Supporting Information:

CTCF-Induced Circular DNA Complexes Observed by Atomic Force Microscopy

Matthew T. Mawhinney[1], Runcong Liu[1], Fang Lu[2], Jasna Maksimoska[2,3], Kevin Damico[2,3], Ronen Marmorstein[2,3,4], Paul M. Lieberman[2], and Brigita Urbanc[1,5]

[1]Department of Physics, Drexel University, Philadelphia, PA, USA

[2]The Wistar Institute, Philadelphia, PA, USA

[3]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[4]Department of Chemistry, University of Pennsylvania, Philadelphia, PA, USA

[5]Faculty of Mathematics and Physics, University of Ljubljana, Slovenia

WT DNA and MT DNA sequences

For AFM imaging, the 941 bp sequence from KSHV with three specific binding sites was used (WT DNA) alongside the mutated sequence without specific binding sites (MT DNA). The three specific binding sites in WT DNA are highlighted in yellow. The three specific binding sites were mutated in MT DNA as marked by lower case letters.

```
WT  GGATGGGATGGAGGGATTGGGGAGGAAGACGTGGTTACGGGGGGTAAGAGTGCCGGTGGA   60
MT  ............................................................   60

WT  GGTAAAGGTGTTGCGGGAGCGGGTGAAGGAATGGGAGCCACCGGTAAAGTAGGACTAGAC  120
MT  ............................................................  120

WT  ACAAATGCTGGCAGCCCGGATGTGAACACTGTGGGACTTCCTGGTATAGGCAAGGTGTGG  180
MT  ............................................................  180

WT  GGTCCACATTCCCGGCCGTCGACGGAGTCGGCGACATGCTTCCTTCGCGGTTGTAGATGT  240
MT  ............................................................  240

WT  AGGTCATCGCCAAGGTCACATCTTTCCGGAGACCTGTTTCGTTTCCTACAACTTCCTCTC  300
MT  ............................................................  300

WT  GTTAAGGGCGCGCCGGTGCTCCGTCCCGACCTCAGGCGCATTCCCGGGGGCGCCATCCTC  360
MT  ............................................................  360

WT  GGGAAATCTGGTCTGACAACCAAAGTAAAATTATGGAGGCGGTGGCAGTATATTCACATT  420
MT  ............................................................  420

WT  ATGCAATACCCGTAGTGACCACAAGGGGGAGCTCTCAGACAATTAAGCGGTTACACACAG  480
MT  ..................ttgt..ttttt.............................  480

WT  TAGCAGGCTGCAGTACCGCCCATGGCCACAGGATGGAGATCGCAGACACTGAAACGCTGA  540
MT  ........................ttgt.tt..tt........................  540

WT  AACACAGCATTAAGCTGCAATACCGCCGATGGCCACCAGATGGCACGCGCCGCCAGCAAA  600
MT  ................................ttgt..t..ttt................  600

WT  TTTAAGTCCTGGTGGCTCACCTGCCAGGTAAACAAGGTTAAAGTGGGTTTGCTGGCCTTG  660
MT  ............................................................  660

WT  CGTTGCCATGGATGCTACCTAGGCAAGTCCAGATATATAATCCGGGCGTGAGAAACAGAA  720
MT  ............................................................  720

WT  ACGGCCAATAACCCATGTTTTTCGAAAACCACCACACACCTTAACACAAATCATGTACAC  780
MT  ............................................................  780

WT  CTGGTATTACTATTTCCCACACATCTTATAGCATTTCAAAGATAAGGGTGCCTTACGGGC  840
MT  ............................................................  840
```

```
WT  CGCCCGAAACAAGTGGGCGGGCGCTACTCACTGTTTATAAGTCAGCCGGACCAAGCTGCT   900
MT  ..........................................................   900

WT  GCTCTTGGGGACGTGACTGCTTCGTGGCGCAGCTGCCTCCA   941
MT  ...........................................   941
```

Protein sequences

The sequence of the 11 ZF CTCF protein domain (residues 266-579) is:

```
266      FQCEL CSYTCPRRSN LDRHMKSHTD ERPHKCHLCG RAFRTVTLLR 310
311 NHLNTHTGTR PHKCPDCDMA FVTSGELVRH RRYKHTHEKP FKCSMCDYAS 360
361 VEVSKLKRHI RSHTGERPFQ CSLCSYASRD TYKLKRHMRT HSGEKPYECY 410
411 ICHARFTQSG TMKMHILQKH TENVAKFHCP HCDTVIARKS DLGVHLRKQH 460
461 SYIEQGKKCR YCDAVFHERY ALIQHQKSHK NEKRFKCDQC DYACRQERHM 510
511 IMHKRTHTGE KPYACSHCDK TFRQKQLLDM HFKRYHDPNF VPAAFVCSKC 560
561 GKTFTRRNTM ARHADNCAG                                  579
```

and the sequence of the 6-8 ZF CTCF protein domain (residues 402 to 494) is:

```
402  SGEKPYECY ICHARFTQSG TMKMHILQKH TENVAKFHCP HCDTVIARKS 450
451 DLGVHLRKQH SYIEQGKKCR YCDAVFHERY ALIQHQKSHK NEKR       494
```

Complementary oligonucleotides for EMSA experiments

The following complementary oligonucleotides with (wt) and without (mt) specific binding sites for CTCF were used:

```
wt (XqYq_CTCF-1951 Bam/HinDII)
oPL3689: GATCCTGCTGTGCCAGGGCGCCCCCTGCTGGCGACTAGGGCAACTA
oPL3660: AGCTTAGTTGCCCTAGTCGCCAGCAGGGGGCGCCCTGGCACAGCAG

mt (XqYq_CTCF-1951 BamHI/HindIII)
oPL4456: GATCCTGCTGTGCCAGAATACAAAATGCTAATAACTAGGGCAACTA
oPL4457: AGCTTAGTTGCCCTAGTTATTAGCATTTTGTATTCTGGCACAGCAG
```

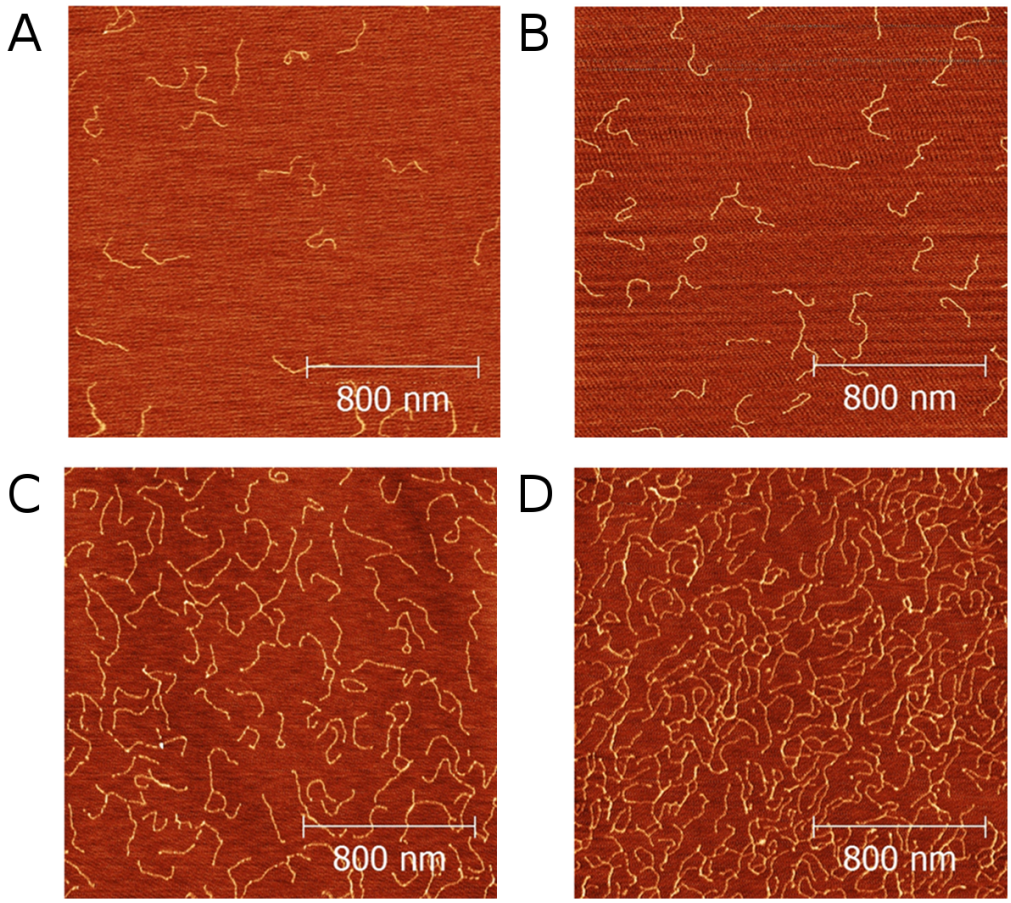

Figure S1: **AFM images of control DNA samples with a variable number density of DNA molecules.** (A-D) Control DNA conformations displaying variations in the number of DNA molecules that adhered to the mica surface. The number of DNA molecules on AFM images increases from (A) through (D).

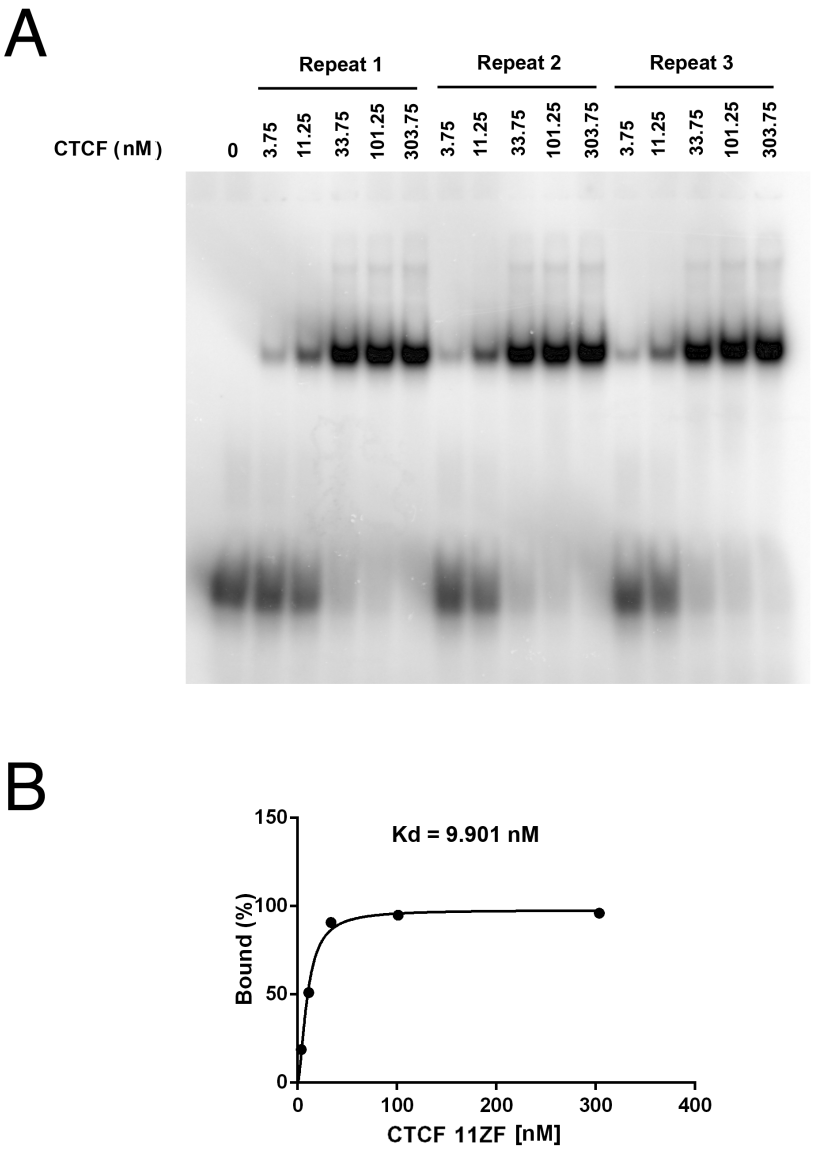

Figure S2: **EMSA of 11 ZF CTCF.** (A) EMSA analysis of 11 ZF CTCF at 0, 3.75, 11.25, 33.75, 101.25, and 303.75 nM incubated with CTCF wt radiolabeled DNA probes and (B) the corresponding binding curves. Error bars (within the black markers) were generated from three repeats. The dissociation constant of 11 ZF CTCF binding to the consensus CTCF binding sites was determined to be 9.901 nM. The experimental conditions for EMSA are described in *Materials and Methods*.

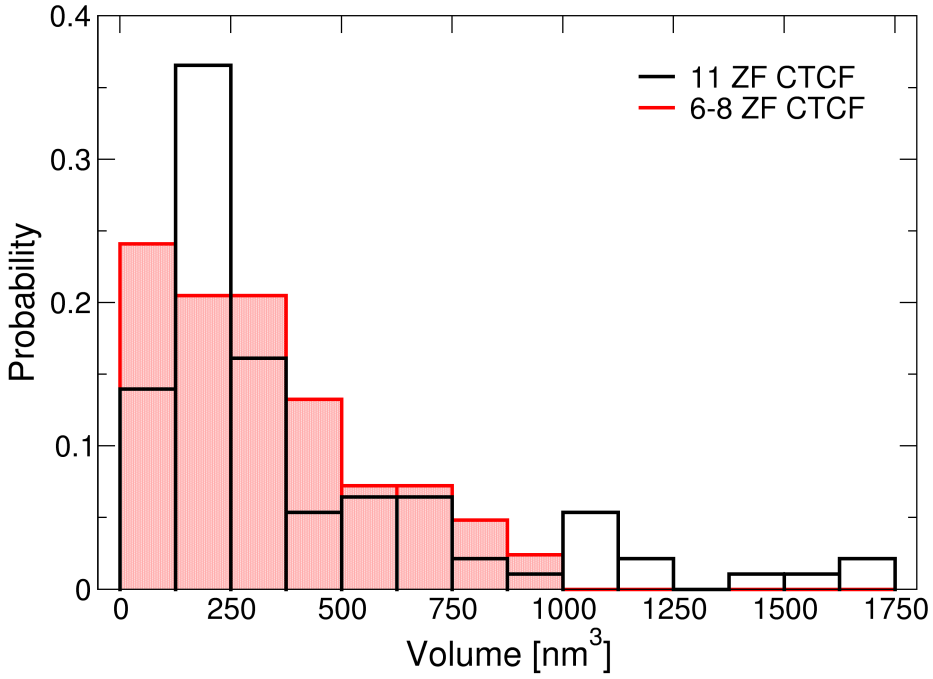

Figure S3: **Volume histograms of the two CTCF protein domains as measured by AFM.** Distributions of measured volumes for 11 ZF CTCF (black) and 6–8 ZF CTCF (red) obtained from AFM images of CTCF control samples. 20% and 13% of all volume measurements on 11 ZF CTCF and 6–8 ZF CTCF, respectively, correspond to molecular weights of dimers and larger multimers.

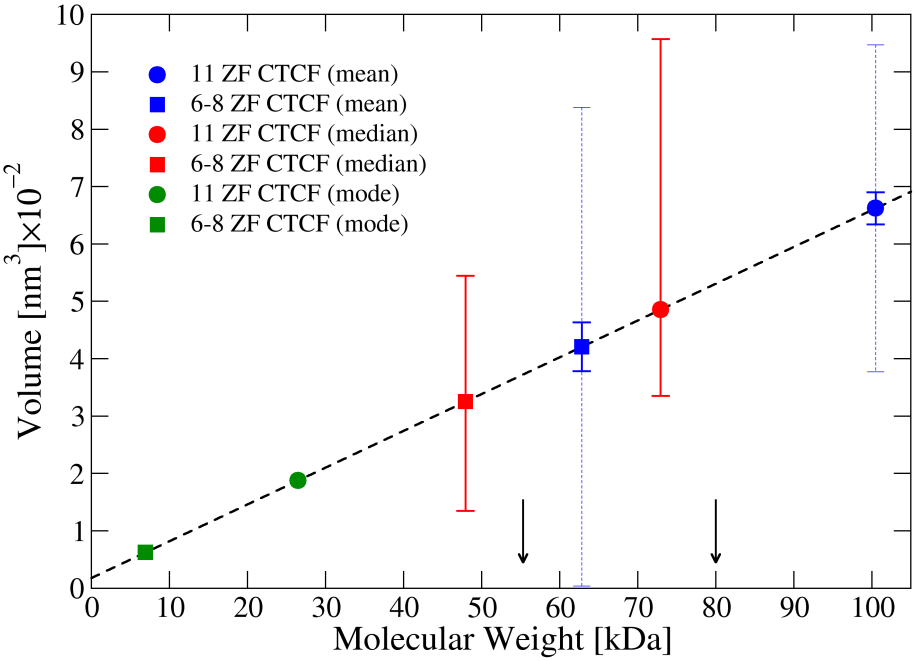

Figure S4: **Molecular weight estimation from volumetric AFM measurements.** The linear dependence of the mean AFM-derived volume on the molecular weight was determined for four protein standards (see *Protein volume standards* in *Materials and Methods*), resulting in $V = 6.41\ \text{nm}^3/\text{kDa} \times (MW) + 17.87\ \text{nm}^3$ with $R^2 = 0.97$ (black dashed line). Molecular weights of 11 ZF and 6–8 ZF CTCF were calculated by extrapolation of the mean, median, and mode of each of the two protein volume distributions. The mean volumes and SEM values of 662.08±28.06 nm^3 and 420.62±42.55 nm^3 resulted in molecular weights of 100.49±39.93 kDa and 62.82±25.69 kDa for 11 ZF CTCF (blue circle) and 6–8 ZF CTCF (blue square), respectively. The standard deviations of 285.12 nm^3 and 416.9 nm^3 are also shown in the blue dashed lines. Using the median volume value of 485.42 nm^3 with an interquartile range (IQR) of 621.97 nm^3 resulted in a molecular weight for 11 ZF CTCF (red circle) of 72.93 kDa with an upper and lower limits of 151.96 kDa and 35.80 kDa. Using the median value of 325.09 nm^3 with an IQR of 409.88 nm^3 resulted in a molecular weight for 6–8 ZF CTCF (red square) of 47.92 kDa with an upper and lower limit of 87.04 kDa and 12.70 kDa. Using the mode values of 187.5 nm^3 and 62.5 nm^3 resulted in molecular weights of 26.46 kDa and 6.91 kDa for 11 ZF CTCF (green circle) and 6–8 ZF CTCF (green square), respectively. The actual molecular weights of 11 ZF CTCF (right arrow) and 6–8 ZF CTCF (left arrow) are indicated by the black arrows.

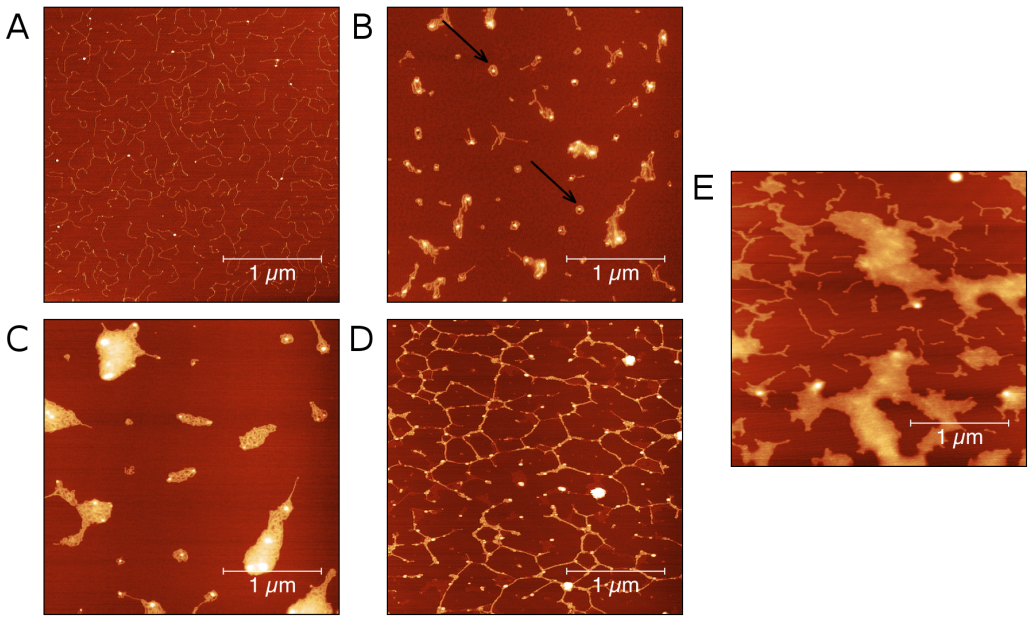

Figure S5: **AFM images of MT DNA morphologies.** Four distinct morphologies of MT DNA in the presence of 11 ZF CTCF observed by AFM: (A) relaxed MT DNA conformations, similar to those found in DNA control, (B) circular DNA/CTCF complexes (see black arrows pointing at representative complexes), (C) mesh-like structures with multiple protein signals, and (D) protein signal-rich DNA network structures spanning the entire image. (E) AFM image of mesh morphology observed in MT DNA control samples in the absence of 11 ZF CTCF.

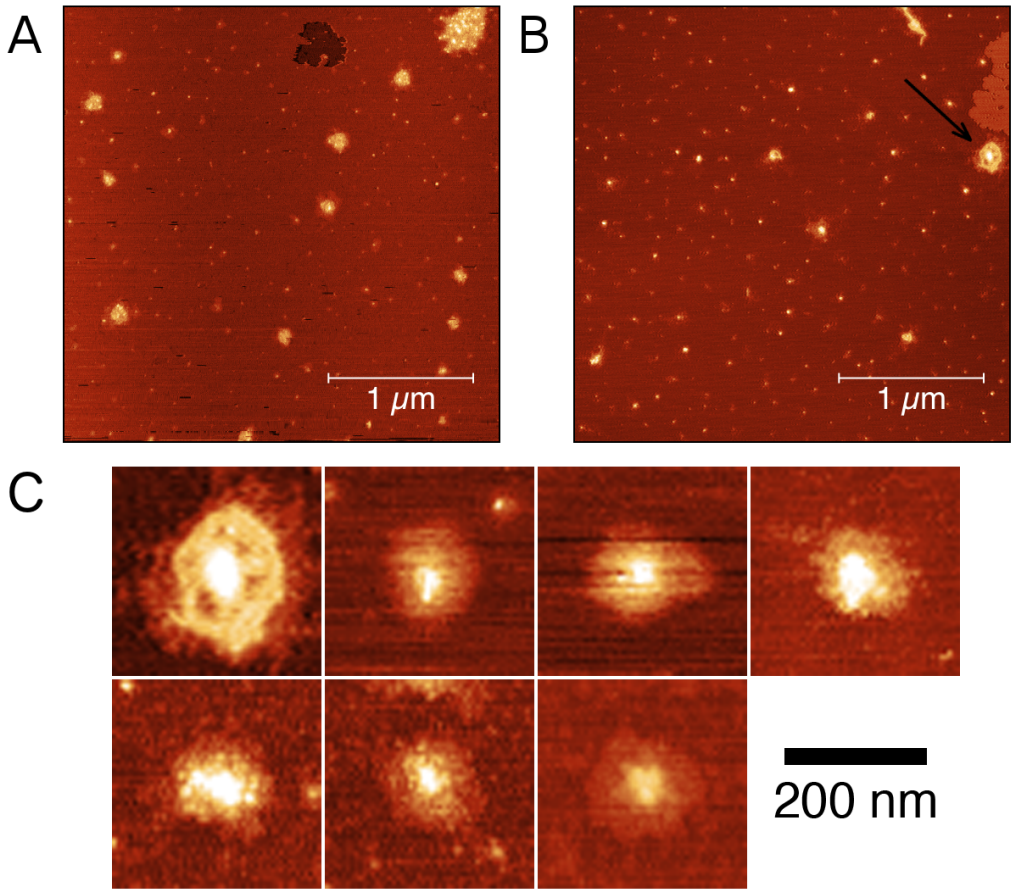

Figure S6: **AFM images of CTCF and DNA/CTCF samples in the 100 mM NaCl buffer.** AFM images of (A) a 11 ZF CTCF control sample, (B) a sample of WT DNA incubated with 11 ZF CTCF, where a black arrow points to a circular complex, and (C) several AFM images of circular DNA/CTCF complexes.