Electronic Supplementary Material (ESI) for Chemical Science.

Cu(II)-based DNA Labeling Identifies the Structural Link Between Activation and Termination in a Metalloregulator

Supporting Information

Joshua Casto[a], Alysia Mandato[a], Lukas Hofmann[b], Idan Yakobov[b], Shreya Ghosh[a], Sharon Ruthstein*[b], Sunil Saxena*[a]

[a]Department of Chemistry, University of Pittsburgh, Pittsburgh, PA 15260, United States and [b]Department of Chemistry , Faculty of Exact Sciences and the Institute of Nanotechnology and Advanced Materials, Bar-Ilan University, Ramat-Gan 5290002, Israel.

AUTHOR INFORMATION

Corresponding Author

*[a]Sunil Saxena – Department of Chemistry, University of Pittsburgh, Pittsburgh PA, 15260, USA. ORCID: 0000-0001-9098-6114, Phone (412) 624-8680. Email: sksaxena@pitt.edu

*[b]Sharon Ruthstein – Department of Chemistry, Faculty of Exact Sciences and the Institute of Nanotechnology and Advanced Materials, Bar-Ilan University, Ramat-Gan, 5290002, Israel. Phone 972 3-7384329. Email: Sharon.ruthstein@biu.ac.il

Methods:

Protein expression and purification:

All *E. coli* CueR protein used for experiments were wild type. Wild type CueR expression was performed using a previously described pET-28a(+) expression vector containing the CueR gene in BL-21 *E. coli* cells[1]. Protein purification was carried out by High Performance Liquid Chromatography (HPLC) using a General Electric Äkta™ Start HPLC with four 5 mL Cytiva HisTrap™ HP columns packed with Ni Sephrarose™ High Performance medium. Protein was loaded onto the column with a buffer of pH 7.0 50 mM HEPES, 150 mM NaCl, 20 mM imidazole, and 0.1 mM TCEP. After column equilibration, CueR was eluted with the same buffer, except prepared with 250 mM imidazole. The collected CueR fractions were buffer exchanged with pH 7.4 50 mM NEM and 150 mM NaCl buffer and stored in 20 % glycerol (v/v) at -80 °C. Prior to sample preparation, frozen CueR was thawed on ice and then buffer exchanged with pH 7.0 50 mM NEM and 150 mM NaCl buffer prepared in D_2O prior to sample preparation.

DPA-DNA+Cu(II) preparation:

Single-strand oligonucleotides containing 2,2'-Dipicolylamine and an abasic site were obtained from Karebay Biochem Inc. The supplier purified the strands using HPLC and characterized them using mass spectrometry. Complementary DNA strands were mixed in addition to 2.5x equivalents of $CuCl_2$ per duplex to ensure saturation of both DPA sites. Sample solutions were then annealed in D_2O to increase Cu(II) chelation. Annealing was done by heating the samples to 90 °C for one minute, 60 °C for three minutes, 50 °C for five minutes, 40 °C for ten minutes, 30 °C for five minutes, and then cooled to 4 °C using a GeneAMP PCR System 9700.

Electrophoresis mobility shift assay (EMSA):

EMSA was performed according to the previously published protocol[2]. Briefly, protein-DNA complexes were formed in 20 μL incubation buffer (25 mM HEPES pH 7.4, 250 mM NaCl, 10% v/v glycerol) for 30 min at RT. The protein+DNA complexes were resolved on 10% (37:1) native polyacrylamide gels which were prerun at 20 mA for 1 h at 4 °C. The samples were loaded onto gel and run at 80 V for one hour at 4 °C. The gel was stained with 1 μg/mL ethidium bromide in running buffer for 15 minutes and developed with the BioRad gel reader.

Final sample preparations:

All CueR samples were prepared to 150 μM of homodimer for 50 μL samples. DPA-DNA was added to CueR for final concentrations of 25 μM or 75 μM for 2:1 and 6:1 concentration ratios of protein+DNA samples, respectively. Next, for samples containing Cu(I), enough Cu(I) was added to final concentrations of 300 μM, 900 μM, and 1200 μM for x1 Cu(I), x3 Cu(I), and x4 Cu(I) samples, respectively. Cu(I) stock solutions were prepared by dissolving tetrakis(acetonitrile)copper(I) hexafluorophosphate salt in anhydrous acetonitrile under argon gas in an air free glove box. Final samples were in 50 mM NEM and 150 mM NaCl buffer in D_2O and 20 % (v/v) d_8 glycerol at a final pH of 7.4. Samples were transferred to 2 mm I.D. x 3 mm O.D. quartz EPR samples tubes and flash frozen with liquid MAPP gas prior to analysis[3].This process was first implemented in the group in 2006[4].

UV/Vis Measurements:

All UV/Vis measurements were performed on a Thermo Scientific NanoDrop 2000 spectrophotometer at room temperature. Absorbance spectra were measured from 220 to 840 nm with a 10 mm path length and 1 μL sample aliquots. Measurements were baseline corrected using a water blank. Bicinchoninic acid (BCA) stock solutions were prepared in water under low

light conditions. Each UV/Vis sample was prepared to 10 µL with a combination of Cu(I), Cu(II), BCA, and CueR in molar equivalents of 1 mM.

EPR experiments:

Continuous Wave (CW) EPR experiments were performed to determine Cu(II) coordination to DPA. CW-EPR experiments were performed with a Bruker ElexSys 580 X-band (~9.4 GHz) FT/CW spectrometer with a Bruker ER4118X-MD5. Each spectrum was acquired at 80 K and contained 1024 data points, with a magnetic field sweep of 2000 G centered at 3100 G, 10.24 ns time constant, 20.48 ms conversion time, 100 kHz modulation frequency, a 4 G modulation amplitude, and an attenuation of 30 dB. Spectra were simulated using the EasySpin software[5].

Pulsed EPR experiments for distance measurements were performed at Q-Band frequency (~35 GHz) using a Bruker E580 spectrometer, 300 W TWT amplifier, ER5106-QT2 resonator, and Bruker B8692690 cryogen free cryostat. The four-pulse Double Electron-Electron Resonance (DEER) sequence $\left[\left(\frac{\pi}{2}\right)_{v1} - \tau_1 - (\pi)_{v1} - (\tau_1 + T) - (\pi)_{v2} - \tau_2 - T - \pi_{v1} - \tau_2 - \text{echo}\right]$ with 16 step phase cycling[6] was used. For all samples, the observer pulses, $\left(\frac{\pi}{2}\right)_{v1}$ and π_{v1}, were determined to be 12 and 24 ns, respectively. The pump pulse, $(\pi)_{v2}$, was set to a 100 ns chirp with a frequency range of -200 to -100 MHz relative to the observer frequency. The pump pulse was stepped out by 24 ns for 232 points over T, to achieve a maximum dipolar evolution time of 5.8 µs. The pump pulse was applied at the magnetic field corresponding to the greatest intensity determined by an echo detected field sweep. DEERNet[7], DEERAnalysis2019[8], and ComparativeDEERAnalyzer[8] software packages were used to determine the distance distributions. Distance distribution background validations were run for 50 trials between 1 and 5 µs of the DEER time traces analyzed in DEERAnalysis2019.

MD simulations:

The Nucleic Acid Builder program in the AMBER software suite[9] was used to generate the B-DNA model. The adenine and thymine residues at the sites to be labeled were replaced in PyMOL[10] by DFT-optimized Cu(II)-DPA[11] and dSpacer structures. Molecular dynamics (MD) simulations were performed on this labeled model using the AMBER parmbsc1 force field[12]. The force field parameters for the added structures were developed and incorporated following the protocol from a previous study[11]. Solvent water was treated using the TIP3P water model[13]. The DPA-DNA was solvated in a 14 Å truncated octahedral water box and neutralized with Na^+ and Cl^- ions.

MD simulations were performed using the pmemd program in the AMBER20 software package. The solvated systems were energy-minimized. The system was thermalized from 0 to 298 K with a constant restraint force on the entire DNA molecule at 1.0 kcal mol^{-1} $Å^{-1}$ over 20 ps. Before production runs, the system was equilibrated at the final temperature and a constant pressure of 1 atm for 1 ns with constant restraint and for 1 ns without restraint. The integration time step for the heating, equilibration, and production runs was 2 fs. Periodic boundary conditions with particle mesh Ewald[14] were applied to account for the long-range electrostatic interactions under NPT (P = 1 atm) conditions. SHAKE[15] was used to restrain all bonds involving hydrogen using a nonbonded cutoff of 10 Å. All visualizations and distance measurements of the simulations were done using VMD18[16].

Sample Characterization:

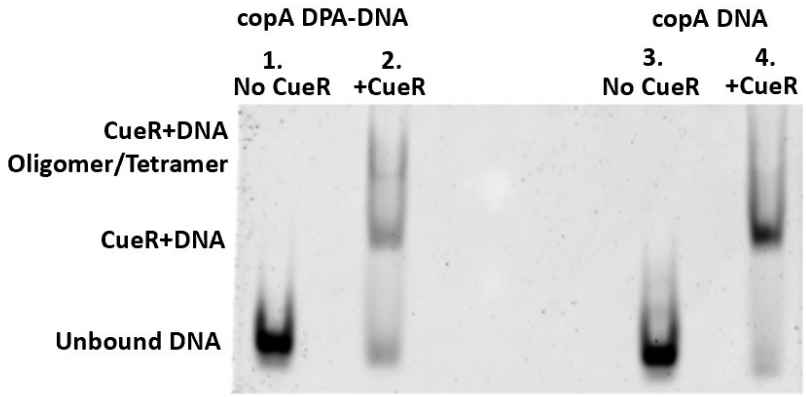

Figure S1. EMSA gel of CueR bound to copA DPA-DNA or an unmodified copA DNA sequence. Lane 1: copA DPA-DNA alone; Lane 2: copA DPA-DNA + 40 μM CueR; Lane 3: copA DNA alone; Lane 4: copA DNA + 40 μM CueR.

Figure S1 shows the EMSA gel stained with ethidium bromide. On the left side the labelled DNA sequence is used whereas on the right side the wild type DNA sequence is used. The EMSA experiment shows the free DNA in absence of protein (lane 1 and 3) and in presence of CueR protein (line 2 and 4). When CueR is present (lane 2 and 4) two bands (bound and un-bound DNA) are still visible. Although an excess protein ratio between CueR dimer and DNA of 2:1 is used, the binding between CueR and DNA is in equilibrium and therefore there will always be a fraction of un-bound DNA present. Additionally, on the left side where the labeled DNA was run with CueR a faint additional band appears above the bound DNA. Thus, it can be concluded that a higher oligomer species aside of the dimer is present.

However, the EMSA only offers a qualitative perspective based on mass separations of DNA and CueR-bound DNA. Therefore, we cannot use EMSA to differentiate between CueR bound to kinked or undistorted DNA.

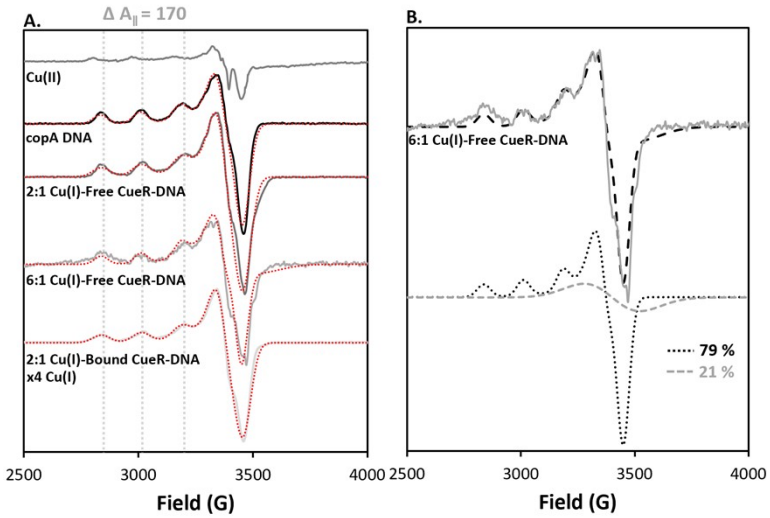

Figure S2. A) CW-EPR spectra for the different sample types. The ratios 6:1 and 2:1 refer to the concentration ratio of CueR dimer to DPA-Cu(II) labeled copA DNA. The 6:1 and 2:1 ratio samples were prepared to 25 and 75 μM Cu(II) respectively. Additional sample preparation information is detailed in the Methods section. Each sample, except for 6:1 Cu(I)-free CueR-DNA, was simulated (red dots) with a single component with $\Delta A_{\parallel}$ = 170 (grey dashed lines) and $g_{\parallel}$= 2.04, which are both characteristic of Cu(II) coordination to DPA. A 25 μM Cu(II) solution spectrum is shown as a comparison to emphasize the lineshape changes of Cu(II) after coordination to DPA. B) The 6:1 Cu(I)-free CueR-DNA sample required a two-component fit. A component characteristic of DPA-Cu(II) coordination (79%), and the other a broadened component (21%) indicative of an aggregated species.

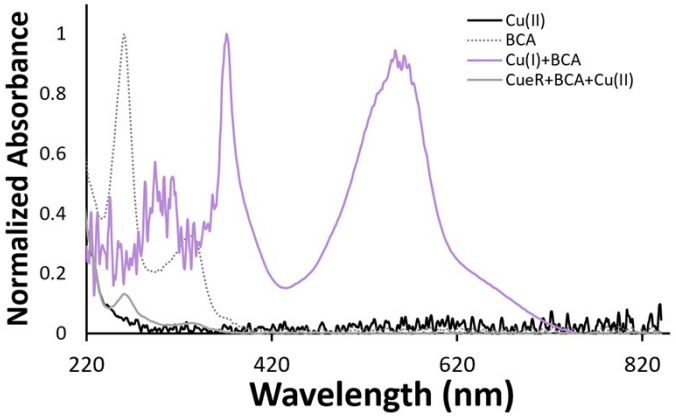

Figure S3. UV/Vis spectra for the samples as shown. All samples were prepared to 10 μL with Cu(II), BCA, Cu(I), and Cu(I)-free CueR added in molar equivalents of 1 mM each. BCA+Cu(I) complexes absorbs strongly at 562 nm (purple). No peaks at 562 nm appear in the Cu(I)-free CueR+BCA+Cu(II).

Figure S3 depicts a UV/Vis bicinchoninic acid (BCA) assay to determine if excess Cu(II) in the CueR samples are being reduced to Cu(I). The BCA-Cu(I) complex absorbs strongly at 562 nm, as shown by the purple spectra. Notably, there is no peak at 562 nm for the CueR+BCA+Cu(II) sample, indicating Cu(II) does not reduce to Cu(I) in the presence of CueR.

Supplemental DEER Data:

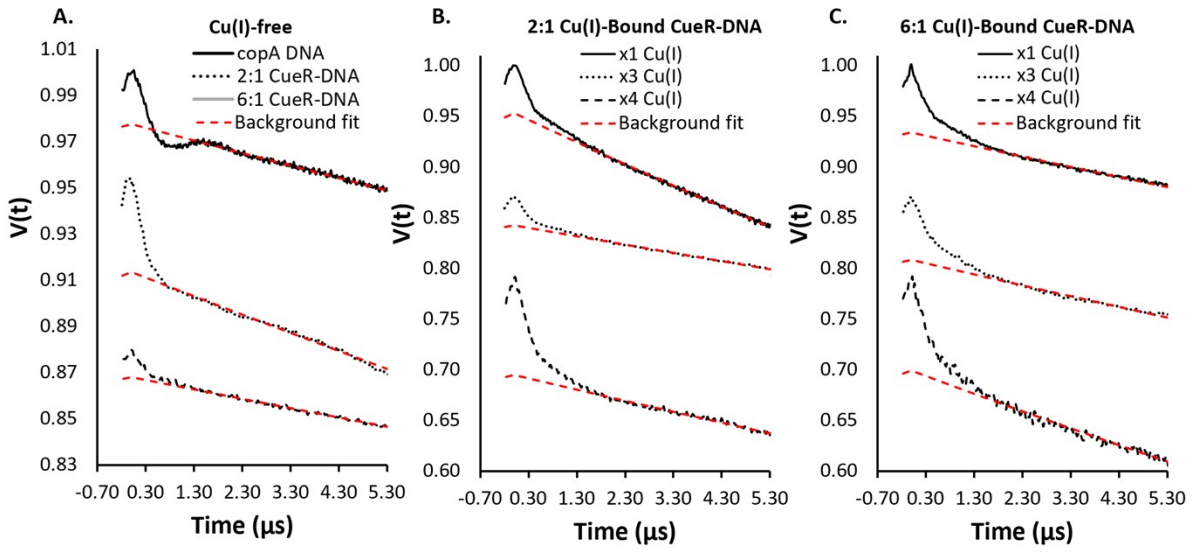

Figure S4. Raw DEER time traces and background fits using DeerAnalysis for each sample as shown. Each time trace was collected for an acquisition time of 5.8 µs over 232 points with a 24 ns step size.

The DEER time traces were collected at a single magnetic field that provided the largest echo. The DPA-DNA label does not exhibit orientational selectivity even at Q-band, permitting distance measurements at only one magnetic field. Prior to this work we demonstrated this case experimentally[17]. Subsequent to this experimental work we have computationally rationalized this experimental observation[11]. In this work we performed long force parameterized MD simulations on DPA-DNA sequences and used quantum mechanical calculations to calculate the direction of the g-tensors. For DPA the coordination to Cu(II) is elastic and there is variation of bond-length and bond-angles, which leads to a large distribution in the direction of $g_{||}$. More importantly, there are two rotatable bonds that link the DPA motif to the DNA backbone, which exhibit large fluctuations. Consequently, the relative orientations of the two Cu^{2+} g-tensors exhibit a large distribution characterized by standard deviations that are in the order of 50^0. Such orientational flexibility is sufficient to wash out orientational effects at Q-band. The case of the DPA label is distinct from the protein dHis-Cu(II) label[18-20,] where orientation effects are possible at Q-band[21]. The distinguishing feature is that the dHis-label does not have a rotatable linker[20].

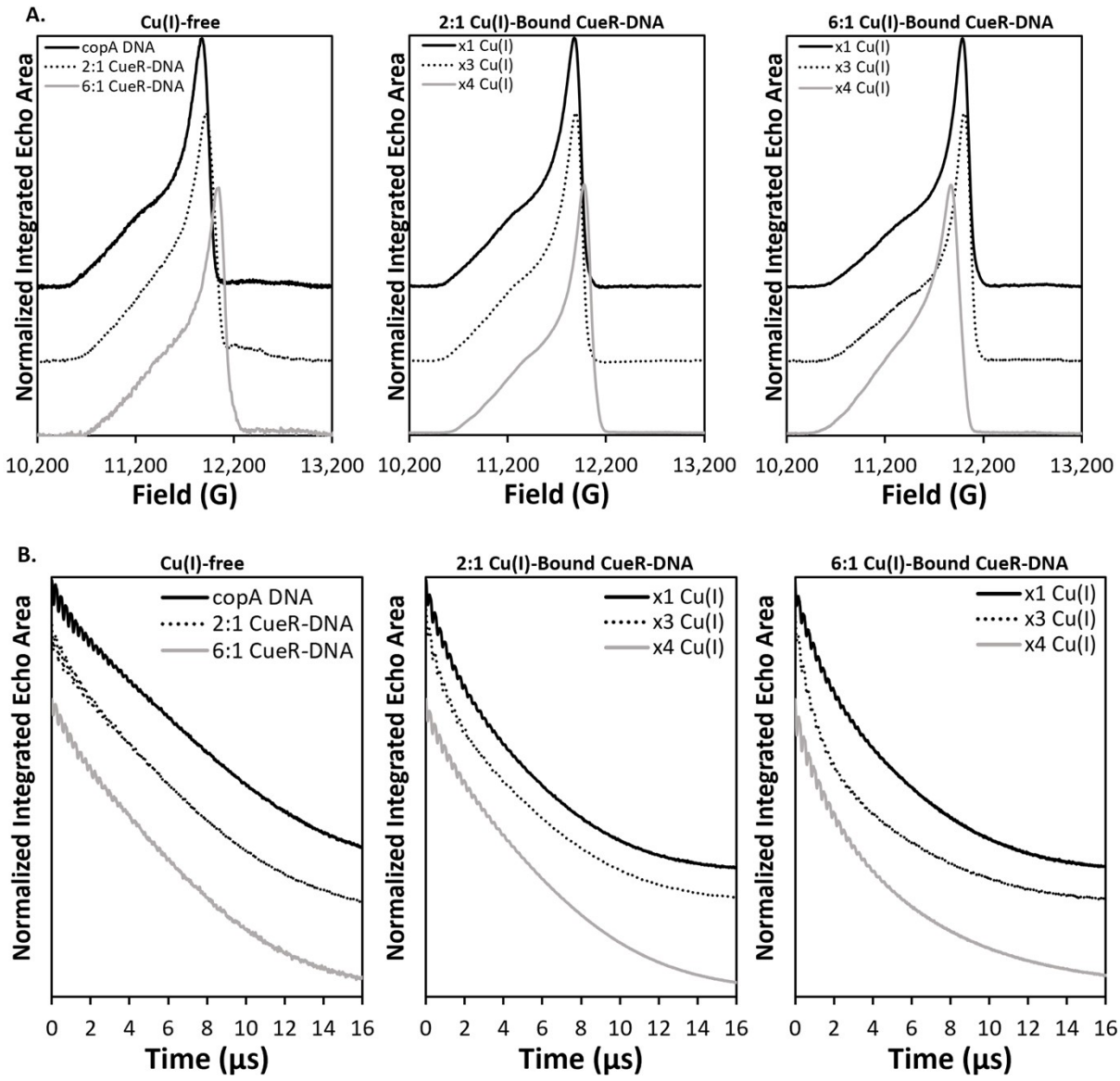

Figure S6. A) Field sweeps collected for each sample type as shown by integrating the area of the Hahn echo as a function of field positions. Field sweeps are shifted along the y-axis for ease of comparison. B) Echo decays for each sample type as shown collected by integrating the area of the Hahn echo as a function of the time between the π/2 and π pulses. Echo decays are shifted along the y-axis for ease of comparison. Modulations in the echo decays are due to the presence of deuterium in solution.

First, the copA DNA and Cu(I)-free CueR-DNA Trial 1 samples were collected using protonated solvent and glycerol. To increase the feasible acquisition times, deuterated solvent and glycerol were used for Trial 2. As such, the protonated and deuterated samples' time trace modulations are in good agreement. Due to the cost and challenging time constraints for long acquisition time, low temperature, and low Cu(II) concentration DEER experiments, the 6:1 Cu(I)-free CueR-DNA replicate (Trial 1) was not run extensively. However, both 6:1 Cu(I)-free CueR-DNA samples show similar time traces for both trials.

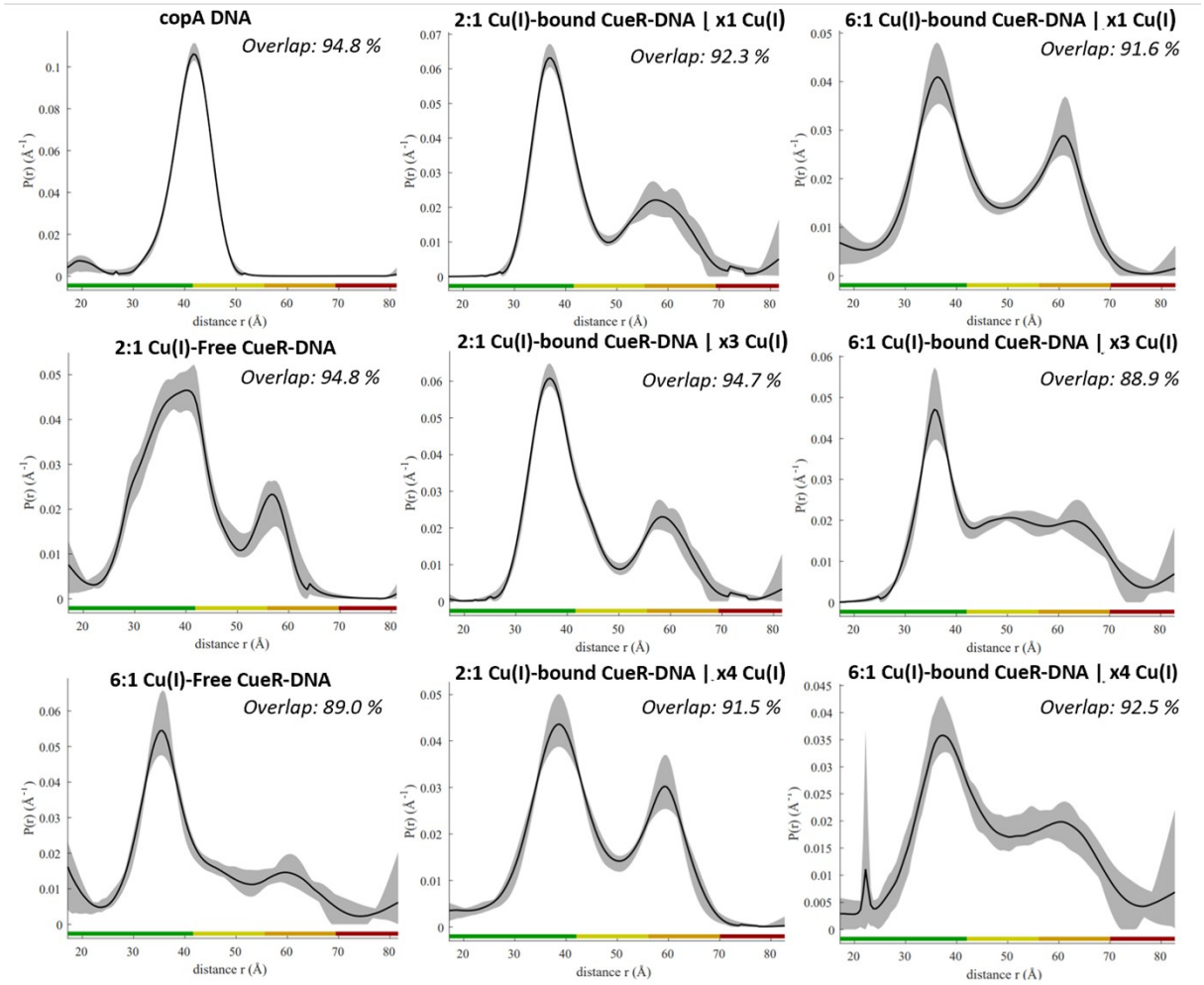

Figure S7. User-independent consensus DEER distance distributions obtained from ComparativeDEERAnalyzer (CDA) software from each sample's DEER time trace as shown. Grey regions are distance uncertainty estimates. The overlap percentage shown in each graph refers to the distance distribution overlap from both automated Tikhonov regularization and neural networking time trace analysis methods.

DeerAnalysis is a two-step analysis method that requires manually input choices for background subtraction and time trace fitting to obtain distance distributions. As such, there is potential for latent user bias in the analysis. Therefore, we also analyzed each time trace using ComparativeDEERAnalyzer (CDA) for a comparison of analysis methods. CDA is an automated program that generates a consensus distance distribution and uncertainty estimate from both DEERNet[7] and DEERLab[22]. DEERNet is an automated fitting program that utilizes a deep neural network and automatically analyzes the contribution to the baseline from intermolecular dipolar

interactions, which removes potential user bias from the analysis. DEERLab is a single step automated fitting program that utilizes Tikhonov regularization, removing all user bias. Each time trace was analyzed with CDA Generally, the returned consensus distance distributions from each CDA analysis resemble distance distributions obtained by DeerAnalysis shown in the main text (Figures 2C and 3B), but there are slight variations in the breadth and populations of the distributions. However, the observed changes in DNA distances as CueR and Cu(I) are introduced into the system on which the conclusions are based are consistent regardless of the DEER analysis program used.

Moreover, in the main text we state that the increasing modulation depth from 1 % in the Cu(I)-free CueR-DNA up to 9 % in the x4 Cu(I)-bound CueR-DNA samples is due to the formation of high order binding complexes. Specifically, two CueR-DNA complexes come together in response to high CueR concentrations and the presence of Cu(I) to accelerate DNA recognition to readily initiate transcription. As the fraction of the high order complex in solution increases with respect to CueR and Cu(I) concentration, the modulation depth also increases. The population of the distance distributions larger than ~ 5 nm in Figure S6 do increase in either intensity or breadth as CueR or Cu(I) concentration increases. However, due to the acquisition time of the DEER time traces collected, confident and quantitative interpretations of the larger distributions cannot be made yet.

Bending Model:

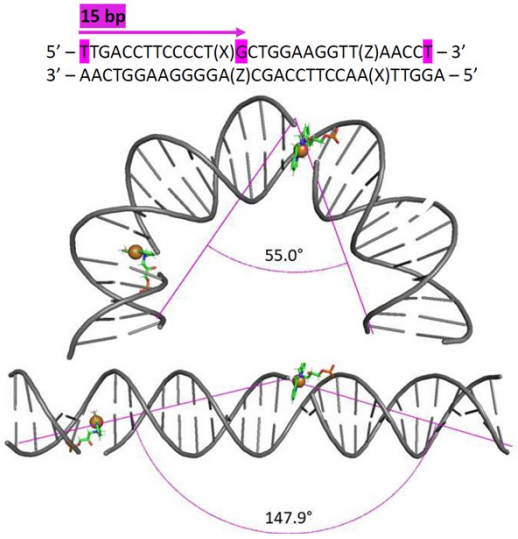

Figure S8. Bending model of the copA DNA duplex. Our model uses the 15th nucleotide from the 5' end to act as the angle origin. Angles were measured from the C4' atoms of the 5' to 3' end on the same strand as our DPA moiety. The top duplex was bent until the Cu(II)-Cu(II) distance was 3.6 nm, based on the EPR data. Bottom is the linear undistorted duplex with the same angle measurement.

The bent structure of DNA was created using the 3D-DART software[23]. DPA-Cu(II) sites were manually added into the duplex using PyMOL. The DNA angle was collected by choosing three points on the structure and measuring the center angle. As seen, upon bending the centered angle shifts from 148° for a linear copA duplex to 55°.

DEER Data Comparisons to Crystal Structures

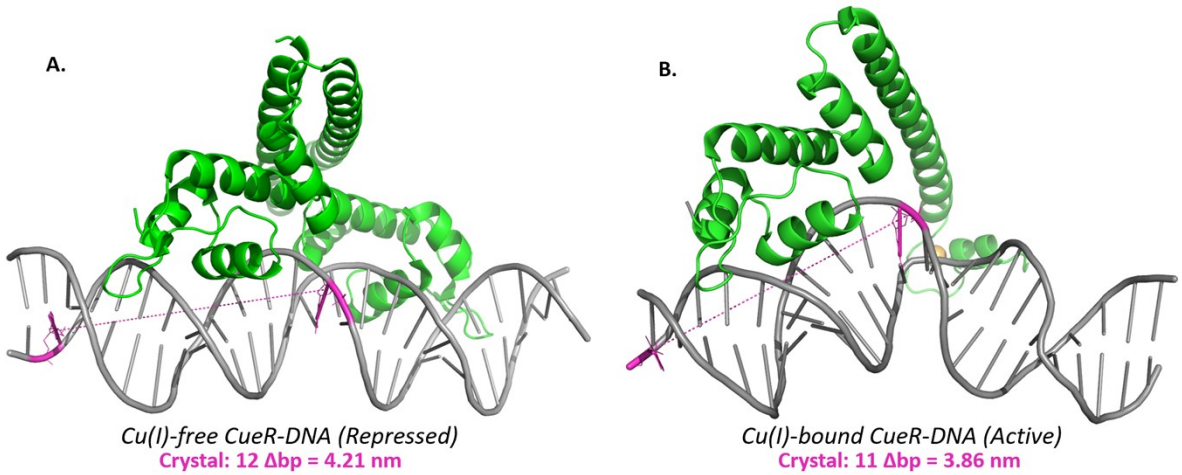

Figure S9. A) Cu(I)-free CueR-DNA (repressed state) crystal structure (PDB: 4WLS). The 12 bp distance measurement, taken from equivalent sites of our DPA moieties, is 4.21 nm. B) Cu(I)-bound CueR-DNA (active state). The 11 bp distance measurement, taken from the equivalent sites of our DPA moieties, is 3.86 nm. The crystal structure did not contain a long enough DNA sequence for us to compare exactly to the 12 bp separation of our DPA-DNA sites.

An equivalent 12 bp measurement on the repressed crystal structure[24] (Figure S7A) to our DPA sites and their respective separation is 4.21 nm. The measurement was taken from C4'-C4' since the C' backbone atom of DPA most closely resembles the C4' site when viewing the duplex from the 5' end. A 4.1 nm distance agrees well with our unbound undistorted copA DNA measurements, although it differs significantly from the observed CueR-DNA 3.6 nm distance for 6:1 protein:DNA samples. For the active CueR crystal structure[24], an 11 bp C4'-C4' separation starting from the equivalent site of our label placed between the CueR promoter regions and the 5' end is 3.86 nm. The crystal structure did not contain a long enough copA DNA sequence for us to compare exactly to the 12 bp separation of our DPA-DNA sites. Collected DEER constraints for all active state samples have most probable distances centered between 3.7-3.8 nm. Therefore, our data is in good agreement with the crystal structure. However, theoretically, the presence of an additional base pair in the crystal structure distance measurement for an exact 12 bp separation comparison between methods would result in a crystal structure distance greater than 3.8 nm. As such, we infer that our DEER results indicate a higher degree of kinking in the active state.

References:

(1) Sameach, H.; Narunsky, A.; Azoulay-Ginsburg, S.; Gevorkyan-Aiapetov, L.; Zehavi, Y.; Moskovitz, Y.; Juven-Gershon, T.; Ben-Tal, N.; Ruthstein, S. *Structure.* **2017**, *25*:7, 988–996.

(2) Hellman, L.; Fried, M. *Nat. Protoc*. **2007**, 2, 1849-1861.

(3) Singewald, L.; Wilkinson, J. A.; Saxena, S. *Bio-protocol*. **2021**, 11:24, e4258.

(4) Pornsuwan, S.; Bird, G.; Schafmeister, C. E.; Saxena, S. *J. Am. Chem. Soc*. **2006**, 128:12, 3876-3877.

(5) Stoll, S.; Schweiger, A. *J. Magn. Reson.* **2006**, *178:1*, 42–55.

(6) Tait, C. E.; Stoll, S. *Phys. Chem. Chem. Phys.* **2016**, *18*:27, 18470–18485.

(7) Worswick, S. G.; Spencer, J. A.; Jeschke, G.; Kuprov, I. *Sci. Adv.* **2018**, *4*:8, 1-17.

(8) Jeschke, G.; Chechik, V.; Ionita, P.; Godt, A.; Zimmermann, H.; Banham, J.; Timmel, C. R.; Hilger, D.; Jung, H. *Appl. Magn. Reson.* **2006**, *30*, 473–498.

(9) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comp. Chem*. **2005**, 1668–1688.

(10) Delano, W. L. PyMOL: An Open-Source Molecular Graphics Tool. **2002**.

(11) Ghosh, S.; Casto, J.; Bogetti, X.; Arora, C.; Wang, J.; Saxena, S. *Phys. Chem. Chem. Phys.* **2020**, *22:46*, 26707–26719.

(12) Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; *Nat. Methods,* **2015**, *13:1*, 55–58.

(13) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79:2*, 926–935.

(14) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98:12*, 10089–10092.

(15) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*:3, 327–341.

(16) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14:1*, 33–38.

(17) Ghosh, S.; Lawless, M. J.; Brubaker, H. J.; Singewald, K.; Kurpiewski, M. R.; Jen-Jacobson, L.; Saxena, S. *Nucleic Acids Res.* **2020**, 48:9, 1–11

(18) Cunningham, T.F.; Putterman, M.R.; Desai, A.; Horne, W.S.; Saxena, S. *Angew. Chem. Int. Ed*. **2015**, 54:21, 6330-6334

(19) Lawless, M. J.; Gosh, S.; Cunningham, T. F.; Shimshi, A.; Saxena, S. *Phys. Chem. Chem. Phys.* **2017**, 19, 20959-20967.

(20) Ghosh, S.; Lawless, M. J.; Rule, G. S.; Saxena, S. *J. Magn. Reson.* **2018**, 286, 163-171.

(21) Gamble Jarvi, A.; Ranguelova, K.; Ghosh, S.; Weber, R.T.; Saxena, *S. J. Phys. Chem B*. **2018**, 122:47, 10669-10677

(22) Ibáñez, L. F.; Jeschke, G.; Stoll, S. *Magn. Reson*. **2020**, 1, 209-224.

(23) van Dijk, M.; Bonvin, A. M. J. J. *Nucleic Acids Res.* **2009**, *37*:2, W235–W239.

(24) Philips, S. J.; Canalizo-Hernandez, M.; Yildirim, I.; Schatz, G. C.; Mondragón, A.; O'Halloran, T. V. *Science.* **2015**, *349:6250*, 877–881.