Supporting Information for

Effective Distance for DNA-Mediated Charge Transport between Repair Proteins

Edmund C. M. Tse[1], Theodore J. Zwang[2], Sebastian Bedoya, Jacqueline K. Barton*

Division of Chemistry and Chemical Engineering, California Institute of Technology,

Pasadena, CA 91125

[1] Current Address: Department of Chemistry, The University of Hong Kong, Pokfulam

Road, Hong Kong SAR

[2] Current Address: Department of Chemistry and Chemical Biology, Harvard University,

Cambridge, MA 02138

*To whom correspondence should be addressed at jkbarton@caltech.edu

Supplemental Methods

    Protein Preparation and Characterization. Wild-type (WT) Endonuclease III

(EndoIII) and Y82A EndoIII mutant were prepared as described previously.[1,2] Crude

proteins were harvested from *E. coli* cells and purified using fast protein liquid

chromatography (FPLC, Bio-Rad NGC) at 4 °C. Protein concentration was quantified

based on the absorbance of the [4Fe4S] cluster in the native oxidation state ($\varepsilon_{410} = 17000$

M^{-1} cm^{-1}) [3] using a 100 Bio UV-visible spectrophotometer (Cary, Agilent).

    Synthesis of Linear DNA Duplexes of Various Lengths. pUC19 (2686 bp),

pT2/SVNeo (5153 bp), pCXWB-EBNA1 (7037 bp), and gag/pol (8922 bp) (Addgene)

were cultured individually in LB containing the appropriate antibiotics. Circular plasmids

were harvested from cell cultures using QIAprep® Miniprep plasmid purification kit

(Qiagen), and then were linearized using Sac1, a restriction enzyme (NEB). *E. coli*

plasmid excision was carried out by incubating plasmid (1 μg) and Sac1 (10 enzyme

units, 1 μL) at 37 °C for 1 hour in 1× CutSmart® Buffer (50 mM potassium acetate, 20

mM tris-acetate, 10 mM magnesium acetate, 100 μg/mL BSA, pH 7.9, 50 μL, NEB) in a

C1000 Touch Thermal Cycler (Bio-rad). Linear DNA duplexes were then separated from

the reaction mixture by a QIAquick PCR purification kit (Qiagen).

    The Polymerase Chain Reaction (PCR) was used to amplify different lengths of

linear DNA using pUC19 (2686 bp) and pCXLE-EGFP (10911 bp) (Addgene) as the

plasmid templates. Benchling was used to design primers to yield 1625, 1999, 3895,

5967, and 7996 bp DNA duplexes (Table S1). PCR was carried out in a 50 μL solution

containing plasmid template (1 μg), forward primer (10 μM, 2.5 μL), backward primer

(10 μM, 2.5 μL), and Q5® Hot Start High-Fidelity 2× Master Mix (25 μL, NEB). PCR

was conducted using a C1000 Touch Thermal Cycler (Bio-rad) using temperature conditions deduced using a melting temperature (T_m) calculator (NEB).

Modeling Protocols. This equilibrium model takes into consideration the number of proteins that are bound to each duplex, and the model assumes that each protein-bound macrostate has its own equilibrium. These equilibria will be influenced by factors that change the protein's effective affinity to DNA such as the redox identity of the [4Fe4S] cluster and the likelihood that the protein will dissociate from DNA via DNA-mediated CT. These factors are not inherently observable and thus may be described as microstates, with all of the possible variations in these characteristics having their own predicted affinities. Combining all of the microstate affinities with the probability of those microstates being observed gives the effective affinity of the macrostate. *Eq. (1)* describes the combination of microstates for a single protein.

$$K_1 = \frac{k_1}{k_{-1}} = N_1\left(\kappa_1^R + \kappa_1^O\right) \qquad (1)$$

Here K_1 is the effective affinity describing a single protein bound to a DNA duplex. This protein can be either in the reduced (R) or oxidized (O) state, each of which has its own affinity for the DNA duplex. The likelihood of a state being in the reduced or oxidized form must also be considered. Combining these give the terms κ_1^R (and κ_1^O) which are the ratios reduced (and oxidized) protein multiplied by the single base pair affinity of a protein in that oxidation state. The term N_1 is a unitless proportionality term that divides the sum of each microstate by the total number of microstates to get an average affinity for all microstates together. For a single protein there is no possibility for DNA-mediated CT so it does not contribute to the affinity. Therefore the percent of reduced and oxidized protein can be determined by relating the known affinity of the reduced and oxidized state and the

percent of DNA with at least one protein via the relationship below in *Eq. (2)*.

$$\theta_1 = \frac{P}{P + \frac{1}{K_1}} \tag{2}$$

Where θ_1 is ratio of DNA duplexes with at least one protein to the total amount of DNA duplexes and P is the protein concentration incubated with the DNA.

Incorporating a second protein becomes significantly more complicated, because now there are combinations of proteins that may have different affinities depending on their location relative to one another as well as their redox identity. *Eq. (3)* describes the combination of microstates for the two-protein case.

$$K_2 = \frac{k_2}{k_{-2}} = N_{2,1}(\kappa_{1R,2R} + \kappa_{1O,2O})$$
$$+ N_{2,2}(\kappa_{1R,2O} + \kappa_{1O,2R}) + N_{2,3}(\kappa_{1,2}^{CT}) \tag{3}$$

with all of the terms having similar definitions as in the first equation. The terms $\kappa_{1R,2R}$ (and $\kappa_{1O,2O}$) represent the affinity contribution from the binding of a second protein when both proteins are reduced or both proteins are oxidized, and are determined from the ratios reduced (and oxidized) protein multiplied by the single base pair affinity of a protein in that oxidation state. The terms $\kappa_{1R,2O}$ and $\kappa_{1O,2R}$ represent the affinity of the second protein that binds in a different oxidation state from the first, but the two proteins are not within CT distance from one another so the affinity of the second protein to DNA is independent of the first. The term $\kappa_{1,2}^{CT}$ represents the affinity of the state in which one oxidized and one reduced protein are within CT distance, which results in the two proteins being able to reduce and oxidize one another and gives them a different effective affinity compared to the case in which the two proteins are not within CT distance. In this equation

the three proportionality terms $N_{2,1}$, $N_{2,2}$, and $N_{2,3}$ combine to give an average affinity for all microstates together, but the relative weight for $N_{2,2}$, and $N_{2,3}$ change depending on the length of DNA and the length that DNA CT can occur. Importantly, as the length of DNA CT approaches zero, $N_{2,2}$ approaches its maximum value and $N_{2,3}$ approaches zero. As the length of DNA CT approaches the length of the DNA duplex, $N_{2,2}$ approaches zero and $N_{2,3}$ approaches its maximum value. For all cases where the length of DNA CT is greater than the length of the DNA duplex, $N_{2,2}$ is zero. In all cases $N_{2,1}$ is insensitive to the length of DNA CT. Similar claims can be made about the proportionality terms with more proteins, but for the sake of simplicity, the N_2 terms are described here.

DNA CT decreases the effective affinity of proteins for DNA relative to the affinity of proteins with oxidized $[4Fe4S]^{3+}$ clusters, which means that when the length of CT is equal to or greater than the length of the DNA duplex, the macrostate affinity value K_2 is its weakest. The effective macrostate affinity will increase as the length of DNA increases beyond the length over which DNA CT can occur. The maximum affinity will occur when the length of DNA CT is so much smaller than the length of DNA, that it is effectively zero. Therefore, for two or more proteins that undergo DNA-mediated CT, a plot of macrostate affinity versus the length of DNA can be described by piecewise functions.

The point at which the change in slope occurs in our data points to a length at which DNA CT is attenuated. We therefore consider the point at which the change in slope occurs as the DNA CT length. There is a variety of conformations for all of the DNA lengths used, and there is no sudden change in conformations available based on sequence that could explain the change in slope that we observe. Additionally the effective affinity should decrease as the length of DNA increases due to the conformational flexibility decreasing hydrodynamic radii per base pair added,[4] which makes it a smaller target for the proteins

to find, therefore it is unlikely to be the underlying cause. Instead, the abrupt change in slope supports our hypothesis. At DNA lengths shorter than the CT length, CT occurs efficiently. At DNA lengths longer than the CT length, DNA CT is significantly attenuated. In addition to the difference in CT efficiency, the effective DNA binding affinity of [4Fe4S] cluster proteins are lower relative to the affinity of proteins with oxidized [4Fe4S]$^{3+}$ clusters in the regime where DNA length is shorter than CT length. The affinity of the CT state can thus be calculated by using *Eq. (3)* and relating the affinity of the reduced and oxidized protein, their proportion as determined above, and the observed affinity for DNA lengths where the length of DNA is greater than the length of DNA CT and thus $N_{2,2}$ is zero. The results of these calculations for the binding of up to five proteins to each DNA duplex are shown below in Supplemental Table S3 and Figure S4.
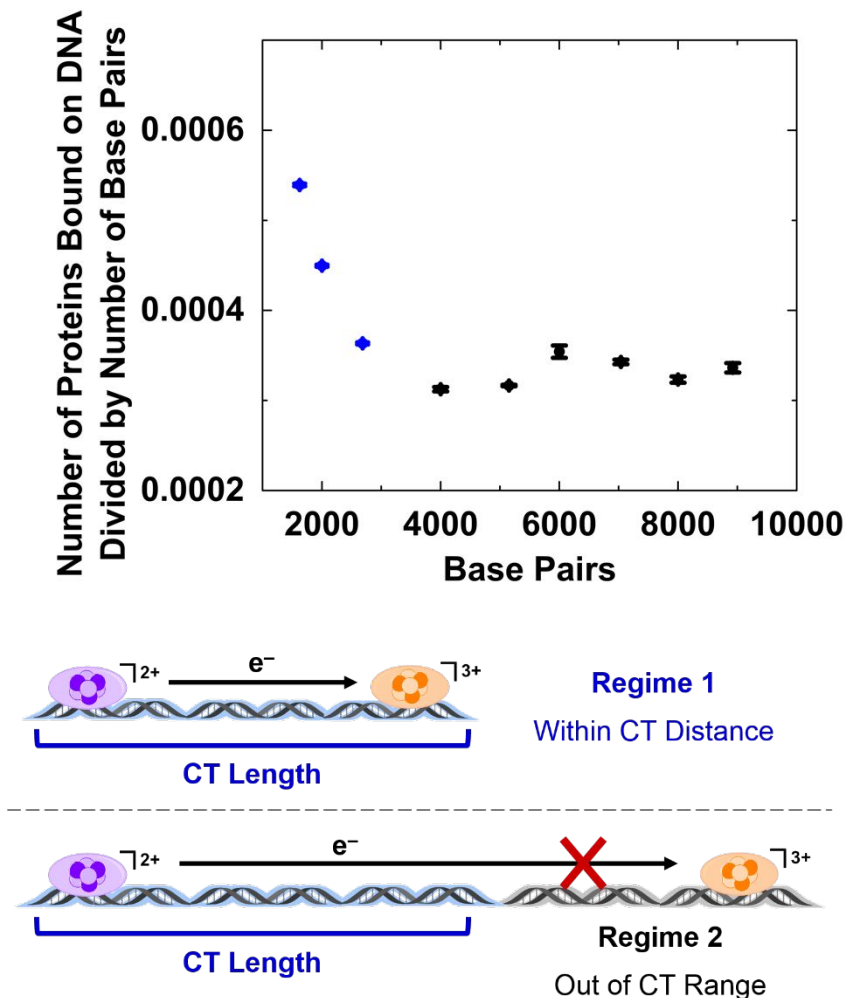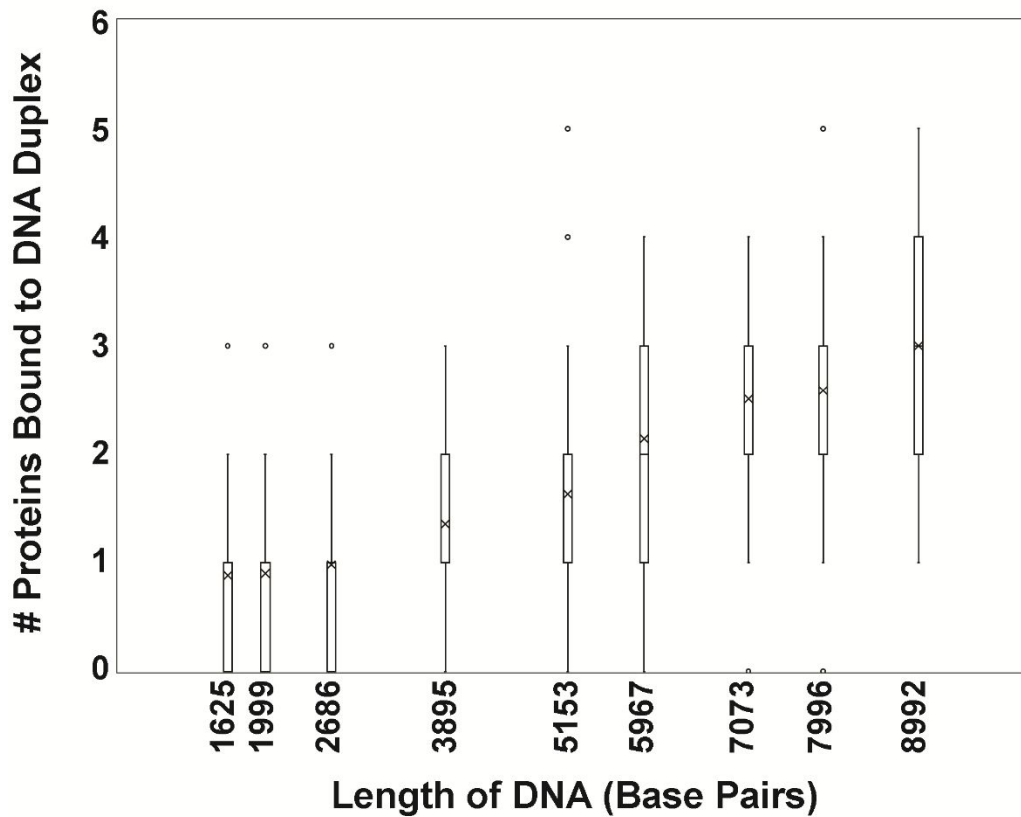
Figure S1a. Plot of number of WT EndoIII bound on dsDNA normalized by the number of base pairs of that particular dsDNA strand versus DNA length in units of base pairs.

In the blue regime (DNA length < CT length), as the DNA length increases, the number of proteins on a piece of DNA is relatively constant at ~1 protein. In the event of two proteins bound on DNA transiently, the proteins will always be within CT range with the assumption that the proteins are in different redox states (2+ and 3+). Hence, the number of protein on DNA normalized by number of bp decreases as DNA length increases.

In the black regime (DNA length > CT length), as the DNA length increases, the probability of proteins on a piece of DNA not in CT range increases, therefore, the number of proteins on DNA increases as DNA length increases. After normalizing to the number of bp, the number of proteins on DNA divided by BP has a different slope or dependence compared to the case with all proteins within CT distance.
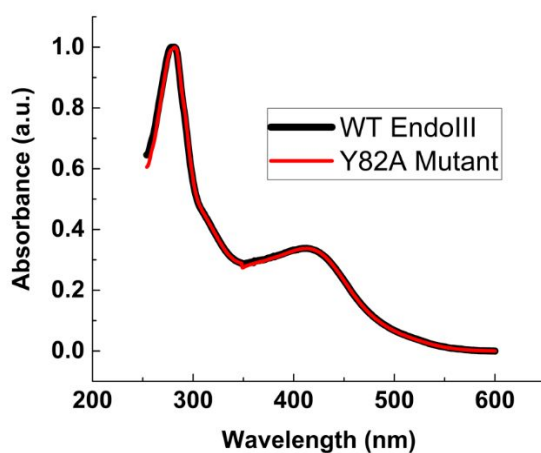

Figure S1b. Box plot of number of WT EndoIII bound on dsDNA versus DNA length in units of base pairs.
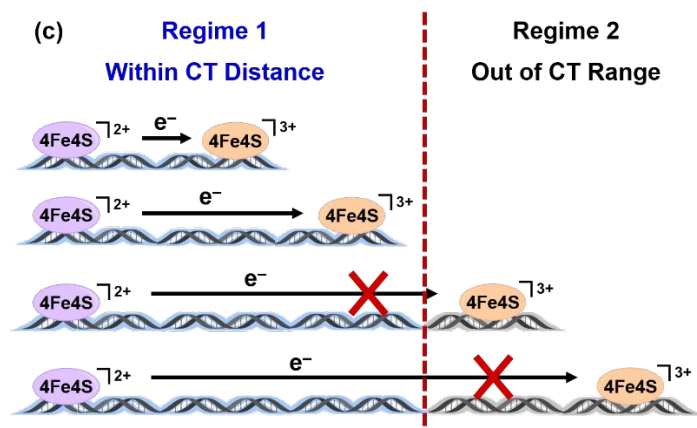
Figure S1c. A cartoon showing that as the DNA lengths increases, the chance of having two [4Fe4S] proteins outside of the effective DNA CT distance increases. The model does not suggest that proteins only bind at the ends of the DNA strands.

Figure S2a. UV-visible spectra of WT EndoIII and Y82A CT-deficient EndoIII mutant. This result shows that WT EndoIII and Y82A CT-deficient EndoIII mutant have very similar [4Fe4S] cluster loading.
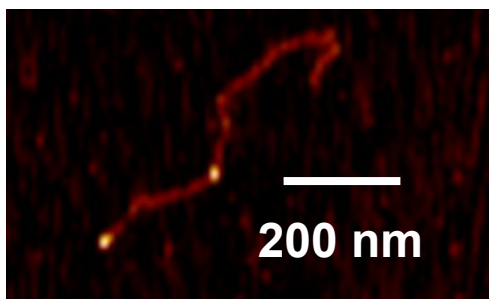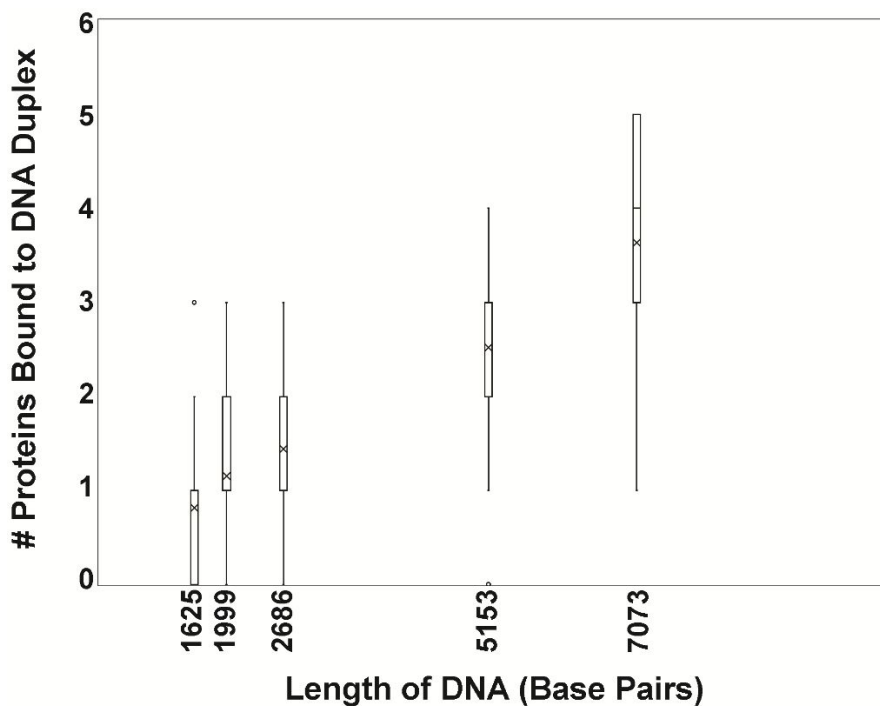

Figure S2b. AFM image of Y82A CT-deficient EndoIII mutants bound on dsDNA.
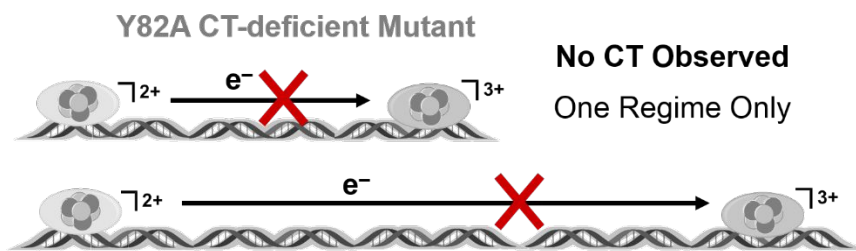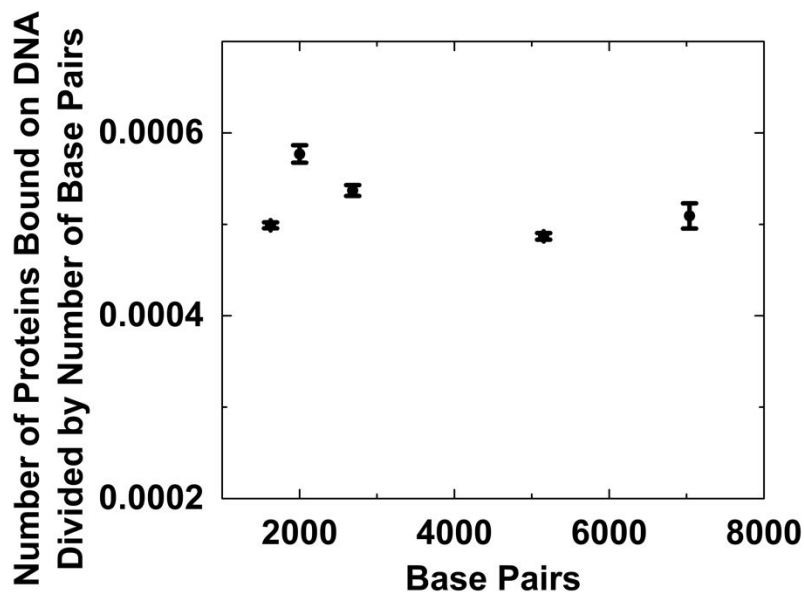
Figure S2c. Plot of number of Y82A CT-deficient EndoIII mutants bound on dsDNA normalized by the number of base pairs of that particular dsDNA strand versus DNA length in units of base pairs.

Only one regime is observed in Figure S2c for the case of Y82A CT-deficient EndoIII mutants. This regime is in reminiscent to the black regime in Figure S1a observed for WT EndoIII in the regime where the DNA length is longer than the CT length. Hence, Y82A exhibits a DNA CT length shorter than the shortest DNA length used in our AFM assay. The dissimilar slopes observed between the line representing the out-of-CT regime for WT EndoIII (Figure 2 black line, slope = 0.00036 bp^{-1}) and the line for Y82A mutant in Figure 3 (slope = 0.00049 bp^{-1}) are likely due to the difference in DNA binding affinity between WT EndoIII and Y82A. Y82A in the native oxidation state binds DNA with a similar

affinity as compared to WT EndoIII. However, the DNA binding affinity of Y82A with oxidized $[4Fe4S]^{3+}$ cluster is not readily obtainable, because chemical oxidants likely results in cluster degradation. Electrochemical oxidation method is also not practical, because Y82A is a CT-deficient mutant, which by definition the [4Fe4S] cluster is not well coupled to the DNA-modified electrode construct. The apparent binding behavior as observed by AFM is modulated by the individual binding affinity of Y82A with reduced $[4Fe4S]^{2+}$ cluster and Y82A with oxidized $[4Fe4S]^{3+}$ cluster as described by *Eq. (1)*. It is important to note, however, that no abrupt change in slope is observed in Figure 3, which supports the notion that the DNA binding affinity of Y82A is independent of each other, and further indicates that the slope change observed for WT EndoIII in Figure 2 is a result of DNA CT. The slope of number of WT EndoIII per DNA length in the short DNA regime is 0.000094 bp^{-1} (Fig. S2e,f).

Figure S2d. Box plot of number of Y82A CT-deficient EndoIII mutants bound on dsDNA versus DNA length in units of base pairs.
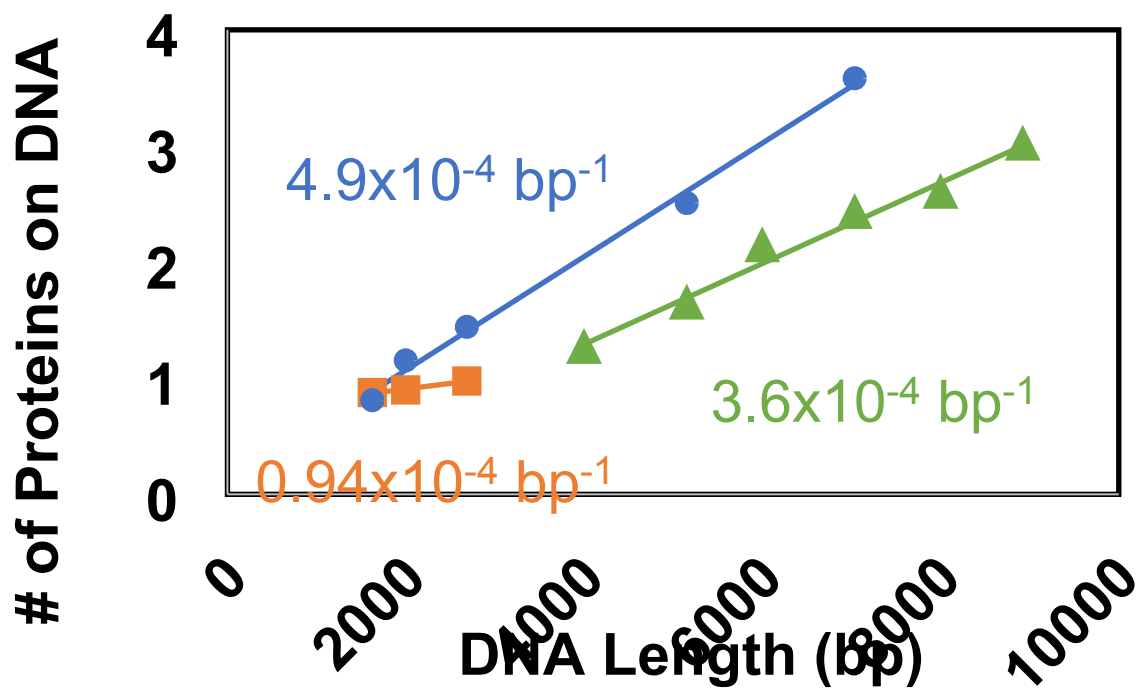
Figure S2e. Plots of number of WT EndoIII (orange and green) and Y82A EndoIII mutant (blue) bound on dsDNA versus DNA length in units of base pairs.
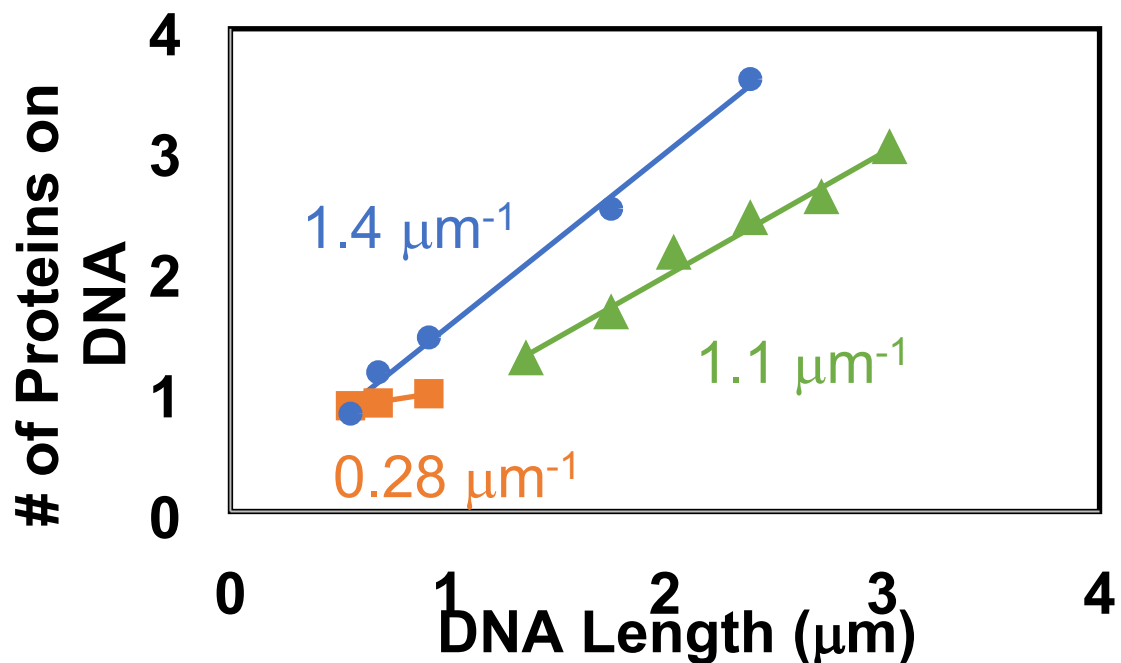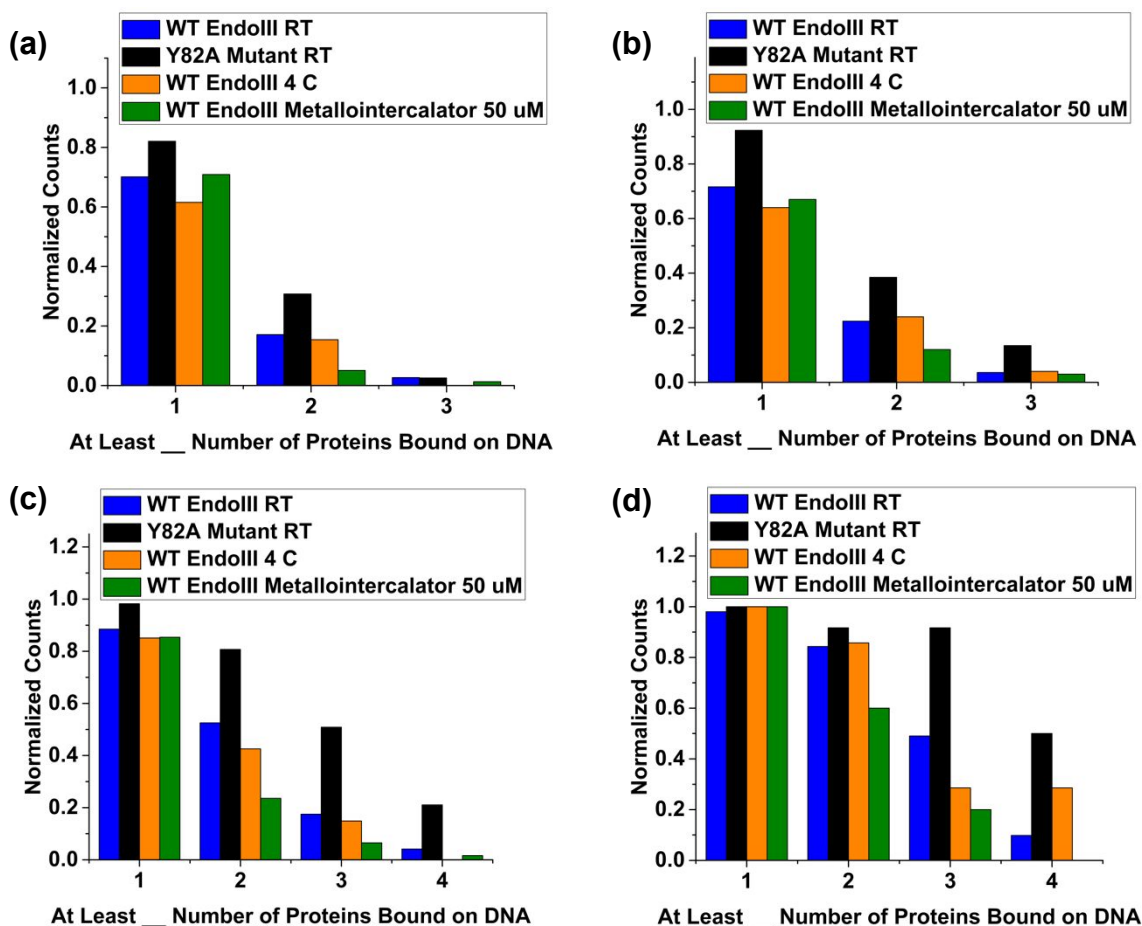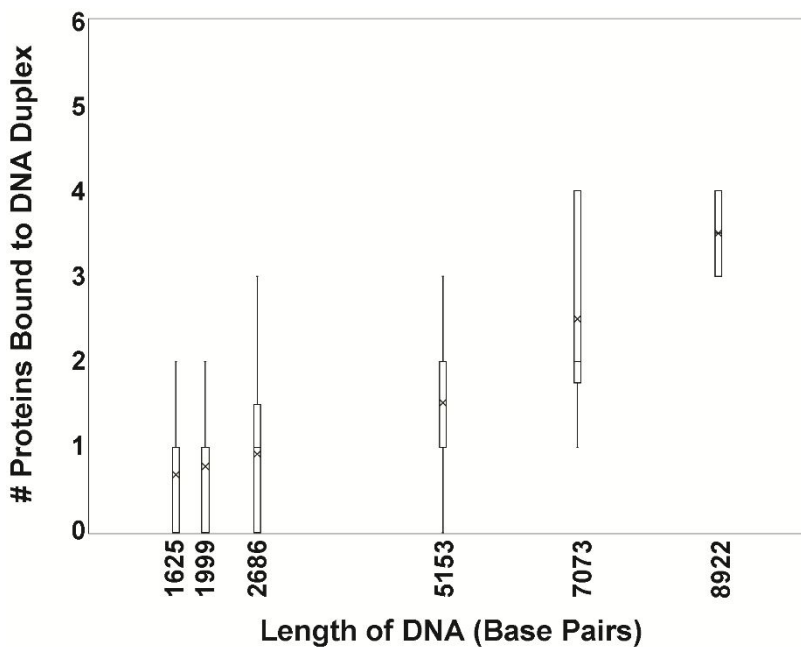

Figure S2f. Plots of number of WT EndoIII (orange and green) and Y82A EndoIII mutant (blue) bound on dsDNA versus DNA length in units of microns.

Figure S3. Normalized counts of proteins having at least 1, 2, 3, or 4 proteins on (a) 0.7, (b) 0.9, (c) 1.8, and (d) 2.4 μm dsDNA. In Figures S3a and b, for cases with DNA lengths shorter than the CT length at RT (blue bars), Y82A (black bars) exhibit a different behavior from the other cases (orange and green bars). In Figures S3c and d, for cases with DNA lengths longer than the CT length at RT, WT EndoIII at 4 °C and in the presence of [Ru(phen)$_2$dppz]Cl$_2$ (50 μM) exhibit a binding behavior that deviates from that found at RT. Upon incubating EndoIII at 37 ºC, the [4Fe4S] cluster degraded and the resulting protein formed aggregates, thus prevented us from collecting meaningful data from subsequent AFM experiments. In the presence of higher concentrations of Ru metallointercalators, lots of DNA-protein aggregates were observed by AFM, thus
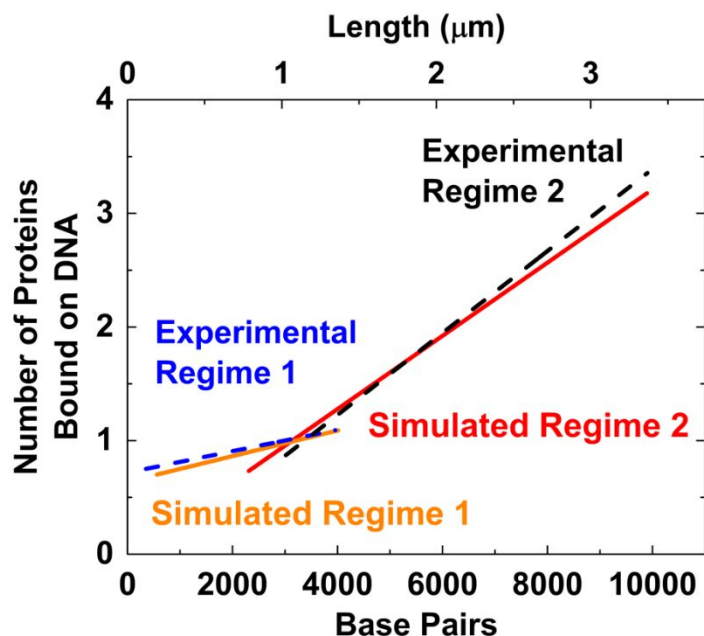
prevented us from deducing the effective DNA CT length for these cases.

Figure S3e. Box plots of number of WT EndoIII bound on dsDNA at 4 ºC vs. DNA length in units of base pairs.
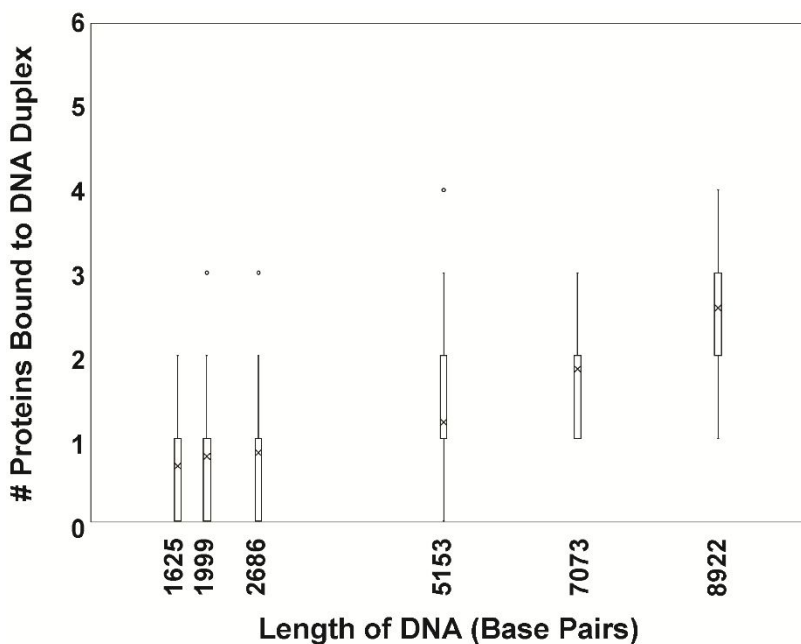

Figure S3f. Box plots of number of WT EndoIII bound on dsDNA in the presence of [Ru(phen)$_2$dppz]Cl$_2$ (50 μM) at RT vs. DNA length in units of base pairs.

Figure S4a. Comparing simulated data to experimental AFM data of wild-type EndoIII binding to DNA at ambient temperature. Orange solid line = simulated DNA binding behavior of wild-type EndoIII within DNA CT range. Red solid line = simulated DNA binding behavior of wild-type EndoIII outside of DNA CT range. Blue dashed line = experimental DNA binding behavior of wild-type EndoIII within DNA CT range. Black dashed line = experimental DNA binding behavior of wild-type EndoIII outside of DNA CT range. The trend resembling a piecewise function observed in the simulated plot in Figure S4a (orange and red solid lines) is similar to the two-slope feature found in the fit of the AFM data (blue and black dashed lines). The similarity between the simulated plot and the experimental plot underscores the ability for this math modelling to describe these collected data.

The equilibrium model equations were used to predict the number of proteins bound on DNA for the different duplex lengths used, assuming that 5.6% of the proteins have

[4Fe4S]$^{3+}$ oxidized clusters and the dissociation constant for the CT state is 4.3×10^{-5} per bp as were determined by experimental data. The trends observed in the simulated plot in Figure S4a is similar to those in Figure 2. The fact that the simulated data shows the same features as the experimental data emphasizes that the intersection of the two regimes can be interpreted as the maximum effective DNA length that DNA CT can occur through under these conditions. Differences in the value determined by the simulation may be due to it being a simple model that does not take into account all of the factors contributing to the binding affinity of proteins to DNA.
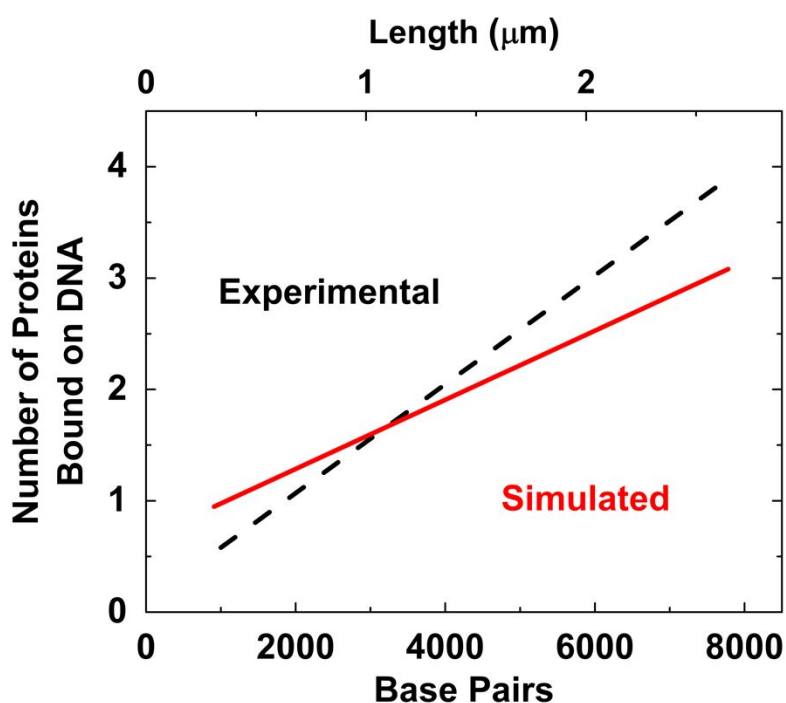
Figure S4b. Comparing simulated data to experimental AFM data of Y82A EndoIII mutant binding to DNA at ambient temperature. Red solid line = simulated DNA binding behavior of Y82A CT-deficient EndoIII mutant. Black dashed line = experimental DNA binding behavior of Y82A CT-deficient EndoIII mutant. The upward sloping trend observed in the simulated plot in Figure S4b (red solid line) is similar to the slope found in the fit of the

AFM data (black dashed line). The dissimilar slopes observed between the simulated plot and the experimental plot are likely due to the assumptions used in the model. The DNA binding affinity of Y82A with reduced $[4Fe4S]^{2+}$ and oxidized $[4Fe4S]^{3+}$ clusters are assumed to be the same as those of WT EndoIII. Y82A in the native state binds DNA with a similar affinity as compared to WT EndoIII. However, the DNA binding affinity of Y82A with oxidized $[4Fe4S]^{3+}$ cluster is not readily obtainable, because chemical oxidants likely results in cluster degradation. Electrochemical oxidation method is also not practical, because Y82A is a CT-deficient mutant, which by definition the [4Fe4S] cluster is not well coupled to the DNA-modified electrode construct. It is important to note, however, that no abrupt change in slope is observed in the simulated data, which supports the DNA binding affinity of Y82A is independent of each other, and further indicates that the slope change observed for WT EndoIII in Figure 2 is a result of DNA CT.

Supplemental Tables

Table S1. Plasmid templates and primer sequences used for constructing DNA duplexes of different lengths.

Amplicon Size (bp)	Plasmid Template	Forward Primer Sequence (5' → 3')	Backward Primer Sequence (5' → 3')
1625	pUC19	GGCATCCGCTTACAGACAAGC	CCGGTAACTATCGTCTTGAGTCC
1999	pCXLE-EGFP	GCAACGTGCTGGTTATTGTG	CACAATTCCACACAACATACGAG
3895	pCXLE-EGFP	CACAACGTCTATATCATGG	CGATTAAGGACCTTGTTATG
5967	pCXLE-EGFP	CACAACGTCTATATCATGG	CTATTAGCTAAACGAAGGAG
7996	pCXLE-EGFP	CACAACGTCTATATCATGG	GAAGCTAGAGTAAGTAGTTC

Table S2. Summary of AFM data analysis.

Condition	DNA Length (BP)	Binding Density	Standard Deviation	Standard Error
WT EndoIII at RT	1625	0.876	0.663	0.00248
	2000	0.899	0.736	0.00247
	2686	0.976	0.786	0.00314
	4000	1.25	0.887	0.00975
	5153	1.63	1.01	0.00463
	6000	2.13	1.04	0.0413
	7037	2.41	0.920	0.0181
	8000	2.59	1.12	0.0273
	8922	3.00	1.02	0.0465
EndoIII Y82A Mutant	1625	0.811	0.580	0.00523
	2000	1.15	0.745	0.0191
	2686	1.44	0.826	0.0159
	5153	2.51	1.07	0.0188
	7037	3.58	1.16	0.0970
WT EndoIII at 4 °C	1625	0.675	0.572	0.0143
	2000	0.769	0.710	0.0273
	2686	0.920	0.862	0.0345
	5153	1.43	0.927	0.0197
	7037	2.43	1.13	0.162
	8922	3.50	0.387	0.0646
EndoIII with Metallointercalators	1625	0.661	0.494	0.00429
	2000	0.772	0.576	0.00729
	2686	0.820	0.716	0.00716
	5153	1.17	0.797	0.00648
	7037	1.80	0.764	0.0306
	8922	2.57	0.782	0.112

Table S3. Equilibrium modeling results with standard errors shown.

Condition	% Oxidized	CT State Affinity per Base Pair
WT EndoIII at RT	5.60% ± 1.67%	$4.32\times10^{-5} \pm 1.42\times10^{-5}$
EndoIII Y82A Mutant	8.53% ± 2.43%	NA
WT EndoIII at 4 °C	1.73% ± 0.14%	$4.30\times10^{-6} \pm 2.26\times10^{-6}$
EndoIII with Metallointercalators	2.04% ± 0.14%	$2.09\times10^{-5} \pm 1.61\times10^{-5}$

Supplemental References

(1)   Grodick, M. A.; Segal, H. M.; Zwang, T. J.; Barton, J. K. DNA-Mediated Signaling by Proteins with 4Fe-4S Clusters is Necessary for Genomic Integrity. *J. Am. Chem. Soc.* **2014,** *136* (17), 6470−6478.

(2)   Romano, C. A.; Sontz, P. A.; Barton, J. K. Mutants of the Base Excision Repair Glycosylase, Endonuclease III: DNA Charge Transport as a First Step in Lesion Detection. *Biochemistry* **2011,** *50* (27), 6133−6145.

(3)   Cunningham, R. P.; Asahara, H.; Bank, J. F.; Scholes, C. P.; Salerno, J. C.; Surerus, K.; Munck, E.; McCracken, J.; Peisach, J.; Emptage, M. H. Endonuclease III is an Iron-sulfur Protein. *Biochemistry* **1989,** *28* (10), 4450−4455.

(4)   Smith, D. E.; Perkins, T. T.; Chu, S. Dynamical Scaling of DNA Diffusion Coefficients. *Macromolecules* **1996,** *29* (4), 1372−1373.