SUPPLEMENTARY INFORMATION

For manuscript

Functional interplay between SA1 and TRF1 in telomeric DNA binding and DNA-DNA pairing

Jiangguo Lin[1,2], Preston Countryman[2], Haijiang Chen[3,4], Hai Pan[2], Yanlin Fan[3], Yunyun Jiang[3], Parminder Kaur[2], Wang Miao[5], Gisele Gurgel[6], Changjiang You[7], Jacob Piehler[7], Neil M. Kad[8], Robert Riehn[2], Patricia L. Opresko[9], Susan Smith[10], Yizhi Jane Tao[3], Hong Wang[2,*]

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

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

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

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

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

[7]Division of Biophysics, *Universität Osnabrück*, Barbarstrasse 11, 49076, *Osnabrück*, Germany

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

[9]Department of Environmental and Occupational Health, University of Pittsburgh, Pennsylvania 15213, USA

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

This Supplementary Information contains:

Supplementary Methods

DNA substrates for single-molecule imaging

pSP73 vector (Promega) derived pSXneo(T2AG3) plasmid DNA containing 270 TTAGGG repeats (T270) was purchased from Addgene and purified from Stbl2 cells (Invitrogen) to retain stable telomeric DNA lengths (62). To generate DNA fragments containing TTAGGG repeats for AFM imaging, the restriction digestion of T270 DNA (10 µg) was carried out at 37 °C for 4 hr using *HpaI* (130 U) in Buffer 4 (New England BioLabs). Final DNA substrate purification was done using a PCR DNA Purification Kit (Qiagen). To generate the DNA substrate containing centromeric sequences, first, a 1.7 kb PCR product with human X-linked alpha-satellite sequences was generated using the forward primer (5' GCC GAT TCA TTA ATG CAG GTT AAC AAA GGC TTT CAG GCT TTT CCA CCA 3'), reverse primer (5'AAT TTC GAT AAG CCA GGT TAA CAA GTG GCT ATT TAG CGG GCT TGG A 3'), and genomic DNA purified from MDA-MB-231 breast cancer cells as the template. The PCR product was gel purified and subsequently cloned into the pSP73 vector (Promega) digested with HpaI. In order to generate longer DNA substrates for the DNA tightrope assay, linearized DNA substrates were ligated using the Quick Ligation™ Kit (New England BioLabs) at room temperature for 30 min. Ligated DNA samples were then purified using phenol-chloroform extraction.

Protein purification

The full length WT SA1 and SA1 R37A R39A mutant (1258 AA) containing a N-terminal 3X Flag tag (DYKDHDGDY KDHDIDYKDD DDK) was overexpressed using a baculoviruse system and purified using the anti-Flag M2 column (GenScript). To purify the 6xHis tagged SA1 (His-SA1), ~2 x 10^8 Sf21 cells were transfected with baculoviruses overexpressing His-SA1 and harvested 48 hr post-infection. Cell pellets were re-suspended in 10 ml of Ni-NTA buffer A (50 mM Tris-HCl, pH 7.5, 300 mM NaCl, 20 mM imidazole, 10% glycerol, 0.1 mM PMSF) and lysed by sonication. Supernatants from the lysates were collected by 35,000 g x 40 min centrifugation and loaded onto 0.2 ml of pre-equilibrated Ni-NTA resin (Qiagen). The supernatant was then constantly mixed with the resin at 4°C for 1 hr. The resin was collected by 1,000 g x 3 min centrifugation and the supernatant (flow through) was removed. The resin was then washed with 5 ml of Ni-NTA Buffer A. The proteins bound to the Ni-NTA agarose were eluted by Ni-NTA buffer B (50 mM Tris-HCl, pH 7.5, 300 mM NaCl, 250 mM imidazole, 10% glycerol, 0.1 mM PMSF). The concentration of the purified protein was determined by the Bradford protein assay using BSA as the protein concentration standard. SA1 was more than 90% pure based on SDS-PAGE analysis. The His_6-tagged single-chain antibody fragment (37 KDa) was overexpressed in *E. coli* and purified using a Ni-NTA agarose column (Qiagen) according to protocols specified by the manufacture.

Fluorescence anisotropy

The telomeric and non-telomeric DNA substrates were made by annealing 5' Alexa488 labeled top and unlabeled complementary stands using the following oligos:

Alexa-Tel7-top:

5'CTGGATCCGTAC**TTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGG**ACAAGAATTCGA3'

Tel7-com:

5'TCGAATTCTTGT**CCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAA**GTACGGATCCAG3'

Scram-top:

5'CTGGATCCGTACAGTGTAAGGGTGAGTGGTTGGGTGTGTGGGTGTTGATGTAGAACAAGAATTCGA3'

Scram-com:

5' TCGAATTCTTGTTCTACATCAACACCCACACACCCAACCACTCACCCTTACACTGTACGGATCCAG3'

Flag-tagged full length SA1 was concentrated 10-fold from 0.110 mg/ml (stock concentration) to 1.1 mg/ml and dialyzed in the DNA binding buffer [20 mM Hepes (pH 7.5), 0.1 mM $MgCl_2$, 0.5 mM DTT, 100 mM KCl] to remove the glycerol in the storage buffer. SA1 was then titrated into the binding solution containing DNA (0.2 nM) until the millipolarization (mP) stabilized. The data were plotted and analyzed by using the equation $P = \{(P_{bound} - P_{free})[protein]/(K_d + [protein])\} + P_{free}$, where P is the polarization measured at a given total protein concentration, P_{free} is the initial polarization of Alexa488-labeled DNA without protein bound, P_{bound} is the maximum polarization of DNA when all DNA molecules are bound by proteins, and [protein] is the protein concentration. The free and total protein concentrations were assumed to be equal since the K_d was at least 10-fold higher than the concentration of Alex488-labeled DNA. The binding curves were fitted by the nonlinear least-squares regression analysis assuming a bimolecular model of pseudo first order such that the K_d values represent the protein concentrations at which half of the DNA substrates are in the bound-state.

Electrophoresis mobility shift assays (EMSAs)

The telomeric DNA substrate used for EMSAs was constructed by annealing a 5' Alexa488 labeled oligo (5'TTA GGG TTAGGG TTAGGG ATG TCC AGC AAG CCA GAA TTC GGC AGC GTA3') with a second unlabeled oligo containing its complementary sequences. TRF1 and SA1 DNA binding reactions were carried out in a buffer containing 20 mM Hepes (pH 7.9), 150 mM KCl, 1 mM MgCl, 0.5 mM DTT, 0.25% NP-40, and 5% glycerol with DNA substrate and protein concentrations indicated in the figure legends (Figure 5A). The reactions were run by gel electrophoresis on a 5% 29:1 (bisacrylamide:acrylamide) native gel at 4°C and 150 V for about 40 min in 1X TBE buffer. The gel was visualized using a Typhoon FLA 7000 Phosphorimager. The percentages of unbound and bound DNA substrate were quantified using ImageQuant software.

Coarse-grained molecular dynamics simulations

We used the molecular dynamics package HOOMD-blue to explore the probable pathway that leads to protein-mediated DNA-DNA pairing under the action of a coarse-grained, simplified functional model of SA1 and TRF1. All molecules are made of 1 nm-wide beads. Bonded interactions within DNA and the protein are represented through a harmonic potential with spring stiffness of 330 k_BT/nm^2. Mutual steric exclusion between all pairs of particles is ensured by truncated shifted Lennard-Jones potentials. DNA is modeled as a chain of bonded beads with harmonic angle potentials. Non-bonded DNA segment pairs interact through a screened Coulomb potential, which raises the persistence length to 50 nm (63). Protein domains mainly consist of "inert" beads that are present only to establish steric structures containing the DNA binding (Myb) domain, linker, dimerization domain for TRF1, and only a DNA binding domain for SA1.

Specific attractive interactions were added to this model by adding specific beads with additional Lennar-Jones interaction terms. Specifically, those interactions are between

1) A bead within the groove of either SA1 or the DNA-binding domain of TRF1 and DNA.
2) A bead at the center of the TRF1 dimerization domain and the corresponding bead at the center of another TRF1 dimerization domain.
3) A bead at the side of SA1 (along the DNA binding groove) and a bead at the side of the DNA-binding domain of TRF1 (in the same location as the bead on SA1).

The model is integrated using a Langevin integrator.

Time steps were ~1, 2, or 4 ps, and the total time scale of the simulation was ~ 3 ms. Results were identical with different time steps. Because DNA was confined to a small simulation box, the effective concentration of DNA (400 mg/l, 6.7×10^{-4}mol/l of bp) was rather high, requiring a similarly high concentration of TRF1 (1.9×10^{-5} mol/l) and SA1 (1.9×10^{-6} mol/l) to create similar ratios of protein to DNA compared to experimental conditions.

Monte Carlo Simulations of protein diffusion on DNA

To determine if the observed slowing of SA1 diffusion at telomere sequences could be due to pausing, Monte Carlo simulations were performed using MATLAB. The model is based on the assumption that protein molecules exist in two states (free diffusing and pausing) and the DNA context affects the equilibrium between these states. We created a model DNA strand with 1.6 kb telomeric sequences spaced 3.8 kb apart by non-telomeric DNA. For every simulation cycle the molecule could either take a 10 bp step with a probability of stepping (equal chance for forwards/backwards) or pausing. When the protein molecule was in regions containing non-telomere DNA, a bias was applied towards the probability of stepping of 92%. Upon entering a region defined as containing telomeric DNA, the probability of stepping was reduced to 20%. Pausing durations were randomly selected from an exponential probability distribution. The number of iterations used was $1x10^4$. However, because the

duration of pausing advanced the step position, consequently the total number of steps simulated was ~10x greater than the iteration number. The trace shown in Supplementary Figure S11 is highly representative of all the traces simulated.

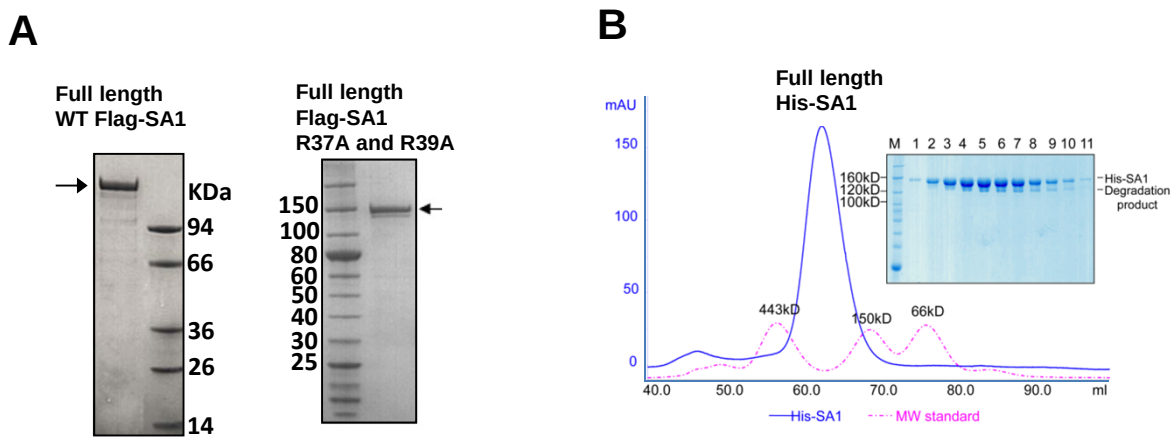
Supplementary Movie Legends

1. **Supplementary Movie S1.** A full length Flag-SA1-QD complex displaying alternation between slow and fast diffusion on a T270 DNA tightrope. The scale bar is 1 μm.
2. **Supplementary Movie S2.** A SA1-N-QD complex displaying alternation between slow and fast diffusion on a T270 DNA tightrope. The scale bar is 1 μm.
3. **Supplementary Movie S3.** A dual color differentially QD-labeled SA1-TRF1 complex showing subdiffusive motion on a T270 DNA tightrope. The scale bar is 1 μm.

Supplementary References

62. Hanish, J.P., Yanowitz, J.L. and de Lange, T. (1994) Stringent sequence requirements for the formation of human telomeres. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 8861-8865.

63. Hsieh, C.C., Balducci, A. and Doyle, P.S. (2008) Ionic effects on the equilibrium dynamics of DNA confined in nanoslits. *Nano letters*, **8**, 1683-1688.

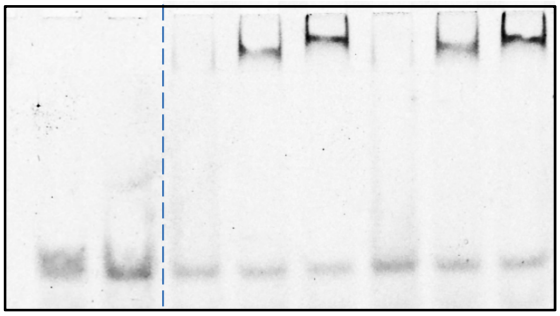


Supplementary Figure S1. Purification of His- and Flag-tagged full length SA1 and characterization of the oligomeric state of SA1 in solution. (**A**) Coomassie blue stained SDS-PAGE gel of Flag-tagged WT SA1 and SA1 R37A R39A mutant (Genscript). (**B**) Evaluation of the oligomeric state of the full length His-SA1 in solution using a gel filtration column. Lanes 1-11, peak fractions from gel filtration.

A

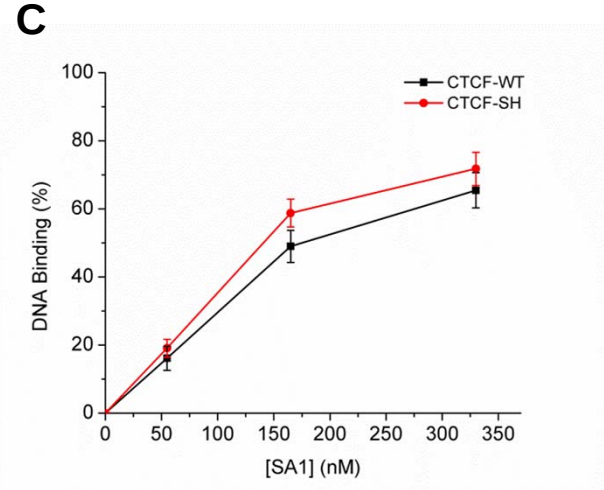
CTCF-WT DNA sequences (CTCF-WT):
5'Cy3/TCCACAATGACATGATGGCCAGCAGAGGGCGCATGGCCCTGGGGACCT3'
Shuffled DNA sequences (CTCF-SH):
5'Cy3/TCCACAATGACATGGCAGCATGGCGTGCGAGAACGGCCCTGGGGACCT3'

B



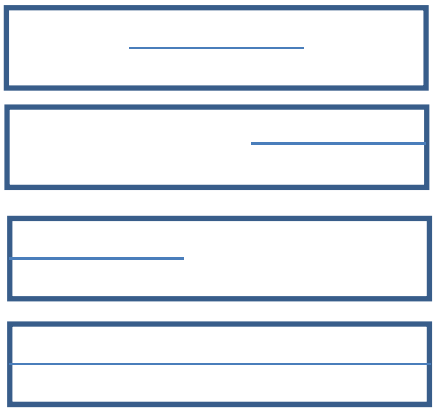
CTCF-SH	+	-	+	+	+	-	-	-
CTCF-WT	-	+	-	-	-	+	+	+
SA1 (nM)	-	-	55	165	330	55	165	330

C



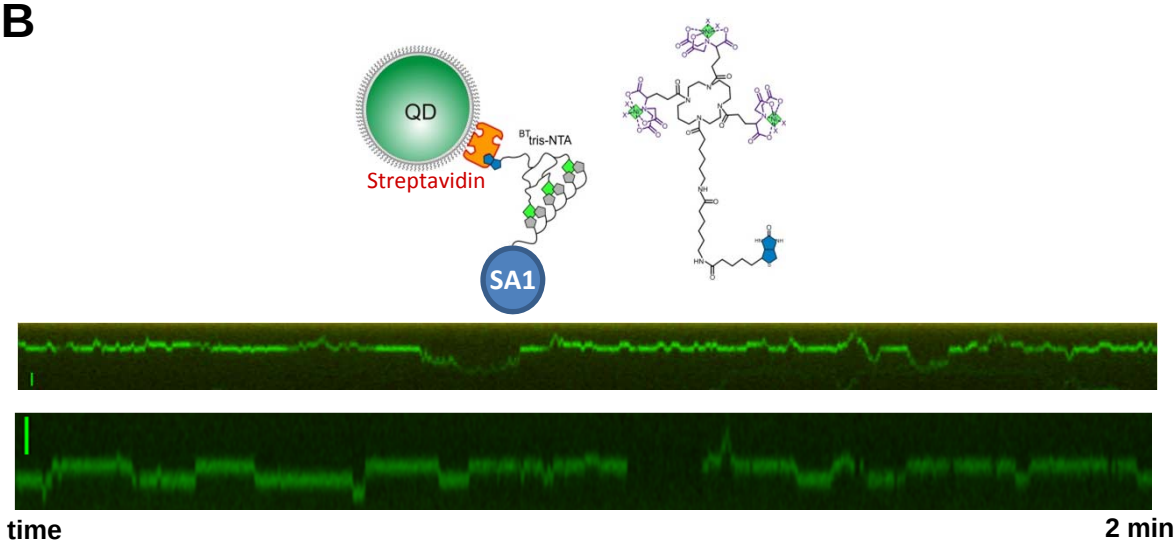
Supplementary Figure S2. SA1 does not specifically bind to consensus CTCF DNA binding sequences. (**A**) The DNA substrate with consensus CTCF DNA binding sequences and control substrate with shuffled sequences used for EMSAs. The sequences from the top strands are shown. (**B**) EMSA of Flag-tagged 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

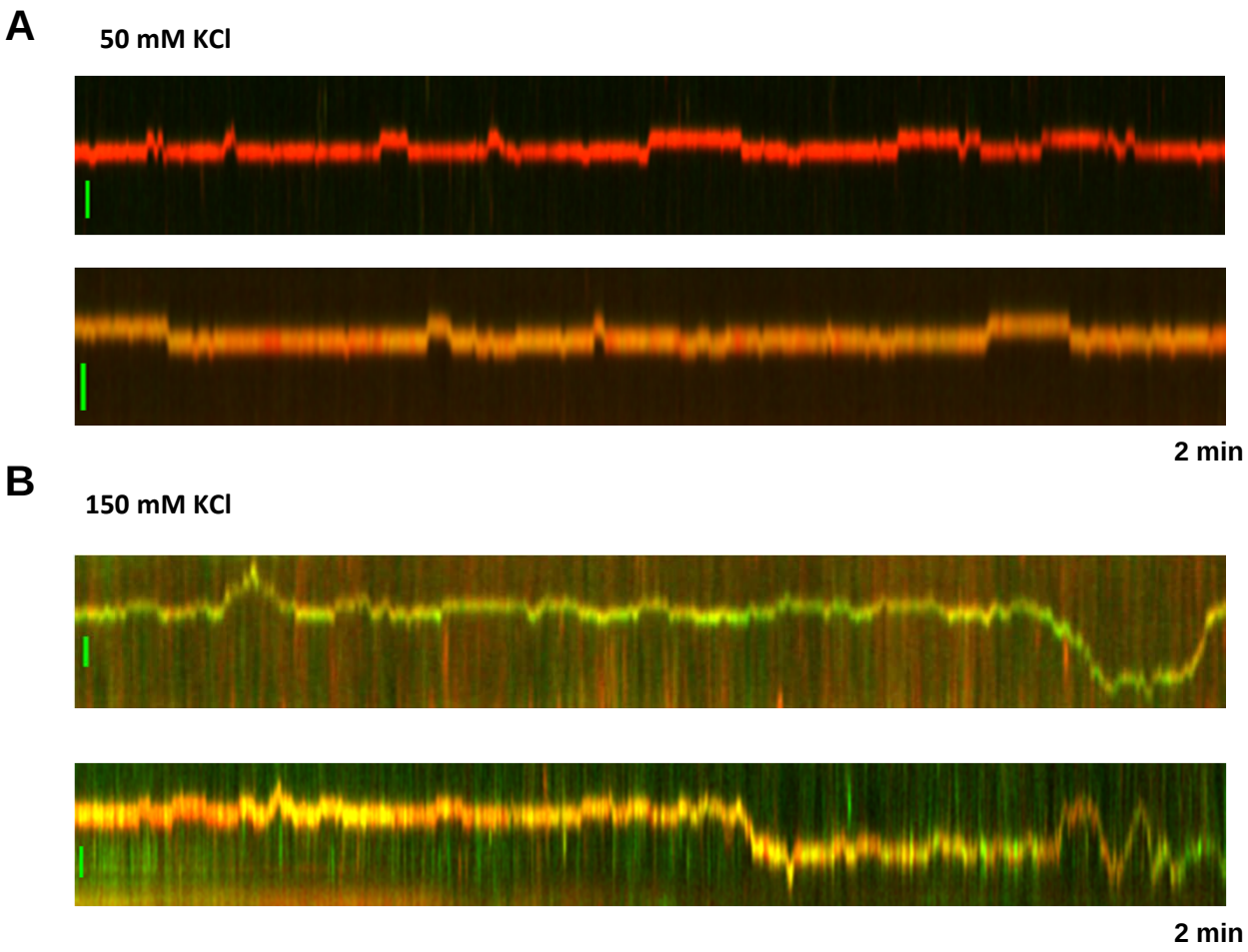


	Flag-SA1 (%, N = 107)	SA1+TRF1 dual-color (%, N = 101)
Type I	0	0
Type II	0	1.3 ±1.0
Type III	21.0 ± 7.0	20.3 ± 10.4
Type IV	78.5 ± 7.0	78.4 ± 9.6

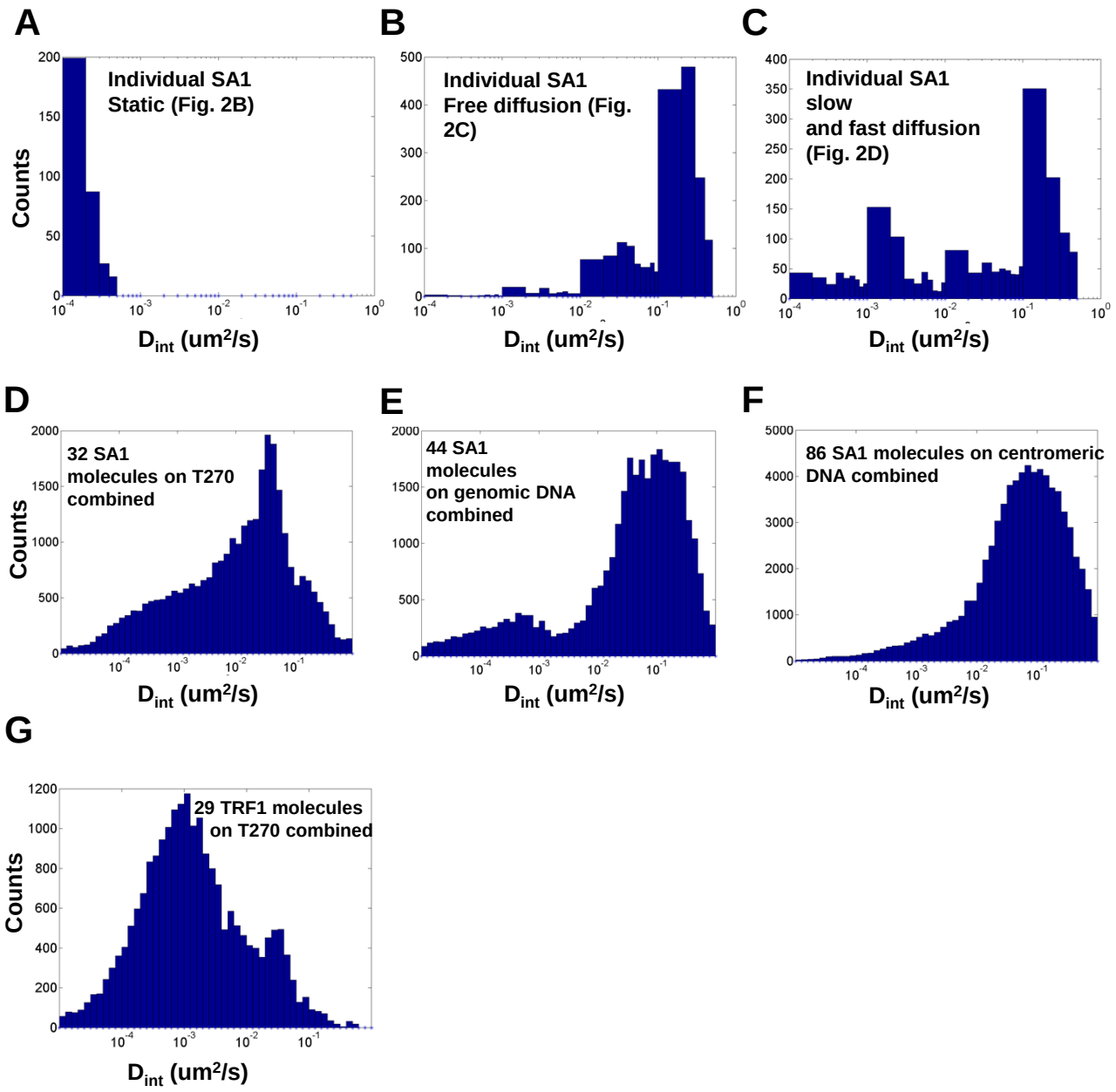
B



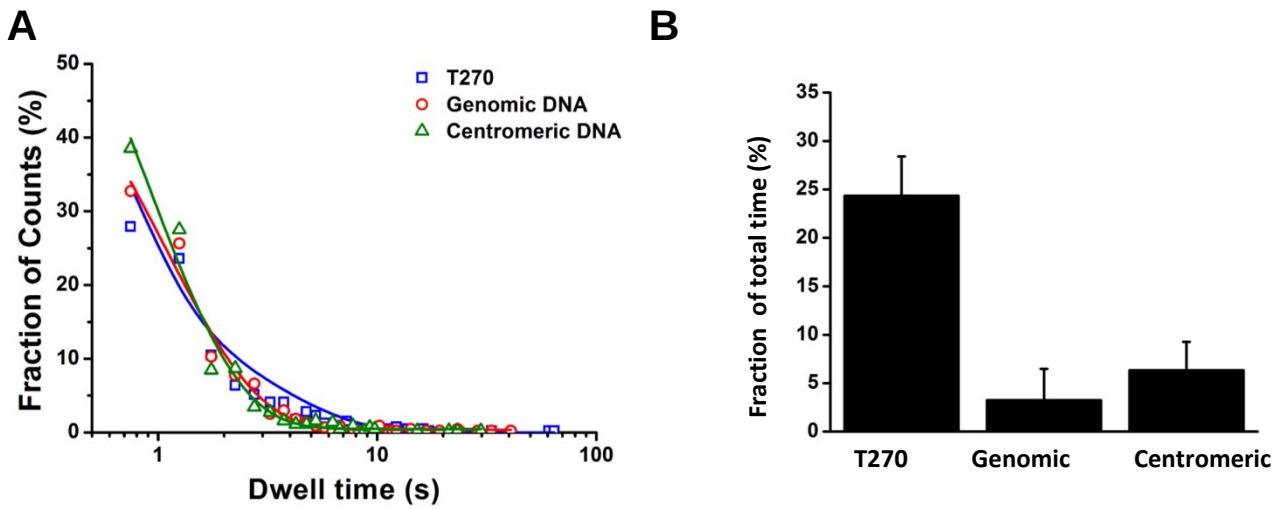
Supplementary Figure S3. Classification of different types of SA1-DNA interactions (A) and QD-labeled His-tagged SA1 on T270 DNA tightropes (B). (**A**) Cartoon drawings of different types of SA1-DNA interactions (left) and statistical analysis of these different types for Flag-SA1 and SA1-TRF1 complexes. 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-nitrilotriacetic acid (^{BT}tris-NTA). Bottom panel: Dynamics of His-SA1-QDs on T270 DNA tightropes. Scale bars (y-axis: 1 μm).



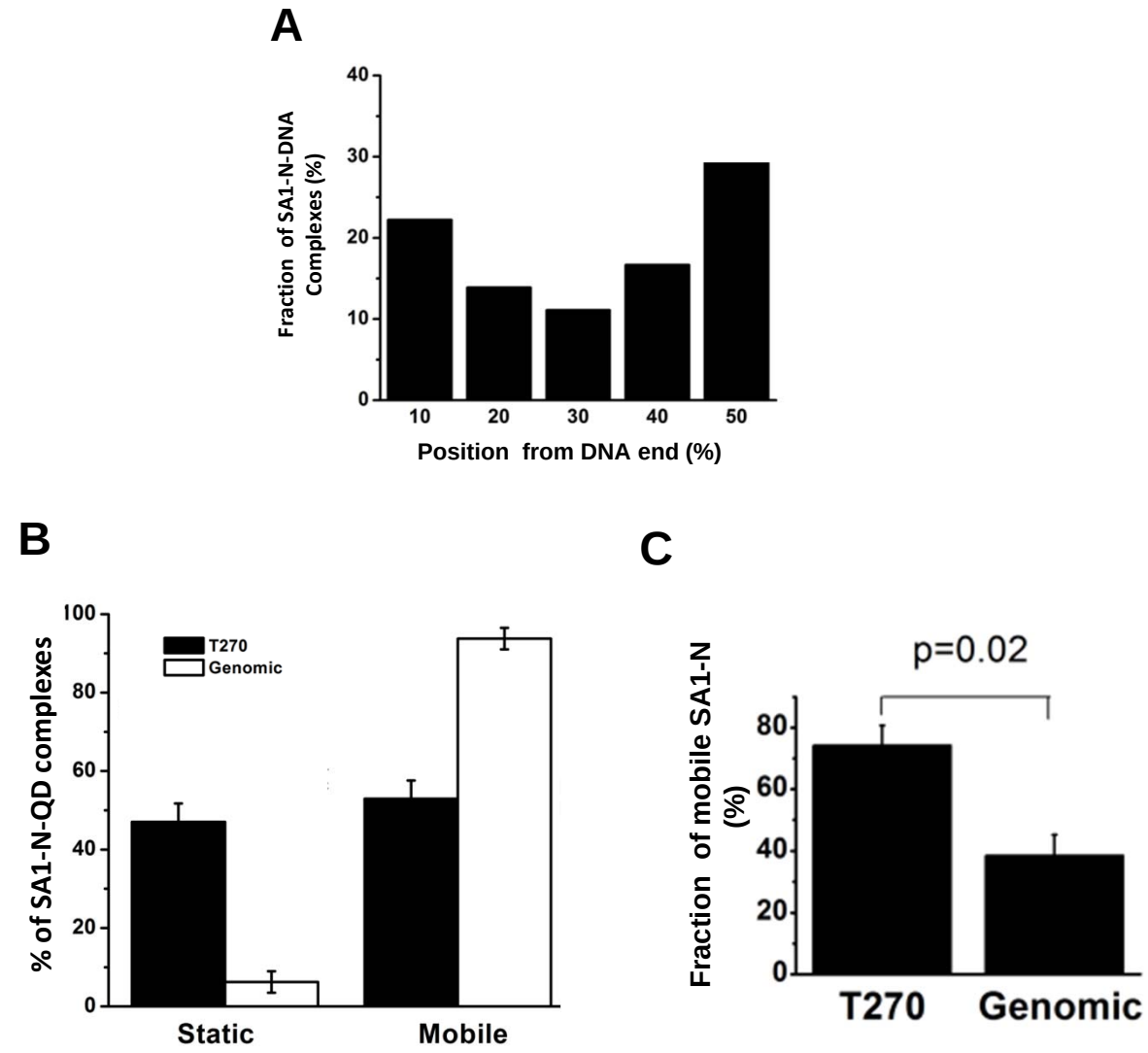
Supplementary Figure S4. The full length Flag-SA1 alternates between fast and slow diffusion at lower (A), and higher (B) salt concentrations. Representative kymographs of Flag-SA1 on T270 DNA tightropes in buffers containing 50 mM KCl (**A**) or 150 mM KCl (**B**). Scale bars (y-axis): 1 μm. Equal molar 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 $MgCl_2$.



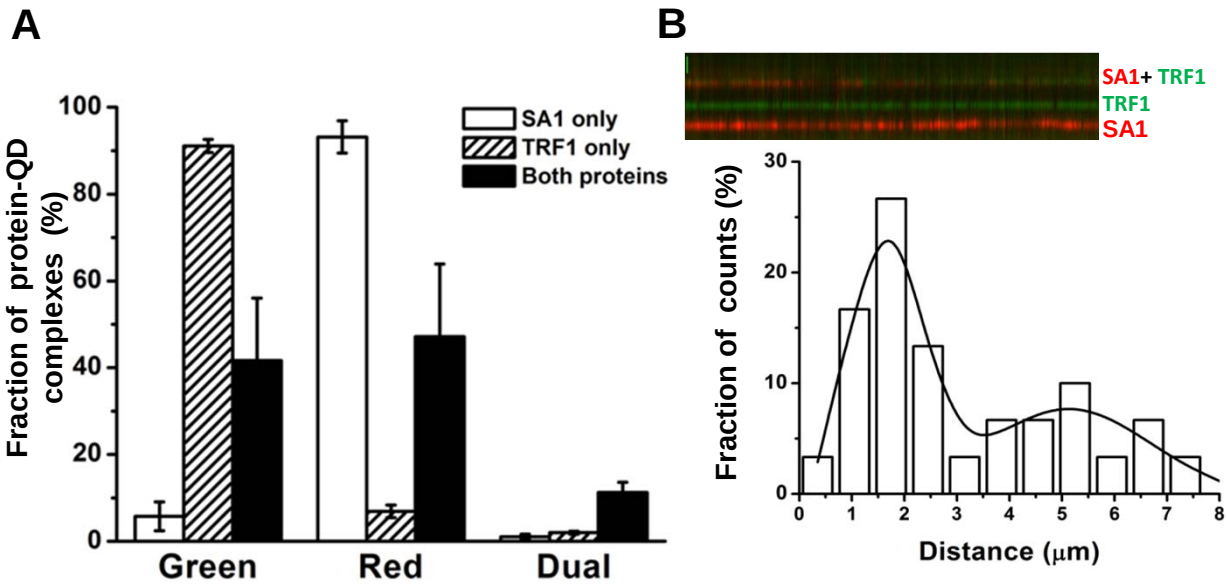
Supplementary Figure S5. Comparison of the distributions of interval based diffusion constants (D_{int}) for SA1 (A-F) and TRF1 (G). (**A**-**C**) Distributions of D_{int} for individual SA1 molecules on T270 DNA tightropes with static (A, kymograph in Fig. 2B), free diffusion (B, kymograph in Fig. 2C), and alternation between fast and slow 1-D diffusion (C, kymograph in Fig. 2D). The SA1 molecules with alternation between fast and slow diffusion show a distinct peak at ~1.0 X10^{-3} μm²/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 1X 10^{-2} μm²/s. In comparison, on genomic and centromeric DNA fast diffusion dominates with only 25% and 18% of D_{int} values less than 1X 10^{-2} μm²/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 1X 10^{-2} μm²/s).



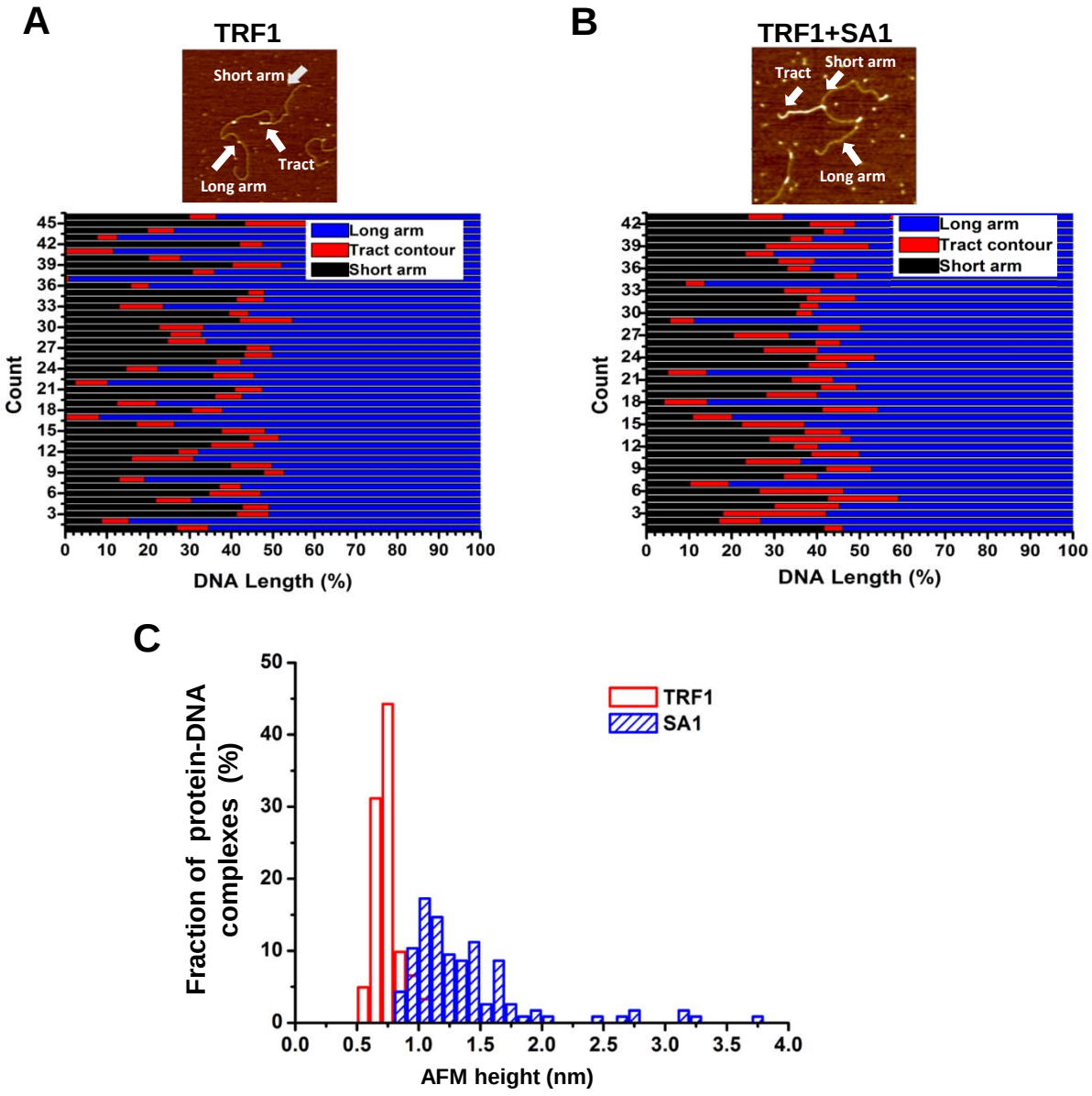
Supplementary Figure S6. Analysis of the dwell time shows distinct DNA binding dynamics by full length WT SA1 on T270 DNA tightropes. (A) Dwell times of individual slow diffusion events on T270 (dwell time: 1.17 s, R^2: 0.97, N = 437), the genomic substrate (dwell time: 0.80 s, R^2: 0.99, N = 437), and the DNA substrate containing centromeric sequences (dwell time: 0.79 s, R^2: 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.



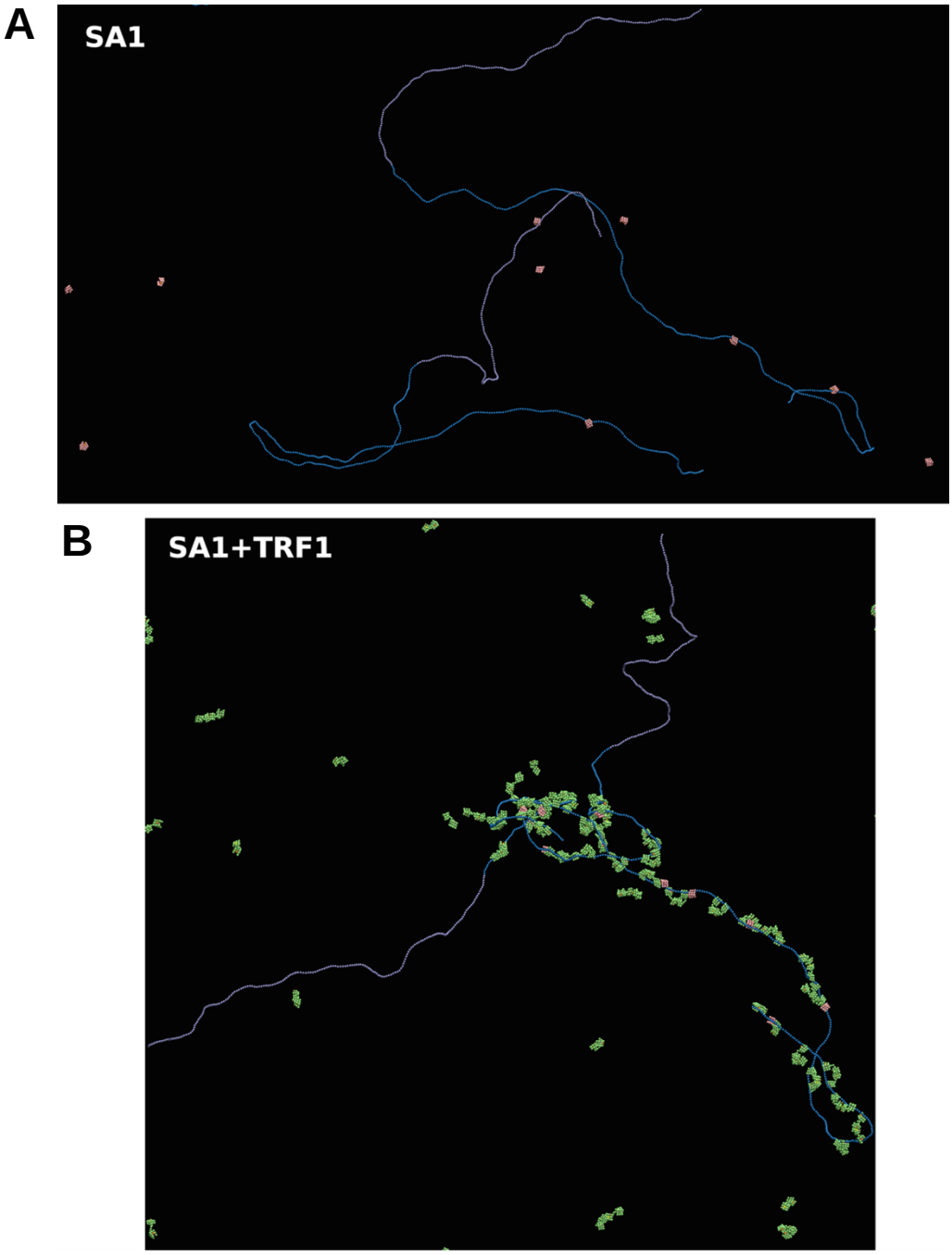
Supplementary Figure S7. The SA1 N-terminal domain (SA1-N) displays alternation between fast and slow diffusion on T270 DNA tightropes. (**A**) The position distribution of SA1-N on T270 DNA based on the analysis of AFM images of SA1-N-T270 complexes. Similar to the full length SA1 (Fig. 1), SA1-N preferentially binds to the telomeric region (35 – 50%) on T270 (N = 36). (**B**) On T270 tightropes, SA1-N displays a higher percentage of static (47.1%) complexes, compared to on genomic DNA (6.2%). (**C**) On T270 tightropes, mobile SA1-N molecules display a higher percentage (74.3%, N = 88) of complexes with slow diffusion events ($D_{in} < 5.0 \times 10^{-3}$ μm^2/s, dwell time > 2.2 s), compared to on genomic DNA (38.6%, N = 103).



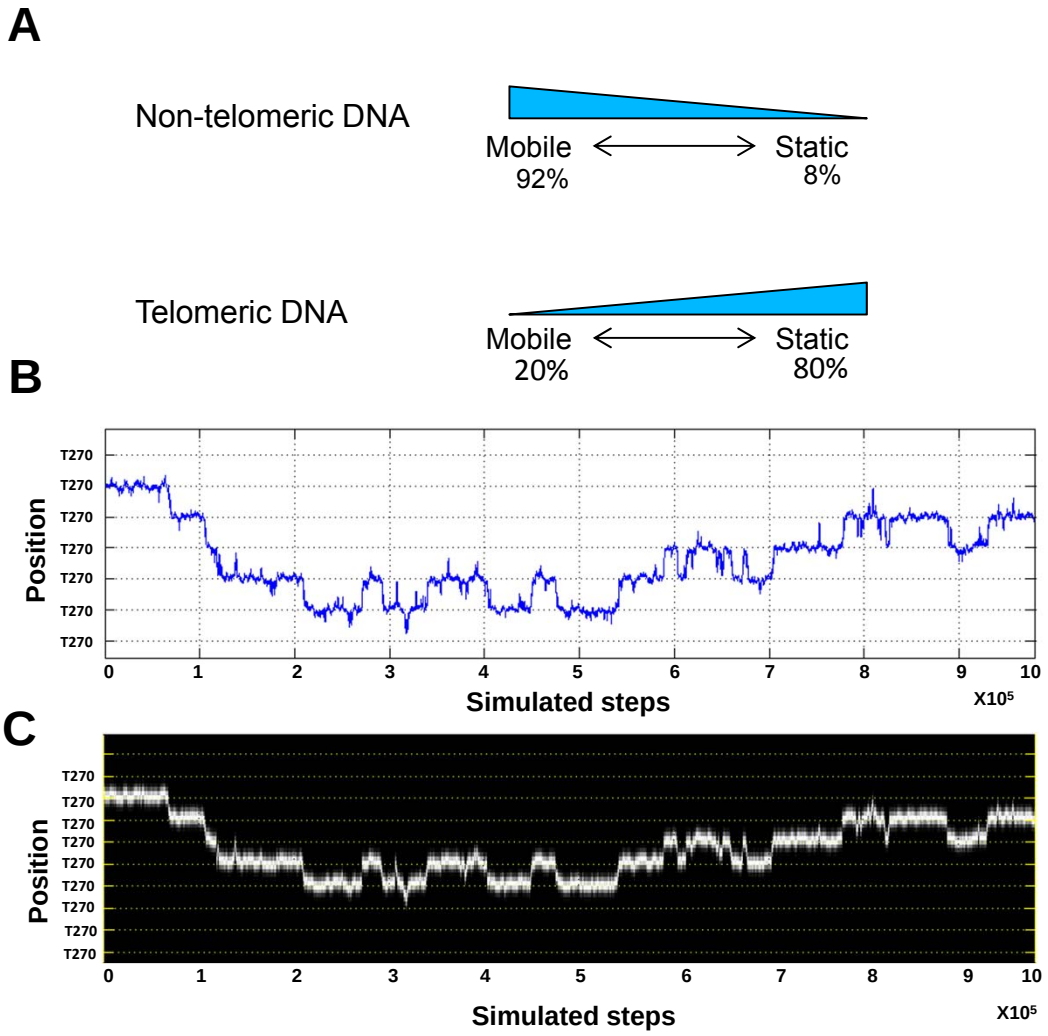
Supplementary Figure S8. Validation of the dual-color-labeling strategy for SA1-TRF1 complexes. (**A**) Formation of dual-color QD-labeled complexes on T270 tightropes depends on the presence of both TRF1 and SA1. In all experiments, ^{BT}tris-NTA, red Ab-QDs, and green SAv-QDs were present in flow cells. 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 (makers 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 ± 0.36 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 tightropes 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 $1x10^{-3}$ μm^2 s^{-1}. Since the equilibrium is rapid and dynamic even on telomeric DNA, SA1 continues to diffuse over telomeric regions, however much more slowly.