veith, R., Veenendaal, A., Siebrasse, J.P., and Kubitscheck, U. (2010). Light sheet microscopy for single molecule tracking in living tissue. *PLoS One* **5**, e11639.

Rust, M.J., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793-795.

Scherz, P.J., Huisken, J., Sahai-Hernandez, P., and Stainier, D.Y.R. (2008). High-speed imaging of developing heart valves reveals interplay of morphogenesis and function. *Development* **135**, 1179-1187.

Schmid, B., Shah, G., Scherf, N., Weber, M., Thierbach, K., Campos, C.P.r., Roeder, I., Aanstad, P., and Huisken, J. (2013). High-speed panoramic light-sheet microscopy reveals global endodermal cell dynamics. *Nat. Commun.* **4**, 2207.

Siedentopf, H., and Zsigmondy, R. (1903). Ueber Sichtbarmachung ultramikroskopischer Teilchen, mit besonderer Anwendung auf Goldrubingläser. *Annal. der Physik* **10**, 1-39.

Stelzer, E.H.K. (2015). Light-sheet fluorescence microscopy for quantitative biology. *Nat. Methods* **12**, 23-26.

Tokunaga, M., Imamoto, N., and Sakata-Sogawa, K. (2008). Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat. Methods* **5**, 159-161.

Tomer, R., Khairy, K., Amat, F., and Keller, P.J. (2012). Quantitative high-speed imaging of entire developing embryos with simultaneous multiview light-sheet microscopy. *Nat. Methods* **9**, 755-763.

Truong, T.V., Supatto, W., Koos, D.S., Choi, J.M., and Fraser, S.E. (2011). Deep and fast live imaging with two-photon scanned light-sheet microscopy. *Nat. Methods* **8**, 757-760.

Turaga, D., and Holy, T.E. (2008). Miniaturization and defocus correction for objective-coupled planar illumination microscopy. *Opt. Lett.* **33**, 2302-2304.

Verveer, P.J., Swoger, J., Pampaloni, F., Greger, K., Marcello, M., and Stelzer, E.H.K. (2007). High-resolution three-dimensional imaging of large specimens with light sheet-based microscopy. *Nat. Methods* **4**, 311-313.

Vettenburg, T., Dalgarno, H.I.C., Nylk, J., Coll-Llado, C., Ferrier, D.E.K., Cizmar, T., Gunn-Moore, F.J., and Dholakia, K. (2014). Light-sheet microscopy using an Airy beam. *Nat. Methods* **11**, 541-544.

Voie, A.H., Burns, D.H., and Spelman, F.A. (1993). Orthogonal-plane fluorescence optical sectioning: Three-dimensional imaging of macroscopic biological specimens. *J. Microsc.* **170**, 229-236.

Wu, Y., Ghitani, A., Christensen, R., Santella, A., Du, Z., Rondeau, G., Bao, Z., Colon-Ramos, D., and Shroff, H. (2011). Inverted selective plane illumination microscopy (iSPIM) enables coupled cell identity lineaging and neurodevelopmental imaging in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 17708-17713.

Chapter 3

Real-time monitoring of mammalian transcription factor dynamics *in singulo* using RLS microscopy [1]

3.1. Abstract

Transcription factor (TF) dynamics is universally known to play a crucial role in regulating gene expression. A thorough elucidation of this highly complex and heterogeneous process requires us to follow the dynamic mobility of individual TF molecules in real time, which still remains challenging in mammalian cell nucleus. We demonstrated the capability of RLS microscopy in directly monitoring the DNA binding dynamics of two key mammalian TFs, glucocorticoid receptor and estrogen receptor-α. By measuring their diffusion constants and DNA-bound fractions under both induced and un-induced conditions, we determined the residence times of the various oligomerization states and mutants of these TFs, and resolved three distinct modes of their interaction with genomic DNA. Using two-color RLS imaging, we also visualized, in real time, colocalization of single molecules of GR and its co-activator GRIP1 as well as of the heterodimeric TF pair BMAL1 and CLOCK, thus allowing us to directly interrogate the spatio-

[1] The work described in this chapter was performed in collaboration with Drs. J. Christof M. Gebhardt, David M. Suter, and Rahul Roy.

temporal interaction of two molecular species *in vivo*. Our measurements and statistical analysis enriched our quantitative understanding of the dynamic properties of mammalian TFs, and could be generally applicable to single-molecule studies of other biomolecular systems in live mammalian cells.

3.2. Introduction: Transcription factor binding as a key aspect of nuclear dynamics

The binding of TFs to specific DNA sequences represents the first layer of eukaryotic gene regulation (Watson et al., 2007). Many transcription factors have been found to exhibit highly dynamic and heterogeneous interactions with genomic DNA. Such interactions have been shown to be crucial for recruiting the transcription machinery to the promoter, and could in turn significantly impact transcriptional outcome (Hager et al., 2009). Previous studies using techniques such as chromatin immunoprecipitation (ChIP) suggest that the various components of the transcription machinery assemble in a sequential manner, during which the TFs may reside on DNA for minutes. This view was challenged by studies suggesting that TF binding could be much more transient, with DNA residence times on the order of seconds or less (Larson, 2011). This led to a new model in which TF binding and unbinding modifies the promoter chromatin while recruiting the transcription machinery, eventually reaching a state competent for initiating transcription.

In an attempt to probe this critical step in gene regulation, a large number of live-cell imaging studies have been undertaken in a variety of eukaryotic systems, ranging from yeast (Larson et al., 2011), to *Drosophila* salivary glands (Yao et al., 2006), and to developing mouse embryos (Kaur et al., 2013). However, in order to fully dissect this highly complex process and to quantify key parameters of the binding interactions between TFs and genomic DNA, direct

visualization of individual TF molecules in action is essential. Unfortunately, given the difficulty in performing single-molecule tracking (SMT) in the mammalian cell nucleus as delineated in Chapter 1, this was not achieved in most of the abovementioned studies.

Besides SMT, two other techniques commonly used to study TF dynamics in live cells are fluorescence recovery after photobleaching (FRAP) and fluorescence correlation spectroscopy (FCS). FRAP monitors the kinetic replenishing of fluorescence in a bleached zone of the sample through the diffusion and rebinding of unbleached molecules from outside of the zone (Mueller et al., 2010), while FCS extracts physical parameters of the diffusing species from the autocorrelation function of fluorescence intensity fluctuations in a small observation volume (Macháň and Wohland, 2014). However, the indirect analysis of TF dynamics via these techniques is error-prone, as experimental conditions including the geometry of the bleached volume, the fraction of freely diffusing molecules, and photophysical properties of the fluorophore must be accurately determined and carefully accounted for (Beaudouin et al., 2006; Mazza et al., 2012; Mueller et al., 2010; Mueller et al., 2012; Sprague et al., 2006; Stavreva et al., 2012). Owing to such complications, the values reported by several previous FRAP and FCS studies on the same TF exhibit significant discrepancy (Sprague et al., 2006; Sprague et al., 2004; Stasevich et al., 2010).

To overcome these limitations, we took advantage of the superior signal-to-background ratio of RLS microscopy and applied it to directly monitor the *in vivo* dynamics of mammalian glucocorticoid receptor (GR) and estrogen receptor-α (ER). GR is a transcription factor that localizes mostly to the cytoplasm in the absence of hormone, but forms homodimers and translocates into the nucleus upon binding to glucocorticoids (Tsai et al., 1988). ER operates through a similar mechanism, although it is constitutively localized to the nucleus (Zava et al.,

1977). Previous studies have shown that dimeric GR and ER bind directly to DNA at regulatory sequences, while the monomer can be indirectly recruited to DNA by other DNA-bound protein complexes (Aagaard et al., 2011). The mode of their interaction with DNA defines whether the target gene is activated or repressed. By performing SMT using RLS microscopy, we found that the residence times of monomeric GR and indirectly bound GR are only 10% and 50%, respectively, of the residence time of the dimeric form. A similar result was obtained for ER. We also demonstrated the capability to perform two-color single-molecule imaging, and used it to directly observe the spatio-temporal colocalization of two pairs of mammalian TFs *in vivo*. The insights revealed by these studies could apply to other mammalian TFs in general.

3.3. Results

3.3.1. Following and quantifying single TF diffusion in real time using RLS microscopy

To optimize single-molecule detection of TF in live mammalian cell nucleus, we first tested the suitability of various fluorescent fusion partners. In principle, the protein fusion tags SNAP and Halo, which can be covalently labeled with organic dyes *in vivo*, offer a very attractive labeling strategy due to the brightness and photostability of organic dyes (Keppler et al., 2003; Los et al., 2008; Mazza et al., 2012). However, we found that both SNAP and Halo proteins exhibit a certain degree of intrinsic binding propensity in the nucleus, which could bias the kinetic analysis of DNA interactions of their fusion partners. We therefore chose the bright fluorescent proteins mEos2 and YPet as labels for transcription factors, as neither of them shows detectable nuclear binding. In addition, we chose the fluorescent proteins EGFP and TagRFP-T as candidates for two-color imaging due to their spectral separation.

With this choice, we expressed a mEos2-GR fusion protein in MCF-7 cells, and performed single-molecule imaging to probe its diffusion in the nucleus both with and without treatment of the hormone analog dexamethasone. We photoactivated only a small subset of mEos2 molecules in the focal plane to limit the number of simultaneously observable molecules, thereby avoiding the overlap of their trajectories (Manley et al., 2008). Under such condition, single mEos2-GR molecules in the cell nucleus could be detected with 10 ms time resolution, and their diffusion trajectories were analyzed (**Figure 3.1 A**). While a small fraction of the trajectories last for tens of frames (**Figure 3.1 A**, bottom), the majority of them tend to be relatively short (less than 10 frames) due to fast photobleaching as well as diffusion of the molecule out of the focal plane (**Figure 3.1 A**, top).

Each time a molecule was photoactivated in the field of view, we computed the cumulative distribution function of its squared displacement during a fixed time interval instead of the mean squared displacement, in order to accurately extract diffusion properties from short trajectories (Schutz et al., 1997). The probability density of squared displacement, $p\left(x^2+y^2\right)$, for a molecule exhibiting purely Brownian diffusion is given by

$$p\left(x^2+y^2\right)=\frac{1}{4\pi D\tau}\exp\left(-\frac{x^2+y^2}{4D\tau}\right), \qquad \textbf{(3.1)}$$

where D denotes the diffusion constant and τ the camera integration time. Integrating this probability density yields the cumulative distribution function, $P\left(x^2+y^2\right)$:

$$P\left(x^2+y^2\right)=\int\frac{1}{4\pi D\tau}\exp\left(-\frac{x^2+y^2}{4D\tau}\right)d\left(x^2+y^2\right)=1-\exp\left(-\frac{x^2+y^2}{4D\tau}\right), \qquad \textbf{(3.2)}$$

or

$$P(X) = \sum_i A_i \left[1 - \exp\left(-\frac{X}{D_i}\right)\right] \quad \textbf{(3.3)}$$

in the case of multiple distinct diffusion components, with $X \equiv \left(x^2 + y^2\right)/4\tau$ and A_i and D_i denoting the fraction and diffusion constant of the i^{th} component, respectively.

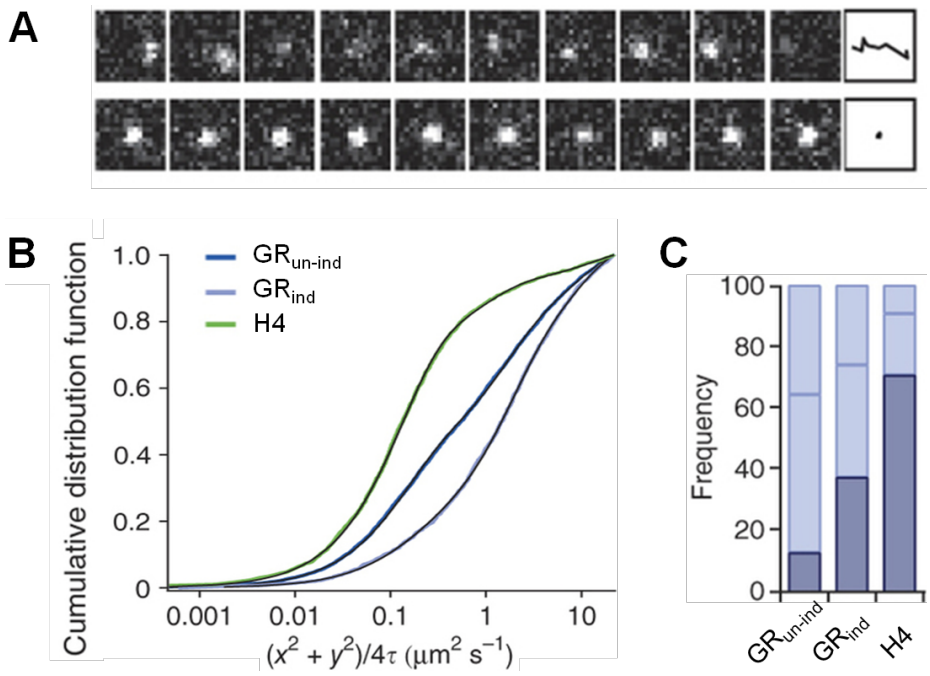

Figure 3.1. Detection and quantification of GR diffusion at single-molecule level. (A) Two representative trajectories showing a fast-diffusing (top) and a DNA-bound (bottom) mEos2-GR molecule in the presence of 100 nM dexamethasone at 10 ms time resolution. The identified traces are shown on the far right. **(B)** Cumulative distribution functions of squared displacements of un-induced GR ($GR_{un\text{-}ind}$), induced GR (GR_{ind}), and histone H4 (H4) as control (for $GR_{un\text{-}ind}$: n = 1644 from 4 cells; for GR_{ind}: n = 3336 from 7 cells; for H4: n = 2020 from 8 cells). Black lines indicate fits using three effective diffusion components (**Equation 3.3**). **(C)** Fractions of molecules exhibiting slow diffusion (*i.e.* DNA-bound, dark blue) as well as the two fast diffusion modes (light blue) for the three cases shown in **(B)**.

Using such analysis, we found that the cumulative distribution functions of GR, both un-induced and induced, deviate significantly from the exponential form expected for purely

Brownian diffusion (**Equation 3.2**), suggesting that a GR molecule undergoes transitions between different diffusion modes (**Figure 3.1 B**). Instead, these distributions can be well fit by **Equation 3.3** with $i = 3$, indicating the presence of three effective components with distinct diffusion constants, D_1, D_2, D_3, and respective fractions A_1, A_2, A_3, as shown in **Table 3.1**. We observed a significantly higher fraction of the slow-diffusing component for induced GR than for un-induced GR (**Figure 3.1 C**).

Table 3.1. Three distinct diffusion components of both un-induced and induced GR.

		Diffusion component		
		1	**2**	**3**
Un-induced	***D*** (μm² s⁻¹)	0.13 ± 0.03	1.6 ± 0.3	8.9 ± 3.0
	A (%)	12 ± 2	52 ± 5	36 ± 6
Induced	***D*** (μm² s⁻¹)	0.13 ± 0.01	1.4 ± 0.2	9.2 ± 2.3
	A (%)	37 ± 2	37 ± 3	26 ± 4

All values are reported as mean ± s.d..

To assign the slow component, we repeated the measurement for a fusion protein of mEos2 to histone H4, which serves as a control since it is stably incorporated into chromatin. We again found three distinct diffusion components, with the slowest one having a diffusion constant similar to that of GR (D_1 = 0.13 ± 0.01 μm² s^{-1}) but constituting the highest fraction (71% ± 4%) among the three components (**Figure 3.1 B** and **C**). This value of D_1 is much larger than the diffusion constants associated with the movement of chromatin in mammalian cells, which have been reported to range from 10^{-4} to 10^{-3} μm² s^{-1} (Akhtar and Gasser, 2007). Furthermore, we calculated a localization error of Δx = 49 nm at the photon count of 27.5 within 10 ms for histone H4 (Thompson et al., 2002). Such displacement corresponds to an apparent diffusion constant of

0.06 $\mu m^2\ s^{-1}$, which is close to the value of D_1. Therefore, we conclude that the apparent slow component arises from the uncertainty in localizing DNA-bound molecules at low signal levels. The larger diffusion constants observed presumably arise from transient non-specific interactions with DNA and spatially restricted diffusion in the nucleus (Bancaud et al., 2009).

With such understanding, we can then use the fraction of the slowest diffusion component as an estimate for the fraction of the transcription factor that is bound to DNA: 12% of residual nuclear GR is bound to chromatin in the absence of hormone treatment, as compared to 37% after dexamethasone induction. At the single-molecule level, these percentages also correspond to the fraction of time a GR molecule is bound to DNA.

3.3.2. Characterization of DNA binding dynamics of GR and ER *in vivo*

Next, we measured the *in vivo* residence time of individual GR dimers bound to DNA in the presence of 100 nM dexamethasone, using the principle of "detection by localization" (Elf et al., 2007). Since mEos2 exhibits prolonged fluorescent dark states that might interfere with residence time measurements, we chose instead the bright yellow fluorescent protein YPet as a label for GR, using a plasmid that allows low expression levels in MCF-7 cells (see Section 3.5 for details). We considered a molecule to be bound to DNA only if it stayed immobile for at least two consecutive frames (Elf et al., 2007).

When monitoring a bound TF molecule, the fast photobleaching of the fluorescent protein label makes it impossible to determine the residence time based on continuous single-molecule tracking, since the fluorescence signal ("on" state) could terminate as a consequence of two parallel Poisson processes, namely photobleaching (with a rate constant k_b) and dissociation (with a rate constant k_{off}). In order to extract these two parameters, we performed time-lapse

imaging with a fixed integration time, τ_{int}, of 50 ms interspersed with dark periods, τ_d, of varying duration, thereby giving a total time-lapse time of $\tau_{tl} = \tau_{int} + \tau_d$ (**Figure 3.2 A**). The apparent photobleaching rate, being proportional to light intensity and thus dependent on the integration time and time-lapse time, is then given by $k_b\tau_{int}/\tau_{tl}$. Therefore, the fluorescence "on"-time of each bound GR molecule follows an exponential distribution with an effective off-rate constant of $k_{eff} = k_b\tau_{int}/\tau_{tl} + k_{off}$:

$$p_1(t) = A\exp\left(-k_{eff}t\right) = A\exp\left[-\left(k_b\frac{\tau_{\mathrm{int}}}{\tau_{tl}} + k_{off}\right)t\right]. \quad \textbf{(3.4)}$$

Fitting the distributions of fluorescence "on"-times to **Equation 3.4** yields k_{eff}. When $k_{eff}\tau_{tl}$ is plotted as function of τ_{tl}, the slope of the graph gives k_{off} while the y-intercept gives $k_b\tau_{int}$ (**Figure 3.2 B**), and the residence time of the TF is then given by k_{off}^{-1}. Applying such analysis to bound YPet-GR dimers, we obtained the dissociation and photobleaching rate constants as k_{off} = 0.69 ± 0.11 s^{-1} and k_b = 26.8 ± 0.5 s^{-1}, as well as an *in vivo* residence time of 1.45 s (**Figure 3.2 B** and **Table 3.2**). The value of k_b for YPet is consistent with that found in a control experiment performed *in vitro* under comparable illumination conditions (**Figure 3.2 C**).

For comparison, we probed the DNA binding of monomeric GR using a point mutant that is capable, upon induction, of nuclear import but not of dimerization (GR^{A458T}) (Heck et al., 1997). Importantly, a simple model with one dissociation rate constant was not sufficient to fit its fluorescence "on"-time distribution (**Figure 3.3 A**). Such deviation suggests the existence of more than one modes of dissociation, which call for the distribution to be fit by a double-exponential function

$$p_2(t) = A_1\exp\left[-\left(k_b\frac{\tau_{\mathrm{int}}}{\tau_{tl}} + k_{off,1}\right)t\right] + A_2\exp\left[-\left(k_b\frac{\tau_{\mathrm{int}}}{\tau_{tl}} + k_{off,2}\right)t\right] \quad \textbf{(3.5)}$$

with two distinct dissociation rate constants $k_{off,1}$ and $k_{off,2}$ and respective fractions A_1 and A_2. Using such fit, we found a major fraction (97% ± 2%) of GR^{A458T} molecules exhibiting a residence

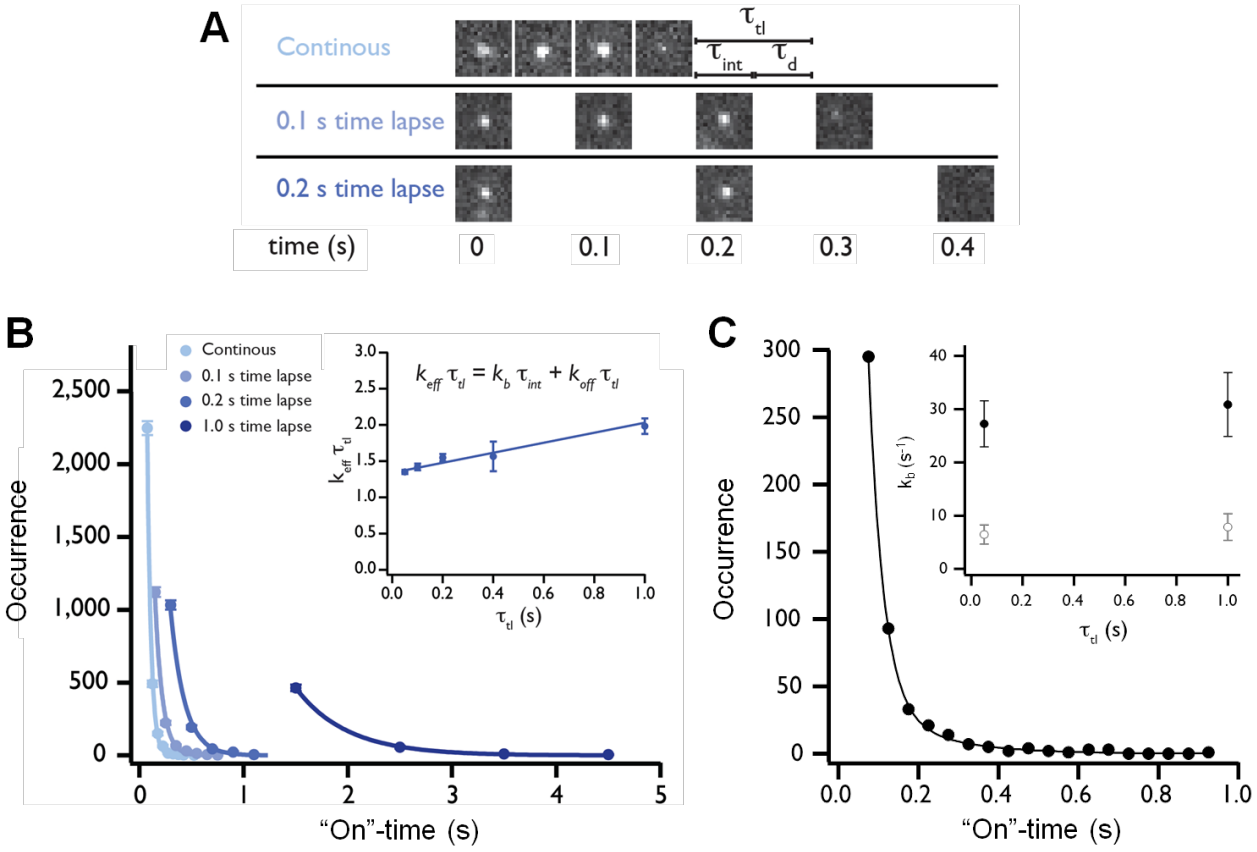

Figure 3.2: Characterization of binding dynamics and residence time of GR dimer *in vivo*. (A) Representative images of single DNA-bound YPet-GR molecules under induced condition during time-lapse imaging with a fixed integration time of 50 ms and varying dark times. **(B)** Histograms of fluorescence "on"-times for bound YPet-GR dimers under various time-lapse conditions: τ_{tl} = 0.05 s (or continuous illumination, n = 2991); τ_{tl} = 0.1 s (n = 1465); τ_{tl} = 0.2 s (n = 1308); and τ_{tl} = 1.0 s (n = 539) (data from a total of 34 cells). Lines denote fits by an exponential distribution with one effective off-rate constant (**Equation 3.4**). Inset shows extracted effective off-rate constant as a function of τ_{tl}. **(C)** Histogram of "on"-times of YPet only in an *in vitro* control experiment performed under continuous illumination. The distribution is best fit with a double-exponential function, yielding two distinct photobleaching rate constants. Inset shows both rate constants obtained at τ_{tl} = 0.05 s (n = 484) and τ_{tl} = 1 s (n = 412). The value of the faster rate constant (solid circles), which accounts for 97% of the molecules in both cases, agrees well with that measured for YPet-GR *in vivo*. Error bars denote ± s.d..

time of 0.15 ± 0.02 s, which is ten-fold shorter than dimeric GR, and a minor fraction (3% ± 2%) exhibiting a residence time of 0.76 ± 0.12 s (**Table 3.2**). To assign these components, we imaged another GR mutant lacking the DNA-binding domain ($GR^{\Delta DBD}$), and observed a single residence time of 0.76 ± 0.35 s, comparable to that of the minor fraction of GR^{A458T} (**Figure 3.3 A** and **Table 3.2**). The DNA-binding capability of the GR^{A458T} mutant (as well as the lack of such for the $GR^{\Delta DBD}$ mutant) was validated by their respective frequency of binding events (**Figure 3.3 B**). We therefore conclude that the binding exhibited by the minor fraction of monomeric GR^{A458T} molecules represents protein-protein interactions rather than direct binding to DNA.

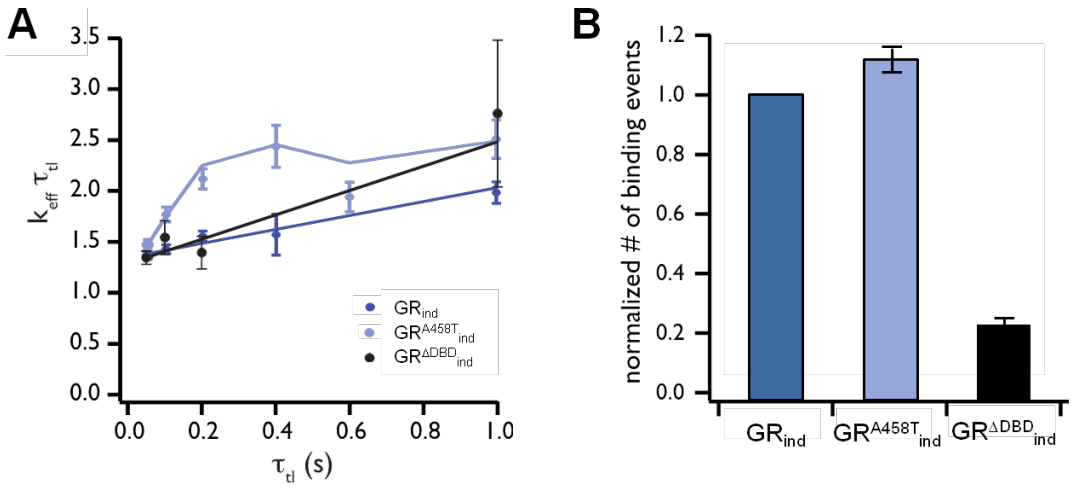

Figure 3.3. Characterization of binding dynamics and residence times of GR mutants *in vivo*. **(A)** Extracted effective off-rate constant as a function of time-lapse time for GR^{A458T} (purple) and $GR^{\Delta DBD}$ (black) in MCF-7 cells induced with 100 nM dexamethasone (for GR^{A458T}: n = 1274 (0.05 s), 846 (0.1 s), 681 (0.2 s), 253 (0.4 s), 260 (0.6 s), 277 (1.0 s), from a total of 29 cells; for $GR^{\Delta DBD}$: n = 584 (0.05 s), 135 (0.1 s), 109 (0.2 s), 35 (1.0 s), from a total of 16 cells). The data for induced GR (blue, from **Figure 3.2 B** inset) is shown for comparison. The distribution for the monomeric GR^{A458T} deviates significantly from single-exponential distribution, and is better fit with a double-exponential function (**Equation 3.5**) with two distinct dissociation rate constants. Error bars denote ± s.d.. **(B)** Relative frequency of binding events observed for GR, GR^{A458T} and $GR^{\Delta DBD}$ (all induced) under continuous illumination, normalized with respect to imaging duration, nuclear area, and average intensity of the nucleus (for GR: n = 8 cells; for GR^{A458T}: n = 9 cells; for $GR^{\Delta DBD}$: n = 6 cells). The value for GR_{ind} is set to 1. Error bars denote ± s.e.m..

In addition to GR, we also measured the residence time of the closely related ER, fused to YPet. Similar to GR, ER can be induced by hormone treatment (100 nM β-estradiol) to dimerize and bind to its cognate DNA sequences. We found a single residence time of 3.85 ± 0.30 s for induced ER; as for un-induced ER, we resolved a major fraction (87% ± 5%) exhibiting a residence time of 0.65 ± 0.14 s, which is six-fold shorter than induced ER, as well as a minor fraction (13% ± 5%) with a longer residence time of 4.35 ± 0.95 s (**Figure 3.4** and **Table 3.2**).

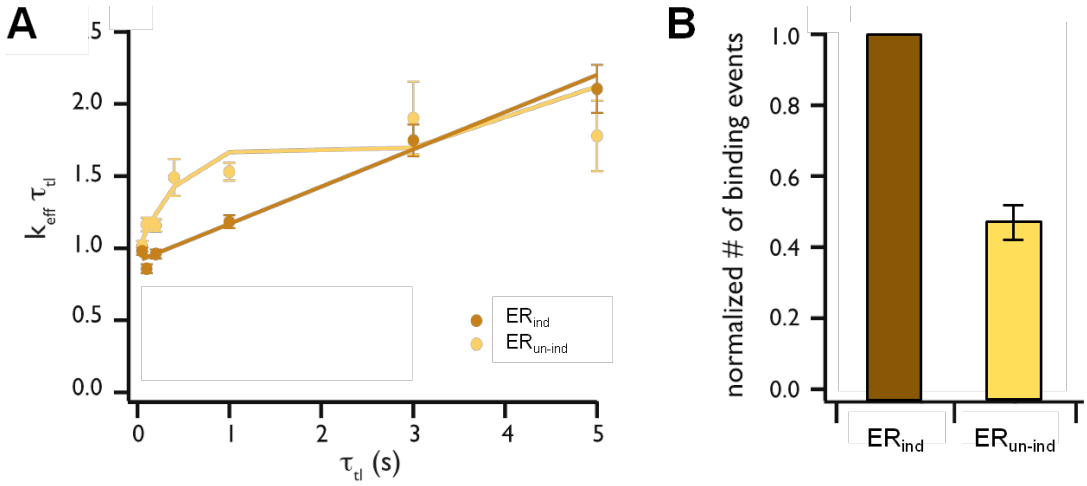

Figure 3.4. Characterization of binding dynamics and residence times of ER *in vivo*. (A) Extracted effective off-rate constant as a function of time-lapse time for induced (with 100 nM β-estradiol, brown) and un-induced (yellow) ER in MCF-7 cells (for ER_{ind}: n = 1848 (0.05 s), 1072 (0.1 s), 1254 (0.2 s), 927 (1.0 s), 393 (3.0 s), 281 (5.0 s), from a total of 7 cells; for $ER_{un\text{-}ind}$: n = 1089 (0.05 s), 813 (0.1 s), 983 (0.2 s), 214 (0.4 s), 876 (1.0 s), 109 (3.0 s), 80 (5.0 s), from a total of 14 cells). Error bars denote ± s.d.. **(B)** Relative frequency of binding events observed for induced and un-induced ER under continuous illumination, normalized with respect to imaging duration, nuclear area, and average intensity of the nucleus (for ER_{ind}: n = 5 cells; for $ER_{un\text{-}ind}$: n = 7 cells). The value for ER_{ind} is set to 1. Error bars denote ± s.e.m..

The results for the various GR and ER species are summarized in **Table 3.2** and **Figure 3.5**. Collectively, these results convincingly demonstrate that RLS microscopy allows us to discriminate, at the single-molecule level, between three different modes of TF binding to DNA:

dimeric, monomeric, or indirect DNA binding through association with other transcription factors or nuclear proteins.

Table 3.2. Residence times of various TFs and co-activator in MCF-7 or U2-OS cells.

Cell line	TF/Co-activator	$t_{res,1}$ (s)	$t_{res,2}$ (s)	$t_{res,1}$ fraction
MCF-7	GR_{ind}	1.45 ± 0.25	-	100%
MCF-7	GR^{A458T}_{ind}	0.15 ± 0.02	0.76 ± 0.12	97% ± 2%
MCF-7	$GR^{\Delta DBD}_{ind}$	0.76 ± 0.35	-	100%
MCF-7	ER_{ind}	3.85 ± 0.30	-	100%
MCF-7	$ER_{un\text{-}ind}$	0.65 ± 0.14	4.35 ± 0.95	87% ± 5%
U2-OS	GR_{ind}	1.75 ± 0.49	-	100%
U2-OS	$GRIP1_{ind}$	1.47 ± 0.67	-	100%

All proteins are fused to YPet. Values are reported as mean ± s.d..

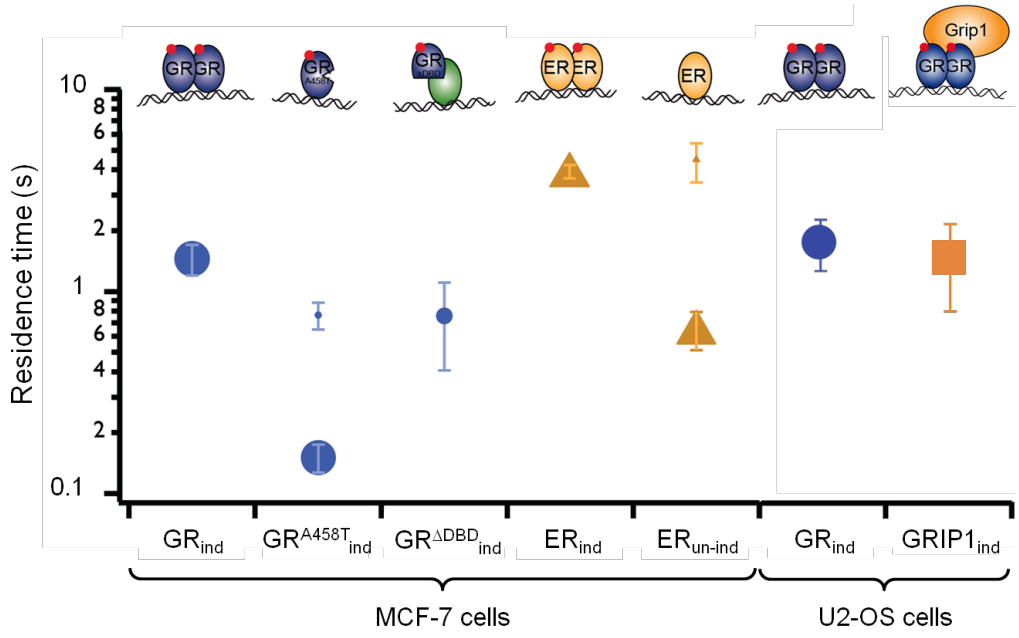

Figure 3.5. Graphical summary of *in vivo* residence times of various TFs and co-activator. The cartoons at the top illustrate the most abundant binding mode for each species. Symbol size is proportional to the fraction of molecules exhibiting that residence time, or in the case of $GR^{\Delta DBD}$, to the frequency of binding events relative to dimeric GR. Error bars denote ± s.d..

3.3.3. Spatio-temporal colocalization of two interacting molecular species in the mammalian nucleus

Lastly, we applied RLS microscopy to probe the interaction of GR and its transcriptional co-activator, GR interacting protein 1 (GRIP1) (Hong et al., 1996), in both space and time (**Figure 3.6**). To this end, we performed imaging in U2-OS cells, in which these factors are not endogenously expressed (Beck et al., 2011), thus allowing the exclusive observation of only the fluorescently labeled GR and GRIP1. YPet-GR fusion protein in induced U2-OS cells exhibited a single residence times of 1.75 ± 0.49 s, comparable to that measured in MCF-7 cells. GRIP1, which is known to interact with dimeric GR in a hormone-regulated manner (Hong et al., 1996), also exhibited a single residence time of 1.47 ± 0.67 s in induced U2-OS cells (**Figures 3.5** and **3.6 B** and **Table 3.2**).

For simultaneous observation of GR and GRIP1, we performed two-color single-molecule imaging by labeling GR with TagRFP-T and GRIP1 with EGFP. We alternated between 488 nm and 560 nm laser excitations with 50 ms integration time in the same light-sheet illumination plane, and detected the spatio-temporal colocalization of single molecules of GR and GRIP1 on DNA (**Figure 3.6 A**). By comparing the numbers of localizations per pixel and per second of GR and GRIP1 alone with the number of colocalization events detected, we estimate that the observed colocalization was ~ 80 times more likely than that expected from chance, thus suggesting that these events indeed reflect interaction between the two molecular species during transcriptional activation.

We also used the same fluorescent protein pair to label BMAL1 and CLOCK, two TFs that are known to bind to DNA as a heterodimer and drive the rhythmic gene expressions that regulate the mammalian circadian clock (Gekakis et al., 1998; Menet et al., 2014). Both proteins

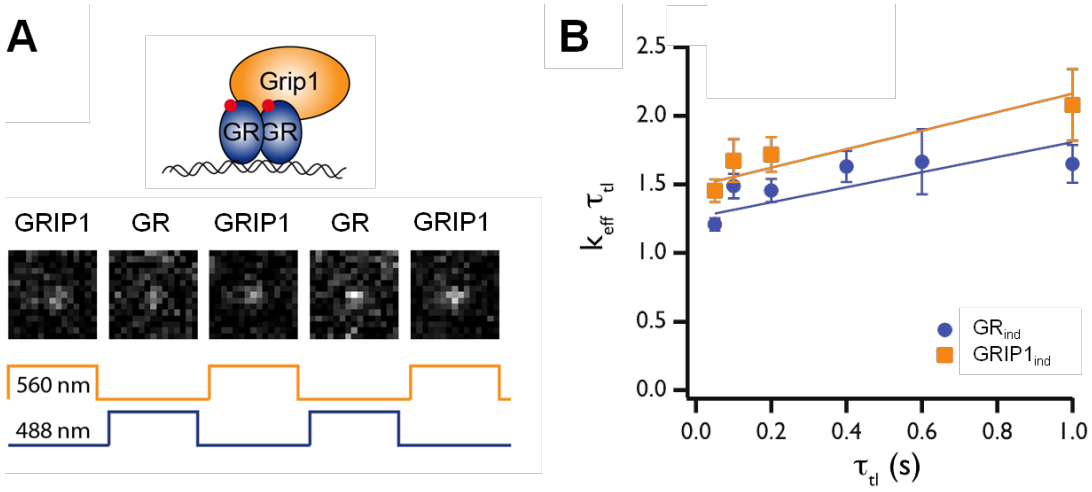

Figure 3.6. Two-color imaging of GR/GRIP1 interaction and their binding dynamics on DNA. **(A)** Representative frames from a two-color movie in which a single TagRFP-T-GR dimer (blue) and a single molecule of EGFP-GRIP1 (orange) was alternately excited and imaged, revealing their colocalization on a DNA locus *in vivo*. **(B)** Extracted effective off-rate constant a as function of time-lapse time for GR and GRIP1 in induced U2-OS cells (for GR_{ind}: n = 987 (0.05 s), 441 (0.1 s), 442 (0.2 s), 306 (0.4 s), 81 (0.6 s), (1.0 s), from a total of 26 cells; for $GRIP1_{ind}$: n = 445 (0.05 s), 196 (0.1 s), 273 (0.2 s), 110 (1.0 s), from a total of 10 cells). Error bars denote ± s.d..

showed colocalization events consistent with the formation of a complex composed of BMAL1, CLOCK, and largely stationary DNA (**Figure 3.7**). Similar to GR and GRIP1, the colocalization events observed were two orders of magnitude more likely than that expected from chance alone. Taken together, these results show that RLS microscopy allows us to directly interrogate the spatio-temporal interaction dynamics between two distinct molecular species at the single-molecule level in a living mammalian cell nucleus.

3.4. Discussion and conclusion

In this study, we adopted a time-lapse approach to characterize the DNA binding dynamics of transcription factors, accomplishing reliable measurements of residence times ranging from 50 ms (as given by the integration time) to several seconds. Since this strategy does not rely on the

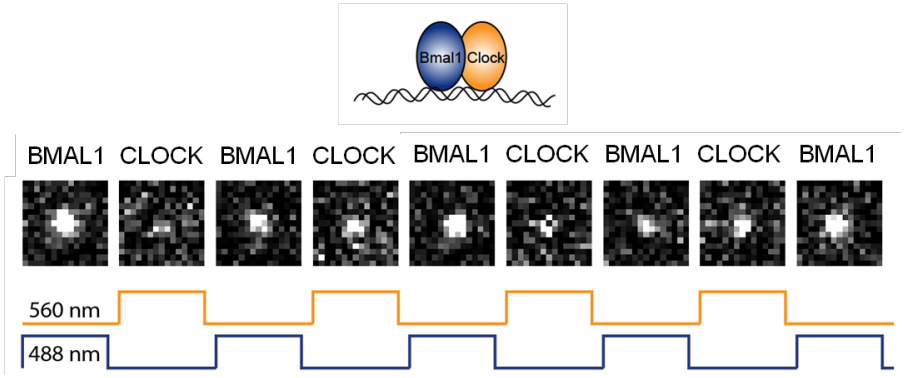

Figure 3.7. Spatio-temporal colocalization of BMAL1 and CLOCK on DNA. Representative frames from a two-color movie are shown, in which a single molecule of TagRFP-T-BMAL1 (blue) and of EGFP-CLOCK (orange) was alternately excited and imaged, revealing their colocalization on a DNA locus *in vivo*.

detection of long continuous traces of a fluorescent molecule, it is therefore well suited for imaging fluorescent protein labels with fast photobleaching kinetics. In fact, our time-lapse approach should prove advantageous compared to analysis of continuous traces even when more photostable organic dyes are used as labels, as long as the residence time of the DNA-binding protein of interest is on the order of seconds. For longer time scales, both approaches have limitations, since cellular movements will prevent the reliable assignment of a bound molecule in time-lapse movies, and continuously fluorescing dye molecules will become increasingly sparse due to photobleaching.

The increase in the residence time of dimeric GR and ER that we observed, compared to the monomeric form, probably reflects stabilization of DNA binding by an associated partner. However, our observations are also compatible with a proportion of molecules remaining in the monomeric form, since the dynamics of a fast-dissociating fraction of molecules cannot be resolved if majority of the molecules dissociates slowly. In contrast, a small fraction of longer

bound molecules was resolved for monomeric GR and ER, which we could assign to an indirect binding mode to other protein factors in the nucleus.

As validation for our approach, the values of effective diffusion constants for GR measured in our study agree very well with those reported for dye-labeled STAT1 in the nucleus (Speil et al., 2011). Our estimates for the DNA-bound fraction as well as residence time of nuclear GR are also similar to previous measurements on STAT1 and p53 (Mazza et al., 2012; Speil et al., 2011), lending further credence to our results. More recently, as a consequence of continuous efforts to improve the analysis procedures of FCS and FRAP data over the years (Mazza et al., 2012), a series of FCS and FRAP studies have obtained a set of values for the *in vivo* DNA-bound fraction and residence time of GR that are consistent with those found by our SMT approach (**Figure 3.8**) (Mazza et al., 2013; Stasevich et al., 2010).

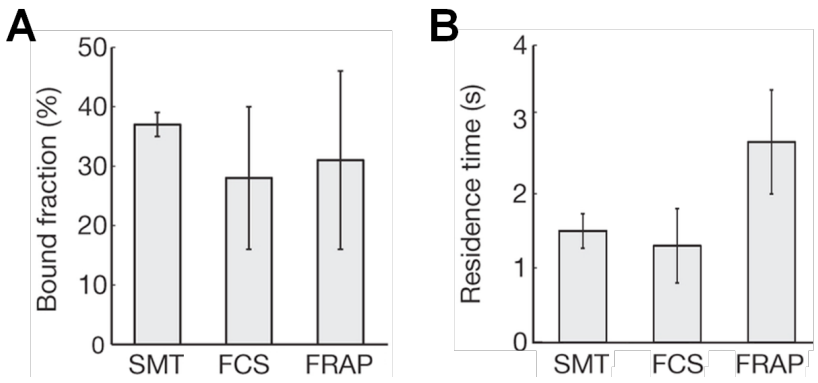

Figure 3.8. Comparison of DNA-bound fraction and residence time of GR obtained using three different techniques: SMT, FCS and FRAP. Error bars denote ± s.d.. (Figure adapted from Mazza et al., 2013, with modifications).

While FCS and FRAP have both been very useful in the study of *in vivo* dynamics of a wide range of biomolecules in mammalian systems, the SMT approach made possible by RLS

microscopy still possesses the advantage of being able to *directly* monitor transcription factor dynamics, without the need for additional calibrations or corrections that are commonly associated with FRAP and other techniques (Mueller et al., 2010). In addition, the nanometer spatial and millisecond temporal accuracy in the single-molecule trajectories accessible with our approach could facilitate high-resolution studies that are unachievable via other means. It is therefore foreseeable that SMT, FRAP and FCS, as a set of self-consistent and complementary techniques, will generate new mechanistic insights into the *in vivo* dynamics of TF binding, transcription initiation, as well as other key molecular processes in the mammalian nucleus for years to come.

3.5. Materials and methods

DNA constructs

The vector containing SNAP-tag (pSNAPf) was purchased from New England Biolabs, and that containing Halo-tag (pHTN HaloTag CMV-neo) was purchased from Promega. The mEos2-GR and mEos2-H4 constructs were generated by fusing GR or H4 to mEos2 (McKinney et al., 2009) in the pSNAPf vector. YPet (Nguyen and Daugherty, 2005), EGFP, and Tag-RFP-T (Shaner et al., 2008) fusion constructs were generated by using pLV-TetO-Oct4 (gift of K. Hochedlinger) as backbone. Briefly, the Oct4 coding sequence was replaced with the coding sequence for each of the fusion proteins. All constructs were generated as N-terminal fusions of the fluorescent protein to the protein of interest. The YPet-GR A458T mutant was generated by site-directed mutagenesis. To generate the DNA-binding domain deletion mutants of GR, PCR products of the sequences both upstream and downstream of the DNA binding domain (as defined in the Uniprot database at http://www.uniprot.org/) were generated and ligated together.

Generation of stable cell lines

MCF-7 cells stably expressing mEos2, mEos2-GR and mEos2-H4 were generated by transfection of the plasmids with Polyplus reagent and selected with puromycin for 2 - 3 weeks. All other cell lines were generated by lentiviral transduction. Briefly, each construct was co-transfected with the packaging plasmids MD2G and PAX2 (gift of D. Trono) in 293T cells using the Lipofectamine 2000 reagent (Life Technologies). Supernatants were collected 48 hours after transfection and filtered through 0.45 μm low-protein-binding filters (Pall Corporation), and 1 - 2 ml of the supernatant was used to transduce 3 - 5 × 10^4 U2-OS or MCF-7 cells.

Cell culture

U2-OS cells were cultured in high glucose DMEM supplemented with 10% fetal bovine serum (FBS), 1% penicillin/streptomycin and 2 mM GlutaMax (all from Gibco). MCF-7 cells were cultured in a medium (mMCF7) comprising of Advanced MEM supplemented with 10% FBS, 1% penicillin/streptomycin, 2 mM L-glutamine, 1 mM sodium pyruvate and 100 mM non-essential amino acids (all from Gibco). To achieve un-induced conditions for GR or ER, cells were grown for at least one day in mMCF7 containing charcoal stripped FBS (mMCF7-) and one day in mMCF7- without phenol red. Just before imaging, OptiMEM was used to wash the cells once and maintain them for imaging. To induce the activity of GR or ER, cells were treated for 30 min with 100 nM dexamethasone or 100 nM β-estradiol, respectively. For the U2-OS cell lines stably expressing BMAL1 and CLOCK fusion constructs, cells were first synchronized by serum shock. Briefly, cells were first suspended in a 1:1 mix of complete culture medium and FBS for one hr before changing to a phenol red-free medium. Cells were incubated for 24 more hr, and then imaged without washing or changing buffer, except for the two-color experiments in which OptiMEM was used for imaging.

Image acquisition

Cells were imaged in Delta-T glass-bottom dishes (Bioptechs) at 36 °C for up to 45 min (GR and GR mutants), 60 min (un-induced ER), and 120 min (induced ER). For the control characterization of YPet photobleaching *in vitro*, His-tag fused YPet protein molecules expressed in *E. coli* and purified with a Ni-NTA column were adsorbed nonspecifically to a cleaned coverslip. During imaging, the reflecting mirror was removed and the focal cross-section of the vertical light-sheet was used for excitation, in order to achieve illumination intensity comparable to *in vivo* experiments.

Image and statistical analysis

Single-molecule tracking was performed essentially as described previously (Elf et al., 2007). Briefly, images were background-subtracted using Matlab (2010b, MathWorks). After additional smoothing, the coarse positions of fluorescent molecules were determined from pixel values exceeding a threshold of 4 times the standard deviation over the background. The refined positions of the molecules were obtained by a two-dimensional Gaussian fitting to the background-subtracted image. A molecule is considered as bound when it is localized for at least two consecutive frames to within 0.5 pixel (or 1 pixel in two-color experiments). Localizations only visible in a single frame were discarded (to avoid counting slowly moving molecules). An allowance of one dark frame within a trajectory was given to account for rare blinking events of the fluorescent protein at an integration time of 50 ms. To avoid bias towards slowly moving molecules that remain visible for longer times, only the first displacement step of each trajectory was counted. On the other hand, to avoid false assignment of fast molecules to another trajectory, an upper limit of 6 pixels was set for the maximum squared displacement for a molecule. This limit was accounted for by replacing the last term in **Equation 3.3** with

$$\left(1-A_1-A_2\right)\left[\exp\left(-\frac{X}{D_3}\right)-\exp\left(-\frac{C_1}{D_3}\right)\right]\Big/\left[\exp\left(-\frac{C_2}{D_3}\right)-\exp\left(-\frac{C_1}{D_3}\right)\right],$$

where the constants C_1 and C_2 are given by the lower and upper limits for the squared displacement of 0 and 6 pixels, respectively (Gebhardt et al., 2006).

When extracting diffusion and rate constants, all fitting was performed in Igor Pro, v. 6.2 (WaveMetrics), with a nonlinear least-square fitting procedure. Errors were calculated as the s.d. of parameters obtained from fits to 2000 random subsets of the displacements, each comprising 80% of the original dataset.

References

Aagaard, M.M., Siersbaek, R., and Mandrup, S. (2011). Molecular basis for gene-specific transactivation by nuclear receptors. *Biochim. Biophys. Acta* **1812**, 824-835.

Akhtar, A., and Gasser, S.M. (2007). The nuclear envelope and transcriptional control. *Nat. Rev. Genet.* **8**, 507-517.

Bancaud, A., Huet, S., Daigle, N., Mozziconacci, J., Beaudouin, J., and Ellenberg, J. (2009). Molecular crowding affects diffusion and binding of nuclear proteins in heterochromatin and reveals the fractal organization of chromatin. *EMBO J.* **28**, 3785-3798.

Beaudouin, J., Mora-Bermudez, F., Klee, T., Daigle, N., and Ellenberg, J. (2006). Dissecting the contribution of diffusion and interactions to the mobility of nuclear proteins. *Biophys. J.* **90**, 1878-1894.

Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymborska, A., Herzog, F., Rinner, O., Ellenberg, J., and Aebersold, R. (2011). The quantitative proteome of a human cell line. *Mol. Syst. Biol.* **7**, 549.

Becker, M., Baumann, C., John, S., Walker, D.A., Vigneron, M., McNally, J.G., and Hager, G.L. (2002). Dynamic behavior of transcription factors on a natural promoter in living cells. *EMBO Rep.* **3**, 1188-1194.

Elf, J., Li, G.W., and Xie, X.S. (2007). Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191-1194.

Gebhardt, J.C.M., Clemen, A.E., Jaud, J., and Rief, M. (2006). Myosin-V is a mechanical ratchet. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8680-8685.

Gekakis, N., Staknis, D., Nguyen, H.B., Davis, F.C., Wilsbacher, L.D., King, D.P., Takahashi, J.S., and Weitz, C.J. (1998). Role of the CLOCK protein in the mammalian circadian mechanism. *Science* **280**, 1564-1569.

Hager, G.L., McNally, J.G., and Misteli, T. (2009). Transcription Dynamics. *Mol. Cell* **35**, 741-753.

Heck, S., Bender, K., Kullmann, M., Gottlicher, M., Herrlich, P., and Cato, A.C. (1997). IκBα-independent downregulation of NF-κB activity by glucocorticoid receptor. *EMBO J.* **16**, 4698-4707.

Hong, H., Kohli, K., Trivedi, A., Johnson, D.L., and Stallcup, M.R. (1996). GRIP1, a novel mouse protein that serves as a transcriptional coactivator in yeast for the hormone binding domains of steroid receptors. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 4948-4952.

Kaur, G., Costa, M.W., Nefzger, C.M., Silva, J., Fierro-González, J.C., Polo, J.M., Bell, T.D.M., and Plachta, N. (2013). Probing transcription factor diffusion dynamics in the living mammalian embryo with photoactivatable fluorescence correlation spectroscopy. *Nat. Commun.* **4**, 1637.

Keppler, A., Gendreizig, S., Gronemeyer, T., Pick, H., Vogel, H., and Johnsson, K. (2003). A general method for the covalent labeling of fusion proteins with small molecules *in vivo*. *Nat. Biotechnol.* **21**, 86-89.

Larson, D.R. (2011). What do expression dynamics tell us about the mechanism of transcription? *Curr. Opin. Genet. Dev.* **21**, 591-599.

Larson, D.R., Zenklusen, D., Wu, B., Chao, J.A., and Singer, R.H. (2011). Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science* **332**, 475-478.

Los, G.V., Encell, L.P., McDougall, M.G., Hartzell, D.D., Karassina, N., Zimprich, C., Wood, M.G., Learish, R., Ohana, R.F., Urh, M., *et al.* (2008). HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* **3**, 373-382.

Macháň, R., and Wohland, T. (2014). Recent applications of fluorescence correlation spectroscopy in live systems. *FEBS Lett.* **588**, 3571-3584.

Manley, S., Gillette, J.M., Patterson, G.H., Shroff, H., Hess, H.F., Betzig, E., and Lippincott-Schwartz, J. (2008). High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nat. Methods* **5**, 155-157.

Mazza, D., Abernathy, A., Golob, N., Morisaki, T., and McNally, J.G. (2012). A benchmark for chromatin binding measurements in live cells. *Nucleic Acids Res.* **40**, e119.

Mazza, D., Mueller, F., Stasevich, T.J., and McNally, J.G. (2013). Convergence of chromatin binding estimates in live cells. *Nat. Methods* **10**, 691-692.

McKinney, S.A., Murphy, C.S., Hazelwood, K.L., Davidson, M.W., and Looger, L.L. (2009). A bright and photostable photoconvertible fluorescent protein. *Nat. Methods* **6**, 131-133.

McNally, J.G., Muller, W.G., Walker, D., Wolford, R., and Hager, G.L. (2000). The glucocorticoid receptor: Rapid exchange with regulatory sites in living cells. *Science* **287**, 1262-1265.

Menet, J.S., Pescatore, S., and Rosbash, M. (2014). CLOCK:BMAL1 is a pioneer-like transcription factor. *Genes Dev.* **28**, 8-13.

Mueller, F., Mazza, D., Stasevich, T.J., and McNally, J.G. (2010). FRAP and kinetic modeling in the analysis of nuclear protein dynamics: What do we really know? *Curr. Opin. Cell Biol.* **22**, 403-411.

Mueller, F., Morisaki, T., Mazza, D., and McNally, J.G. (2012). Minimizing the impact of photoswitching of fluorescent proteins on FRAP analysis. *Biophys J.* **102**, 1656-1665.

Nguyen, A.W., and Daugherty, P.S. (2005). Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nat. Biotechnol.* **23**, 355-360.

Schutz, G.J., Schindler, H., and Schmidt, T. (1997). Single-molecule microscopy on model membranes reveals anomalous diffusion. *Biophys J.* **73**, 1073-1080.

Shaner, N.C., Lin, M.Z., McKeown, M.R., Steinbach, P.A., Hazelwood, K.L., Davidson, M.W., and Tsien, R.Y. (2008). Improving the photostability of bright monomeric orange and red fluorescent proteins. *Nat. Methods* **5**, 545-551.

Speil, J., Baumgart, E., Siebrasse, J.P., Veith, R., Vinkemeier, U., and Kubitscheck, U. (2011). Activated STAT1 transcription factors conduct distinct saltatory movements in the cell nucleus. *Biophys J.* **101**, 2592-2600.

Sprague, B.L., Muller, F., Pego, R.L., Bungay, P.M., Stavreva, D.A., and McNally, J.G. (2006). Analysis of binding at a single spatially localized cluster of binding sites by fluorescence recovery after photobleaching. *Biophys J.* **91**, 1169-1191.

Sprague, B.L., Pego, R.L., Stavreva, D.A., and McNally, J.G. (2004). Analysis of binding reactions by fluorescence recovery after photobleaching. *Biophys J.* **86**, 3473-3495.

Stasevich, T.J., Mueller, F., Michelman-Ribeiro, A., Rosales, T., Knutson, J.R., and McNally, J.G. (2010). Cross-validating FRAP and FCS to quantify the impact of photobleaching on *in vivo* binding estimates. *Biophys J.* **99**, 3093-3101.

Stavreva, D.A., Varticovski, L., and Hager, G.L. (2012). Complex dynamics of transcription regulation. *Biochim. Biophys. Acta.* **1819**, 657-666.

Thompson, R.E., Larson, D.R., and Webb, W.W. (2002). Precise nanometer localization analysis for individual fluorescent probes. *Biophys J.* **82**, 2775-2783.

Tsai, S.Y., Carlstedt-Duke, J., Weigel, N.L., Dahlman, K., Gustafsson, J.A., Tsai, M.J., and O'Malley, B.W. (1988). Molecular interactions of steroid hormone receptor with its enhancer element: Evidence for receptor dimer formation. *Cell* **55**, 361-369.

Watson, J.D., Baker, T.A., Bell, S.P., Gann, A., Levine, M., and Losick, R. (2007). Molecular Biology of the Gene, 6th ed. (San Francisco, Benjamin Cummings).

Yao, J., Munson, K.M., Webb, W.W., and Lis, J.T. (2006). Dynamics of heat shock factor association with native gene loci in living cells. *Nature* **442**, 1050-1053.

Zava, D.T., Chamness, G.C., Horwitz, K.B., and McGuire, W.L. (1977). Human breast cancer: Biologically active estrogen receptor in the absence of estrogen? *Science* **196**, 663-664.

Chapter 4

Super-resolution imaging of RNA polymerase II-mediated transcription with RLS-SRM [1]

4.1. Abstract

The transcription of eukaryotic genes by RNA polymerase II (RNAP II) has long been proposed to occur heterogeneously in discrete "transcription factories" inside the nucleus. Yet, estimates of the number of such "factories", their size, as well as the stoichiometry of their constituents have been subject to a great deal of variability, primarily due to the inability of previous studies to visualize these structures with sufficient spatial resolution and molecular specificity, as well as to accurately count the copy number of RNAP II molecules in them. Using RLS-SRM, we mapped the spatial organization of RNAP II-mediated transcription inside the mammalian nucleus at sub-diffraction-limit resolution. To quantify their global extent of clustering, we also developed a spatio-temporal clustering algorithm capable of molecular counting with single-copy accuracy by leveraging on the blinking photophysics of rhodamine-based dyes. Application of such analysis to RLS-SRM images revealed that majority (> 70%) of the

[1] The work described in this chapter was performed in collaboration with Drs. Rahul Roy, J. Christof M. Gebhardt, and David M. Suter.

transcription foci originate from single RNAP II molecules, and no significant clustering between RNAP II molecules was detected within the length scale of the reported diameters of "transcription factories". Colocalization measurements of RNAP II molecules equally labeled by two spectrally distinct dyes confirmed the primarily unclustered distribution, arguing against the prevalent existence of "transcription factories" as previously proposed. Our study suggests that the model of concerted transcription of mammalian genes in these "factories" needs to be reconsidered, while presenting quantitative mapping and stoichiometric characterization tools to further probe into the functional significance of RNAP II (un)clustering in gene regulation.

4.2. Introduction: Spatial organization of mammalian transcription and the "transcription factories" hypothesis

RNAP II, arguably one of the most important protein species in the nucleus, is the main workhorse of mammalian gene transcription, being responsible for the synthesis of majority of the pre-mRNA transcripts as well as most of the snRNAs and microRNAs. Composed of 12 subunits with a total molecular weight of ~ 550 kDa, the biochemical and structural mechanisms of its action have been worked out to excruciating detail, thanks to many decades of intense research (Cramer, 2004; Herbert et al., 2008; Kornberg, 2007; Liu et al., 2013; Svetlov and Nudler, 2013). Interestingly, the "geography" of this key enzyme, *i.e.* how it is distributed inside the nucleus, has been subject to substantial controversy.

The spatial organization of eukaryotic transcription by RNAP II has long been proposed to be heterogeneous (Chakalova and Fraser, 2010; Edelman and Fraser, 2012; Papantonis and Cook, 2010; Papantonis and Cook, 2013; Rieder et al., 2012; Sutherland and Bickmore, 2009). Such spatial heterogeneity was first suggested based on the observation that nascent mRNA

transcripts are not uniformly distributed in the nucleus, but rather occur in discrete foci termed "transcription factories" (**Figure 4.1**) (Jackson et al., 1993; Wansink et al., 1993). Yet, estimates of the number of such factories, their size, as well as the stoichiometry of their constituents have been subject to a great deal of variability. Immunofluorescence and electron microscopy studies performed on a variety of human and mouse cell types have found in each nucleus 300 - 8000 of such "factories", each proposed to contain 4 - 60 RNAP II molecules and with a mean diameter ranging from 40 - 130 nm (**Table 4.1**) (Brown et al., 2008; Eskiw and Fraser, 2011; Eskiw et al., 2008; Iborra et al., 1996; Jackson et al., 1998; Osborne et al., 2004; Pombo et al., 1999).

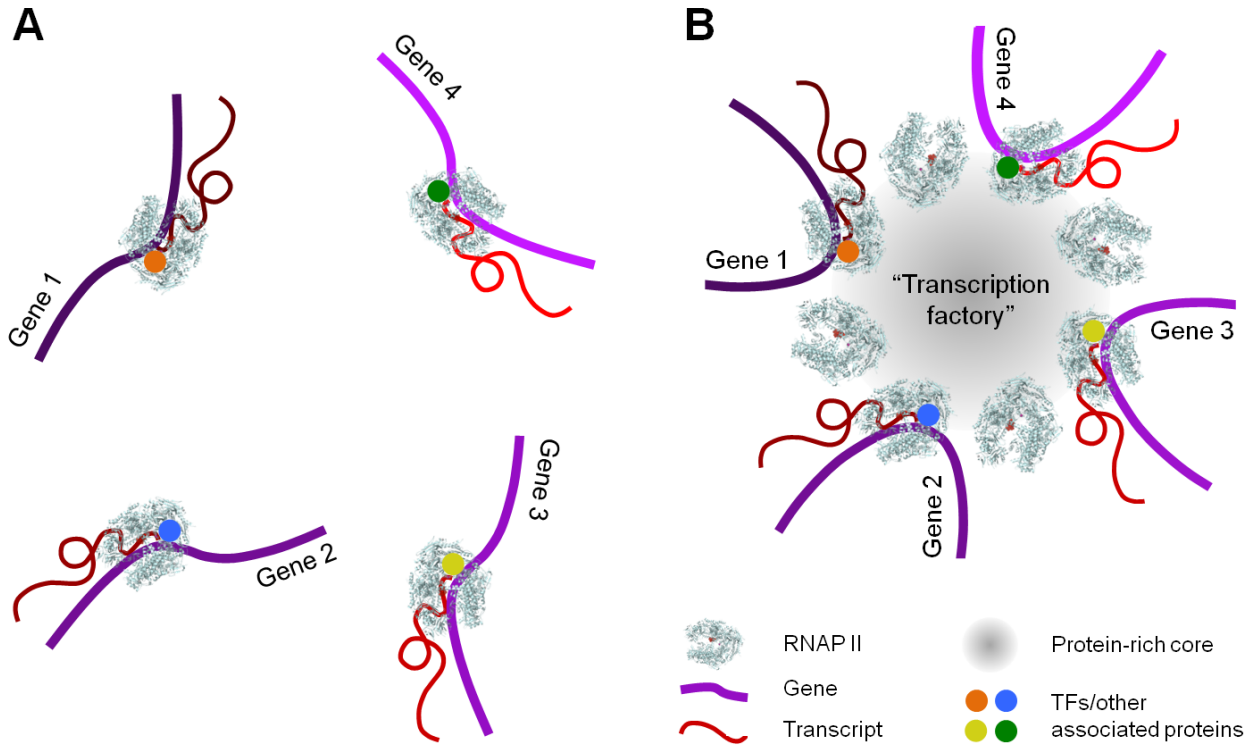

Figure 4.1. The "transcription factories" hypothesis. Two models of the spatial distribution of RNAP II in mammalian nucleus (not drawn to scale), in which transcription is carried out by either **(A)** unclustered individual RNAP II molecules or **(B)** multiple molecules clustered into spatially discrete "factories" that pull together genes that need to be transcribed (The crystal structure of RNAP II used follows Bushnell et al., 2002).

Table 4.1. Number, size, and composition of "transcription factories" in both human and mouse cells previously reported in literature.

Cell type	Visualization technique	No. of factories	Diameter of factory (nm)	No. of RNAP II per factory	Reference
HeLa	Confocal microscopy	300 - 500	-	60 (inferred)	Jackson et al., 1993
Human bladder/skin fibroblast	Epi-fluorescence microscopy	Several hundred	-	-	Wansink et al., 1993
HeLa	Electron microscopy	2000 - 2700	71	4 - 20 (inferred)	Iborra et al., 1996
HeLa	Electron microscopy	2400	80	30 (inferred)	Jackson et al., 1998
HeLa	Confocal & cryo-electron microscopy	8000	40	8 (inferred)	Pombo et al., 1999
HeLa	Electron spectroscopic imaging	-	87	-	Eskiw et al., 2008
Human erythroblast	Confocal microscopy	1138 - 1897	-	-	Brown et al., 2008
Mouse embryonic fibroblast	Confocal microscopy	2000 (for Ser5-RNAP II only)*	-	-	Osborne et al., 2004
Mouse erythrocyte	Light & electron microscopy	545	130	-	Eskiw et al., 2011

*: only RNAP II phosphorylated at the Ser5 position are included.

Such inconsistency in estimates has led to suspicions about the existence of these "factories". In order to ascertain whether, and if so to what extent do mammalian RNAP II molecules exist in such clustered form, an imaging method with high spatial resolution, molecular specificity, as well as the ability to accurately count the copy number of RNAP II

molecules in these transcription foci is needed. Among the previously used techniques, electron microscopy, which lacks labeling specificity in general, is not suitable for the task. Conventional optical microscopy is also incapable of resolving the structure of these foci down to single-molecule level, especially given that RNAP II is present at very high density in the mammalian nucleus (ranging from 60,000 to 75,000 molecules per nucleus (Jackson et al., 1998; Jackson et al., 2000)). Moreover, while some previous super-resolution studies have been able to probe sub-nuclear structures in mammalian cells (Baddeley et al., 2010; Bohn et al., 2010), they do not provide absolute quantification in molecular copy numbers. Aside from less-than-complete labeling of molecular species, the difficulty stems primarily from the fact that the blinking photophysics of a dye results in multiple detection events for the same molecule over the course of image acquisition. Assignment of these blinking events to a specific molecule is often difficult, especially in dense structures where intermolecular distances are comparable to localization accuracy. Spatial pairwise correlation analysis can predict protein density at length scales larger than the localization accuracy (Sengupta et al., 2011; Veatch et al., 2012), but the absolute copy number of probes within the length scale of localization accuracy (~ 25 nm laterally and ~ 50 nm axially) is difficult to ascertain.

In this chapter, we applied our RLS-SRM technique to probe the spatial organization of RNAP II-mediated transcription at sub-diffraction-limit resolution. Leveraging on the blinking photophysics of rhodamine-based dyes, we also developed a density-based clustering algorithm that pools multiple localizations in super-resolution images based on their spatial as well as temporal proximity, so as to be able to count the number of RNAP II molecules in transcription foci with single-copy accuracy. Contrary to previous proposals, we found that majority of the RNAP II molecules do not form multi-molecular clusters inside the nucleus. Our

global quantification study provides clear and convincing answers to the controversy, and suggests that the model for coordinated transcription of mammalian genes in prevalently existing "transcription factories" needs to be revisited (Chakalova and Fraser, 2010; Cook, 2002; Edelman and Fraser, 2012; Papantonis and Cook, 2010; Papantonis and Cook, 2013; Rieder et al., 2012; Sutherland and Bickmore, 2009). At the same time, the general applicability of the methods introduced in this Chapter should open doors to the quantitative characterization of important biomolecular species in other mammalian nuclear structures.

4.3. Results

4.3.1. One-to-one labeling of RNAP II molecules *in situ*

Even before imaging, an essential prerequisite for accurately determining the structure and composition of transcription foci is the ability to perform quantitative labeling of RNAP II molecules. While immuno-tagging of the highly accessible C-terminal domain (CTD) or nascent mRNAs containing many degenerate sites for antibody binding ensures high efficiency of labeling, the high possibility of having multiple antibodies binding to each RNAP II molecule (which contains 52 heptad repeats in its CTD) or several dye molecules per secondary antibody make accurate determination of RNAP II copy number difficult.

To circumvent these complications and ensure one-to-one labeling, we generated a U2-OS cell line expressing a α-amanitin-resistant mutant of the major subunit of RNAP II, RPB1, fused to a SNAP tag (Darzacq et al., 2007; Nguyen et al., 1996). SNAP tag is a small protein tag that reacts specifically with cell-permeable benzylguanine derivatives containing an organic dye of our choice. This reaction can take place inside a living cell, leading to irreversible covalent labeling of each RPB1 with a single dye molecule (**Figure 4.2 A**). When these cells are

maintained in the presence of the drug α-amanitin, the endogenous RPB1 are degraded (Nguyen et al., 1996), and the cells thrive off the transcription by the exogenously produced mutant RPB1 molecules, each of which is labeled by exactly one dye molecule (**Figure 4.2 B**).

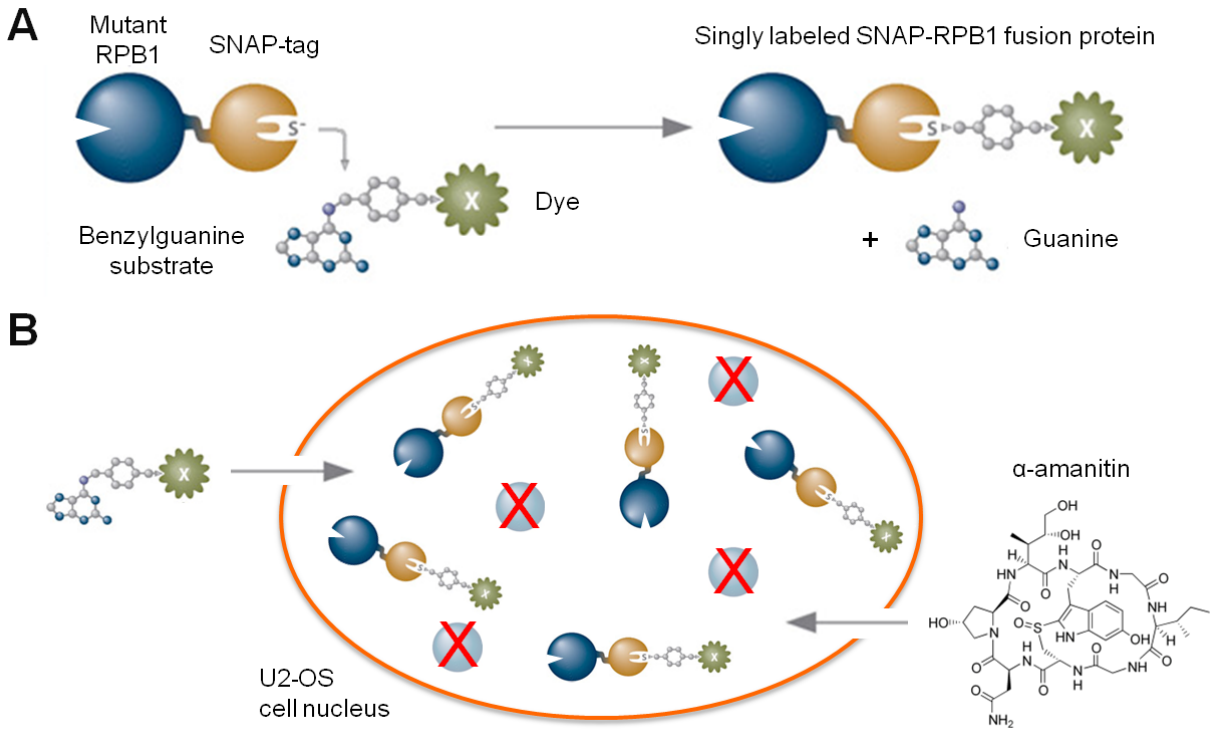

Figure 4.2. One-to-one labeling of RNAP II molecules using SNAP tag. (A) Reaction between SNAP-tagged mutant RPB1 and a dye-labeled benzylguanine substrate, leading to singly labeled RNAP II molecules. **(B)** Schematic diagram showing the labeling reaction inside a live cell nucleus in the presence of α-amanitin, which results in the degradation of all endogenous RPB1 molecules (marked by red crosses) and allows only the labeled molecules to be detected. (Figure adapted from https://www.neb.com/tools-and-resources/feature-articles/snap-tag-technologies-novel-tools-to-study-protein-function, with modifications).

4.3.2. Accurate molecular counting with spatio-temporal clustering analysis

In order to quantify the copy number of RNAP II molecules within the transcription foci, we took advantage of the fact that the photoblinking events of many rhodamine-based dyes, such

as TMR and Atto 550, tend to be clustered temporally, *i.e.* characterized by large bursts of "on" events followed by long dark periods (corresponding to the "off" state) or photobleaching, in contrast to those of cyanine-based dyes such as Alexa 647, which tend to be more evenly distributed in time (**Figure 4.3**). Using the dye TMR, we developed a density-based clustering

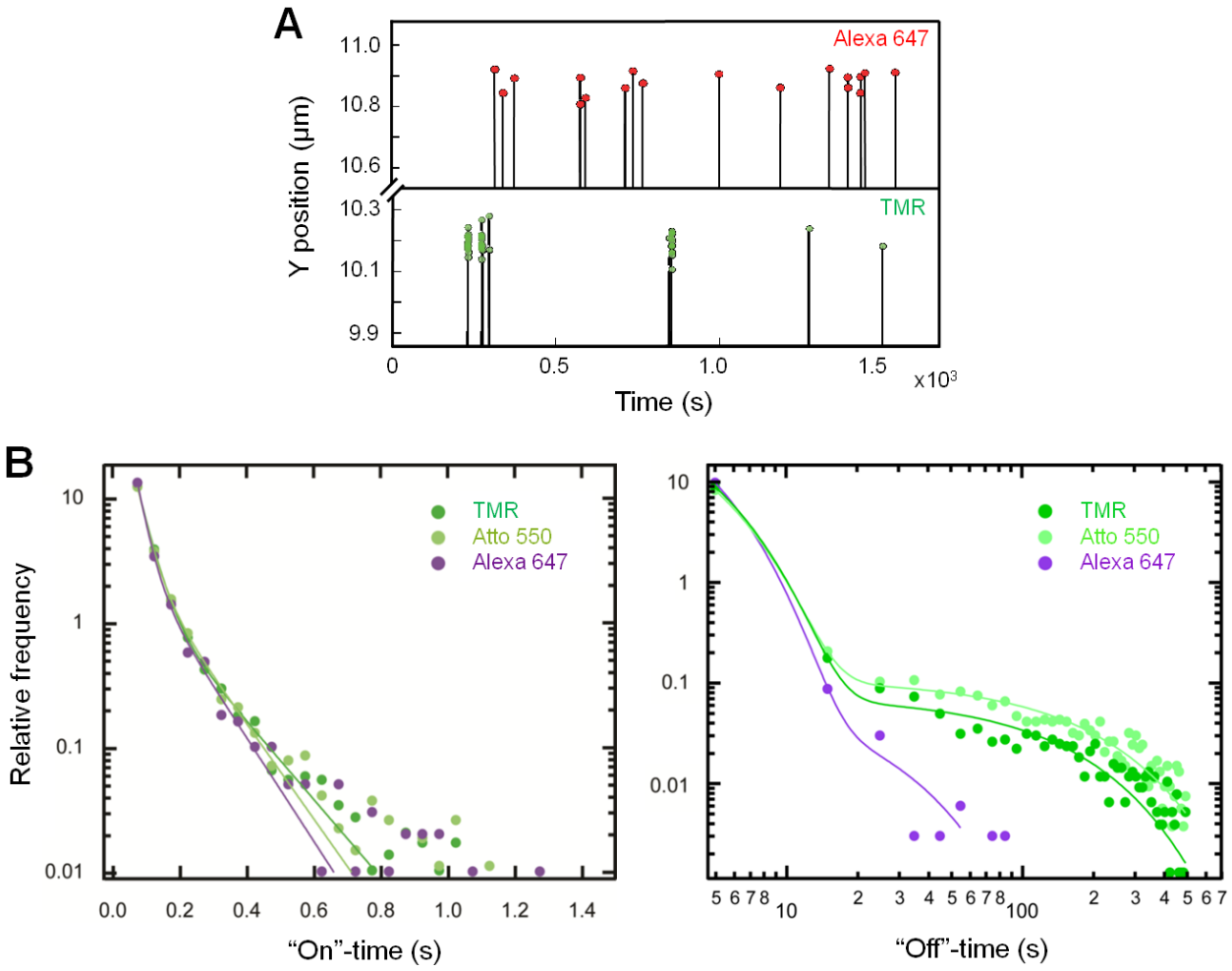

Figure 4.3. Blinking kinetics of rhodamine-based dyes enables spatio-temporal clustering analysis. (A) Comparison of representative blinking time traces between a Alexa 647 (top) and a TMR (bottom) dye molecule, imaged *in vitro* on a coverslip. **(B)** Distributions of the "on"-time (left) and "off"-time (right) for TMR, Atto 550 (both rhodamine-based) and Alexa 647 (cyanine-based). While all three dyes remain "on" for similar periods of time, TMR and Atto 550 tend to exhibit much longer "off"-times, making their blinking trajectories appear clustered in time and allowing temporal clustering to be performed.

algorithm that pools multiple localizations in the RLS-SRM dataset based on their proximity not only in space but also in time, so as to accurately assign localizations to spatio-temporal (st-) clusters (**Figure 4.4 A** and Section 4.5), each of which arises from a "burst" event of a single dye.

The temporal clustering of TMR localizations was calibrated using artificial clusters of dye molecules imaged under conditions similar to cellular imaging. To that end, we designed four short DNA constructs, each labeled with a TMR dye, as well as a long backbone construct complementary to the four constructs. The backbone was separately hybridized to either one or multiple (2 to 4) short constructs, mimicking transcription foci with one or multiple RNAP II molecules. The average number of dyes per construct molecule was determined by counting the number of photobleaching steps for hundreds of construct molecules (**Figure 4.4 B**); these values were compared with the average number of st-clusters per spatial cluster determined from spatio-temporal clustering analysis of the construct molecules. This calibration allowed us to establish an unambiguous correspondence between the average number of st-clusters observed in images of TMR-labeled RNAP II and the average copy number of RNAP II molecules in the transcription foci (**Figure 4.4 C** and Section 4.5).

4.3.3. RLS-SRM imaging and stoichiometric characterization of mammalian transcription foci

Equipped with the two abovementioned capabilities, we then applied RLS-SRM to resolve the spatial organization of TMR-labeled RNAP II in mammalian nucleus beyond the diffraction limit with single-copy accuracy. Upon fixation and imaging (see Section 4.5 for details), the distribution of TMR localizations in the nucleus displays a punctate pattern with discrete foci (**Figure 4.5 A**), consistent with previous findings using immunofluorescence labeling of RNAP

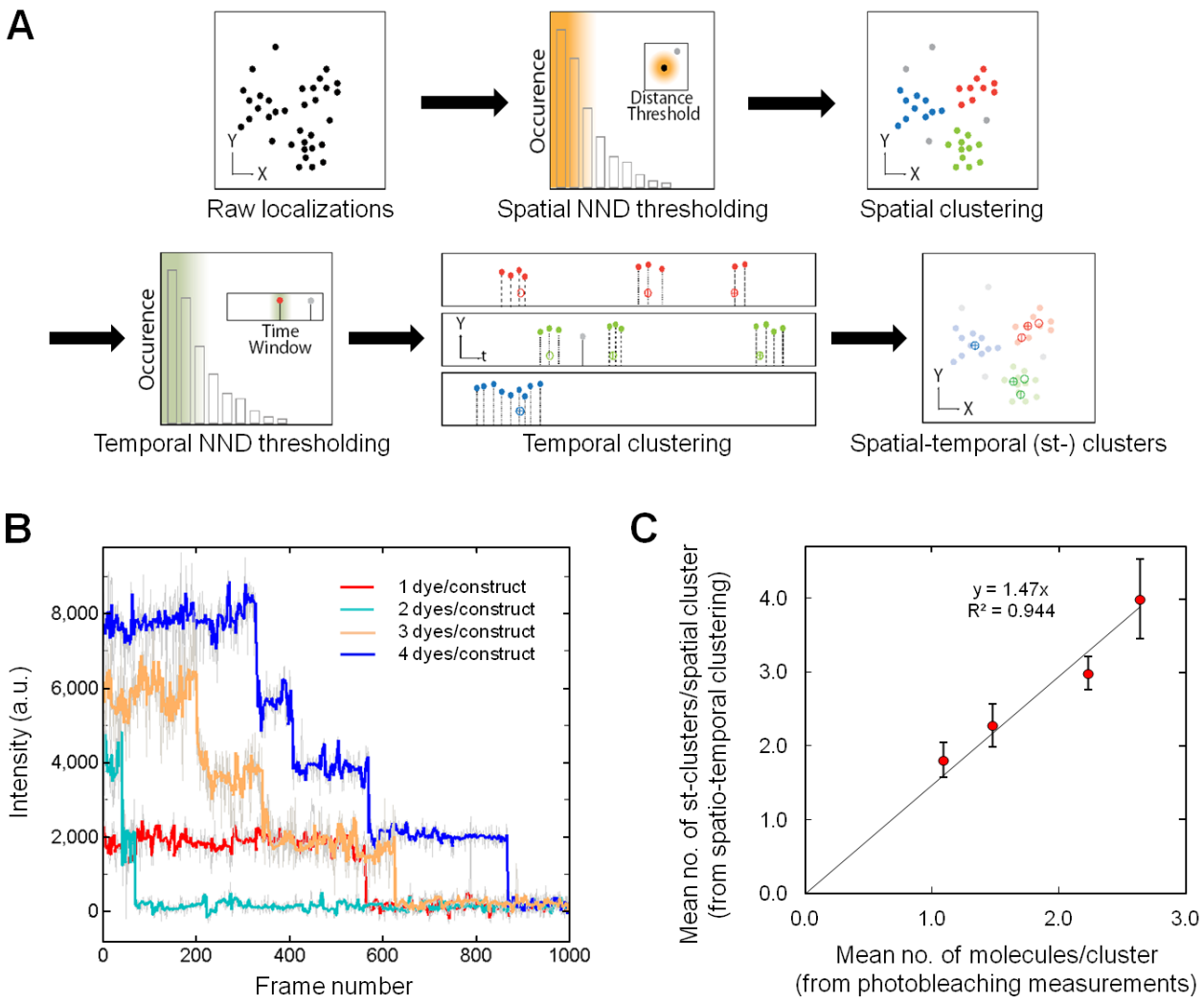

Figure 4.4. Counting molecules in RLS-SRM images based on spatio-temporal clustering analysis. (A) Flow chart of spatio-temporal clustering analysis based on spatial and temporal nearest neighbor distance (NND) thresholding of TMR localizations, leading to the identification of st-clusters. **(B)** Representative photobleaching time traces (denoised from the gray raw traces) for each of the TMR-labeled DNA constructs (mimicking transcription foci with one or multiple RNAP II molecules) used to calibrate the temporal clustering algorithm. **(C)** A plot of the average number of st-clusters per spatial cluster (as determined from spatio-temporal clustering analysis) against the average number of dyes per construct molecule (as determined by counting photobleaching steps) allows us to correlate the average number of st-clusters observed with the average copy number of RNAP II molecules in the transcription foci. Error bars denote ± s.d. (n = 3).

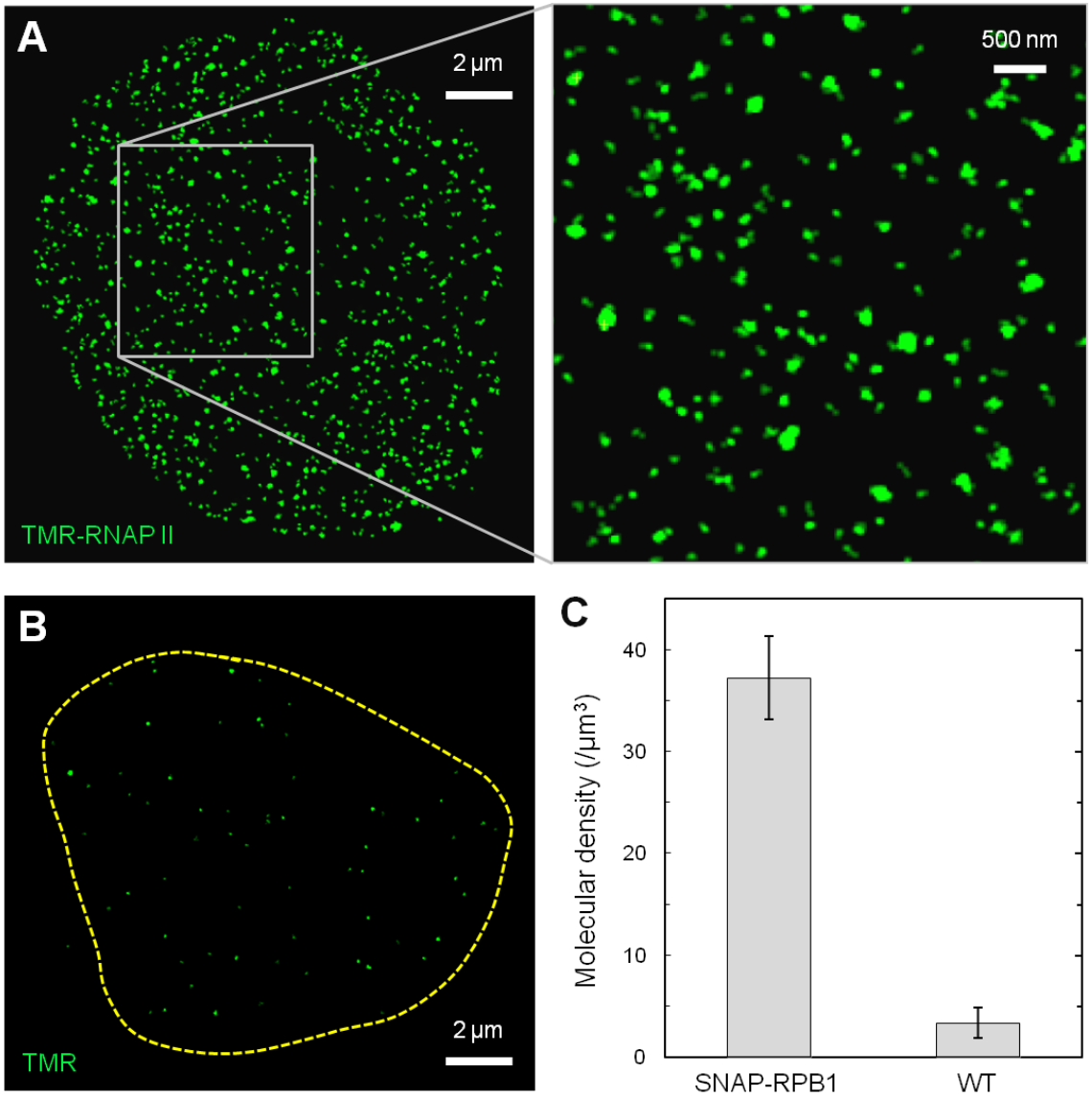

Figure 4.5. RLS-SRM imaging of singly labeled RNAP II molecules in mammalian nucleus. **(A)** Distribution of SNAP-RPB1 molecules in a thin optical section of a fixed U2-OS cell nucleus labeled with TMR. Inset shows a zoomed-in area, where single transcription foci are discernible. **(B)** Distribution of nonspecifically bound TMR molecules in the nucleus of a wild-type (WT) U2-OS cell not expressing SNAP-RPB1 that has been similarly labeled and imaged as in **(A)**. The yellow dotted line delineates the nuclear boundary. **(C)** Histogram of the density of TMR molecules detected in **(A)** and **(B)** (for SNAP-RPB1: n = 12,482 from 8 cells; for WT: n = 533 from 3 cells). The density observed in WT cells is 8.8% of that in SNAP-RPB1-expressing cells. Error bars denote ± s.d..

II molecules (Iborra et al., 1996; Jackson et al., 1993; Pombo et al., 1999; Wansink et al., 1993). As a control, when wild-type U2-OS cells not expressing SNAP-RPB1 were similarly labeled and imaged, the density of TMR molecules detected was < 9% of that in SNAP-RPB1-expressing cells, suggesting minimal nonspecific labeling by TMR (**Figure 4.5 B** and **C**).

Applying the spatio-temporal clustering analysis to the super-resolution images of TMR-labeled transcription foci (**Figure 4.6 A**), we found that each of the observed foci contains on average 2.1 ± 0.4 st-clusters which, according to our calibration, corresponds to an average of 1.4

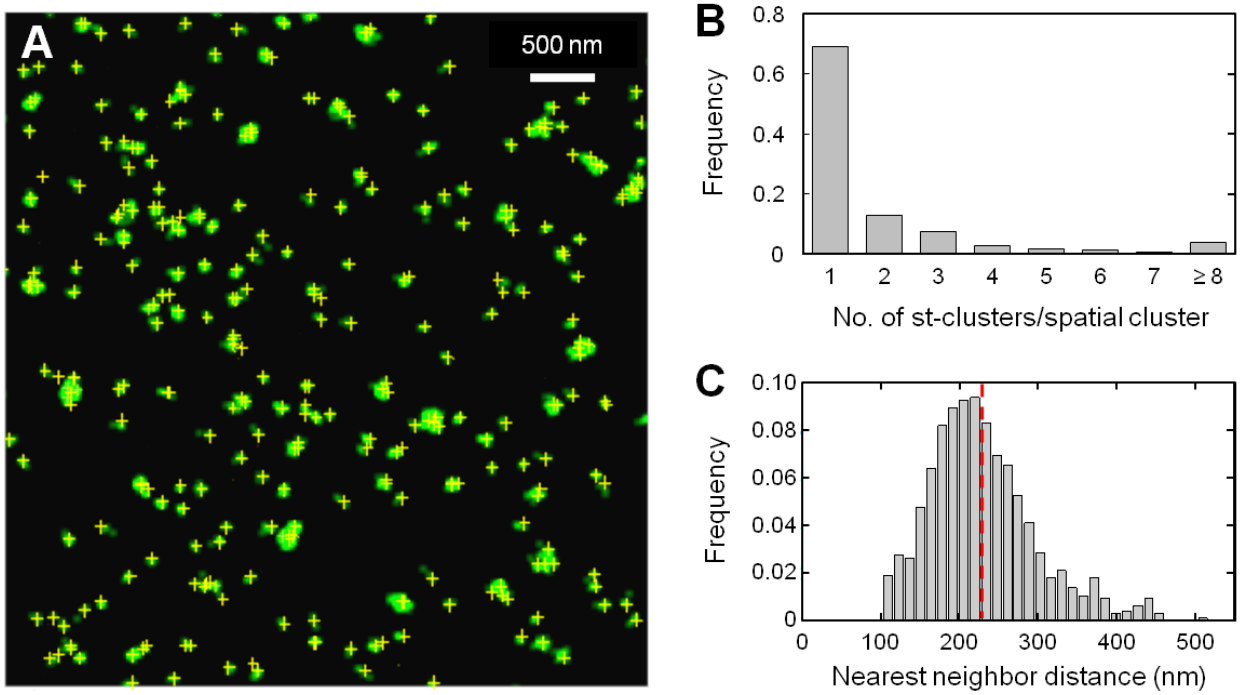

Figure 4.6. Spatial organization of RNAP II molecules shows no significant clustering. (A) The same image of transcription foci as shown in the inset of **Figure 4.5 A**, on which the centroid positions of all st-clusters identified (indicated by yellow crosses) are super-imposed. **(B)** Distribution of the number of st-clusters in transcription foci indicates that at least 70% of the foci consist of only one RNAP II molecule (n = 4465). **(C)** Distribution of spatial nearest neighbor distance between transcription foci shows that majority of the RNAP II molecules do not associate with each other within the reported diameters of "transcription factories" (40 - 130 nm). Dotted red line indicates mean.

molecules of RNAP II per focus. Approximately 70% of the foci consist of only 1 st-cluster, which most likely originate from only one RNAP II molecule, whereas the fraction with 4 or more st-clusters is minimal (< 10%, **Figure 4.6 B**). In contrast, RNA polymerase I (RNAP I) molecules has been shown to form clusters in mammalian nucleoli (Boisvert et al., 2007; Garcia-Blanco et al., 1995; Gilbert et al., 1995; Scheer and Rose, 1984; Shaw and Jordan, 1995). When the

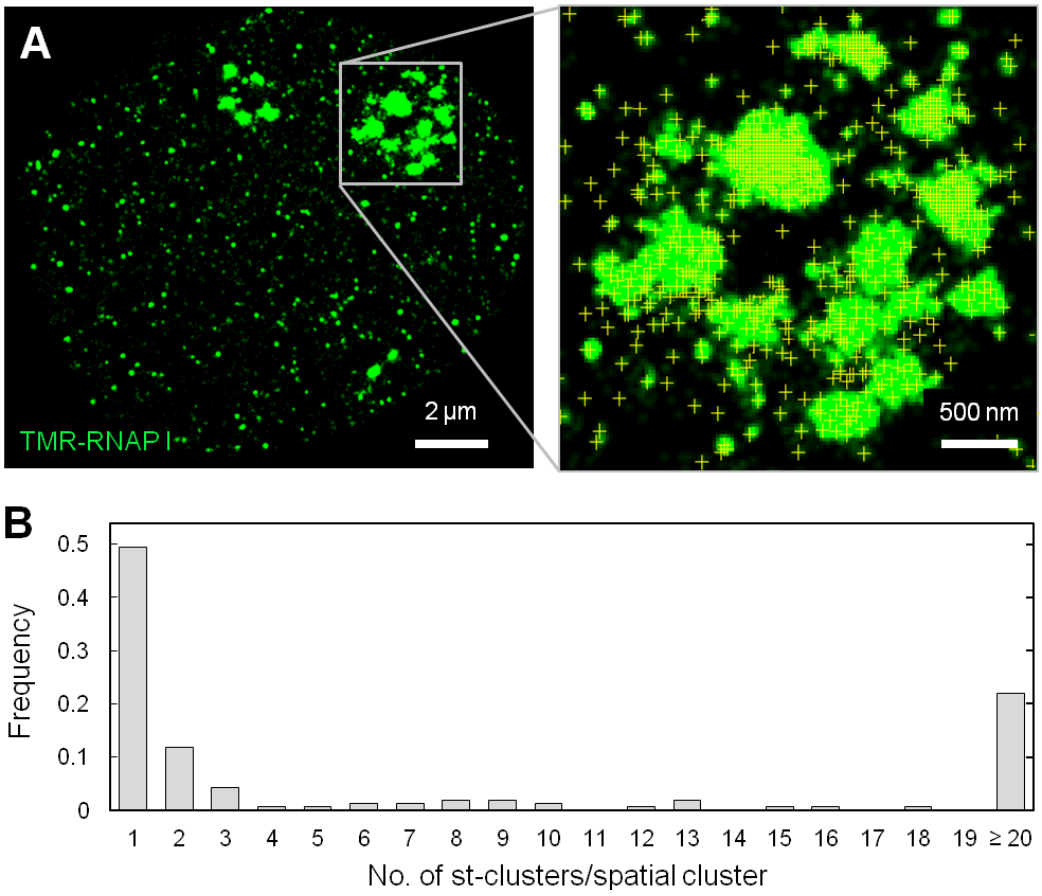

Figure 4.7. RLS-SRM imaging and spatio-temporal clustering analysis of RNAP I molecules reveal clustering in nucleoli. (A) Distribution of RNAP I molecules in an optical section of a wild-type U2-OS cell nucleus stained with a TMR-labeled antibody against human RPA194, the largest subunit of RNAP I that is not present in RNAP II. Inset shows a zoomed-in area where large clusters of RNAP I molecules are discernible; yellow crosses indicate the centroid positions of the st-clusters identified. **(B)** Histogram of the number of st-clusters in RNAP I foci, with a significant fraction (~ 22%) containing ≥ 20 st-clusters.

same analysis was applied to images of RNAP I in wild-type U2-OS cells that have been immunofluorescently labeled with TMR, a significant fraction (~ 22%) of the foci containing more than 20 st-clusters per focus was observed in the nucleolar regions (**Figure 4.7**), suggesting that our analysis is capable of detecting multi-molecular clusters when they exist. Moreover, the spatial nearest neighbor distance between the RNAP II foci was found to have a mean of 230 nm (**Figure 4.6 C**), exceeding the diameter range of "transcription factories" reported (40 - 130 nm) by a factor of 2 to 6 (Eskiw and Fraser, 2011; Eskiw et al., 2008; Iborra et al., 1996; Jackson et al., 1998; Pombo et al., 1999). These results collectively suggest that majority of the RNAP II molecules exist *in singulo* inside the mammalian nucleus.

4.3.4. Ensuring near-complete imaging of all RNAP II molecules

An obvious prerequisite for making any claim on the presence or absence of RNAP II clusters is that we can detect and image most, if not all, of the molecules illuminated by RLS-SRM. To exhaustively image all dye molecules in the optical section and thereby ensure accurate counting, we continuously acquired images until most of the dye molecules were bleached (usually after ~ 50,000 frames for each optical section), as evidenced by the plateauing of the cumulative distribution of total localizations detected (**Figure 4.8 A**). In addition to optical sectioning through RLS-SRM, we also restricted, based on fluorescence intensity and ellipticity, the final set of localizations used to reconstruct the super-resolution images to a central section of the illumination volume ~ 300 nm in thickness (FWHM), in which our confidence of near-complete imaging of all molecules present is highest (**Figure 4.8 B**).

To estimate the labeling and detection efficiency of the SNAP-RPB1 molecules under our imaging conditions, we generated a U2-OS cell line expressing a nuclear localized SNAP tag

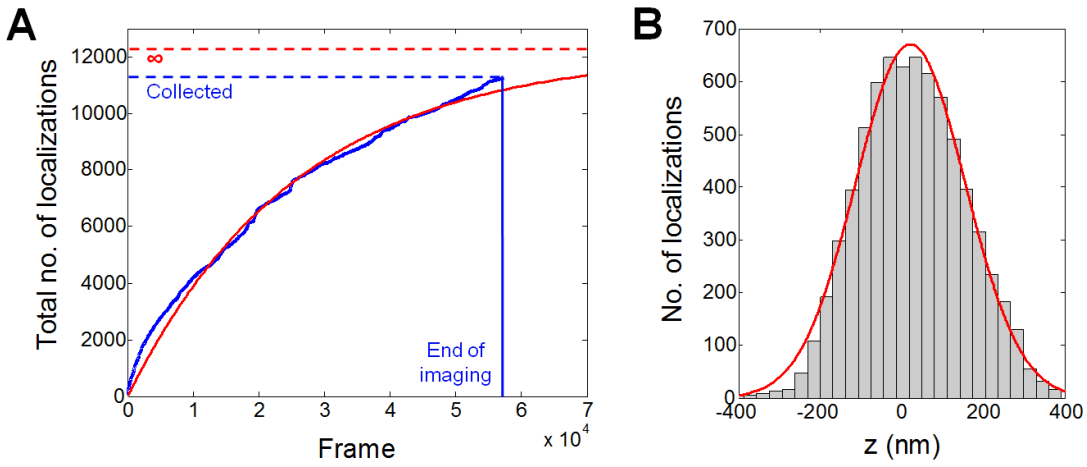

Figure 4.8. Acquisition and selection of fluorophore localizations ensure near-complete imaging of all RNAP II molecules present. (A) Cumulative distribution of the number of localizations collected (blue) over the course of imaging. Fitting the distribution to a single exponential function (red) shows that at the end of image acquisition, > 90% of all dyes in the optical section have been exhaustively imaged. **(B)** Distribution of the axial positions of localizations used to reconstruct a super-resolution image of RNAP II. Fitting the distribution with a Gaussian function (red) gives a FWHM thickness of ~ 300 nm.

fused to a Halo tag. By labeling the two tags with two spectrally distinct dyes (SNAP tag with SiR (Lukinavičius et al., 2013) and Halo tag with TMR) and performing two-color RLS-SRM imaging, we found ~ 52% of colocalization between the two tags (**Figure 4.9**). The total number of molecules detected for each color was similar, suggesting comparable labeling efficiency for both tags. From these results, we calculate an overall labeling and detection efficiency of $\sqrt{0.52} = 72\%$. Moreover, the density of RNAP II molecules observed was found to be 37.2 ± 4.1 molecules per μm^3, which translates to a total of 80,200 ± 8,800 RNAP II molecules in the entire nucleus (see Section 4.5), in good agreement with values reported in previous studies (Jackson et al., 1998; Jackson et al., 2000; Pombo et al., 1999). Therefore, we conclude that we indeed exhaustively imaged most of the labeled RNAP II molecules in the illuminated optical section.

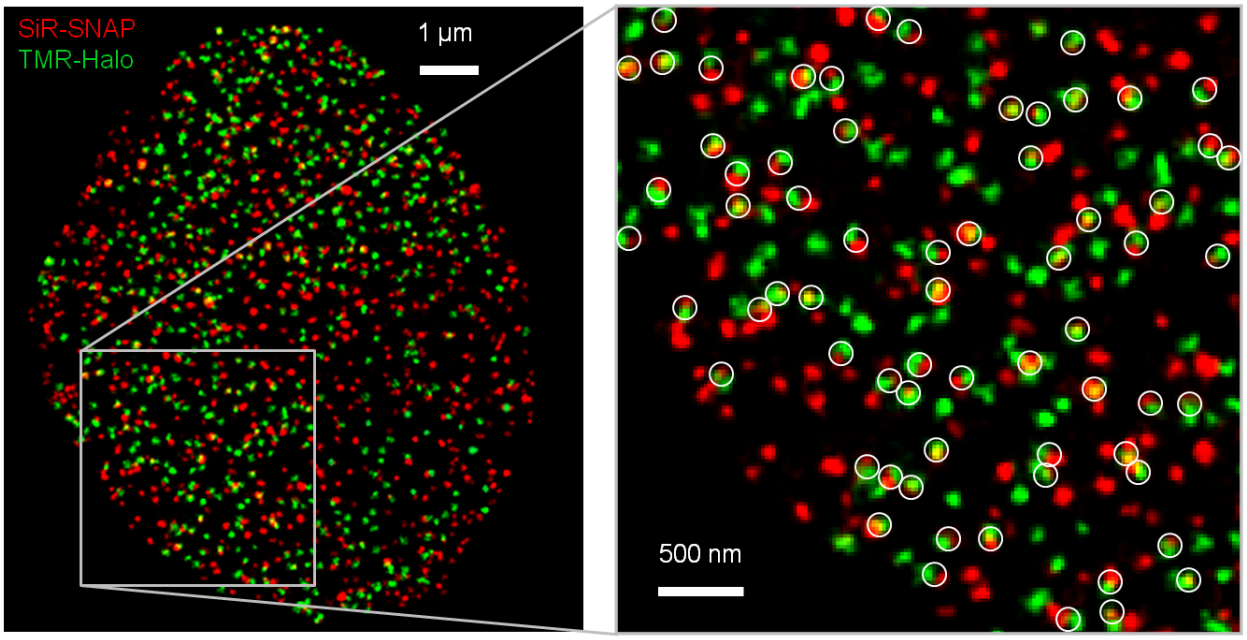

Figure 4.9. Determining labeling and detection efficiency using doubly labeled SNAP-Halo fusion protein. A two-color super-resolution image of an area of a U2-OS cell nucleus is shown, in which the SNAP tag moiety of the fusion protein was labeled with SiR (red) and the Halo tag moiety with TMR (green). Inset shows a zoomed-in area, in which ~ 52% of the two tags colocalize with one another (circled in white).

4.3.5. Quantifying global extent of RNAP II clustering with two-color colocalization

As an orthogonal approach to validate our spatio-temporal clustering results, we labeled the SNAP-RPB1 molecules in a cell nucleus simultaneously with SiR and TMR dyes; the labeling conditions were fine-tuned so that the population labeled with either dye was approximately equal (see Section 4.5 for details). Thus, if there is clustering of two or more RNAP II molecules, at least half of them will be revealed as colocalized signals between the two dyes, whereas the rest (colocalization between two or more molecules labeled with the same color) will not. Two-color RLS-SRM imaging and colocalization analysis detected 17.9% ± 1.0% of the molecules that colocalize with each other (**Figure 4.10**), thus yielding a maximum of 35.8% ± 2.0% of the

clusters that contain more than one RNAP II molecule, in good agreement with our spatio-temporal clustering result of 30% or less of RNAP II molecules existing in the clustered form.

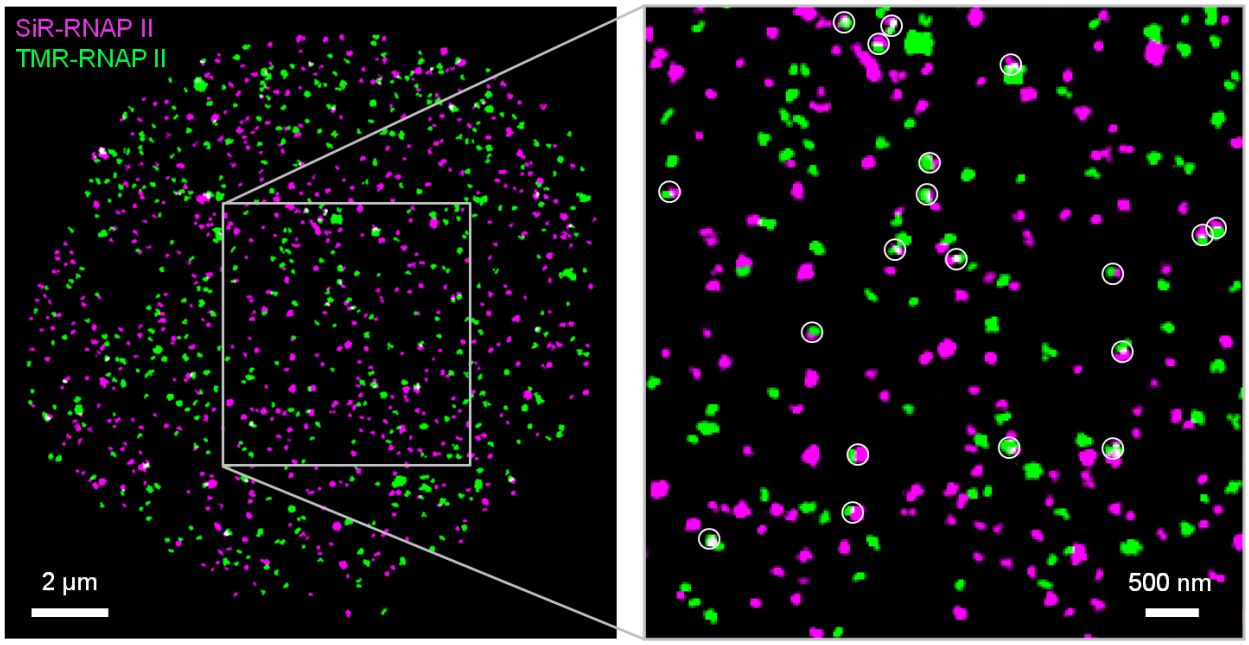

Figure 4.10. Validating RNAP II clustering with two-color colocalization. A super-resolution image of a U2-OS cell nucleus is shown, in which the SNAP-RPB1 molecules are simultaneously and equally labeled with either SiR (cyan) or TMR (green). Inset shows a zoomed-in area, in which 17.9% ± 1.0% (n = 8,929 from 6 cells) of the molecules colocalize with each other (circled in white).

4.4. Discussion and conclusion

Despite the tremendous progress made in super-resolution microscopy over the last decade, imaging inside the mammalian nucleus has until recently remained challenging. More importantly, most of the super-resolution studies performed were limited in terms of quantitative characterization of molecular copy numbers, due to difficulties stemming from non-stoichiometric labeling of biomolecules and inability to assign localization events to specific

molecules. While intracellular counting of fluorescent protein molecules based on similar principles as used in this Chapter have recently been achieved in bacteria (Lee et al., 2012) and yeast (Puchner et al., 2013) systems, no such feat has yet been accomplished in the mammalian nucleus, primarily due to the difficulty of imaging single molecules inside the nucleus with sufficient signal-to-background ratio.

By using RLS-SRM as well as implementing spatio-temporal clustering analysis on super-resolution images, we demonstrated the capability of stoichiometric counting of a key nuclear biomolecular species with single-copy accuracy. Besides the substantial background suppression, another important advantage of using RLS, as opposed to epi- or HILO illumination in conventional SRM, lies in its ability to minimize photobleaching of probes from out-of-focus "on" events before they are imaged, thus preventing undercounting of molecules in cases where the entire axial range of the cell needs to be imaged.

With the capability provided by RLS-SRM, we resolved the spatial organization of RNAP II-mediated transcription down to single-molecule level, and quantified the global extent of clustering among RNAP II molecules in the mammalian nucleus. While we found that at least 70% of transcription foci consist of only one RNAP II molecule, we should bear in mind that given the molecular density observed, a fraction of the RNAP II molecules are expected to colocalize with each other by chance. To estimate this fraction, we performed a stochastic simulation of the spatial distribution of 2500 RNAP II molecules equally labeled with two different dyes and randomly scattered over an area of 15 μm × 15 μm, mimicking the molecular density observed in U2-OS nucleus (**Figure 4.11**). At this density, 22.2% of the molecules colocalize with each other by chance, making the effective fraction of non-random transcription foci consisting of at least two RNAP II molecules to be only 12.9% ± 2.7%, or namely > 85% of

the RNAP II molecules exist *in singulo*, free of any nonrandom association with neighboring molecules. The lack of significant clustering within the length scale of the reported diameters of "transcription factories" further corroborates our claim that the mammalian nuclear distribution of RNAP II is primarily unclustered.

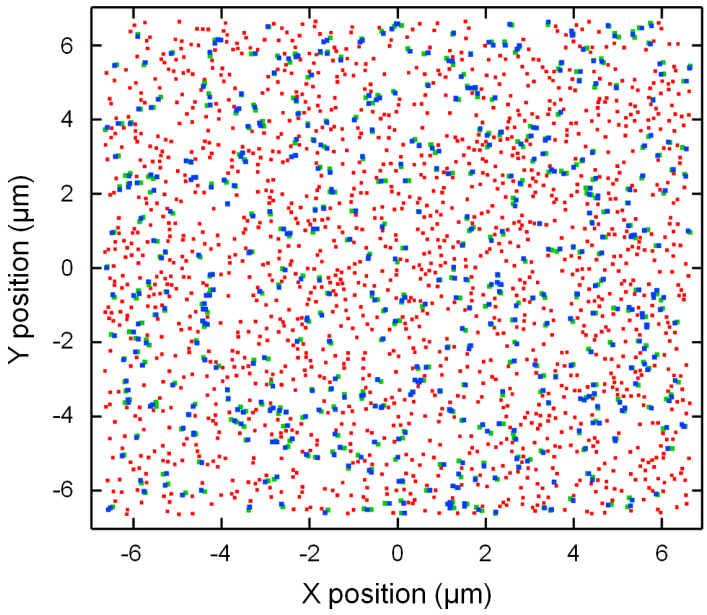

Figure 4.11. Stochastic simulation of RNAP II distribution reveals that a fraction of the molecules will colocalize by chance. At the experimentally observed molecular density, 22.2% of the molecules will colocalize by chance (blue/green) while the rest remain in singular form (red). Modeling the RNAP II distribution as a spatial Poisson distribution gives a probability of colocalization by chance of 3.95×10^{-7} per molecule per μm^2.

The "transcription factory" model posits that active genes that are located spatially apart from each other might be pulled together into preassembled "factories", each consisting of 4 - 30 molecules of RNAP II, and slide through the "factories" as they are transcribed (Chakalova and Fraser, 2010; Edelman and Fraser, 2012; Papantonis and Cook, 2010; Papantonis and Cook, 2013; Rieder et al., 2012; Sutherland and Bickmore, 2009). Such a model challenges the conventional

view that considers genomic DNA as the central scaffold around which the transcriptional machinery is assembled. The observation that the majority of RNAP II molecules exist in a solitary fashion without significant clustering among them suggests that the model of prevalently existing "transcription factories" that transcribe mammalian genes in a concerted fashion needs to be reconsidered. The fact that these "factories" were observed in previous studies could possibly be due to the lack of molecular specificity in electron microscopy and spectroscopy studies (Eskiw and Fraser, 2011; Eskiw et al., 2008; Iborra et al., 1996; Jackson et al., 1998; Pombo et al., 1999) or limited resolution in fluorescence microscopy studies (Brown et al., 2008; Jackson et al., 1993; Osborne et al., 2004; Pombo et al., 1999; Wansink et al., 1993), since insufficient resolution could potentially create artificial "transcription factories" out of images of spatially separate RNAP II molecules. To demonstrate such effect, we binned the raw RNAP II localizations from an actual RLS-SRM dataset into 10 nm bins, which are then convoluted by a Gaussian function representing a lateral FWHM resolution of ~ 30 nm. The pairwise correlation function calculated for the binned data is largely a Delta function convoluted by the resolution, as expected for randomly distributed RNAP II molecules. In contrast, when the same dataset is binned into larger bins (50 nm and 100 nm) to mimic the resolution of conventional fluorescence microscopes, the pairwise correlation function exhibits non-trivial components at larger correlation distances, suggesting artificial clustered structures as a result of insufficient resolution (**Figure 4.12**). Lastly, it should be noted that in most of the previous studies, the average number of RNAP II molecules in each "factory" was inferred (from the ratio between the total number of RNAP II molecules in the nucleus and the number of "factories" detected) rather than directly visualized (Iborra et al., 1996; Jackson et al., 1993; Jackson et al., 1998; Pombo et al., 1999), thus contributing to the large variability in its value (see **Table 4.1**).

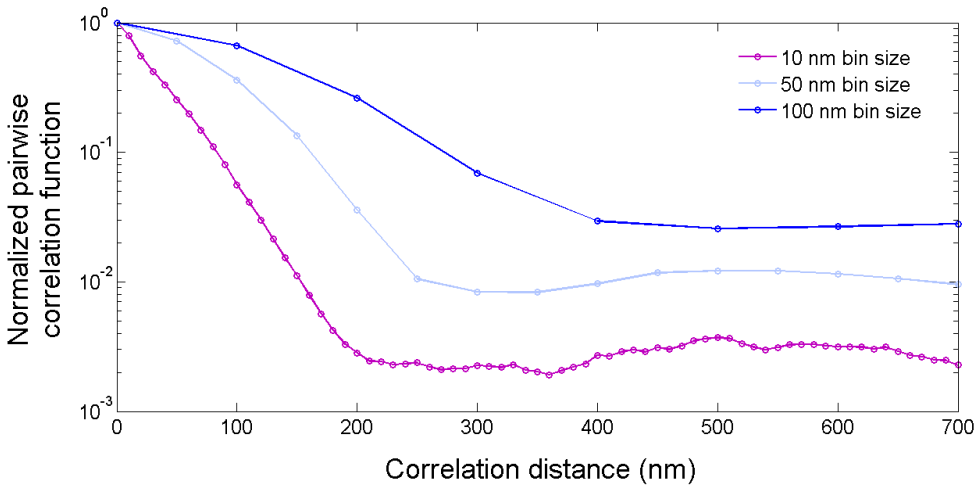

Figure 4.12. Insufficient resolution could result in artificial clustering of separate molecules. Pairwise correlation function calculated for a super-resolution dataset of RNAP II localizations binned into 10 nm bins (purple) resembles a Delta function convoluted by the resolution of RLS-SRM, characteristic of a random distribution of RNAP II molecules. However, when binned into larger bins (50 nm and 100 nm, azure and blue respectively), the pairwise correlation function exhibits non-trivial components at larger correlation distances, indicating the emergence of artificial clustered structures.

Nevertheless, a small but non-negligible fraction of the foci we observed consisted of two or more RNAP II molecules. This observation could be due to several reasons. For example, more than one RNAP II molecule might be transcribing on the same gene or on two neighboring genes that are too close to be resolved by our method. Also, a considerable fraction of human genes are organized bidirectionally, with two promoters within short distance from each other (Adachi and Lieber, 2002); in such cases, paused RNAP II molecules could be localized within distances below our resolution limit. Despite these plausible scenarios, we cannot exclude the possibility that a small fraction of RNAP II molecules do form functional clusters. In fact, our results are in line with the recent observations in live cells of transient RNAP II assembly to short-lived clusters (Cisse et al., 2013), which provides a complementary

time-averaged perspective to our ensemble-averaged view of RNAP II clustering dynamics. These two approaches, being two manifestations of the ergodic principle, present essentially the same picture: while the evanescent nature of this assembly might result in a small subset of RNAP II molecules colocalizing with each other at a given moment, on a global scale the majority of the molecules remain unclustered. These studies, together with the observation of similar "transcription factories" in bacteria (Weng and Xiao, 2014), suggest that RNAP II clustering indeed plays important roles in regulating gene expression across the species, with the difference being only the extent to which this phenomenon occurs in each organism. As such, the approaches developed in this Chapter presents quantitative imaging and analysis tools to further characterize RNAP II clustering dynamics in a variety of eukaryotic systems, as well as to correlate the spatial localization patterns with their functionality.

4.5. Materials and methods

DNA constructs and stable cell line generation and culture

The SNAP-RPB1 plasmid was generated from the YFP-aPolII plasmid (gift of X. Darzacq) by replacing the coding sequence of YFP with that of a SNAP tag from pSNAPf (New England Biolabs). Human osteosarcoma (U2-OS) cells (ATCC HTB-96) stably expressing SNAP-RPB1 were generated by transfection with Polyplus reagent and selected with 25 μg/ml α-amanitin (Sigma-Aldrich) for 2-3 weeks. The selected cells were subsequently cultured in high-glucose DMEM supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin and 2 mM GlutaMAX (all from Gibco), and in the presence of 25 μg/ml α-amanitin. The SNAP-Halo construct was generated by removing the coding sequence of Oct4-GFP from the pLV-TetO-Oct4 plasmid (gift of K. Hochedlinger) and ligating it to the coding sequences of a SNAP tag, a

nuclear localization signal, and a Halo tag in frame. A U2-OS cell line stably expressing the SNAP-Halo fusion protein was generated by lentiviral transduction as described in Chapter 3.

SNAP and Halo labeling

SNAP-RPB1 cells were passaged onto an imaging dish and allowed to settle for 3 ~ 4 hr such that they just began to attach. The culture medium was then changed to one with 1 μM of SNAP-TMR substrate (SNAP-cell TMR-Star, New England Biolabs). After incubation for 30 min at 37 °C, the cells were washed twice and incubated with fresh medium for 30 min to remove the free substrate. To enhance the impact of optical sectioning, cells were immediately fixed with 3.2% (v/v) paraformaldehyde (PFA) in PBS buffer for 15 min before their nuclei had flattened out. For two-color SNAP-RNAP II labeling, 1 μM of SNAP-SiR substrate (gift of K. Johnsson) and 0.5 μM of SNAP-TMR substrate were simultaneously added to the medium to achieve approximate 50 : 50 labeling of the SNAP molecules with either one of the two dyes. For SNAP-Halo labeling, 1 μM of SNAP-SiR substrate and 5 μM of Halo-TMR substrate (HaloTag TMR ligand, Promega) were simultaneously added to the medium.

Calibration of clustering algorithm using hybridized DNA constructs

To calibrate for the temporal clustering of TMR localizations, we designed four 20-nt DNA oligonucleotides as well as an 83-nt biotinylated backbone DNA complementary to the four oligonucleotides (with a spacer of 1 nt between each). The backbone was separately hybridized to either one or multiple (2 to 4) short oligonucleotides, each of which labeled with a single TMR dye. The hybridized DNA constructs were surface-immobilized to cleaned coverslips at low molecular densities that allow unambiguous identification of single construct molecules, and the average number of dyes per construct were determined by counting the number of

photobleaching steps for each construct molecule over the period of imaging. By comparing this value with the average number of st-clusters per construct as determined from our spatio-temporal clustering analysis of super-resolution images acquired under identical conditions as cellular imaging, we established a conversion relation for the average number of dye molecules, and thus the average copy number of RNAP II molecules, in the transcription foci.

Immunofluorescence labeling of RNAP I molecules

For the control experiment of detecting RNAP I clustering, wide-type U2-OS cells were fixed for 15 min with a solution of 3.2% (v/v) PFA and 0.1% (v/v) glutaraldehyde in PBS buffer. The sample was then treated with freshly prepared 0.1% (w/v) sodium borohydride for 7 min to quench the fluorescence background. Subsequently, the sample was incubated for 1 hr with a blocking buffer containing 5% (w/v) BSA, 0.05% (w/v) yeast tRNA, 0.05% (w/v) salmon sperm DNA, 5% (v/v) glycerol, 0.05% (w/v) sodium azide, 0.1% (v/v) Tween-20, and 0.5% (v/v) Triton X-100 in PBS buffer. The blocked sample was then stained with a mouse monoclonal antibody against human RPA194 (RPA194(C-1):sc-48385, Santa Cruz) for 3 hr, followed by a TMR-labeled goat anti-mouse secondary antibody for 45 min. The secondary antibody was labeled by reacting with TMR-NHS ester in 0.1 M sodium bicarbonate for 30 min on a rotating mixer while protected from light. The reaction mix was purified on a Sephadex G-25 medium column (GE Healthcare), and the labeling efficiency was determined spectrophotometrically. After staining, the sample was post-fixed with 0.5% PFA in PBS buffer for 5 min before imaging.

Image acquisition

Cells to be imaged were first identified and positioned under low laser intensities. The solution was then changed to an imaging buffer containing 10% glucose (m/v), 10 mM NaCl, 0.143 M β-

mercaptoethanol, 56 mg/ml glucose oxidase and 4 mg/ml catalase in 50 mM Tris (pH 8.0). Epi-illumination intensity was then increased to ~ 5 - 10 kW cm^{-2} to rapidly switch most of the dyes off before RLS-SRM imaging at a frame rate of 20 Hz and a power density of ~ 5 - 10 kW cm^{-2} along the cross-section of the light-sheet. RLS illumination was transiently alternated with epi-illumination to suppress the out-of-focus molecules to dark states. Images were continuously acquired until most of the dye molecules in the optical section were bleached (usually after ~50,000 frames), as evidenced by the plateauing of the cumulative distribution of total localizations detected within the illumination volume (**Figure 4.8 A**).

Image analysis

Image analysis was performed using custom-written software in Matlab (2010b, MathWorks). The image of each single molecule was fitted to a 2D elliptical Gaussian profile, whose centroid position and ellipticity provided the lateral and axial positions of each activated fluorescent molecule. Lateral drift between frames was corrected by tracking one or multiple fluorescent beads (TetraSpeck, Life Technologies) placed in the sample prior to imaging. Localizations were selected based on intensity and ellipticity to restrict the final set of localizations used to build the super-resolution image to a central section of the illumination volume ~ 300 nm in thickness (FWHM, **Figure 4.8 B**). Local density-based spatio-temporal clustering was then performed on these localizations to assign them to individual molecules. The raw localizations were first thresholded based on their NND in space, and grouped into spatial clusters (consisting of at least two localizations not more than three times the resolution apart from each other). The localizations in each spatial cluster were then thresholded based on their NND in time, and further grouped into spatio-temporal (st-) clusters (consisting of localizations that are less than 1.5 s (upper limit of the "on"-time for TMR) away from each other). The final super-resolution

images were reconstructed by binning the localizations (bin size equal to the localization accuracy determined separately from single dye molecules) and running a 2D Gaussian filter with a standard deviation of unit bin size.

The cross-sectional area of the nucleus imaged was determined using a custom-written algorithm, and multiplied by the thickness of the optical section to obtain the illuminated volume. The density of RNAP II molecules was calculated by dividing the total number of molecules detected by the illuminated volume in which they are contained. The volume of the nucleus was calculated assuming that a standard, unflattened U2-OS cell nucleus can be modeled by a sphere with a diameter of 15 μm (average of measurements from 8 cells), from which the fraction occupied by the nucleolar region (~ 1/27 (Iborra et al., 1996)) was excluded. The total number of RNAP II molecules in the nucleus was then estimated by multiplying the molecular density with the extra-nucleolar volume of the nucleus, adjusting for the small fraction of nonspecific binding (see **Figure 4.5 C**), and dividing by the estimated labeling and detection efficiency. Colocalization between two molecules was defined as their centroid positions being not more than 100 nm apart from each other.

References

Adachi, N., and Lieber, M.R. (2002). Bidirectional gene organization: A common architectural feature of the human genome. *Cell* **109**, 807-809.

Baddeley, D., Chagin, V.O., Schermelleh, L., Martin, S., Pombo, A., Carlton, P.M., Gahl, A., Domaing, P., Birk, U., Leonhardt, H., *et al.* (2010). Measurement of replication structures at the nanometer scale using super-resolution light microscopy. *Nucleic Acids Res.* **38**, e8.

Bohn, M., Diesinger, P., Kaufmann, R., Weiland, Y., Muller, P., Gunkel, M., von Ketteler, A., Lemmer, P., Hausmann, M., Heermann, D.W., *et al.* (2010). Localization microscopy reveals expression-dependent parameters of chromatin nanostructure. *Biophys. J.* **99**, 1358-1367.

Boisvert, F.-M., van Koningsbruggen, S., Navascues, J., and Lamond, A.I. (2007). The multifunctional nucleolus. *Nat. Rev. Mol. Cell Biol.* **8**, 574-585.

Brown, J.M., Green, J., das Neves, R.P., Wallace, H.A.C., Smith, A.J.H., Hughes, J., Gray, N., Taylor, S., Wood, W.G., Higgs, D.R., *et al.* (2008). Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *J. Cell Biol.* **182**, 1083-1097.

Bushnell, D.A., Cramer, P., and Kornberg, R.D. (2002). Structural basis of transcription: α-Amanitin-RNA polymerase II cocrystal at 2.8 Å resolution. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1218-1222.

Chakalova, L., and Fraser, P. (2010). Organization of transcription. *Cold Spring Harb. Perspect. Biol.* **2**, a000729.

Cisse, I.I., Izeddin, I., Causse, S.Z., Boudarene, L., Senecal, A., Muresan, L., Dugast-Darzacq, C., Hajj, B., Dahan, M., and Darzacq, X. (2013). Real-time dynamics of RNA polymerase II clustering in live human cells. *Science* **341**, 664-667.

Cook, P.R. (2002). Predicting three-dimensional genome structure from transcriptional activity. *Nat. Genet.* **32**, 347-352.

Cramer, P. (2004). RNA polymerase II structure: From core to functional complexes. *Curr. Opin. Genet. Dev.* **14**, 218-226.

Darzacq, X., Shav-Tal, Y., de Turris, V., Brody, Y., Shenoy, S.M., Phair, R.D., and Singer, R.H. (2007). In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **14**, 796-806.

Edelman, L.B., and Fraser, P. (2012). Transcription factories: Genetic programming in three dimensions. *Curr. Opin. Genet. Dev.* **22**, 110-114.

Eskiw, C.H., and Fraser, P. (2011). Ultrastructural study of transcription factories in mouse erythroblasts. *J. Cell Sci.* **124**, 3676-3683.

Eskiw, C.H., Rapp, A., Carter, D.R., and Cook, P.R. (2008). RNA polymerase II activity is located on the surface of protein-rich transcription factories. *J. Cell Sci.* **121**, 1999-2007.

Garcia-Blanco, M.A., Miller, D.D., and Sheetz, M.P. (1995). Nuclear spreads: I. Visualization of bipartite ribosomal RNA domains. *J. Cell Biol.* **128**, 15-27.

Gilbert, N., Lucas, L., Klein, C., Menager, M., Bonnet, N., and Ploton, D. (1995). Three-dimensional co-location of RNA polymerase I and DNA during interphase and mitosis by confocal microscopy. *J. Cell Sci.* **108**, 115-125.

Herbert, K.M., Greenleaf, W.J., and Block, S.M. (2008). Single-molecule studies of RNA polymerase: Motoring along. *Annu. Rev. Biochem.* **77**, 149-176.

Iborra, F.J., Pombo, A., Jackson, D.A., and Cook, P.R. (1996). Active RNA polymerases are localized within discrete transcription "factories" in human nuclei. *J. Cell Sci.* **109**, 1427-1436.

Jackson, D.A., Hassan, A.B., Errington, R.J., and Cook, P.R. (1993). Visualization of focal sites of transcription within human nuclei. *EMBO J.* **12**, 1059-1065.

Jackson, D.A., Iborra, F.J., Manders, E.M.M., and Cook, P.R. (1998). Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Mol. Biol. Cell* **9**, 1523-1536.

Jackson, D.A., Pombo, A., and Iborra, F.J. (2000). The balance sheet for transcription: An analysis of nuclear RNA metabolism in mammalian cells. *FASEB J.* **14**, 242-254.

Kornberg, R.D. (2007). The molecular basis of eukaryotic transcription. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12955-12961.

Lee, S.-H., Shin, J.Y., Lee, A., and Bustamante, C. (2012). Counting single photoactivatable fluorescent molecules by photoactivated localization microscopy (PALM). *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17436-17441.

Liu, X., Bushnell, D.A., and Kornberg, R.D. (2013). RNA polymerase II transcription: Structure and mechanism. *Biochim. Biophys. Acta* **1829**, 2-8.

Lukinavičius, G., Umezawa, K., Olivier, N., Honigmann, A., Yang, G., Plass, T., Mueller, V., Reymond, L., Corrêa Jr, I.R., Luo, Z.-G., *et al.* (2013). A near-infrared fluorophore for live-cell super-resolution microscopy of cellular proteins. *Nat. Chem.* **5**, 132-139.

Nguyen, V.T., Giannoni, F., Dubois, M.F., Seo, S.J., Vigneron, M., Kedinger, C., and Bensaude, O. (1996). In vivo degradation of RNA polymerase II largest subunit triggered by alpha-amanitin. *Nucleic Acids Res.* **24**, 2924-2929.

Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W., *et al.* (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.* **36**, 1065-1071.

Papantonis, A., and Cook, P.R. (2010). Genome architecture and the role of transcription. *Curr. Opin. Cell Biol.* **22**, 271-276.

Papantonis, A., and Cook, P.R. (2013). Transcription factories: Genome organization and gene regulation. *Chem. Rev.* **113**, 8683-8705.

Pombo, A., Jackson, D.A., Hollinshead, M., Wang, Z., Roeder, R.G., and Cook, P.R. (1999). Regional specialization in human nuclei: Visualization of discrete sites of transcription by RNA polymerase III. *EMBO J.* **18**, 2241-2253.

Puchner, E.M., Walter, J.M., Kasper, R., Huang, B., and Lim, W.A. (2013). Counting molecules in single organelles with superresolution microscopy allows tracking of the endosome maturation trajectory. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 16015-16020.

Rieder, D., Trajanoski, Z., and McNally, J. (2012). Transcription factories. *Frontiers Genet.* **3**, 221.

Scheer, U., and Rose, K.M. (1984). Localization of RNA polymerase I in interphase cells and mitotic chromosomes by light and electron microscopic immunocytochemistry. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1431-1435.

Sengupta, P., Jovanovic-Talisman, T., Skoko, D., Renz, M., Veatch, S.L., and Lippincott-Schwartz, J. (2011). Probing protein heterogeneity in the plasma membrane using PALM and pair correlation analysis. *Nat. Methods* **8**, 969-975.

Shaw, P.J., and Jordan, E.G. (1995). The nucleolus. *Annu. Rev. Cell Dev. Biol.* **11**, 93-121.

Sutherland, H., and Bickmore, W.A. (2009). Transcription factories: Gene expression in unions? *Nat. Rev. Genet.* **10**, 457-466.

Svetlov, V., and Nudler, E. (2013). Basic mechanism of transcription by RNA polymerase II. *Biochim. Biophys. Acta* **1829**, 20-28.

Veatch, S.L., Machta, B.B., Shelby, S.A., Chiang, E.N., Holowka, D.A., and Baird, B.A. (2012). Correlation functions quantify super-resolution images and estimate apparent clustering due to over-counting. *PLoS One* **7**, e31457.

Wansink, D.G., Schul, W., van der Kraan, I., van Steensel, B., van Driel, R., and de Jong, L. (1993). Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. *J. Cell Biol.* **122**, 283-293.

Weng, X., and Xiao, J. (2014). Spatial organization of transcription in bacterial cells. *Trends Genet.* **30**, 287-297.

Chapter 5

Spatio-temporal dynamics of mammalian DNA replication at sub-diffraction-limit resolution [1]

5.1. Abstract

The integrity and accuracy of the genomic information in the cell is maintained across generations through DNA replication. Given the large number of sites from which DNA synthesis could originate in the mammalian nucleus, the coordination of the replication process in both space and time is critical for ensuring that each and every nucleotide in the genome gets replicated once and only once during the cell cycle. However, our knowledge of the spatial organization of DNA replication as well as its temporal spreading dynamics at ultra-high resolution is lacking. Here, using RLS-SRM, we super-resolved individual replication domains (RDs) in Hela cell nucleus, and quantitatively characterized their physical morphology and distribution across different stages of the S-phase. We further probed the spatio-temporal dynamics of RDs using multi-color imaging of newly synthesized DNA in two consecutive labeling windows, and detected a reversal in the radial direction of RD spreading as replication

[1] The work described in this chapter was performed in collaboration with Qian Peter Su under the co-supervision of Prof. Yujie Sun from Peking University.

proceeds from early to late S-phase. Our results support a model in which the progression of replicons along the chromosomes drives the spreading of RDs in specific spatial patterns, and point to the role of the extent of replication completion, together with the nuclear lamina, in constraining the distribution and dynamics of DNA synthesis in the mammalian nucleus.

5.2. Introduction: Organization of mammalian genome replication in space and time

The genome of a eukaryotic cell consists of long linear double-stranded DNA molecules that are packaged into highly compact chromatin fibers. The faithful and complete duplication of the genome during S-phase is a fundamental process that every mitotic cell must go through during each cell cycle, so as to ensure that an exactly identical copy of the genetic information is passed down to the daughter cells. Through decades of work in biochemistry and genetics, we now have a comprehensive understanding of the molecular mechanisms of DNA replication as well as the components of the replication machinery (Perumal et al., 2010; van Oijen and Loparo, 2010). Moreover, we also possess substantial knowledge on the genomic organization of DNA replication in the sequence space, in which replication is initiated at discrete chromosomal loci known as replication origins, which demarcate the genome into multiple replicons (Jacob, 1993; Jacob et al., 1963). Aided by interactions with many DNA and protein elements, replication initiates by first unwinding the duplex DNA at these locations, followed by an unzipping process assisted by helicases as the DNA gets copied by DNA polymerase. Gradually, these replicons grow in size along the chromosomes, and their ultimate fusion with each other marks the completion of the replication process (Watson et al., 2007), which typically takes 8 - 10 hr (Jackson, 1995). Direct measurements on spread-out DNA fibers have shown that replication origins are spaced at roughly 100 ~ 150 kb apart from each other, thus giving a total of 30,000 - 50,000 origins across the entire genome (Jackson and Pombo, 1998; Maya-Mendoza et al., 2012).

Unlike in bacteria and lower eukaryotes, the specific sequences that define replication origins in higher organisms such as mammals are less distinctive (Robinson and Bell, 2005). As such, an alternative approach to probe these key genomic sites is to visualize them inside the nucleus via imaging. Over the past decades, a variety of fluorescence imaging approaches have been developed to probe the spatial and temporal organization of replicating DNA, both in fixed cells (Aten et al., 1992; Fox et al., 1991; Gratzner, 1982; Nakamura et al., 1986; Nakayasu and Berezney, 1989; O'Keefe et al., 1992; Sadoni et al., 1999; Visser and Aten, 1999; Visser et al., 2000) as well as in live cells (Bornfleth et al., 1999; Manders et al., 1999; Schermelleh et al., 2001). These studies have revealed DNA replication as an intricately coordinated process that follows a highly organized spatio-temporal pattern (Aladjem, 2007). While on the whole replication origins tend to be fired stochastically in time throughout the S-phase (Chagin et al., 2010; Dijkwel et al., 2002), the spatial distribution of replication inside the mammalian nucleus exhibits a punctate pattern that changes dynamically over the course of S-phase (Hozák et al., 1993; Rhind and Gilbert, 2013). These discrete puncta, termed replication foci or factories, are believed to contain multiple replicons that are fired in concert.

More recently, advances in high-throughput sequencing techniques have shed important light on the asynchronous nature of replication origin activation on a genome-wide scale. Studies on replication timing have shown that many replication origins often fire in a temporally specific manner, in that certain regions of the genome tend to always replicate early while other regions always replicate late in S-phase (Pope and Gilbert, 2013; Rhind and Gilbert, 2013; Smith and Aladjem, 2014). These temporally distinct chromosomal regions in such replication timing profiles are termed replication domains (Ryba et al., 2010), which could be considered as the counterpart of replication foci in sequence space. Moreover, it was found that

actively transcribed euchromatin tends to replicate early, whereas the transcriptionally silent heterochromatin tends to replicate in late S-phase (Chagin et al., 2010). A recent Hi-C study has further shown that replication loci that are brought into close proximity to each other via higher-order chromatin structure could share similar replication timing (Ay et al., 2014).

While these studies have illuminated important details about the coordination of replication, most of them were limited in the spatial resolution they provide. Although super-resolution imaging of DNA in the mammalian nucleus has recently been demonstrated (Raulf et al., 2014), to date no study has been performed to examine mammalian DNA replication at resolutions beyond the diffraction limit, primarily due to the challenges delineated in Chapter 1. We therefore seek to reveal in this chapter, using RLS-SRM, the spatio-temporal organization and dynamics of replication in mammalian cell nucleus at sub-diffraction-limit resolution. Without preference for any particular nomenclature intended, we shall use the term replication domains (RDs) to denote the distinct foci of DNA synthesis sites observed (either in Cartesian space or in sequence space) across different stages of S-phase.

5.3. Results

5.3.1. Labeling and RLS-SRM imaging of DNA replication across S-phase

To label newly replicated DNA during specific stages of the S-phase, we first synchronized a population of Hela cells to the G1/S boundary over two contiguous cell cycles using successive treatment of the drugs thymidine and aphidicolin (Harper, 2005; Jackson, 1995; Ma and Poon, 2011), and then released them into S-phase to commence replication (**Figure 5.1 A**, and see Section 5.5 for details). Single-cell measurements on DAPI-labeled cells using flow cytometry indicates that majority (~ 70%) of the cells have been arrested in the S-phase 30 min after release

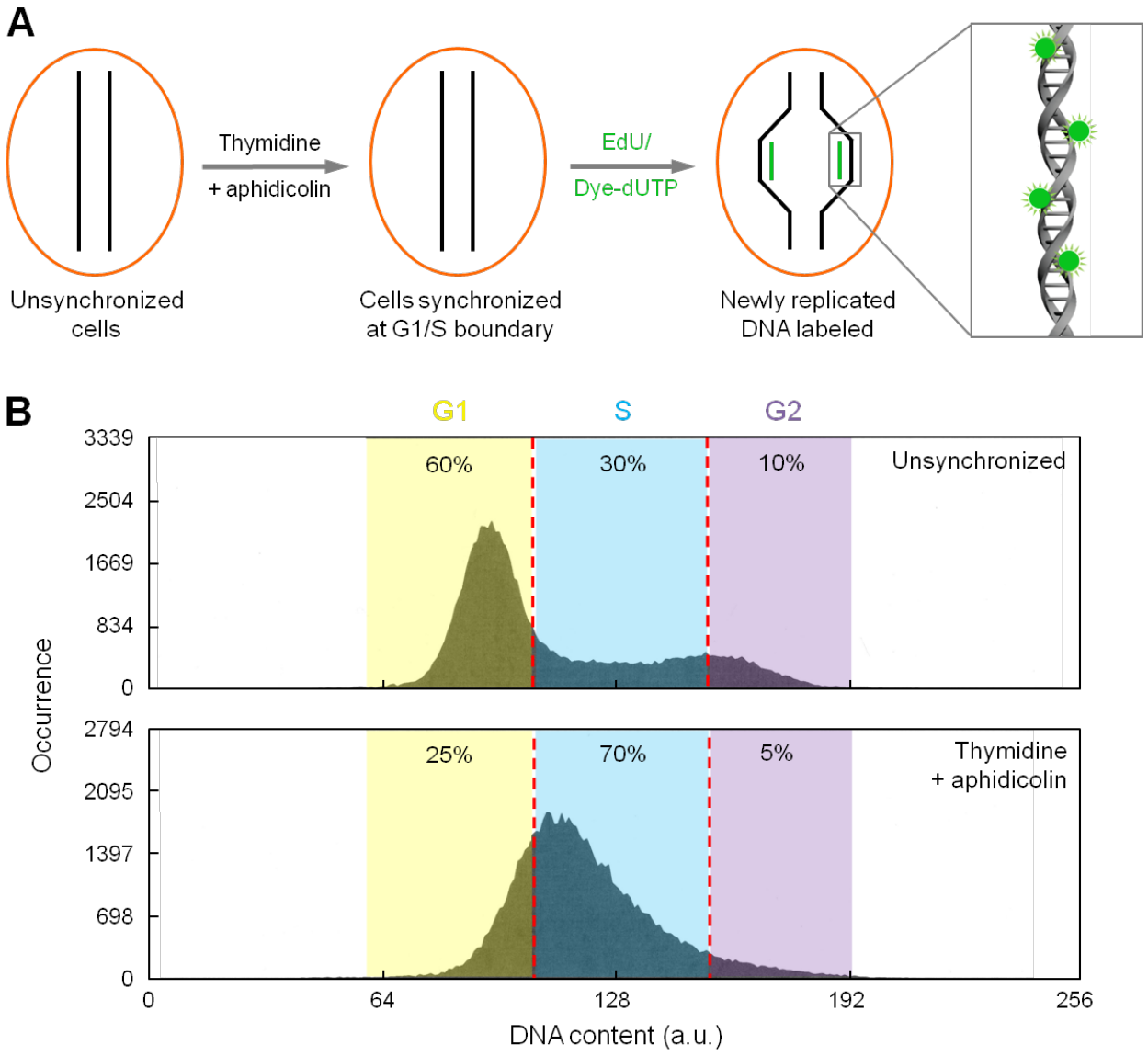

Figure 5.1. Cell cycle synchronization and fluorescent labeling of newly replicated DNA. (A) Experimental scheme in which unsynchronized mammalian cells are first synchronized to the G1/S boundary with thymidine/aphidicolin treatment; upon release into S-phase, the DNA replicated during specific stages are then fluorescently labeled using short pulses of modified thymidine analogues (shown in green in inset). **(B)** Histograms of DNA content of single cells (as a measure of cell cycle progression) in an unsychronized population (top) as well as a synchronized population 30 min after release from drug treatment (bottom). Numbers at the top of each panel denote the approximate fractions of the cell population in each stage of the cell cycle: G1 (yellow), S (blue), and G2 (purple). (Inset of **(A)** adapted, with modifications, from https://tools.lifetechnologies.com/content/sfs/manuals/mp10637.pdf).

from drug treatment (**Figure 5.1 B**). At different stages during the S-phase, a short 30-min pulse of dye-labeled dUTP or EdU (both thymidine analogues) was then supplied to the cells (see Section 5.5), which incorporated them into the new DNA synthesized during the pulse period (Dimitrova, 2009).

Upon labeling and fixation, the cells were imaged with RLS-SRM as described in Chapter 4. In order to accurately quantify the replication sites revealed in super-resolution images, we adopted a hierarchical thresholding strategy to rigorously define the ultra-structural features resolved (**Figure 5.2 A**). After the centroid positions of each blinking event above an intensity threshold, *i.e.* the localizations, are identified, the raw image is de-noised and thresholded according to spatial nearest neighbor distance to group the localizations into clusters, which denote structures with ≥ 3 localizations within a volume with lateral dimensions of 60 nm and axial dimension of 120 nm (*i.e.* three times the resolution of the RLS-SRM setup). Each cluster of localizations most likely represents one dye molecule that has blinked multiple times during image acquisition. These clusters are further grouped into domains, which are higher-order structures consisting of ≥ 3 clusters whose centroid positions are within 100 nm apart from each other. This distance threshold was set to provide adequate capability for resolving individual domains and yet avoid segregating each domain into artificial sub-structures (**Figure 5.2 B**).

5.3.2. Quantitative characterization of replication domains in mammalian nucleus

Using RLS-SRM and the abovementioned thresholding analysis, we resolved in Hela cell nucleus individual domains of DNA replication, whose morphology and distribution changes significantly across the S-phase, as measured at five distinct stages (**Figure 5.3 A**). At the beginning of S-phase, a relatively low number of small RDs are sparsely scattered throughout

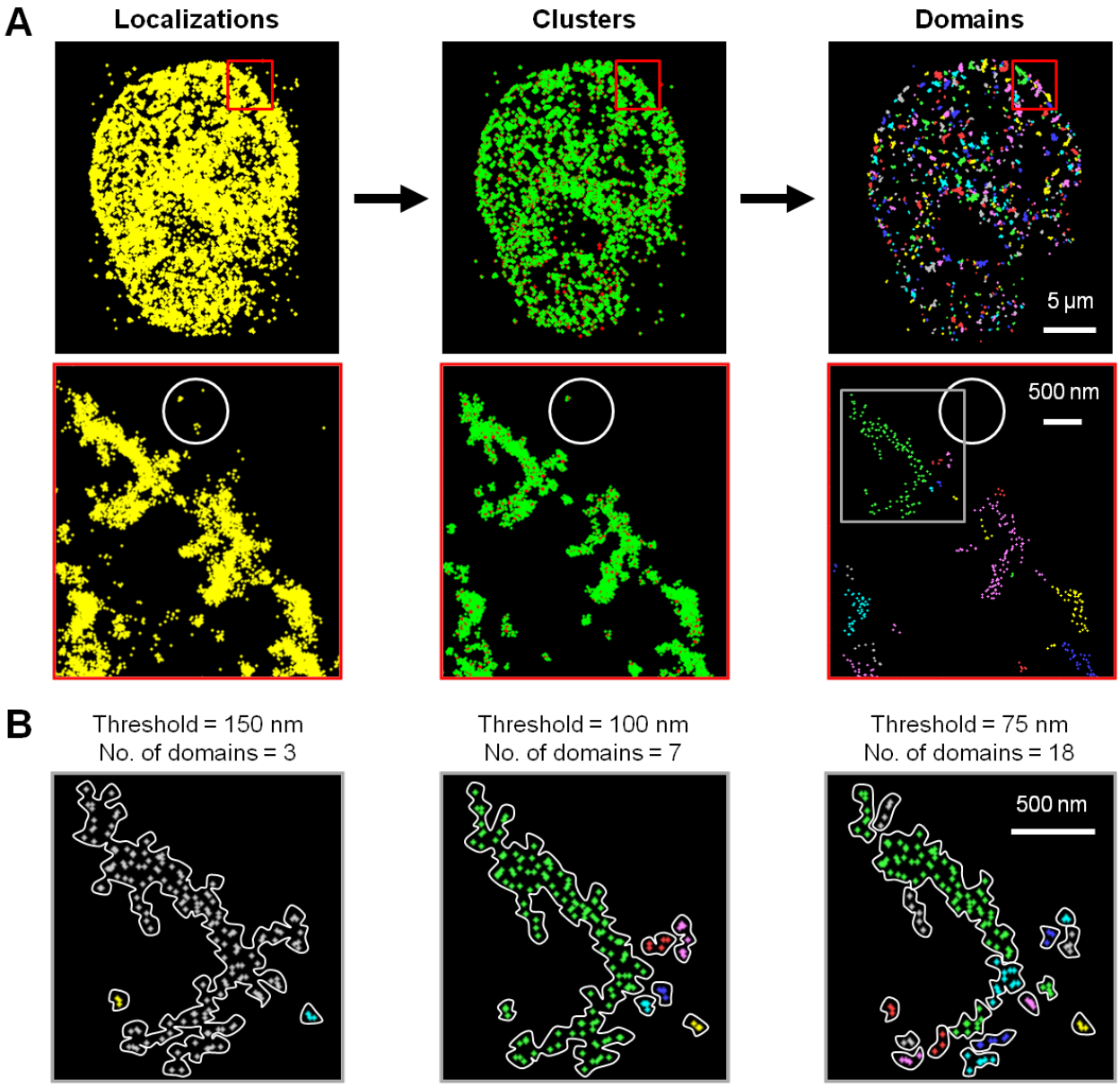

Figure 5.2. Definitions used in quantifying super-resolution images of DNA replication. (A) A hierarchical algorithm is used to identify raw localizations (left, shown in yellow), clusters (middle, with the localizations that constitute the clusters shown in green and their centroid positions shown in red), and domains (right, with each color of the rainbow denoting one domain). Insets show a zoomed-in area. An illustration of the thresholding procedure is shown in the white circle in each inset, in which several sparse localizations are filtered out when identifying clusters, which are in turn filtered out when identifying domains because they are too far apart from each other. **(B)** Effect of distance threshold on domain identification. The domains identified from an area of the inset in **(A)** (grey box) are shown as a function of the inter-cluster distance threshold. A threshold of 100 nm is used for all subsequent analyses.

the nucleus. Their abundance as well as size increases as the cell progresses through the S-phase. At mid S-phase, the RDs begin to congregate around the nuclear membrane and nucleoli. Such congregation persists to a certain extent till the end of S-phase, when only a handful of very large domains remain. Such punctate patterning and morphological changes are in qualitative agreement with previous observations using conventional optical microscopy as well as genome-wide mapping techniques (Chagin et al., 2010; Peric-Hupkes and van Steensel, 2010).

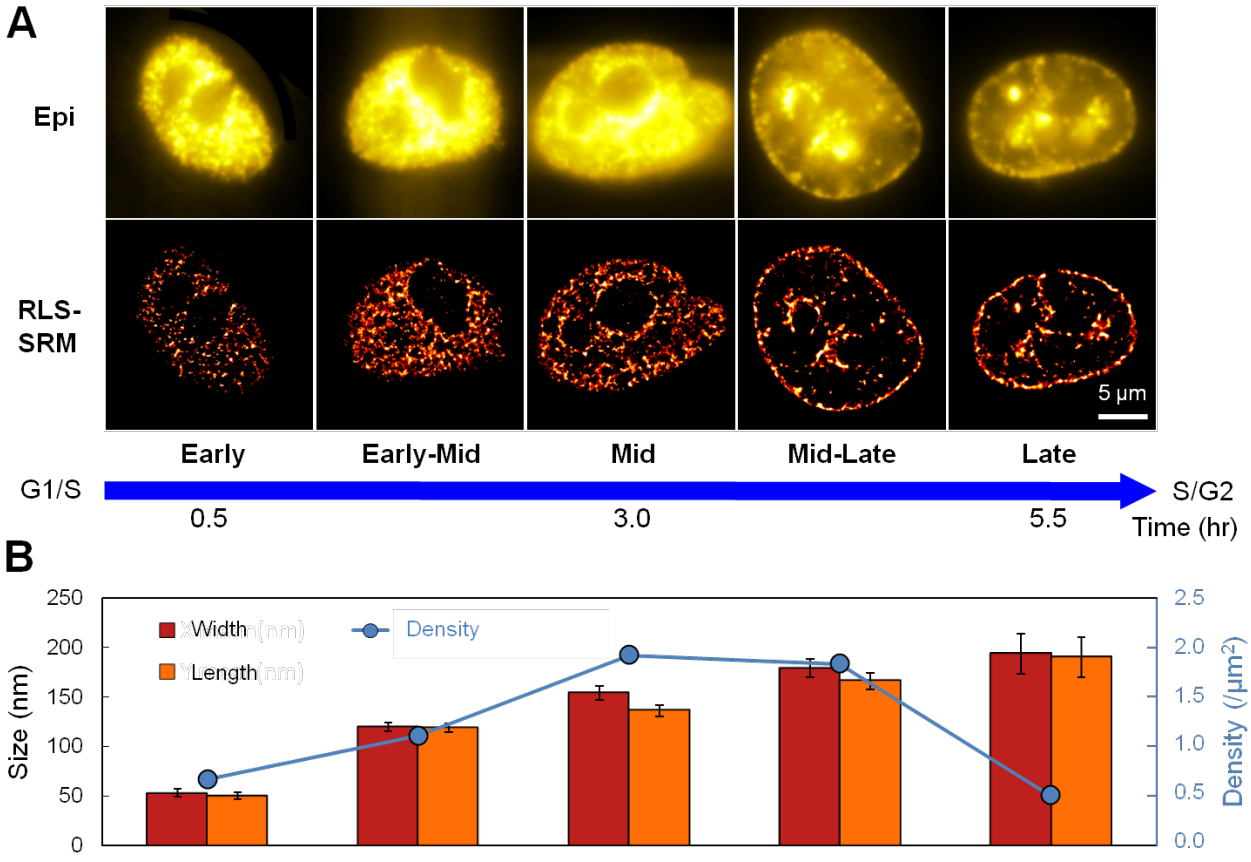

Figure 5.3. RLS-SRM imaging and quantitative characterization of replication domains across S-phase. (A) Corresponding epi (top) and RLS-SRM (bottom) images of replication sites in single Hela cell nuclei at five different time points across the S-phase (blue arrow), corresponding to early, early-mid, mid, mid-late, and late stage of DNA replication, respectively. **(B)** Histograms of the size (both width and length) of replication domains (red and orange bars, respectively) as well as of their density (blue circles) during each of the five stages imaged in **(A)**. Error bars denote ± s.e.m..

The physical characteristics of RDs are more quantitatively measured in **Figure 5.3 B**, which shows that the mean lateral dimension (both length and width) of the domains steadily increases from ~ 50 nm in early S-phase to ~ 200 nm in late S-phase, while their density rises from ~ 0.7/μm^2 in early S-phase to ~ 2.0/μm^2 in mid S-phase, and drops back to ~ 0.5/μm^2 towards the end of S-phase. There is, however, no substantial difference between the width and length of these domains (where length is measured along the axis in which the spatial span of a RD is largest, and width is measured along the axis orthogonal to length), in accordance with our expectation that replication should not adopt any preferred directionality in Cartesian space, other than radially inward or outward from their points of origin.

Moreover, the distribution of domain area indicates that while during early S-phase most of the domains remain small (tens of nm^2 in area), a significant fraction of the domains that remain in late S-phase have expanded in size by more than an order of magnitude (**Figure 5.4 A**). This observation leads us to wonder whether such an increase in domain size is due to multiple replicons coming together spatially or to each domain replicating a longer stretch of DNA. To distinguish between these two scenarios, we measured the density of localization clusters that constitute each RD in super-resolution images (both per unit area as well as per domain). We found that while their density per unit area remains more or less constant across the S-phase, the mean number of clusters per domain, which is indicative of the length of replicated DNA that each domain contains, increases from early to late S-phase by as much as 4-fold (**Figure 5.4 B**). Given that the spatial congregation of multiple replicons will lead to an increase in RD density per unit area, our findings suggest that the enlargement of replication domains as the cell progresses through the S-phase is a consequence of longer DNA sequences being replicated per domain.

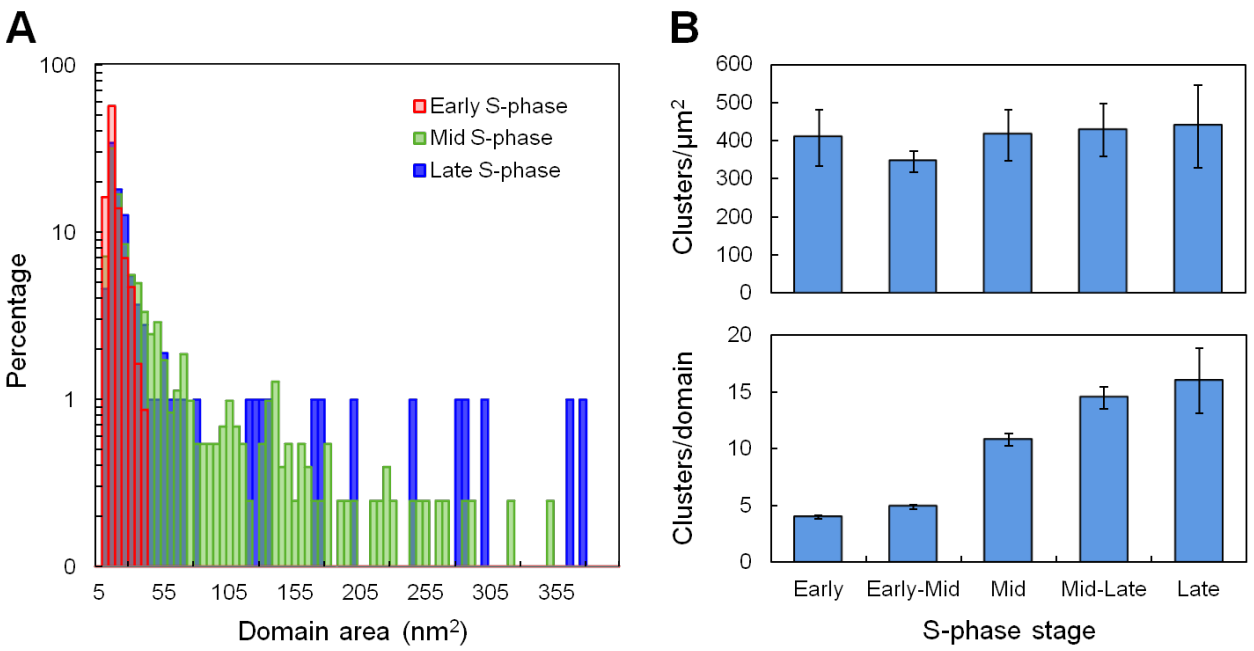

Figure 5.4. Increase in domain size is a consequence of each RD replicating longer sequences of DNA. (A) Histogram of domain area in early, mid, and late S-phase, exhibiting a distinctively higher fraction of large domains as the cell progresses through the S-phase. **(B)** Histograms of the density of localization clusters constituting the replication domains at each stage of S-phase, measured in per µm² (top) as well as per domain (bottom). Error bars denote ± s.e.m..

Lastly, from these five snapshots, a total of 1695 replication domains are observed during the five 30-min labeling windows in these cell nuclei. Scaling for the mean volume of the nucleus (~ 11 times the imaging volume) as well as the duration of S-phase (~ 8 hr), the total number of replication domains produced during S-phase in an entire nucleus stands at ~ 59,700. This number, in light of the fact that a Hela cell has 76 - 80 chromosomes instead of the 46 in normal human cells (Macville et al., 1999), agrees well with previous estimate of 30,000 - 50,000 active replication origins in a normal human nucleus during each cell cycle (Maya-Mendoza et al., 2012; Méchali, 2010).

5.3.3. Spatial-temporal dynamics of replication domains spreading

The above results provide a static picture of the distribution of mammalian RDs. In order to further probe the dynamic spreading patterns of these domains in space and time, we labeled the newly replicated DNA during two consecutive 30-min time windows at three key stages of the S-phase using two spectrally distinct dyes, Alexa 647 and Atto 550 (**Figure 5.5**). In addition,

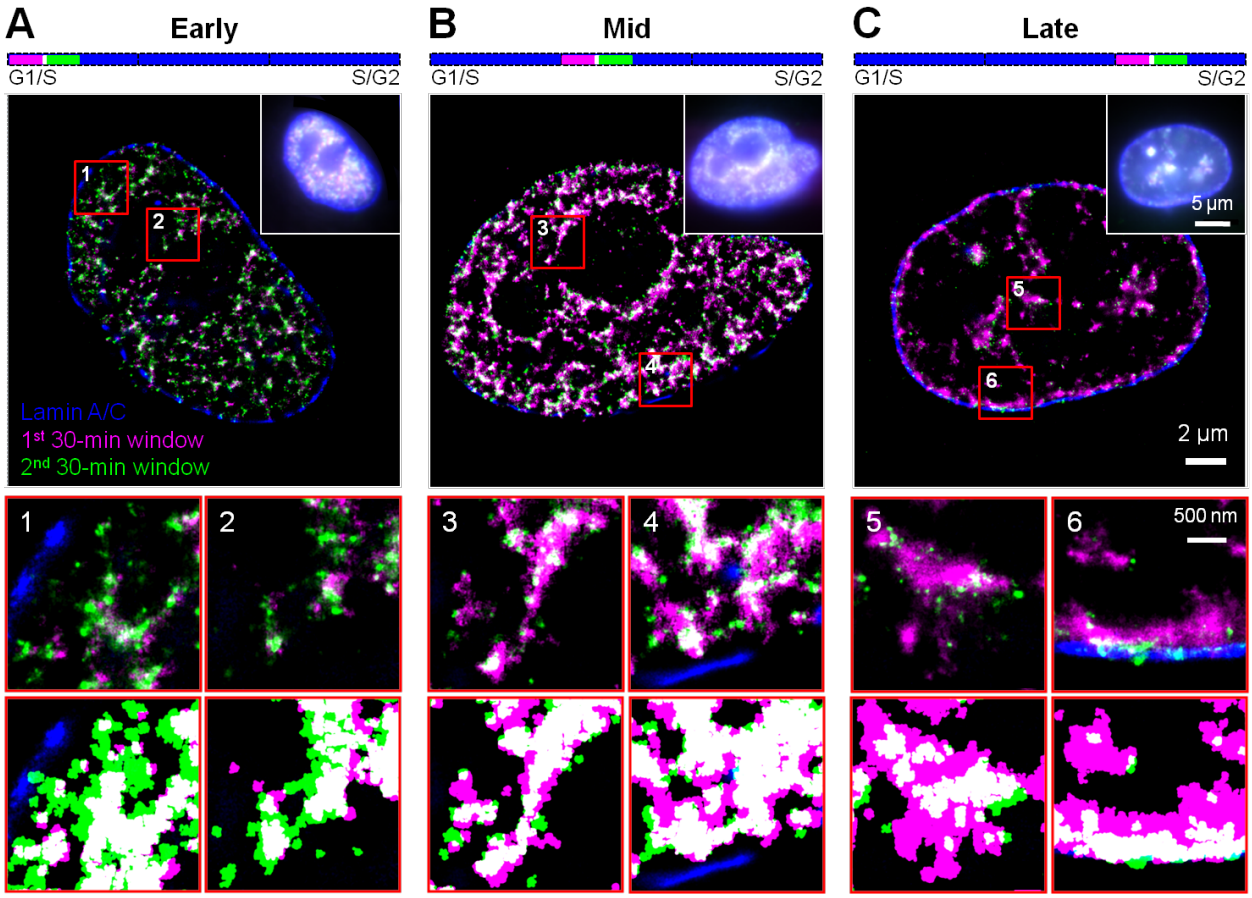

Figure 5.5. Spatio-temporal dynamics of RD spreading across S-phase. Newly replicated DNA in two consecutive windows at the beginning of early **(A)**, mid **(B)**, and late **(C)** S-phase were labeled (pink and green), together with lamin A/C (blue) to demarcate the nuclear lamina. Multi-color RLS-SRM images of single labeled nuclei are shown in comparison with epi images of the same nuclei (top right corner), and bars at the top indicate the relative temporal position of the labeling windows during the respective stages of S-phase (not drawn to scale). Insets (numbered 1 through 6) show zoomed-in areas, with the top row showing standard contrast and the bottom row showing enhanced contrast between the two colors.

lamin A/C in the nuclear lamina (NL) was immunofluorescently labeled with a third dye, Atto 488. Upon multi-color RLS-SRM imaging of the labeled cells, the relative spatial spread of the replication domains during the two time windows were correlated to reveal a unique pattern in the spatial progression of RDs. In early S-phase, the DNA synthesized during the second window is spatially more spread out and envelopes around the DNA synthesized during the first window, a trend that can be more conspicuously visualized by enhancing the contrast between the two colors. In other words, the direction in which RDs spread is radially outward from their points of origin. Such pattern is diminished in mid S-phase, and a reversal in spreading direction was observed during late S-phase, as the DNA synthesized during the first window now envelopes around that synthesized during the second window, indicating radially inward spreading.

Such spreading dynamics can be quantitatively validated by measuring the density of localization clusters that constitute the RDs in each of the two time windows (**Figure 5.6 A**). During early S-phase, the fraction of domains containing a larger number of clusters per domain (which is indicative of the length of replicated DNA it contains) is higher in the second window. Such trend in distribution is reversed in late S-phase, in which the domains in the first window tend to contain more clusters per domain. In addition, during early S-phase, as much as 94.4% of the RDs in the first time window colocalize with those in the second, as compared to only 76.5% in the second window that colocalize with those in the first (**Figure 5.6 B**), indicating a larger spatial span for the domains in the second window. However, when late S-phase is reached, the spatial span of the RDs in the second window retreats to within those in the first window, as evidenced by the fact that 94.8% of the domains in the second window colocalize with those in the first, but only 64.7% in the first window colocalize with those in the second.

Lastly, a similar reversal in spreading direction was observed when the order of the two dyes was swapped (**Figure 5.7**), thereby eliminating the possibility that such a trend could be the consequence of labeling and detection artifacts associated with the specific dyes used.

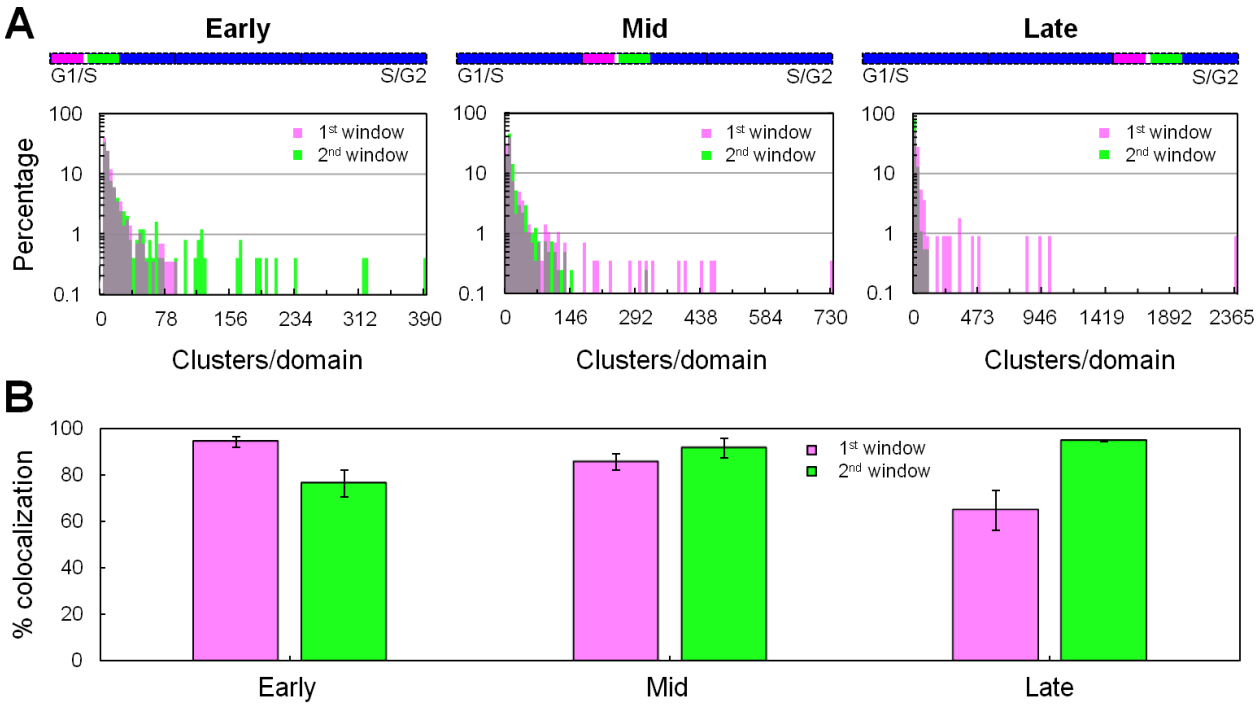

Figure 5.6. Quantification of cluster density and spatial colocalization between domains in consecutive time windows. (A) Histograms of localization density per domain in each labeling window at early (left), mid (middle), and late (right) S-phase. Bars at the top indicate the relative temporal position of the labeling windows during the respective stages of S-phase (not drawn to scale). **(B)** Histogram showing the fractions of domains in the first window that colocalize with those in the second window (pink), and vice versa (green), during each of the three stages in **(A)**. Error bars denote ± s.e.m..

5.4. Discussion and conclusion

The coordination of DNA replication is critical for the complete and faithful duplication of genomic information, and thus needs to be controlled precisely in both space and time. While we now have a detailed understanding of the organization of the non-dividing genome during

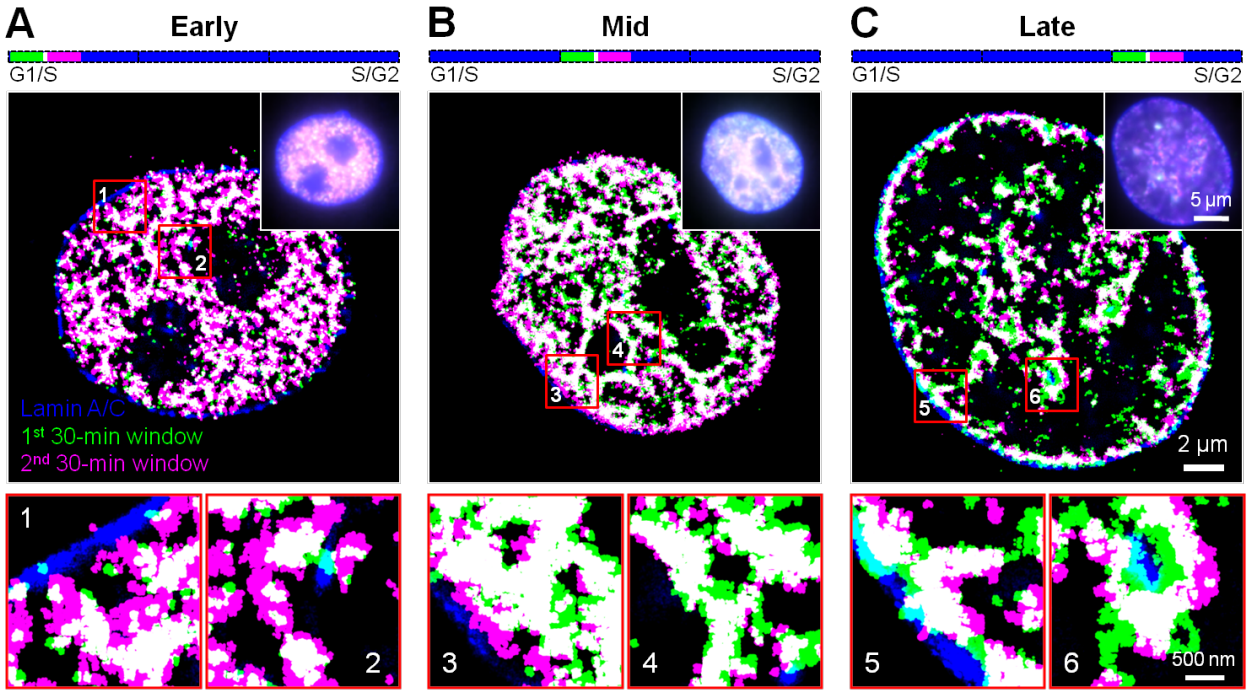

Figure 5.7. Reverse labeling confirms the spatial pattern of RD spreading. Newly replicated DNA in two consecutive windows at the beginning of early **(A)**, mid **(B)**, and late **(C)** S-phase were labeled (green and pink), together with lamin A/C (blue) to demarcate the nuclear lamina. Multi-color RLS-SRM images of single labeled nuclei are shown in comparison with epi images of the same nuclei (top right corner), and bars at the top indicate the relative temporal position of the labeling windows during the respective stages of S-phase (not drawn to scale). Insets (numbered 1 through 6) show zoomed-in areas with enhanced contrast between the two colors, demonstrating the same spreading pattern as in **Figure 5.5,** in which the two dyes used for labeling were added in the opposite order.

interphase, both in Cartesian space (in which various chromosomes are compartmentalized into discrete "territories" (Cremer and Cremer, 2010)) as well as in sequence space (in which large, megabase-sized local interaction domains termed "topological domains" exist within chromosomes (Dixon et al., 2012)), our knowledge of genome organization during replication is more limited, especially at ultra-high spatial resolution. By applying RLS-SRM, we performed super-resolution mapping of mammalian DNA replication sites on a global scale. Taking advantage of the superior SBR of our technique, we super-resolved single replication domains

within the highly dense structures of replicated DNA. These domains, as pervasive and compartmentalized structural units of the genome in Cartesian space that are replicated in a temporally specific manner, resemble the replication domains in sequence space revealed by replication timing profiles (Pope and Gilbert, 2013). However, given the fact that the total number of domains we detected is on par with the estimated number of mammalian replicons, whereas the domains in replication timing profiles are larger (~ 1 Mb) chromosomal regions believed to consist of multiple replicons, the structures we observed could also represent individual replicons fired during each labeling duration, which have thus far been difficult to detect due to insufficient resolution of the imaging techniques used.

Moreover, we also quantitatively characterized the physical morphology and distribution of these replication domains throughout S-phase, revealing distinct organizational features. Firstly, we observed substantial aggregation of RDs around the nuclear membrane and nucleoli during mid and late S-phase, in agreement with previous findings that peripheral and NL-associated DNA tend to replicate later in S-phase (Hozák et al., 1993; Peric-Hupkes and van Steensel, 2010). NL is also known to be involved in the assembly of pre-replication complex in replication foci (Hutchison et al., 1994; Wilson and Coverley, 2013), in line with its roles in organizing chromatin loops and anchoring chromatin to the nuclear envelope. We also quantified a steady increase in RD size across S-phase, which agrees well with measurements from genome-wide replication timing analysis (Ryba et al., 2011; Ryba et al., 2010). We found that such increase could be the result of each RD replicating longer DNA sequences, which could in turn be attributed to the fewer total number of replication sites towards late S-phase.

Our sub-diffraction-limit observations and quantifications suggest a model for the spatial-temporal morphological and spreading dynamics of mammalian replication domains

that hinges on the involvement of both the NL and the extent of replication progression (**Figure 5.8**). During early S-phase, RDs originate sporadically in the nucleoplasm; their small size is reflective of the sizes of early replicons, and they tend to propagate radially outward as they spread into the unreplicated chromosomal regions around each replication origin. As the cell progresses through mid S-phase, RDs increase in size as well as in number, and tend to associate with the NL and nucleoli, which provide both structural anchors for the replicating chromatin as well as spatial constraints for their propagation. Toward the end of S-phase, the replicons (and thus the RDs) become much larger in size but fewer in number. Since the amount of unreplicated DNA is diminishing and adjacent replicons are rapidly fusing with one another, the direction for the propagation of the remaining RDs would therefore be towards these fusion loci, resulting in an appearance of radially inward spreading.

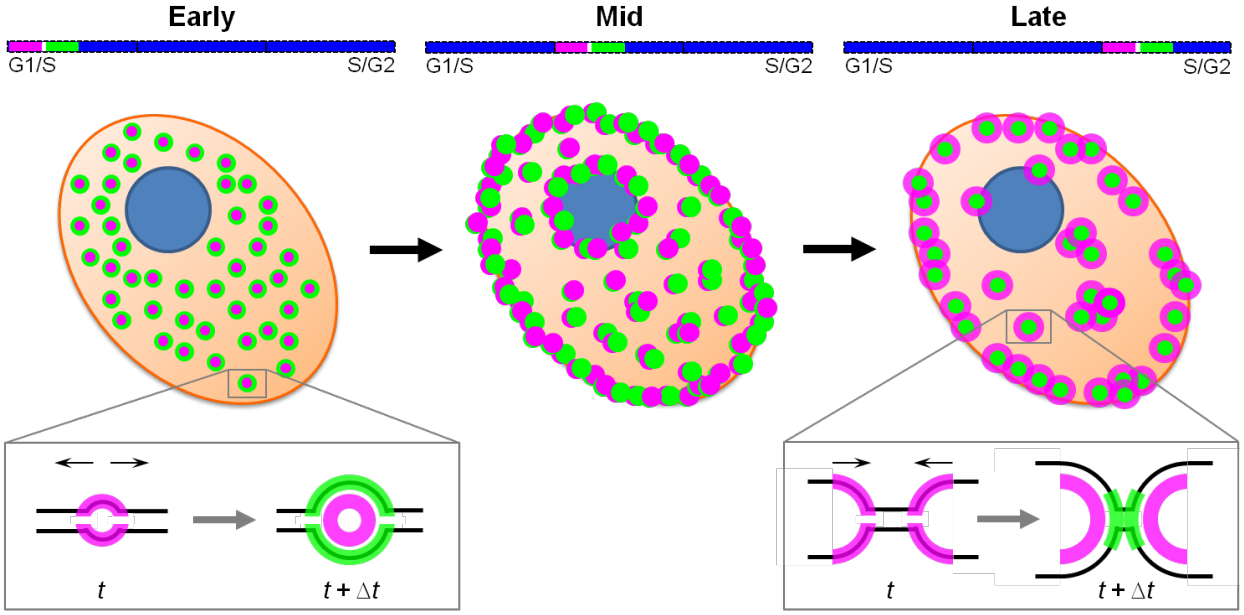

Figure 5.8. A model of morphological and spreading dynamics of mammalian RDs. Bars at the top indicate the relative temporal position of the labeling windows (pink and green) during the respective stages of S-phase (not drawn to scale). Insets show zoomed-in views of a RD during early (left) and late (right) S-phase, illustrating their respective direction of propagation.

In addition to nuclear lamina and the extent of replication progression, a number of other factors could also act to constrain the distribution and timing of RDs. First and foremost is chromatin structure, since loosely packed euchromatic DNA are known to replicate early while the highly condensed heterochromatin tends to replicate late (Chagin et al., 2010). Along that line, CTCF, which is involved in maintaining chromatin architecture (including demarcating the boundaries between transcriptionally active and repressed chromatin) could also play a role in delineating replication domains. Replication foci have also been found to localize to the anchor of DNA loops (Courbet et al., 2008), further pointing to the role of CTCF, which anchors DNA loops to the nuclear matrix, in spatially constraining RDs. The same could be said of cohesin, given that it shares a large fraction of DNA binding sites with CTCF, and its depletion in human cells has been shown to reduce the number of active replication origins and increase inter-origin distance (Guillou et al., 2010). Lastly, other nuclear species such as PCNA (which plays a central role in coordinating the events at the replication fork) and Rif1 (whose depletion leads to the loss of mid S-phase replication foci (Yamazaki et al., 2012)), as well as epigenetic determinants could also play a role in dictating the location and timing of chromosomal replication (Smith and Aladjem, 2014).

Overall, the work described in this chapter elucidates the ultra-structural features of replication domains in the mammalian nucleus in both space and time, and paves the way for further quantitative investigations into the organization and dynamics of replication at the super-resolution level. For example, the correlation between the RDs observed here in Cartesian space and those in sequence space revealed by replication timing profiles could be established by co-labeling specific sequences in these domains using DNA fluorescence *in situ* hybridization (FISH) and measuring their colocalization. By sequentially labeling RDs at the same stage in

contiguous cell cycles with multiple colors, we could also address the issue of the extent to which replication originates from the same sites in different cell cycles. Lastly, replication initiation has been found to often occur adjacent to transcription initiation events, and replicator elements can contribute to transcriptional activation (Cayrou et al., 2011; Martin et al., 2011). By combining the approaches adopted here with those in Chapter 4, we could probe the extent to which replication and transcription are spatially colocalized or segregated (Chakalova et al., 2005), thereby generating novel insights into these two critically important nuclear processes.

5.5. Materials and methods

Cell culture and synchronization

Hela cells were routinely cultured in high-glucose DMEM supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin (all from Gibco). The duration of the cell cycle was determined by counting the number of cells every 8 hr for a continuous period of 72 hr. The cells were synchronized to the G1/S boundary by treating with, sequentially, 2 mM thymidine for 15 hr, followed by incubation in normal culture medium for 10 hr (interim release period), and then by treatment with 2 μg/ml aphidicolin for 15 hr. The synchronization efficiency was determined by measuring the cellular DNA content profile in a population of DAPI-stained cells fixed at 30 min after release into S-phase, using a fluorescence-activated cell sorter (MoFlo, Beckman Coulter) at excitation wavelength of 340 - 380 nm.

Replication domains labeling and immuno-staining

At specific times after releasing the cells from G1/S boundary, replication sites in the nucleus were labeled by introducing short pulses of thymidine analogues. Dye-labeled (either Alexa 647

(Invitrogen) or Atto 550 (MolBi Tech)) dUTPs were delivered into the nucleus via transfection using the FuGENE 6 reagent (Promega), while EdU, being uncharged, was directly added into the culture medium and subsequently labeled with a dye using the Click-iT EdU Imaging Kit (Life Technologies). When performing two-color labeling, the cells were washed with medium for three times after incubating with the first dye before adding in the second dye. Upon labeling, the sample was fixed with 4% (v/v) paraformaldehyde (PFA) and 0.1% (v/v) glutaraldehyde in PBS buffer for 15 min. To stain the nuclear lamina, the cells were blocked and permeabilized with 5% (w/v) BSA and 0.5% (v/v) Triton X-100 in PBS buffer for 30 min, and incubated with mouse anti-Lamin A/C antibodies for 60 min. The primary antibody was visualized using a donkey anti-mouse secondary antibody labeled with Atto 488 (Sigma), using a procedure similar to that described in Chapter 4. After staining, the sample was post-fixed with 0.5% PFA in PBS buffer for 5 min.

Super-resolution imaging and data analysis

RLS-SRM imaging was performed as described in Chapter 4. The super-resolution images of Alexa 647- and Atto 550-labeled replication domains were continuously acquired for up to 100,000 frames, and those of Atto 488-labeled nuclear lamina were acquired for up to 50,000 frames. Bright-field images of 3 μm-sized glass beads placed in the sample dish prior to imaging were acquired every 2 s for correcting lateral drift between the frames.

RLS-SRM image analysis, drift correction, and quantifications were performed using Insight3 (gift of B. Huang) and custom-written software in Matlab (2010b, MathWorks), similar to those described in Chapter 4. Density-based spatial clustering was performed on the detected localizations to group them into clusters (consisting of ≥ 3 localizations not more than 60 nm

(laterally) and 120 nm (axially) apart from each other). These clusters are further grouped into domains (consisting of ≥ 3 clusters whose centroid positions are within 100 nm apart from each other). Colocalization was defined as the centroid positions of two structures being not more than 100 nm apart from each other.

References

Aladjem, M.I. (2007). Replication in context: Dynamic regulation of DNA replication patterns in metazoans. *Nat. Rev. Genet.* **8**, 588-600.

Aten, J.A., Bakker, P.J.M., Stap, J., Boschman, G.A., and Veenhof, C.H.N. (1992). DNA double labelling with IdUrd and CldUrd for spatial and temporal analysis of cell proliferation and DNA replication. *Histochem. J.* **24**, 251-259.

Ay, F., Bailey, T.L., and Noble, W.S. (2014). Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res.* **24**, 999-1011.

Bornfleth, H., Edelmann, P., Zink, D., Cremer, T., and Cremer, C. (1999). Quantitative motion analysis of subchromosomal foci in living cells using four-dimensional microscopy. *Biophys. J.* **77**, 2871-2886.

Cayrou, C., Coulombe, P., Vigneron, A., Stanojcic, S., Ganier, O., Peiffer, I., Rivals, E., Puy, A., Laurent-Chabalier, S., Desprat, R.*, et al.* (2011). Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res.* **21**, 1438-1449.

Chagin, V.O., Stear, J.H., and Cardoso, M.C. (2010). Organization of DNA replication. *Cold Spring Harb. Perspect. Biol.* **2**, a000737.

Chakalova, L., Debrand, E., Mitchell, J.A., Osborne, C.S., and Fraser, P. (2005). Replication and transcription: Shaping the landscape of the genome. *Nat. Rev. Genet.* **6**, 669-677.

Courbet, S., Gay, S., Arnoult, N., Wronka, G., Anglana, M., Brison, O., and Debatisse, M. (2008). Replication fork movement sets chromatin loop size and origin choice in mammalian cells. *Nature* **455**, 557-560.

Cremer, T., and Cremer, M. (2010). Chromosome territories. *Cold Spring Harb. Perspect. Biol.* **2**, a003889.

Dijkwel, P.A., Wang, S., and Hamlin, J.L. (2002). Initiation sites are distributed at frequent intervals in the Chinese hamster dihydrofolate reductase origin of replication but are used with very different efficiencies. *Mol. Cell. Biol.* **22**, 3053-3065.

Dimitrova, D.S. (2009). Visualization of DNA replication sites in mammalian nuclei. *Methods Mol. Biol.* **521**, 413-436.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380.

Fox, M.H., Arndt-Jovin, D.J., Jovin, T.M., Baumann, P.H., and Robert-Nicoud, M. (1991). Spatial and temporal distribution of DNA replication sites localized by immunofluorescence and confocal microscopy in mouse fibroblasts. *J. Cell Sci.* **99**, 247-253.

Gratzner, H. (1982). Monoclonal antibody to 5-bromo- and 5-iododeoxyuridine: a new reagent for detection of DNA replication. *Science* **218**, 474-475.

Guillou, E., Ibarra, A., Coulon, V., Casado-Vela, J., Rico, D., Casal, I., Schwob, E., Losada, A., and Méndez, J. (2010). Cohesin organizes chromatin loops at DNA replication factories. *Genes Dev.* **24**, 2812-2822.

Harper, J.V. (2005). Synchronization of cell populations in G1/S and G2/M phases of the cell cycle. *Methods Mol. Biol.* **296**, 157-166.

Hozák, P., Hassan, A.B., Jackson, D.A., and Cook, P.R. (1993). Visualization of replication factories attached to a nucleoskeleton. *Cell* **73**, 361-373.

Hutchison, C.J., Bridger, J.M., Cox, L.S., and Kill, I.R. (1994). Weaving a pattern from disparate threads: Lamin function in nuclear assembly and DNA replication. *J. Cell Sci.* **107**, 3259-3269.

Jackson, D.A. (1995). S-phase progression in synchronized human cells. *Exp. Cell Res.* **220**, 62-70.

Jackson, D.A., and Pombo, A. (1998). Replicon clusters are stable units of chromosome structure: Evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. *J. Cell Biol.* **140**, 1285-1295.

Jacob, F. (1993). The replicon: Thirty years later. *Cold Spring Harb. Symp. Quant. Biol.* **58**, 383-387.

Jacob, F., Brenner, S., and Cuzin, F. (1963). On the regulation of DNA replication in bacteria. *Cold Spring Harb. Symp. Quant. Biol.* **28**, 329-348.

Ma, H.T., and Poon, R.Y.C. (2011). Synchronization of HeLa cells. *Methods Mol. Biol.* **761**, 151-161.

Macville, M., Schröck, E., Padilla-Nash, H., Keck, C., Ghadimi, B.M., Zimonjic, D., Popescu, N., and Ried, T. (1999). Comprehensive and definitive molecular cytogenetic characterization of HeLa cells by spectral karyotyping. *Cancer Res.* **59**, 141-150.

Manders, E.M.M., Kimura, H., and Cook, P.R. (1999). Direct imaging of DNA in living cells reveals the dynamics of chromosome formation. *J. Cell Biol.* **144**, 813-822.

Martin, M.M., Ryan, M., Kim, R., Zakas, A.L., Fu, H., Lin, C.M., Reinhold, W.C., Davis, S.R., Bilke, S., Liu, H., *et al.* (2011). Genome-wide depletion of replication initiation events in highly transcribed regions. *Genome Res.* **21**, 1822-1832.

Maya-Mendoza, A., Olivares-Chauvet, P., Kohlmeier, F., and Jackson, D.A. (2012). Visualising chromosomal replication sites and replicons in mammalian cells. *Methods* **57**, 140-148.

Méchali, M. (2010). Eukaryotic DNA replication origins: Many choices for appropriate answers. *Nat. Rev. Mol. Cell Biol.* **11**, 728-738.

Nakamura, H., Morita, T., and Sato, C. (1986). Structural organizations of replicon domains during DNA synthetic phase in the mammalian nucleus. *Exp. Cell Res.* **165**, 291-297.

Nakayasu, H., and Berezney, R. (1989). Mapping replicational sites in the eucaryotic cell nucleus. *J. Cell Biol.* **108**, 1-11.

O'Keefe, R.T., Henderson, S.C., and Spector, D.L. (1992). Dynamic organization of DNA replication in mammalian cell nuclei: Spatially and temporally defined replication of chromosome-specific alpha-satellite DNA sequences. *J. Cell Biol.* **116**, 1095-1110.

Peric-Hupkes, D., and van Steensel, B. (2010). Role of the nuclear lamina in genome organization and gene expression. *Cold Spring Harb. Symp. Quant. Biol.* **75**, 517-524.

Perumal, S.K., Yue, H., Hu, Z., Spiering, M.M., and Benkovic, S.J. (2010). Single-molecule

studies of DNA replisome function. *Biochim. Biophys. Acta* **1804**, 1094-1112.

Pope, B.D., and Gilbert, D.M. (2013). The replication domain model: Regulating replicon firing in the context of large-scale chromosome architecture. *J. Mol. Biol.* **425**, 4690-4695.

Raulf, A., Spahn, C., Zessin, P., Finan, K., Bernhardt, S., Heckel, A., and Heilemann, M. (2014). Click chemistry facilitates direct labelling and super-resolution imaging of nucleic acids and proteins. *RSC Adv.* **4**, 30462–30466.

Rhind, N., and Gilbert, D.M. (2013). DNA replication timing. *Cold Spring Harb. Perspect. Biol.* **5**, a010132.

Robinson, N.P., and Bell, S.D. (2005). Origins of DNA replication in the three domains of life. *FEBS J.* **272**, 3757-3766.

Ryba, T., Battaglia, D., Pope, B.D., Hiratani, I., and Gilbert, D.M. (2011). Genome-scale analysis of replication timing: From bench to bioinformatics. *Nat. Protocols* **6**, 870-895.

Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T.C., Robins, A.J., Dalton, S., and Gilbert, D.M. (2010). Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res.* **20**, 761-770.

Sadoni, N., Langer, S., Fauth, C., Bernardi, G., Cremer, T., Turner, B.M., and Zink, D. (1999). Nuclear organization of mammalian genomes: Polar chromosome territories build up functionally distinct higher order compartments. *J. Cell Biol.* **146**, 1211-1226.

Schermelleh, L., Solovei, I., Zink, D., and Cremer, T. (2001). Two-color fluorescence labeling of early and mid-to-late replicating chromatin in living cells. *Chromosome Res.* **9**, 77-80.

Smith, O.K., and Aladjem, M.I. (2014). Chromatin structure and replication origins: Determinants of chromosome replication and nuclear organization. *J. Mol. Biol.* **426**, 3330–3341.

van Oijen, A.M., and Loparo, J.J. (2010). Single-molecule studies of the replisome. *Annu. Rev. Biophys.* **39**, 429-448.

Visser, A.E., and Aten, J.A. (1999). Chromosomes as well as chromosomal subdomains constitute distinct units in interphase nuclei. *J. Cell Sci.* **112**, 3353-3360.

Visser, A.E., Jaunin, F., Fakan, S., and Aten, J.A. (2000). High resolution analysis of interphase chromosome domains. *J. Cell Sci.* **113**, 2585-2593.

Watson, J.D., Baker, T.A., Bell, S.P., Gann, A., Levine, M., and Losick, R. (2007). Molecular Biology of the Gene, 6th ed. (San Francisco, Benjamin Cummings).

Wilson, R.H.C., and Coverley, D. (2013). Relationship between DNA replication and the nuclear matrix. *Genes Cells* **18**, 17-31.

Yamazaki, S., Ishii, A., Kanoh, Y., Oda, M., Nishito, Y., and Masai, H. (2012). Rif1 regulates the replication timing domains on the human genome. *EMBO J.* **31**, 3667-3677.

Chapter 6

Coupling conformational dynamics to catalysis in DNA polymerase [1]

6.1. Abstract

In this chapter, we turn our focus to a different form of dynamics on a finer spatio-temporal scale, namely the conformational dynamics within individual protein molecules, which has been shown to have a critical impact on many enzymatic reactions. In the realm of nuclear biology, an example of such conformational motion is the nucleotide-induced closing of the finger domain of DNA polymerase, whose catalytic action is crucial for DNA replication. Such large-amplitude molecular motion is often not fully accessible to either direct experimental monitoring or molecular dynamics simulations. However, a coarse-grained model can offer an informative alternative. Here we investigate the dynamics of T7 DNA polymerase catalysis using a Langevin-type Gaussian Network Model incorporating detailed structural information for the open conformation of the enzyme. Our single-parameter model, albeit simplified, remarkably captures the induced conformational dynamics of DNA polymerase upon

[1] The work described in this chapter was performed in collaboration with Prof. Hao Ge from Peking University.

deoxynucleoside triphosphate binding, and reveals its close coupling to the advancement towards transition state along the coordinate of the target reaction, which contributes to significant lowering of the activation energy barrier. Furthermore, analysis of stochastic catalytic rates suggests that when the activation energy barrier is low and when nonequilibrium relaxation towards the closed form dominates the catalytic rate, one must appeal to a picture of two-dimensional free energy surface in order to account for the full spectrum of catalytic modes. Our semi-quantitative study illustrates the precise role of conformational dynamics in achieving transition-state stabilization in the DNA polymerization reaction, and suggests that such approach possesses the potential to furnish significant mechanistic insights into the functioning of a variety of enzymatic systems.

6.2. Introduction

6.2.1. Conformational dynamics as a key driving force of enzymatic catalysis

It is now universally known that enzymes are the indispensable workhorse for a wide variety of biological activities inside a living organism, and mechanistic insights into how they carry out their functions is of utmost importance to an understanding of life at the molecular level. The working of most enzymes can be understood in the context of the general Michaelis-Menten reaction scheme, in which binding between the enzyme molecule and the substrate molecule forms an enzyme-substrate complex, which then undergoes the catalytic reaction to form the product molecule and regenerate the enzyme (Michaelis and Menten, 1913). Among the large number of mechanisms put forward to account for the drastic acceleration of enzyme-catalyzed reactions, the "transition-state stabilization" theory proposed by Pauling, in which the increase in reaction rate is achieved mainly through the lowering of the free energy activation barrier for

the catalytic step, is probably the most well accepted (Pauling, 1946). However, such purely thermodynamic picture could not fully account for the catalytic power of enzymes, as demonstrated by the large number of catalytic antibodies designed based on this principle (Hilvert, 2000).

More recently, with the advent of a wide range of new experimental approaches, in particular single-molecule techniques that allow real-time observation of individual enzyme molecules at work, a plethora of previously unattainable dynamic information related to enzyme catalysis have been uncovered (Abbondanzieri et al., 2008; Adachi et al., 2007; Eisenmesser et al., 2002; Furuike et al., 2008; Gershenson, 2009; Kern et al., 2005; Kohen et al., 1999; Kohen and Klinman, 1999; Lu et al., 1998; Luo et al., 2007; Nagel and Klinman, 2006; Noji et al., 1997; Yang et al., 2003). It has been shown that an enzyme molecule undergoes conformational motions during catalysis, which could take place across a broad range of time scales from femtoseconds up to 100 s (Kamerlin and Warshel, 2009; Min et al., 2005a; Min et al., 2007). In cases where conformational dynamics occur on a time scale comparable to that of enzymatic reactions, as demonstrated by single-molecule and NMR experiments (Eisenmesser et al., 2005; Kern et al., 2005; Min et al., 2005b; Yang et al., 2003), the classical transition-state theory and Kramers' rate formula would be insufficient. In many cases, such slow conformational motions could be the rate-limiting step for the reaction.

Realizing the fact that the enzyme molecule is structurally flexible, Haldane has proposed, at a very early time, the concept of "strain energy", the release of which upon substrate binding distorts the enzyme-substrate complex and leads to the lowering of activation energy barrier (Haldane, 1930). This idea was further developed by Koshland as the "induced fit" theory, which suggests that the structure of the active site of the enzyme as well as that of the substrate

itself could be gradually modified by the interactions between them (Koshland, 1958). These theories point to the close coupling between the intrinsic reaction coordinate (IRC) and the conformational coordinate of the enzyme molecule, and suggest that it is such dynamic conformational rearrangements that provide the mechanistic basis for "transition-state stabilization".

Intense efforts have since been undertaken to understand the effects of conformational motions on enzyme catalysis (Garcia-Viloca et al., 2004; Hammes, 2002; Hammes-Schiffer and Benkovic, 2006; Kamerlin and Warshel, 2009). Recently, Bagchi and colleagues formulated a general theoretical framework that accounts for such effects using a two-dimensional (2D) reaction free energy surface incorporating both the IRC and the conformational coordinate (Min et al., 2007). However, to date the applicability of this framework has not been quantitatively demonstrated in a specific catalytic system.

6.2.2. Polymerase conformational motions during DNA replication

In this chapter, we aim to demonstrate, semi-quantitatively, the exact role of conformational dynamics in achieving enzymatic catalysis in the context of DNA replication, and show how it is closely coupled to the forward dynamics towards the transition state along the IRC. To this end, we decide to use T7 DNA polymerase, an extremely well characterized enzyme, as our model system.[2] DNA polymerase is a family of enzymes that catalyze the addition of a deoxynucleoside triphosphate (dNTP) substrate to a newly synthesized DNA chain, a crucial

[2] We choose this enzyme of non-mammalian origin as our model system mainly because it is the single best-studied case among all the DNA polymerases, for which multiple crystal structures of both open and closed forms exist. Numerous studies have shown that both the key structural motifs as well as the nucleotide-induced conformational change of the DNA polymerase are well conserved across the species, thus making our findings equally applicable to the mammalian system.

step in genome replication across the species. The three-dimensional structure of the enzyme resembles that of a partially closed right hand (Watson et al., 2007). Upon formation of the correct base pairing between the template DNA and the incoming dNTP, the exposed 3' hydroxyl group of the nascent DNA chain will attack the α-phosphate of the substrate with the help of two metal ions bound to the palm domain. At the same time, several residues in the fingers domain of the polymerase form contact with the substrate, forcing one of the helices, named the O-helix, to move towards the dNTP, ultimately forming a closed conformation of the polymerase active site (Doublie et al., 1998; Langer, 1969) (**Figure 6.1**). Such large and distinctive conformational change associated with catalysis presents us with an excellent opportunity to study the relationship between the two (Doublie et al., 1998; Luo et al., 2007).

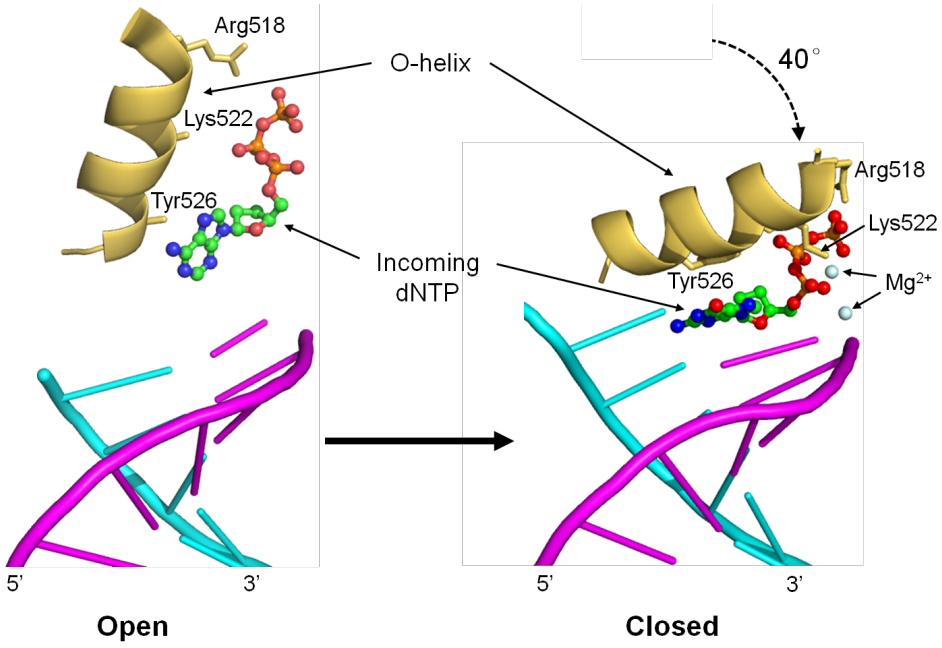

Figure 6.1. Structural mechanism of DNA synthesis by T7 DNA polymerase. Upon binding of the incoming dNTP (shown in ball-and-stick) to the DNA chain (cyan and magenta) in the open conformation (left), the O-helix (yellow) of the polymerase swings 40° clockwise to form the closed conformation (right) of the active site. (Figure follows Watson et al., 2007, with modifications).

6.2.3. Gaussian Network Model and its utility in probing conformational dynamics

To date, it is not yet possible to experimentally probe the dynamics of enzyme catalysis at the atomic level. Even if the technical challenge of observing the dynamics of a single enzyme molecule with full atomic detail could be overcome, one would quickly be buried in the sea of noisy information generated for each atom. In addition, there is no way to gradually adjust the physical parameters, such as electrostatic effects or strength of hydrogen bonds, of a real enzyme. While these difficulties could be surmounted *in silico* using molecular dynamics simulation, the millisecond timescale that could be reached by this approach is insufficient for following most catalytic reactions to completion. Moreover, it is also difficult, in a detailed molecular dynamics simulation, to choose and adjust a set of collective model parameters that are reflective of the conformational change upon substrate binding.

In view of these issues, we have chosen a coarse-grained Langevin-type collective model which could be regarded as the dynamic counterpart of the static Gaussian Network Model (GNM), with only one parameter (Bahar, 1997; Tirion, 1996). As a recent advancement of the classic polymer models for proteins (Flory et al., 1976), the GNM represents each amino acid residue of a protein by a node corresponding to the α-carbon atom, and each nucleotide in DNA by three nodes corresponding to the P, C2 and C4′ atoms. The spatial interactions between the nodes are modeled by harmonic springs. This model is relatively easy to build and compute, and has already been validated for its ability to catch the slowest vibrational modes in proteins, which are consistent with the experimental B-factors measured in crystallographic studies (Bahar, 1997; Tirion, 1996). Those slowest normal modes predicted by this model have been found to be closely related to enzyme catalysis; in particular, these slowest modes correspond to the hinge region in enzymes that undergoes crucial conformational change upon substrate

binding (Bahar et al., 1998; Bahar, 1997). Systematic structural analyses of many classes of enzymes have also revealed that a few of the slowest normal modes could play more than 50% of the role in the ligand-binding process for catalysis, which leads us to surmise that nature would likely make effective use of these high-entropic modes (Delarue and Sanejouand, 2002; Watson et al., 2007). These analyses suggest our GNM model could potentially capture key modes of motion that are functionally important for the enzymatic activity of T7 DNA polymerase, and that the conformational changes corresponding to the slowest modes in the GNM model would not be fully dissipated by other faster modes of dynamics not captured by our model.

In this chapter, we first build the static GNM for T7 DNA polymerase, and convert it into a Langevin-type dynamic model, incorporating only the open-form structure information. We then study the conformational dynamics of the enzyme molecule immediately following the binding of the dNTP substrate, and show how conformational distortion is closely coupled to the forward catalytic step along the chemical coordinate. Moreover, we show that the traditional Kramers' one-dimensional (1D) reaction-rate formula for enzyme catalysis, which normally focuses only on the structure of the enzyme-substrate complex and activation energy barrier height of the stabilized transition state, may not be sufficient in cases where the activation energy barrier is relatively low. Instead, one should appeal to a two-dimensional (2D) free energy surface description that incorporates more details on how to couple catalysis with conformational dynamics, in order to obtain more accurate reaction rates. Our results, based on specific structural information of a real enzymatic system, provide semi-quantitative support for the theoretical framework of enzyme-assisted reaction using a multisurface free-energy description (Min et al., 2007).

6.3. Results

6.3.1. Quantifying interactions in the polymerase-substrate complex

The catalytic process of T7 DNA polymerase consists of two crucial and nearly simultaneous steps: the breaking of the bond between α-phosphate and β-phosphate in dNTP (henceforth named αP-βP bond), and the formation of a covalent bond between the α-phosphate of dNTP and the exposed 3′ hydroxyl group of the nascent DNA chain (henceforth named αP-OH bond). In order to quantify these two steps, we model the interactions between the incoming dNTP and the polymerase by introducing five new springs into the structure of the enzyme-substrate complex (**Figure 6.2**), representing respectively: the αP-OH bond; the hydrogen bonds between the β-phosphate and Lys522 residue on the O-helix (henceforth named βP-Lys bond), between the β-phosphate and Arg518 residue (henceforth named βP-Arg bond [3]), and between the base of the dNTP and its cognate base on the template (henceforth named base pair bond); as well as the stacking interaction between the dNTP base and Tyr526 residue on the O-helix (henceforth named base-Tyr bond). Since the energy of hydrogen bond as well as stacking interaction is approximately 3 - 4 kcal/mol, we set all of these bonds to have the same tunable spring constant k, with the exception of the αP-OH bond, which has a spring constant c. Note that all other bonds within the DNA polymerase (assumed to exist between any two nodes less than 10 Å apart) are assigned a fixed spring constant of 1, in accordance with the conventional practice in static GNM (Yang et al., 2005). While such assignment will not allow us to differentiate the stronger covalent bonds from other weaker interactions, simulation shows that assigning all the covalent bonds with a larger spring constant do not significantly alter the results of our model

[3] In fact, it is the γ-phosphate that contacts the Arg518 residue, but for simplicity we assign the interaction to β-phosphate instead.

(data not shown), in agreement with results from earlier studies using GNM (Bahar, 1997; Haliloglu et al., 1997; Tirion, 1996). Compared to these bonds, the reasonable range for parameter k would be 0 ~ 1, whereas that for parameter c would be much larger than 1 due to strong electrostatic effects during the process of forming a covalent bond.

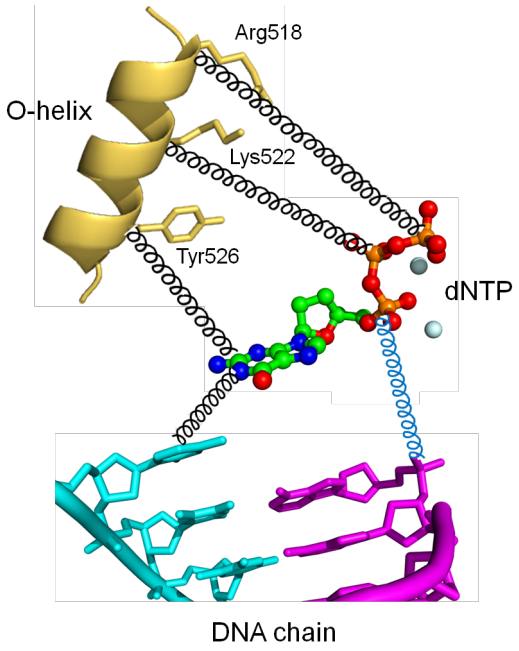

Figure 6.2. Key interactions introduced into the structure of the polymerase-substrate complex. The hydrogen bonds and stacking interaction (spring constant k) are shown in black, and the αP-OH bond (spring constant c) is shown in blue. The target αP-βP bond (spring constant b) is not shown.

Another issue worth taking note of while constructing the model is the strength of the target αP-βP bond, to which we assign a spring constant b with value 3, since it is a high-energy covalent bond that must be broken ultimately. Doing so would make it difficult to pull apart the target bond mechanically as well as lower the activation energy barrier for the reaction (see **Figure 6.3** and Section 6.3.2). However, since it is very possible that the conformational rearrangements during the formation of the closed conformation of the active site have already

weakened this high-energy bond prior to its breakage, we decide to assign a value of 1 to its spring constant after the substrate has come into the pocket of the DNA polymerase. Hence, the final lowering of the activation energy barrier is due to the cooperative effects of weakening the target bond and pulling it apart.

6.3.2. Modeling target bond breakage and activation energy barrier lowering

Upon substrate binding, *i.e.* formation of the abovementioned five interactions, the polymerase-substrate complex would no longer be at equilibrium, and there must be some restoring force pushing it towards a new equilibrium state, namely the closed conformation. As a result, some of the bonds, either covalent or noncovalent, would be transiently but significantly elongated and even broken, as a result of the dual actions of the spring-like mechanical effects arising from conformational dynamics and other more static chemical interactions. The central issue we would like to probe here is whether the interplay among the various interactions upon bond elongation would result in the breaking of only the desired αP-βP bond.

The new equilibrium position for each node and the lengths of four key bonds reflective of conformational changes within the polymerase-substrate complex are calculated as a function of k (**Figure 6.3**), demonstrating each of the two cases of whether the αP-βP bond has already been weakened or not in the closed conformation prior to catalysis. For proper catalysis, we would expect the lengths of these bonds not to deviate significantly from their equilibrium values except for the αP-βP bond, which is expected to be pulled broken. Given that the equilibrium length of this bond is approximately 3 Å, we set the threshold for bond breaking to be 8 Å. Our simulations show that only the αP-βP bond is significantly elongated and eventually broken during the conformational change from the open form to the closed form (**Figure 6.3**), in

excellent agreement with expectation, as long as the electrostatic effects of the two metal ions are set to be much stronger than the hydrogen bonds and stacking interaction (*i.e.* the case of c = 10). It is worth noting that such "pulling" effects have already been envisioned by biologists when such mechanisms are presented in textbooks (see, for example, Figure 8-7 in Watson et al., 2007), although evidence for such effects is hard to obtain from structural information alone (see **Figure 6.1**). Therefore, our simulation strongly supports this picture of "pulling the bond" while furnishing it with much more quantitative information.

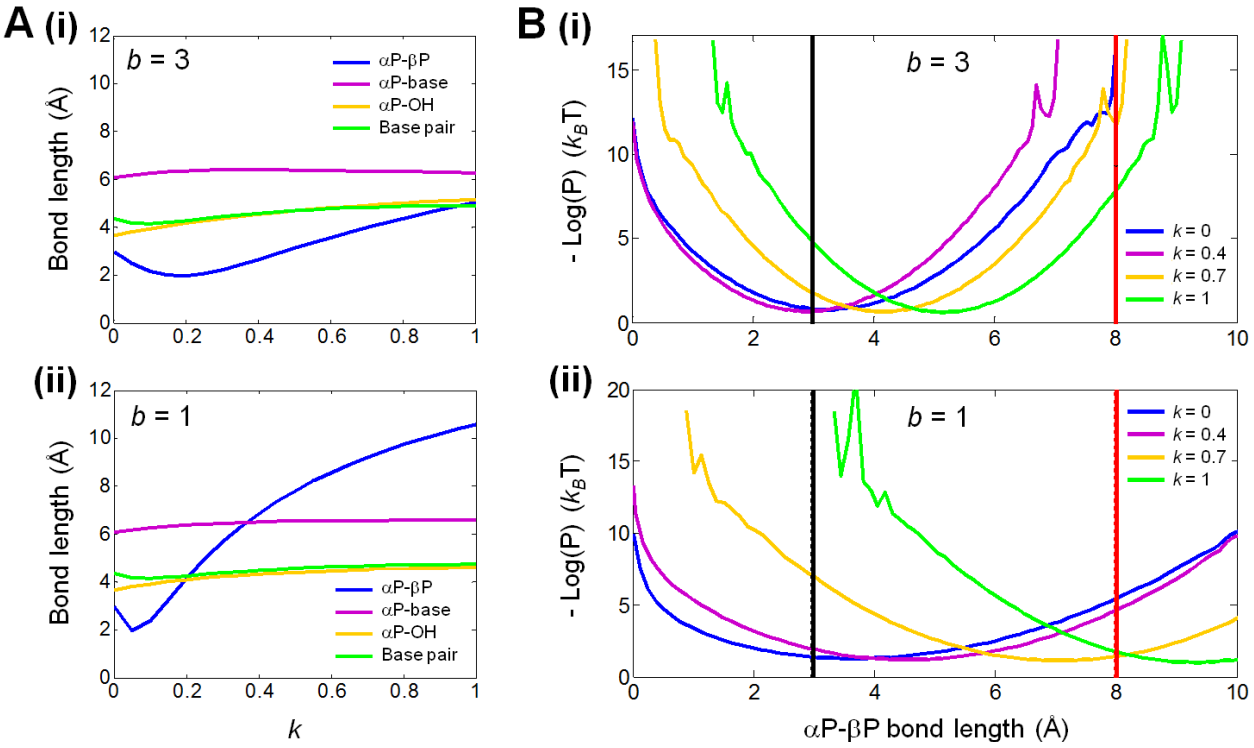

Figure 6.3. Modeling target bond breakage and activation energy barrier lowering during DNA polymerization. (A) Distortion of the polymerase upon substrate binding when the αP-βP bond is either un-weakened (*i.e.* b = 3, **(i)**) or weakened (*i.e.* b = 1, **(ii)**) in the closed conformation. The lengths of four key bonds are shown as a function of k for the case of strong electrostatic interaction (c = 10)**. (B)** Potential of mean force as a function of αP-βP bond length, using different values of k. The activation energy barrier of the reaction can be lowered to different extents for both un-weakened **(i)** and weakened **(ii)** αP-βP bond**.** The black and red lines denote the equilibrium (3 Å) and broken (8 Å) length of the αP-βP bond, respectively.

Moreover, if the αP-βP bond is not weakened prior to catalysis (*i.e.* b = 3), the target αP-βP bond would not be significantly elongated even if the electrostatic effects of the two metal ions are strong (**Figure 6.3 A (i)**). However, numerical simulation still suggests that the activation energy barrier could be lowered from ~ 15 k_BT to ~ 8 k_BT (**Figure 6.3 B (i)**), solely as a consequence of the spring-like mechanical effects. Such reduction is not sufficient for a typical enzymatic reaction, and thus there must be some other mechanisms not included in the present model that would help weaken the target αP-βP bond. It is likely that these interactions are also a consequence of the optimized structure of the closed conformation. When such mechanisms are taken into consideration (by setting b = 1), then the desired αP-βP bond is significantly elongated and eventually broken (**Figure 6.3 A (ii)**). However, if this is the only effect at work (without the distortion of the enzyme), the activation energy barrier could only be lowered to ~ 5 k_BT (**Figure 6.3 B (ii)**, k = 0). It is only through the combined effects of αP-βP bond weakening and the spring-like conformational distortion of the enzyme that this barrier could be further lowered to 3.2 k_BT, 0.28 k_BT, or even made to vanish (**Figure 6.3 B (ii)**, k = 0.2, 0.4 and 0.7, respectively). Therefore, we have shown that the hydrogen bonds and stacking interaction that operate during the conformational change would significantly help to advance the catalytic reaction by further lowering the activation energy barrier for bond breaking beyond that achievable by standard chemical mechanisms. Note that in our model, the induced conformational dynamics is always energetically downhill, and would not meet any barrier along its trajectory.

6.3.3. Catalytic time and 2D free energy surface of the polymerization reaction

In order to obtain more detailed information on the rate of polymerase catalysis, we compute the stochastic first passage time for the reaction, *i.e.* the time in simulation starting from

substrate binding up to the breaking of the αP-βP bond (**Figure 6.4**). For simplicity, we set the value of the spring constant c to be 10 and that of b to be 1, *i.e.* we assume strong electrostatic interaction, and that the αP-βP bond has been pre-weakened before substrate binding. We found that the mean first passage time is significantly reduced even in the presence of relatively weak hydrogen bond interactions between the dNTP and the polymerase (*i.e.* when k is small), corresponding to an acceleration of catalysis (**Figure 6.4**).

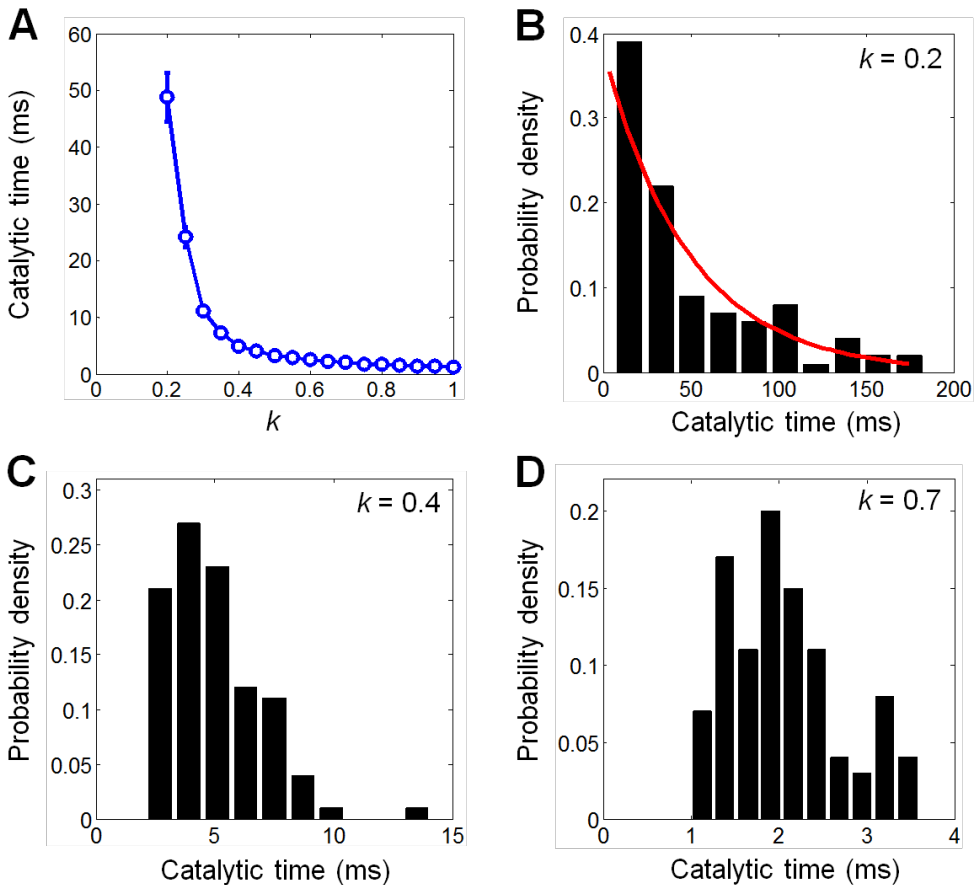

Figure 6.4. Catalytic time of the polymerization reaction. (A) Mean first passage time of the catalytic reaction as a function of k. **(B-D)** Histograms of the simulated catalytic times show that Kramers' rate formula (as exemplified by exponentially distributed catalytic time) holds when the hydrogen bond interactions is weak (k = 0.2) and the activation energy barrier is high **(B)**. As the strength of hydrogen bond interactions increases (k = 0.4 or 0.7), a Gaussian-like distribution for the catalytic time indicative of the existence of an intermediate emerges **(C** and **D)**.

Although the catalytic time has a reduced mean in the presence of stronger interactions, it still exhibits large variations around this mean (**Figure 6.4 B** to **D**). It is interesting to observe that when the activation energy barrier is high (*e.g.* $k = 0.2$), the catalytic time could be well fit by an exponential distribution, whereas when the barrier is extremely low (*e.g.* $k = 0.7$), the distribution becomes Gaussian-like, suggesting the existence of a two-step mechanism with an intermediate (Lu et al., 1998). Note that the simulated catalytic time does not include the time for the substrate binding process, which is entirely diffusion-driven and could simply be regarded as a first-order rate process. The emergence of a Gaussian-like distribution of catalytic time is due to the fact that the entire catalytic process, starting from substrate binding, consists of at least two steps: a nonequilibrium downhill relaxation process for the polymerase-dNTP complex to reach the closed conformation, and the crossing of the lowered activation energy barrier along the IRC of the catalyzed reaction. In the overdamped case, the probability density of the relaxation time is not necessarily monotonic, since it must peak at the deterministic relaxation time.

Therefore, there are perhaps two possible types of catalysis for DNA polymerase: one with a relatively high energy barrier where the catalytic time is dominated by the crossing of the activation energy barrier and is nearly exponentially distributed according to the celebrated Kramers' rate formula (Hanggi et al., 1990; Kramers, 1940), and the other with extremely low energy barrier where the catalytic time is dominated by the nonequilibrium relaxation from the open conformation to the closed conformation of the polymerase. In the former mode, the dynamics of the conformational change is fast compared to that of barrier crossing along the IRC; hence the catalytic rate is entirely barrier-dominant and Kramers' rate formula holds. In the latter case, however, the time scale for the conformational motion is comparable to or even

much slower than that for barrier crossing. Hence, we must appeal to a free energy surface description incorporating both the conformational dynamics and the catalytic reaction, in order to calculate a more accurate rate of catalysis (Langer, 1969; Min et al., 2007).

Such a 2D free energy surface is computed for our model (**Figure 6.5**), in which the horizontal *X* axis denotes the IRC as represented by the length of the αP-βP bond, and the vertical *Q* axis denotes the length of the βP-Arg bond, which is considered as the conformational coordinate for the enzyme. Regardless of whether the high-energy αP-βP bond is pre-weakened, it is evident that in the presence of the newly formed hydrogen bond and stacking interactions after substrate binding, the induced distortion of the polymerase could simultaneously push the polymerase-dNTP complex to undergo a drastic forward step towards the transition state (**Figure 6.5**). The optimal reaction trajectory towards the global free energy minimum of the surface, which is slanted along both coordinates, is exactly the consequence of such kinetic coupling as predicted in previous model (Min et al., 2007).

Lastly, as mentioned above, the simulated rate accounts only for the second step from the encounter complex (E··S) to the transition state (ES≠) in the well-known Michaelis-Menten catalytic scheme (Cornish-Bowden, 2004; English et al., 2006; Michaelis and Menten, 1913):

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} E \cdot\cdot S \xrightarrow{k_2} E + P \qquad \textbf{(6.1)}$$

where *E*, *S*, and *P* denote the enzyme, substrate, and product, respectively. Under this scheme, the velocity of the polymerization reaction is given by

$$v = \frac{k_2[S]}{\left(k_{-1} + k_2\right)/k_1 + [S]}, \qquad \textbf{(6.2)}$$

and the catalytic power is $k_1 k_2 / \left(k_{-1} + k_2\right)$. Here we could simply assume that the product release

time after crossing the activation energy barrier is negligible, and hence k_2 is just the catalytic rate we have simulated, which increases with the spring constant k. Therefore, it is clear that both the reaction velocity and the catalytic power are monotonically increasing with k_2, and consequently increasing with the hydrogen bond strength k.

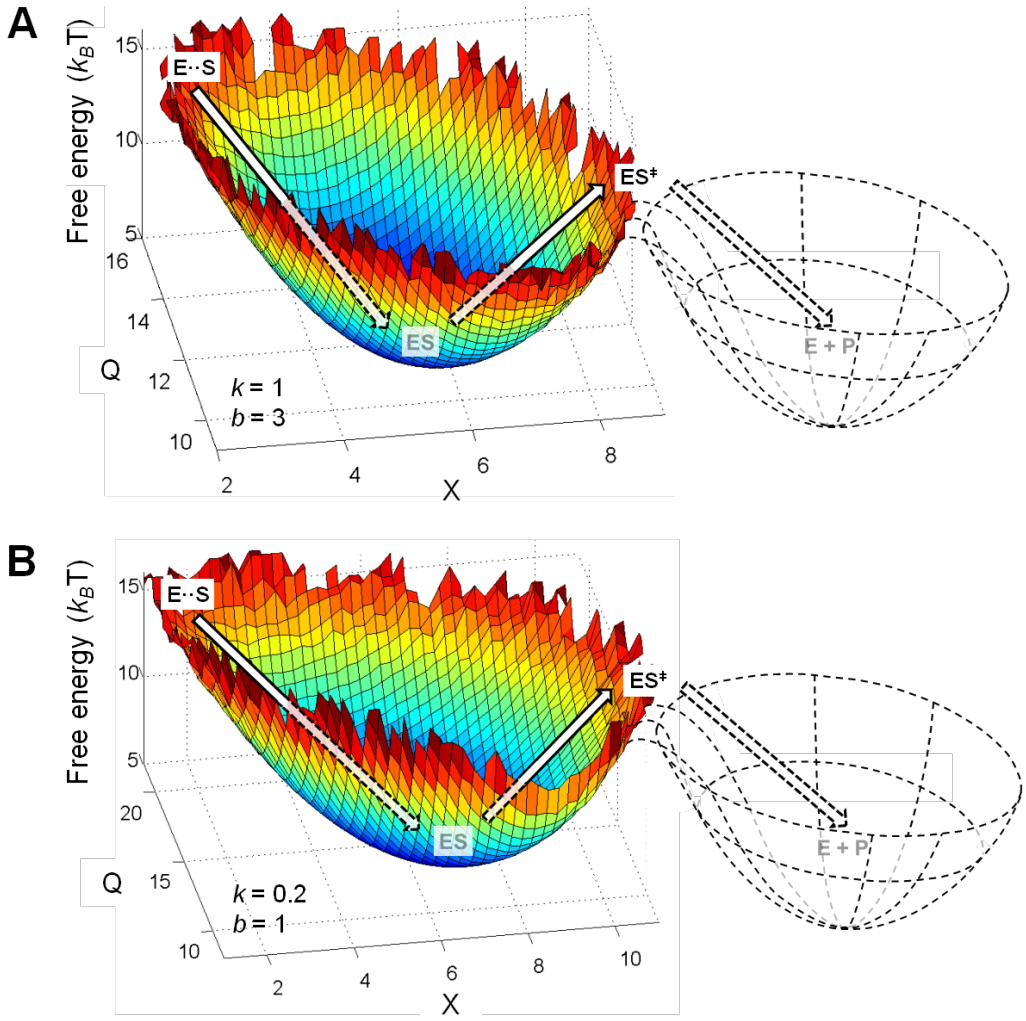

Figure 6.5. Effective 2D free energy surface for DNA polymerase catalysis. Different strengths of hydrogen bonds are used to generate the surface: **(A)** k = 1, b = 3; and **(B)** k = 0.2, b = 1. The X and Q axes denote the lengths of the αP-βP bond and the βP-Arg bond, respectively. The push of the polymerase-dNTP complex towards the transition state by the induced distortion in the enzyme is clearly shown. The additional well (dashed lines) illustrates the process of product release after crossing the transition state, which could not be directly simulated in our model.

6.4. Discussion and conclusion

Model validity and its discontents

Among the anticipated issues regarding the Langevin-type GNM used in this study, one primary argument concerns how could such a harmonic model possibly allow us to infer the correct mechanism for enzyme catalysis. Since ligand binding is an active process rather than the result of spontaneous thermal fluctuations, this kind of dynamic anharmonicity does not contradict with the harmonic approximation of the local structures in the system. Furthermore, as a result of induced fit, the activation energy barrier along the pathway of ligand binding might be low, thereby making the harmonic approximation hold for a large range of spatial and temporal scales (Ma, 2005). On the other hand, the frictional constant would be several orders different for the vibrational motion and the large conformational motion, since the latter will feel strong solvent damping (Ma, 2005). Therefore, it is possible to apply such a model describing the vibrational motion of the native state of a protein at the picosecond time scale in the linear response regime to study the microsecond/millisecond conformational changes associated with enzyme catalysis. In fact, a model based on similar principles has already been used to show that competing native interactions in adenylate kinase can account for the conformational transitions between the open and closed forms of the enzyme (Whitford et al., 2007).

Positional dependence of spring constants

When calculating the first passage time of the catalytic reaction, we assigned fixed values to the spring constants. More generally, we could also consider the case when the value of the spring constant k is dependent on the distance between the two ends of the spring. However, due to

the similar qualitative behavior of such a broad parameter range for k, there should not be any serious problem in assuming a constant value for k. This is also not a problem for the calculations of the new equilibrium positions in the closed conformation; what we really need is the final spring constant in the active site rather than its varying value during conformational change. Hence, although the approximated spring constant must be quite low when the O-helix and incoming dNTP are far from each other at the beginning, it should be comparable to that of other hydrogen bonds after they finally form their own interaction in the closed conformation.

Specificity of DNA polymerase

It has been shown recently that the conformational dynamics of DNA polymerase upon substrate binding is very sensitive to the correctness of base pairing (Luo et al., 2007). This is intuitively true if one views the conformational dynamics as a result of the physical interactions between the incoming dNTP and the O-helix in the finger domain as well as the hydrogen bonds formed between the correct base pair. Without correct base pairing, the interactions between the O-helix and the phosphate groups in the dNTP alone would not drive the O-helix to move close to the DNA template, but would only produce local fluctuations. Our model could not explain this phenomenon well because we have already incorporated the correct equilibrium positions of the incoming dNTP where the correct base pair forms, which would forbid the substrate from moving towards the O-helix by itself under the strong electrostatic effects of the αP-OH bond. However, in reality, if no correct base pair is formed, the probability of the αP-OH bond formation would be significantly lowered since it is such a short-range local interaction. Hence in this case, the two unpaired bases would be far from each other, implying that the fidelity of the DNA polymerase would also be reflected by its conformational dynamics.

Induced fit and spring-like mechanical effects

Theoretically, to achieve superior specificity, an enzyme and its substrate should possess almost completely complementary structures. However, we now know very well that it is impossible to have an enzyme and its substrate to fit perfectly into one another, *i.e.* the "lock and key" model (Fischer, 1894), for the enzyme molecule is structurally flexible, and hence Koshland's "induced fit" model (Koshland, 1958). In our model, it could be explicitly seen that transition state stabilization is achieved entirely through the coupling between the conformational change of the O-helix and other more static chemical interactions. This is in agreement with the previously proposed general model of enzyme catalysis, in which motion along the conformational coordinate of the enzyme accommodates the substrate to accomplish activation energy lowering (Min et al., 2007). Although at the atomic level, such spring-like mechanical effects are also chemical (mainly electrostatic) in nature, this kind of dynamical forces has not been given much attention in previous studies of enzyme mechanisms. The main result of such coupling is not simply the positioning of the incoming nucleotide into close contact with the catalytic metal ions (Watson et al., 2007), but also the lowering of the activation energy barrier. Such coupling could even eliminate the activation energy barrier altogether in certain cases. This is why the reaction rate could be tremendously enhanced.

It is well-known that conformational change upon substrate binding is almost inevitable during enzyme catalysis, and the present study illustrates the role of conformational dynamics in enzyme catalysis. In this model, the incoming substrate would not first wait for the formation of the optimized conformation of the active site and then start the catalysis reaction. Instead, these two dynamic processes are strongly coupled together due to the semi-flexible nature of

the enzyme polymer which exerts spring-like mechanical effects on the catalytic reaction. This coupling constitutes the key difference between a reaction with and without the enzyme.

Dynamic realization of transition-state stabilization and related rate formula

By building a Langevin-type model, the transition-state stabilization is dynamically realized in a rigorous manner. Once the activation energy barrier is significantly lowered, the semi-classical transition-state theory for calculating the reaction rate may no longer be applicable, since the slow conformational fluctuations take place on a similar time scale as enzymatic reactions (Eisenmesser et al., 2005; Koshland, 1958; Min et al., 2005b; Yang et al., 2003). Therefore, the nonequilibrium relaxation time for the system to slide down to the active site would be the rate-limiting step during catalysis. In this case, the Kramers' rate formula would not be accurate. On the other hand, the quantum effects of tunneling would for sure enhance the reaction rate, but they are also sensitive to the distance between the donor and the acceptor; as a precondition, the two groups must first be pushed together for tunneling through the conformational change of enzyme to occur, and these effects result in the weakening of the αP-βP bond. Thus, the mechanism of hydrogen tunneling still falls within the framework of the present study, although its full treatment would involve more details of quantum mechanics.

Conformational dynamics: equilibrium fluctuation vs. nonequilibrium relaxation

It is evident that conformational dynamics of the enzyme upon substrate binding must deviate from equilibrium fluctuations, and undergo a spontaneous nonequilibrium relaxation process. At equilibrium state without substrate binding, the closed conformation of the polymerase could hardly be arrived at only through thermal fluctuations, and hence nonequilibrium downhill conformational dynamics on a two-dimensional free energy surface after substrate

binding is precisely what is called "reorganization" by Warshel et al. (Kamerlin and Warshel, 2009). However, the substrate does not passively wait for the accomplishment of the reorganization process and then start the catalytic reaction; instead, it simultaneously starts its journey towards the transition state with the help of conformational dynamics. Of course, it is hard to pinpoint which factor causes or helps the other, since they are both the result of electrostatic pre-organization in a general sense.

In conclusion, we used a Langevin-type dynamic model to dissect the conformational dynamics of T7 DNA polymerase. By treating conformational dynamics explicitly on equal footing as reaction kinetics along the IRC, we showed that the mechanism for lowering the reaction barrier through large conformational change during catalysis has already been encoded in the topology of the open-form structure of the enzyme. We found that the conformational dynamics of the enzyme molecule immediately after the binding of dNTP is closely coupled to the forward catalytic step, and results in significant acceleration of the catalytic reaction. In particular, we showed that the αP-OH bond can be formed concomitantly as the αP-βP bond is broken, thereby lending evidence to such "pulling" effect that have long been envisioned but so far unsubstantiated from structural studies alone. The spring-like mechanical effects exerted by the enzyme molecule on the substrate could provide an alternative mechanism for the realization of transition-state stabilization, beyond the conventional chemical mechanisms for enzyme catalysis. Moreover, we have shown that the traditional Kramers' reaction-rate formula for enzyme catalysis could not fully describe reactions for which the activation energy barrier is sufficiently low. Instead, a more accurate 2D free energy surface description that incorporates details on how to achieve catalysis through conformational dynamics is needed (Min et al., 2007). Overall, our semi-quantitative approach combining Langevin-type dynamic modeling

and real structural information could be applied to many other enzymes, so as to shed mechanistic insights on their multifarious modes of catalytic actions.

6.5. Materials and methods

Dynamic Gaussian Network Model construction

Atomic coordinates of the open form of T7 DNA polymerase are obtained from the Protein Data Bank (accession number 1SL0), and the connection matrix Γ is generated from the online server, http://ignm.ccbb.pitt.edu/GNM_Online_Calculation-t.htm (hosted by Ivet Bahar's group at University of Pittsburgh), using a default cutoff distances of 10 Å for both amino acid residues and nucleotides (Yang et al., 2005). The results obtained are found to be insensitive to the value of cut off distance used from 7 Å to 12 Å (data not shown). Using the connection matrix, the overdamped Langevin equation for the open conformation of the enzyme can be written as

$$\eta \frac{dX_{pol}}{dt} = -\frac{k_B T \gamma}{2} \frac{\partial \left[\left(X_{pol} - X_{open}^{eq} \right)^T \Gamma_{open} \left(X_{pol} - X_{open}^{eq} \right) \right]}{\partial X_{pol}} + \xi(t), \qquad \textbf{(6.3)}$$

where γ is the renormalization factor, and the diffusion constant D is defined according to the Fluctuation-Dissipation Theorem

$$2D = 2\eta k_B T. \qquad \textbf{(6.4)}$$

The spring constant is set to $\eta = k_B T$. Upon substrate binding, the new connection matrix incorporates the coordinates of the incoming dNTP and the five induced interactions shown in **Figure 6.2**, while keeping all other interactions unchanged. The new equilibrium positions of the nodes in the polymerase-dNTP complex in the closed conformation is then taken to be the solution of the equation

$$\frac{\partial V_{binding}(X)}{\partial X}\Big|_{X=X_{closed}} = 0, \tag{6.5}$$

where

$$V_{binding}(X) = \left(X_{pol} - X_{open}^{eq}\right)^T \Gamma_{open} \left(X_{pol} - X_{open}^{eq}\right) + E_{\text{int}eraction} + \left(X_{dNTP} - X_{dNTP}^{eq}\right)^T \Gamma_{dNTP} \left(X_{dNTP} - X_{dNTP}^{eq}\right) \tag{6.6}$$

is the overall potential of the polymerase-dNTP complex in the modified Langevin equation

$$\eta \frac{dX}{dt} = -\frac{k_B T \gamma}{2} \frac{\partial V_{binding}(X)}{\partial X} + \xi(t). \tag{6.7}$$

The interaction energy $E_{\text{int}eraction}$ includes all five interactions shown in **Figure 6.2**, and Γ_{dNTP} represents the connectivity between the base, αP and βP atoms in the incoming dNTP molecule. The related variance of the Gaussian-distributed new equilibrium positions is then

$$\left\langle \Delta X \Delta X^T \right\rangle = \frac{\Gamma^{-1}}{\gamma}, \tag{6.8}$$

and the variance of the bond length is

$$\left\langle \Delta R \Delta R^T \right\rangle = \frac{3\Gamma^{-1}}{\gamma}, \tag{6.9}$$

which is used to calculate the activation energy barrier as well as the 2D free energy surface for the polymerization reaction.

Simulation and image preparation

All simulations were performed in Matlab (2010b, MathWorks) using the simplest Euler form (difference equations) for the Langevin dynamics. All structural figures were prepared using PyMOL (Schrödinger, Inc.).

References

Abbondanzieri, E.A., Bokinsky, G., Rausch, J.W., Zhang, J.X., Le Grice, S.F.J., and Zhuang, X. (2008). Dynamic binding orientations direct activity of HIV reverse transcriptase. *Nature* **453**, 184-189.

Adachi, K., Oiwa, K., Nishizaka, T., Furuike, S., Noji, H., Itoh, H., Yoshida, M., and Kinosita, K. (2007). Coupling of rotation and catalysis in F1-ATPase revealed by single-molecule imaging and manipulation. *Cell* **130**, 309-321.

Bahar, I., Atilgan, A.R., Demirel, M.C., and Erman, B. (1998). Vibrational dynamics of folded proteins: Significance of slow and fast motions in relation to function and stability. *Phys. Rev. Lett.* **80**, 2733.

Bahar, I., Atilgan, A. R., Erman, B. (1997). Direct evaluation of thermal fluctuations in protein using a single parameter harmonic potential. *Folding Design* **2**, 173-181.

Cornish-Bowden, A. (2004). Fundamentals of Enzyme Kinetics, 3rd ed. (London, Portland Press).

Delarue, M., and Sanejouand, Y.H. (2002). Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model. *J. Mol. Biol.* **320**, 1011-1024.

Doublie, S., Tabor, S., Long, A.M., Richardson, C.C., and Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* **391**, 251-258.

Eisenmesser, E.Z., Bosco, D.A., Akke, M., and Kern, D. (2002). Enzyme dynamics during catalysis. *Science* **295**, 1520-1523.

Eisenmesser, E.Z., Millet, O., Labeikovsky, W., Korzhnev, D.M., Wolf-Watz, M., Bosco, D.A., Skalicky, J.J., Kay, L.E., and Kern, D. (2005). Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438**, 117-121.

English, B.P., Min, W., van Oijen, A.M., Lee, K.T., Luo, G., Sun, H., Cherayil, B.J., Kou, S.C., and Xie, X.S. (2006). Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited. *Nat. Chem. Biol.* **2**, 87-94.

Fischer, E. (1894). Einfluss der configuration auf die wirkung der enzyme. *Ber. Dtsch. Chem. Ges.* **27**, 2985-2993.

Flory, P.J., Gordon, M., and McCrum, N.G. (1976). Statistical thermodynamics of random networks. *Proc. Roy. Soc. London A* **351**, 351-380.

Furuike, S., Hossain, M.D., Maki, Y., Adachi, K., Suzuki, T., Kohori, A., Itoh, H., Yoshida, M., and Kinosita, K. (2008). Axle-less F1-ATPase rotates in the correct direction. *Science* **319**, 955-958.

Garcia-Viloca, M., Gao, J., Karplus, M., and Truhlar, D.G. (2004). How enzymes work: Analysis by modern rate theory and computer simulations. *Science* **303**, 186-195.

Gershenson, A. (2009). Single molecule enzymology: Watching the reaction. *Curr. Opin. Chem. Biol.* **13**, 436-442.

Haldane, J.B.S. (1930). Enzymes (London, New York, Longmans, Green).

Haliloglu, T., Bahar, I., and Erman, B. (1997). Gaussian dynamics of folded proteins. *Phys. Rev. Lett.* **79**, 3090-3093.

Hammes, G.G. (2002). Multiple conformational changes in enzyme catalysis. *Biochemistry* **41**, 8221-8228.

Hammes-Schiffer, S., and Benkovic, S.J. (2006). Relating protein motion to catalysis. *Annu. Rev. Biochem.* **75**, 519-541.

Hanggi, P., Talkner, P., and Borkovec, M. (1990). Reaction-rate theory: Fifty years after Kramers. *Rev. Mod. Phys.* **62**, 251-341.

Hilvert, D. (2000). Critical analysis of antibody catalysis. *Annu. Rev. Biochem.* **69**, 751-793.

Kamerlin, S.C.L., and Warshel, A. (2009). At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins: Struct. Funct. Bioinf.* **78**, 1339-1375.

Kern, D., Eisenmesser, E.Z., Wolf-Watz, M., and Thomas, L.J. (2005). Enzyme dynamics during catalysis measured by NMR spectroscopy. In *Methods in Enzymology* (Academic Press), vol. 394, pp. 507-524.

Kohen, A., Cannio, R., Bartolucci, S., and Klinman, J.P. (1999). Enzyme dynamics and hydrogen tunnelling in a thermophilic alcohol dehydrogenase. *Nature* **399**, 496-499.

Kohen, A., and Klinman, J.P. (1999). Hydrogen tunneling in biology. *Chem. Biol.* **6**, R191-R198.

Koshland, D.E. (1958). Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **44**, 98-104.

Kramers, H.A. (1940). Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7**, 284-304.

Langer, J.S. (1969). Statistical theory of the decay of metastable states. *Annals Phys.* **54**, 258-275.

Lu, H.P., Xun, L., and Xie, X.S. (1998). Single-molecule enzymatic dynamics. *Science* **282**, 1877-1882.

Luo, G., Wang, M., Konigsberg, W.H., and Xie, X.S. (2007). Single-molecule and ensemble fluorescence assays for a functionally important conformational change in T7 DNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12610-12615.

Ma, J. (2005). Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure* **13**, 373-380.

Michaelis, L., and Menten, M. (1913). Die kinetik der invertinwirkung. *Biochem. Z.* **49**, 333-369.

Min, W., English, B.P., Luo, G., Cherayil, B.J., Kou, S.C., and Xie, X.S. (2005a). Fluctuating enzymes: Lessons from single-molecule studies. *Acc. Chem. Res.* **38**, 923-931.

Min, W., Luo, G., Cherayil, B.J., Kou, S.C., and Xie, X.S. (2005b). Observation of a power-law memory kernel for fluctuations within a single protein molecule. *Phys. Rev. Lett.* **94**, 198302.

Min, W., Xie, X.S., and Bagchi, B. (2007). Two-dimensional reaction free energy surfaces of catalytic reaction: Effects of protein conformational dynamics on enzyme catalysis. *J. Phys. Chem. B* **112**, 454-466.

Nagel, Z.D., and Klinman, J.P. (2006). Tunneling and dynamics in enzymatic hydride transfer. *Chem. Rev.* **106**, 3095-3118.

Noji, H., Yasuda, R., Yoshida, M., and Kinosita, K. (1997). Direct observation of the rotation of F1-ATPase. *Nature* **386**, 299-302.

Pauling, L. (1946). Molecular architecture and biological reactions. *Chem. Eng. News* **24**, 1375-1377.

Tirion, M.M. (1996). Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* **77**, 1905-1908.

Watson, J.D., Baker, T.A., Bell, S.P., Gann, A., Levine, M., and Losick, R. (2007). Molecular Biology of the Gene, 6th ed. (San Francisco, Benjamin Cummings).

Whitford, P.C., Miyashita, O., Levy, Y., and Onuchic, J.N. (2007). Conformational transitions of adenylate kinase: Switching by cracking. *J. Mol. Biol.* **366**, 1661-1671.

Yang, H., Luo, G., Karnchanaphanurach, P., Louie, T.-M., Rech, I., Cova, S., Xun, L., and Xie, X.S. (2003). Protein conformational dynamics probed by single-molecule electron transfer. *Science* **302**, 262-266.

Yang, L.-W., Liu, X., Jursa, C.J., Holliman, M., Rader, A.J., Karimi, H.A., and Bahar, I. (2005). iGNM: a database of protein functional motions based on Gaussian Network Model. *Bioinformatics* **21**, 2978-2987.