Conformational heterogeneity allows access to DNA in longer Transcription Activator-Like Effector (TALE) arrays.

Kathryn Geiger-Schuller[1,2], Jaba Mitra[3], Taekjip Ha[1,4,5,6], & Doug Barrick[1]*

[1]T.C. Jenkins Department of Biophysics and Program in Molecular Biophysics, Johns Hopkins University, 3400 N. Charles St, Baltimore MD 21218.

[2]Current address: Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA 02142.

[3]Materials Science and Engineering, University of Illinois Urbana-Champaign, Urbana IL 61801.

[4]Department of Physics, Center for the Physics of Living Cells and Institute for Genomic Biology, University of Illinois at Urbana Champaign, Urbana IL 61801.

[5]Department of Biophysics and Biophysical Chemistry and Department of Biomedical Engineering, Johns Hopkins University, Baltimore MD 21205.

[6]Howard Hughes Medical Institute, Baltimore, MD, 21205, USA

*Corresponding Author: barrick@jhu.edu, (410) 516-0409

Keywords: TALE repeat, Single-molecule biophysics, FRET, Functional instability, Deterministic modeling

Abstract

Transcription activator-like effectors (TALEs) bind DNA through an array of tandem 34-residue repeats. Here, we examine the kinetics of DNA binding for a set of TALE arrays with identical repeats of varying length (and repeat number) using single molecule microscopy. Using a new deterministic modeling approach to test kinetic models consistent with data, we find evidence for conformational heterogeneity in both the free- and DNA-bound TALE arrays. We connect these results with previous work demonstrating populations of partly folded TALE states. TALEs forming less than one superhelical turn around DNA access partly folded open states that inhibit DNA binding, whereas TALEs forming more than one turn access partly folded open states that facilitate DNA binding. Overall, we find that increasing repeat number results in significantly slower interconversion between and among DNA-free and DNA-bound states. These findings highlight the role conformational dynamics can play in facilitating the assembly of large complexes.

38 **Introduction**

39 Transcription activator-like effectors (TALEs) are bacterial proteins containing a
40 domain of tandem DNA-binding repeats as well as a eukaryotic transcriptional activation
41 domain[1,2]. The repeat domain binds double stranded DNA with a register of one repeat
42 per base pair. Specificity is determined by the sequence identity at positions twelve and
43 thirteen in each TALE repeat, which are referred to as repeat variable diresidues
44 (RVDs)[3–5]. This specificity code has enabled design of TALE-based tools for
45 transcriptional control[6–10], DNA modifications[11], in-cell microscopy[12,13], and genome
46 editing (TALENs)[14,15].

47 TALE repeat domains wrap around DNA in a continuous superhelix of 11.5 TALE
48 repeats per turn[16,17]. Because TALEs contain on average 17.5 repeats[18], most form
49 over 1.5 full turns around DNA. Many proteins that form rings around DNA require
50 energy in the form of ATP to open or close around DNA, yet TALEs are capable of
51 wrapping around DNA without energy from nucleotide triphosphate hydrolysis. One
52 possibility is that TALEs bind DNA through an energetically accessible open
53 conformation. Consistent with this possibility, we previously demonstrated that TALE
54 arrays can populate partly folded or broken states[19].

55 Consensus TALEs (cTALEs) are homopolymeric arrays composed of the most
56 commonly observed residue at each of the 34 positions of the repeat[19]. In addition to
57 simplifying analysis of folding and conformational heterogeneity, the consensus
58 approach simplifies analysis of DNA binding, eliminating contributions from sequence
59 heterogeneity and providing an easy means of site-specific labeling.

60 Here we characterize DNA binding kinetics of cTALEs using total internal
61 reflection fluorescence single-molecule microscopy. We find that consensus TALE
62 arrays bind to DNA reversibly, with high affinity. Analysis of the dwell-times of the on-
63 and off-states reveals multiphasic binding and unbinding kinetics, suggesting
64 conformational heterogeneity in both the free and DNA bound state. Here we describe a
65 deterministic optimization analysis that supports such a model, and provides rate
66 constants for both conformational changes and binding. Comparing the dynamics
67 observed here to previously characterized local unfolding suggests that locally unfolded
68 states inhibit binding of short cTALE arrays (less than one full superhelical turn around
69 DNA), whereas they promote binding of long arrays (more than 1 full superhelical turn).
70

71 **Results**

72 **cTALE design**

73 Consensus TALE (cTALE) repeat sequence was design previously described[19].
74 To avoid self- association of cTALE arrays, we fused arrays to a conserved N-terminal
75 extension of the PthXo1 gene. Although the sequence of this domain differs from a
76 TALE repeat sequence, the structure of this domain closely mimics four TALE
77 repeats[17,20] and is required for full transcriptional activation[20]. In this study, all repeat
78 arrays contain this solubilizing N-terminal domain.
79

80 **cTALE local instability promotes population of partly folded states**

81 Figure 1A depicts types of partly folded states of a generic repeat protein. In the
82 fully folded state, all repeats are folded, and all interfaces are intact. In the end-frayed
83 states, one or more terminal repeats are unfolded and all interfaces, except the
84 interface(s) between the unfolded and adjacent folded repeat(s), are intact. In the
85 internally unfolded states, a central repeat is unfolded and all interfaces, except the
86 interfaces involving the unfolded repeats, are intact. In the interfacially ruptured state, all
87 repeats are folded but one interface is disrupted due to local structural distortion.
88 Figure 1B shows calculated free energy differences between various partly
89 folded states and the fully folded repeat array for two different RVDs (NS and HD) in an
90 otherwise identical consensus sequence background, using the intrinsic and interfacial
91 engeries we determined previously[19]. The distribution of partly folded states is
92 calculated for different 20-repeat arrays containing two types of TALE arrays (with the
93 NS RVD in red and with the HD RVD in blue) as well as consensus ankyrin arrays
94 (cAnk in black). For cTALE arrays, end frayed states are within a few *RT*, internally
95 unfolded states are highest in energy, and interfacially ruptured states fall energetically
96 between end frayed and internally unfolded states.
97 Changing the RVD sequence affects the distribution of these partly folded states:
98 HD repeat containing arrays are more likely to internally unfold or interfacially rupture
99 than NS repeat containing arrays. Both types of cTALEs are more likely to populate
100 many of these partly folded states than cAnk is to populate even the lowest energy
101 partly folded state, the end frayed state. Thus, compared to ankyrin repeats, cTALEs

102 are locally unstable, meaning they are likely to form partly folded states. As these
103 states disrupt the superhelix, they may facilitate DNA binding.
104
105 **Single-molecule studies of cTALE binding to DNA**
106 To ask if cTALE local instability is relevant for DNA binding kinetics, DNA binding
107 trajectories were measured using single molecule total internal reflection fluorescence
108 (smTIRF). Figure 2A shows a schematic of the smTIRF experiments performed to
109 measure DNA binding. For site-specific cTALE labeling, R30 is mutated to cysteine in a
110 single repeat. Position 30 is frequently a cysteine in naturally occurring TALEs (in earlier
111 folding studies, arginine was chosen in the consensus sequence to avoid disulfide
112 formation)[19]. This cysteine was Cy3 (FRET donor)-labelled using maleimide chemistry,
113 and was attached to biotinylated slides via the C-terminal His_6 tag and α-Penta•His
114 antibodies. At salt concentrations below 300 mM NaCl, cTALEs aggregate. Because
115 DNA binding is weak at high salt concentrations, measuring binding kinetics in bulk at
116 high salt is not possible. However, tethering cTALEs to the quartz slide at high salt
117 prevents self-association, even at the low salt concentrations required to study DNA
118 binding kinetics. A histogram of $NcTALE_8$ (8 NS-type repeats and the N-terminal
119 domain) labeled via a cysteine in the first repeat shows a single peak at zero FRET
120 efficiency, as expected for donor-only constructs (Figure 2B).
121 To test for DNA binding to tethered cTALE constructs, we added Cy5 (FRET
122 acceptor)-labeled 15 bp-long DNA ($Cy5.A_{15}/T_{15}$) to tethered $NcTALE_8$. This results in a
123 new peak at a FRET efficiency of 0.45, indicating that DNA binds directly to cTALE

124 arrays. As DNA concentration in solution is increased, the peak at 0.45 FRET efficiency
125 increases in population (Figure 2C-D), suggesting a measurable equilibrium between
126 free and bound DNA rather than saturation or irreversible binding. In support of this,
127 single molecule time trajectories show interconversion between bound and unbound
128 states, providing access to rates of binding and dissociation.

129 As expected for reversible complex formation, the peak at 0.45 FRET efficiency
130 can be competed away by adding mixtures of labeled and unlabeled DNA to pre-formed
131 cTALE-labelled DNA complexes (schematic shown in Figure 2E; pre-formed complex
132 shown in Figure 2F, competition data shown in Figure 2G-H). Challenging pre-formed
133 complexes with a mixture of 5 nM unlabeled DNA and 15 nM labeled DNA results in a
134 slight decrease in the population of the peak at 0.45 FRET (compare Figures 2F and
135 2G). Challenging with a mixture of 50 nM unlabeled DNA and 15 nM labeled DNA
136 further decreases the peak at 0.45 FRET (Figure 2H).

137

138 **cTALE arrays display multiphasic DNA-binding kinetics.**

139 In addition to the short smTIRF movies used to generate smFRET histograms
140 from many molecules, long movies were also collected to examine the extended
141 transitions of individual molecules between the low- and high-FRET (0 and 0.45) (Figure
142 3A-B). A transition from low to high FRET (0 to 0.45) indicates that the acceptor
143 fluorophore on DNA moved close enough to the donor on the protein for FRET and is
144 likely a binding event. A transition from high to low FRET (0.45 to 0.0) indicates the
145 acceptor fluorophore on DNA moved too far away from the donor on the protein for

146 FRET and is likely an unbinding event. Low-FRET states show low colocalization with
147 signal upon direct excitation of the acceptor, confirming that high-FRET states are DNA-
148 bound states and low-FRET states are DNA-free states (Figure S1). These long single
149 molecule traces show both long- and short-lived low- and high-FRET states, indicating
150 that kinetics are multi-phasic (Figure 3A-B). Binding events (transitions from low to high
151 FRET) become more frequent as bulk DNA concentration increases (compare
152 representative traces at 1 nM dsDNA to 15 nM dsDNA; Figure 3A and Figure 3B).
153 Cumulative distributions generated from dwell times in the low FRET state at a given
154 DNA concentration are best-fit by a double-exponential decay, indicating a minimum of
155 two kinetic phases associated with binding events (Figure 3C). Cumulative distributions
156 generated from dwell times in the high FRET state are best-fit by a double-exponential
157 decay, indicating that there are a minimum of two kinetic phases for unbinding as well
158 (Figure 3D).

159 The rate constant for the fast phase in DNA binding shows a linear increase with
160 DNA concentration (Figure 3E), indicating that this step involves an associative binding
161 mechanism. The slope of these rate constants as a function of DNA concentration gives
162 a bimolecular rate constant of 5.9×10^{8} $nM^{-1}s^{-1}$, close to the diffusion limit. The rate
163 constant for the slower phase (0.59 s^{-1}) is independent of DNA concentration indicating
164 a unimolecular isomerization mechanism (Figure 3E).

165 In contrast, neither of the two fitted rate constants for transitions from high to low
166 FRET (0.45 to 0.0; unbinding events) depends on DNA concentration, suggesting that

167 unbinding involves two (or more) unimolecular processes (Figure 3F). The rate
168 constants of these two phases are 1.2 s^{-1} and 0.13 s^{-1} respectively.
169 To rule out kinetic contributions of TALEs threading axially onto the ends of short
170 DNAs, binding kinetics were measured with capped double-helical DNA sites. Capped
171 DNA was generated by forming 5'digoxygenin-A_5-Cy5-A_{15} duplexed with 5'-digoxygenin-
172 T_{26} and adding a three-fold molar excess of anti-Digoxygenin. Low and high FRET dwell
173 time cumulative distributions generated from capped DNA-binding kinetics are bi-
174 phasic, similar to distributions from uncapped DNA (Figure S2). The DNA concentration-
175 independent rate constant for binding is roughly the same for capped DNA as for
176 uncapped DNA (compare $FRET_{L\rightarrow H}$ red and blue triangles in Figure S2), as are the
177 dissociation rate constants (compare $FRET_{H\rightarrow L}$ red and blue triangles in as well as
178 $FRET_{H\rightarrow L}$ red and blue circles in Figure S2). The rate constant for bimolecular binding of
179 capped DNA decreases compared to that for uncapped DNA (compare $FRET_{L\rightarrow H}$ red
180 and blue circles in Figure S2), which is consistent with the expected decrease in the
181 rate of diffusion of the larger capped DNA. To assess the effect of molecular weight
182 increase on diffusion of capped versus uncapped DNA, Sednterp[21], a program
183 commonly used to estimate sedimentation and diffusion properties of biomolecules, was
184 used to estimate maximum diffusion coefficients. Including the two antibodies bound on
185 the ends of capped DNA (320 kDa total) gives an estimated diffusion coefficient of
186 4.7×10^{-7} cm^2s^{-1}, which is much lower than the estimated diffusion coefficient of the
187 uncapped DNA (1.5×10^{-6} cm^2s^{-1}). This ~3.6-fold decrease in the diffusion constant for

188 capped DNA is similar to the 6.7-fold decrease in the bimolecular rate constant for
189 binding of capped DNA (Figure S2).
190
191 **Longer cTALEs have slower DNA binding and unbinding kinetics**
192 To examine how increasing the length of the cTALE array influences DNA
193 binding, we generated Cy3-labelled constructs with 16 and 12 cTALE repeats, and
194 measured binding to a longer Cy5-labelled DNA (A_{23}/T_{23}). Because we did not observe
195 FRET for these longer constructs, a fluorescence colocalization microscopy protocol
196 was used instead (Figure S3). In this protocol, Cy3 was first imaged for ten camera
197 frames (1017.5 msec total) to identify positions of single TALE molecules. Then a long
198 time series of fluorescence images of Cy5 signal were collected through directly exciting
199 Cy5 on the DNA, and time trajectories of Cy5 signal were generated from the initially
200 identified single TALE positions.
201 Increasing the number of cTALE repeats from 8 to 12 and 16 dramatically affects
202 DNA binding kinetics. Long movies collected over a range of DNA concentrations show
203 short- and long-lived Cy5 signal on and off states, indicating a level of kinetic
204 heterogeneity similar to $NcTALE_8$ (Figure S3). Single molecule traces were analyzed
205 using a thresholding filter (see Materials and Methods and Figure S3) to identify states
206 and dwell times. Cumulative distributions were generated from dwell times at low Cy5
207 signal (unbound states, with lifetimes representing binding kinetics), and at high Cy5
208 signal (bound states, with dwell times representing unbinding kinetics). As with the eight
209 repeat constructs, unbound cumulative distributions for these longer TALE arrays are

210 best-fit by double exponential decays, particularly at high DNA concentrations (compare
211 the cumulative distribution at low DNA concentration, Figure S4A, to cumulative
212 distribution at 5 nM DNA, Figure S4B). Bound cumulative distributions for longer TALE
213 arrays are best-fit by double exponential decays (Figure S4C-D). All apparent rate
214 constants are much smaller for $NcTALE_{16}$ and $NcTALE_{12}$ (green/black circles and
215 triangles, Figure 4A-B) compared to $NcTALE_{8}$, indicating that binding and unbinding is
216 impeded by increasing the length of the binding surface between cTALEs and their
217 cognate DNA (Figure 4C). To address whether differences in binding kinetics are
218 related to experimental differences between colocalization and FRET assays,
219 alternating laser experiments were performed by switching between FRET and
220 colocalization detection (every 5 frames) within single molecule trajectories (Figure S1).
221 Changes in FRET and colocalization signals occurred simultaneously according to
222 single molecule time traces, showing that differences in binding and unbinding kinetics
223 of short and longer cTALEs are not due to differences in colocalization and FRET
224 assays (Figure S1).
225
226 **A deterministic approach to modeling cTALE-DNA binding kinetics.**
227 To determine how the kinetic changes above are partitioned into underlying
228 kinetic steps in binding, we fitted various kinetic models to the cumulative distributions
229 for binding and unbinding. In addition to providing information about the mechanism of
230 binding, this approach allows us to estimate the underlying microscopic rate constants
231 and compare them for different constructs. This approach is generally applicable to

232 studies of complex single molecule kinetics. Numerical integration was used to calculate
233 the relative population of cTALE states as a function of time (Figures 5A-C and 5G-H),
234 given a binding mechanism, an associated set of rate laws, and a set of initial
235 conditions. Cumulative distributions of unbound dwell times represent the distribution of
236 times single molecules spent in the unbound state before transitioning into the bound
237 state, allowing us to split the kinetic scheme when fitting to single-molecule dwell times.

238 Among the various models tested, the model that is most consistent with the data
239 has two unbound DNA-free states and two DNA-bound states. This is consistent with
240 alternating laser experiments showing that DNA is only colocalized when cTALEs are in
241 the high FRET state (Figure S1). This four-state model includes a TALE isomerization
242 step in the absence of DNA from a DNA-binding incompetent conformation (which we
243 refer to as TALE) to DNA-binding competent conformation (which we refer to as TALE*).
244 The DNA-binding competent TALE* conformer binds and unbinds DNA (called TALE*
245 when DNA free and TALE*~DNA when DNA-bound). Before unbinding, a fraction of
246 TALE*~DNA isomerizes to a longer-lived DNA-bound state called TALE‡~DNA.

247 Based on this mechanism, the rate laws for binding are given in equations 1a -
248 1d.

249 $$\frac{d[TALE]}{dt} = -k_1[TALE] + k_{-1}[TALE^*] \quad (1a)$$

250 $$\frac{d[TALE^*]}{dt} = k_1[TALE] - k_{-1}[TALE^*] - k_2[TALE^*][DNA] \quad (1b)$$

251 $$\frac{d[TALE^* \sim DNA]}{dt} = k_2[TALE^*][DNA] \quad (1c)$$

252 $$K_{eq,DNA-free} = \frac{k_1}{k_{-1}} \quad (1d)$$

253 Since the single-molecule dwell-time histograms of the unbound states are insensitive
254 to the isomerization after DNA binding, the equation describing the time evolution of the
255 long-lived bound state ($TALE^{\ddagger}$~DNA) is not relevant to our analysis of unbound-state
256 lifetimes.

257 To determine microscopic rate constants k_1, k_{-1}, and k_2, equations 1a-1c were
258 numerically integrated in Matlab, and the fraction of TALE*~DNA as a function of time
259 was fitted to the low-FRET cumulative distributions ($NcTALE_8$; Figure 5D-E) or to the no
260 colocalization cumulative distributions ($NcTALE_{16}$; Figure 5J-K). Microscopic rate
261 constants were adjusted to reduce sum of the squared residuals between the
262 concentration of TALE*~DNA (the direct product of binding) as a function of time and
263 single-molecule cumulative distributions. In both cases, cumulative distributions at
264 different bulk DNA concentrations were fitted globally. Initial fractions of TALE and
265 TALE*~DNA were set to zero, and the initial fraction of TALE* was set to one.
266 Confidence intervals (CI) were estimated by bootstrapping (Table 1; mean and 68% CI
267 from 2000 or 8000 bootstrap iterations).

268 Rate laws for dissociation are given in equations 2a - 2d

269 $$\frac{d[TALE^* \sim DNA]}{dt} = -k_{-2}[TALE^* \sim DNA] - k_3[TALE^* \sim DNA] + k_{-3}[TALE^{\ddagger} \sim DNA] \quad (2a)$$

270 $$\frac{d[TALE^{\ddagger} \sim DNA]}{dt} = k_3[TALE^* \sim DNA] - k_{-3}[TALE^{\ddagger} \sim DNA] \quad (2b)$$

271 $$\frac{d[TALE^*]}{dt} = k_{-2}[TALE^* \sim DNA] \quad (2c)$$

272 $$K_{eq,DNA-bound} = \frac{k_3}{k_{-3}} \quad (2d)$$

273 As with the system of equations above (1a-d), the equation describing the time
274 evolution of the binding-incompetent free state (TALE) is not relevant to our analysis of
275 bound-state lifetimes.

276 To determine microscopic rate constants k_{-2}, k_{-3}, and k_3, equations 2a-2c were
277 numerically integrated in Matlab, and the fraction of TALE* as a function of time was
278 fitted to the high-FRET cumulative distributions ($NcTALE_8$; Figure 5F) or to the low
279 colocalization cumulative distributions ($NcTALE_{16}$; Figure 5L). Microscopic rate
280 constants were adjusted to reduce sum of the squared residuals between the
281 concentration of TALE* (the direct product of dissociation) as a function of time and
282 single-molecule cumulative distributions. In both cases, cumulative distributions at
283 different bulk DNA concentrations were fitted globally. The initial fraction of
284 TALE*~DNA conformer was set at one; all other initial fractions were set to zero.
285 Confidence intervals were estimated by bootstrapping (Table 1; mean and 68% CI from
286 2000 iterations).

287 Fitted curves reproduce the experimental cumulative distributions for binding and
288 unbinding (Figure 5), both for the short and long cTALE arrays, with reasonably small
289 residuals, over a range of DNA concentrations. Generally, fitted rate constants have
290 confidence intervals of 10% or smaller (Table 1).

291 Comparison of microscopic rate constants for 8, 12, and 16 repeats show some
292 significant differences. The bimolecular microscopic binding rate constant, k_2, is slightly
293 larger for 8 repeats than for 12 and 16 repeats (1.1, 0.31, and 0.39 $nM^{-1}s^{-1}$ for 8, 12, and

294 16 repeats respectively). However, microscopic unbinding rate constant, k_{-2}, is higher
295 for 8 repeat cTALEs (0.66 s^{-1} for $NcTALE_8$ versus 0.13 s^{-1} for $NcTALE_{12}$ and 0.299 s^{-1} for
296 $NcTALE_{16}$). Also, bound state isomerization (interconversion between TALE*~DNA and
297 $TALE^{‡}$~DNA) is 5-10 times slower for 16 and 12 repeat cTALEs than 8 repeat cTALEs.
298 The value of $K_{eq,\ DNA\text{-}free}$, which is a measure of the equilibrium proportion of the
299 unbound TALE that is DNA-binding competent (TALE*) to that which is binding-
300 incompetent (TALE), is larger for cTALEs with 8 repeats ($K_{eq,\ DNA\text{-}free}$ = 1.32) than for
301 cTALEs with 12 and 16 repeats ($K_{eq,\ DNA\text{-}free}$ = 0.11 and $K_{eq,\ DNA\text{-}free}$ = 0.61, respectively).
302

303 **Discussion**

304 By measuring DNA-binding kinetics of cTALE arrays that form 0.7, 1, and 1.4
305 superhelical turns, we probe the functional relevance of locally unfolded TALE states.
306 We describe a novel method to glean mechanistic details from complex single molecule
307 kinetics. In our simplified cTALE system, we find conformational heterogeneity in both
308 DNA free and DNA-bound states. We find that association is slowed in arrays
309 containing one full turn of repeats or more. Because most natural and designed TALEs
310 contain more than a full turn of repeats, this finding has important implications for design
311 of high affinity TALE molecules, suggesting that placement of destabilized repeats at
312 specific positions may increase activity.
313

314 **cTALEs containing NS RVD bind DNA with high affinity**

315 NS is an uncommon RVD in natural TALEs. Previous reports suggest that NS is
316 fairly nonspecific, but may bind with higher affinity than other common RVDs (NG, NI,
317 NN, and HD)[5]. Our fitted rate constants can be used to calculate the apparent K_d (K_{app})
318 calculated from using equation as follows:

319
$$\begin{aligned} K_{app} &= \frac{[TALE-DNA+TALE^{*}-DNA]}{[DNA][TALE+TALE^{*}]} \\ &= \frac{K_{eq,DNA-free}K_2 + K_{eq,DNA-free}K_2K_{eq,DNA-bound}}{1+K_{eq,DNA-free}} \\ &= \frac{\frac{k_1}{k_{-1}} \times \frac{k_2}{k_{-2}}(1+\frac{k_3}{k_{-3}})}{1+\frac{k_1}{k_{-1}}} \end{aligned} \quad (3)$$

320 where $K_2 = k_2/k_{-2}$.

321

322 Using fitted rate constants from Table 1 in the final equality in equation 3 gives values
323 for K_{app} of 2.5 nM for the 8 repeat cTALE array, 0.5 nM for the 12 repeat cTALE array,
324 and 1.0 nM for the 16 repeat cTALE array. Increasing the number of repeats has a
325 modest affect on the apparent K_d due to the increased population of binding
326 incompetent DNA free TALE in the 12, and 16 repeat arrays. This affinity change is
327 small compared to a previous report studying length dependence on affinity of designed
328 TALEs (dTALEs) showing the K_d of a dTALE decreased by a factor of two with the
329 addition of only 1.5 repeats[22].

330 TALEs are believed to read out sequence information from one strand[3]. Due to
331 the asymmetry of our DNA sequences (poly-dA base-paired with poly-dT), in principle,
332 the FRET efficiency contains information on the binding orientation (and thus strand

333 preference). However, based on the crystal structure of the DNA-bound state of TAL-
334 effector PthXo1[17], we estimate that the distance between the donor site of $NcTALE_8$
335 (repeat 1) to the 5' acceptor site on the DNA (Cy5-A_{15}/T_{15}) should be similar for both the
336 dA-sense or dT-sense orientations (Figure S5A). Thus, the FRET data does not
337 discriminate between the two modes of binding for the eight-repeat construct. However,
338 for the 16 repeat NS RVD cTALE arrays, the PthXo1 model suggests very different
339 distances (25 Å versus 73 Å for the dT-sense or dA-sense respectively, Figure S5B)
340 between the donor site (TALE repeat 14) and the acceptor site (5' Cy5-A_{23}/T_{23}). To
341 restrict the number of binding positions available to longer cTALE arrays, the 23 base
342 pair DNA used for $NcTALE_{16}$ measurements (as well as DNA depicted in Figure S5B)
343 has the same number of additional base pairs as repeats (8 additional repeats and 8
344 additional base pairs) compared to the 15 base pair DNA used for $NcTALE_8$
345 measurements (as well as DNA depicted in Figure S5A). While we limited the number of
346 available binding positions, it may be possible for cTALEs to slide along DNA. However,
347 taking into account the four repeat N-terminal capping domain, there are only three
348 available base pairs in the bound complex. Thus we don't expect the distance
349 measurements to change by more than 10Å (~3 base pairs) if sliding occurs. The
350 observation that there is colocalization but no measurable FRET when $NcTALE_{16}$ is
351 bound to DNA suggests that cTALEs containing the NS RVD prefer adenine (the dA-
352 sense mode) compared with thymine bases, consistent with previous reports[3].
353

354 **Conformational heterogeneity in the unbound state could be the result of local**
355 **unfolding**

356 The cumulative distributions of dwell-times in Figure 3 provide clear evidence for
357 conformational heterogeneity in both the free and DNA-bound cTALEs. Although the
358 deterministic modeling supports such heterogeneity, puts it in the framework of a
359 molecular model, and provides a means to determine the microscopic rate and
360 equilibrium constants, such analysis provides little information about the structural
361 nature of TALE conformational heterogeneity.

362 Figure 6 shows a model of cTALE conformational change consistent with DNA
363 binding kinetics. In this model there are four TALE states. DNA-free cTALEs comprise
364 both incompetent and binding competent states. DNA-bound cTALES comprise
365 encounter and locked complexes with DNA. For 8 repeat cTALE arrays, the DNA-
366 binding competent state is more highly populated than the DNA-binding incompetent
367 state. In this reaction scheme, the DNA-binding incompetent state can be regarded as
368 an off-pathway conformation that inhibits DNA binding (Figure 6A).

369 Because the 8 repeat cTALE array does not form multiple turns of a superhelix,
370 unfolding to bind DNA is not required. In the model in Figure 6, the binding competent
371 state is the fully folded conformation, whereas the binding incompetent state includes
372 partly folded conformations. Consistent with this interpretation, increasing populations of
373 partly folded states through addition of 1M urea and through entropy enhancing
374 mutations decreases apparent binding rates of 8 repeat cTALEs (Figure S6). This is

375 also consistent with a partly folded DNA-binding incompetent state in shorter cTALE
376 arrays.
377 For 12 and 16 repeat cTALE arrays, the DNA-binding incompetent state is more
378 highly populated than the DNA-binding competent state. In the model in Figure 6, the
379 DNA-binding competent state is a high-energy conformation required for DNA binding
380 (Figure 6B-C). Because 12 and 16 repeat cTALEs are expected to form 1 and 1.4 turns
381 (excluding the N-terminal domain), we hypothesize that the binding competent state
382 includes some partly folded states that allow access to DNA. Not all partly folded states
383 open the array to access DNA; therefore, the binding incompetent state includes some
384 nonproductive partly folded states in addition to the fully folded state.
385 In arrays containing 12 or more cTALEs, the binding competent and binding
386 incompetent states likely include mixtures of many specific partly folded states. Because
387 the types of partly folded states are unknown, connecting equilibria between binding
388 competent and binding incompetent states to calculated partly folded equilibria (using
389 folding free energies similar to Figure 1) is challenging. Future work towards
390 understanding the structural characteristics of the binding competent state in TALE
391 arrays of one or more turns would inform which partly folded states to include in the
392 calculation, making this comparison meaningful.
393
394 **Conformational heterogeneity in the bound state**
395 Previous reports show that TALEs have multiple diffusional modes when
396 searching nonspecific DNA[23]. Our work suggests that cTALEs have multiple binding

397 modes (encounter and locked states in Figure 6) indicating cTALEs undergo a
398 conformational change also when bound to specific DNA sequences. Table 1 shows
399 that microscopic rate constants for transition into and out of longer lived locked bound
400 states become much slower in 12 and 16 repeat cTALEs compared with 8 repeat
401 cTALEs (k_{-3} and k_3). These rate constants decrease much more than the microscopic
402 unbinding rate constant (the k_{-2} values are 0.66 s^{-1}, 0.13 s^{-1}, and 0.299 s^{-1} for $NcTALE_8$,
403 $NcTALE_{12}$, and $NcTALE_{16}$ respectively) suggesting a large conformational change that
404 depends on the number of repeats. Although the model does not provide information on
405 structure of this conformational change, it is possible this conformational change
406 involves a slinky motion to decrease helical rise. Consistent with this hypothesis, crystal
407 structures of TALEs in the free and DNA-bound state show 11.5 repeats per turn in both
408 states, but the helical rise decreases upon binding[16]. Another possibility involves
409 specific interaction with RVDs and bases in the major groove of DNA. Crystal structures
410 show little deformation of DNA structure, so bending of DNA seems unlikely. While we
411 can only hypothesize about the structural nature of the conformational changes,
412 deterministic simulations show that cTALEs bind DNA through short encounters which
413 occasionally become long-lived locked conformations (Figure 6). Taken together, these
414 findings indicate that functional stability plays a crucial role in cTALE DNA binding, and
415 demonstrate the importance of conformational dynamics in complex assembly.
416

417 **Materials and Methods**

418 **Cloning, expression, purification, and labeling**

419 Consensus TALE repeat constructs were cloned with C-terminal His_6 tags via an
420 in-house version of Golden Gate cloning[24]. TALE constructs were grown in BL21(T1R)
421 cells at 37°C to an OD of 0.6-0.8 and induced with 1 mM IPTG. Following cell pelleting
422 and lysis, proteins were purified by resuspending the insoluble material in 6M urea, 300
423 mM NaCl, 0.5 mM TCEP, and 10 mM $NaPO_4$ pH 7.4. Constructs were loaded onto a Ni-
424 NTA column. Protein was eluted using 250 mM imidazole and refolded during buffer
425 exchange into 300 mM NaCl, 30% glycerol, 0.5 mM TCEP, and 10 mM $NaPO_4$ pH 7.4.
426 Labelling of cTALE arrays followed a previously reported protocol[25]. $NcTALE_8$
427 and $NcTALE_{12}$ were labeled at residue R30C in the first repeat, while $NcTALE_{16}$ was
428 labeled at residue R30C in the fourteenth repeat. 1 mg protein was loaded onto 500 uL
429 NiNTA spin column. The column as washed with 10 column volumes of 300 mM NaCl,
430 0.5 mM TCEP, and 10 mM $NaPO_4$ pH 7.4. Tenfold molar excess Cy3 maleimide dye
431 was resuspended in 10 μL DMSO and added to column. The column was rocked at
432 room temperature for 30 minutes, then at 4°C overnight. Cy3-labeled protein was eluted
433 with 250 mM imidazole, 300 mM NaCl, 30% glycerol, 0.5 mM TCEP, and 10 mM $NaPO_4$
434 pH 7.4. Protein was stored in 300 mM NaCl, 30% glycerol, 0.5 mM TCEP, and 10 mM
435 $NaPO_4$ pH 7.4 at -80°C.

436 **Oligonucleotides**

437 Sequences used for binding studies were 5'-Cy5-A_{15}-3' and 5' T_{15}-3' duplex
438 (Cy5-A_{15}/T_{15}) for 8 repeat binding studies, and 5'-Cy5-A_{23}-3' and 5' T_{23}-3' duplex (Cy5-
439 A_{23}/T_{23}) for 12 and 16 repeat binding studies. DNA was annealed at 5 μM concentration
440 with 1.2-fold molar excess unlabeled strand in 10 mM Tris pH 7.0, 30 mM NaCl.

441 **Single-molecule detection and data analysis**

442 Biotinylated quarts slides and glass coverslips were prepared as previously
443 described[25]. Cy3-labeled cTALEs were immobilized on biotinylated slides taking
444 advantage of neutravidin interaction with biotinylated α-penta•His antibody which binds
445 the His_6 cTALE tag. Slides were pretreated with blocking buffer (5 μL yeast tRNA, 5 μL
446 BSA, 40 μL T50) before addition of 250 pM labeled cTALE. Cy5-labeled duplex DNA
447 was mixed with imaging buffer (20 mM Tris pH 8.0, 200 mM KCl, 0.5 mg mL^{-1} BSA, 1
448 mg mL^{-1} glucose oxidase, 0.004 mg mL^{-1} catalase, 0.8% dextrose and saturated Trolox
449 ~1mg mL^{-1}) and molecules were imagined using total internal reflection fluorescence
450 microscopy. The time resolution was 50 msec for $NcTALE_8$ and 100 msec for $NcTALE_{16}$
451 and $NcTALE_{12}$. Collection and analysis was performed as previously described[26].

452 **FRET histograms**

453 A minimum of 20 short movies were collected, and the first 5 frames (50 msec
454 exposure time) were used to generate smFRET histograms. FRET was calculated as
455 $I_A/(I_A+I_D)$ where I_A and I_D are donor-leakage and background corrected fluorescence
456 emission of acceptor (Cy5) and donor (Cy3) fluorophores. In competition experiments,
457 unlabeled DNA with the same sequence as labeled DNA was mixed at indicated
458 concentrations with labeled DNA prior to imaging.

459 **Dwell time analysis**

460 Long movies were collected with 50 msec exposure time for $NcTALE_8$ and 100
461 msec exposure time for $NcTALE_{16}$ and $NcTALE_{16}$. At least 20 representative traces at
462 each DNA concentration were selected and dwell times were determined by fitting as

463 previously described using HaMMy[27] for FRET in $NcTALE_8$. Dwell times in $NcTALE_{12}$
464 and $NcTALE_{16}$ colocalization experiments are determined by using a thresholding
465 procedure for Cy5 excitation (Figure S3). The algorithm used to identify low and high
466 emission states here is slightly different than previously described thresholding
467 algorithms[28]. To reduce the number of incorrectly identified transitions arising from
468 increased background and noise at higher Cy5-labeled DNA concentrations, a
469 thresholding algorithm with two limits was implemented (see Figure S3). All FRET and
470 colocalization data are well described by models with two distinct states (0.0 FRET and
471 ~0.45 FRET as well as low colocalization and high colocalization). Dwell times of the
472 same state (low versus high FRET or low versus high colocalization) for all traces at a
473 given DNA concentration are compiled, and cumulative distribution is generated with
474 spacing equal to imaging exposure time.

475 To determine apparent rate constants using model-independent analysis,
476 cumulative distributions were fitted with single and double exponential decays (Figure 3
477 and 4). Observed rates from exponential decay fits were plotted as a function of DNA
478 concentration. Apparent rate constants were calculated as slope of DNA concentration-
479 dependent observed rates or average of DNA concentration-independent observed
480 rates.

481 **Deterministic modeling**

482 Equations 1a-1c and 2a-2c were numerically integrated using the ODE15s and
483 ODE45 solver in MATLAB. Microscopic rate constants were adjusted to minimize the
484 sum of squared residuals between ODE-determined concentration of bound or free

TALE and single molecule cumulative distributions using lsqnonlin in MATLAB. 68% confidence intervals were estimated by performing 2000 or 8000 bootstrap iterations in which residuals from the best fit of the model to the data were randomly re-sampled (with replacement) and re-fitted.

Acknowledgements

The authors thank members of the Barrick and Ha lab for their input on this work. The authors acknowledge the support of the Center for Molecular Biophysics at Johns Hopkins and Dr. Katherine Tripp for instrumental and technical support. Support to KGS was provided by NIH training grant T32-GM008403. Support for this project was provided by NIH grant 1R01-GM068462 to DB and GM112659 to TH.

497 **References**

498 1. Kay, S., Hahn, S., Marois, E., Hause, G. & Bonas, U. A bacterial effector acts as a
499 plant transcription factor and induces a cell size regulator. *Science* **318,** 648–651
500 (2007).
501 2. Römer, P. *et al.* Plant pathogen recognition mediated by promoter activation of the
502 pepper Bs3 resistance gene. *Science* **318,** 645–648 (2007).
503 3. Boch, J. *et al.* Breaking the code of DNA binding specificity of TAL-type III effectors.
504 *Science* **326,** 1509–1512 (2009).
505 4. Moscou, M. J. & Bogdanove, A. J. A simple cipher governs DNA recognition by TAL
506 effectors. *Science* **326,** 1501 (2009).
507 5. Miller, J. C. *et al.* Improved specificity of TALE-based genome editing using an
508 expanded RVD repertoire. *Nat. Methods* **12,** 465–471 (2015).
509 6. Cong, L., Zhou, R., Kuo, Y.-C., Cunniff, M. & Zhang, F. Comprehensive interrogation
510 of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat.*
511 *Commun.* **3,** 968 (2012).
512 7. Geissler, R. *et al.* Transcriptional activators of human genes with programmable
513 DNA-specificity. *PloS One* **6,** e19509 (2011).
514 8. Li, Y., Moore, R., Guinn, M. & Bleris, L. Transcription activator-like effector hybrids for
515 conditional control and rewiring of chromosomal transgene expression. *Sci. Rep.* **2,**
516 897 (2012).
517 9. Mahfouz, M. M. *et al.* Targeted transcriptional repression using a chimeric TALE-
518 SRDX repressor protein. *Plant Mol. Biol.* **78,** 311–321 (2012).

519 10. Zhang, F. *et al.* Efficient construction of sequence-specific TAL effectors for
520 modulating mammalian transcription. *Nat. Biotechnol.* **29,** 149–153 (2011).
521 11. Maeder, M. L. *et al.* Targeted DNA demethylation and activation of endogenous
522 genes using programmable TALE-TET1 fusion proteins. *Nat. Biotechnol.* **31,** 1137–
523 1142 (2013).
524 12. Miyanari, Y., Ziegler-Birling, C. & Torres-Padilla, M.-E. Live visualization of
525 chromatin dynamics with fluorescent TALEs. *Nat. Struct. Mol. Biol.* **20,** 1321–1324
526 (2013).
527 13. Ma, H., Reyes-Gutierrez, P. & Pederson, T. Visualization of repetitive DNA
528 sequences in human chromosomes with transcription activator-like effectors. *Proc.*
529 *Natl. Acad. Sci. U. S. A.* **110,** 21048–21053 (2013).
530 14. Li, T. *et al.* TAL nucleases (TALNs): hybrid proteins composed of TAL effectors
531 and FokI DNA-cleavage domain. *Nucleic Acids Res.* **39,** 359–372 (2011).
532 15. Christian, M. *et al.* Targeting DNA Double-Strand Breaks with TAL Effector
533 Nucleases. *Genetics* **186,** 757–761 (2010).
534 16. Deng, D. *et al.* Structural basis for sequence-specific recognition of DNA by TAL
535 effectors. *Science* **335,** 720–723 (2012).
536 17. Mak, A. N.-S., Bradley, P., Cernadas, R. A., Bogdanove, A. J. & Stoddard, B. L.
537 The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* **335,**
538 716–719 (2012).
539 18. Boch, J. & Bonas, U. Xanthomonas AvrBs3 family-type III effectors: discovery
540 and function. *Annu. Rev. Phytopathol.* **48,** 419–436 (2010).

541 19. Geiger-Schuller, K. & Barrick, D. Broken TALEs: Transcription Activator-like
542 Effectors Populate Partly Folded States. *Biophys. J.* **111,** 2395–2403 (2016).
543 20. Gao, H., Wu, X., Chai, J. & Han, Z. Crystal structure of a TALE protein reveals an
544 extended N-terminal DNA binding region. *Cell Res.* **22,** 1716–1720 (2012).
545 21. edited by S. E. Harding, A. J. R. *Analytical ultracentrifugation in biochemistry and*
546 *polymer science*. (Royal Society of Chemistry, 1992).
547 22. Rinaldi, F. C., Doyle, L. A., Stoddard, B. L. & Bogdanove, A. J. The effect of
548 increasing numbers of repeats on TAL effector DNA binding specificity. *Nucleic Acids*
549 *Res.* (2017). doi:10.1093/nar/gkx342
550 23. Cuculis, L., Abil, Z., Zhao, H. & Schroeder, C. M. Direct observation of TALE
551 protein dynamics reveals a two-state search mechanism. *Nat. Commun.* **6,** 7277
552 (2015).
553 24. Cermak, T. *et al.* Efficient design and assembly of custom TALEN and other TAL
554 effector-based constructs for DNA targeting. *Nucleic Acids Res.* **39,** e82 (2011).
555 25. Rasnik, I., Myong, S., Cheng, W., Lohman, T. M. & Ha, T. DNA-binding
556 Orientation and Domain Conformation of the E.coli Rep Helicase Monomer Bound to
557 a Partial Duplex Junction: Single-molecule Studies of Fluorescently Labeled
558 Enzymes. *J. Mol. Biol.* **336,** 395–408 (2004).
559 26. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nat.*
560 *Methods* **5,** 507–516 (2008).
561 27. McKinney, S. A., Joo, C. & Ha, T. Analysis of single-molecule FRET trajectories
562 using hidden Markov modeling. *Biophys. J.* **91,** 1941–1951 (2006).

563 28. Blanco, M. & Walter, N. Analysis of Complex Single Molecule FRET Time
564 Trajectories. *Methods Enzymol.* **472,** 153–178 (2010).
565
566

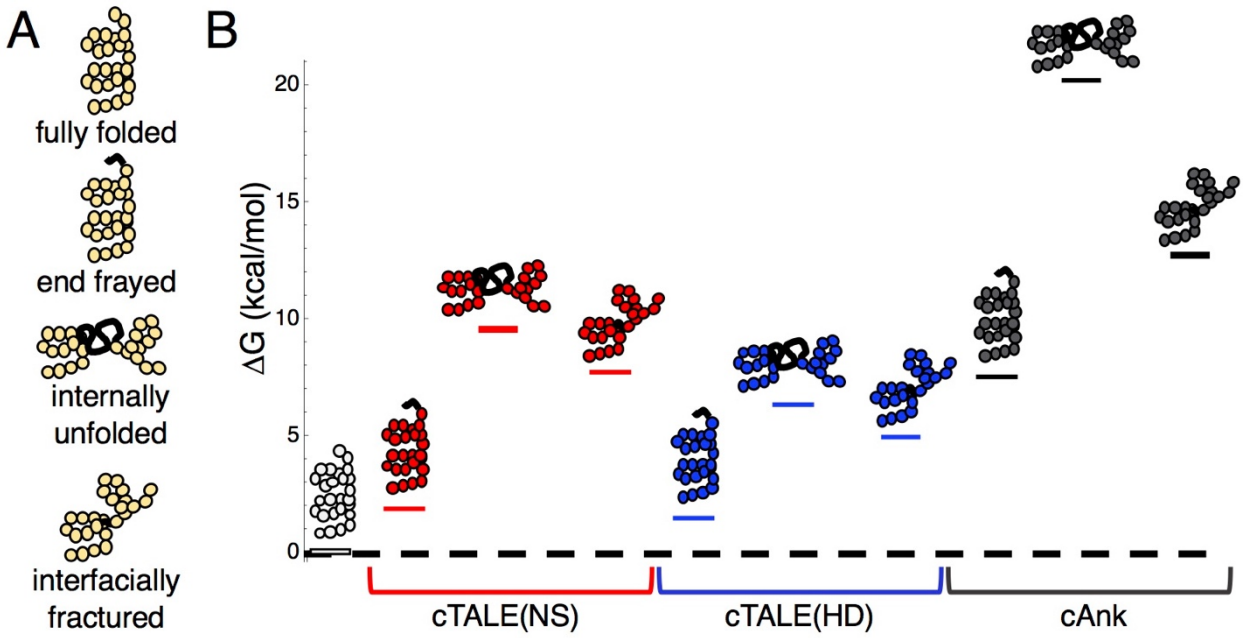
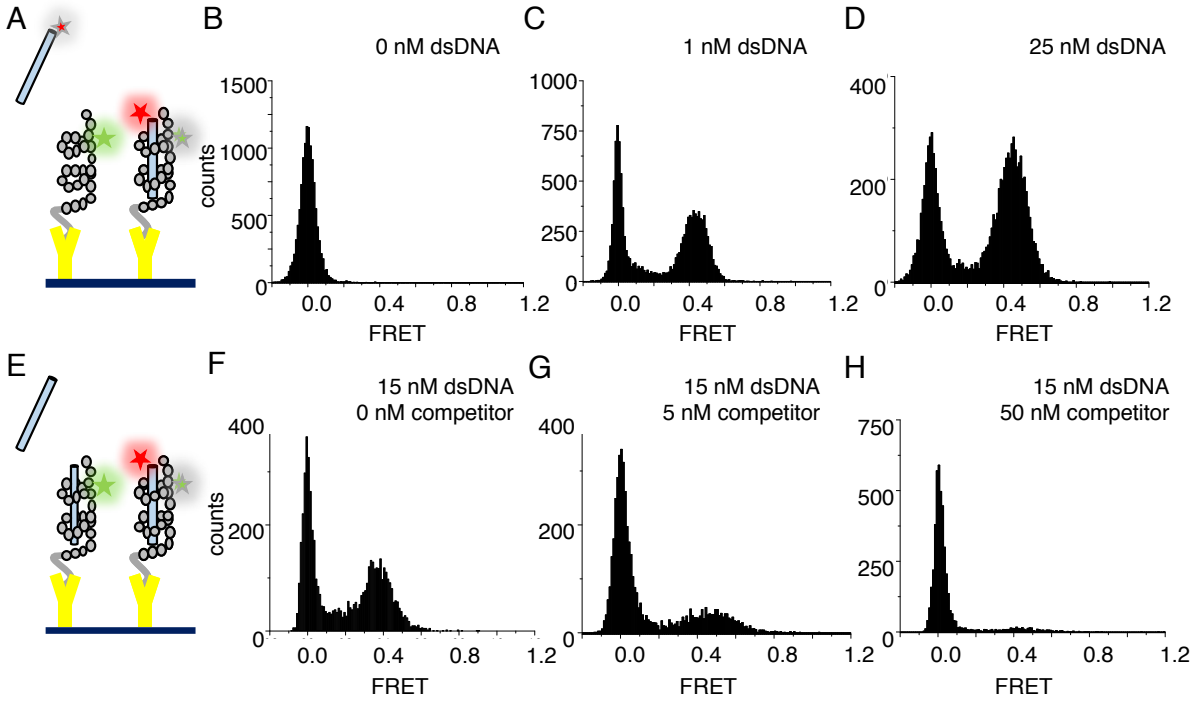


Figure 1. cTALEs populate partly folded states. (A) Cartoon of different partly folded TALE conformational states. End-frayed states have one (or more) terminal repeats unfolded. Internally unfolded states have a central repeat unfolded. Interfacially fractured states have a disrupted interface between adjacent repeats. (B) Free energies of partly folded states, calculated from previously published measurements and analysis[19], relative to the fully folded state, for consensus TALE repeats with the NS repeat-variable diresidue sequence (cTALE(NS), red), consensus TALE repeats with the HD repeat-variable diresidue sequence (cTALE(HD), blue), and consensus ankyrin repeats (cAnk, black).



578

579 **Figure 2. cTALEs bind dsDNA reversibly.** (A) Schematic of single-molecule FRET

580 assay, with donor-labelled cTALE attached to a surface, and acceptor-labelled DNA free

581 in solution. (B-D) Single molecule FRET histograms show the appearance of a peak at

582 a FRET efficiency of 0.45 with increasing labelled DNA, consistent with a DNA-bound

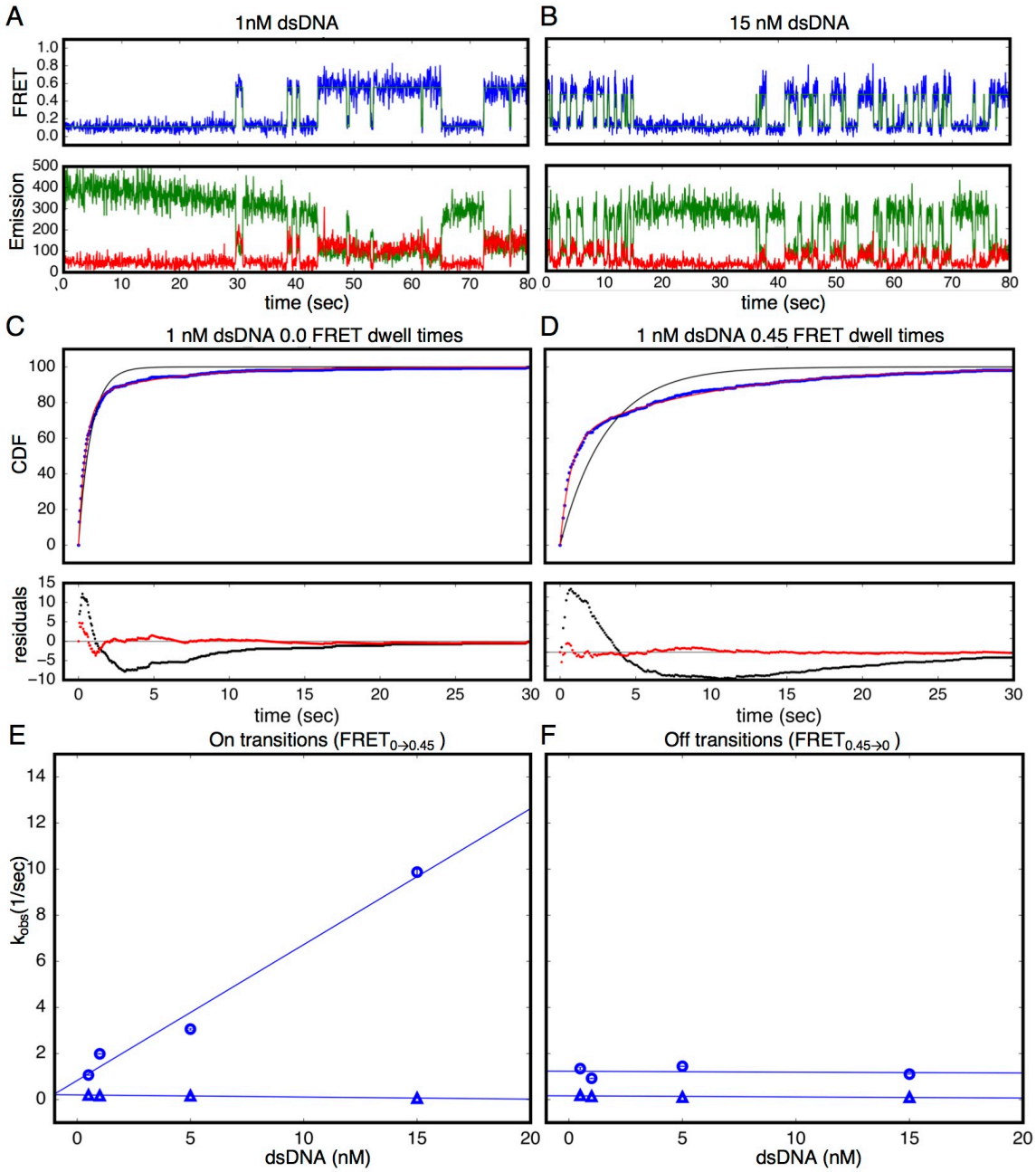
583 cTALE. (E) Schematic of single-molecule FRET competition assay, with donor-labelled

584 cTALE attached to a surface, and acceptor-labelled DNA as well as competitor

585 unlabeled DNA free in solution. (F-H) Single molecule FRET histograms show the

586 disappearance of the peak at 0.45 FRET efficiency with increasing unlabeled competitor

587 DNA. Conditions: 20 mM Tris pH 8.0, 200 mM KCl.



588

589 **Figure 3. Single Molecule (SM) kinetics show multiple phases in binding and**

590 **unbinding kinetics.** (A-B) Long time trajectories showing transitions between low- and

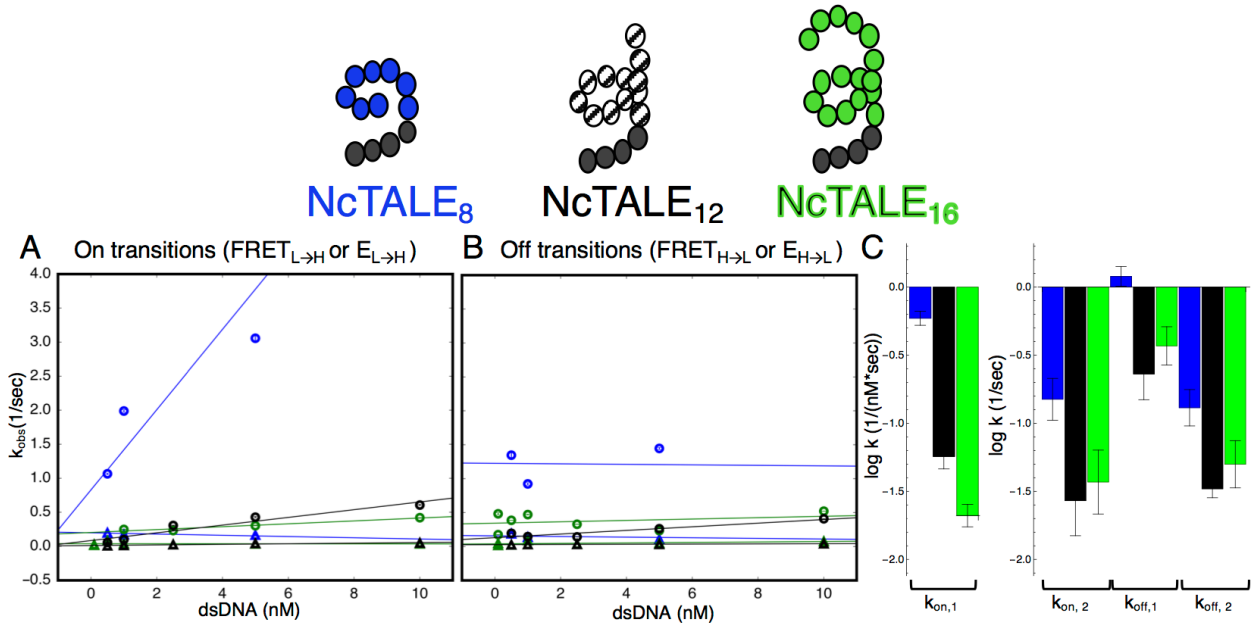
591 high-FRET states (efficiencies of 0 and 0.45). The top panel shows calculated FRET

592 efficiency in blue and two-state Hidden Markov Model fit in green[27]. The bottom panels

593 show Cy3 and Cy5 fluorescence emission in green and red respectively. At low DNA

594 concentration (A), the low FRET state predominates. As DNA concentration is
595 increased (B), more time is spent in the high FRET state, because the dwell times in the
596 low FRET state are shorter. At low DNA concentrations, there appears to be long- and
597 short-lived high-FRET states. Likewise, at near-saturating DNA concentrations, there
598 appear to be long and short-lived low FRET states. (C, D) Cumulative distributions of
599 low- and high-FRET dwell times (blue circles). Fits to single-exponentials (black) show
600 large nonrandom residuals (lower panels), consistent with the heterogeneity noted in (A)
601 and (B). Double-exponentials (red) give smaller, more uniform residuals. (E) Apparent
602 association rate constants as a function of DNA concentration. The apparent rate
603 constants for the fast phase depend on DNA concentration (blue circles), indicating a
604 bimolecular step binding event. The apparent rate constants for the slow phase do not
605 depend on DNA-concentration (blue triangles), suggesting an isomerization event. (F)
606 Apparent dissociation rate constants as a function of DNA concentration (phase 1
607 shown in blue circles, and phase 2 shown in blue triangles). Neither phase shows a
608 DNA concentration dependence, indicating a dissociation and/or isomerization events.
609 Conditions: 20 mM Tris pH 8.0, 200 mM KCl.
610

611



612

613 **Figure 4. A 16-repeat TALE protein binds and unbinds DNA more slowly than an**

614 **eight repeat protein.** (A) Apparent association rate constants as a function of DNA

615 concentration for an 8 repeat cTALE (blue), a 12 repeat cTALE (black), and a 16 repeat

616 cTALE (green). 8 repeat TALE kinetics are measured by FRET ($FRET_{L\rightarrow H}$) while 12 and

617 16 repeat TALE kinetics are measured by colocalization ($E_{L\rightarrow H}$). The apparent rate

618 constants for the fast phase of binding are DNA concentration dependent (blue, black,

619 and green circles), indicating a bimolecular binding event. The DNA concentration-

620 dependence is greatest (larger slope) for the 8 repeat cTALE. The apparent rate

621 constants for the slow phase do not depend on DNA-concentration (blue, black, and

622 green triangles), suggesting an isomerization event. (B) Apparent dissociation rate

623 constants as a function of DNA concentration (phase 1 shown in circles, and phase 2

624 shown in triangles). Neither phase shows a DNA concentration dependence, indicating

625 a dissociation and/or isomerization events. Rate constants for all phases are slower for

626 the 12-repeat construct (black) and 16-repeat construct (green) than for the 8-repeat

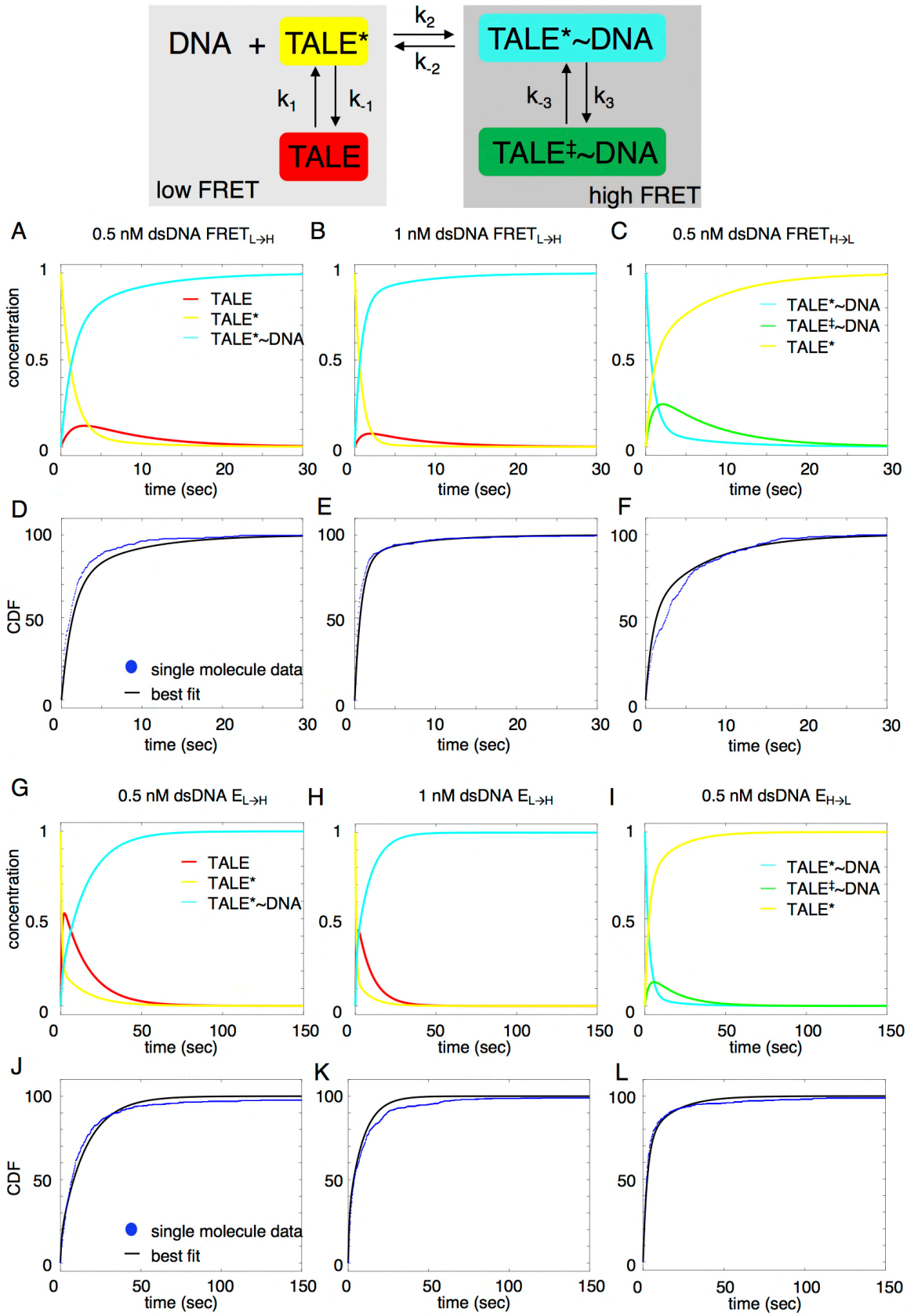
627 construct (blue), particularly for the bimolecular binding step. (C) Log_{10} of rate constants

628 for 8 (blue), 12 (black), and 16(green) repeat cTALEs. Units of the bimolecular binding

629 rate constant are $nM^{-1}s^{-1}$, other unimolecular rate constants have units s^{-1}. Conditions:

630 20 mM Tris pH 8.0, 200 mM KCl.

631


DNA + TALE*
k2
k-2
TALE*~DNA
k1
k-1
k-3
k3
TALE
TALE‡~DNA
low FRET
high FRET
A
0.5 nM dsDNA FRETL→H
B
1 nM dsDNA FRETL→H
C
0.5 nM dsDNA FRETH→L
concentration
time (sec)
TALE
TALE*
TALE*~DNA
TALE‡~DNA
D
E
F
CDF
single molecule data
best fit
G
0.5 nM dsDNA EL→H
H
1 nM dsDNA EL→H
I
0.5 nM dsDNA EH→L
J
K
L

632

633 **Figure 5. Deterministic simulations provide evidence for conformational**

634 **heterogeneity in the unbound state.** The model most consistent with data is shown at

635 the top. Unbound TALEs can exist in DNA-binding competent (TALE*) or DNA-binding

636 incompetent (TALE) states. DNA-bound TALEs can exist in short-lived (TALE*~DNA) or

637 long-lived ($TALE^{‡}$~DNA) DNA-bound states. Cumulative distributions of dwell-times

638 (shown as blue points) from 8 repeat single-molecule time trajectories (A-F) and 16

639 repeat single-molecule time trajectories (G-L) were analyzed with the model (best-fit

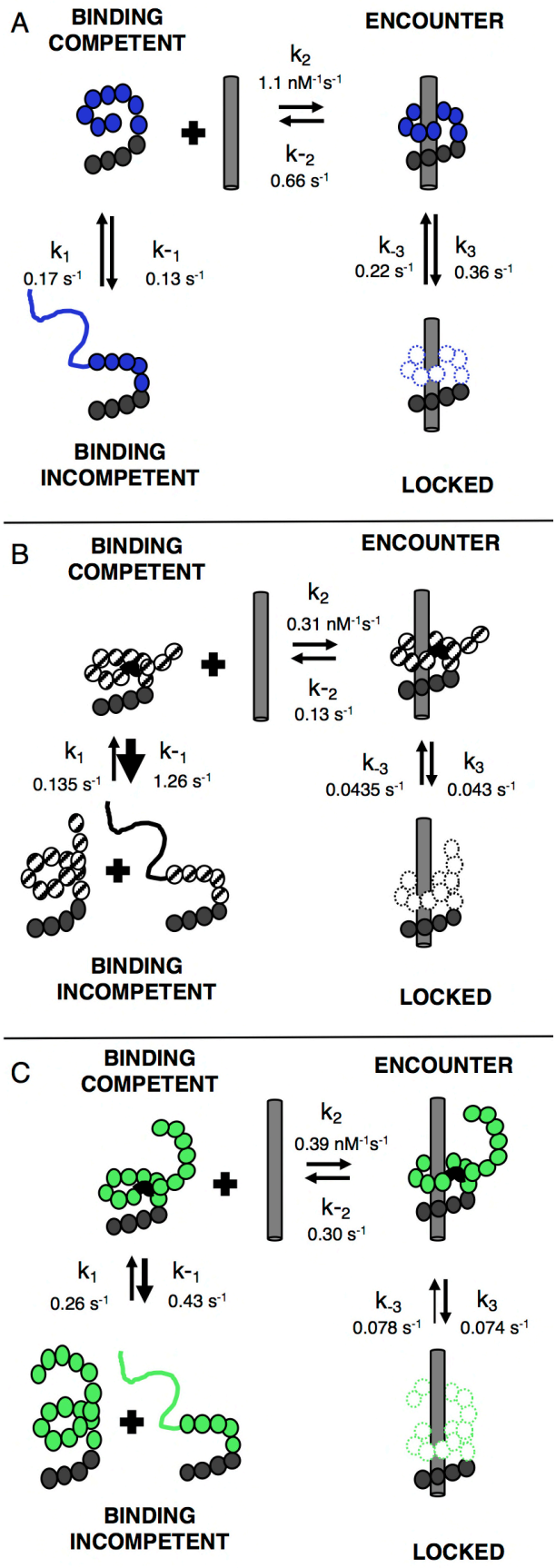
640 shown in black). (A-C and G-I) Populations of states as a function of time, generated by

641 numerical integration in Matlab. (D-F and J-L) Cumulative distributions in blue circles

642 and best fit lines are shown in black. Best-fit microscopic rate constants and 68%

643 confidence intervals are listed in Table 1.

644


A
BINDING COMPETENT
ENCOUNTER
k2
1.1 nM-1s-1
k-2
0.66 s-1
k1
0.17 s-1
k-1
0.13 s-1
k-3
0.22 s-1
k3
0.36 s-1
BINDING INCOMPETENT
LOCKED
B
BINDING COMPETENT
ENCOUNTER
k2
0.31 nM-1s-1
k-2
0.13 s-1
k1
0.135 s-1
k-1
1.26 s-1
k-3
0.0435 s-1
k3
0.043 s-1
BINDING INCOMPETENT
LOCKED
C
BINDING COMPETENT
ENCOUNTER
k2
0.39 nM-1s-1
k-2
0.30 s-1
k1
0.26 s-1
k-1
0.43 s-1
k-3
0.078 s-1
k3
0.074 s-1
BINDING INCOMPETENT
LOCKED

645

646 **Figure 6. TALEs with multiple superhelical turns must break to bind DNA.** Single-
647 molecule FRET studies and deterministic modeling support a model where TALEs exist
648 in four states: binding incompetent, binding competent, encounter complex, and locked.
649 In this model, for TALEs that form less than one full superhelical turn (8 repeats, A),
650 partly folded states are off-pathway and slow down binding. For longer TALEs that form
651 one (12, B) or more (16, C) complete superhelical turns, partial unfolding is required for
652 binding. DNA-bound TALEs form both encounter complexes and higher-affinity locked
653 conformations. Dynamics of long (12 and 16-repeat; B-C) TALEs bound to DNA are
654 significantly slower than for the shorter (8-repeat; A) TALE.
655

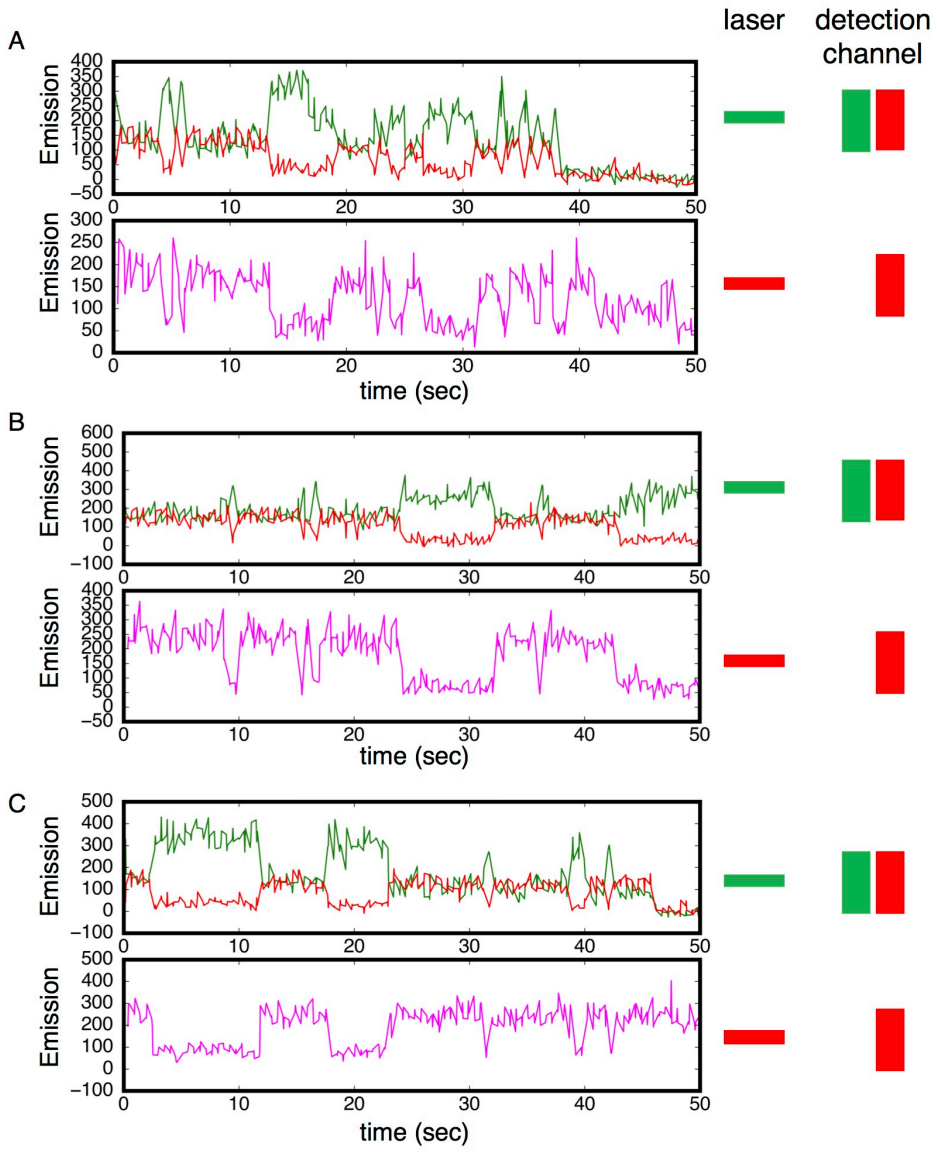
Table 1. Kinetic parameters obtained from deterministic simulation fits.

	k_1 (sec^{-1})	k_{-1} (sec^{-1})	$K_{eq,\ DNA\text{-}free}$	k_2($sec^{-1}nM^{-1}$)	k_{-2} (sec^{-1})	k_3 (sec^{-1})	k_{-3} (sec^{-1})	$K_{eq,\ DNA\text{-}bound}$ (nM^{-1})
$NcTALE_8$ [a]	0.17 [0.16, 0.18]	0.13 [0.12, 0.14]	1.32 [1.26, 1.39]	1.1 [1.08, 1.12]	0.66 [0.65, 0.67]	0.36 [0.35, 0.37]	0.222 [0.218, 0.227]	1.62 [1.58, 1.66]
$NcTALE_{12}$ [b]	0.135 [0.133, 0.137]	1.26 [1.09, 1.34]	0.11 [0.10, 0.12]	0.31 [0.28, 0.33]	0.130 [0.129, 0.131]	0.043 [0.042, 0.044]	0.0435 [0.0428, 0.0442]	0.99 [0.97, 1.00]
$NcTALE_{16}$ [a]	0.26 [0.25, 0.27]	0.43 [0.39, 0.47]	0.61 [0.57, 0.64]	0.39 [0.38, 0.41]	0.299 [0.298, 0.300]	0.074 [0.073, 0.076]	0.078 [0.076, 0.079]	0.96 [0.95, 0.97]

68% confidence intervals shown in brackets are from 2,000[a] and 8,000[b] iterations of bootstrap analysis.

656

657 **Supplemental Material**



658

659 **Figure S1. Alternating laser experiments show agreement between cTALE$_8$ FRET**

660 **and colocalization kinetics.** (A-C) Three representative time trajectories. In these

661 trajectories, excitation alternated between red and green with 250 msec at each color.

662 Top panels show detection of red and green light resulting from green laser excitation.

663 High red (and low green) fluorescence result from FRET. Lower panels show detection

664 of red light resulting from red laser excitation. There is strong correlation between

665 periods of high FRET in the upper panels (high red fluorescence, low green

666 fluorescence) and periods of high red fluorescence (from DNA binding) in the lower

667 panel. Conditions: 200 mM KCl, 20 mM Tris pH 8.0.

668

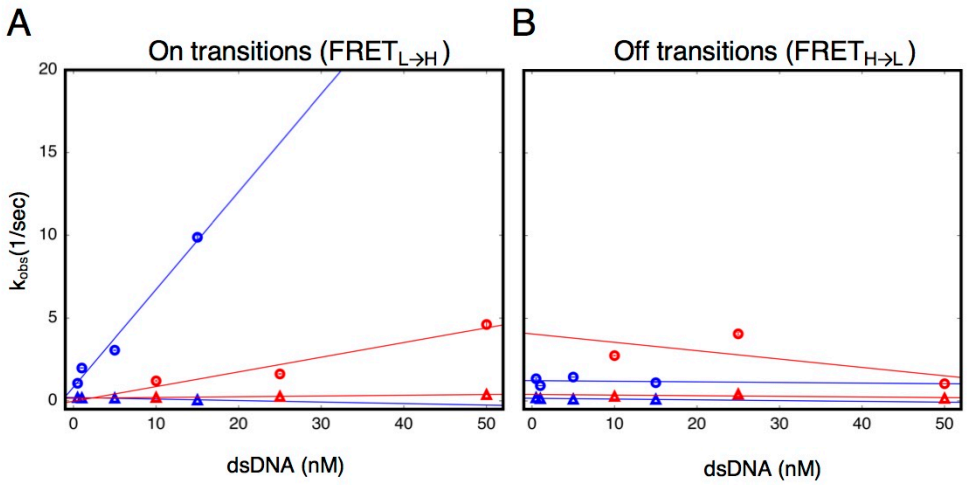
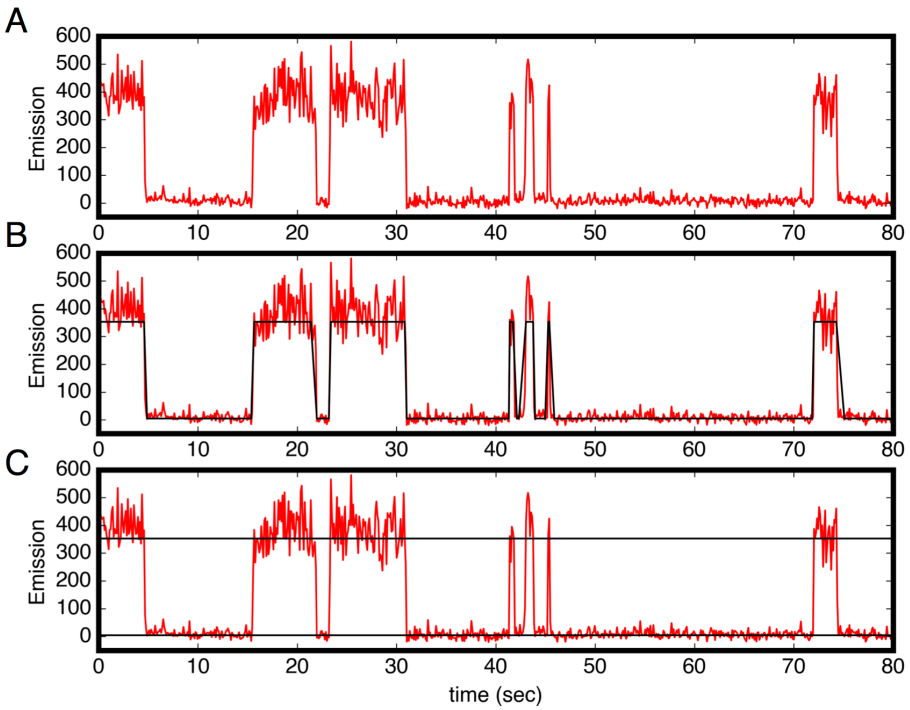


Figure S2. cTALEs do not slide onto ends of short dsDNA. Apparent association and dissociation rate constants as a function of DNA concentration for an 8 repeat cTALE array (NcTALE$_8$) binding to uncapped (blue) and capped (red) DNA, measured by single molecule FRET dwell time analysis. (A) Rate constants for conversion from the low to the high FRET state (FRET$_{L\rightarrow H}$). Rate constants for the faster phase increase with DNA concentration (circles), indicating a bimolecular event. The DNA concentration-dependence is stronger (larger slope, k=0.59 ± 0.08 nM^{-1} sec^{-1}) with uncapped DNA than with capped DNA (k=0.09 ± 0.05 nM^{-1} sec^{-1}), which is likely a result of the faster diffusion of small, uncapped DNA (10 kDa) compared to large, capped DNA (320 kDa). Rate constants for the slower phase are not DNA-concentration dependent (triangles). (B) Rate constants for conversion from the high to the low FRET state (FRET$_{H\rightarrow L}$). Neither phase shows a DNA concentration dependent, indicating unimolecular steps. Conditions: 200 mM KCl, 20 mM Tris pH 8.0.



684

685 **Figure S3. Colocalization trajectories show TALE-DNA binding and unbinding**

686 **events.** (A-C) One representative time trajectory and colocalization analysis for a

687 sixteen repeat TALE, $NcTALE_{16}$, incubated with 1 nM Cy5-A_{15}/T_{15}. In the colocalization

688 protocol, both the green and red lasers were used in the first ten frames (not shown) to

689 identify single molecule locations. For all subsequent frames, Cy5-labeled DNA was

690 continuously excited with the red laser. Red light emission images (from Cy5) were

691 collected with time steps of 100 msec. Trajectories were generated from sites of

692 red/green colocalization in the first ten frames, and were corrected for background Cy5

693 emission (red trajectories above). For each trajectory, we set two thresholds for low and

694 high Cy5 emission (C). A molecule is designated to be in the high emission (DNA-

695 bound) state above the high threshold, and in the low emission state below the low

696 threshold. From this state-assignment procedure, dwell-times were determined in the

697 high (bound) and low (unbound) states (C). Conditions: 200 mM KCl, 20 mM Tris pH

698 8.0.

699

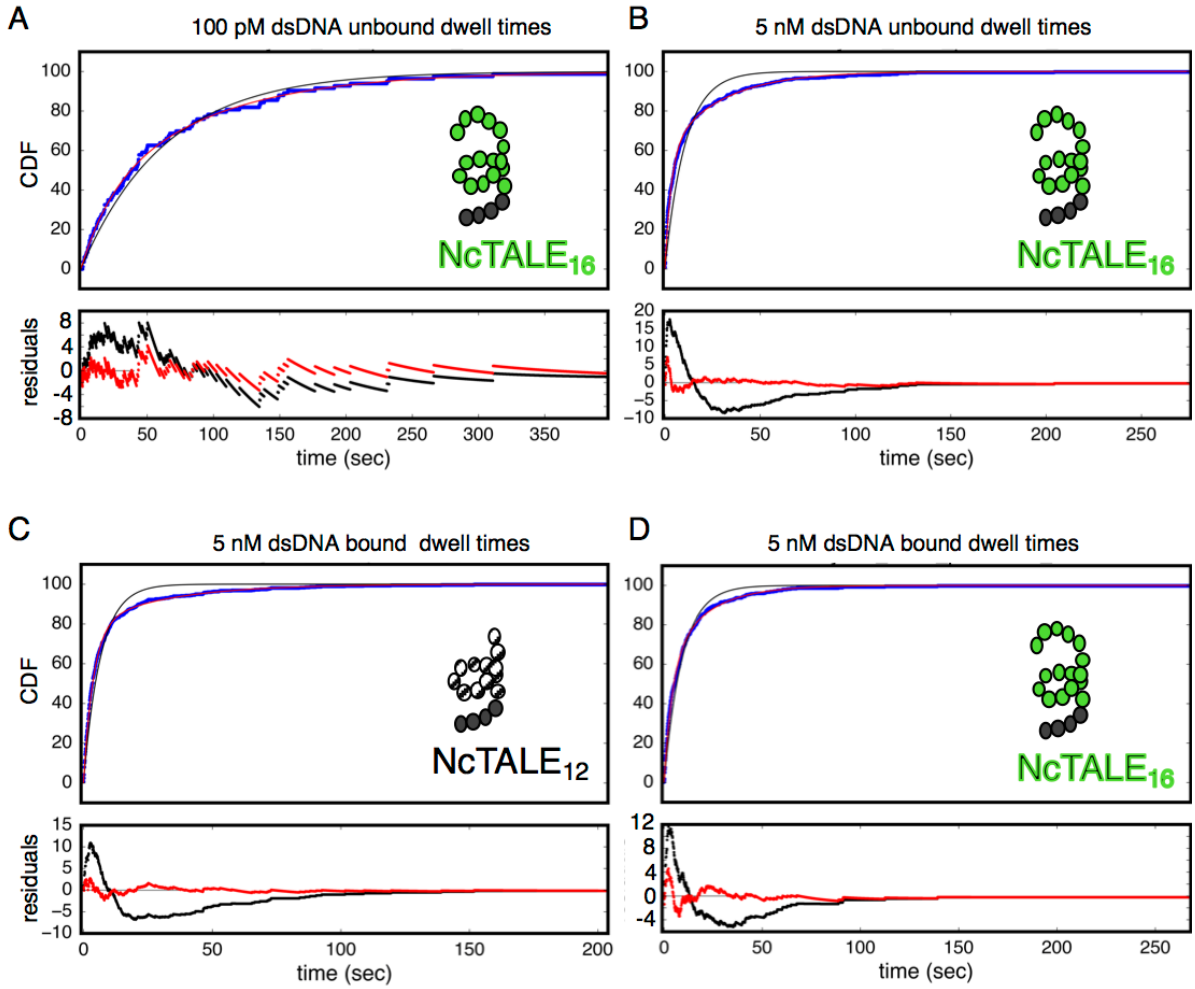
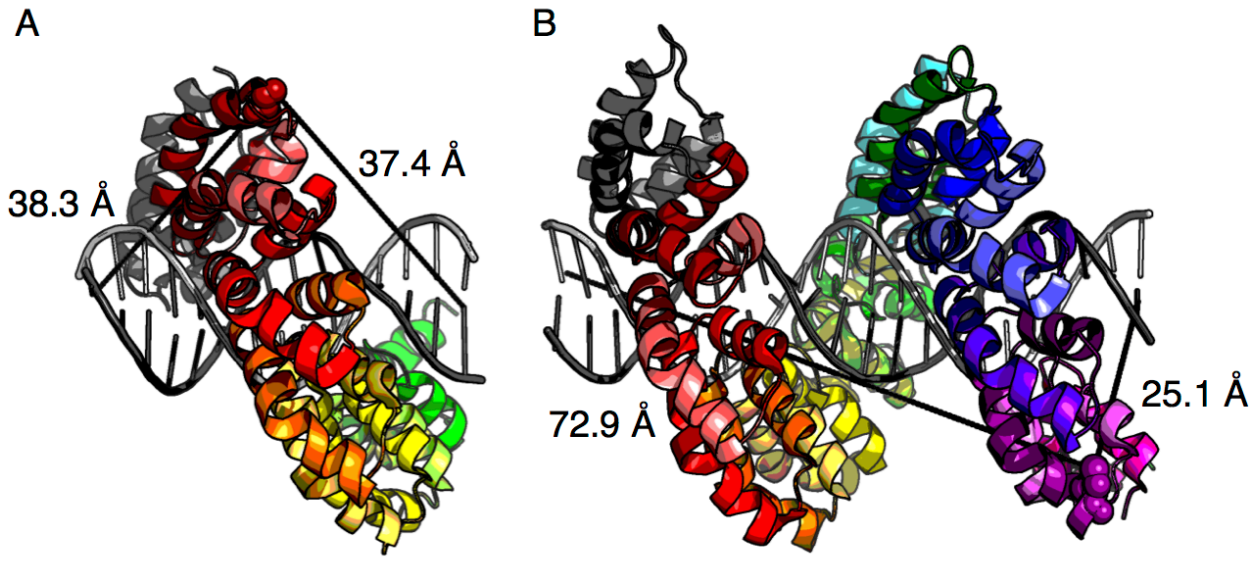


Figure S4. Bound and unbound lifetimes of 16- and 12-repeat TALE proteins are consistent with multiphasic binding and unbinding. (A, B) Cumulative distributions of low-FRET dwell times (blue circles) for a 16-repeat construct at low and high DNA concentration. Fits to single-exponentials (black) show large nonrandom residuals (lower panels) which are most pronounced at high DNA concentrations, consistent with the binding heterogeneity observed in eight repeat cTALEs (Figure 3). Double-exponentials (red) give smaller, more uniform residuals. (C, D) Cumulative distributions of high-FRET dwell times (blue circles) for 16- and 12-repeat constructs at high DNA

concentration. Fits to single-exponentials (black) show large nonrandom residuals (lower panels), consistent with the dissociation heterogeneity observed in eight repeat cTALEs. Double-exponentials (red) give smaller, more uniform residuals.



713

714 **Figure S5. Distance estimates between labeling sites for NcTALE$_8$ and NcTALE$_{16}$**

715 **and the 5' ends of bound DNA.** The crystal structure of PthXo1 (PDB: 3UGM) [17] in

716 complex with DNA (colored by repeat) as a model of donor-acceptor fluorophore

717 distances . (A) The distance from C30 in the first repeat (the site of labelling in

718 NcTALE$_8$) to the 5' base of the sense strand (light grey) is 38.3Å; the distance to the 5'

719 base of the antisense strand (dark grey) is 37.4Å. (B) The distance from C30 in the

720 fourteenth repeat (the site of labelling in NcTALE$_{16}$) to the 5' base of the sense strand

721 (light grey) is 72.9Å; the distance to the 5' base of the antisense strand (dark grey) is

722 25.1Å. Note that although there is likely to be some variation these distances since the

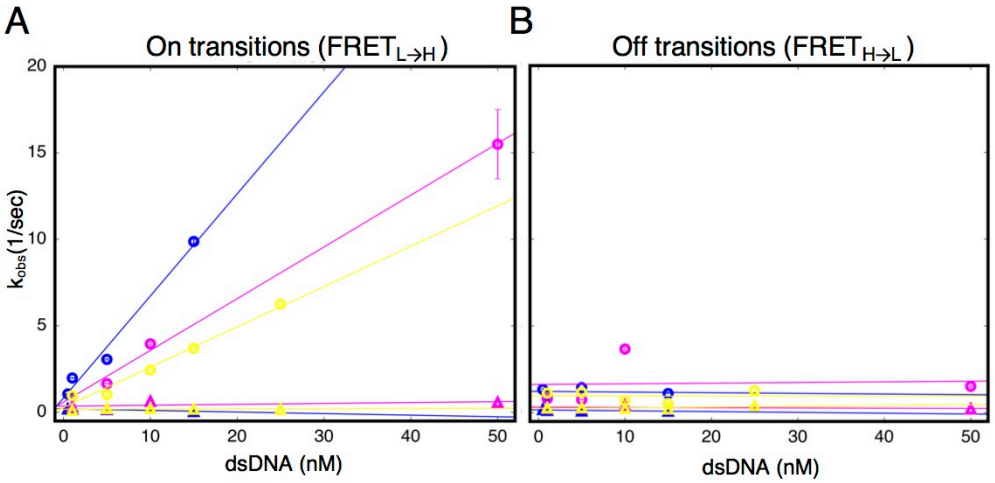
723 DNAs used here are longer (15 and 23 bases) than the TALE arrays (8 and 16 repeats,

724 each with a four-repeat N-capping domain), the differences in donor-acceptor distances

725 for the 16 repeat construct (B) are likely to be robust to registry shifts of a few bases.

726

727



728

729 **Figure S6. Urea and destabilizing mutations decrease apparent binding rate of**

730 **cTALE$_8$.** (A) Apparent association rate constants as a function of DNA concentration for

731 an 8 repeat cTALE in 0 M urea (blue), 1 M urea (pink), and with destabilizing point

732 mutations (yellow; mutations substitute leucine at position 1 in the fourth repeat with a

733 glycine). The DNA concentration dependence is strongest (steepest slope) in the

734 absence of urea and destabilizing point substitutions (circles). Rates in the slower

735 phase (triangles) appear unaffected by urea or mutational destabilization. (B) Apparent

736 dissociation rate constants as a function of DNA concentration (phase 1 shown in

737 circles, and phase 2 shown in triangles). Rates in the both phases appear unaffected by

738 urea and mutational destabilization. Conditions: 200 mM KCl, 20 mM Tris pH 8.0.