ABSTRACT

COUNTRYMAN, PRESTON JAMES. Single-Molecule Studies of Sequence and Structure Dependent DNA Binding by SA1 and SA2. (Under the direction of Dr. Hong Wang).

Single-molecule interactions between proteins and DNA can reveal information about biological function that is difficult to obtain using bulk assay methods. Sister chromatid cohesion is mediated by the cohesin complex, which also plays important roles in diverse biological processes including maintenance of 3-D chromatin organization, DNA replication fork restart, and DNA double-strand break repair. In vertebrates, the core cohesin complex consists of a tripartite ring assembled from Smc1, Smc3, Rad21, and the stromal antigen subunit (SA) SA1 or SA2. All current models attribute cohesin DNA binding exclusively to the entrapment of ds-DNA by the cohesin ring subunits. In this thesis, we use single-molecule atomic force microscopy (AFM) and fluorescence imaging to characterize the static and dynamic interactions of cohesin subunits SA1 and SA2 in the presence of different DNA sequences and structures. AFM imaging revealed that SA1 had a weak affinity for telomeric regions of DNA, while SA2 showed no such sequence dependence. Fluorescence imaging captured on telomeric DNA tightropes showed that SA1 uses two unique binding modes: a fast mode to scan non-telomeric regions, and a slow mode once the telomere region was encountered. Individual SA1 were able to transition between these two distinct diffusive modes. Additionally, AFM studies showed that both SA1 and SA2 had a strong predilection for DNA ends. SA2 also had a large affinity for ss-DNA regions placed in the middle of longer ds-DNA. Fluorescence imaging revealed that SA2 binding to gap DNA tightropes was very strong, often trapping any mobile SA2 that encountered a gap region. Our work has shown that SA1 and SA2 are able to bind to DNA directly and may act as structural anchors for the entire cohesin complex to successfully bind to DNA. In addition, we applied single-
molecule imaging techniques to study proteins involved in the telomere maintenance pathway. Human telomeres are maintained by the shelterin protein complex in which TRF1 and TRF2 bind directly to duplex telomeric DNA. Using single-molecule fluorescence imaging of quantum dot labeled TRF1 and TRF2, we study how these proteins locate TTAGGG repeats on DNA tightropes. Unlike the stable and static associations observed for other proteins at specific binding sites, TRF proteins possess reduced binding stability marked by transient binding and slow 1-D diffusion on specific telomeric regions. We propose that the TRF proteins use 1-D sliding to find protein partners and assemble the shelterin complex, which in turn stabilizes the interaction with specific telomeric DNA.
Single-Molecule Studies of Sequence and Structure Dependent DNA Binding by SA1 and SA2

by
Preston James Countryman

A thesis submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Physics

Raleigh, North Carolina

2016

APPROVED BY:

_______________________________  ________________________________
Dr. Hong Wang                  Dr. Robert Riehn
Committee Chair

_______________________________  ________________________________
Dr. Keith Weninger             Dr. Guozhuo Xu
DEDICATION

To my friends, family, and wife for their support and love
BIOGRAPHY

Preston James Countryman was born on November 22, 1986 in Daly City, California to Patricia Ellen Countryman and Preston Otis Countryman. He spent his adolescence living in the Bay Area, and moved to Sacramento at the age of 10. By 2004 he had enrolled at California State University at Chico and would go on to receive majors in both physics and history. In summer 2010 he departed California to begin graduate school in physics at North Carolina State University. Here he developed a passion for single-molecule analysis of biological systems and met his lovely wife, Dr. Colleen Lanz Countryman.
ACKNOWLEDGMENTS

Committee members: Hong Wang, (NC State, Physics), Robert Riehn (NC State, Physics), Keith Weninger (NC State, Physics), Guozhuo Xu (NC State, Biochemistry)


Collaborators: Ingrid Tessmer, Patricia Opresko, Susan Smith, Neil M. Kad, Jacob Piehler, Yizhi Jane Tao

Advisors: Karen Daniels (NC State, Physics), Adam Hall (Wake Forest, Medicine)
TABLE OF CONTENTS

LIST OF FIGURES .............................................................................................................................................. vii

CHAPTER 1: Fundamentals of Atomic Force Microscopy .......................................................... 1

1.1 Introduction ........................................................................................................................................... 1

1.2 Atomic Force Microscopy (AFM) ................................................................................................... 1

   Modes of operation ................................................................................................................................. 2

   Sample preparation for AFM imaging .................................................................................................. 4

   AFM imaging in liquids ......................................................................................................................... 6

   Force analysis using AFM .................................................................................................................... 7

   Imaging improvements ............................................................................................................................ 8

   AFM studies of protein-protein and protein-DNA interaction ........................................................... 9

   Data processing of AFM images ........................................................................................................... 11

CHAPTER 2: Fundamentals of fluorescence microscopy .......................................................... 14

2.1 Introduction .......................................................................................................................................... 14

2.2 Theory behind TIRF ............................................................................................................................ 14

2.3 Penetration depth and Oblique Angle Fluorescence ...................................................................... 17

2.4 Single molecule tracking ................................................................................................................... 17

2.5 Quantum dot properties and fluorescent dyes ................................................................................. 23

2.6 Protein-QD conjugation methods ..................................................................................................... 24

2.7 Gap DNA preparation ....................................................................................................................... 25

2.8 Specifications of in lab fluorescence microscope ............................................................................. 27

CHAPTER 3: Single-particle tracking of quantum-dot labeled proteins on DNA ........ 28

3.1 Introduction .......................................................................................................................................... 28

3.2 Improved positional accuracy through Gaussian fitting ................................................................. 28

3.3 Methods of data analysis .................................................................................................................... 29

3.4 Prediction of diffusion constants and stepping rates based on Stokes-Einstein relation ................. 34

3.5 Prediction of additional energy barriers ............................................................................................ 36

CHAPTER 4: Cohesin SA2 is a sequence independent DNA binding protein but recognizes DNA ends and gaps ................................................................. 38

4.1 Cohesin background ............................................................................................................................. 38
4.2 Summary ........................................................................................................................................... 39

CHAPTER 5: Functional interplay between SA1 and TRF1 in telomeric DNA binding and DNA-DNA pairing .......................................................... 49

5.1 Background ...................................................................................................................................... 49

5.2 Summary .......................................................................................................................................... 50

CHAPTER 6: Telomere maintenance pathway .................................................................................... 58

6.1 Introduction to telomeres ................................................................................................................. 58

6.2 Summary .......................................................................................................................................... 63

REFERENCES .......................................................................................................................................... 74

APPENDICES ......................................................................................................................................... 91

Appendix A: Cohesin SA2 is a sequence independent DNA binding protein that recognizes DNA ends and gaps ........................................................................................................... 92

Appendix B: Functional interplay between SA1 and TRF1 in telomeric DNA binding and DNA-DNA pairing ................................................................................................................................. 131

Appendix C: TRF1 and TRF2 use different mechanisms to find telomeric DNA but share a novel mechanism to search for protein partners at telomeres ................................................. 184
LIST OF FIGURES

Figure 1: Cartoon representation of the principal mechanism behind AFM scanning ........ 12
Figure 2: Example force vs distance curve that mirrors the type of interaction between the
scanning tip and sample during AFM measurements using specific imaging modes ........ 13
Figure 3: SA2 does not show binding preference for telomeric or centromeric DNA
sequences, but recognizes DNA ends. ................................................................. 42
Figure 4: SA2 specifically binds to ssDNA gaps......................................................... 44
Figure 5: SA2 switches between searching and recognition modes on DNA tightropes
containing ssDNA gaps.................................................................................. 46
Figure 6: Comparison of the distribution of interval based diffusion constants (Dint) for SA2
on different DNA tightropes. ........................................................................ 48
Figure 7: Full length SA1 alternates between fast and slow 1-D diffusion on T270 DNA
tightropes........................................................................................................ 54
Figure 8: Comparison of the distributions of interval based diffusion constants (Dint) for SA1
TRF1. .................................................................................................................. 56
Figure 9: TRF1 and TRF2 strike a balance between target search and specificity .......... 68
Figure 10: Computer simulations of diffusion by modeling random walk of proteins on a 1-D
DNA lattice using PythonTM programming language........................................ 70
Figure 11: Dynamics of TRF2ΔB-QDs on λ DNA and the ligated T270 DNA ............ 72
CHAPTER 1: Fundamentals of Atomic Force Microscopy

1.1 Introduction

Interactions between proteins and DNA are fundamental in diverse biological processes, including DNA repair, replication, and transcription (Billingsley, Bonass et al. 2012, Qiu, DeRocco et al. 2012, Stratmann and van Oijen 2014). Revealing the structure and dynamics behind protein-DNA interactions is the key to improving the understanding about these complex biological systems. Single-molecule studies can provide addition about population distributions lost in bulk biochemical assays that only report the average system behavior. Importantly, single-molecules techniques allow for the observation of biologically rare important events or conformations that are not reported in bulk assays. This work focuses on applying single-molecule atomic force microscopy (AFM) to investigate the conformational properties of the protein-DNA interactions involved in telomere maintenance pathways and sister chromatid cohesion.

1.2 Atomic Force Microscopy (AFM)

Atomic Force Microscopy (AFM), also known as Scanning Force Microscopy was invented during the 1980s (Binnig, Quate et al. 1986, Hansma, Elings et al. 1988). Over the last two decades, different AFM imaging modes were introduced, making it one of the most versatile single-molecule imaging and manipulation techniques capable of examining
physical properties of biological samples in air or liquids with high surface resolution (Bustamante and Rivetti 1996).

Modes of operation

AFM is a near-field approach, wherein a sample surface is directly probed by a sharp tip (several nanometers in diameter) attached to a flexible cantilever. A laser is reflected off the back of the cantilever into a four quadrant photodiode. During the scanning process, the cantilever is bent towards the surface due to attractive tip-surface interactions, or reflected away from the surface due to repulsive surface interactions. These interactions are captured by the deviation from the central location of the photodiode (Figure 1: Cartoon representation of the principal mechanism behind AFM scanning). The most commonly used AFM imaging modes include contact mode, intermittent contact mode (or oscillating mode), and non-contact mode. The height profile is measured directly from the deviation of the reflected laser onto the four-quadrant photodiode with sub nanometer precision capability.

Contact mode requires constant direct contact between the AFM tip and the sample surface. The shear forces from the tip dragging over the surface increase the odds of damaging or moving individual protein and DNA molecules, which leads to sample degradation or image smearing. Tapping mode remedies this by oscillating the cantilever tip (close to its resonant frequency) during the scanning process, allowing for surface interaction solely at the apex of the bottom oscillation. The photodiode converts the reflected laser into a sinusoidal electric signal, which is then converted into an RMS value for the amplitude of the
oscillations. The feedback loop keeps the amplitude of oscillation constant as the tip scans over the surface, whereas in contact mode the angle of cantilever deflection is what is held constant. The brief interaction between the tip and surface greatly reduces stress forces, allowing for superior imaging of protein and DNA systems (Bustamante and Rivetti 1996).

In order to ensure proper imaging of the sample, the tip-sample distance needs to be held constant with sub-nanometer uncertainty. This is accomplished by using a frequency modulated signal to dampen external forces that affect the tip-sample distance (Martin, Williams et al. 1987, Rode, Oyabu et al. 2009).

A further development known as non-contact mode relies solely on attractive and repulsive forces between the surface and the tip without actually making contact with the sample, thus best preserving the integrity of the sample. These non-contact forces dampen the amplitude of oscillation, which is corrected by the feedback loop in order to keep the amplitude constant. This stress reduction makes tapping imaging the ideal candidate for imaging dense protein/DNA systems.

During scanning, AFM tip-sample interactions are mediated by combinations of forces including long-range electrostatic interactions, medium range attractive van der Waals forces (~10 nm), and short range repulsive Bohr forces (~0.1 nm) (Leite and Herrmann 2005). The repulsive Bohr force that dominates very close interactions occurs when two molecules begin overlapping electron orbitals, thus the repulsive force is a direct consequence of the Pauli Exclusion Principal. A standard model used to describe the forces encountered by the AFM tip is the Lennard-Jones potential,
\[ U = \frac{C_1}{x^{12}} - \frac{C_2}{x^6} \]  

where \( C_1 \) and \( C_2 \) are the corresponding constants of the repulsive and attractive forces encountered by the tip and \( x \) represents the distance between the sample and tip. As the tip draws closer to the surface, it falls into the potential well, resulting in an attractive force on the tip (Figure 2: Example force vs distance curve that mirrors the type of interaction between the scanning tip and sample during AFM measurements using specific imaging modes). The attractive \( x^{-6} \) term is often described as the van der Waals interaction because it is the direct product of interaction between molecules on the tip and molecules on the surface (more on molecular interactions later). If the separation between the tip and sample continues to decrease, the repulsive \( x^{-12} \) term dominates, resulting in a net repulsive force on the tip (Leite Herrmann 2005).

AFM imaging of biological samples can be done either in air or in liquid. Liquid imaging is more complicated due to hydrodynamic interactions between the oscillating tip and the sample (Bustamante and Rivetti 1996). However the phase change between the cantilever vibration and the drive signal can be held constant via the feedback loop instead of the amplitude, which is called phase intermittent contact mode. Examining the phase difference after sample interaction yields insight into the elasticity or friction of the sample. It has also been shown to improve resolution during liquid imaging of samples (Argaman, Golan et al. 1997).

Sample preparation for AFM imaging
An ideal surface for AFM imaging must be compatible with the sample and exceptionally flat. While the latter issue is important to ensure that only the sample height is being measured, the former issue is much more important for biological samples. A variety of surfaces have been used to immobilize biomolecules, such as glass, silicon, and mica (Wagner 1998). Muscovite mica is the most commonly used substrate. It has a negatively charged surface that can easily be made atomically flat via a simple peeling process with household scotch tape. It is important to have cations such as Mg\(^{2+}\), Ni\(^{2+}\), or Ca\(^{2+}\) to act as salt bridges to adhere the negatively charged DNA or proteins to the mica surface (Hansma and Laney 1996). Another method is to chemically treat the mica surface to prepare a positively charged surface for depositing negatively charged samples without buffer limitations. For example, a silicon or mica surface incubated with (3-aminopropyl)triethoxysilane (APTES) forms a self-assembled monolayer with siloxane bonds. The free, positively charged amino groups are then able to attach to negatively charged elements (Tessmer, Kaur et al. 2013).

Before DNA is deposited on the surface, it is floating in an aqueous 3D environment. Upon deposition to mica, the DNA becomes restricted to 2D motion. Given the right buffer conditions, DNA is quickly adsorbed onto the surface (5 minutes for 6 kBp ds DNA). Rather than observing projections of the 3D conformation of proteins or DNA onto the mica surface, the equilibration process creates a stable 2D environment (Rivetti, Guthold et al. 1996). Longer DNA chains (> 20 persistence lengths) also encounter volume exclusion, which deviates the measured length of the DNA from the theoretical prediction. For mica surfaces
without surface modification. ~20 ul of the sample is deposited on the surface, followed by a brief incubation time (< 30 seconds). The sample is then gently (to avoid striating the DNA on the surface) blown dry with nitrogen gas before being loaded onto the AFM.

AFM imaging in liquids

Liquid imaging offers the ability to observe real-time dynamics of biological systems while under physiologically relevant conditions. This can be for used for studying single-molecule systems, interactions between macromolecules, or live cells. Liquid imaging is particularly useful to analyze behavior of lipids for phase transitions or lipid mixtures (Alessandrini and Facci 2014). Note that for a nitrogen gas dried mica surface, there is a thin layer of water covering the sample from water molecules in the air. Similarly, AFM tips made of silicon or silicon nitride are also hydrophilic, and also have a surface layer of water molecules (Lyubchenko 2011). This layer of water also induces an attractive force, known as the capillary force, between the tip and the sample at close distances. In contrast, since the AFM tip is always immersed during liquid imaging, this drastically reduces the capillary forces on the sample, leading to images with higher resolution (Drake, Prater et al. 1989). Lower capillary forces allows for higher scan speeds and minimized sample destruction (Hamon, Curmi et al. 2010). However there are complications in getting biological samples to properly adsorb onto the surface for sufficient time frames and resolving the biological conformational changes temporally. Recent advances in high-speed AFM imaging through minimization of vibrations in the z-scanner and optimization of the feedback controller allow
for ~250 nm scans to be completed in a few hundred milliseconds (Ando, Uchihashi et al. 2008). Improvements to the electronic circuits, cantilever deflection detection system, and other electronic circuits have also improved scanning speed (Ando 2012).

Force analysis using AFM

Another useful feature provided by an AFM is the ability to probe the physical interactions between biomolecular assemblies or cells using force spectroscopy. The cantilever tip essentially acts as a spring that can test the force between the tip and the sample using Hooke’s Law

\[ F = k \times x \]  

(2)

where \( F \) is the force measured by the tip deflection, \( k \) is the spring constant of the cantilever, and \( x \) is the displacement. The spring constant \( k \) can be calculated using

\[ k = \frac{Ywt^3}{4L^3} \]  

(3)

with cantilever dimensions width \( w \), thickness \( t \), length \( L \), and a Young’s modulus \( Y \) (Giessibl 2003). Another method used to calculate the spring constant describes the motion of the oscillating cantilever through the use of equipartition theorem. The kinetic energy stored in the system is equal to one half of the thermal energy of the system, and the deflection is related back to the force through Hooke’s law (eq. 2) (Cook, Schaffer et al. 2006). Typical spring constants range from 0.01 N m\(^{-1}\) to 100 N m\(^{-1}\), providing a sensitivity down to \( 10^{-11} \) N. In force spectroscopy, the displacement is typically measured as a function of the cantilever deflection from the photodiode sensor in volts and the deflection sensitivity of the cantilever,
from which the force is then calculated (Zlatanova, Lindsay et al. 2000). Usually in Atomic Force Spectroscopy (AFS) experiments, the sample is immobilized on the surface while the interacting partner is adhered to the AFM tip. Analysis of a force versus distance curve can provide insight into the forces necessary to break or establish the bond between the two biological samples, examples of which include protein-protein interactions, DNA secondary structures, or ligand-receptor interactions (Lynch, Baker et al. 2009, Han, Qin et al. 2012, He, Hu et al. 2015). The force versus distance curve generated by this technique can also provide insight into intra-molecular forces.

Imaging improvements

There are many avenues of improvement for AFM technology to increase the resolution of images. Perhaps the most obvious starting point is the tip used to scan over the sample. A tip with a smooth shape has some radius $R$ near the apex of curvature. This tip curvature makes it difficult to resolve features between two objects on the surface located close to each other. The limit of resolution for the AFM tip depends on the distance between the surface objects ($d$), the height of the objects ($h$), the deformation height of the feature between objects ($\Delta z$), and finally $R$ (Lyubchenko 2011). Tip sharpness is also a vital concern in order to minimize the amount of van der Waals interactions by ensuring the surface molecules interact with as few tip molecules as possible. Early silicon cantilevers contained integrated tips etched in the [0 0 1] crystalline direction, which maximized the sharpness (Marcus 1990). However, the tip must be stable enough to avoid damage if it makes contact
with the sample or experiences an excess of force. Tips with [1 1 1] orientation proved more structurally stable due to increased bonding with surrounding tip molecules (Giessibl, Hembacher et al. 2000). Carbon nanotubes offer another suitable material for cantilever and tip construction. In addition to improved tip sharpness, carbon nanotubes have improved elasticity, which limits the maximum force it can deliver to biological samples (Wong, Joselevich et al. 1998). Controlling the length and orientation of carbon nanotubes is difficult however, thus the process of creating high quality tips from this method is expensive and tedious (Buchoux, Aime et al. 2009). Other factors that improve image resolution include tip-sample interactions, proper immobilization and adsorption on the surface of the sample, pH balance of the buffer, and ionic strength of the buffer (Mou, Czajkowsky et al. 1995, Muller and Engel 1997, Muller, Fotiadis et al. 1999). Changes in temperature can cause the cantilever to bend and possible alter the eigenfrequency of the tip (Giessibl 2003).

AFM studies of protein-protein and protein-DNA interaction

Due to the ease of sample preparation and low impact on biological samples, imaging protein-protein or protein-DNA interactions with AFM is an excellent way to examine single-molecule interactions in systems. As previously mentioned, on a mica surface, DNA molecules re-equilibrates on the surface in 2D, allowing for relevant information to be gathered from AFM images.

1. The recognition of specific DNA sequences and structures by protein machineries.
Proteins can also have a specific predilection for specific regions of DNA based on the sequence of DNA (Kaur, Wu et al. 2014, Lin, Countryman et al. 2014). Other proteins localize to regions of DNA damage, such as nicks, gaps, or breaks (Sukhanova, Abrakhi et al. 2016). Proteins can also serve as stabilizing agents for secondary DNA structures such as loops or G quadruplexes (Wang, Finzi et al. 2009). Analysis of protein position on DNA can yield information about the specificity of the protein to these regions, additionally providing the disassociation constant (Wang, Tessmer et al. 2008).

2. **The size, shape, and volume of protein-DNA or protein-protein complexes.**

In order to extract meaningful information from AFM scans, a volume calibration curve using known proteins can be used to create a linear relationship between volume measured by the AFM and the molecular weight of proteins (Wyman, Grotkopp et al. 1995, Ratcliff and Erie 2001, Yang, Wang et al. 2003, Sukhanova, Abrakhi et al. 2016). This relationship additionally allows for examination of protein aggregation and protein-DNA volume interactions (Daban 2011, Kaur, Wu et al. 2014). Additionally, the disassociation constants of proteins can be obtained from analysis of the AFM images (Ratcliff and Erie 2001, Yang, Wang et al. 2003).

3. **The processing of DNA by proteins.**

AFM scans can yield information about the searching or targeting mechanisms for proteins to locate specific locations along DNA by mapping the instances of protein-DNA interaction (Chen, Haushalter et al. 2002, Shlyakhtenko, Lushnikov et al. 2012). It can also
serve as a tool to investigate the effects of protein binding to DNA conformation, such as the DNA bending angle (Janicijevic, Sugasawa et al. 2003, Yang, Wang et al. 2003).

Recall that for all of these instances, the protein-DNA samples are held immobile on the surface. Thus the images generated by the AFM resemble instances in time rather than live, dynamic processes. Just like a movie, enough consecutive still images can give an idea for a larger pathway, process, or reaction. Each AFM scan takes a relatively small amount of time (on the order of minutes), making it possible to generate enough images to have a statistically robust sample size. Importantly, images from different depositions must be scanned and analyzed to avoid concluding results from sampling noise, deposition pollutants, or strange surface interactions. Additional controls, depending on the nature of the experiment in question, are also necessary to avoid inclusion of artifacts into the data analysis.

Data processing of AFM images

A variety of platforms and programs exist, mostly generated by the scientific community, to maximize the acquired data from each AFM scan. Many free programs, such as Gwyddion or ImageJ, can be used or customized in order to extract information such as height, area, volume, and length of AFM scans (both of which are used in our lab). More sophisticated and expensive software is necessary to analyze the complex interactions between protein-DNA complexes. Matlab, Igor, and LabView are the three used by our lab to trace the backbone of DNA and map assorted protein-DNA interactions, although there are
many acceptable substitutes. Our lab uses Gwyddion to extract information such as height, area, volume, and length of DNA.

Figure 1: Cartoon representation of the principal mechanism behind AFM scanning [provided by (Tessmer, Kaur et al. 2013)]
Figure 2: Example force vs distance curve that mirrors the type of interaction between the scanning tip and sample during AFM measurements using specific imaging modes [adapted from (Leite and Herrmann 2005)]
CHAPTER 2: Fundamentals of fluorescence microscopy

2.1 Introduction

Over the years, Total Internal Reflection Fluorescence (TIRF) has become a staple of the biological community in order to examine a variety of systems under physiologically relevant conditions. Unlike AFM imaging in air, TIRF experiments provide dynamic information about the system in question. The highly adaptable system allows a plethora of samples, such as live cells, DNA suspended tightropes, and microtubules, to be viewed using fluorescent materials such as quantum dots or fluorescent dyes. The field is rapidly evolving, with revolutionary new adaptations of the basic technique altering the technical landscape nearly annually.

2.2 Theory behind TIRF

James Clerk Maxwell revolutionized modern science with his unification of light, electric fields, and magnetic fields. Previously these concepts were considered individually, but Maxwell was able to unify them in just four equations:

\[ \nabla \cdot D = \rho_f \quad (4) \]

\[ \nabla \times E = -\frac{\partial B}{\partial t} \quad (5) \]

\[ \nabla \cdot B = 0 \quad (6) \]

\[ \nabla \times H = J_f + \frac{\partial D}{\partial t} \quad (7) \]
Armed with the fact that light is constructed of time varying electric and magnetic fields, Maxwell’s equations make it possible to establish boundary conditions for light as it impinges on a surface:

\[ \hat{n} \cdot (D_2 - D_1) = \sigma_f \]  
\[ (8) \]

\[ \hat{n} \times (E_2 - E_1) = 0 \]  
\[ (9) \]

\[ \hat{n} \cdot (B_2 - B_1) = 0 \]  
\[ (10) \]

\[ \hat{n} \times (H_2 - H_1) = K_f \]  
\[ (11) \]

If there are no free charges on the surface and no surface currents, then equations (8-11) state that the normal components of \(D\) and \(B\) are continuous and the tangential components of \(E\) and \(H\) are continuous. Error! Reference source not found. Error! Reference source not found. Error! Reference source not found.

When a ray of light impinges on a surface, there are three total rays to ultimately consider: the initial ray, the reflected ray, and the transmitted (or refracted) ray. Snell’s Law (eq. 12) shows that the angle of the reflected ray with respect to the normal of the impinged surface is equal to that of the initial ray (although this can also be deduced by applying Maxwell’s boundary conditions).

\[ \theta_i = \theta_r \]  
\[ (12) \]

The transmitted light ray is governed by the law of refraction (eq. 13)

\[ n_1 \sin \theta_1 = n_2 \sin \theta_2 \]  
\[ (13) \]

where \(n\) is the index of refraction of the material the ray is moving through, and the numeration applies to the initial and refracted wave respectively. A curious instance occurs
when the ray moves from a higher index of refraction to a lower index of refraction \((n_1 > n_2)\). As \(\theta_1\) increases, \(\theta_2\) also increases, until the condition \(\sin \theta_2 \leq 1\) is met. Thus a real angle for the transmitted light ray only exists for incoming light below a certain angle. This angel is defined as the critical angle

\[
\sin \theta_c = \frac{n_2}{n_1}
\]

(14)

For angles greater than the critical angle, all light is reflected, which gives rise to total internal reflection. At the critical angle, the refracted light ray travels parallel to the surface. The refracted wave also has an amplitude that decreases perpendicular to the surface in material 2. This phenomena is known as an evanescent wave, and describes the decaying intensity of the light wave in the transmitted medium for incoming light at the critical angle. The amplitude of the evanescent wave decreases exponentially, and has a limited penetration depth (eq. 15) into the second medium

\[
\delta = \frac{1}{K} = \frac{(\lambda_2/2\pi)}{\left[\left(\frac{n_1}{n_2}\right)^2 \sin^2 \theta_i - 1\right]^{1/2}}
\]

(15)

where \(\lambda_2 = 2\pi/k_2 = 2\pi c/n_2\omega\) is the wavelength of a normal plane wave in the medium. Finally the application of Maxwell’s boundary conditions (8-11) reveals that the reflected wave has the same magnitude of electric field as the initial wave. Due to the conservation of energy between the initial and reflected wave, this is known as total reflection.
2.3 Penetration depth and Oblique Angle Fluorescence

The fundamental mechanic of TIRF is to allow a small amount of energy to penetrate into a closed system in order to activate a fluorescent device related to some relevant biomolecule (more on fluorescent devices later). In principal, a biological sample is placed on or close to a surface. A highly collimated light beam is incident on the opposite side of the sample at the critical angle, causing total internal reflection to occur. The evanescent waves allow for an energy transfer to the fluorescent devices to occur, but the penetration depth of the evanescent waves into the sample is limited. The typical penetration depth is on the order of hundreds of nanometers (Konopka and Bednarek 2008). While there are many adjustable parameters in order to determine the penetration depth of the evanescent wave (eq. 15), the angle of the incident laser beam has the most experimental variability. Oblique Angle Fluorescence (OAF) is a variant on the TIRF technique where sub-critical angles are used instead of the true critical angle. This technique generates an increased signal-to-noise ratio and increased penetration depth into the material (Kad, Wang et al. 2010). Furthermore, variable OAF studies allow for imaging of cells by mapping fluorescence as a function of penetration depth into the material by simply adjusting the sub-critical angle of the incoming beam (Konopka and Bednarek 2008).

2.4 Single molecule tracking

The primary motivation is single molecule tracking, or the ability to follow the path of a single protein as it traverses a DNA substrate in order to examine specific interactions
that are missed by bulk assay experiments. DNA is ordinarily tangled up in the nucleus at very high concentration. A worm-like chain model can describe the DNA as a length of N polymers each of which are free to rotate under thermal mechanics (Nelson, Radosavljević et al. 2004). If the concentration of DNA is high, as it is in the nucleus, the proximity of other DNA restrict the mobility of each individual N polymers constructing the DNA, effectively restraining the available motion for each chain, and straightening the molecule. A variety of experimental TIRF setups exist in order to establish to examine dynamic protein-DNA interactions. Most of these setups involve tethering either the protein or DNA to a support in order to stabilize the system while avoiding surface adsorption to extract physiologically relevant data. A technique pioneered by Eric Greene’s lab does this by etching small channels into a silica microscope slide and occupying these regions with lipid bilayers to resemble a cell like environment (Silverstein, Gibb et al. 2014). Importantly, the head groups of these lipids have been functionalized with biotin. DNA with a biotinylated end is flown across the lipid filled channel and adheres to the surface through a streptavidin linkage. The DNA are pushed against the channel wall from the input flow and the silica slide acts as a physical barrier to prevent lipid diffusion out of the channel, creating a DNA curtain. These tethered DNA molecules are stretched out by this flow and tethered to the surface, thus allowing for simultaneous data collection of many protein-DNA interactions using TIRF. This model has control over the DNA concentration as well, which is necessary to isolate individual interactions for statistical analysis. Variants on this technique include using a ‘zig-zag’ barrier to prevent the overlap of nearby DNA molecules, anchoring the other end of the

Diffusion is a particularly interest problem of note. Many proteins are tasked with locating a specific DNA sequence or region along lengths of DNA many orders of magnitude longer than the location of interest. There are four diffusion methods utilized by proteins in order to search for targets: random three dimensional diffusion, 1D diffusion along the DNA substrate, 1D hopping in order to bypass obstacles, and intersegmental transfer (Berg, Winter et al. 1981, von Hippel and Berg 1989, Gorman and Greene 2008). Any searching mechanism that employs hopping, sliding, or intersegmental transfer is referred to as facilitated diffusion, for any of these processes increases the target association rates above that of 3D collision based diffusion. A protein can use any amalgamation of these processes in order to locate the target, and single molecule analysis is a fantastic method for testing each method. DNA substrates can be designed with protein specific sequences to test the affinity and stability of the interaction as well as the method of diffusion to locate the region
Roadblocks such as nucleosomes can be added to the DNA to in order to test the ability of specific proteins to bypass obstacles (Brown, Kim et al. 2016).

Another method of examining protein-DNA interactions, and the method used for all experiments for this thesis, is to establish DNA tightropes in a flow cell by tethering the DNA to beads (Kad, Wang et al. 2010). First, poly-L-lysine is used to coat 5 micrometer silica beads. The coated beads are washed with a 20-fold excess of DI H$_2$O and centrifuged twice at 16000G to wash the beads. They are then diluted in 100 microliters of DI H$_2$O and introduced into the flow cell for 5 minutes. The negatively charged oxygen of the carbonyl groups on the poly-L-lysine attract cations in the solution of the flow cell, which will ultimately create a salt bridge similar to that of the AFM in order to attract DNA. These beads are introduced into a closed flow cell at a concentration experimentally deduced to produce optimal coverage on the surface. Finally the beads are washed with the buffer used in the TIRF experiments to remove any weakly tethered beads.

The flowcell is constructed by drilling two holes 15 mm apart through a glass microscope slide. Plastic tubing is cut, expanded at one end using a heat gun, and pulled through the holes until taut. The tubes are then glued into place using UV curing adhesive, which also seals any air pockets between the holes and tubing. Excess tubing is cleaved away using a box cutter. A coverslip is meanwhile prepared by thoroughly bleaching and cleaning. This is followed with amine functionalization with 3-aminopropyl-triethoxysilance in 2% dry acetone and washed thoroughly with DI H$_2$O. Once the coverslips have been treated, they are sandwiched together along with PEG-SVA for ~4 hours before they are separated, washed,
and dried. The coverslip is then attached to the glass slide with adhesive tape and sealed shut. Finally the flow cell is flushed with DI H₂O.

After the beads are introduced and incubated, the flow cell are connected at one end to a syringe pump (operated at 0.15 ml/min unless otherwise specified) and at the other end to a 1.5 mL centrifuge tube that acts as a reservoir. The flow cell is again flushed with whatever imaging buffer is used for the experiment, taking precaution to avoid introducing air bubbles into the system. The DNA is then introduced into the system by injecting it into the reservoir and withdrawing the solution into the flow cell. Once inside the flow cell, the DNA is moved back and forth across the beads for ~15 minutes. The DNA sticks to the poly-L-lysine covered beads during this process, and continues in the direction of flow until it encounters another bead. If the DNA strikes another bead, it adheres again, creating a region between the two beads of taught DNA. The DNA remains taught even after the flow is shut off, allowing for observation under an absence of external flow.

In order to examine the stability and extension of DNA in the flow cell, it is necessary to evaluate DNA as an ideal polymer composed of N units with flexible linker ends allowing for random bend angles. In solution, each linker region is subject to thermal fluctuations, causing each bend angle to be a result of a random walk. The force required to extend the DNA must overcome the thermal fluctuations, and can be described according to the worm like chain model as follows

\[
F = \frac{k_B T}{4A} \cdot \left( \frac{1}{4 \left( 1 - \frac{x}{L} \right)^2} - \frac{1}{4} + \frac{x}{L} \right)
\]  

(16)
where $k_B$ is Boltzmann’s constant, $T$ is temperature, $L$ is the length of the molecule, and $A$ is the moment of inertia of the DNA (Bustamante, Marko et al. 1994). For DNA there is a predilection towards self-avoidance due to the negative phosphate backbone. This model is well established and was used as a basic framework for single-molecule studies to stretch ds DNA (Smith, Finzi et al. 1992, Bensimon, Simon et al. 1995, Cluzel, Lebrun et al. 1996, Bouchiat, Wang et al. 1999). These studies concluded that a force of 2-3 pN was sufficient to stretch ds DNA to ~90% of its contour length. The DNA can become overextended if ~60 pN forces are applied, and DNA has a theoretical breaking limit of ~500 pN. In the region between 5-60 pN, the DNA stretches but maintains its integrity and helicity, thus it is essential to be within this regime for DNA within the flow cell. The force due to two dimensional laminar flow can be estimated using the following

$$\frac{dp}{dx} = -12n \cdot \frac{V}{B^2}$$

(17)

where $dp/dx$ is the pressure drop, $n$ is the viscosity of the buffer in the flow cell, $V$ is the velocity of the flow rate (flow rate divided by the cross sectional area), and $B$ is the height of the flow cell (~120 micrometers) (Kad, Wang et al. 2010). Given the beads used are ~5 micrometers in diameter and a flow rate of 0.15 ml/min, the estimated force applied to the DNA is below 10 pN, well within the bounds for DNA stability and stretching.

Once the DNA is suspended and extended between the beads, protein complexes tagged with quantum dots or fluorescent dyes can be used to examine to protein-DNA interaction. Like the DNA curtain method, the DNA tightrope assay can also provide insight into protein diffusion capabilities and protein-DNA interactions. Examples include protein
recognition of DNA damage sites (Kad, Wang et al. 2010); telomere proteins binding strongly to telomeric regions of DNA (Lin, Countryman et al. 2014); investigation which proteins in complex process are involved with direct DNA binding (Hughes, Wang et al. 2013); and a glycosylase involved in base excision repair (Blainey, van Oijen et al. 2006).

2.5 Quantum dot properties and fluorescent dyes

Fluorescence probes are widely used in studying biological systems. However organic fluorophores photobleach fairly rapidly and have overlapping emission lines. The former matter is difficult for long temporal monitoring of a system, while the latter issue makes multi-color imaging a difficult proposition. Quantum dots serve as an alternate method of imaging biological materials and offer decreased photobleaching effects and sharp emission spectra.

Quantum dots are spherical, crystalline particles on the nanometer scale, typically composed of periodic group II-VI (ex. CdSe) or III-V (ex. InP) materials (Alivisatos, Gu et al. 2005). Photoexcitation of the quantum dot leads to the creation of an exciton, or electron-hole pair, which is confined to the volume of the quantum dot. Both the electron and the hole are trapped within the respective parabolic energy wells from the respective conduction and valence bands of the material. This occurs only in materials where the particle is smaller than the exciton Bohr radii (Wilson, Szajowski et al. 1993, Chan and Nie 1998).

\[ \delta = \frac{4E_F}{3N} \] (18)
For quantum dots, the spacing between energy levels ($\delta$ in eq. 18) depends on the Fermi energy of the metal $E_F$ and the number of electrons in the metal $N$ (Murray, Kagan et al. 2000). By controlling the size of the quantum dot it is possible to control the spacing between energy levels, and the resulting photon that is produced by excited electrons dropping down from higher energy bands (Bruchez, Moronne et al. 1998, Murray, Kagan et al. 2000). These studies also show quantum dots have broad absorption spectra, allowing for excitation by a wide range of wavelengths, and a narrow, symmetric emission spectra. Typically a layer of ZnS coats the CdSe cores to protect the core from oxidation, prevent leeching of the core into the solution, and increased luminescence (Medintz, Uyeda et al. 2005). The ZnS layer has a significantly larger band gap than the core material, effectively confining excitation to the core (Bruchez, Moronne et al. 1998).

2.6 Protein-QD conjugation methods

While the quantum dots are the source of light emission for single molecule experiments, the biological sample is the point of interest for the experiment. Thus it is necessary to attach, or conjugate, the quantum dot to either the protein or DNA for single-molecule tracking.

It is possible to functionalize quantum dots in order to improve aqueous solubility and stability with a variety of methods. Ligand exchange replaces the hydrophobic surface ligands with hydrophilic bifunctional ligands (Medintz, Clapp et al. 2003). EDC, or 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide, is condensed on the surface, allowing exposed
carboxyl groups on the quantum dot to react to amines. This method is prone to excess aggregation due to the numerous amount of available binding sites on the functionalized surface (Medintz, Uyeda et al. 2005). It is possible to directly bind proteins to the quantum dot by attaching thiolated peptides or polyhistidide (HIS) residues on the quantum dot surface corresponding to the protein. Avidin-biotin conjugation can also be used to directly relate the quantum dot to the sample in question, or through an intermediary such as an antibody (Bruchez, Moronne et al. 1998).

Our experiments utilizing cohesin molecule SA2 unless otherwise specified used streptavidin coated quantum dots to attach to a biotinylated multivalent chelator tris-nitrilotriacetic acid (BT-tris-NTA) (Reichel, Schaible et al. 2007). The histidine heads of the 6x histidine (His$_6$) tagged stromal antibody 2 (SA2) bind with to the NTA with subnanomolar affinity.

2.7 Gap DNA preparation

The functions of proteins in many important biological pathways are mediated through their specific binding to unique DNA sequences and structures. ssDNA gaps are generated during DNA replication on the lagging strand and during DNA repair. To study protein binding to ssDNA gaps, it was necessary for us to generate DNA substrates with ssDNA gaps at defined locations. To achieve this we used nickase enzyme Nt.BSTNBI to nick plasmid DNA pSCW01 at 4 sites located within 37 bp of each other. The nickase enzyme only cuts one strand of the DNA, allowing the resultant region between nick sites to
drift away thermally, leaving a region of ss DNA on the plasmid (Tessmer, Kaur et al. 2013).

In order to prevent the corresponding oligomers from re-annealing to the plasmid DNA template, we introduced complimentary ss DNA strands in 50X molar excess to the solution after the nicking process to act as an annealing template. The solution was allowed to cool before centrifugation with a 10000 MW filter in order to remove the excess oligomers in the solution. The DNA was then linearized using Scal-HF (NEB) and purified. To test the gapping efficiency we digested the gapped substrate using three enzymes (PstI, NcoI-HF, BamHI NEB) that bind in the region between nick sites of the original plasmid. Agarose gel based restriction digestion analysis showed ~85-95% of the resulting DNA was gapped.

Stretching ligated gapped DNA inside flow cells containing poly-L-lysine treated beads did not yield DNA tightropes. The solution was to ligate the standard pSCW01 linear DNA and gap the DNA inside of the flow cell. The procedure follows the same basic steps: (i) introduce nickase enzyme Nt.BstNBI into the flow cell and nick the pSCW01 tightropes (at 55°C); (ii) flush the flow cell with buffer solution and heat to remove the regions between the nicked DNA (at 60°C); (iii) flush the system with 300 µl 1M MgCl₂ to remove any nickase enzyme from the DNA as well as any excess ss- and dsDNA; (iv) flush the system with 3 ml of buffer to re-establish biologically relevant buffer conditions within the flow cell.

In our case we combined the first two steps since the annealing temperature was proximate to the temperature required for the nickase enzyme activation. The DNA tightropes treated with the aforementioned DNA gapping procedures are refractory to restriction enzyme digestion targeting the regions between the nicked sites, confirming the presence of ssDNA regions.
2.8 Specifications of in lab fluorescence microscope

Fluorescence imaging was carried out with an inverted microscope (Nikon Ti-E) equipped with a solid state laser (20 mW Sapphire DPSS), a 100X objective with a numerical aperture of 1.49 (APO TIRF, Nikon) and an electron multiplied (EM) CCD camera (iXon DU897, Andor Technology).
CHAPTER 3: Single-particle tracking of quantum-dot labeled proteins on DNA

3.1 Introduction

While images gathered from the AFM imaging in air are static, real-time fluorescence microscopy imaging enables tracking of individual protein diffusion, binding, and unbind from DNA tightropes. The following section will illustrate how tracking quantum dot labeled proteins on DNA tightropes provides insight into their biological functions.

3.2 Improved positional accuracy through Gaussian fitting

As mentioned previously, in our fluorescence imaging experimental setup, the signal is obtained by capturing photons emitted by quantum dots with an EMCCD camera. However barriers such as the diffraction limit and the binning of captured photons into pixels place restrictions on the quality of image obtained from the experiment. While the resolution of the image is always lost, it is possible to obtain the true center of the object by using Gaussian fits. Begin with a one-dimensional pixel array where $i$ and $j$ are the indices of the pixel, $S_{ij}$ is the intensity of the corresponding pixel, and $N_{ij}$ is the expected value of a Gaussian fit with width $s$ centered at $x_o, y_o$ (in pixel units) (Thompson, Larson et al. 2002). $N_{ij}$ would have the following form:

$$N_{ij} = \exp \left( - \frac{(i - x_o)^2}{2s^2} - \frac{(j - y_o)^2}{2s^2} \right) \quad (19)$$

The goal is to minimize the variance in the sample of a normally distributed sample, which can be accomplished by minimizing the Chi-squared sum of the Gaussian fit.
\[ 0 = \frac{d}{dx} \sum (S_{ij} - N_{ij})^2 \]  

(20)

\[ 0 = \sum S_{ij}(i - x_0)N_{ij} - \sum (i - x_0)N_{ij}^2 \]  

(21)

The second term in eq. 21 has odd symmetry, and is approximately zero, yielding

\[ x_0 = \frac{\sum iS_{ij}N_{ij}}{\sum S_{ij}N_{ij}} \]  

(22)

for locating the center of the particle. Typically two Gaussian fits are used: the first is a least-squares fit to a Gaussian distribution taking into account background noise as well as photon-counting noise of the system (Cheezum, Walker et al. 2001); the second ignores the photon-counting noise and fits the intensity signal only according to the background (Thompson, Larson et al. 2002). The following expression determines the uncertainty of the position due to photon counting noise and pixilation noise

\[ \sigma_i = \sqrt{\frac{s^2 + \frac{a^2}{12} + \frac{4s^2b^2\sqrt{\pi}}{aN^2}}{N}} \]  

(23)

where \( s \) is the standard deviation of the intensity fit, \( a \) is the pixel size, \( b \) is the standard deviation of the background, and \( N \) is the number of collected photons (Thompson, Larson et al. 2002, Selvin, Lougheed et al. 2007, Kad, Wang et al. 2010).

3.3 Methods of data analysis

Once the position of the quantum-dot protein complex in each video frame is determined, the next step is to extract statistical information from tracking the particle position. A simple model to apply is the Brownian motion of a particle undergoing one
30 dimensional diffusion. It is worth noting that this model of random diffusion doesn’t apply to motor proteins as they undergo directed diffusion. The one dimensional probability distribution function is as follows

\[ P(x, t) = \frac{1}{\sqrt{4\pi Dt}} \exp\left(-\frac{(x - x_0)^2}{4Dt}\right) \]  

where \( x \) is the particle position at time \( t \), \( x_0 \) is the particle’s original position, and \( D \) is the diffusion constant in S.I. units of \( \text{m}^2\text{s}^{-1} \). We also require the mean squared displacement (MSD), which is given by eq. 25

\[ \text{MSD} = \langle (x(t) - x_0)^2 \rangle \]  

\[ \langle (x - x_0)^2 \rangle = \langle x^2 \rangle + x_0^2 - 2x_0\langle x \rangle \]  

By expanding out the ensemble average we obtain eq. 26 without the explicit time dependence. At this point there are two options to obtain the MSD: 1) use brute calculating force to obtain \( \langle x^2 \rangle \) and \( \langle x \rangle \); 2) use the moment generating function to obtain the MSD in terms of parameters from the probability density function (eq. 24). Let us generate an expression for the characteristic function \( G(k) \) and expand the exponential

\[ G(k) = \langle e^{ikx} \rangle = \int e^{ikx} P(x, t \mid x_0) \, dx \]  

\[ G(k) = \sum_{m=0}^{\infty} \frac{(ik)^m}{m!} \mu_m \]  

Taking a natural log of eq. 28 provides the cumulant generating function
\[
\ln(G(k)) = \sum_{m=1}^{\infty} \frac{(ik)^m}{m!} K_m
\]

where \(K_m\) is the mth cumulant of the position \(x\). The cumulants are related to the moments of the system, with the first cumulance equal to the mean \(\langle x \rangle\) and the second cumulance equal to the variance of the system \(\langle x^2 \rangle\). By substituting eq. 24 into eq. 27 and completing the integration, the following is obtained

\[
G(k) = \exp(ikx_o - k^2Dt)
\]

After taking the natural log of eq. 30, one can compare terms of \(m\) with eq. 29 in order to obtain values for the first and second cumulants. Thus it is possible to also express the MSD in the following

\[
\text{MSD} = \langle (x(t) - x_o)^2 \rangle = 2Dt
\]

where \(t\) is the time interval spacing between particle instances (more on this shortly). Eq. 31 was derived assuming the particle undergoes random diffusive behavior, however this is not always the case. Particles can exhibit sub-diffusive behavior if they are confined or pause during the random walk process, and super-diffusive behavior if their motion is directed (such as with protein motors). Therefore we must account for this with the adjustment of a diffusive parameter to eq. 31, yielding eq. 32

\[
\text{MSD} = \langle (x(t) - x_o)^2 \rangle = 2Dt^\alpha
\]

We can also expand eq. 25 to calculate the MSD as a function of the time interval (frame rate) as follows
\[ MSD(n\Delta t) = \frac{1}{N-n} \sum_{i=1}^{N-n} [(x_{i+n} - x_i)^2 + (y_{i+n} - y_i)^2] \]

where \( N \) is the total number of frames of the particle path, \( n \) is the number of frames within each time interval, \( \Delta t \) is the time between frames, and \( x_i \) and \( y_i \) are the centered positions as determined by the Gaussian fit outlined in the previous section. From eq. 32 it is possible to generate a relationship for MSD vs. time in order to obtain values for the diffusion constant \( D \) and for the alpha factor \( \alpha \) (Nechyporuk-Zloy, Dieterich et al. 2008, Fan, Sheen et al. 2015).

A common practice is to assume that the motion of the tracked particle is Brownian, thus the \( \alpha \) term is set to 1 when plotting MSD vs. time plots in order to calculate diffusion constants (Kad, Wang et al. 2010, Lin, Countryman et al. 2014, Nenninger, Mastroianni et al. 2014). However particles bound by barriers or stuck in energy wells have sub-diffusive behavior and \( \alpha \) is less than 1, thus \( \alpha \) also needs to be a free parameter (Choi, Margraves et al. 2007). Superdiffusive molecules have an \( \alpha \) greater than 1 (Nechyporuk-Zloy, Dieterich et al. 2008, Dunderdale, Ebbens et al. 2012).

Our experiments capture typically 2-4 minutes of data with a 50 ms frame time for individual protein molecules. When a protein-quantum dot complex alternates between different binding modes, it becomes difficult to isolate individual diffusive modes and to calculate their dwell times. Building on work from the Wallace group, I developed a sliding-window method to isolate local diffusion modes that are all averaged together in bulk MSD calculations (Espenel, Margeat et al. 2008). Error propagates quickly in MSD calculations, thus usually only the first 10-30% of data points are used for fitting the diffusion constant.
and alpha factor (Espenel, Margeat et al. 2008, Lin, Countryman et al. 2014). It is vital that in reducing the window of available data points for MSD analysis [eq. 33 N=10 (Espenel, Margeat et al. 2008) paper] that one does not examine merely nearest neighbor stepping dynamics.

Our sliding window analysis uses a 40 frame window (a 50 ms frame rate over 2 seconds) to calculate interval-based diffusion constants (Dint) over the full data sets (2 minutes unless otherwise stated). Each calculation is based on shifting the calculation window by 1 frame, thus there is an overlap of data with neighboring Dint values to improve the continuity of the calculation. The Dint, α, and r² values are calculated based on fitting MSD vs. time data based on eq. 33, where N = 3. The fit is then re-calculated for N=4 and the r² parameters are compared to each other. If the quality of fit is improved, the program continues to increase N until it reaches the value of N=12. If the r² value decreases compared to the N-1 fit or if the r² value is below 0.8, the value of 1 is added to a counter value P. P resets if the r² value improves. This counter value returns the fit values for the N-P fit if the counter value reaches a value of 2 or higher (values of 3 can be achieved for excessively poor consecutive fits). This iterative process allows us to optimally examine short range diffusive modes that would be otherwise lost in bulk MSD analysis while also maximizing the collectable data according to the quality of fit. Additionally it is entirely automated, removing any human bias outside of initial conditions (such as r² values). This method was used to analyze behavior of WT and mutant cohesin SAI and SA2 proteins.
3.4 Prediction of diffusion constants and stepping rates based on Stokes-Einstein relation

We attach quantum dots to proteins in order to observe the diffusion along DNA tightropes. This means that forces acting on the protein within the flow cell, such as viscous drag forces, are also affecting the quantum dot it is attached to. Therefore, it is necessary to prove that proteins retain their function and do not have their biological function impaired by the conjugation of quantum dots. Studies have shown that quantum dot conjugation to DNA repair proteins does not inhibit the ability of the protein to diffuse on DNA curtains (Gorman, Fazio et al. 2010, Gorman and Greene 2013). For our system, a flexible His linker conjugates the quantum dots to the proteins of interest. The flexible linker can be modeled as a Hookean spring, and the DNA-protein interaction can be modeled as a sinusoidal interaction potential (Bonnet and Desbiolles 2011). In this model, the protein and the label have unique viscous drag coefficients. As the protein slides along the DNA, it is bound to the label by the flexible linker with spring constant $k$. If the protein is not extremely mobile ($D << 1 \mu m^2/sec$), and if the relaxation time of sliding across the DNA ($\tau_s$) is much greater than the relaxation time of the linker ($\tau_l$), the motion of the protein is decoupled from the Brownian motion of the label. Both of these assumptions fall into the purview of our experiments. Thus the model for the protein and quantum dot label contains a flexible linker, rather than a stiff rod-like system where the label severely alters the protein diffusion capabilities.

The expected diffusion constants for protein alone can be estimated based on the Stokes-Einstein relation. Based on recent measurements, the hydrodynamic radii of red (655 nm) and green (565 nm) Sav-QDs are assumed to be 11.5 and 9.5, respectively (Arnspang,
Brewer et al. 2012). The estimated radius of free TRF2 is estimated as 10 nanometers based on crystal structure of the Myb type and dimerization domains (Fairall, Chapman et al. 2001, Court, Chapman et al. 2005). The expected upper limit for diffusion constants for a single red QD labeled TRF2 sliding on DNA without curvilinear motion is $17.8 \mu m^2/s$. This is 187-468 fold higher than the measured diffusion constants for TRF proteins (between $3.8e^{-2}$ and $9.5e^{-2} \mu m^2/s$). This calculation supports the notion that TRF proteins slide on DNA while rotating, following along the DNA phosphate backbone.

Assuming protein rotation around the DNA helix, the expected upper limits for diffusion constants are based on a modified version of the Stokes-Einstein relation

$$D_{1,cal} = \frac{K_B T}{6\pi \eta a [1 + \left(\frac{4}{8}\right) \cdot (2\pi)^2 \cdot \left(\frac{a}{3.4 \cdot 10^{-9}}\right)^2]}$$ (34)

where $\eta$ is the viscosity of the medium, $a$ the radius of the particle, $K_B$ is Boltzmann’s constant, and $T$ is the temperature.

The expected upper limits for diffusion constants and stepping rates for TRF2 proteins (with one red QD) with rotation-coupled diffusion are 0.021 $\mu m^2/s$ and 365492 steps/s, respectively. These numbers correspond to a diffusion rate of 0.42 $\mu m^2/s$ and a stepping rate of 7315292 steps/s without QDs.

The stepping rate is calculated by assuming the diffusion constant to occur as a series of steps of a single base pair using the following relationship

$$k = \frac{\text{steps}}{\text{sec}}(n) = \frac{2D}{l_{bp}^2}$$ (35)
where $k$ is the stepping rate, $D$ is the diffusion constant, and $l_{bp}$ is the length of one ds DNA base pair (Hughes 1995).

3.5 Prediction of additional energy barriers

Furthermore, reduced diffusion constants relative to predicted values for unbiased 1-D diffusion indicates diffusion barriers present along the DNA. These barriers can include DNA target sequences, obstructions, or alternative DNA structures (G quadruplex, nicks, gaps, legions). It is possible to calculate the energy barriers to protein stepping based on the difference between the observed and predicted stepping rates and by applying the Arrhenius equation (Hughes 1995).

$$ k = \exp\left(-\frac{E_A}{K_B t}\right) $$  \hspace{1cm} (36)

$$ E_A = -\ln(k) \cdot K_B t $$  \hspace{1cm} (37)

Then the additional energy barrier between two different regions of DNA can be calculated as follows: (Kad, Wang et al. 2010)

$$ E_{A1} - E_{A2} = -\ln\left(\frac{k_1}{k_2}\right) \cdot K_B t $$  \hspace{1cm} (38)

The relative free binding energy between the two regions can be defined as follows:

$$ \Delta G_{bind} = K_B T \cdot \ln\left(\frac{k_1}{k_2}\right) $$  \hspace{1cm} (39)

where $k_1$ and $k_2$ are the equilibrium association constants for the two respective DNA substrates. If both substrates share the same equilibrium association constants, the Gibbs energy can be defined in terms of protein lifetime on the DNA substrates.
\[ \Delta G_{\text{bind}} = K_B T \cdot \ln\left(\frac{\tau_1}{\tau_2}\right) \] (40)

where \( \tau_1 \) and \( \tau_2 \) are the lifetimes on the respective DNA substrates.
CHAPTER 4: Cohesin SA2 is a sequence independent DNA binding protein but recognizes DNA ends and gaps

4.1 Cohesin background

The cohesin complex plays a role in a plethora of genome maintenance pathways. In eukaryotes the cohesin complex is integral for the proper segregation and alignment of sister chromatid (Michaelis, Ciosk et al. 1997, Remeseiro, Cuadrado et al. 2013). Improper segregation of chromosomes can quickly lead to aneuploidy or genetic diseases (Kim, Kim et al. 2012, Taylor, Platt et al. 2014). Cornelia de Lange syndrome is caused by cohesin disruption, and results in altered transcriptional regulation while maintaining standard chromatid separation, indicating that mutations or defects of Cohesin elements affect a variety of processes independent of DNA cohesion (Peters, Tedeschi et al. 2008).

The core cohesin complex consists of a tripartite ring structure with two arm proteins Structural Maintenance of Chromosomes 1 and 3 (Smc1 and Smc3) and a bridge protein Rad21 in addition to Stromal Antibody 1 or 2 (SA1 or SA2) in vertebrates. This ring architecture makes Cohesin an ideal candidate for maintaining cohesion between sister chromatid separation during mitosis, or regulating transcription by establishing and maintaining chromatin loops (Remeseiro and Losada 2013). Cohesin can contain either SA1 or SA2, but not both. Mutation in any of these core proteins or their regulator genes can lead to a variety of human diseases, grouped as cohesinopathies, and cancer. SA2 is frequently found to be either mutated, depleted, or otherwise obstructed in tumors (Kim, Kim et al. 2012, Taylor, Platt et al. 2014). While other proteins such as sororin, Esco1, and Esco2 are

4.2 Summary

Our paper on the DNA binding properties of cohesin protein SA2 is currently in submission. The paper in its current form is available in the appendix. Below there will be a brief summary of the paper.

Our group recently published results showing that the SA1 protein bound to ds DNA by itself and displayed a weak affinity for telomeric portions of DNA. Yet the DNA binding properties of SA2 are as of yet unknown. Using a combination of single-molecules studies with Atomic Force Microscopy (AFM) and fluorescence imaging, we were able to show that SA2 is also able to bind to ds DNA directly. AFM scans were performed on DNA substrates containing telomere regions, centromere regions, or random genomic sequences. SA2 showed no predilection for either telomeric or centromeric regions of DNA. Our initial theory was that SA2 would have a similar binding mechanism to centromere DNA that SA1 showed for telomeric DNA, yet this was not the case. Fluorescence studies carried out with
the same DNA substrates indicated that SA2 had no change in diffusive behavior in the presence of telomere, centromere, or genomic DNA sequences.

Curiously, our AFM results saw that SA2 showed a strong affinity for ends of DNA regardless of the sequence, length, or presence of overhang (Figure 3: SA2 does not show binding preference for telomeric or centromeric DNA sequences, but recognizes DNA ends.). Given the alternative roles of cohesin outside of cohesion, we postulated that SA2 had an affinity for regions of DNA damage. We tested the binding of SA2 to new DNA substrates that contained either nicks or gaps using AFM. SA2 had no binding affinity towards a single nick site, however severely nicked DNA did decrease the overall end binding of the SA2 protein. Introduction of a single stranded region of DNA 37 base pairs long significantly shifted the binding affinity of SA2. SA2 showed a significant affinity for the gapped region of the gapped DNA in addition to strong end binding (Figure 4: SA2 specifically binds to ssDNA gaps.). Fluorescence studies conducted with the ligated gapped DNA substrate revealed periodic binding along the DNA tightrope at intervals corresponding to the DNA length. SA2 binding to the gapped substrate was overwhelmingly static. There was a significant difference in the diffusion constant and alpha factor for SA2 on the gapped substrate as compared to all other DNA substrates (Figure 5: SA2 switches between searching and recognition modes on DNA tightropes containing ssDNA gaps.).

To conclude if SA2 exhibited a unique diffusive search mode along gapped DNA, we used a 2 second sliding window analysis (see section 3.3 for more details) to extract instant diffusion constants ($D_{int}$) for each mobile instance. Mobile SA2 on gapped DNA had a much
higher predilection for extremely low values ($D_{\text{int}} < 10^{-4} \, \mu\text{m}^2/\text{sec}$), corresponding to regions where the SA2 encountered gap regions of DNA (Figure 6: Comparison of the distribution of interval based diffusion constants ($D_{\text{int}}$) for SA2 on different DNA tightropes.). Unlike SA1 on telomere DNA tightropes, SA2 showed no additional diffusive modes in the presence of gapped DNA. Our work has shown that SA2 is successfully able to bind to ds and ss DNA independent of other proteins or cofactors and shows a high affinity for DNA ends and gaps. Thus SA2 could act as a structural anchor for the cohesin complex to successfully bind to the DNA.

Specific contributions of co-authors: Parminder Kaur assisted in generating AFM images. LabView code used to analyze macroscopic MSD values was created by Jiangguo Lin. Dianwen Wang and Parminder Kaur assisted in protein purification and gel analysis. Xuechun Wang assisted with protein position analysis on AFM data sets. Christian White assisted with protein purification. All cohesin proteins were provided by Yizhi Jane Tao. Protocol for gapping DNA adapted from Ingrid Tessmer. Hong Wang designed the research. All authors contributed to the interpretation and manuscript preparation.
Figure 3: SA2 does not show binding preference for telomeric or centromeric DNA sequences, but recognizes DNA ends.

(A) Three DNA substrates used for AFM imaging. (B-C) Representative AFM images of SA2 on the centromeric (B, Cen-mid) or telomeric (C) DNA substrates. (D) Position distributions of SA2 on DNA substrates containing telomeric (T270 N = 283), centromeric
sequences close to one end (Cen-end, N = 275), or in the middle (Cen-mid, N = 298). Each data set was from three independent experiments. The error bars represent SEM.
Figure 4: SA2 specifically binds to ssDNA gaps.

(A) Generation of linear gapped DNA substrate. Gapped DNA was created using pUC19 derived from pSCW01 plasmid (2030 bp) that contains 4x Nt.BstNBI nicking sites. After restriction digestion, the resulting ssDNA gap is 37 nt long and located 470 nt (23%) from one end of the linear DNA fragment. (B) Representative AFM images of the full length SA2 complex binding to the linear gapped DNA substrate. The contour length of the linear gapped
DNA was measured as $L_c = 622.48 \pm 41.3$ nm. (C) Statistical analysis of the position distribution of the full length SA2 complex on the linear gapped ($N = 251$) and non-gapped ($N = 201$) DNA, as well as full length SA1 on the gapped DNA substrate ($N = 295$). Each data set was obtained from at least 2 independent experiments. The error bars represent SEM. The solid line is Gaussian fit to the data with the peak centered at 25.1% $\pm$ 1.1%. 
Figure 5: SA2 switches between searching and recognition modes on DNA tightropes containing ssDNA gaps.

(A) Comparison of SA2 diffusion constants and alpha factors on gapped DNA ($N = 28$), nicked $\lambda$ DNA ($N = 20$), and non-nicked $\lambda$ ($N = 20$) DNA. Final SA2 concentration was at 0.5 nM in the flow cell. (B) Kymographs of SA2 showing SA2 alternating between 1-D diffusion and stable binding on gapped DNA tightropes. (C) Percentages of time windows
(40 frames or 2 s) with $D_{\text{int}}$ less than $10^{-4}$ displayed by mobile SA2 on gapped, $\lambda$, non-gapped control, centromeric, and telomeric DNA tightropes.
Figure 6: Comparison of the distribution of interval based diffusion constants ($D_{\text{int}}$) for SA2 on different DNA tightropes.

$D_{\text{int}}$ distributions for all mobile SA2 molecules observed on centromeric (A, $N = 48$), telomeric (B, $N = 52$), $\lambda$ (C, $N = 41$), and gapped DNA (D, $N = 31$).
CHAPTER 5: Functional interplay between SA1 and TRF1 in telomeric DNA binding and DNA-DNA pairing

5.1 Background

Proper chromosome alignment and segregation during mitosis depends on cohesion between sister chromatids. In vertebrates, the core cohesin complex consists of a tripartite ring assembled by Smc1, Smc3, Rad21, and either SA1 or SA2 (Remeseiro and Losada 2013). The cohesin complex is distributed along centromere and telomere regions of chromosome arms (Wendt, Yoshida et al. 2008). Telomeres are nucleoprotein structures that prevent the degradation or fusion of linear chromosome ends by preventing them from activating the DNA damage response and double-strand DNA break repair mechanisms (Blackburn 2005, Palm and de Lange 2008, Holohan, Wright et al. 2014, Lin, Kaur et al. 2014). Surprisingly, cohesin rings do not play a major role in sister telomere cohesion. This role is replaced by SA1 and telomere binding proteins TRF1 and TIN2, however the DNA binding mechanism of SA1 and the unique telomere cohesion mechanism are poorly understood. Using single-molecule fluorescence, we discovered that SA1 displays two-state binding on DNA: searching via free 1-D diffusion and recognition through sub diffusive sliding at telomere regions. SA1 and TRF1 together form longer DNA-DNA pairing tracts than with TRF1 alone, as revealed by atomic force microscopy. These results suggest that at telomeres cohesion relies on the molecular interplay between TRF1 and SA1 to promote DNA-DNA pairing, while along chromosomal arms the core cohesin assembly might also depend on SA1 1-D diffusion on DNA and sequence specific DNA binding.
5.2 Summary

Recently we demonstrated that the N-terminal domain of SA1 (1-72 AA, SA1-N) binds to a DNA substrate containing telomere sequences (Bisht, Daniloski et al. 2013). To understand the SA1 DNA binding mechanism, we obtained full-length His- and Flag-tagged SA1 proteins. We carried out binding of SA1 on a telomeric DNA substrate containing 270 TTAGGG repeats (T270, 5.4 kb) and a control substrate (3.8 kb) containing only the non-telomeric (genomic) DNA sequences from T270. AFM imaging indicated that a higher percentage (41.9%) of SA1 bound at the telomeric region on the T270 DNA substrate compared to the same locations along the genomic DNA substrate (27.0%).

To understand how SA1 dynamically achieves DNA binding specificity for telomeric sequences, we used oblique angle fluorescence microscopy imaging of QD-labeled proteins on T270 DNA tightropes containing alternating telomeric and genomic regions (see previous chapters or Materials and Methods section of paper in appendix for more detail). The binding of SA1-QDs molecules to DNA was long lived, with 78.5% (N=107) of the complexes remaining on T270 tightropes after 2 minutes. SA1 on T270 DNA revealed two populations: static and mobile. While some mobile SA1 molecules displayed free 1-D diffusion on T270 DNA throughout the entire observation period (2 min), a subpopulation of mobile SA1 molecules alternated between periods of slow and fast diffusion. Some SA1 molecules repeatedly slowed down at the same regions along T270 DNA tightropes, manifested as distinct peaks in a position histogram of SA1 along DNA. These distances were consistent
with the boundaries of the telomeric region (1.6 kb) and the spacing between two adjacent telomeric regions on T270 (5.4 kb).

To further compare the dwell times of slow diffusion events displayed by mobile SA1 on different DNA substrates, I developed the ‘sliding window’ (40 frame, 2s) MSD analysis method to calculate a time interval-based diffusion constant (\(D_{\text{int}}\), bottom panels in Fig. 1C and 1D). Distinct from static (< 0.5 \(\times 10^{-3}\) \(\mu\text{m}^2/\text{sec}\), Figure 7: Full length SA1 alternates between fast and slow 1-D diffusion on T270 DNA tightropes.A) and fast free diffusion modes (> \(10^{-1}\) \(\mu\text{m}^2/\text{sec}\), Figure 7: Full length SA1 alternates between fast and slow 1-D diffusion on T270 DNA tightropes.B), this analysis indicated that mobile SA1 molecules with fast and slow diffusion on T270 show a distinct peak at \(\sim10^{-3}\) \(\mu\text{m}^2/\text{sec}\) (Figure 7: Full length SA1 alternates between fast and slow 1-D diffusion on T270 DNA tightropes.C). Therefore, for calculating dwell times on DNA tightropes we used \(D_{\text{int}}\) value of 5.0 \(\times 10^{-3}\) \(\mu\text{m}^2/\text{sec}\) as the threshold to identify slow diffusion events. This \(D_{\text{int}}\) based analysis showed that the dwell times of individual SA1 slow diffusion events on T270 (1.17 s) are significantly (p < 0.05) longer than on genomic (0.80 s) and centromeric (0.79 s) DNA. The percentage of mobile SA1 molecules showing long slow diffusion events (\(D_{\text{int}} < 5.0 \times 10^{-3}\) \(\mu\text{m}^2/\text{sec}\) for longer than 2.2 s) on T270 DNA (51.2%) was at least 2-fold higher than on genomic (17.6%) or centromeric DNA (24.1%). SA1 spent a significantly larger percentage of time (24.4%) in the slow diffusion mode on T270 DNA than on genomic (3.2%) and centromeric DNA (6.3%). SA1 also displayed significantly (p < 0.005) slower diffusion constants on T270 DNA (0.04 ± 0.01 \(\mu\text{m}^2/\text{sec}\) in comparison with DNA substrates contain
genomic (0.14 ± 0.03 µm²/sec) or centromeric (0.11 ± 0.02 µm²/sec) DNA sequences. Additionally the alpha factor for SA1 on T270 was significantly smaller than on genomic DNA, which suggests protein pausing amid free diffusion at telomere sequences.

SA1 contains a unique AT-hook motif at its N-terminal domain which is not present on SA2 (Bisht, Daniloski et al. 2013). To determine whether or not SA1 slow diffusion events depend on its unique N-terminal domain, we purified SUMO-tagged SA1 N-terminal fragment (SA1-N). Analysis of the binding position of SA1-N from AFM images revealed that SA1-N binds specifically to telomere regions. Consistent with results from AFM imaging, incubation of SA1-N-QDs with T270 DNA tightropes resulted in substantial DNA binding. A significantly higher percentage of SA1-N molecules (74.3%) showed slow diffusion events (dwell time > 2.2 s) on T270 DNA than on genomic DNA (38.6%).

SA1 interacts directly with TRF1 through its N-terminal domain (Canudas, Houghtaling et al. 2007). To evaluate how TRF1 affects the dynamics of SA1 on DNA, we directly images their interactions. Flag-SA1 and His-TRF1 proteins were orthogonally conjugated with red Ab-QDs and green SAv-QDs via antibody sandwich and BT-tris-NTA linkage strategies respectively. A higher population of dual-colored SA1-TRF1-QD complexes (~60%) were static than single-colored SA1-QD alone (~35%) on T270 DNA tightropes. SA1-TRF1 complexes exhibited significantly reduced diffusion ranges compared to SA1 alone. Furthermore, the distance between dual-color QD-labeled SA1-TRF1 complexes and green QD-labeled TRF1 was consistent with that of SA1-TRF1 complexes binding to telomeric regions.
Consistent with previous results, we observed that TRF1 molecules formed protein tracts that mediate DNA-DNA pairing on T270 DNA (Griffith, Bianchi et al. 1998). In the presence of both TRF1 and SA1 the average SA1-TRF1 mediated DNA-DNA pairing tract length increased significantly. The location of SA1 on TRF1-mediated DNA-DNA pairing tracts was random.

**Specific contributions provided by the author:** I created and implemented the code used for analyzing all dwell time and $D_{\text{int}}$ measurements for SA1 and TRF1 proteins and their corresponding mutants or fragments. Figures included in this dissertation from the published work are direct contributions from the author. Examples of $D_{\text{int}}$ vs. time plots or group $D_{\text{int}}$ histograms can be seen in figures from this chapter.
Figure 7: Full length SA1 alternates between fast and slow 1-D diffusion on T270 DNA tightropes.

(A) Ligated T270 DNA substrate (top panel) and the DNA tightrope assay setup (bottom left panel). (B-D) Dynamics of full length Flag-SA1 on T270 DNA tightropes. Kymographs of SA1 molecules being static (B), showing free 1-D diffusion (C), and alternating between fast and slow 1-D diffusion (D) on T270 DNA. Scale bars (y-axis): 1 µm. Equimolar concentrations of red (655 nm) and green (565 nm) QDs were used in SA1 conjugation. The bottom panels in C and D show corresponding plots of diffusion constants ($D_{int}$) based on the
40-frame sliding window (2 s) MSD analysis. Each frame is 50 ms. The histogram on the right side of the top panel in D shows the position distribution of SA1 along the T270 DNA tightrope. (E) The distribution of pair-wise distance between nearest SA1 slow diffusion position (N = 22).
Figure 8: Comparison of the distributions of interval based diffusion constants ($D_{\text{int}}$) for SA1 TRF1.

(A-C) Distributions of $D_{\text{int}}$ for individual SA1 molecules on T270 DNA tightropes with static (A, kymograph in Fig. 1B), free diffusion (B, kymograph in Fig. 1C), and alternation between fast and slow 1-D diffusion (C, kymograph in Fig. 1D). The SA1 molecules with
alternation between fast and slow diffusion show a distinct peak at ~1.0 X 10^{-3} \text{ \mu m}^2/\text{sec}. (D-F) Distributions of D_{int} for all mobile SA1 molecules observed on T270 (D), genomic (E), and centromeric (F) DNA tightropes. On T270 DNA, SA1 molecules alternate between fast and slow diffusion with 46% of D_{int} values less than 10^{-2} \text{ \mu m}^2/\text{sec}. In comparison, on genomic and centromeric DNA fast diffusion dominates with only 25% and 18% of D_{int} values less than 10^{-2} \text{ \mu m}^2/\text{sec}, respectively. (G) The distribution of D_{int} for all (N = 29) mobile TRF1 molecules on T270 reported in a previous study (Lin, Countryman et al. 2014). On T270 DNA, the majority (84%) of TRF1 D_{int} values are consistent with slow diffusion (< 10^{-2} \text{ \mu m}^2/\text{sec}).
CHAPTER 6: Telomere maintenance pathway

6.1 Introduction to telomeres

Telomeres play important roles in maintaining the stability of linear chromosomes (de Lange 2002, Cech 2004, Blackburn 2005, de Lange 2005, Sfeir 2012). The telomeric structure allows a cell to distinguish between natural chromosome ends and double-stranded DNA breaks. As such, telomeres prevent the inappropriate activation of DNA damage signaling pathways, which can lead to cell cycle arrest, senescence, or apoptosis (Smogorzewska and de Lange 2004). Loss of telomere function can activate DNA repair processes, leading to nucleolytic degradation of natural chromosome ends and end-to-end fusions. Telomere dysfunction and associated chromosomal abnormalities have been strongly associated with age-associated degenerative diseases and cancer (Wright and Shay 2000, Maser and DePinho 2002). Great progress has been made in the last 20 years in understanding telomere biology in model systems, including ciliates, yeast, drosophila, plants, and mouse (Giraud-Panis, Pisano et al. 2010, Lewis and Wuttke 2012, Wellinger and Zakian 2012).

In a typical human somatic cell, the length of tandem repeats of telomeric DNA TTAGGG is ~2-15 kb with a 3’ overhang of ~100-200 nt (Makarov, Hirose et al. 1997, Wright and Shay 2000). This 3’-overhang serves as a substrate for the reverse transcriptase telomerase, which replicates the telomeric sequence by using an internal RNA subunit as a template to direct DNA synthesis (Cech 2004, Blackburn 2005, Nandakumar and Cech
2013). A specialized protein complex, called shelterin or telosome, binds to and protects telomeres at chromosome ends (Verdun and Karlseder 2007). In humans, this complex consists of six core proteins: duplex TTA-GGG repeat binding factor-1 (TRF1) and -2 (TRF2), the single-stranded telomeric DNA binding protein protection of telomeres-1 (POT1), TRF1-interacting nuclear protein 2 (TIN2), POT1- and TIN2-organizing protein (TPP1), and transcriptional repressor/activator protein RAP1 (Cech 2004, Songyang and Liu 2006, Verdun and Karlseder 2007). The RPA-like CTC1-STN1-TEN1 complex binds to ssDNA and protects telomeres independently of the POT1 protein, and acts as a terminator of telomerase (Miyake, Nakamura et al. 2009, Chen, Redon et al. 2012). Shelterin proteins interact with numerous protein factors, including proteins involved in DNA recombination and repair, such as ERCC1-XPF, WRN, BLM, and DNA-PK (Zhu, Niedernhofer et al. 2003, Opresko, Otterlei et al. 2004, Verdun and Karlseder 2007, Bombarde, Boby et al. 2010). Adding to the complexity of telomere structures, telomeric repeat-containing RNA (TERRA) was identified as an integral part of telomeric heterochromatin (Luke and Lingner 2009, Feuerhahn, Iglesias et al. 2010). TERRA is associated with TRF2 and a large number of RNA-binding proteins, and is implicated in telomere structural maintenance and heterochromatin formation (Deng, Norseen et al. 2009, Lopez de Silanes, Stagno d'Alcontres et al. 2010).

Telomeres can adopt different types of open or closed (capped) conformation. Telomeric DNA from humans, as well as from several other organisms, can be arranged into T-loops, in which the 3’ overhang invades the upstream double-stranded region (Griffith,
Telomeres can be maintained by the recombination-dependent alternative lengthening of telomeres (ALT) pathway in telomerase-negative tumors. The ALT pathway is accompanied by the generation of duplex or single-stranded DNA circles formed from telomeric repeat sequences (t-circles) (Tomaska, Nosek et al. 2009). G-rich sequences have been shown to form discrete four-stranded structures termed G-quadruplexes in vitro (Fry 2007). Studies using an engineered, structure-specific G-quadruplex antibody provided evidence that G-quadruplex DNA exists in telomeres in vivo (Maizels 2006, Biffi, Tannahill et al. 2013, Lam, Beraldi et al. 2013). Stable G-quadruplex DNA plays important roles in the regulation of telomere extension and organization, as well as pairing of homologous chromosomes (Fry 2007).

TRF1 and TRF2 share 30% homology, bind to duplex telomeric DNA with high affinity, and are essential for the maintenance of functional telomeres (Court, Chapman et al. 2005). Together with tankyrase, TRF1 functions as a negative regulator of telomere length: overexpression of TRF1 leads to telomere shortening (van Steensel and de Lange 1997). TRF1 promotes efficient replication of TTAGGG repeats and prevents replication fork stalling at telomeres (Sfeir, Kosiyratrakul et al. 2009). TRF2 is also involved in telomere length regulation (Smogorzewska, van Steensel et al. 2000), and is crucial in capping and protecting chromosome ends (Ancelin, Brun et al. 1998, de Lange 2002). Removal of TRF2 from the telomeres leads to loss of the 3’ overhang, covalent fusion of telomeres, and induction of ATM and p53 dependent apoptosis (van Steensel, Smogorzewska et al. 1998,
Karlseder, Broccoli et al. 1999). Overexpression of TRF2 in telomerase-negative cells prevents short telomeres from fusing and delays the onset senescence (van Steensel, Smogorzewska et al. 1998). Furthermore, TRF2 plays important roles in the assembly of telomeric chromatin (Benetti, Schoeftner et al. 2008). Importantly, post-translational modification of TRF1 and TRF2 regulate their functions, including DNA binding, dimerization, localization, degradation, and interactions with other proteins (Walker and Zhu 2012). In addition, several groups have reported DNA binding and gene regulation functions for both TRF1 and TRF2 proteins outside of telomeres (Bradshaw, Stavropoulos et al. 2005, Fouche, Cesare et al. 2006, Mao, Seluanov et al. 2007, Zhang, Pazin et al. 2008, Simonet, Zaragosi et al. 2011, Yang, Xiong et al. 2011, Bosco and de Lange 2012).

Both TRF1 and TRF2 contain a TRFH domain that mediates homodimerization and a C-terminal Myb type domain that mediates sequence-specific binding to telomeric DNA (Court, Chapman et al. 2005). In both proteins, the DNA binding domain and the dimerization domain are joined together by long linkers (~100 amino acids in TRF1 and 150 amino acids in TRF2). The dimerization domains from human TRF1 and TRF2 have the same α-helical architecture (Fairall, Chapman et al. 2001). However, TRF1 and TRF2 dimerization interfaces feature unique interactions that prevent heterodimerization. Solution structures of Myb domains of TRF1 and TRF2 bound to DNA with the sequence GTTAGGGTTAGGG show that both proteins recognize the central AGGGTT sequence through both hydrophilic and hydrophobic interactions between the protein and DNA (Hanaoka, Nagadoi et al. 2005). The DNA binding domain of TRF2 has ~four-fold weaker
DNA binding activity than that of TRF1 (equilibrium constants $K_d$ 750 nM vs 200 nM respectively). A single amino acid change (lysine on TRF2 to arginine on TRF1) is the main contributor to this difference in binding affinity.

The binding of TRF1 and TRF2 to telomeric DNA has been studied using EM and AFM (Griffith, Bianchi et al. 1998, Bianchi, Stansel et al. 1999, Griffith, Comeau et al. 1999, Stansel, de Lange et al. 2001, Yoshimura, Maruyama et al. 2004, Amiard, Doudeau et al. 2007). TRF1 was observed as a dimer or a tetramer, while TRF2 was observed as a dimer on DNA in the EM micrographs (Griffith, Bianchi et al. 1998, Bianchi, Stansel et al. 1999, Griffith, Comeau et al. 1999). TRF1 has extreme spatial flexibility and can induce DNA looping by binding to two distant half-sites (Bianchi, Stansel et al. 1999). In addition, TRF1 forms protein filaments on longer telomeric repeats ($\geq$ 27 repeats) and promotes parallel pairing of telomeric tracts (Griffith, Bianchi et al. 1998). This activity may play important roles in promoting T-loop formation (in conjunction with TRF2) and sister telomere associations (in conjunction with SA1) (Bisht, Daniloski et al. 2013). In vitro, TRF2 can remodel linear telomeric DNA into T-loops (Stansel, de Lange et al. 2001, Yoshimura, Maruyama et al. 2004), and displays domain B dependent binding to DNA junctions at replication forks and the center of Holliday junctions (Fouche, Cesare et al. 2006). Recent evidence suggested that TRF2 condenses DNA and generates positive supercoiling on DNA, thereby favoring strand invasion (Amiard, Doudeau et al. 2007). The DNA condensation activity is from the TRFH domain. Unlike TRF2, TRF1 alone lacks the ability to condense DNA due to the repression by its acidic domain (Poulet, Pisano et al. 2012). TRF2 can
simultaneously bind to TERRA and telomeric G-quadruplex DNA (Biffi, Tannahill et al. 2012), and a TERRA-like RNA molecule greatly reduces its ability to condense DNA (Poulet, Pisano et al. 2012).

TRF1 and TRF2 are the only shelterin proteins that bind to duplex telomeric DNA with high affinity. Dynamic movements on DNA, such as 1-dimensioinal (1-D) sliding (translocating while maintaining continuous DNA contact), jumping and hopping (microscopic dissociation and rebinding events), are essential for a protein to achieve its function inside cells where nonspecific DNA is in vast excess and bound by histones and other proteins (Berg, Winter et al. 1981, von Hippel and Berg 1989, Gorman and Greene 2008, Tafvizi, Mirny et al. 2011). How TRF1 and TRF2 find their cognate sites and protein partners to form the shelterin complex, and regulate the functions of proteins involved in DNA repair and cell-cycle progression are not fully understood (de Lange 2010).

6.2 Summary

Here we used single-molecules fluorescence imaging to study the dynamics of quantum dot (QD) labeled TRF1 and TRF2 proteins on λ DNA and DNA substrates containing alternating telomeric and non-telomeric sequences. To determine whether TRF1 and TRF2 slide or hop, we evaluated the effect of ionic conditions on the dynamic interactions between the QD labeled TRF proteins and DNA. TRF1 appears to bind directly to telomeric sequences with very little 1-D searching through non-telomeric DNA, whereas TRF2 possesses a significant component of 1-D search. Increasing the salt concentration
should not affect the diffusion constants of a sliding process, but should elevate the diffusion constants of hopping (Berg, Winter et al. 1981, Komazin-Meredith, Mirchev et al. 2008, Gorman, Fazio et al. 2010). While TRF1 followed a trend of decreasing diffusion constants as the ionic strength increased, TRF2 was highly motile on λ DNA across all ionic strengths and showed no change in diffusion constant. TRF1 also showed a slight increase in the α factor proportional to the ionic strength, while TRF2 did not show any significant variation with ionic strength, suggesting an unbiased random walk. Additionally, both TRF1 and TRF2 bound to the ligated T270 DNA tightropes with regular spacing. For both TRF1 and TRF2, the distribution of the distances between nearest-neighbor binding fit well to the sum of two Gaussian distribution functions centered at ~1.6 and 3.2 µm (consistent with the expected spacing of the telomeric regions).

We also examined how far single molecules of TRF1 and TRF2 could slide on the ligated non-telomeric DNA versus ligated T270 DNA. TRF1 and TRF2 displayed distinct diffusion ranges on telomeric DNA, but not on non-telomeric DNA. The diffusion range was invariant across all time windows (~10-100 s), ruling out the possibility that the short range diffusion observed was due to shorter video lengths. Instead, this finding suggests that once the molecules are within a telomeric region, they tend to remain there. We explored the possibility that short range diffusion was caused by multiple proteins binding to the same telomeric region and restricting 1-D sliding. However, at lower TRF2 concentration, the short diffusion range did not change. We noted that in many cases TRF proteins binding to telomere repeats would be confined to diffuse due to the higher affinity for telomeric
sequences. To ensure that this confinement would not artificially reduce the apparent diffusion constant, I designed a 1-D diffusion of proteins on a linear DNA lattice of unlimited length versus a 1.6 kb total length, which mimics the (TTAGGG)$_{270}$ region. These simulations revealed that confinement within 1.6 kb DNA does not significantly reduce the observed diffusion constant at the (TTAGGG)$_{270}$ region (Figure 10: Computer simulations of diffusion by modeling random walk of proteins on a 1-D DNA lattice using Python$^{\text{TM}}$ programming language.). In addition, camera based time-averaging was not a major contributor to the observed slower diffusion constants at the telomeric region under these experimental conditions.

Using a truncation mutant, we localized this 1-D searching activity to the basic domain of TRF2. The basic domain at the N-terminus of TRF2 permits its binding to model replication forks and four-way junctions independent of telomere sequences (Fouche, Cesare et al. 2006). In addition, the absence of this domain leads to a diminished ability of TRF2 to localize to model telomere ends and to facilitate T-loop formation (Fouche, Cesare et al. 2006). We purified and imaged a basic domain deletion mutant of TRF2 (TRF2ΔB) on λ DNA and the ligated T270. Compared to full length TRF2, TRF2ΔB-QDs had a higher specificity for the telomeric sequences on T270 DNA substrate and lower affinity to DNA ends (Figure 11: Dynamics of TRF2ΔB-QDs on λ DNA and the ligated T270 DNA.). Interestingly, the diffusion constant ($9.1 \pm 1.8 \times 10^{-2} \mu m^2/sec$) and $\alpha$ factor ($0.93 \pm 0.04$) of TRF2ΔB on λ DNA were not significantly different from those of full length TRF2. On T270 DNA, the majority of motile TRF2ΔB (90%) was found with a diffusion range consistent
with length of the telomeric region on T270 DNA, suggesting that TRF2ΔB directly associates with telomeric DNA from solution and not by diffusion from a non-telomeric region. The dynamics of TRF2ΔB over the (TTAGGG)_270 region were similar to those of full length TRF2, with a similar diffusion range (0.47 ± 0.03 µm) and diffusion constant (0.27 ± 0.01 x 10^{-2} µm^2/sec at 125 mM NaCl and 0.26 ± 0.01 x 10^{-2} µm^2/sec at 225 mM NaCl). These observations suggest that the basic domain of TRF2 normally facilitates its 1-D search on non-telomeric DNA. The reduced degree of TRF2 localization to the telomeric region due to deletion of the basic domain demonstrates the importance of 1-D diffusion in the TRF2 telomeric target site search.

On telomeric DNA both TRF1 and TRF2 diffuse slowly due to higher energy barriers to diffusion; and they possess longer attached lifetimes at telomeric repeats compared to non-telomeric DNA sequences. These observations indicate that there is preferential binding to telomeric DNA but the affinity is not high enough to prevent TRF proteins from diffusing along TTAGGG repeats. We postulate that this allows TRF1 and TRF2 to find their protein partners locally, and that this is a more general mechanism for coupling the energy from multiple weak DNA binding components to ensure high binding specificity on long repetitive sequences.

**Specific contributions of the author:** I performed a variety of both AFM and TIRF experiments for the TRF1 and TRF2 proteins, including the data analysis. I performed all TIRF experiments for the TRF2ΔB protein and the resulting data analysis. I also executed the protein purification for TRF1 and TRF2 proteins used for the experiments. Finally, I created
a custom Python script to explore the impact of frame rate and hard scattering barriers on simulated TRF1 and TRF2 molecules diffusing on a 1-D DNA lattice.
Figure 9: TRF1 and TRF2 strike a balance between target search and specificity.

(A) TRF1 and TRF2 can undertake a 1-D search on DNA consistent with rotation-coupled diffusion along the DNA helix. The small ovals represent the basic and acidic domains of TRF1 and TRF2. The blue and purple lines represent non-telomeric and telomeric DNA, respectively. TRF1 relies more on 3-D search and the majority of TRF2ΔB molecules bind to the telomeric region directly from solution, forgoing the 1-D component of the search. (B) The energy landscape along the positions at telomeric and non-telomeric sequences. The diffusion constant and lifetime measurements are consistent with ~2.8-3.6 $k_B T$ higher energy barriers to diffusion at the telomeric sequences in comparison with non-telomeric sequences.
The additional energy barrier at the non-telomeric and telomeric junction represents the activation energy needed for conformational change/DNA binding domain switching on proteins to achieve specific binding.
Figure 10: Computer simulations of diffusion by modeling random walk of proteins on a 1-D DNA lattice using Python\textsuperscript{TM} programming language.

The diffusion constants used for simulation were that of TRF2-QDs on λ DNA (A) and on the telomeric DNA (B) at 125 mM ionic strength. A plot for MSD vs Δt is generated and an example of a trajectory of a protein (insert) are presented for (left, A) DNA with unlimited length and (right, A) DNA with 1.6 kb length at 1460436 steps/s stepping rate (corresponding
to TRF2-QDs on λ DNA). A plot for MSD vs Δt and an example of a trajectory of a protein (insert) are also presented for (left, B) DNA with unlimited length and (right, B) DNA with 1.6 kb length at 46713 steps/s (corresponding to the rate of TRF2-QDs at the telomeric region). The fitting parameters were constructed such that only the initial linear portion of the MSD vs. Δt plots were used for calculating the diffusion constant. For diffusion with confinement, a protein walks along a 1-D DNA lattice with two totally reflecting barriers. The numbers in each plot are the mean and standard deviation of the simulated data. The number of particles simulated for each case is indicated in the parenthesis.
Figure 11: Dynamics of TRF2ΔB-QDs on λ DNA and the ligated T270 DNA.

(A) Position distributions of TRF2ΔB-QDs on the linear T270 substrate (n = 51). Among the protein-QDs on DNA, ~66% TRF2ΔB-QDs bound to the telomeric regions (35% to 50% from DNA ends). Kymographs of TRF2ΔB-QDs on λ DNA (B) and the ligated T270 (C). The scale bar is 1 μm. Protein-QD-DNA reactions were carried out at 125 ionic strength. (D)
Diffusion range distributions of TRF2ΔB-QDs on λ DNA (white bars, n = 21) and the ligated T270 DNA (striped bars, n = 30). The binding affinity of TRF2ΔB to λ DNA tightropes was significantly lower as indicated by lower average numbers of TRF2ΔB-QDs on DNA in the field of view (1.4 ± 0.2 vs 13.3 ± 2.2) and lower average numbers of protein-QDs on DNA tightropes between two beads (1.1 ± 0.1 vs 3.8 ± 0.3). Dual color QD labeling confirmed that TRF2ΔB can form dimers or higher order oligomeric species, but at a significantly lower percent (14%) compared with the full length protein (79%). The fraction of motile protein-DNA complexes decreased from 95% for full length TRF2 to 65% (n = 40) for TRF2ΔB at 125 mM NaCl and from 89% to 74% (n = 23) at 225 mM ionic strength.
REFERENCES


APPENDICES
Appendix A: Cohesin SA2 is a sequence independent DNA binding protein that recognizes DNA ends and gaps
Cohesin SA2 is a sequence independent DNA binding protein that recognizes DNA ends and gaps

Preston Countryman¹, Yanlin Fan², Parminder Kaur³, Hai Pan⁴, Xuechun Wang⁵, Jiangguo Lin⁶,¹, Haijiang Chen⁶, Dianwen Wang⁴, Christian White⁴, Changjiang You⁷, Nicolas Wirth⁷, Ingrid Tessmer⁷, Jacob Pielehe⁷, Robert Riehn¹, Yizhi Jane Tao², Hong Wang⁵,⁶,*

¹Physics Department, ²Biomedical Engineering Department, ³Center for Human Health and the Environment, North Carolina State University, Raleigh, North Carolina, NC 27695, USA
²Department of BioSciences, Rice University, Houston, Texas, USA
³School of Bioscience and Engineering, South China University of Technology, Guangzhou, Guangdong 510006, P. R. China
⁶Division of Biophysics, Universität Osnabrück, Barbarstrasse 11, 49076, Osnabrück, Germany
⁷Rudolf Virchow Center for Experimental Biomedicine, University of Würzburg, Josef-Schneider-Str. 2, 97080, Würzburg, Germany

*Correspondence should be addressed to H.W. (hong_wang@ncsu.edu); phone: 919-513-7203; fax: 919-515-6538
The cohesin complex plays crucial roles in diverse genome maintenance pathways. In vertebrates, the core cohesin complex consists of a tripartite ring assembled by Smc1, Smc3, Rad21 and the stromal antigen subunit (SA), SA1 or SA2. Mutations in core cohesin subunits or their regulatory factors lead to a variety of human diseases, known as cohesinopathies, and cancers. The most frequently mutated cohesin subunit in tumors is SA2. Recently, we discovered that SA1 binds to dsDNA and displays weak specificities for telomeric sequences. However, DNA binding properties of SA2, which underlie its functions in diverse biological pathways, are unknown. Here, using single-molecule atomic force and fluorescence microscopy imaging as well as fluorescence anisotropy, we establish that SA2 binds to both dsDNA and ssDNA, albeit with a higher binding affinity for ssDNA. While SA1 does not specifically bind to either centromeric or telomeric DNA, it recognizes DNA ends and ssDNA gaps with high specificities. While SA2 by itself mainly exists as monomers in solution, it forms higher order oligomers on DNA, and can switch from 1D unbiased diffusion (searching) to stable binding (recognition) at ssDNA gaps. Importantly, this study uncovers a new role for SA2 as a cohesin anchoring protein that stabilizes it at specific DNA structures including DNA ends and ssDNA gaps.
In eukaryotes, proper chromosome alignment and segregation during mitosis depend on cohesion between sister chromatids (1-3). Cohesion is mediated by the cohesin complex, which also plays important roles in diverse biological processes including double-strand DNA repair, re-start of stalled replication forks and maintenance of 3-D chromatin organization (4,5). In vertebrates, cohesin consists of heterodimeric ATPases SMC1 and SMC3, a kleisin subunit Rad21 (also known as Scc1) and stromal antigen subunit (SA) SA1 (STAG1) or SA2 (STAG2) (3). Electron microscopy, crystallography, and biochemical assay based studies support the notion that cohesin binds to DNA by topological embrace (6-9). The N- and C- termini of SMC proteins fold back on themselves, forming anti-parallel coiled-coils with one half of an ABC-type ATPase head on one end and a dimerization domain at the other (7,10). The hinge domains of SMC1 and SMC3 dimerize, while RAD21 stabilizes the two SMC ATPase heads forming the cohesin trypitile ring. RAD21 also interacts with SA1/SA2 (11). SA1 and SA2 share 70% sequence homology, and exist in separate cohesin complexes, with SA2 being more abundant than SA1 (12-14). Cohesin loading onto DNA depends on ATP hydrolysis by the SMC heads and is promoted by the cohesin loader complex (9,15-17). In addition to the core cohesin subunits, several cohesin regulatory factors have been discovered that play important roles in the loading, stability, and cleavage of cohesin during different phases of the cell cycle (18-24).

Germline mutations in core cohesin subunits or their regulatory genes are associated with a spectrum of human diseases collectively called “cohesinopathies” and an increased incidence of cancer (4,25,26). Somatic mutations of the SA2 gene and loss of SA2 protein expression have been reported in multiple cancer cell lines including urothelial bladder carcinomas, Ewing’s sarcomas glioblastomas, and malignant melanomas (25). Importantly, targeted inactivation of SA2 in a human cell line with a stable karyotype leads to chromosomal instability and aneuploidy (27).

Despite great progress made since the discovery of the cohesin complex, many fundamental questions regarding the structure and assembly of cohesin remain unanswered (28,29). This includes how cohesin binds to chromatin to establish sister chromatid cohesion (30). Previous observations support the notion that cohesin can exist in two chromatin bound states, being either cohesive or noncohesive for tethering sister chromatids (28). It is not fully understood why cohesin loading onto chromatin (in a non-cohesive state) is necessary, but not sufficient to tether sister chromatids (in a cohesive state) (28). Various models including one ring, twin-ring handcuffs, bracelet oligomers, and C-clamps, have been proposed for cohesin assembly on DNA (8,22,29,31-37). It is known that establishment of sister chromatid cohesion
in the S phase requires acetylation of SMC3 by the cohesin acetyltransferases ESCO1/ESCO2 as well as binding of CDCA5 (Sorokin) (38-40). Several previous studies also support a DNA replication coupled cohesion model. These studies identified the association of cohesion establishment factors (Eco1/ctf7p in *S. cerevisiae*) with DNA replication machineries (RFC clamp loader complex and DNA helicases) as the molecular interactions driving the cohesive state (28,30,39,41-47). However, these models have not taken into consideration that SA2 plays important roles both in loading and unloading of cohesin from chromatin. It is known that SA2 phosphorylation by the polo-like kinase 1 (Plk1) leads to the removal of cohesin from chromatin (48).

In addition, how cohesin DNA binding is spatially controlled along the genome is poorly understood. In budding and fission yeast, outside the high density binding region around each centromere, the low density binding occurs at cohesin-associated regions (CARs) (49-51). In fission yeast, cohesin is detected in regions of strong transcription activity and responds to transcription by downstream translocation and accumulation at convergent transcriptional terminators (52,53). Furthermore, DNA double-strand break (DSB) induction leads to establishment of sister chromatid cohesion in the G2 phase, which facilitates the DNA repair process (54-58). It is thought that following the induction of DSBs, cohesin is recruited to the region surrounding the DSB as well as genome wide through the DNA damage response pathway and chromatin remodeling (59,60). This process is facilitated by cohesin regulators (54,59-62). Recently, it was shown that cohesin suppresses the end joining of distal DNA double-strand breaks, but not close ones (63). Recently, it was also found that *S. pombe* cohesin can bypass DNA-bound proteins (with diameters of ~10.6 nm) (64). In addition, the *S. pombe* cohesin ring is able to slide on DNA with a diffusion constant approaching the theoretical limit for free 1-D diffusion and fall off from free DNA ends (64). These observations raise an important question: how does the cohesin complex promote stable cohesion during DNA DSB repair without sliding off from DNA ends? Furthermore, SA1 and SA2 have differential roles during DSB repair, as well as during sister chromatid cohesion at telomeres and centromeres (65,66). Specifically, depletion of SA2, but not SA1, significantly decreases homologous recombination repair and affects the repair pathway choice (67). While SA2 is important for cohesion at the centromere, SA1 is enriched at telomeres and plays a major role in sister telomere cohesion in a cohesin ring-independent manner (65,66).

As the exact function of SA1/SA2 remains elusive, understanding SA2 function at the single-molecular level is crucial to uncover the mechanisms underlying the function of cohesin in diverse biological pathways. Recently, we discovered that SA1 binds to dsDNA and shows
weak specificity for telomeric sequences (68). These new results raise an important question as to whether or not SA2 specifically recognizes unique DNA sequences and/or structures. To address this question, we applied fluorescence anisotropy and two complementary single-molecule imaging techniques, atomic force microscopy (AFM) and fluorescence imaging of quantum dot-(QD-)labeled proteins, to investigate the binding of SA2 to specific DNA sequences (centromeric or telomeric) and structures (nick, ssDNA gap and end). We discovered that SA2 binds to both ss- and dsDNA, albeit with a higher binding affinity for ssDNA. AFM imaging reveals that SA2 displays high binding specificity for both DNA ends and ssDNA gaps. In contrast to SA1 (68), the 1-D diffusion dynamics of SA2 on DNA is sequence independent. Strikingly, SA2 can switch between two DNA binding modes in one binding event: searching through unbiased 1-D diffusion on dsDNA and recognition through stable binding at ssDNA gaps. Importantly, these results strongly suggest a new model for cohesin assembly that takes into consideration SA2 1-D diffusion on DNA and its ability to specifically recognize ssDNA gaps and ends.

Materials and Methods

RESULTS

SA2 specifically binds to DNA ends

Studying the DNA binding properties of SA1 and SA2 is essential for advancing our understanding of the function of the cohesin complex in diverse genome maintenance pathways. To investigate whether or not SA2 is a sequence specific DNA binding protein, we purified His-tagged full length SA2 (Figure 1A). First, we evaluated the oligomeric state of SA2 using a previously established method that correlates the volume measured from AFM images with the molecular weight of proteins (76,84,85). SA2 molecules (141 KDa) displayed AFM volumes consistent with being predominantly in the monomeric state (Figure S1A). This result is consistent with our previous study of SA2 using gel filtration chromatography (11). Next, to investigate SA2 DNA binding, we carried out fluorescence anisotropy experiments using Alexa 488-labeled double- (66 bp) or single-stranded (66 nt) DNA substrates. These experiments showed that SA2 bound to genomic ds-, telomeric ds-, and genomic ss-DNA substrates with equilibrium dissociation constants (K_d) of 35.5 (± 0.8), 22.5 (± 1.0), and 8.3 (± 0.5) nM, respectively (Figure 1B).

Ensemble based biochemical assays such as fluorescence anisotropy and electrophoresis mobility shift assays (EMSAs) only provide average binding affinities for DNA substrates. These
assays cannot differentiate sequence specific DNA binding from DNA end binding. In contrast, from AFM images of protein-DNA complexes, a direct measurement of DNA binding specificity for unique sequences as well as that for DNA structures such as ends can be obtained through statistical analysis of binding positions of protein complexes on individual DNA fragments (75).

To evaluate SA2-DNA binding specificity, we applied AFM imaging of SA2 in the presence of linear DNA fragments containing either centromeric or telomeric sequences (Figure 2A, Materials and Methods). Two centromeric DNA substrates (4.1 kb) used for AFM imaging contain the α-satellite centromeric sequences that are either close to one end of the linearized DNA (Cen-end DNA) or at the middle (Cen-mid DNA) (Fig. 2A). For the telomeric DNA substrate (T270 DNA), the (TTAGGG)$_{270}$ sequences make up approximately 30% of the total DNA length (5.4 kb) and are located at the middle of the linearized T270 DNA (Figure 2A). SA2 displayed AFM heights (1.41 ± 0.30 nm, mean ± SD, Figures 2B&C, S1B) that were significantly taller than that of dsDNA alone (0.70 ± 0.08 nm, mean ± SD) (86). This large difference in heights enabled unambiguous identification of SA2 molecules on DNA. Statistical analysis of the binding position of SA2 on DNA revealed that SA2 did not bind specifically to either the centromeric or telomeric sequences (Figure 2D). However, on all three DNA substrates the majority of SA2 molecules were bound at the DNA ends. Furthermore, DNA end binding by SA2 was independent of the internal DNA sequence or position of the centromeric region (Figure 2D). On three DNA substrates containing either blunt (T270 and Cen-mid) ends or ends with overhangs (4 nt 3' overhang on Cen-end DNA), SA2 displayed similar fractional occupancies at DNA ends. These results indicated that DNA end binding by SA2 was independent of single-stranded overhangs.

To further quantify the SA2 binding specificity for DNA ends, we applied the analysis based on the fractional occupancies of SA2 at DNA ends (75). The SA2 binding specificity for DNA ends (K$_{a$,$\text{end}}$/K$_{a$,$\text{random}}$) are 2945 (± 77), 2604 (± 68), 2129 (± 76), respectively, for T270, Cen-end, and Cen-mid DNA substrates. In addition, in contrast to SA2 alone, SA2 formed higher-order oligomeric states with sizes ranging from 277 nm$^3$ to 2668 nm$^3$ on DNA (Figure S1C). Based on the calibration curve relating protein molecular weights and AFM volumes (76), these AFM volumes correspond to approximately 1 to 13 SA2 molecules. Taken together, these results from fluorescence anisotropy and AFM imaging clearly show that SA2 binds to both ssDNA and dsDNA, albeit with a higher affinity for ssDNA than for dsDNA. In addition, SA2 does not specifically bind to centromeric or telomeric sequences. However, it binds DNA ends with high specificity that are independent of DNA sequences and single-stranded overhangs.

**SA2 binds to ssDNA gaps with high specificity**
Given the strong binding affinities for both ds- and ss-DNA displayed by SA2 (Figure 1), we speculated that SA2 might bind to ssDNA gaps, which are generated after Fen1 removes the RNA primers at the DNA replication fork (87). To investigate SA2 binding to ssDNA gaps, we used a previously established method to generate a linear substrate containing an ssDNA gap (37 nt) flanked by dsDNA arms (Figure 3A). This method was based on the generation of four closely spaced nicks using DNA nickase and subsequent removal of short ssDNA between nicked sites using complementary oligos (70,71). After restriction digestion of the circular gapped DNA, the ssDNA gap is at 470 nt (23%) from one end of the DNA (blunt end, Figure 3A and Figure S2, Materials and Methods). Based on diagnostic restriction digestion at the gapped region, the DNA gapping efficiencies were typically 85 to 95% (Figure S2B). To further confirm the presence of the ssDNA gap, the position distribution of mitochondria single-strand DNA binding protein (mtSSB) on this DNA substrate was analyzed. mtSSB protein predominantly bound to the expected ssDNA region on the gapped DNA substrate, while its binding on the nicked DNA substrate was random (Parminder Kaur et al., unpublished data). Taken together, these results established the presence of an ssDNA gap at the defined location on the linear gapped DNA substrate.

Next, to study whether or not SA2 specifically binds to ssDNA gaps, we directly compared the SA2 binding on non-gapped (without nickase treatment) to gapped DNA substrates (Figure 3B&C). AFM imaging showed that on the non-gapped DNA substrate, SA2 predominantly bound to the DNA ends and its distribution at internal sites along the linear DNA fragment was random (Figure 3C). This is consistent with position distributions of SA2 on telomeric and centromeric DNA substrates (Figure 2D). In stark contrast, the presence of an ssDNA gap shifted the SA2 binding from the DNA end to a region consistent with the location of the ssDNA gap (23%) (Figure 3C). Analysis of the fractional occupancies of SA2 on DNA demonstrated that SA2 displayed high binding specificity (S = 2116 ± 30) for the ssDNA gap. In addition, compared to the size of SA2 molecules positioned outside the gapped regions (1096 ± 117 nm²), at the ssDNA gaps SA2 formed larger complexes with a broader size distribution (1458 ± 232 nm², Figure S1C).

Since DNA nicking is the intermediate step for generating DNA gaps, we further tested whether or not SA2 specifically binds to DNA nicks. First, to evaluate if SA2 displays binding specificity for individual nick sites, we generated a linear DNA substrate (517 bp) containing a single nick site at 37% from one DNA end (72). DNA nicking was confirmed by the observation of slower mobility of nicked DNA than its non-nicked counterpart under gel electrophoresis.
(Figure S3A). On the nicked DNA substrate, SA2 displayed preferential binding to DNA ends (Figure S3A). In stark contrast with what was observed on the gapped DNA substrate, SA2 molecules were randomly distributed along the nicked DNA substrate (Figure S3A). Furthermore, on a DNA substrate containing 4 nick sites spatially separated from one another, AFM imaging further established that SA2 did not show preference for nicked sites (Figure S3B). In addition, with C-terminal domain deletion, SA2 1-1051 retains binding affinities, with $K_d$ of 28.0 and 47.3 nM for both ss- and dsDNA, respectively (Figure S4). Consistent with these results, AFM imaging showed that SA2 1-1051 retains DNA binding specificities for DNA ends ($S = 3382 \pm 133$) and ssDNA gaps ($S = 3637 \pm 128$) (Figure S4). In contrast, while SA1 also displayed ssDNA binding affinities with $K_d$ of 36.0 ($\pm$ 1.0) nM, AFM imaging showed its high binding specificity for DNA ends ($S = 2094 \pm 38$), but not for the 37-nt ssDNA gap (Figure 3C). In summary, these results show that SA2 displays high binding specificity for ssDNA gaps, but not DNA nicks. In addition, SA2 with C-terminal domain deletion retains binding specificity for DNA ends and ssDNA gaps.

**SA2 carries out sequence-independent unbiased 1-D diffusion on dsDNA**

Dynamic movements on DNA, such as 1-dimensional (1-D) sliding, jumping, and hopping, are essential for a protein to find its target sites on DNA (88-93). To understand how proteins dynamically achieve DNA binding specificity, we developed a DNA tightrope assay based on oblique angle total internal reflection fluorescence microscopy (TIRFM) imaging of QD-labeled proteins on DNA stretched between micron-sized silica beads (Materials and Methods) (77-80). DNA tightropes (at an elongation of ~90% of the contour length) are formed between poly-L-Lysine treated silica microspheres using hydrodynamic flow (Figure 4A) (77). To generate longer DNA substrates with specific sequences that can span between silica microspheres, we ligated linear DNA fragments containing genomic, telomeric, or centromeric DNA sequences (Figure 2A) (77). Recently, using the DNA tightrope assay, we observed that QD-labeled SA1 displays slow subdiffusive events amid fast 1-D unbiased diffusion in a telomeric sequence dependent manner (68).

To study SA2-DNA binding dynamics, the streptavidin-coated QD was conjugated to His-S2A using biotinylated multivalent chelator tris-nitrilotriacetic acid ($^{6}$tris-NTA) as the linker (Figure 4B) (74). The three Ni-NTA moieties on the circular scaffold of the tris-NTA adaptor bind to a His-tag with subnanomolar affinities (74,94). AFM imaging revealed that QDs in the presence of only $^{6}$tris-NTA did not have significant binding affinities for DNA. The addition of His-tagged SA1 to the $^{8}$tris-NTA-QD reaction led to loading of QDs onto DNA, indicating that
QD binding to DNA tightropes was mediated through SA2. In addition, SA2-QDs retained DNA binding specificity toward ssDNA gaps (Figure S6). To monitor SA2 binding on DNA in real time, QD-labeled SA2 molecules were introduced into the flow cell using a syringe pump after DNA tightropes were established between poly-L-Lysine treated silica microspheres. Then the flow was stopped, allowing freely diffusing SA2 molecules in solution to bind to DNA tightropes. On all DNA substrates, SA2-QD molecules on DNA were long lived, with ~80% of SA2-QD complexes remaining on DNA tightropes after 2 minutes (N = 277). The positions of SA2-QDs were tracked by Gaussian fitting to intensity profiles to obtain the diffusion constant (88,77-79). Importantly, at the same protein concentrations (5 nM in the flow cell), the diffusion constants of SA2 on λ DNA and DNA tightropes containing either telomeric or centromeric sequences are indistinguishable (Figure 4D, Table 1). In addition, the diffusive exponent (alpha factor) was calculated to determine whether SA2 displayed subdiffusive motion on DNA. An alpha factor of 1 indicates an unbiased random walk and a value less than 1 indicates periods of pausing in the random walk process (subdiffusion) (82). Recently, we found that SA1 shows telomeric sequence dependent subdiffusive behavior on DNA, manifested by an alpha factor much smaller than 1 (alpha factor: 0.69 ± 0.03 on telomeric DNA) (68). SA2 displayed free 1-D diffusion on centromeric DNA (alpha factor = 0.96 ± 0.022, N = 65) and λ DNA (alpha factor = 0.93 ± 0.038, N = 48) tightropes (Table 1). In comparison, the alpha factors displayed by SA2 on telomeric DNA tightropes were slightly (p = 0.01) lower (0.86 ± 0.028, N = 55). In summary, fluorescence imaging of QD-labeled SA2 on DNA tightropes directly shows that SA2 carries out sequence independent 1-D diffusion on DNA tightropes containing telomeric, centromeric, or genomic sequences. These results are consistent with random position distributions of SA2 on both telomeric and centromeric DNA substrates shown in AFM images.

SA2 becomes static and can switch between dsDNA and ssDNA gap binding modes

To study SA2 DNA binding dynamics on DNA tightropes containing gaps, we introduced ssDNA gaps after anchoring ligated DNA between silica microspheres. Generation of ssDNA gaps on DNA tightropes was done by introducing nickase and complementary oligos in the flow cell, followed by heating it at 55°C, and washing with buffers to remove nickase, and excess short ss- and dsDNA (Figure 3A). Compared to telomeric (46%), centromeric (23%), and non-gapped control DNA (36%), on DNA tightropes containing ssDNA gaps, a significantly (p < 10^-6) higher percentage of SA2 molecules were static (80.0 ± 7.5%, N = 55, Figure 5B&C). In addition, the density of SA2 on gapped DNA tightropes increased with higher SA2 concentrations (compare Figure 5B top and bottom panels). To evaluate whether or not the static SA2 binding events occurred at the gapped region, we measured the distance between
nearest neighbor SA2-QD pairs. The distribution of this distance shows three distinct peaks centered at 0.72, 1.23, and 1.87 μm, respectively (Figure 5D), which are consistent with the expected spacing between ssDNA gaps on the ligated DNA tightropes (Figure 5A). In stark contrast, on DNA tightropes containing nicks, the spacing between nearest neighbor SA2-QD pair was random (Figure S3C).

To further confirm that DNA binding dynamics of SA2 on gapped DNA tightropes is distinctly different from that on nicked DNA, we compared the diffusion constant and alpha factor of mobile SA2 on DNA containing ssDNA gaps (Figure 3A) and λ DNA (untreated and nicked) tightropes (Figure 6A). We introduced nicked sites by incubating λ DNA with Nt.BstNBI nickase. To remove nickase, nicked λ DNA was further purified using phenol chloroform extraction before being introduced into the flow cell. λ DNA has over 40 Nt.BstNBI nickase sites, with spatial separation ranging from 13 bp to over 2000 bp. To observe mobile SA2 complexes on DNA tightropes, the final SA2-QD concentration in the flow cell (0.6 nM) was kept the same across all DNA substrates but lower than the standard concentration (5 nM, Fig. 4 and Table 1). On gapped DNA tightropes (D = 0.01 ± 0.003 μm²/s, and alpha factor = 0.70 ± 0.05), SA2 showed a significant decrease in the diffusion constant and alpha factor (p < 0.02) compared to untreated λ (D = 0.13 ± 0.032 μm²/s, and alpha factor = 0.96 ± 0.03) or nicked λ DNA tightropes (D = 0.082 ± 0.030 μm²/s, and alpha factor = 0.94 ± 0.04, Figure 6A).

Interestingly, a subpopulation of mobile SA2 molecules (N = 21 out of 150) alternated between mobile and static binding modes (Figure 6B). The pair-wise distance between nearest neighbor SA2 static binding positions was 0.60 (± 0.19) um (N = 21), which is consistent with the spacing between two adjacent ssDNA gaps on DNA tightropes. To further compare SA2 DNA binding dynamics on different DNA substrates, we calculated a time interval-based diffusion constant (Dint, Figure S7) by mobile SA2 using a "sliding window" (40-frame, 2 s) MSD analysis (83). This analysis indicated that distinct from the free 1-D diffusion mode (~1.0 X 10⁻² μm²/s, Figure S7), mobile SA2 molecules displayed an additional population with Dint values centered at ~1.0 X 10⁻⁴ μm²/s on gapped DNA tightropes (Supplementary Fig. S7D). Furthermore, we used Dint value of 1 X 10⁻⁴ μm²/s as the threshold value to identify individual static binding events. This value is based on the diffusion constants measured from static protein-QDs on DNA tightropes (68). This analysis indicated that compared to other DNA substrates (nicked and non-gapped), on the gapped DNA tightropes, mobile SA2 molecules displayed a significantly (p = 0.002) higher percentage of time windows (40-frame, 2 s) in the static binding mode (Figure 6C).
Taken together, fluorescence imaging of QD-labeled SA2 establishes that SA2 alternates between two DNA binding modes on gapped DNA – 1-D unbiased diffusion on dsDNA and stable binding at ssDNA gaps. This result is consistent with AFM imaging showing specific binding for ssDNA gaps by SA2.

**SA2 forms higher-order oligomeric complexes and can bypass diffusion barriers on DNA**

In AFM images, while SA2 alone mainly existed as monomers, SA2 formed higher-order oligomers on DNA (Figure S1). Consistent with these observations using AFM, SA2-QDs with brighter intensities were observed to break up into multiple fainter ones (red arrows, Figure 7). This observation indicated that the brighter SA2 complexes were higher-order oligomers. To determine how SA2 dynamically forms higher-order oligomeric complexes on DNA, we analyzed instances where a mobile SA2 molecule encountered additional stationary or mobile SA2 molecules. The overwhelming majority (92%, N = 49) of SA2-SA2 interactions on DNA were collisions that did not form complexes. However, there were cases (8%) of initial separate mobile SA2 molecules that collided and then diffused in synchronicity with brighter intensity than individual molecules (white arrows, Figure 7A). These results show that the assembly and disassembly of higher-order SA2 complexes on DNA are dynamic.

Proteins that maintain continuous close contact with DNA during sliding are unable to circumnavigate obstacles posed by another protein on DNA. In contrast, a hopping mechanism in which a protein micro-dissociates and re-associates with DNA within a distance comparable to or greater than the dimension of DNA-bound proteins could enable it to transverse these diffusion barriers. Previously, single-molecules imaging has revealed hopping by a DNA repair protein (Mlh1-Pms1) and P53 (95,96). We observed instances of mobile SA2 molecules (N = 4 out of 49 colliding SA2 pairs) bypassing another DNA-bound SA2 molecules (Figure 7B). This bypass frequency is comparable with what was observed with Mlh1-Pms1 (95).

In summary, the assembly and disassembly of higher-order SA2 protein complexes on DNA are dynamic. Combining with the observation that SA2 by itself mainly exists in the monomeric form, these results imply that SA2 binds directly to DNA as monomers from the solution. The assembly of higher-order SA2 complexes on DNA is promoted through 1-D diffusion and direct interactions on DNA. SA2 hops on DNA, enabling it to bypass diffusion barriers posed by another DNA-bound proteins.

**Discussion**

Despite the importance of SA2 in multiple genome maintenance pathways, the function of SA2 has been elusive. Using single-molecule imaging in combination with biochemical
characterization of SA2, we establish that SA2 binds to both ds- and ss-DNA, albeit with a higher affinity for ssDNA. It does not specifically bind to centromeric or telomeric sequences, but shows high binding specificity for DNA ends and ssDNA gaps. SA2 can switch between free 1-D diffusion on dsDNA and stable binding at ssDNA gaps in one binding event. These findings have important implications for understanding the function of cohesin in diverse genome maintenance pathways.

SA2 DNA binding in the context of sister chromatin cohesion

Cohesin is required for sister chromatin cohesion at the time of DNA replication or shortly thereafter (97). Cohesin induced in G2 or M phases still associates with chromosomes but does not promote cohesion between sister chromatids (Uhlmann and Nasmyth, 1998). However, protein-DNA structures that direct the loading of cohesin at replication forks and the timing of cohesion events relative to the progression of the DNA replication fork are not fully understood (98). It was suggested that cohesin establishment occurs in concert with lagging strand Okazaki fragment processing events mediated by the Fen1 flap endonuclease (99). Previous studies of cohesin loading onto DNA had been focused on the three ring-subunits and their regulators. The results from this study shed new light on a previously uncharacterized function of SA2 in DNA binding. ssDNA gaps between Okazaki fragments are created on the lagging strand during DNA replication. Specific binding toward ssDNA gaps by SA2 could facilitate the loading of the core cohesin complex at replication forks. SA2 can switch between 1-D diffusion searching on dsDNA and stable binding at ssDNA gaps in one binding event. Diffusion across dsDNA and ssDNA regions without dissociation would allow individual SA2 molecules to navigate on the lagging strand to form multi-protein SA2 complexes. Recently, single-molecule imaging of QD-labeled S. pombe cohesin complexes on DNA curtains showed that DNA-bound core cohesin complexes displayed a central hole that was estimated to be larger than ~10.6 nm, but less than ~19.5 nm (64). In light of these new results, to accommodate both the leading and lagging strands during DNA replication, the bracelet and handcuff models in which each cohesin complex binds to separate DNA strands are most appealing. However, the findings from this study strongly suggest another possibility with SA2 and the cohesin ring subunits binding to separate DNA strands. In this model, it is likely that SA2 binds to the ssDNA gaps on the lagging strand. In other words, sister chromatid cohesion relies on both ssDNA gap binding by SA2 and sequence/structure independent entrapment of dsDNA within the cohesin tripartite ring. This model is consistent with our finding that SA2 binding at the ssDNA gaps does not promote DNA-DNA pairing. The percentage of DNA-DNA pairing mediated by SA2 observed in
AFM images were less than 1% and 2% for gapped (N = 343) and non-gapped DNA (N = 435), respectively.

**DNA binding by SA2 in the context of 3-D genome architecture formation**

One important function of cohesin in interphase chromatin is to mediate chromatin remodeling loops that bring together enhancer and promoter DNA regions for gene expression and insulation (100-102). Cohesin shares many binding sites with CCCTC binding factor (CTCF) (103), a mediator of long-range chromatin interactions. Importantly, the C-terminal tail of CTCF protein directly binds to the SA1/SA2 while the rest of the cohesin complex is recruited later. A prevailing notion is that cohesin plays a role in the stabilization of chromatin loop regulation (104). 1-D diffusion of SA2 on DNA would enable its search for the CTCF protein on DNA. Furthermore, results from this study strongly suggest that SA2 1-D diffusion might play an important role in protein-mediated DNA extrusion for promoting topologically associating domains (TADs) (102).

**SA2 in DNA DSB repair**

Cohesin localizes to dsDNA breaks induced by radiation, enzyme digestion, or DNA replication through DNA lesions (56,57,59,61,105-108). Participation of cohesin in DNA DSB repair depends on its ability to hold together sister chromatids (58). Previous studies found that localization of SA1 and SA2 to DNA damage sites is strongly correlated with the presence of Mre11 and/or Rad50 (59). The recruitment of Smc1 and Rad21 also vanished with depletion of SA2 but not SA1, providing strong evidence that SA2 is necessary for the other cohesin subunits to accumulate at DNA damage sites (60). Despite these connections, the mechanism underlying cohesin recruitment to regions of DNA damage is poorly understood. Our observations reported in this study suggest that SA2 DNA end binding could provide the "structure anchor" for the cohesin complex to stably bind at the DNA ends. Without anchoring at the DNA ends by SA2 and the cohesin ring subunits, one single-ended DSB could interact with another distant one, causing intra- and inter-chromosomal rearrangements through non-homologous end-joining. Furthermore, the recruitment of SA2 to DNA breaks may also facilitate homologous chromosomal DNA pairing during dsDNA break repair. In other words, the results from this study are consistent with a model in which loading of SA2 at the DNA double-strand break in conjunction with entrapment of its sister chromatid within the cohesin ring restrict the trafficking of DNA extremities to prevent intra- and inter-chromosomal rearrangements.

**Overlapping and unique roles of SA1 and SA2**

13
It is well established that SA1 and SA2 have overlapping as well as unique functions. One major difference between SA1 and SA2 proteins is found in the first 75 amino acids of their N-terminal domains (66). Only SA1 contains an AT-hook motif at its N-terminal domain. Cohesin SA1 plays a more prominent role than SA2 in regulation of gene expression (109). In a recent study, we propose a model in which SA1 is the "DNA sequence guide" (using its AT-hook motif) and directs the loading of the core cohesin complex at AT-rich DNA sequences along the genome (68). In this study, we discover that unlike SA1, SA2 does not recognize specific DNA sequences. However, in parallel with DNA sequence dependent binding by SA1, SA2 might play a more prominent role in anchoring the cohesin complex at specific structures including ssDNA gaps. Fluorescence anisotropy experiments demonstrate that SA2 displays a 4-fold higher binding affinity for ssDNA compared to dsDNA. On the other hand, SA1 displays comparable DNA binding affinity for double-stranded telomeric sequences \((K_d = 34.0 \pm 5.8 \text{ nM})\) and ssDNA \((K_d = 36.0 \pm 1.0 \text{ nM})\). While both SA1 and SA2 bind to ssDNA gaps, the specificity of SA1 for ssDNA gaps can be masked by its preference for AT-rich sequences. In summary, we propose that both SA1 and SA2 function as the DNA sequence/structure anchor for the cohesin complex, while the unique roles of SA2 and SA1 are manifested by the difference in their DNA binding properties. Future single-molecule studies are necessary to build a complete model for cohesin assembly that takes into consideration DNA binding by SA1 and SA2.

**SUPPLEMENTAL INFORMATION**

Supplemental information that includes 7 Supplemental figures can be found with this article online.

**ACKNOWLEDGEMENTS**

We would like to thank the Weninger and Riehn groups at NCSU for technical support. This work was supported by the National Institutes of Health (R00ES016758 to H.W., R01GM107559 to R.R. and H.W., Deutsche Forschungsgemeinschaft (Forschungszentrum FZ82 to I.T., J.P., P30 ES025128 to CHHE at NCSU, and the Welch Foundation [C-1565 to Y.J.T.].

**AUTHOR CONTRIBUTIONS**

P.C., P.K., Y.F., H. P., J.L., H.C., D.W., X.W., C.W., N.W., and I.T. performed the experiments. P.C. wrote the Matlab code for analysis. I.T., J.P., R.R., J.T., and H.W. were involved in the design of the study and writing of the paper.

*Conflict of interest statement.* None declared.
FIGURE LEGENDS

Figure 1. SA2 binds to both ds- and ss-DNA substrates. (A) SDS-PAGE gel of purified full length SA2. (B) Fluorescence anisotropy experiments showing concentration-dependent binding of SA2 to dsDNA (66 bp) and ssDNA (66 nt) substrates. The data were fitted to the law of mass action. DNA substrates are labeled with Alexa 488. The equilibrium dissociation constants are 35.5 (± 0.8 SEM) and 8.3 (± 0.5, SEM) nM, respectively, for ds- and ss-DNA substrates (two independent experiments).

Figure 2. SA2 does not show binding preference for telomeric or centromeric DNA sequences, but recognizes DNA ends. (A) Three DNA substrates used for AFM imaging. (B-C) Representative AFM images of SA2 on the centromeric (B, Cen-mid) or telomeric (C) DNA substrates. (D) Position distributions of SA2 on DNA substrates containing telomeric (T270, N = 283), centromeric sequences close to one end (Cen-end, N = 275) or in the middle (Cen-mid, N = 298). Each data set was from three independent experiments. The error bars represent SEM.

Figure 3. SA2 specifically binds to ssDNA gaps. (A) Generation of the linear gapped DNA substrate. Gapped DNA was created using pUC19 derived pSCW01 plasmid (2030 bp) that contains 4x Nt. BstNBI nicking sites. After restriction digestion, the resulting ssDNA gap is 37 nt long and at 470 nt (23%) from one end of the linear DNA fragment. (B) Representative AFM images of the full length SA2 complex binding to the linear gapped DNA substrate. The contour length of the linear gapped DNA was measured as Lc = 622.48 ± 41.3 nm. (C) Statistical analysis of the position distribution of the full length SA2 complex on the linear gapped (N = 251) and non-gapped (N = 201) DNA, as well as full length SA1 on the gapped DNA substrate (N = 295). Each data set was obtained from at least 2 independent experiments. The error bars represent SEM.

Figure 4. SA2 displays similar binding dynamics on DNA substrates containing centromeric, telomeric, or random sequences. (A) Schematic of the DNA tightrope assay. (B) The QD conjugation strategy: a His-NTA-biotin-QD sandwich method for conjugating His-tagged SA2 to QDs. (C) Representative kymographs of QD-labeled SA2 on centromeric (top), telomeric (middle), and λ DNA tightropes (bottom). In all reactions, SA2 protein was incubated with both red (655 nm) and green (565 nm) QDs at equal molar concentrations. (D) Diffusion constants of SA2 on centromeric (Cent-DNA, D = 0.10 ± 0.018 μm²/s, N = 48), telomeric (T270, D = 0.096 ± 0.018 μm²/s, N = 53), or λ (D = 0.095 ± 0.017 μm²/s, N = 48) DNA tightropes. The error bars represent SEM.
Figure 5. SA2 stably binds at ssDNA gaps. (A) Schematic of the DNA tightropes with ssDNA gaps at defined spacing. (B) Representative kymograph of SA2 on the ligated DNA tightropes containing gaps at the low (0.6 nM) and standard (5 nM) protein concentrations. The ssDNA gaps were generated by heating and introduction of complementary oligos after the DNA tightropes were formed. Equal molar concentrations of red and green QDs were present in the conjugation reaction. (C) The percentage of static SA2 molecules on telomeric, centromeric, non-gapped control, and gapped DNA tightropes. The final SA2 concentration in the flow cell was 0.6 nM. (D) Statistical analysis of the spacing between SA2-QD complexes on the gapped DNA tightropes (N = 149). The line represents Gaussian fit to the data (R² > 0.93) with peaks centered at 0.72, 1.23, and 1.87 μm, respectively.

Figure 6. SA2 switches between searching and recognition modes on DNA tightropes containing ssDNA gaps. (A) Comparison of SA2 diffusion constants and alpha factors on gapped DNA (N = 28), nicked λ DNA (N = 20), and non-nicked λ (N = 20) DNA. Final SA2 concentration was 0.5 nM in the flow cell. *: p < 0.02; **: p < 0.001; ***p < 0.0005. (B) Kymographs of SA2 showing SA2 alternating between 1-D diffusion and stable binding on gapped DNA tightropes. (C) Percentages of time windows (40 frames/2 s) with Dm values less than 10⁻⁴ for mobile SA2 on gapped, λ, non-gapped control, centromeric (Cen-DNA), and telomeric (T270) DNA tightropes.

Figure 7: SA2 forms dynamic higher-order oligomeric complexes on DNA and can bypass another DNA-bound SA2 complex. (A) Kymographs of SA2-QDs on DNA showing dis-assembly (red arrows) and re-assembly (white arrows) of SA2 complexes on DNA tightropes. (B) Examples of SA2 bypassing diffusion barriers posed by another DNA-bound SA2 molecules. Top: SA2 was blocked by a DNA-bound SA2-QD (left arrow) and then bypassed it (right arrow).
### Table 1: Summary from analysis of diffusion dynamics of QD-labeled full length WT SA2 on DNA tightropes.

<table>
<thead>
<tr>
<th>DNA</th>
<th>WT SA2-QDx (5 nM)</th>
<th>WT SA2-QDx (0.6 nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$D$ ($\mu m^2/s$)</td>
<td>Alpha factor</td>
</tr>
<tr>
<td>Centromeric</td>
<td>$0.10 \pm 0.018$</td>
<td>$0.96 \pm 0.022$</td>
</tr>
<tr>
<td></td>
<td>($N = 48$)</td>
<td>($N = 65$)</td>
</tr>
<tr>
<td>T270</td>
<td>$0.096 \pm 0.018$</td>
<td>$0.96 \pm 0.028$</td>
</tr>
<tr>
<td></td>
<td>($N = 53$)</td>
<td>($N = 55$)</td>
</tr>
<tr>
<td>$\lambda$ DNA</td>
<td>$0.095 \pm 0.017$</td>
<td>$0.93 \pm 0.038$</td>
</tr>
<tr>
<td></td>
<td>($N = 45$)</td>
<td>($N = 48$)</td>
</tr>
<tr>
<td>Nicked $\lambda$ DNA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gapped DNA</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**Note:**

$D$: diffusion constant (mean $\pm$ SEM)
REFERENCES


107. Atienza, J.M., Roth, R.B., Rosette, C., Smylie, K.J., Kammerer, S., Rehbock, J., Ekblom, J. and 
Denissenko, M.F. (2005) Suppression of RAD21 gene expression decreases cell growth and 
enhances cytotoxicity of etoposide and bleomycin in human breast cancer cells. *Molecular 
cancer therapeutics*, 4, 361-368.


contributions of cohesin-S14 to cohesion and gene expression: implications for cancer and 
A. Nickase treatment
1. Nickase treatment
2. Addition of complementary oligos
3. Heat destabilization and annealing

B. SA2 + gapped linear DNA
4. Removal of short dsDNA and ssDNA via centrifugation/filtration
5. Restriction Digestion

C. Fraction of protein-DNA complexes (%) vs. Position from DNA end (%)
- SA2 + linear gapped DNA
- SA2 + linear non-gapped DNA
- SA1 + linear gapped DNA
Figure S1. Evaluation of the oligomeric state of SA2 in solution and on DNA. (A) AFM volume of SA2 alone in solution. SA2 volume was measured using the Gwyddion software. The solid line is the Gaussian fit to the data (N = 173) with the peak centered at 146 nm$^3$. The expected AFM volume of full length SA2 monomer (141 KDa) is ~150 nm$^3$ based on the calibration curve $V = 1.45 \text{ MW} - 21.59$, where $V$ is the AFM volume (nm$^3$) and MW (KDa) is the molecular weight of the protein. (B) SA2 AFM height distribution on the gapped DNA substrate (N = 173). (C) Comparison of AFM volumes of SA2 on the non-gapped and gapped DNA substrates.
Figure S2. Characterization of the DNA substrate with an ssDNA gap for AFM studies. (A) AFM images of circular gapped DNA substrate (left panel) and the linearized gapped DNA substrate with the ssDNA gap at 470 bp from one end. Circular DNA was digested with Scal to generate the linear gapped DNA substrate. Circular gapped DNA substrate exists as dimers and displays contour length (1354 ± 29 nm, mean ± SD) that is approximately twice that of linearized fragment (622.48 ± 41.3 nm, mean ± SD). (B) Quantification of DNA gapping efficiency using restriction digestion. Circular DNA was digested with NcoI, BamHI and PstI, which are located between nicked sites. The percentages of digestion were approximately 60%, 5%, respectively, for nicked and gapped DNA.
Figure S3. SA2 binding to DNA substrates containing either a single or multiple nick sites. (A) Characterization of linear DNA fragment (517 bp) containing a single nick (top panel) and binding positions of SA2 on this DNA substrate (N = 169, bottom panel). (B) Positions of SA2 on linear DNA substrates before (N = 51) and after (N = 101) treatment with nickase. The linear nicked DNA substrate (pUC19 derived) contains a total of 5 nicks: at 10% (2 nicks), 20%, 38%, and 43% from one end (purple arrows). (C) Spacing between nearest neighbor pairs of SA2-QDs (N = 71) on nicked pSGVO1 DNA tightropes is random.
Figure S4. The C-terminal deletion SA2 mutant (SA2 1-1051 AA) binds specifically to ssDNA gaps. (A) An AFM image of SA2 1-1051 binding to the linear gapped DNA substrate. (C) Statistical analysis of the position distribution of SA2 1-1051 complexes on the gapped DNA (N = 303).
Figure S6. Full length SA1 binds to ssDNA, recognizes DNA ends, but does show preferential binding to DNA nicks. (A) Fluorescence anisotropy experiments showing concentration-dependent binding of SA1 to the ssDNA (66 nt) substrate. The data were fitted to the law of mass action. DNA substrates are labeled with Alexa 488. The equilibrium dissociation constant is 36.5 ± 0.2 nM (two independent experiments). (B and C) Analysis of the binding position of SA1 on the telomeric T270 DNA substrate including DNA end binding (B, N = 200), and the linear DNA substrate containing a single nick (N = 97).
Figure S6. QD-labeled SA2 retains specific binding to DNA ends and ssDNA gaps. (A) An AFM image of QD-labeled full length SA2 on the linear gapped DNA. White arrows point to a SA2-QD complex binding to a location consistent with an ssDNA gap. (B) Statistical analysis of the AFM height of SA2-QDs on DNA (N = 109). The solid line shows the Gaussian fit ($R^2 > 0.72$) with the peaks centered at 1.9 ± 1.25 nm and 5.8 ± 0.2 nm respectively. (C) The position distribution of SA2-QDs on the gapped DNA substrate (N = 201).
Supplementary Figure S7. Comparison of the distribution of interval based diffusion constants ($D_{int}$) for SA2 on different DNA tightropes. $D_{int}$ distributions for all mobile SA2 molecules observed on centromeric (A, N = 48), telomeric (B, N = 52), λ (C, N = 41), and gapped DNA (D, N = 31).
Appendix B: Functional interplay between SA1 and TRF1 in telomeric DNA binding and DNA-DNA pairing
Functional interplay between SA1 and TRF1 in telomeric DNA binding and DNA-DNA pairing

Jiangguo Lin¹,², Preston Countryman², Haijiang Chen³,⁴, Hai Pan², Yanlin Fan⁵, Yunyun Jiang⁶, Parminder Kaur³, Wang Miao⁸, Gisele Gurgel⁶, Changjiang You⁷, Jacob Plehler⁷, Neil M. Kad⁹, Robert Riehn⁷, Patricia Opresko⁸, Susan Smith¹⁰, Yizhi Jane Tao³, Hong Wang²

¹School of Bioscience and Engineering, South China University of Technology, Guangzhou, Guangdong 510006, P. R. China

²Physics Department, ⁶Biomanufacturing Training and Education Center, North Carolina State University, Raleigh, North Carolina, NC 27695, USA

³Department of BioSciences, Rice University, Houston, TX 77005, USA

⁴Institute of Microbiology and College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang, 310058, P. R. China

⁵Department of Neurology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, 450014, P. R. China

⁶Division of Biophysics, Universität Osnabrück, Barbarstrasse 11, 49076, Osnabrück, Germany

⁷School of Biosciences, University of Kent, Canterbury, Kent, CT2 7NJ, United Kingdom

⁸Department of Environmental and Occupational Health, University of Pittsburgh, Pennsylvania 15213, USA

¹⁰Kimmel Center for Biology and Medicine at the Skirball Institute, Department of Pathology, New York University School of Medicine, New York, NY 10016.

*Correspondence should be addressed to H.W. (hong_wang@ncsu.edu)
Proper chromosome alignment and segregation during mitosis depend on cohesion between sister chromatids. Cohesion is thought to occur through the entrapment of DNA within the tripartite ring (Smc1, Smc3, and Rad21) with enforcement from a fourth subunit (SA1/SA2). Surprisingly, cohesin rings do not play a major role in sister telomere cohesion. Instead, this role is replaced by SA1 and telomere binding proteins (TRF1 and TIN2). Neither the DNA binding property of SA1 nor this unique telomere cohesion mechanism is understood. Here, using single-molecule fluorescence imaging, we discover that SA1 displays two-state binding on DNA: searching by 1-dimensional free diffusion versus recognition through subdiffusive sliding at telomeric regions. The AT-hook motif in SA1 plays dual roles in modulating nonspecific DNA binding and subdiffusive dynamics over telomeric regions. TRF1 tethers SA1 within telomeric regions that SA1 transiently interacts with. SA1 and TRF1 together form longer DNA-DNA pairing tracts than with TRF1 alone, as revealed by atomic force microscopy imaging. These results suggest that at telomeres cohesion relies on the molecular interplay between TRF1 and SA1 to promote DNA-DNA pairing, while along chromosomal arms the core cohesin assembly might also depend on SA1 1-D diffusion on DNA and sequence specific DNA binding.
INTRODUCTION

In eukaryotes, proper chromosome alignment and segregation during mitosis depend on cohesion between sister chromatids (1-3). Cohesion is mediated by the cohesin complex, which also plays important roles in other diverse biological processes, including double-strand DNA repair and maintenance of 3-dimensional chromatin organization (4,5). In vertebrates, the core cohesin complex consists of a tripartite ring assembled by Smc1, Smc3, Rad21 (also known as Scc1) and the stromal antigen subunit (SA) SA1 (STAG1) or SA2 (STAG2) (3).

In addition to association at centromeres, cohesin complexes are distributed at low densities along chromosome arms (6). This observation implies a low coverage of cohesin rings at telomeres. Telomeres are nucleoprotein structures that prevent the degradation or fusion of linear chromosome ends by preventing them from activating the DNA damage response and double-strand DNA break repair machineries (7-10). Human telomeres contain ~2 to 20 kb of TTAGGG repeats and a G-rich 3' overhang (11). In humans, a specialized protein complex called shelterin (consisting of TRF1, TRF2, POT1, TIN2, TPP1, and RAP1) regulates telomerase access, DNA damage response, and sister chromatid cohesion at telomeres (12-16).

Aging or disease associated telomere shortening contributes to genome instability and cancer progression by inducing chromosome end resection, fusion, and breakage (17). G-quadruplex (G4) and intermediate structures present during G4 formation cause chromosome fragility and replication fork stalling at telomeres (18,19). However, the cohesion process can counteract these effects by facilitating the restart of stalled replication forks (5,20). This function highlights the important role that the cohesion process plays at telomeres.

Our previous studies revealed that SA1 is required for telomere cohesion whereas, SA2 is required at centromeres (21). Depletion analysis showed that telomeres relied heavily on SA1 and to a lesser extent on the ring for cohesion (22). While deletion of cohesin ring subunits or SA2 dramatically decreases cohesion at centromeres, it does not significantly affect sister telomere association (22). Furthermore, SA1, not SA2, functionally interacts with TRF1 and
TIN2 (23). Beyond its function at telomeres, SA1 is enriched at promoters and CCCTC-binding factor (CTCF) sites, which in turn determines the distribution of cohesin complexes along chromosomes (24). However, the DNA binding properties of SA1, are unknown, but have important implications for advancing our understanding of the mechanism underlying sister chromatid cohesion and its contribution to chromosome architecture determination (25). Furthermore, it is unclear if SA1 specifically recognizes telomeric DNA sequences, or if TRF1 influences SA1’s interactions with telomeric DNA. Importantly, it is not fully understood how sister telomere cohesion is achieved through SA1 in conjunction with the shelterin proteins TRF1 and TIN2.

Here we used fluorescence imaging to study quantum dot- (QD-) labeled SA1 on DNA containing alternating telomeric and non-telomeric sequences. This platform was used to investigate how SA1 achieves DNA binding specificity for telomeric sequences by itself and in partnership with TRF1. We discovered that SA1 displays two-state binding on DNA: fast searching using 1-dimensional unbiased diffusion and reading (recognition) at telomeric regions using a slow subdiffusive sliding mechanism. The N-terminal domain of SA1 (SA1-N) containing the AT-hook motif mediates both the nonspecific binding and subdiffusive diffusion modes. Monte Carlo simulations using a two-state model for SA1 (free 1-D diffusing during searching and pausing during reading) suggest that the slow subdiffusive behavior can be explained by higher probabilities of pausing events at telomeric sequences. Furthermore, we found that the presence of TRF1 tethers SA1 within the telomeric region, while individual SA1 molecules diffuse through multiple telomeric and non-telomeric regions on DNA tightropes. Using AFM imaging we found that TRF1 and SA1 together promote longer protein-mediated DNA-DNA pairing tracts compared with TRF1 alone. Taken together, these results directly revealed the molecular interplay between SA1 and TRF1 in telomeric DNA binding and in promoting DNA-DNA pairing during sister telomere cohesion. Importantly, these results strongly suggest a new
model for cohesin assembly that takes into consideration the 1-D diffusion of SA1 on DNA and its sequence specific binding.

MATERIALS AND METHODS

Protein-QD conjugation
Streptavidin-conjugated quantum dots (SAv-QDs) and secondary antibody-coated quantum dots (Ab-QDs) were purchased from Invitrogen. Flag and SUMO antibodies were purchased from Sigma and Santa Cruz Biotechnology, respectively. For labeling N-terminal 3X Flag-tagged WT SA1, SA1 R37A R39A (GenScript), or SUMO-tagged SA1-N (SA1-N, 1-72 AA), QDs (1 µL of 1 µM) were incubated with the primary antibodies (1 µL of 1 µM) for 20 min. Proteins (1 µL, 0.7 µM Flag-SA1 or 1 µM SUMO-SA1-N) were added to the solution and incubated for an additional 20 min. For single-color QD labeling of N-terminal His<sub>6</sub>-tagged full length SA1 (His-SA1), 1 µL of SAv-QD (1 µM, Invitrogen) was incubated with 1 µL of the multivalent chelator tris-nitrotriacetic acid (tris-NTA, 2 µM) for 20 min (26). His-SA1 protein (1 µL of 1 µM) was then added to the SAv-QD-NTA solution and incubated for an additional 20 min. Experiments were carried out using equal molar concentration of green (565 nm) and red (655 nm) QDs. A higher percentage of dual-color labeling of Flag-SA1 molecules (46%, N = 107) was present compared to His-SA1 (10%, N = 67). This difference is most likely due to the simultaneous binding of two Ab-QDs to the 3X Flag tag on SA1. For dual-color differential labeling of Flag-SA1 and N-terminal His-tagged TRF1 (His-TRF1), additional His-tagged single-chain antibody fragments (2 µL of 5 µM) and dithiothreitol (DTT, 5 µL of 2 mM) were added sequentially to the SA1-QD solution and incubated for 20 min at each step. The addition of DTT and the antibody fragments is to prevent His-TRF1 from nonspecifically binding to Ab-QDs through the metal-histidine coordination (27). The Flag-SA1-QDs and His-TRF1-QDs were prepared separately, then mixed and incubated for an additional 20 min. All samples were diluted 200X for WT SA1 and 20X for SA1 R37A R39A mutant protein, before being introduced into the flow cell in the imaging buffer (20 mM Tris (pH
7.5), 100 mM KCl, and 0.1 mM MgCl\(_2\). Formation of dual-color QD-labeled complexes on T270 tightropes depends on the presence of both TRF1 and SA1, which shows that the crosstalk between His-TRF1 and Ab-QDs is not significant.

**AFM imaging and analysis**

All DNA and protein samples were diluted in 1X AFM buffer [25 mM HEPES–KOH (pH 7.5), 25 mM NaOAc, and 10 mM Mg(OAc)\(_2\)] before being deposited onto freshly cleaved mica surface (SPI Supply). Then, the samples were washed with MilliQ water and dried under a stream of nitrogen gas. The final protein, QD, and DNA concentrations were 6.7 nM, 6.7 nM, and 0.58 µg/ml, respectively. When QDs were not included, the final concentrations of SA1 and TRF1 proteins were 30 nM and 50 nM respectively. All images were collected in the AC mode using a MFP-3D-Bio AFM (Asylum Research). Pointprobe\(^\text{6}\) PPP-FMR probes (Nanosensors) with spring constants at ~2.8 N/m (nominal value) were used. All images were captured at a scan size of 1-3 µm × 1-3 µm, a scan rate of 1-2 Hz, and a resolution of 512 pixels × 512 pixels. The positions of proteins and protein-QDs on DNA were analyzed using the software from Asylum Research.

**Fluorescence imaging and analysis**

The oblique angle total internal reflection microscopy based particle tracking of QD-labeled proteins on DNA tightropes was described previously (28).

The mean square displacement (MSD) as a function of time interval is given by:

\[
MSD(n\Delta t) = \frac{1}{N-n} \sum_{i=n}^{N} \left[ (x_{i+n} - x_i)^2 + (y_{i+n} - y_i)^2 \right]
\]

**Equation 1**

where N is the total number of frames in the trajectory, n is the number of frames for different time intervals, \(\Delta t\) is the time between frames, and \(x_i\) and \(y_i\) are the positions of the protein-QD in the frame \(i\). The 1-D diffusion constant \(D\) and diffusion exponent (alpha factor) were analyzed by a custom routine developed in LabView Software based on the following equation (29):

\[
MSD = 2Dt^{\alpha}
\]

**Equation 2**
A protein was categorized as being mobile if the diffusion constant was greater than $5 \times 10^{-4}$ $\mu$m$^2$/s and $R^2$ value from data fitting using Equation 2 was greater than 0.8. To detect slow diffusion events on DNA based on the time interval-based diffusion constant ($D_{\text{int}}$), we developed a custom MATLAB code to execute “sliding window” (40-frame, 2 s) MSD analysis (34).

RESULTS

Full length SA1 binds specifically to telomeric DNA sequences

Recently, we demonstrated that the N-terminal domain of SA1 (1-72 AA, SA1-N) binds to a DNA substrate containing telomeric sequences (22). However, whether or not the full length SA1 binds specifically to telomeric DNA was unknown. To understand the SA1 DNA binding mechanism, we obtained full length His- and Flag-tagged SA1 proteins using the baculovirus/insect cell expression system (Supplementary Methods, Supplementary Fig. S1A). On a gel filtration column, the full length His-SA1 eluted at a mean volume consistent with a monomeric protein (Supplementary Fig. S1B). Fluorescence anisotropy experiments demonstrated that the equilibrium dissociation constant of SA1 for a duplex DNA substrate containing 7 TTAGGG repeats ($K_d = 34.0 \pm 5.8$ nM, mean ± SEM) was ~3-fold lower than that obtained for a DNA substrate of the same length but with scrambled sequences ($K_d = 104.0 \pm 13.6$ nM, Fig. 1A). In contrast, SA1 did not show specificity for CTCF consensus sequences (Supplementary Fig. S2). To study SA1 binding to longer telomeric substrates, we carried out AFM imaging and statistical analysis (Materials and Methods) of the binding position of SA1 on a telomeric DNA substrate containing 270 TTAGGG repeats (T270, 5.4 kb) and a control substrate (3.8 kb) containing only the non-telomeric (genomic) DNA sequences from T270 (Fig. 1B). AFM imaging indicated that a higher percentage (41.9%) of SA1 bound at the telomeric region on the T270 DNA substrate compared to the same locations along the genomic DNA substrate (27.0%). In summary, both fluorescence anisotropy and AFM imaging showed that full
length SA1 binds to telomeric sequences with weak (~2-3 fold) specificity.

**SA1 alternates between fast and slow diffusion on DNA containing telomeric sequences**

Dynamic movements on DNA, such as 1-dimensional (1-D) sliding (translocation while maintaining continuous DNA contact), jumping, and hopping (microscopic dissociation and rebinding events) are essential for a protein to find its target sites on DNA (30-35). To further understand how SA1 dynamically achieves DNA binding specificity for telomeric sequences, we used oblique angle fluorescence microscopy imaging of QD-labeled proteins on T270 DNA tightropes containing alternating telomeric and genomic regions (Materials and Methods) (28,36-38). Hydrodynamic flow was used to stretch DNA and suspend ligated T270 DNA strands between poly-L-lysine coated silica microspheres at an elongation of ~90% of the DNA contour length (Fig. 2A, Supplementary Methods) (28). We conjugated Flag-SA1 to secondary antibody-coated QDs (Ab-QDs) using an antibody sandwich method (Fig. 1C) (26). AFM imaging revealed that Ab-QDs without SA1 did not have significant binding affinity for T270 DNA, and SA1-QDs retained binding specificity for telomeric sequences (Fig. 1C). The binding of SA1-QDs molecules to DNA was long lived, with 78.5% (N = 107) of SA1-QD complexes remaining on T270 DNA tightropes after 2 minutes (Supplementary Fig. S3A).

Analysis of SA1 on T270 DNA revealed two populations (Fig. 2B-D): static and mobile molecules. Surprisingly, while some mobile SA1 molecules displayed free 1-D diffusion on T270 DNA throughout the entire observation period (2 min, Fig. 2C), a subpopulation of mobile SA1 molecules alternated between periods of slow and fast diffusion (Fig. 2D, Supplementary Movie S1). His-SA1-QDs also displayed bimodal diffusion behavior (fast and slow diffusion) on T270 DNA (Supplementary Fig. S3B). These results indicate that this type of bimodal diffusion behavior (fast and slow diffusion) is independent of QD conjugation strategies.

The diffusion range of SA1 molecules on T270 DNA tightropes covered distances from ~0.18 to 8.55 μm (~6 ligated T270 molecules), with some SA1 molecules visiting more than one
telomeric segment (1.6 kb each, Fig. 2D). Strikingly, some SA1 molecules repeatedly slowed down at the same regions along T270 DNA tightropes, manifested as distinct peaks in the position histogram of SA1 along DNA (Fig. 2D, top panel). The distribution of the pair-wise distances between these peaks exhibited two populations centered at 0.5 and 1.5 μm (Fig. 2E). These distances are consistent with the boundaries of the telomeric region (1.6 kb) and the spacing between two adjacent telomeric regions on T270 DNA (5.4 kb), respectively.

Furthermore, we also observed the alternation between fast and slow diffusion of SA1 at lower (50 mM KCl) and higher (150 mM KCl) salt concentrations (Supplementary Fig. S4). Collectively, these results show that SA1 alternates between fast and slow diffusion on DNA tightropes containing telomeric and non-telomeric sequences. The pairwise distance between slow diffusion events on T270 DNA indicates that slow diffusion events are more likely to occur at telomeric regions.

Long SA1 slow diffusion events depend on telomeric sequences

To investigate whether the slow diffusion events displayed by SA1 on DNA depend on the presence of TTAGGG repeats, we imaged QD-labeled SA1 on DNA tightropes containing only the genomic sequence portion of the T270 DNA substrate or centromeric DNA sequences (Fig. 3A). On genomic and centromeric DNA tightropes, SA1 molecules showed different DNA binding dynamics compared with T270 DNA (Fig. 3). The percentage of SA1 molecules displaying static binding on T270 DNA was ~3-fold and 9-fold higher than on genomic and centromeric DNA, respectively (Fig. 3B).

To further compare the dwell times of the slow diffusion events displayed by mobile SA1 on different DNA substrates, we used “sliding window” (40-frame, 2 s) MSD analysis to calculate a time interval-based diffusion constant (D_m, bottom panels in Fig. 2C and 2D) (39). Distinct from static (< 0.5 X 10^{-3} μm^2/s, Supplementary Fig. S5A) and fast free diffusion modes (> 10 X 10^{-2} μm^2/s, Supplementary Fig. S5B), this analysis indicated that mobile SA1 molecules with fast
and slow diffusion on T270 DNA show a distinct peak at \( \sim 1.0 \times 10^{-3} \ \mu \text{m}^2/\text{s} \) (Supplementary Fig. S5C). Therefore, for calculating dwell times on DNA tightropes we used \( D_{\text{int}} \) value of \( 5.0 \times 10^{-3} \ \mu \text{m}^2/\text{s} \) as the threshold to identify individual slow diffusion events. This \( D_{\text{int}} \)-based analysis showed that the dwell times of individual SA1 slow diffusion events on T270 (1.17 s) are significantly (\( p < 0.05 \)) longer than on genomic (0.80 s) and centromeric (0.79 s) DNA (Supplementary Fig. S6A). The percentage of mobile SA1 molecules showing long slow diffusion events (\( D_{\text{kin}} < 5.0 \times 10^{-3} \ \mu \text{m}^2/\text{s} \) for longer than 2.2 s) on T270 DNA (51.2%) was at least \( \sim 2\)-fold higher than on genomic (17.6%) or centromeric DNA (24.1%, Fig. 3C). Furthermore, among all mobile SA1 molecules analyzed, SA1 spent a significantly larger percentage of time (24.4%) in the slow diffusion mode on T270 DNA than on genomic (3.2%) and centromeric DNA (6.3%, Supplementary Fig. S6B).

Using MSD analysis we further compared the dynamics of SA1 on different DNA substrates. SA1 displayed significantly (\( p < 0.005 \)) slower diffusion constants on T270 DNA (0.04 \( \pm \) 0.01 \( \mu \text{m}^2/\text{s} \)) in comparison with DNA substrates containing genomic (0.14 \( \pm \) 0.03 \( \mu \text{m}^2/\text{s} \)) or centromeric (0.11 \( \pm \) 0.02 \( \mu \text{m}^2/\text{s} \)) DNA sequences (Table 1). In addition, we calculated the diffusive exponent (alpha factor) to determine whether SA1 displays subdiffusive motion on DNA (Table 1). An alpha factor of 1 indicates an unbiased random walk and a value less than 1 indicates periods of pausing in the random walk (subdiffusion) (29). On T270 DNA, SA1 displayed significantly (\( p < 0.001 \)) smaller alpha factors (0.69 \( \pm \) 0.03) compared to genomic (0.89 \( \pm \) 0.02) and centromeric (0.82 \( \pm \) 0.02) DNA substrates (Table 1). In summary, fluorescence imaging of QD-labeled proteins established that SA1 alternates between slow and fast diffusion on DNA. These slow diffusion events are telomere sequence dependent. Additionally, the alpha factor for SA1 on T270 was significantly smaller than on genomic DNA, which suggests protein pausing amid free diffusion at telomeric sequences.

**SA1 slow diffusion events are mediated through its N-terminal domain**
SA1 contains a unique AT-hook motif at its N-terminal domain, which is not present on SA2 (22). The AT-hook domain has been proposed to serve as an accessory domain for transcription factors to bind specific DNA sequences/structures (40). To determine whether or not SA1 slow diffusion events depend on its unique N-terminal domain, we purified the SUMO-tagged SA1 N-terminal fragment (SA1-N) (22). Analysis of the binding position of SA1-N on T270 DNA in AFM images demonstrated that SA1-N binds specifically to the telomeric regions (Fig. 4A and Supplementary Fig. S7A). These results are consistent with previous electrophoresis mobility shift assays (EMSAs) indicating SA1-N binding to telomeric sequences (22).

For fluorescence imaging, we labeled N-terminal SUMO-tagged SA1-N with Ab-QDs through a primary antibody against the SUMO tag (Fig. 4B, Supplementary Methods). Consistent with results from AFM imaging, incubation of SA1-N-QDs with T270 DNA tethers resulted in substantial DNA binding. Importantly, SA1-N also displayed slow diffusion events on T270 DNA (Fig. 4C, Supplementary Movie S2). The static binding and long slow diffusion events (> 2.2 s) on T270 displayed by SA1-N also were telomeric sequence dependent (Supplementary Fig. S7B&C). A significantly higher percentage of SA1-N molecules (74.3%) showed slow diffusion events (dwell time > 2.2 s) on T270 DNA than on genomic DNA (38.6%). Consistent with these results, the diffusion constant of mobile SA1-N on T270 DNA (0.06 ± 0.01 μm²/s) was significantly (p = 0.001) lower than that on genomic DNA (0.12 ± 0.02 μm²/s, Table 1). Furthermore, the alpha factor of SA1-N 1-D diffusion (0.74 ± 0.03) was significantly smaller (p < 0.001) on T270 than on genomic DNA (0.89 ± 0.02, Table 1).

Previously, it was shown that the mutations at the central core sequence (KRKRGGRP) in the SA1 AT-hook motif (SA1-N R37A R39A) significantly reduces its binding to telomeric DNA (22). To further understand the role of SA1’s AT-hook motif in DNA binding, we obtained the full length Flag-tagged SA1 R37A R39A mutant (Supplementary Fig. S1A). Fluorescence anisotropy experiments showed that the double mutations at the AT-hook domain reduced the
$K_d$ of SA1 for genomic DNA and telomeric DNA by approximately 9.3- and 6.5-fold, respectively (Fig. 4D). Intriguingly, on the control DNA tightropes containing nonspecific genomic and centromeric sequences, QD-labeled SA1 R37A R39A diffused significantly slower (0.03 ± 0.01 μm$^2$/s, $p < 0.0003$) than the WT protein (0.11 ± 0.02 μm$^2$/s, Table 1). These results demonstrated that the core AT-hook motif plays an important role in nonspecific DNA binding and in promoting mobility of SA1 on DNA. Meanwhile, it is worth noting that SA1 R37A R39A still showed specificity (~4X) for telomeric sequences (7 TTAGGG repeats), even though the overall DNA binding was reduced (Fig. 4D). At the single-molecule level, QD-labeled SA1 R37A R39A also displayed alternating slow and fast diffusion events on T270 DNA tightropes (Fig. 4C). However, compared to the WT protein, the ability of SA1 R37A R39A to carry out subdiffusive diffusion on DNA is significantly compromised. Compared to the WT protein (64%), the percentage of mobile SA1 R37A R39A molecules showing slow diffusion events (> 2.2 s, 34%, $N = 86$) was significantly decreased for both telomeric and centromeric DNA substrates. In addition, the alpha factor of SA1 R37A R39A on T270 (0.79 ± 0.05) was significantly higher ($p < 0.003$, Table 1) than that for WT SA1 (0.69 ± 0.03). In addition, with R37A R39A mutations, the difference between alpha factors displayed by SA1 on T270 (0.79 ± 0.05) and the control DNA substrate (0.83 ± 0.03) was diminished (Table 1). Collectively, the fluorescence anisotropy and single-molecule fluorescence imaging results demonstrated the dual roles of the central core sequence (KRKRGGRP) at the AT-hook motif in achieving high affinity nonspecific SA1 DNA binding and modulating telomere sequence dependent subdiffusive behavior on DNA.

**SA1 becomes subdiffusive or static within telomeric regions in the presence of TRF1**

SA1 interacts directly with TRF1 through its N-terminal domain (23). EMSAs using a DNA substrate containing 3 TTAGGG repeats showed that TRF1 and SA1 together induced a supershift relative to TRF1-DNA and SA1-DNA complexes (Fig. 5A). This result suggests that TRF1 and SA1 can interact simultaneously with the same piece of DNA.

To evaluate how TRF1 affects the dynamics of SA1 on DNA, we directly imaged their
interactions. Flag-SA1 and His-TRF1 proteins were orthogonally conjugated with red Ab-QDs and green SAv-QDs via antibody sandwich and βTris-NTA linkage strategies, respectively (Fig. 5B and Movie S3). Under these conditions, 11.2% of the total protein-QDs on T270 DNA tightropes were dual-color labeled, which were dependent on the presence of both TRF1 and SA1 (Supplementary Fig. S8A). A higher population of dual-colored SA1-TRF1-QD complexes (~60%) were static than the single-colored SA1-QD alone (~35%) on T270 DNA tightropes (Fig. 5D). On T270 DNA tightropes, the majority (68.2%) of the mobile SA1-TRF1 complexes diffused within a range less than the length of telomeric region (~0.5 μm) during the entire observation time window (2 mins, bottom panel of Fig. 5C, Fig. 5E). In stark contrast, the majority (73.3%) of SA1 alone molecules (single-color labeled) diffused through multiple telomeric and non-telomeric regions (with diffusion ranges > 0.5 μm, Fig. 5C top panel, Fig. 5E). These results suggest that the narrow diffusion range displayed by SA1-TRF1 is due to interactions between TRF1 and SA1. To further confirm that the confined motion displayed by SA1-TRF1 is telomeric sequence dependent, we studied the diffusion range of dual-colored SA1-TRF1-QDs on genomic DNA. Strikingly, on genomic DNA, only 6.7% of mobile SA1-TRF1 protein complexes were confined to a range less than 0.5 μm (Fig. 5E). Furthermore, the distance between dual-color QD-labeled SA1-TRF1 complexes and green QD-labeled TRF1 was consistent with that of SA1-TRF1 complexes binding to telomeric regions (Supplementary Fig. S8B). Therefore the restricted diffusion range is not a property of SA1-TRF1-QDs, but instead is related to the DNA binding energy landscape over telomeric DNA sequences. In addition, the diffusion constant and alpha factor of mobile SA1-TRF1 complexes were significantly (p < 0.05) smaller on T270 DNA than on genomic DNA tightropes (Fig. 5F). Taken together, these results reveal that TRF1 is required to hold SA1 at the telomeric region.

**SA1 and TRF1 together facilitate DNA-DNA pairing**
TRF1 forms protein filaments on DNA and promotes parallel pairing of telomeric tracts (41). To elucidate the role of SA1-TRF1 interactions in sister telomere cohesion, we used AFM to investigate whether or not SA1 influences TRF1 mediated DNA-DNA pairing. Consistent with previous results (41), we observed that TRF1 molecules formed protein tracts that mediate DNA-DNA pairing on T270 DNA (Fig. 6A). TRF1 protein tracts displayed average heights of 0.73 (± 0.10, mean ± SD) nm and average tract lengths of 66 (± 4) nm (Fig. 6D and Supplementary Fig. S9). In contrast to TRF1 (50 nM), no protein tracts were observed when SA1 alone (30 nM) was incubated with T270 DNA. Importantly, in the presence of both TRF1 and SA1 (Fig. 6B&C), the average SA1-TRF1 mediated DNA-DNA pairing tract length increased significantly to 92 (± 7) nm (Fig. 6D and Supplementary Fig. S9). Due to its larger molecular weight, SA1 (142 kDa) on T270 DNA can be identified as individual proteins with heights (1.39 ± 0.50 nm, mean ± SD) significantly (p < 0.05) higher than TRF1 protein tracts (Supplementary Fig. S9C). The location of SA1 on TRF1-mediated DNA-DNA pairing tracts was random. In the presence of SA1 R37A R39A and TRF1, the protein-mediated tract on T270 substrate (78 ± 7 nm, N = 16) was comparable with TRF1 alone. This result is consistent with our previous observation based on telomere Fluorescence in situ hybridization (FISH) showing that SA1 R37A R39A mutations abrogate the ability of SA1 to induce persistent cohesion at telomeres (22). In summary, SA1 and TRF1 together promote DNA-DNA pairing and the enhancement of TRF1-mediated DNA-DNA pairing depends on DNA binding by SA1.

**Coarse-grained molecular dynamics (MD) simulations of synergistic effects of TRF1 and SA1 DNA binding**

To further test the model that SA1 and TRF1 together bind to telomeric DNA and promote DNA-DNA pairing, we performed coarse-grained molecular dynamics (MD) simulations using HOOMD-blue (Supplementary Methods) (42). SA1 is modeled as a cubic, rigid domain that carries a DNA-binding site within a groove (Fig. 7A). The groove enables SA1 to diffuse along
DNA, but prevents bound molecules from bypassing each other. A previous electron microscopy study showed that a TRF1 dimer can simultaneously bind two DNA binding sites with only loose constraints on the distance or orientation between these two sites (43). Therefore, we constructed a TRF1 model containing a DNA binding domain similar to SA1 (representing the Myb domain), a flexible linker, and a dimerization domain which also carries an anionic group (charge bead representing the acidic N-terminal domain, Fig. 7B). We modeled the interaction between SA1 and TRF1 by allowing heterodimerization between the two proteins at the DNA-binding domain. Two 1800 bp DNA strands were modeled as semi-flexible strings of anionic 1-nm beads, which carry blocks corresponding to high affinity binding regions representing telomeric DNA sequences (1200 bp) and low affinity binding regions representing genomic DNA (600 bp). Based on previous studies of TRF1 DNA binding and fluorescence anisotropy measurements of SA1 DNA binding affinity (Fig. 1A) (44,45), DNA binding constants where chosen so that binding energies followed the order: (TRF1 on genomic) < (SA1 on genomic) < (SA1 on telomeric) < (TRF1 on telomeric DNA).

Initially, we used MD simulations to establish TRF1 DNA binding modes that included TRF1 dimer-mediated DNA-DNA pairing (Fig. 7C, left panels) (41). Next, we simulated SA1-DNA binding (Supplementary Fig. S10A) and tested two computational scenarios: SA1+TRF1 with heterodimerization (Fig. 7C, right panels, Supplementary Fig. S10B) and SA1+TRF1 without heterodimerization. Under the simulation conditions, when only SA1 was present in the simulations, 34% of SA1 was bound to DNA (Supplementary Fig. S10A). In stark contrast, in the scenario of heterodimerization between SA1 and TRF1, all of the SA1 molecules were associated with telomeric DNA and bound to TRF1 (Supplementary Fig. S10B). After binding to DNA, SA1 molecules diffused along the DNA, until captured by a TRF1 molecule, which was typically part of a TRF1 dimer. Meanwhile, in the scenario of an absence of heterodimerization between TRF1 and SA1, only 18% of SA1 was bound to DNA. Importantly, the results were identical at time steps of ~1, 2 and 4 ps. Thus, the simulations are robust and consistent with
the experimental data (Fig. 5) showing that the interaction with TRF1 strongly enhances SA1 binding specificity to telomeric sequences. We note that the protein-DNA interactions are dynamic in the sense that SA1 recruitment also influences TRF1 binding affinity, although the net effect is relatively small due to the excess of TRF1 in the current MD simulation models.

**DISCUSSION**

Cohesin rings do not play a major role in sister telomere cohesion (22). Instead, this role is taken over by the cohesin subunit-SA1 and shelterin proteins. Shelterin proteins normally protect telomeres from DNA damage responses. However, their role in sister telomere cohesion is unclear. Here, we used single-molecule imaging to shed new light on how SA1 achieves telomeric DNA binding specificity, and how TRF1 modulates SA1 DNA binding dynamics. These results demonstrated that SA1 and TRF1 function together in binding to telomeric DNA and promoting DNA-DNA pairing.

**Monte Carlo simulations suggest a two-state model for SA1 DNA binding**

SA1 and TRF1 have distinct binding patterns along the chromosomes. *In vivo*, TRF1 is found at telomeres (46). This property derives from the high propensity of TRF1 to remain within telomeric regions due to a rougher diffusion energy landscape within these regions compared to genomic sequences (28). In contrast, SA1 has wider roles *in vivo* exemplified by its distribution along chromosome arms (47). Consistent with the roles of SA1 *in vivo*, in this study we found that SA1 diffuses through both telomeric and non-telomeric regions. The diffusion constant of SA1-QDs on T270 tightropes (∼ 0.04 to 0.11 μm²/s) was comparable with QD-labeled DNA repair proteins including Mlh (0.137 μm²/s), Mlh1-Pms1 (0.02 – 0.99 μm²/s), and MutS (0.036 μm²/s) (48). Based on the Stokes-Einstein relation, the diffusion constant of free SA1 was estimated to be ∼0.22 μm²/s (using the hydrodynamic radius of SAv-QDs: 10.5 nm; and the dimension of SA1 ortholog, yeast Scc3: radius at 6.2 nm) (49-51).
Furthermore, the alternating telomeric and non-telomeric DNA sequences on DNA tightropes allow us to pinpoint specific DNA binding events by correlating the distances between repeated transient DNA binding dynamics with specific DNA sequences. Importantly, for SA1 on T270, we observed defined distances between slow diffusion events consistent with spacing between telomeric regions. These results demonstrate that SA1 diffuses more slowly over telomeric sequences. In addition, on T270 SA1 diffusion shows alpha factors less than one, suggesting that SA1 pauses amid free diffusion on DNA. These slow diffusion events are not due to the hydrodynamic drag on QDs and the flexible linker between SA1 and QDs. Using dissipative particle dynamics (DPD) simulations, we found that the linker extension was independent of the nanoparticle size, and under no circumstances did a nanoparticle tag lead to pausing (Riehn, unpublished results). To provide a mechanistic basis for the observed slow diffusion at telomeric sequences, we applied Monte Carlo simulations and modeled SA1 existing in one of two DNA binding states: 1) statically bound (recognition or reading) mode; 2) freely diffusing (search) mode (Supplementary Methods, Supplementary Fig. S11). We assigned less frequent pausing (8%) on non-telomeric DNA for SA1. As SA1 passes into a telomeric region the equilibrium is shifted towards pausing with a 80% probability of entering the paused state. By shifting the equilibrium between these two states we were able to reproduce the diffusive behavior of SA1 with kymographs displaying periods of slow diffusion that coincide with the positions of the telomeric regions (Supplementary Fig. S11). The strong qualitative correlation between results from Monte Carlo simulations and single-molecule fluorescence imaging suggests that the two-state model is a plausible mechanism for the observed fast and slow diffusion events on telomeric DNA. Based on these Monte Carlo simulations, an important prediction is that DNA sequence dependent SA1-DNA interaction energy landscape will determine the statistical probability of SA1 being in the sub-diffusive/pausing state. Future experiments need to be carried out to test the SA1-DNA interaction energy landscape involving high frequency SA1 binding sites in vivo (47).
SA1 N-terminal AT-hook domain has dual roles in both specific and nonspecific DNA binding

The AT-hook is a small DNA binding motif that has been identified in proteins that play important roles in modulating chromatin structure or functioning as transcription factors (40). The NMR structure of an archetypal AT-hook protein HMG-I(Y) reveals that the AT-hook forms a C-shaped structure with the concave surface inserted into the DNA minor groove (52). These nonspecific interactions include hydrogen bonds or electrostatic interactions with the DNA phosphate groups. Similar to full length SA1, a higher percentage of SA1-N displays long slow diffusion events on T270 DNA compared to genomic DNA. These results indicate that the slow diffusion events displayed by full length SA1 are mediated at least partly through its N-terminal domain. Our results from using SA1 R37A R39A suggest that, similar to Lac repressor, the AT-hook motif in SA1 plays dual roles in nonspecific DNA binding as well as in mediating pausing events on DNA for achieving specific DNA binding. For Lac repressor, the same set of residues involved in the highly specific DNA binding mode can shift and twist to participate in nonspecific DNA binding (53). Our results are also consistent with the notion that thermally driven conformational fluctuations in the DNA binding domains are effectively coupled to the 1-D diffusion-mediated target sequence search (54). It has been suggested that there are at least two different conformations of DNA binding domains. A DNA binding domain less ordered in structure enables the fast-diffusing state, while more ordered in structure facilitates recognition of specific DNA binding sites. Furthermore, the function of the AT-hook domain in SA1 is reminiscent of the key “wedge residue” of Fpg, Nel, and Nth DNA glycosylases, which modulates the diffusive behavior of these proteins in searching for DNA lesions (39). A two-state search model has also been proposed for the transcription activator-like effector (TALE) proteins (55). Future experiments could further investigate whether the two-state (searching and recognition) model is universal for other DNA binding proteins containing the AT-hook motif.
TRF1 confines the diffusion of SA1 within telomeric regions

As a result of the fast alternation between static and free diffusion, SA1 alone can diffuse across multiple telomeric and non-telomeric regions and only temporarily slows down at telomeric sequences. This study provides the experimental evidence for how TRF1 makes SA1 a dedicated telomere binding protein. The observation that TRF1 tethers SA1 at telomeric DNA regions and reduces its probability of entering into non-telomeric DNA regions offers a mechanism by which SA1 can switch from the role of a protein functioning along chromosome arms to one specific to telomeres.

Consistent with previous results (41), we observed TRF1-mediated DNA-DNA pairing in our AFM images. Interestingly, we also found that the addition of SA1 significantly increased (p < 0.05) the DNA-DNA pairing tract length, as compared to TRF1 alone. This enhancement depends on SA1 DNA binding. SA1 R37A R39 with diminished DNA binding affinity failed to enhance TRF1-mediated DNA-DNA pairing. Under our imaging conditions (6.7 nM DNA, 30 nM TRF1 and 50 nM SA1), while TRF1 molecules formed closely spaced molecules over the DNA-DNA pairing region, SA1 sparsely decorated TRF1 protein tracts at random positions. It is worth noting that the protein concentrations used in our single-molecule experiments are comparable with estimated TRF1 concentration in vivo (56). Coarse-grained MD simulations suggest a model where TRF1 stabilizes SA1 at telomeric sequences and enhances SA1 telomeric DNA binding specificity. Due to their combined DNA binding energy, SA1-TRF1 complexes become more stable than either TRF1 or SA1 alone on telomeric DNA. One mechanism of SA1 action would be to enhance TRF1 dimer-DNA linkages contributing to the stability and therefore the increase in the length of the DNA-DNA pairing region. Future studies are needed to understand how TIN2 modulates SA1-TRF1-DNA structures and sister telomere cohesion in the context of heterochromatin protein 1γ (HP1γ) and nucleosomes (23,57).

SA1 DNA binding in the context of the core cohesin complex
Cohesin-SA1 and SA2 subunits play different roles in sister chromatid cohesion, DNA repair, and transcription regulation (24,58). SA1 deletion causes embryonic lethality (47), demonstrating that SA2 cannot replace the function of SA1. Besides its function at telomeres, SA1 is enriched at promoters and CTCF binding sites. However, by itself, SA1 does not display specificity for CTCF consensus sequences. These results suggest that SA1 localizes at CTCF binding sites through protein-protein interactions. SA1 also displays slow diffusion events on genomic DNA, albeit with shorter dwell times and lower frequencies than on the T270 DNA substrate. This raises the possibility that, much like at telomere sequences, frequent pausing of SA1 at AT-rich promoters can lead to targeting of the core cohesin complexes to these regions. A longer dwell time of SA1 at AT-rich sequences in its recognition (reading) mode could permit the assembly of specific structures to carry out unique functions across the genome. Therefore, we propose a model in which SA1 is the “DNA sequence guide” (using its AT-hook motif) for the core cohesin complex. It directs the loading of the core cohesin complexes at specific sequences along the genome. Finally, these results support a revised model for cohesin assembly that requires both 1-D searching on DNA and DNA sequence specific binding by SA1.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR online, including Supplementary Methods, Supplementary Figures S1–11, Supplementary Legends for Movies S1–3, and Supplementary References [59-60].

FUNDING
This work was supported by the National Institutes of Health [R01ES022944 to P.L.O., R00ES016758 to H.W, R01GM107559 to R.R. and H.W, R01CA116352 to S.S., P30 ES025128 to CHHE at NCSU], and the Welch Foundation [C-1565 to Y.J.T]. H.C. was supported by the China Scholarship Council.

ACKNOWLEDGEMENTS
We would like to thank the Weninger and Riehn groups at NCSU for technical support, Wei Qian at the University of Pittsburgh for purifying the genomic DNA, and Christian White at NCSU for cloning the centromeric DNA substrate. We also thank Justin Lormand for purifying the TRF1.

Conflict of interest statement. None declared.
FIGURE LEGENDS

Figure 1. Full length SA1 and SA1-QDs bind to telomeric DNA sequences with weak specificity. (A) Fluorescence anisotropy experiments showing concentration-dependent binding of 3xFlag-tagged SA1 (Flag-SA1) to DNA substrates containing 7 TTAGGG repeats (Tel7, red dots) or scrambled DNA sequences (blue dots). The data were fitted to the law of mass action (Supplementary Methods). The equilibrium dissociation constants are 34.0 ± 5.8 and 104.0 ± 13.6 nM, respectively, for DNA substrates with telomeric and scrambled sequences (two independent experiments). (B and C) Representative AFM images (left panels) and statistical analysis of the full length Flag-SA1 (B) and SA1-QDs (C) (white arrows) binding to the T270 DNA substrate. Schematic drawings of the T270 DNA substrate and the QD conjugation strategy (an antibody-sandwich method for the full length Flag-SA1) are shown in the top panels in B and C. The purple arrows point to the estimated boundaries between the genomic and telomeric sequences. Each data set was from at least three independent experiments with error bars representing SEM.

Figure 2. Full length SA1 alternates between fast and slow 1-D diffusion on T270 DNA tightropes. (A) Ligated T270 DNA substrate (top panel) and the DNA tightrope assay setup (bottom left panel). (B-D) Dynamics of the full length Flag-SA1 on T270 DNA tightropes. Kymographs of SA1 molecules being static (B), showing free 1-D diffusion (C), and alternating between fast and slow 1-D diffusion (D) on T270 DNA. Scale bars (y-axis): 1 μm. Equimolar concentrations of red (655 nm) and green (565 nm) QDs were used in SA1 conjugation. The bottom panels in C and D show corresponding plots of diffusion constants (D_{mm}) based on the 40-frame sliding window (2 s) MSD analysis. Each frame is 50 ms. The histogram on the right side of the top panel in D shows the position distribution of SA1 along the T270 DNA tightrope. (E) The distribution of pair-wise distance between nearest SA1 slow diffusion positions (N = 22).
Figure 3. Long slow diffusion events by SA1 on DNA depend on the presence of telomeric sequences. (A) Dynamics of the full length Flag-SA1 on DNA tightropes containing centromeric sequences. Top: a schematic drawing of the ligated DNA substrate containing centromeric sequences. Bottom: a kymograph showing a mobile QD-labeled Flag-SA1 complex on the centromeric DNA tightrope. The scale bar (y-axis): 1 μm. (B) Comparison of the static and mobile SA1 populations on telomeric, genomic, and centromeric DNA tightropes. The percentages of static SA1 molecules are 36% (N = 84), 11% (N = 63), and 4% (N = 97) on telomeric, genomic, and centromeric DNA, respectively. (C) The percentages of SA1 (out of all mobile molecules) with long slow diffusion events (> 2.2 s) on telomeric, genomic, and centromeric DNA. The mean values are 51.2% (N = 54), 17.8% (N = 56), and 24.1% (N = 93), for telomeric, genomic and centromeric DNA substrates, respectively (from 2 or 3 independent experiments). The error bars represent SEM.

Figure 4. The AT-hook motif in SA1 mediates nonspecific DNA binding and subdiffusive dynamics on telomeric T270 DNA substrate. (A) A representative AFM image showing SA1-N (white arrow) binding to T270 DNA. The purple arrows point to the estimated boundaries of the telomeric region. (B) QD conjugation strategy: an antibody-sandwich method for Sumo-tagged SA1-N. (C) Kymographs of QD-labeled SA1-N (green, top panel) and full length SA1 R37A R39 A (red, bottom panel) on T270 DNA tightropes. Scale bars (y-axis): 1 μm. (D) Fluorescence anisotropy experiments showing concentration-dependent binding of 3×Flag-tagged SA1 R37A R39A to DNA substrates containing 7 TTAGGG repeats (Tel7, left panel) or scrambled DNA sequences (right panel). The data were from two independent experiments.

Figure 5. TRF1 tethers SA1 within a telomeric region on DNA. (A) EMSA of TRF1 (350 nM) and SA1 (350 nM) in the presence of the Alexa 488-labeled DNA substrate containing 3 TTAGGG repeats (5 nM). (B) Schematic drawing of QD-labeling for TRF1 using the βTris-NTA linkage strategy. (C) Kymographs of Flag-SA1 and His-TRF1 proteins differentially conjugated
with red Ab-QDs and green SAv-QDs using the antibody sandwich (Fig. 1C) and 89Tris-NTA linkage strategies, respectively. SA1-QDs (43.8 nM) and TRF1-QDs (62.5 nM) were incubated together for 20 minutes before being diluted and injected into the flow cell containing DNA tethropes. (D) On T270 tethropes, dual-color labeled SA1-TRF1 complexes display a higher percentage of static complexes, compared to SA1 alone (red QD-labeled). The percentage of static complexes is 35% (±4%, N = 44) for SA1 only, and 60% (±14%, N = 47) for SA1-TRF1 complexes. (E) Dual-color QD-labeled SA1-TRF1 complexes are confined within a short range on T270 DNA, but not on genomic DNA (G-DNA). Dual-color labeled SA1-TRF1 complexes on T270 display a higher percentage of complexes (68.2%, N = 44) with diffusion range less than 0.5 μm, compared to SA1 alone on T270 (red QD-labeled, 26.7%, N = 43), or dual colored SA1-TRF1 on the genomic DNA (6.7%, N = 30). (F) On T270, dual color-labeled SA1-TRF1 complexes show slower diffusion constants (D = 0.0016 ± 0.0006 μm²/s, N = 44) and smaller alpha factors (0.69 ± 0.04), compared to complexes on genomic DNA (G-DNA, D = 0.03 ± 0.009 μm²/s, alpha factor = 0.91 ± 0.04, N = 30). Each data set was from at least three independent experiments.

**Figure 6. SA1 facilitates TRF1-mediated DNA-DNA pairing.** (A-B) AFM images of T270 DNA (6.7 nM) in the presence of TRF1 only (50 nM) (A), or both TRF1 (50 nM) and SA1 (30 nM) (B). The white arrows point to the protein complex mediated DNA-DNA pairing. (C) A 3-D image of the zoomed region from the top panel in B. (D) Protein tract lengths in the presence of only TRF1 (white bars) or both SA1 and TRF1 (blue bars). SA1 and TRF1 together increase the DNA-DNA pairing tract lengths to 92 ± 7 nm (N = 50) from 66 ± 4 nm (N = 40) for TRF1 alone. Each T270 DNA length was normalized to the mean length of the T270 DNA substrate (1.70 μm).

**Figure 7. MD simulations of TRF1 and SA1 DNA binding.** (A and B) Coarse-grained structural models for SA1 (A) and TRF1 (B) used in the simulations. (C) TRF1 dimer mediated
DNA-DNA pairing in the absence (left panels) or presence of interactions with SA1 (right panels). Examples of zoomed regions from MD simulations show single (top panels) and multiple TRF1-mediated DNA-DNA pairing (bottom panels). Overview pictures from MD simulations of SA1 alone and SA1+TRF1 are shown in Supplementary Fig. S10.
Table 1: Summary from analysis of diffusion dynamics of QD-labeled full length WT SA1 SA1-N, and SA1 R37A R39A on DNA tightropes.

<table>
<thead>
<tr>
<th>DNA</th>
<th>WT SA1-QDs</th>
<th>SA1-N-QDs</th>
<th>SA1 R37A R39A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D (µm²/s) Alpha factor N %</td>
<td>D (µm²/s) Alpha factor N %</td>
<td>D (µm²/s) Alpha factor N %</td>
</tr>
<tr>
<td>T270</td>
<td>0.04 ± 0.01 (0.02 to 0.06) 0.69</td>
<td>0.06 ± 0.01 (0.04 to 0.08) 0.74</td>
<td>0.03 ± 0.01 (0.01 to 0.05) 0.79</td>
</tr>
<tr>
<td>Genomic</td>
<td>0.14 ± 0.03 (0.08 to 0.20) 0.89</td>
<td>0.12 ± 0.02 (0.08 to 0.16) 0.89</td>
<td>0.03 ± 0.01 (0.01 to 0.05) 0.83</td>
</tr>
<tr>
<td>Centromeric</td>
<td>0.11 ± 0.02 (0.07 to 0.15) 0.82</td>
<td>- 0.82 (N = 97) 75% 4%</td>
<td>0.03 ± 0.01 (0.01 to 0.05) 0.83</td>
</tr>
</tbody>
</table>

Note:

D: diffusion constant (mean ± SEM)

%: percentage of mobile molecules with long slow diffusion events (> 2.2 s).

N: numbers of molecules analyzed for diffusion constant and alpha factors calculations.

The numbers in the parenthesis represent confidence intervals at the 95% confidence level.
References


Figure 1

A

B

C

Tet7
Scrambled DNA

(ITAAGG)_n - (ITAAGG)_m

1.9 kb 1.6 kb 1.9 kb

Fraction of DNA molecules [%]

Position from DNA end (%)
Figure 2

A

B

C

D

E

Fast and slow diffusion

D_{int} (\mu m^2/s)

Frame number

Fraction of count (%)

Distance between slow diffusion (\mu m)
Figure 3

A

Centromere sequences

4.1 kb

1.7 kb

2.4 kb

1.7 kb

2.4 kb

1.7 kb

SA1 + centromere DNA

B

Fraction of SA1 molecules (%)

T270  Genomic  Centromeric

Static  Mobile

C

Diffusion constant (µm²/s)

Centromeric  Genomic  T270
Figure 6

A
Flag-SA1 + His-TRF1 + T270
110 s
Flag-SA1 + His-TRF1 + G-DNA
120 s

B
Fraction of molecules (%)

C
Diffusion Range (μm)

D
Diffusion constant (μm²/s)

SA1+TRF1 on T270
SA1+TRF1 on G-DNA
SA1 only on T270
Figure 7

A. TRF1 + T270

B. TRF1 + SA1 + T270

C."

D. Fraction of molecules (%) vs DNA-DNA pairing tract length (nm)
Figure 8

A

Specific DNA-binding bead

SA1 - TRF1 binding bead

B

Dimerization domain

Linker domain

DNA binding domain

Dimerization bead

Charge bead

C

Single link TRF1 only

Multiple links TRF1 only

Single link TRF1 + SA1

Multiple links TRF1 + SA1
Supplementary Figure S1. Purification of His- and Flag-tagged full length SA1 and characterization of the oligomeric state of SA1 in solution. (A) Coomassie blue stained SDS-PAGE gel of Flag-tagged WT SA1 and SA1 R37A R39A mutant (Genscript). (B) Evaluation of the oligomeric state of the full length His-SA1 in solution using a gel filtration column. Lanes 1-11, peak fractions from gel filtration.
**Supplementary Figure S2. SA1 does not specifically bind to consensus CTCF DNA binding sequences.** (A) The DNA substrate with consensus CTCF DNA binding sequences and control substrate with shuffled sequences used for EMSAs. The sequences from the top strands are shown. (B) EMSA of Flag-tagged WT SA1 in the presence of the Cy3-labeled DNA substrates containing consensus CTCF DNA binding sequences or shuffled DNA sequences. The dotted line indicates that the picture is combined from two regions in one gel. (C) Quantification of the percentage of DNA binding based on EMSAs. The error bars are standard deviations from two independent experiments.
A

<table>
<thead>
<tr>
<th>Type</th>
<th>Flag-SA1 (%; N = 107)</th>
<th>SA1+TRF1 dual-color (%; N = 101)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Type II</td>
<td>0</td>
<td>1.3 ± 1.0</td>
</tr>
<tr>
<td>Type III</td>
<td>21.0 ± 7.0</td>
<td>20.3 ± 10.4</td>
</tr>
<tr>
<td>Type IV</td>
<td>78.5 ± 7.0</td>
<td>78.4 ± 9.6</td>
</tr>
</tbody>
</table>

B

Supplementary Figure S3. Classification of different types of SA1-DNA interactions (A) and QD-labeled His-tagged SA1 on T270 DNA tightropes (B). (A) Cartoon drawings of different types of SA1-DNA interactions (left) and statistical analysis of these different types for Flag-SA1 and SA1-TRF1 complexes. Type I: a protein binds and then releases; Type II: a protein binds and doesn’t leave; Type III: a protein is bound at the beginning of the movie but releases; Type IV: a protein is bound from the beginning to end of the movie. Reliable attached lifetime measurements could only be obtained from analysis of the Type I interactions, which were not obtained for this study. (B) Top panel: conjugating His-tagged SA1 to SAv-QDs was achieved using the biotinylated multivalent chelator tris-nitrotriacetic acid (9 tris-NTA). Bottom panel: Dynamics of His-SA1-QDs on T270 DNA tightropes. Scale bars (y-axis: 1 μm).
Supplementary Figure S4. The full length Flag-SA1 alternates between fast and slow diffusion at lower (A), and higher (B) salt concentrations. Representative kymographs of Flag-SA1 on T270 DNA tightropes in buffers containing 50 mM KCl (A) or 150 mM KCl (B). Scale bars (y-axis): 1 μm. Equal molar concentrations of red (655 nm) and green (565 nm) Ab-QDs were used in the SA1 conjugation. All buffers containing additional 20 mM Tris (pH 7.5) and 0.1 mM MgCl₂.
Supplementary Figure S6. Comparison of the distributions of interval based diffusion constants ($D_{int}$) for SA1 (A-F) and TRF1 (G). (A-C) Distributions of $D_{int}$ for individual SA1 molecules on T270 DNA tightropes with static (A, kymograph in Fig. 2B), free diffusion (B, kymograph in Fig. 2C), and alternation between fast and slow 1-D diffusion (C, kymograph in Fig. 2D). The SA1 molecules have alternation between fast and slow diffusion show a distinct peak at $\sim 1.0 \times 10^{-5}$ $\mu m^2/s$. (D-F) Distributions of $D_{int}$ for all mobile SA1 molecules observed on T270 (D), genomic (E), and centromeric (F) DNA tightropes. On T270 DNA, SA1 molecules alternate between fast and slow diffusion with 46% of $D_{int}$ values less than $1 \times 10^{-2} \mu m^2/s$. In comparison, on genomic and centromeric DNA fast diffusion dominates with only 25% and 18% of $D_{int}$ values less than $1 \times 10^{-2} \mu m^2/s$, respectively. (G) The distribution of $D_{int}$ for all (N = 29) mobile TRF1 molecules on T270 reported in a previous study (Lin et al., 2014). On T270 DNA, the majority (84%) of the TRF1 $D_{int}$ values is consistent with slow diffusion (less than $1 \times 10^{-2} \mu m^2/s$).
Supplementary Figure S6. Analysis of the dwell time shows distinct DNA binding dynamics by full length WT SA1 on T270. (A) Dwell times of individual slow diffusion events on T270 (dwell time: 1.17 s, R²: 0.97, N = 437), the genomic substrate (dwell time: 0.80 s, R²: 0.99, N = 437), and the DNA substrate containing centromeric sequences (dwell time: 0.79 s, R²: 0.97, N = 438). The data sets were fitted with a single exponential decay function. (B) SA1 spends a higher percentage of time in the slow diffusion mode out of the total dwell time on T270 DNA (24.4%, N = 18), compared to on genomic (3.2%, N = 8) and centromeric (6.3%, N = 21) DNA substrates. Error bars: SEM.
Supplementary Figure S7. The SA1 N-terminal domain (SA1-N) displays alternation between fast and slow diffusion on T270 DNA tigrotes. (A) The position distribution of SA1-N on T270 DNA based on the analysis of AFM images of SA1-N-T270 complexes. Similar to the full length SA1 (Fig. 1), SA1-N preferentially binds to the telomeric region (35 – 50%) on T270 (N = 36). (B) On T270 tigrotes, SA1-N displays a higher percentage of static (47.1%) complexes, compared to on genomic DNA (6.2%). (C) On T270 tigrotes, mobile SA1-N molecules display a higher percentage (74.3%, N = 88) of complexes with slow diffusion events ($D_{in} < 5.0 \times 10^{-3} \mu m^2/s$, dwell time > 2.2 s), compared to on genomic DNA (38.8%, N = 103).
Supplementary Figure S8. Validation of the dual-color-labeling strategy for SA1-TRF1 complexes. (A) Formation of dual-color QD-labeled complexes on T270 tightropes depends on the presence of both TRF1 and SA1. In all experiments, βIIHS-NTA, red Ab-QDs, and green SAv-QDs were present in flow cells. The percentages of green, red, and dual-color labeled QD complexes observed in the presence of SA1 only (white bars, N = 190), TRF1 only (striped bars, N = 358), or both TRF1 and SA1 (N = 235) are shown. Error bars: SEM based on at least three independent experiments. The percentage of dual-color QD-labeled complexes on T270 (11.2%) in the presence of both TRF1 and SA1 is 5.6- and 10-fold higher than when only SA1 (2.0%) or TRF1 (1.1%) was present, respectively. (B) Dual-color QD-labeled SA1-TRF1 complexes were present at telomeric regions on T270. The distribution of pair-wise distance between green QD-labeled TRF1 (markers for telomeric regions) and dual-color QD-labeled SA1-TRF1 complexes is consistent with the spacing between telomeric regions on T270. The data set (N = 30) was fitted with a double Gaussian function with peaks centered at 1.6 ± 0.08 μm (distance between two nearest neighbor (TTAGGG)$_{270}$ regions) and 5.1 ± 0.38 μm. The insert shows a representative kymograph (120 s) used for the distance analysis.
Supplementary Figure S9. Comparison of TRF1 and SA1-TRF1 DNA binding. Protein mediated DNA-DNA pairing tract lengths on T270 DNA in the presence of only TRF1 (A) or both SA1 and TRF1 (B). (C) Comparison of the AFM height of TRF1 and SA1 only on T270 DNA. The AFM height of TRF1 and SA1 on T270 DNA is 0.73 (± 0.1, mean ± SD, N = 61) nm, and 1.39 (± 0.5, mean ± SD, N = 116) nm, respectively.
Supplementary Figure S10. Overview of the MD simulations of SA1 (A) and SA1-TRF1 (B) DNA binding. Regions representing genomic (600 bp) and telomeric DNA (1200 bp) DNA are shown in purple and dark blue, respectively. The SA1 and TRF1 models are shown in Fig. 7.
Supplementary Figure S11. Monte Carlo simulations of two-state DNA binding by SA1 on the DNA substrate with alternating telomeric and non-telomeric regions. (A) SA1 is hypothesized to exist in one of two DNA binding states: mobile or static. The static period is exponentially distributed with a duration of ten steps. The DNA substrate used in the simulations mimics the ligated T270 DNA (1.6 kb telomeric sequences spaced 3.8 kb apart by non-telomeric regions, Fig. 2A). (B) The positional trace of a simulated SA1 molecule on DNA mimicking ligated T270 DNA tightrope with alternating telomeric and non-telomeric regions. (C) An artificial kymograph created by converting each positional step in (B) to a Gaussian distribution. The horizontal grid lines correspond to the telomeric regions. The stepping rate is ~173000 s⁻¹ for a diffusion constant of 1x10⁻⁵ µm² s⁻¹. Since the equilibrium is rapid and dynamic even on telomeric DNA, SA1 continues to diffuse over telomeric regions, however much more slowly.
Appendix C: TRF1 and TRF2 use different mechanisms to find telomeric DNA but share a novel mechanism to search for protein partners at telomeres
TRF1 and TRF2 use different mechanisms to find telomeric DNA but share a novel mechanism to search for protein partners at telomeres

Jiangguo Lin\(^1\), Preston Countryman\(^2\), Noah Buncher\(^2\), Parminder Kaur\(^1\), Longjiang E\(^3\), Yiyun Zhang\(^4\), Greg Gibson\(^5\), Changjiang You\(^6\), Simon C. Watkins\(^7\), Jacob Piehler\(^8\), Patricia L. Opresko\(^9\), Neil M. Kad\(^10\), Hong Wang\(^11\)

\(^1\)Physics Department, \(^2\)Industry and System Engineering, North Carolina State University, Raleigh, North Carolina, NC 27695, USA

\(^3\)Department of Environmental and Occupational Health, \(^4\)Center for Biologic Imaging, University of Pittsburgh Graduate School of Public Health, Pittsburgh, Pennsylvania 15219, USA

\(^5\)Electrical and Computer Engineering, University of North Carolina at Charlotte, Charlotte, North Carolina 28223, USA

\(^6\)Division of Biophysics, Universität Osnabrück, Barbarstrasse 11, 49076, Osnabrück, Germany

\(^7\)Department of Biology, University of Essex, Colchester, Essex CO4 3SQ UK

To whom correspondence should be addressed. Tel: +1 919 5137203; Fax: +1 919 5156538; Email: hong_wang@ncsu.edu

Correspondence may also be addressed to Neil M. Kad. Tel: +44 1206 873010; Fax: +44 1206 872592; Email: nkad@essex.ac.uk

Keywords: single-molecule, telomeres, TRF1, TRF2, fluorescence imaging, atomic force microscopy, protein-DNA interactions

Running title: Distinct Dynamics of TRF1 and TRF2 on Telomeric DNA
ABSTRACT

Human telomeres are maintained by the shelterin protein complex in which TRF1 and TRF2 bind directly to duplex telomeric DNA. How these proteins find telomeric sequences among a genome of billions of base pairs and how they find protein partners to form the shelterin complex remains uncertain. Using single-molecule fluorescence imaging of quantum dot labeled TRF1 and TRF2, we study how these proteins locate TTAGGG repeats on DNA tightropes. By virtue of its basic domain TRF2 performs an extensive 1-D search on non-telomeric DNA, whereas TRF1’s 1-D search is limited. Unlike the stable and static associations observed for other proteins at specific binding sites, TRF proteins possess reduced binding stability marked by transient binding (~9–17 seconds) and slow 1-D diffusion on specific telomeric regions. These slow diffusion constants yield activation energy barriers to sliding ~2.8–3.6 K_bT greater than those for non-telomeric DNA. We propose that the TRF proteins use 1-D sliding to find protein partners and assemble the shelterin complex, which in turn stabilizes the interaction with specific telomeric DNA. This ‘tag-team proofreading’ represents a more general mechanism to ensure a specific set of proteins interact with each other on long repetitive specific DNA sequences without requiring external energy sources.
INTRODUCTION

Telomeres play a crucial role in maintaining the stability of linear chromosomes (1,2). Loss of telomere function can activate DNA repair processes, leading to nucleolytic degradation of natural chromosome ends and their end-to-end fusion (3). Telomere dysfunction and associated chromosomal abnormalities have been strongly associated with age-related degenerative diseases and cancer (4,5). In a typical human somatic cell, the telomeric repeat sequence TTAGGG is ~2-15 kb in length with a 3’ overhang of ~100-200 nt (6). This 3’-overhang serves as a substrate for the reverse transcriptase telomerase, which replicates the telomeric sequence by using an internal RNA subunit as a template to direct the DNA synthesis (1,7-9). A specialized protein complex, shelterin (or telosome) binds to and protects the chromosome ends (2,10). The shelterin complex in humans consists of six core proteins: TRF1, TRF2, POT1, Tin2, TPP1, and RAP1 (11).

TRF1 and TRF2 are the only proteins in the shelterin complex that make high affinity contact with double-stranded telomeric DNA (12,13). TRF1 negatively regulates telomere length and promotes telomere replication (14). Whereas, TRF2 caps and protects chromosome ends (11), in addition to regulating telomere length (15). Removal of TRF2 from the telomeres results in loss of the 3’ overhang, covalent fusion of telomeres, and induction of ATM and p53 dependent apoptosis (16,17). Both TRF1 and TRF2 contain a TRFH domain that mediates homodimerization and a Myb type domain that sequence-specifically binds to telomeric DNA (Figure 1A) (12). However, these two proteins differ at their N-termini, where TRF1 and TRF2 are rich in acidic and basic residues, respectively. Previous electron microscopy (EM) and atomic force microscopy (AFM) studies established that both TRF1 and TRF2 play important architectural roles at telomeres (18-21). TRF1 forms protein filaments on longer telomeric repeats (>27 repeats) and promotes parallel pairing of telomeric tracts (19). In vitro, TRF2 can remodel linear telomeric DNA into T-loops (20).

A previous cell based study of TRF1 and TRF2 using fluorescence recovery after photobleaching and fluorescence loss in photobleaching suggested that TRF1 and TRF2 interact with telomeres in a dynamic fashion (22). Although TRF1 and TRF2 are proposed to have extra-telomeric functions, they preferentially localize to the TTAGGG repeat sequences whether these target sites are at interstitial regions or at chromosome ends (23-25). Once telomeric sequences are located, TRF1 and TRF2 must find protein partners to form the shelterin complex and to regulate the functions of other DNA binding proteins at telomeres (26-28). Despite recent advancements in the understanding of functions of TRF1 and TRF2, it is still
unclear how TRF1 and TRF2 are able to find telomeric sequences and protein partners in a genome of billions of base pairs.

Accumulating evidence suggests that a protein can use one-dimensional (1-D) sliding (correlated translocation while maintaining continuous DNA contact), jumping (non-correlated detachment and reattachment), or hopping (correlated detachment and reattachment) to navigate through the vast excess of nonspecific DNA sequences in vivo (29-32). Investigations of DNA binding dynamics on nonspecific DNA at the single-molecule level have significantly advanced our understanding of how proteins with diverse functions conduct their target DNA search (31,33,34). However, the paradoxical requirements of rapid search at nonspecific sites and stability at target sites have been primarily investigated in theoretical studies (35-38), direct comparisons of the protein binding energy landscape at nonspecific sites and target sites from single-molecule experimental data are still lacking.

Here we used single-molecule fluorescence imaging to study the dynamics of quantum dot (QD) labeled TRF1 and TRF2 proteins on λ DNA and DNA substrates containing alternating regions of telomeric and non-telomeric sequences. TRF1 appears to bind directly to telomeric sequences with very little 1-D searching through non-telomeric DNA, whereas TRF2 possesses a significant component of 1-D search. Using a truncation mutant, we localized this 1-D searching activity to the basic domain of TRF2. On telomeric DNA both TRF1 and TRF2 diffuse slowly due to higher energy barriers to diffusion; and they possess longer attached lifetimes at telomeric repeats compared to non-telomeric DNA sequences. These observations indicate that there is preferential binding to telomeric DNA but the affinity is not high enough to prevent TRF proteins from diffusing along TTAGGG repeats. We postulate that this allows TRF1 and TRF2 to find their protein partners locally, and that this is a more general mechanism for coupling the energy from multiple weak DNA binding components to ensure high binding specificity on long repetitive sequences.

MATERIALS AND METHODS

Protein purification

Recombinant N-terminal His6-tagged TRF1 and TRF2 were purified using a baculovirus/insect cell expression system and an AKTA Explorer FPLC (GE Healthcare) as described previously (39). TRF2ΔB was purified using a bacterial expression system (40). Protein concentrations were determined using the Bradford assay. Proteins used in this study are more than 90% pure based on SDS-PAGE and Coomassie staining. Proteins are active in binding to the telomeric
DNA substrate containing 3 TTAGGG repeats based on electrophoresis mobility shift assays (EMSAs).

**DNA substrates**

λ DNA was purchased from New England BioLabs. Other DNA substrates used in this study are shown in Figures 1B and S1. pSXneo(T2AG3) plasmid DNA containing 270 TTAGGG repeats was a gift from Dr. Peter Lansdorp (University of British Columbia) (41); pGTK4 plasmid derived Tel10 plasmid is 5994 bp long and contains 10 TTAGGG repeats and was prepared as described previously (42). To generate DNA fragments containing TTAGGG repeats for AFM imaging, digestion of T270 DNA (10 μg) was carried out at 37 °C for 4 hrs using HpaII (130 U) in Buffer 4 (New England BioLabs). For Tel10 plasmid, digestions were carried out using XbaI (100 U) in Buffer 4. For fluorescence imaging, linearized plasmids were ligated to generate longer DNA substrates using a Quick Ligation™ Kit (New England BioLabs). The ligation reactions were done at room temperature for 15 mins. The non-telomeric DNA substrate without the (TTAGGG)270 sequence was gel purified after the digestion of pSXneo(T2AG3) with BglII and XbaI. Final DNA substrate purification was done using an Illustra GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare).

**Protein-QD conjugation**

Streptavidin-conjugated quantum dots (Sav-QDs) were purchased from Invitrogen. Biotinylated multivalent chelator tris-nitrilotriacetic acid (6^tris-NTA) was prepared according to the previous reports (43,44). The TRF-DNA reaction buffer contains 50 mM Hepes (pH 7.5) and varying concentrations of NaCl (25, 50, 75 and 100 mM). The total ionic strengths are 75, 125, 175, and 225 mM, respectively (45).

For single color QD labeling of His6-tagged TRF1 or TRF2, 1 μL of red QD (655 nm, 1 μM, Invitrogen, hydrodynamic radius: 11.5 nm) was incubated with 1 μL of 6^tris-NTA (2 μM) for 20 mins (46). 1 μL of proteins (2 μM) were then added to the QD-NTA solution and incubated for an additional 20 mins. For dual-color QD labeling, 1 μL of red (1 μM) and green QDs (565 nm, 1 μM, hydrodynamic radius: 9.5 nm) were incubated with 2 μL of 6^tris-NTA (2 μM) (46). TRF1 or TRF2 (2 μL, 2 μM) was added to the solution and incubated for additional 20 mins. For fluorescence imaging, unless otherwise specified, protein-NTA-QD solutions were diluted 200-fold before being drawn into the flow cell using a syringe pump (model SP260p, World Precision Instruments) at 300 μl/ml flow rate. The final protein concentration was 3.3 nM for both TRF1 and TRF2. Protein concentrations and ionic strengths of the buffer used in this study are
comparable to physiological conditions (Supplementary Text). For AFM imaging of TRF2-QDs in the presence of monoclonal TRF2 antibody (ImageneX Corporation), the Ab:TRF2:NTA:QD ratio was 1:1:2:1 or 5:1:2:1, and reactions were carried out at room temperature for 30 mins after the addition of antibodies.

AFM imaging and image analysis

All DNA and protein samples were diluted 10-fold in 1X AFM buffer [25 mM NaOAc, 25mM HEPES–KOH (pH 7.5) and 10 mM Mg(OAc)$_2$] before deposition onto a freshly cleaved mica (SPI Supply). The samples were then washed with MilliQ water and dried under a stream of nitrogen gas. All images were collected in tapping mode using a MFP-3D-Bio AFM (Asylum Research). Pointprobe$^\text{®}$ PPP-FMR probes (Nanosensors) with spring constants at $\sim$2.8 N/m (nominal value) were used. All images were captured at a scan size of $1 \text{ m} \times 1 \text{ m}$, a scan rate of 1–2 Hz, and a resolution of $512 \times 512$ pixels. The position of TRF proteins on DNA was analyzed using the software from Asylum Research.

Fluorescence imaging and analysis of fluorescence microscopy data

Fluorescence imaging was carried out with an inverted microscope (Nikon Ti-E) equipped with an encoded motorized stage, perfect focus system (PFS), and a Ti-TIRF E motorized illuminator unit. Fluorescence imaging was performed by excitation at 488 nm using a solid state laser (20 mW Sapphire DPSS), a 100X objective with a numerical aperture of 1.49 (APO TIRF, Nikon) and 1.5X additional magnification. The laser power was controlled by using neutral density filters. The excitation beam was reflected into the objective through a TIRF filter set containing zt488dc and ET500LP filters. For simultaneous imaging of green (565 nm) and red (655 nm) QDs, a dual view simultaneous imaging system (DV2, Photometrics) was used in combination with a T605LPXR dichroic beamsplitter (Chroma) and a bandpass filter ET655/40m (Chroma). The images were captured using an electron multiplied (EM) CCD camera (Ixon DU897, Andor Technology) operated at -60°C, with an EM gain of $\sim$250 and a frame rate of 20 Hz. Construction of the flow cell was carried out according to a procedure described previously (33,47,48). Silica beads (5 μm, Polysciences) were first treated with poly-L-lysine hydrobromide (2500 μg/mL, M.W. > 300 KDa, Wako Chemicals). λ DNA or ligated DNA substrate (5 μg/mL) were stretched, unless otherwise specified, under hydrodynamic flow at 300 μl/min flow rate using a syringe pump. Extended DNA strands anchored between two poly-L-lysine-coated beads formed DNA tightropes. After introducing the protein-QDs into the flow cell, all data collection was performed in the absence of any further buffer flow. The presence of YOYO-1 on
DNA significantly reduced the diffusion constant, alpha factor and the percentage of motile protein-QD complexes on DNA at certain salt conditions. Consequently, all data analysis was done using movies collected from using unstained DNA tightropes (Supplementary Text).

**Statistical analysis**

Single-factor ANOVA and Student-t Tests were used for statistical analysis.

**RESULTS**

**TRF1- and TRF2-quantum dot (QD) conjugates are functional in DNA binding**

Fluorescent labeling of TRF1 and TRF2 was achieved by conjugating 6x histidine (His_{6}) tagged TRF1 and TRF2 to streptavidin-conjugated quantum dots using the biotinylated multivalent chelator tris-nitriotriacetic acid (^{6}Tnis-NTA) (44) (Figure 1B, Materials and Methods). The multiple Ni-NTAs on the circular scaffold of the tris-NTA adaptor bind the His-tag with subnanomolar affinity, resulting in a bound lifetime in the range of hours (43,44). Importantly, we applied a previously established method based on AFM imaging to characterize the stoichiometry of QD-TRF complexes (49,50). AFM imaging revealed that using TRF2 antibody marking the presence of TRF2 (TRF2:Ab = 1:1 or 1:5), among the QDs displayed TRF2-Ab complexes (24%), 90% (n = 39) possessed only one TRF2-Ab complex (Figure S2).

QDs alone exhibited minimal nonspecific binding to DNA as confirmed by AFM (Figure 1C). As expected, addition of QD labeled TRF1 or TRF2 to DNA containing two stretches of (TTAGGG)_{35}, connected by a short linker region (T270 DNA, Figure 1B, Materials and Methods) resulted in substantial binding (Figure 1D,E). Further AFM image analysis revealed that both TRF1- and TRF2-QDs bound preferentially to the telomeric DNA sequences on both the T270 and Tel10 DNA substrates (Supplementary Figure S3).

**TRF1 and TRF2 diffuse one-dimensionally on non-telomeric DNA**

To study the dynamics of individual TRF1 and TRF2 molecules on DNA using oblique-angle fluorescence microscopy, we applied a DNA tightrope assay (Figure 2A) (33). DNA strands are suspended between poly-L-lysine coated microspheres at an elongation of ~90% DNA contour length using hydrodynamic flow (47). This process isolates DNA from the surface and does not require continuous buffer flow for the observation of protein-DNA interactions. QDs did not bind to DNA tightropes alone or in the presence of TRF proteins without ^{6}Tnis-NTA. However, with both ^{6}Tnis-NTA and His_{6}-tagged TRF1 or TRF2, QDs were observed on DNA throughout the
visual field (Figure 2B,C). Both TRF1- (Supplementary Movie S1) and TRF2-QDs (Supplementary Movie S2) showed clear 1-D diffusion on DNA, which was tracked by Gaussian fitting to kymographs (particle position versus time plots, Supplementary Information) (33,47).

To determine whether TRF1 and TRF2 slide or hop, we evaluated the effect of ionic conditions on the dynamic interactions between the QD labeled TRF proteins and DNA. Increasing the salt concentration should not affect the diffusion constants of a sliding process, but should elevate the diffusion constants of hopping (29,51,52). We performed experiments at 75, 125, 175, and 225 mM ionic strengths (Materials and Methods). The fraction of motile TRF1 proteins ranged from 15% to 33% (Supplementary Figure S4A) and followed a trend of decreasing diffusion constants as the ionic strength increased (7.5 to 3.8 × 10⁻² μm²/s), such that the difference between the highest and lowest salt was statistically significant (p = 0.017; Supplementary Figure S4B and Table 1). In contrast, TRF2 was highly motile on λ DNA across all ionic strengths and showed no significant change in diffusion constant (8.4 to 9.5 × 10⁻² μm²/s, Supplementary Figure S4B and Table 1). TRF2 diffused substantially faster than TRF1 at all ionic strengths showing statistical significance at ionic strengths between 125 and 225 mM.

In addition to the diffusion constant, we also measured the diffusive exponent (α factor, Supplementary Information). An α factor of 1 indicates an unbiased random walk, larger than 1 indicates directed motion and less than 1 indicates periods of pausing in the random walk (sub-diffusion) (53). TRF1 showed a slight trend toward increasing α factor from 0.65 to 0.89 with increasing ionic strength (Supplementary Figure S4C and Table 1); this result suggests pausing at low ionic strength, which is abrogated by salt. For TRF2, however, the α factor was consistently close to 1 and did not show any significant variation with ionic strength, suggesting an unbiased random walk. Dual color labeling of the TRF proteins allowed us to assess whether protein hopping could enable bypass of other DNA-bound proteins that act as diffusion barriers (Supplementary Figure S5). Neither TRF1 nor TRF2 could bypass differentially labeled proteins of the same species on DNA, which is consistent with a TRF2 sliding mechanism and suggests that TRF1 also navigates DNA by sliding (Supplementary Information).

Next, we measured the attached lifetimes of protein-QD complexes on DNA. Firstly however, we classified the protein-DNA interactions into four types based on how they behaved during a movie. Type I: protein binds and then releases; Type II: proteins binds and doesn’t leave; Type III: protein is bound at the beginning of the movie but releases; Type IV: protein is bound from the beginning to end of the movie (Figure 2D and Supplementary Table S1). Reliable attached lifetime measurements could only be obtained from analysis of the Type I interactions. The lifetimes of both TRF1 and TRF2 on λ DNA decreased with increasing ionic
strength, ranging from 1.8 s (175 mM) to 0.3 s (225 mM) for TRF1 and from 10 s (75 mM) to 3.4 s (225 mM) for TRF2 (Table 1 and Figure S4D). These results are consistent with salt-sensitive electrostatic interactions between TRF proteins and DNA and increased probability of dissociation from DNA during sliding as the ionic strength increases (54).

In summary, these results demonstrate that both TRF1 and TRF2 slide on DNA in search of their target DNA binding sites. TRF2 is a canonical slider, whereas TRF1 also appears to slide but may alter its conformation with salt.

**TRF1 and TRF2 bind specifically to telomeric sequences on DNA tightropes**

To examine the dynamics of TRF1 and TRF2 binding to telomeric DNA sequences, we ligated linearized T270 DNA to generate long DNA substrates with alternating (TTAGGG)$_{220}$ telomeric and non-telomeric regions (Figure 3A). The lengths of these DNA tightropes ranged from ~2.1 to ~22 µm, consistent with ligation of 2 to 12 of 5.4 kb T270 DNA fragments (Supplementary Figure S6A). TRF1 and TRF2 bound to the ligated T270 DNA tightropes with regular spacing (Figure 3B, and Supplementary Movies S3 and S4). For both TRF1 and TRF2, the distributions of the distances between adjacent binders fit well to the sum of two Gaussian distribution functions centered at ~1.6 and 3.2 µm (Figure 3C). These findings are consistent with the expected spacing of the telomeric regions (Figure 3B). In contrast, on the ligated non-telomeric DNA, the distribution of TRF2 spacing was broad (Figure 3C), and no examples of three or more bound protein-QDs on individual DNA tightropes with a spacing of ~1.6 or 3.2 µm were observed for either TRF1 or TRF2. As an additional control the telomeric repeats were spaced further apart using a 5.99 kb long DNA substrate containing only 10 TTAGGG repeats (Tel10, Supplementary Figure S7, and Supplementary Movie S5) and, as expected, adjacent bound TRF2 molecules were further apart (1.9 µm, ~95% contour length) than on T270.

We also examined how far single molecules of TRF1 and TRF2 could slide on the ligated non-telomeric DNA versus ligated T270 DNA (Figure 4, Supplementary Movies S3 and S4). On T270 DNA, TRF1 displayed one major population with diffusion ranges centered on 0.38 µm (Figure 4D). TRF2 exhibited two distinct populations centered on 0.5 and 1.2 µm at 125 mM ionic strength, and on 0.5 and 1.5 µm at 225 mM ionic strength (Figure 4 and Supplementary Figure S6B). But on non-telomeric DNA, no clear peak was evident (white bars, Figure 4D). Approximately 90% (n = 29) of TRF1 and 73% (n = 30) of TRF2 diffused in a short range (<850 nm). The diffusion range was invariant across all time windows (~10~100 s, Supplementary Figure S8), ruling out the possibility that the short range diffusion observed was due to shorter video lengths. Instead, this finding suggests that once the molecules are within a
telomeric region, they tend to remain there. We explored the possibility that short range diffusion was caused by multiple proteins binding to the same telomeric region and restricting 1-D sliding. However, at a lower TRF2 concentration, the short diffusion range did not change (compare Figure 4D and Supplementary Figure S6C). Therefore, the two diffusion range populations could be assigned to diffusion of TRF proteins over the (TTAGGG)_{20} telomeric regions (0.5 μm, ~90% contour length) and the non-telomeric spacers (1.2 μm, ~90% contour length), respectively (Figure 3B). For TRF2, transitions were observed between telomeric and non-telomeric regions or even between two adjacent T270 repeats, which were more frequent at 225 mM ionic strength (white arrows, Figure 4C). These events provided the peak with diffusion range centered at ~1.5 μm (Supplementary Figure S6B).

Taken together, the regular spacing between QD labeled TRFs demonstrated that TRF1 and TRF2 bind specifically to the telomeric regions on both T270 and Tel10 DNA substrates. These results also showed that compared to TRF2, TRF1 undergoes a greater number of direct binding events from solution to the (TTAGGG)_{20} region, forgoing a 1-D search (Figure 4D).

**TRF1 and TRF2 exhibit slower dynamics on telomeric DNA**

To quantify the diffusion constants at the (TTAGGG)_{20} telomeric region, we selectively analyzed TRF1 and TRF2 on the ligated T270 DNA tightropes with at least 3 or more protein-QDs in a row spaced at the length of non-telomeric spacers (1.5 to 1.7 μm, Figures 3 and 4). TRF1 and TRF2 diffused at ~0.15–0.22 × 10^{-2} μm²/s and 0.27–0.29 × 10^{-2} μm²/s at the (TTAGGG)_{20} region, respectively. These rates are ~17–37 and ~30 fold slower, for TRF1 and TRF2, respectively, compared with those on λ DNA at the same ionic strength (Tables 1 and 2). We noted that in many cases TRF proteins binding to telomere repeats would be confined to diffuse within this region due to the higher affinity for telomeric sequences (Figure 4 and Supplementary Figure S8). To ensure that this confinement would not artificially reduce the apparent diffusion constant, we simulated 1-D diffusion of proteins on a linear DNA lattice of unlimited length versus a 1.6 kb total length, which mimics the (TTAGGG)_{20} region (Supplementary Text). These simulations revealed that confinement within 1.6 kb DNA does not significantly reduce the observed diffusion constant at the (TTAGGG)_{20} region (Supplementary Figure S9). In addition, camera based time-averaging was not a major contributor to the observed slower diffusion constants at the telomeric region under these experimental conditions (Supplementary Text). An alternative fitting method to simultaneously determine the diffusion constant and confined DNA length also provided similar results (Supplementary Information) (55). Furthermore, the diffusion constants of TRF2-QDs (0.31 ± 0.003 × 10^{-2} μm²/s, n = 37) on
DNA tightropes formed under a 12X slower flow rate (25 μl/min) are not significantly different from those on DNA tightropes stretched at a higher flow rate (300 μl/min) (Table 2). Under this condition, DNA tightropes were under less tension with final extension to only ~88% of DNA contour length (Figure S6D). These results suggest that under these conditions, diffusion constants of TRF2 do not vary significantly with the amount of tension on dsDNA tightropes.

We observed that TRF1 and TRF2 can directly dissociate from telomeric regions or through non-telomeric regions (Supplementary Figure S10). Overall, we found that the relative proportions of Type I (protein binds and releases) and Type IV (protein is bound from the beginning to end of the movie) protein-DNA interactions observed during the experimental time course depended on the DNA substrate (Supplementary Table S1). For TRF1 on T270 DNA, the vast majority of molecules were Type IV, indicating a considerably longer attachment. Consistent with this result, the average lifetime of Type I TRF1 bound to the telomeric sequences on T270 DNA was ~31-fold longer than that for λ DNA (9.2 s vs. 0.3 s, 225 mM ionic strength, Tables 1 and 2). TRF2 behaved quite differently, showing a less pronounced difference between the proportions of Type I and Type IV complexes on λ DNA and T270 DNA. Furthermore, the attached lifetimes for Type I TRF2 complexes was only ~3-fold longer at the telomeric regions on T270 DNA compared with λ DNA (10.3 s vs 3.4 s; Tables 1 and 2). It is worth noting that the lifetimes of TRF proteins on DNA are longer than the QD blinking rate (56), ruling out artifacts from QD blinking in the lifetime measurement. In summary, compared with binding to non-telomeric DNA, both TRF1 and TRF2 possess distinctly slower detachment and diffusional dynamics on the telomeric DNA.

The basic domain is essential for the 1-D search by TRF2

The basic domain at the N-terminus of TRF2 permits its binding to model replication forks and four-way junctions independent of telomere sequences (57). In addition, the absence of this domain leads to a diminished ability of TRF2 to localize to model telomere ends and to facilitate T-loop formation (57). We created and imaged a basic domain deletion mutant of TRF2 (TRF2ΔB) on λ DNA and the ligated T270 (Supplementary Figure S11). Compared to full length TRF2, TRF2ΔB-QDs have higher specificity for the telomeric sequences on T270 DNA substrate and lower affinity to DNA ends (compare Figure S11A and Figure S3B). Furthermore, relative to the full length TRF2, the fraction of motile protein-DNA complexes decreased by approximately 1.5-fold for TRF2ΔB (Supplementary Figure S11 legend). Interestingly, the diffusion constant (9.1 ± 1.8 ×10^{-2} μm²/s) and α factor (0.93 ± 0.04) of TRF2ΔB on λ DNA were not significantly different from those of full length TRF2 (Table 1). However, the percentage of
complexes undergoing long-range diffusion (10% at 125 mM ionic strength) was significantly lower (p = 0.01) than for full length protein (27%) at the same ionic strength (Supplementary Figure S11D). On T270 DNA, majority of motile TRF2ΔB (90%) was found with a diffusing range consistent with length of the telomeric region on T270 DNA, suggesting that TRF2ΔB directly associates with telomeric DNA from solution and not by diffusion from a non-telomeric region. Since the frequency of TRF2ΔB DNA binding was lower than the full length protein (1.1 vs. 3.8 molecules/bead pair), it was not possible to restrict the analysis to those tightropes with three adjacent bound molecules. Therefore, we treated all short range diffusion (<850 nm) by TRF2ΔB on the ligated T270 as diffusion over the telomeric region. The dynamics of TRF2ΔB over the (TTAGGG)4 region were similar to those of full length TRF2, with a similar diffusion range (0.47 ± 0.03 μm, Supplementary Figure S11D) and diffusion constant (0.27 ± 0.01 × 10⁻² μm²/s at 125 mM and 0.26 ± 0.01 × 10⁻² μm²/s at 225 mM). These observations suggest that the basic domain of TRF2 normally facilitates its 1-D search on non-telomeric DNA. The reduced degree of TRF2 localization to the telomeric region due to deletion of the basic domain demonstrates the importance of 1-D diffusion in the TRF2 telomeric target site search (Supplementary Figure S11).

DISCUSSION
TRF1 and TRF2 are the only scaffolding shelterin proteins that bind directly to duplex telomeric DNA. The results presented here from single-molecule imaging of TRF1 and TRF2 dynamics on telomeric and non-telomeric DNA provide for the first time a fundamental understanding of the mechanisms that drive the dynamics of shelterin assembly/disassembly at telomeres.

TRF2 performs 1-D searching more effectively than TRF1 to find telomeric sequences
Rotational tracking along DNA during which a protein follows a helical track along the DNA to maintain optimal contact has been inferred for several DNA binding proteins (58). The measured diffusion constants for TRF1 and TRF2 obtained using the DNA tightrope assay were consistent with rotational tracking of the DNA helix (Table 1 and Supplementary Text), although slightly higher than the predicted upper limit for this motion (2.1 × 10⁻² μm²/s, Supplementary Text). This discrepancy could be due to the flexible linkage mediated by the His-tag and 86TNTA between TRF proteins and QDs (59). The measured diffusion constants together with the lack of observed barrier bypass events in dual color experiments (Supplementary Figure S5) demonstrated that both TRF1 and TRF2 track the DNA helix to maintain optimum contact between their DNA binding surfaces and the DNA (Figure 5A). However, the attached lifetime of
Type I TRF1 at the non-telomeric region was 10-fold shorter than that of TRF2 (0.3 s vs. 3.4 s at 225 mM, Table 1). These results are consistent with a significantly lower percent of TRF1 molecules exhibiting long range diffusion compared with TRF2 (Figure 4). This difference between TRF1 and TRF2 is partly due to the sequences at the N-termini of TRF proteins (Figure 1A). For TRF2 this region contains a basic domain, the deletion of which (TRF2ΔB) led to a clear reduction in the percentage of motile protein complexes on λ DNA. Importantly, it was observed that 90% of TRF2ΔB molecules underwent short-range diffusion consistent with the length of the telomeric regions (Figure S11). This result suggests that the majority of the TRF2ΔB molecules found the telomeric region directly from solution, forgoing the 1-D component of the search (Figure 5A). These results support the notion that domain B facilitates the association of TRF2 to non-specific DNA and this results in sliding subsequently. However, the diffusion constant and α factor of TRF2ΔB were not significantly different from the full length protein (Table 1). We speculate that TRF2ΔB containing the Myb type domain has weak DNA binding affinity for non-telomeric DNA. On non-telomeric λ DNA, the DNA binding energy landscapes are similar for full length TRF2 and TRF2ΔB, leading to similar diffusion constants. However, it is unclear whether in full length TRF2, nonspecific DNA binding is solely dependent of domain B or combination of this domain and the Myb domain. TRF1 behaved similarly to TRF2ΔB, perhaps as a consequence of also lacking the basic domain. Therefore, unlike TRF1, TRF2 can bind to non-telomeric sequences and use a 1-D search to more efficiently locate telomeric DNA.

Comparing the 1-D diffusion of TRF1 and TRF2 on non-telomeric and telomeric DNA

We found that, in general, TRF2 slides faster than TRF1 at non-telomeric sequences (Supplementary Figure S4B and Table 1). The diffusive exponent was less than 1 only for TRF1 at lower ionic strengths, consistent with sub-diffusive motion or pausing during diffusion (Supplementary Figure S4C and Table 1). Together, these observations indicate that TRF2’s diffusion is consistent with the canonical description of sliding. However, TRF1’s behavior changed with salt in a manner that was inconsistent with a solely electrostatic mediated protein-DNA interaction (60), and suggesting a possible conformational rearrangement induced by salt at the DNA binding interface. This rearrangement could lead to obstacles to diffusion and/or traps within the binding energy landscape or escape time (53).

The Myb type DNA binding domain of TRF2 has a 4-fold weaker DNA binding affinity than the Myb domain in TRF1 (equilibrium dissociation constants Kd: 750 vs. 200 nM, respectively) (61). The diffusion constant of TRF1 was ~2-fold slower than that of TRF2 within telomeric repeats.
(125 mM ionic strength, Table 2). This result is equivalent to ~0.6 $K_BT$ increase in the roughness of the DNA binding landscape or ~2-fold change in affinity. While these results are consistent with the stronger binding to the telomeric sequences by TRF1 Myb domain, other domains on TRF proteins could also indirectly influence the DNA binding dynamics of these two proteins over the telomeric regions. Furthermore, the difference in the dynamics of the TRF proteins between telomeric DNA and non-telomeric DNA is due to inherent sequence effects and therefore likely represents the situation in vivo. This is further supported by the ionic conditions used in our experiments which were chosen to represent those encountered in vivo (Supplementary Text).

**TRF1 and TRF2 strike a balance between search and specificity**

TRF proteins face a unique challenge. They must find both their cognate sites and protein partners to form the shelterin complex, and to regulate the functions of a myriad of proteins involved in telomere maintenance and cell-cycle progression (26). For example, TRF1 and TRF2 both bind to TIN2 to form a ternary complex of TRF1, TRF2, and TIN2 (27,28). Importantly, TRF2 is a protein hub interacting with several DNA binding proteins that play important roles in DNA repair, including WRN, Ku70-Ku80, and ERCC1-XPF (26,39,62,63). This requires that TRF proteins retain specificity for their DNA target site but also the ability to slide within the telomeric regions to encounter protein partners to form protein complexes.

The binding energy of a protein along DNA contains a series of local energy minima separated by energy barriers. Protein sliding on DNA has been modeled as a particle diffusing along a rough potential energy landscape. The roughness of the landscape reduces the diffusion constant from the theoretical maximum determined by solution viscosity. We found that the diffusion of TRF1 and TRF2 was ~17- to 37-fold slower at telomeric regions compared with non-telomeric λ DNA, corresponding to ~2.8 to 3.6 $K_BT$ increase in the roughness of the energy landscape (Supplementary Text and Figure 5B). Also, the TRF1 and TRF2 attached lifetimes within telomeric sequences were ~31- and 3-fold longer, respectively, compared with those on λ DNA (225 mM ionic strength, Table 2). These differences correspond to an increase of ~3.4 $K_BT$ (for TRF1) and 1.1 $K_BT$ (for TRF2) in relative binding energy at the telomeric regions (Supplementary Information). Taken together, the relative activation energy barriers based on the diffusion constants and lifetimes are not only consistent with each other, but also close to the estimated minimal roughness of the energy landscape at specific binding sites (~6.6 $K_BT$) for a genome size of 3 x $10^9$ bp (Supplementary Information) (35).
Interestingly, the percentage of TRF2 arriving at the (TTAGGG)$_{20}$ region (73%, 125 mM; Figure 4D) was lower than the simulated equivalent situation assuming TRF2 first binds to the non-telomeric spacer (98%, n = 500). This discrepancy is consistent with an additional activation energy barrier between telomeric and non-telomeric regions, likely due to a switch within TRF2 from a nonspecific binding mode to a specific recognition mode (Figure 5B) (64). Noticeably, for TRF2, this energy barrier was lower at 225 mM ionic strength than at 125 mM, since more proteins arrived at telomeric regions from the non-telomeric spacers (Figure 4D and Supplementary Figure S6B), consistent with the desolvation of electrostatic residues required for DNA binding.

In contrast to the metastable and dynamic nature of the TRF protein binding to telomeric sequences (Figure 5), other systems characterized by single-molecule imaging show long-lived stable binding to specific sequences. For example, the mismatch repair protein, MutSα binds to a mismatch (+ADP) with a half-life of 9.6 ± 1.5 min (36); and the average lifetime of the Type III restriction enzyme EcoP15I on DNA with specific binding sites was ~180 s (38). The primary differences between these systems are the target DNA sites. For TRF proteins, the target is a long repetitive sequence whereas for other systems target sites consist of much shorter non-repetitive DNA. We propose that TRF proteins utilize the combined free energy of binding from the association of multiple TRF proteins in the same region to increase binding specificity and stability. For example, TRF1 and TRF2 linked by TIN2 would increase the total affinity for telomeric sequences by summing the interaction energies of TRF1 and TRF2. We postulate that in vivo the diffusional properties of TRF proteins at the telomeric regions enable these proteins to search for their protein partners, such as another TRF-TIN2 complex, to assemble stable shelterin complexes on telomeric substrates. In this putative model of partner search, we expect that long distance searching is unlikely due to DNA-bound obstacles such as nucleosomes and other DNA-binding proteins. Rather 1-D diffusion represents a relatively local search mechanism which increases the probability of partner encounter during the attached period. In cells the intrinsic dynamics of TRF1 and TRF2 could potentially be important for regulating the assembly and disassembly of shelterin complexes, and switching between different telomere structures (capped and uncapped states).

In summary, using QD-conjugated proteins, DNA tightropes embedded with site specific sequences, AFM, and fluorescence imaging, we reveal that TRF1 and TRF2 use different mechanisms to find telomeric DNA but share a novel mechanism to search for protein partners at telomeres. Based on these results, we postulate a general mechanism for how multi-protein
complexes strike a balance between achieving specificity and target search, in a process we define as ‘tag-team proofreading’. In this model, proteins first form weak transient complexes with their cognate DNA sequences, and then rely on the additive energies of binding provided by partner proteins to generate higher specificity.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online, including Supplementary Text, Supplementary Legends for Movies S1–5, Table S1, Supplementary Figures S1–11, and Supplementary References [65-73].

ACKNOWLEDGMENTS

We would like to thank R. Keller and M. Nazareth at the University of Essex, Y. Jin at the Zhejiang University, Drs. K. Weninger, R. Riehn, S.F. Lim, K. Daniels at the North Carolina State University, and D. Erle at the University of North Carolina at Chapel Hill for providing critical comments and technical support.

FUNDING

This work was supported by the BBSRC [BB/I003460/1 to N.M.K.] and by the National Institutes of Health [ES0515052 to P.L.O. and 4R00ES016758 to H.W].

Conflict of interest statement. None declared.
REFERENCES


of TRF1 and TRF2 regulate their ability to condense telomeric DNA. *Nucleic acids research*, 40, 2566-2576.


and brus, L.E. (1996) Fluorescence intermittency in single cadmium selenide

The basic domain of TRF2 directs binding to DNA junctions irrespective of the presence


protein translocation on nucleic acids. 3. The Escherichia coli lac repressor--operator

TRF1 of their telomeric DNA-bound structures and DNA-binding activities. *Protein Sci*,
14, 119-130.

(2003) ERCC1/XPF removes the 3' overhang from uncapped telomeres and represses
formation of telomeric DNA-containing double minute chromosomes. *Molecular cell*, 12,
1489-1498.

and Songyang, Z. (2009) TRF2 functions as a protein hub and regulates telomere


Table 1: Summary of the diffusion constant, α factor, and lifetime of Sav-QD (655 nm) labeled TRF1 and TRF2, on λ DNA, at different ionic strengths.

| Ionic Strength (mM) | TRF1 | | TRF2 | |
|--------------------|------|--|------|--|------|
|                    | D ($10^{-5}$ μm$^2$/s) | α Factor | Lifetime | D ($10^{-5}$ μm$^2$/s) | α Factor | Lifetime |
| 75                 | 7.5±1.2 (51) | 0.65±0.04 (51) | - | 8.9±0.9 (59) | 0.94±0.05 (59) | 10±0.1 (104) |
| 125                | 5.5±1.4 (37) | 0.72±0.05 (37) | - | 8.4±0.9 (54) | 0.95±0.06 (54) | 2.5±0.1 (106) |
| 175                | 4.9±1.0 (40) | 0.72±0.06 (40) | 1.8±0.1 (53) | 9.5±0.1 (63) | 0.82±0.03 (63) | 4.8±0.1 (107) |
| 225                | 3.8±1.2 (33) | 0.89±0.07 (33) | 0.3±0.01 (12) | 9.5±0.1 (66) | 0.84±0.04 (66) | 3.4±0.1 (54) |
| 125-TRF2AR         | 9.1±1.8 (21) | 0.93±0.04 (21) | - | 9.1±1.8 (21) | 0.93±0.04 (21) | - |

Note: The numbers in the parentheses indicate the total number of complexes analyzed.

Lifetime was measured for complexes showing both protein binding and release events within the video frame (Type I, Figure 2D). Data are presented as mean ± standard error.
Table 2: Summary of the diffusion constant and lifetime of TRF1- and TRF2-QDs on the ligated T270 DNA substrates.

<table>
<thead>
<tr>
<th>DNA</th>
<th>Ionic strength (mM)</th>
<th>TRF1</th>
<th>TRF2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Telomeric</td>
<td>Non-telomeric</td>
</tr>
<tr>
<td></td>
<td>D \times 10^2 \mu m^2/s</td>
<td>Lifetime (s)</td>
<td>D \times 10^2 \mu m^2/s</td>
</tr>
<tr>
<td>Tel270</td>
<td>125</td>
<td>0.15±0.02 (22)</td>
<td>17.3±0.2 (15)</td>
</tr>
<tr>
<td>Tel270</td>
<td>225</td>
<td>0.22±0.04 (21)</td>
<td>9.2±0.2 (9)</td>
</tr>
<tr>
<td>Tel10</td>
<td>125</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: Proteins were labeled with equal molar amount of red (655 nm) and green (565 nm) QDs.
The numbers in the parentheses indicate the total number of complexes analyzed. Lifetime was measured for complexes showing both protein binding and release events within the video frame (Type I, Figure 2D). Data are presented as mean ± standard error.
FIGURE LEGENDS

Figure 1. TRF1- and TRF2-QDs retain DNA binding activity. (A) Schematic representations of the domain structures of TRF1 and TRF2. A: Acidic domain, B: Basic domain, M: Myb type domain. (B) Schematic representations of TRF1- and TRF2-QD conjugates (left), 8^Ttris-NTA compound (middle), and the DNA substrate (T270) with 2 tandem (TTAGGG)_{155} repeats connected by a short linker region (right, 5.4 kb in length). (C-E) Representative AFM images of DNA in the presence of (C) only QDs and 8^Ttris-NTA compound, (D) TRF1-QDs, or (E) TRF2-QDs. The scale bar is 200 nm. White arrows point to QDs bound to DNA. The numbers in C-E indicate the percent of DNA molecules bound with QDs in each condition. The total numbers of complexes analyzed were 200, 250, and 250, for no protein, TRF1-QDs and TRF2-QDs, respectively.

Figure 2. DNA tightrope assay based oblique-angle fluorescence imaging of TRF1- and TRF2-QDs on λ DNA tightropes. (A) A schematic drawing of the DNA tightropes (green lines) bound with QD (red ball) labeled proteins (green balls) between silica beads (large white balls). The drawing is not to scale. (B and C) Representative fluorescence images of red (655 nm) QD-conjugated His_{6}-TRF1 (B) and His_{6}-TRF2 (C) on λ DNA (stained with YoYo1). (D) Classification of different types of protein-DNA interactions observed with TRF2-QDs on λ DNA for attached lifetime measurement.

Figure 3. TRF1- and TRF2-QDs bind specifically to telomeric sequences on DNA tightropes. (A) A representative fluorescence image of DNA tightropes formed using ligated linear T270 DNA containing telomeric sequences (stained with YoYo1). (B) A schematic drawing of the ligated T270 DNA substrate (top) and representative fluorescence images of dual color (655 and 565 nm) labeled TRF1- (middle) and TRF2-QDs (bottom) on the ligated T270 DNA substrate. (C) Measured distances between two adjacent TRF1- (n = 96, top) and TRF2-QDs (bottom, n = 96) on the ligated T270 substrate (blue bars), and between TRF2-QDs on the non-
telomeric DNA substrate (bottom, white bars, n = 204). The lines in the top and bottom panels are double Gaussian fits to the data, which have $R^2$ of 0.99 and 0.95, respectively.

**Figure 4.** TRF1 and TRF2 show different diffusional properties over telomeric region versus non-telomeric regions. (A-C) Kymographical analysis of dual color (655 and 565 nm) labeled TRF1 (A, 125 mM ionic strength) and TRF2 (B:125 and C:225 mM ionic strengths) on the ligated T270 DNA. The panel left to the vertical white line shows a schematic drawing of the ligated T270 substrate with telomeric (purple) and non-telomeric sequences (blue), and a fluorescence image of the DNA with protein-QDs. The horizontal white lines indicate the estimated center of the telomeric region based on the spacing between adjacent QDs. The white arrows in C indicate TRF2 diffusing between two adjacent telomeric sequences. (D) The diffusion range distributions of TRF1- (top, n = 29) and TRF2-QDs (bottom, n = 28) on the ligated T270 substrate (blue bars), and TRF2-QDs on the non-telomeric DNA (bottom, white bars, n = 77). Diffusion ranges below and beyond 850 nm are categorized into short (telomeric) and long range (non-telomeric), respectively. The lines in the top and bottom panels of D are single and double Gaussian fits to the data, respectively, which have $R^2$ of 0.90 and 0.96, respectively.

**Figure 5.** TRF1 and TRF2 strike a balance between target search and specificity. (A) TRF1 and TRF2 can undertake a 1-D search on DNA consistent with rotation-coupled diffusion along the DNA helix. The small ovals represent the basic and acidic domains of TRF1 and TRF2. The blue and purple lines represent non-telomeric and telomeric DNA, respectively. TRF1 relies more on 3-D search and majority of the TRF2ΔB molecules bind to the telomeric region directly from solution forgoing the 1-D component of the search. (B) The energy landscape along the positions at telomeric and non-telomeric sequences. The diffusion constant and lifetime measurements are consistent with ~2.8–3.6 $k_B T$ higher energy barriers to diffusion at the telomeric sequences in comparison with non-telomeric sequences (Tables 1 and 2).
additional energy barrier at the non-telomeric and telomeric junction represents the activation energy needed for conformational change/DNA binding domain switching on proteins to achieve specific binding.
Figure 1
Figure 3

A

B

5.4 kb/1.6 μm
5.4 kb/1.6 μm

1.8 kb
1.8 kb
3.2 kb
3.2 kb
6.6 kb

TRF1
TRF2

2 μm

30
20
10
0

% of molecules

0 1 2 3 4 5 6 7
Distance (μm)

C

TRF1 on ligated T370

TRF2 on ligated non-kleismic DNA
Figure 5

A

- REF
  
  0.3 s
  3.8 x 10^2 μm^2/s

- TSE1
  
  3.2 s
  0.22 x 10^2 μm^2/s

WT

  3.4 s
  9.5 x 10^2 μm^2/s

WT/ΔB

  10.3 s
  0.29 x 10^2 μm^2/s

B

Position

ω

2.8 - 3.6 T
Table S1

Percentages of four types of protein-DNA interactions for TRF1 and TRF2 on λ DNA and the ligated T270 DNA substrate at different ionic strengths.

<table>
<thead>
<tr>
<th>Ionic Strength (mM)</th>
<th>DNA</th>
<th>TRF1-QDs (%)</th>
<th>TRF2-QDs (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>I</td>
<td>II</td>
</tr>
<tr>
<td>75</td>
<td>λ DNA</td>
<td>179</td>
<td>5</td>
</tr>
<tr>
<td>125</td>
<td>λ DNA</td>
<td>63</td>
<td>7.9</td>
</tr>
<tr>
<td>175</td>
<td>λ DNA</td>
<td>210</td>
<td>30</td>
</tr>
<tr>
<td>225</td>
<td>λ DNA</td>
<td>291</td>
<td>44.3</td>
</tr>
<tr>
<td>125</td>
<td>T270</td>
<td>184</td>
<td>10.9</td>
</tr>
<tr>
<td>225</td>
<td>T270</td>
<td>110</td>
<td>4.5</td>
</tr>
</tbody>
</table>

Note: The examples of each type of kymograph are shown in Fig. 2D. The total video length was 2 minutes for λ DNA and 4 mins for the ligated T270 DNA substrate. Increasing the video length to 4 mins for λ DNA does not change the lifetime measurement for the Type 1 complexes.
**Figure S1.** DNA substrates used for AFM and fluorescence imaging (not including λ DNA). The purple and blue/green regions represent telomeric and non-telomeric sequences, respectively. Ligated DNA substrates have heterogeneous lengths.
Figure S2. Stoichiometry of TRF2-QDs revealed by AFM imaging. (A) A schematic representation of TRF2 primary antibody binding to a TRF2-QD. (B) AFM images of TRF2-QDs in the presence of TRF2 primary antibody (TRF2:Ab = 1:1). White arrows point to TRF2-QDs with a single antibody. (C) Cross section analysis of a TRF2-QD complex binding to TRF2 antibody. The section analysis on the right is from the path drawn in the AFM image on the left (red line).
Figure S3. TRF1- and TRF2-QDs bind specifically to the telomeric sequences. (A) Position distributions of TRF1-QDs (black bars, n = 103) and unlabeled TRF1 (gray bars, n = 85) on the T270 telomeric substrate, and TRF1-QDs on the non-telomeric DNA substrate (white bars, n = 13). (B) Position distributions of TRF2-QDs (black bars, n = 106), unlabeled TRF2 (gray bars, n = 76) on the T270 telomeric substrate, and TRF2-QDs on the non-telomeric DNA substrate (white bars, n = 100). The (TTAGGG)$_{270}$ sequence makes up approximately 30% of the total DNA length and is located in the middle of the linearized T270 DNA (between 35% to 50% from one end). Among the protein-QDs on DNA, ~40.6% TRF1-QD and ~52.6% of TRF2-QD bound to the telomeric regions (excluding complexes bound right at the end of the DNA). The small peak at 24% is consistent with the locations of previously discovered preferred TRF1 and TRF2 binding sequence (CCATTC) over the non-telomeric region. (C and D) Position distributions of TRF1-QDs (C, n = 67) and TRF2-QDs (D, n=51) on linear Tel10 DNA substrate with (TTAGGG)$_{10}$ sequence located at 42% from one DNA end.
**Figure S4.** Diffusional properties of TRF1 and TRF2 on λ DNA. (A) Percentages of motile TRF1- and TRF2-QDs on λ DNA. n is the number of complexes analyzed. TRF proteins were labeled with red QDs (655 nm). (B) Diffusion constants of TRF1- (white bars) and TRF2-QDs (black bars) at different ionic strengths. (C) Alpha factors of TRF1- (white bars) and TRF2-QDs (black bars) at different ionic strengths. The numbers of the complexes analyzed were 51, 37, 40, and 33 for TRF1-QDs, and 59, 54, 63, and 66 for TRF2-QDs, at 75, 125, 175, 225 mM ionic strengths, respectively. (D) The effect of ionic strength on the lifetimes of TRF1- (white bars) and TRF2-QDs (black bars) on λ DNA. The numbers of the complexes analyzed were 63 and 128 for TRF1-QDs, and 104, 106, 107, and 95 for TRF2-QDs with increasing ionic strengths. Across all ionic strengths, compared to TRF2, TRF1 showed lower affinity for nonspecific λ DNA, indicated by lower average numbers of protein-QDs (1.1 ± 0.1 vs. 13.3 ± 2.2) in each field of view (27 μm x 55 μm) and lower average numbers of TRF1-QDs on DNA tightropes(s) between two beads (1.1 ± 0.1 vs. 3.8 ± 0.3).
**Figure S5.** Dual color labeling of TRF1 and TRF2. A representative fluorescence image (A) and kymograph (B) of dual-color labeled TRF1-QDs on λ DNA. A representative ORF image (C) and kymograph (D) of dual-color labeled TRF2-QDs on λ DNA. TRF1 or TRF2 was incubated with equal molar of red (655 nm) and green (565 nm) QDs. The kymographs shown in (B) and (D) are from the boxed regions in (A) and (C), respectively. Among the observed TRF2-QD complexes encountering another protein on DNA (n = 26), we did not observe any protein barrier bypass events for TRF2. Due to the lower binding affinity of TRF1 on DNA with non-telomeric sequences, even though we did not observe bypass when TRF1-QD complexes encountered barriers posed by other proteins (n = 3 out of 87 binding events), the number of events was not sufficient to make any statistical comparisons. (E) Percentages of dual color QD labeled TRF1 and TRF2.

**Table:** Percentages of dual color labeled TRF1 and TRF2 on λ DNA.

<table>
<thead>
<tr>
<th>Complexes</th>
<th>TRF1-QDs</th>
<th>TRF2-QDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number</td>
<td>141</td>
<td>43</td>
</tr>
<tr>
<td>Dual color (%)</td>
<td>19%</td>
<td>79%</td>
</tr>
<tr>
<td>Red (%)</td>
<td>41%</td>
<td>10%</td>
</tr>
<tr>
<td>Green (%)</td>
<td>39%</td>
<td>11%</td>
</tr>
</tbody>
</table>
Figure S6. Additional information on the length of ligated T270 DNA tightropes, the diffusion range distributions, and distances between two adjacent TRF2-QDs on the ligated T270 substrate. (A) Measured ligated T270 DNA lengths between two anchoring silica beads (n = 163). (B) The diffusion range distribution of TRF2-QDs at 225 mM ionic strength (n = 41). The line in the plot is double Gaussian fit to the data with R² of 0.97 and peaks centered at 0.5 and 1.5 μm. (C) The diffusion range of TRF2-QDs on ligated T270 DNA at 0.68 nM TRF2 and 0.33 nM QD concentrations (n = 33). The line in the plot is single Gaussian fit to the data with R² of 0.97 and the peak centered at 0.5 μm. Individual DNA tightropes with one or two QDs (no contact during recording) were analyzed. (D) The distances between adjacent TRF2-QDs on DNA tightropes formed at 25 μl/ml flow rate (n = 121). The line in the plot is double Gaussian fit to the data with R² of 0.97 and peaks centered at 1.5 and 2.7 μm. The spacing between adjacent TRF2-QDs indicate that DNA was stretched to ~88% of its contour length.
Figure S7. Dynamics of TRF2-QDs on the ligated Tel10 DNA. (A) Kymographs of TRF2-QDs on the ligated Tel10 DNA. On this substrate, there were two populations: static complexes with apparent diffusion constant of $2.5 \pm 0.05 \times 10^{-6} \mu m^2/s$ ($n = 10$) and complexes with long diffusion range (>850 nm, white arrows). (B) Measured distance between two adjacent TRF2-QDs on ligated T10 DNA substrate ($n = 35$). The line in (B) is Gaussian fit to the data, which has $R^2$ of 0.98. Consistent with a larger spacing (5.99 kb) between two adjacent telomeric regions on the Tel10 DNA substrate, static TRF2-QDs were spaced at longer distances ($1.9 \pm 0.049 \mu m$) compared to that on the ligated T270 substrate ($1.6 \pm 0.01 \mu m$, Fig. 3).
Figure S8. Short diffusion ranges of TRF1- and TRF2-QDs at (TTAGGG)_270 telomeric regions are narrowly distributed independent of the video length. The diffusion range over time for TRF1-QDs at 125 mM ionic strength (A) and 225 mM strength (B), and TRF2-QDs at 125 mM ionic strength (C), and 225 mM strength (D) on λ DNA (red dots) and the ligated T270 DNA (blue squares). The numbers of complexes plotted for TRF1 are 37 and 33 for λ DNA, 26 and 21 for T270 DNA, at 125 and 225 mM ionic strength, respectively. The numbers of complexes plotted for TRF2 are 48 and 66 for λ DNA, 22 and 34 for T270 DNA, at 125 and 225 mM ionic strength, respectively. On λ DNA, the diffusion ranges are widely distributed between approximately 0.5 to 9 μm. The time scale (from 0 to 120 s) is from the cropped videos used for the diffusion constant and diffusion range analysis.
Figure S9. Computer simulations of diffusion by modeling random walk of proteins on a 1-D DNA lattice using Python™ programming language. The diffusion constants used for simulation were that of TRF2-QDs on λ DNA (A) and on the telomeric DNA (B) at 125 mM ionic strength. A plot for MSD vs. Δt obtained and an example of a trajectory of a protein (Insert) are presented for (left, A) DNA with unlimited length and (right, A) DNA with 1.6 kb length at 1460436 steps/s stepping rate (corresponding to TRF2-QDs on λ DNA). A plot for MSD vs. Δt and an example of a trajectory of a protein (Insert) are also presented for (left, B) DNA with unlimited length and (right, B) DNA with 1.6 kb length at 46713 steps/s (corresponding to the rate of TRF2-QDs at the telomeric region). The fitting parameters were constructed such that only the initial linear portion of the MSD vs. Δt plots was used for calculating the diffusion constant. For diffusion with confinement, a protein walks along a 1-DNA lattice with two totally reflecting barriers. The numbers in each plot are the mean and standard deviation of the simulated data. The number of particles simulated for each case is indicated in the parentheses.
**Figure S10.** Dissociation events of TRF1 and TRF2-QDs on the ligated T270 DNA. Representative kymographs of TRF1- (A) and TRF2-QDs (B) on the ligated T270 DNA at 125 mM ionic strength. The scale bar is 1 μm. The yellow arrows point to dissociation events either directly from the telomeric region or through the non-telomeric region. The purple arrows point to the transient dissociation and rebinding at the same telomeric region.
**Figure S11.** Dynamics of TRF2ΔB-QDs on λ DNA and the ligated T270 DNA. (A) Position distributions of TRF2ΔB-QDs on the linear T270 DNA substrate (n = 51). Among the protein-QDs on DNA, ~66% TRF2ΔB-QDs bound to the telomeric regions (35% to 50% from DNA ends). Kymographs of TRF2ΔB-QDs on λ DNA (B) and the ligated T270 DNA (C). The scale bar is 1 μm. Protein-QD-DNA reactions were carried out at 125 ionic strength. (D) Diffusion range distributions of TRF2ΔB-QDs on λ DNA (white bars, n = 21) and the ligated T270 DNA (striped bars, n = 30). The binding affinity of TRF2ΔB to λ DNA tightropes was significantly lower as indicated by lower average numbers of TRF2ΔB-QDs on DNA in the field of view (1.4 ± 0.2 vs. 13.3 ± 2.2) and lower average numbers of protein-QDs on DNA tightropes between two beads (1.1 ± 0.1 vs. 3.8 ± 0.3). Dual color QD labeling confirmed that TRF2ΔB can form dimers or higher order oligomeric species, but at a significantly lower percent (14%) compared with the full length protein (79%). The fraction of motile protein-DNA complexes decreased from 95% for full length TRF2 to 65% (n = 40) for TRF2ΔB at 125 mM and from 89% to 74% (n = 23) at 225 mM ionic strength.