

Experimental Validation of the ALLNOX Program for Studying Protein-Nucleic Acid Complexes

Yuan Ding^{1,†}, Venkatesan Kathiresan¹, Xiaojun Zhang¹, Ian S. Haworth², and Peter Z. Qin^{1,3,*}

¹Department of Chemistry, University of Southern California, Los Angeles, California, USA

²Department of Pharmacology and Pharmaceutical Sciences, University of Southern California, Los Angeles, California, USA

³Department of Biological Sciences, University of Southern California, Los Angeles, California, USA

[†]Current address: Illumina Inc., San Diego, California, USA

*Corresponding author: Peter Z. Qin, TRF 119, 3430 S. Vermont Ave., Los Angeles, CA 90089, USA; Tel: (213) 821-2461; Fax: (213) 740-2701; Email: pzq@usc.edu

Abstract

Measurement of distances between spectroscopic labels (e.g., spin labels, fluorophores) attached to specific sites of biomolecules is an important method for studying biomolecular complexes. ALLNOX ('Addition of Labels and Linkers') has been developed as a program to model inter-label distances based on an input macromolecule structure. Here, we report validation of ALLNOX using measured distances between nitroxide spin labels attached to specific sites of a protein-DNA complex. The results demonstrate that ALLNOX predicts average inter-spin distances that matched with values measured with pairs of labels attached at the protein, the DNA, or one at the protein and one at the DNA. This establishes a solid foundation for using spin labeling in conjunction with ALLNOX to investigate complexes without high-resolution structures. With its high degree of flexibility for the label or the target biomolecule, ALLNOX provides a useful tool for investigating the structure-function relationship in a large variety of biological molecules.

Introduction

Site-directed spin labeling (SDSL) is a biophysical technique that provides unique information on the structure and dynamics of biomacromolecules.¹⁻³ In SDSL, a spin label, which is a chemically stable radical, is attached to a specific site within a macromolecule. Information is derived by monitoring the behaviors of the spin label using electron paramagnetic resonance (EPR) spectroscopy. The most commonly used spin labels are nitroxide radicals,¹⁻³ although others such as trityl radicals⁴ and EPR-active metal ions (e.g., Gd^{3+} ,⁵ Cu^{2+} ,⁶) have gained interest recently. The distance between a pair of spin labels can be measured by determining the dipolar interaction strength between the unpaired electrons.¹⁻³ Inter-spin distances below 2 nm can be determined using continuous-wave (cw-) EPR,⁷ and with the development of pulsed EPR spectroscopy, the range has been expanded to beyond 10 nm.^{8, 9} The measured distances provide direct structural constraints, which are highly informative in investigating structures of the target molecules and conformational changes in different functional states.

For the majority of spin labels used, the unpaired electron resides at a functional group that is not native to the target macromolecule, and the measured inter-spin distances differ from the corresponding distances between the points of label attachment. A successful approach for correlating measured inter-spin distances to a given conformation of the target macromolecule is to model the spin label at the desired sites of a target structure, computationally identify the allowable ensemble of conformers, compute the expected inter-spin distances, and then compare the expected distances to those measured experimentally. Examples of the first generation of these programs include NASNOX, which computes expected distances for a nitroxide (designated as R5, Figure 1A) attached to DNA or RNA,^{10, 11} and MMM,¹² PRONOX,¹³ and MtsslWizard,¹⁴ which compute inter-label distances for a nitroxide (designated as R1) attached on proteins. These programs have been validated by studies on target molecules with reported X-ray or NMR structures,¹⁰ and have been successfully applied in a number of studies to investigate complexes with no reported high-resolution structures.^{15, 16}

These first generation programs can only handle one fixed label (i.e., R1 or R5) on a fixed type of biological molecule (i.e., protein or nucleic acids). To overcome these limitations, the ALLNOX ('Addition of Labels and Linkers') program was developed¹⁷ to permit addition of any label to any site on any target biomolecule. ALLNOX implements a "label-linker-target" strategy to model a desired spectroscopic label at a specific site of a biomacromolecule. A user may upload a "label"

moiety as a pdb file (e.g., pyrroline ring of a nitroxide, Figure 1B), define a “linker” (e.g., “-C-S-C-”, Figure 1B), then specify the “target” for attachment (e.g., “C α ” of an amino acid, Figure 1B). The program models the label onto the target macromolecule structure (as specified by an input pdb file), generates rotamers of the label by stepwise variations of the rotatable torsional angles in the linker, identifies the rotamers without steric collision with the target macromolecule, and then outputs the inter-label distances. ALLNOX can examine a variety of labels (different spin labels, fluorophores, etc.) directly on protein and/or nucleic acids.¹⁷ Such flexibility is highly beneficial for evaluation of experimental data in biochemical complexes, especially those containing multiple molecular types.

In this work, we report validation of ALLNOX using EPR measured distances obtained with a protein-DNA complex with reported crystal structures (Figure 1C). The target macromolecule is a complex of a DNA duplex and the DNA-binding domain of GATA3, which is a member of the mammalian GATA family of transcription factor specifically recognizing DNAs with a consensus sequence of “GATA”.^{18, 19} The DNA-binding domain of GATA3 contains two highly conserved zinc fingers, which are referred to as the N-terminal zinc finger (N-finger) and C-terminal zinc finger (C-finger) (Figure 1C & 1D). The C-finger and its adjacent basic region are necessary and sufficient for binding to the cognate consensus DNA sequence WGATAR (W=A/T, R=A/G).^{18, 19} On the other hand, the N-finger has been shown to participate in interactions with protein partners, but can also bind to DNA independently with a slightly different preferred sequence (i.e., GATC).^{20, 21} On certain DNA sequences with two proximal GATA sites, the N-finger can participate in DNA binding together with the C-finger, leading to a greatly increased binding affinity.^{22, 23}

Members of the GATA family have been characterized by both X-ray crystallography and NMR.²³⁻²⁷ In particular, the Chen lab reported crystal structures of full GATA3 DNA binding domain (Figure 1C) containing both zinc fingers and the adjacent basic regions (Figure 1D) bound to DNAs containing composite GATA sites (Figure 1E).²³ This provides an ideal case to test ALLNOX, as the complex contains different macromolecule types (protein and DNA) that require different labeling chemistry. In this report, we attached the R5 nitroxide label to DNA targets (Figure 1A), and the R5p nitroxide label to the protein fragment (Figure 1B). A pulsed EPR scheme, Double Electron Resonance (DEER),⁸ was employed to measure inter-spin distances between pairs of labels attached to protein, to DNA, and one at the protein and the other at the

DNA. In parallel, ALLNOX was used to predict expected inter-nitroxide distances using a crystal structure of a DNA-bound GATA3 fragment (pdb id 4HCA,²³ Figure 1C). In all data sets examined, ALLNOX predicted average distances within 0.15 nm (1.5 Å) of the measured value. The results validate the accuracy of ALLNOX, thus establishing its utility for SDSL studies.

Materials and Methods

Protein mutagenesis, expression and purification

To introduce a cysteine for R5p attachment at a desired site, mutagenesis was carried out using vector pET-28a (kanamycin resistant) encoding the “double-finger” GATA3 fragment (designated as “DF”, Figure 1D), which included human GATA3 amino acids 260-371 fused with a 6X His-tag.²³ Threonine 280 (N-finger) and isoleucine 362 (C-tail) were individually (i.e., constructs “T280” and “I362”, respectively) or simultaneously (i.e., construct “DM”) mutated to cysteine for subsequent spin label attachment. Site-directed mutagenesis was carried out using the Quikchange kit. Mutant plasmids with sequence verified were transformed into BL21(DE3)-Gold competent cell for protein expression. The overexpressed protein was purified using affinity chromatography (i.e., Ni-NTA) followed by cation exchange chromatography. Details of mutagenesis, protein expression and purification are described in Supplementary Information (SI) section S1.

Spin labeling of proteins

The R5p label (Figure 1B) is covalently attached to a cysteine with a thioether linkage, which is stable in the presence of certain reducing agents [e.g., TCEP (tris(2-carboxyethyl)phosphine)] used to maintain activities of most nucleic acid-binding proteins.²⁸ Each labeling reaction was carried out in a solution containing 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, pH 7.5), 500 mM NaCl, 0.5 μM zinc acetate, 0.5 mM TCEP, 0.5 mM 3-Iodomethyl-1-oxyl-2,2,5,5-tetramethylpyrroline (i.e., the “R5 precursor”, prepared as reported previously^{29, 30}], and 10-50 μM protein. The reaction mixture was incubated with constant shaking at 4°C in the dark overnight. After incubation, the unreacted precursor was removed using cation-exchange chromatography, and the purified labeled protein fractions were concentrated to desired volume. The efficiency of protein spin labeling was estimated using a spin counting procedure described previously,³¹ and was found to be ~ 75% for the T280, I362, and DM constructs (SI, section S2.2).

Spin labeling of DNAs

The R5 label (Figure 1A) was attached to a DNA strand following protocols published previously.^{29, 30} Briefly, all DNA oligonucleotides were synthesized by solid-phase chemical synthesis (Integrated DNA Technologies, Coralville, IA, USA), and the desired phosphate for spin labeling is chemically modified to a phosphorothioate in this step. The R5 precursor was then reacted with the phosphorothioate group in a solution containing ~0.5 mM crude DNA, 200 mM R5 precursor, 0.1 M 2-(N-morpholino)ethanesulfonic acid (MES) (pH 5.8), and 40% (v/v) acetonitrile. The reaction mixture was incubated in the dark at room temperature for 24 hours under constant shaking. Labeled DNA was purified by anion-exchange high-pressure liquid chromatography, followed by desalting using a reverse-phase column. Desalted oligonucleotides were lyophilized, re-suspended in water and stored at -20°C. The final concentrations of each DNA strand were determined by absorbance at 260 nm using extinction coefficients of: pX, 228,300 M⁻¹·cm⁻¹; pY, 210,100 M⁻¹·cm⁻¹; sX, 201,300 M⁻¹·cm⁻¹; and sY, 169,700 M⁻¹·cm⁻¹.

EPR sample assembly

DNA duplexes were prepared following an annealing protocol as reported previously.^{29, 32} Briefly, a duplex stock was prepared by annealing the R5-labeled strand with the complementary strand at a ratio of 1:1.1, or annealing two R5-labeled strands at a ratio of 1:1. After mixing appropriate amounts of the respective strands, the mixture was heated at 95°C for 1 minute, and then cooled down at room temperature for 1 minute. An appropriate amount of salt was then added to the mixture to reach a final concentration of 50 mM Tris-HCl (Tris(hydroxymethyl)aminomethane hydrochloride) (pH 7.5) and 100 mM NaCl. The solution was left standing at room temperature for >1 hour to allow duplex formation.

To assemble a GATA3 DF/DNA complex, appropriate amounts of DNA duplex and protein were mixed in a high salt solution containing 50 mM HEPES, pH 7.5, 500 mM NaCl, 0.5 μM zinc acetate, 0.5 mM TCEP, with the protein concentration ranging from 10-50 μM and the total solution volume being >500 μL. The ratio of protein/DNA varied depending on the spin-labeled construct. The mixture was then dialyzed against a low salt solution containing 50 mM HEPES (pH 7.5), 200 mM NaCl, 0.5 μM zinc acetate and 0.5 mM TCEP at 4°C overnight to allow DNA/protein binding. The dialyzed complex was concentrated to the desired volume with complex concentrations ranging from 200-300 μM.

DEER spectroscopy

Each final DEER sample contained 50 mM HEPES, pH 7.5, 200 mM NaCl, 0.5 μ M zinc acetate, and 40% (v/v) glycerol. The total volume was at least 25 μ L, with spin concentrations being $\sim$ 200 μ M. Samples were loaded into round glass capillaries (2.0 mm i.d. $\times$ 2.4 mm o.d., Vitrocom, Inc., Mountain Lakes, NJ) sealed at one end, flash-frozen in liquid nitrogen, and used immediately for DEER measurements.

DEER measurements were carried out at 78 K on a Bruker ELEXSYS E580 X-band spectrometer following previously reported procedure and parameters.¹⁵ Briefly, a dead time free four-pulse scheme⁸ was used, with the pump pulse frequency set at the center of the nitroxide spectrum and the observer frequency being approximately 70 MHz higher. The observer π pulse was 32 ns, and the pump π pulse was usually set at 16 ns. Accumulation time in each measurement ranged from 12 to 24 hours with 512 shots per point. Inter-spin distances were computed from the resulting dipolar echo evolution data using the Tikhonov regularization module in DeerAnalysis2013.³³ From the computed distance distribution, average distance ($\langle R_{\text{deer}} \rangle$) and width of distribution (σ_{deer}) of a selected population were calculated as reported previously.¹⁰

ALLNOX computation of inter-spin distances

Inter-spin distances were computed using the web-based ALLNOX program.¹⁷ A reported GATA3/DNA complex crystal structure (pdb id 4HCA.pdb,²³ Figure 1C) was formatted as described¹⁷ and used as the input structure. R5 labels attached to the DNA (Figure 1A) were modeled using the standard “R5-CS-nucP” Label–Linker–Target Assembly.¹⁷ R5p labels attached to the protein (Figure 1B) were modeled with a “Custom” Label–Linker–Target Assembly, with the “Label Selection” set as “R1(R5)”, the “Target Selection” set as “aaC α ”, and the “Linker” set as “CSC”.¹⁷ The “Conformation Mode” was set to “Automatic”, with each rotatable torsion angle (three for R5 and four for R5p, respectively) varied between 60°, 180°, and 300°. The clash criterion “vdW” was set at the default value of 0.75.

The ALLNOX output file contains a list of accepted and rejected nitroxide rotamers, as well as the ensemble of individual distances between each pair of allowable rotamers.¹⁷ Using the output distances between the nitrogen atoms of the nitroxide (i.e., “distance N – N”) as the

predicted distances, the average of the ensemble was computed as the predicted distance ($\langle R_{\text{nox}} \rangle$), and the standard deviation of the ensemble was computed as the width of the distribution (σ_{nox}).

Note that the current version of ALLNOX program only supports modeling of one R5 diastereomer,¹⁷ specifically the R5 attached to the “OP1” atom of the target nucleotide specified in the user input pdb file. In this work a scheme was developed to account for the presence of two R5 diastereomers at a DNA labeling site (Figure 1A) for distances measured between two DNA sites or between protein and DNA. First, the properly formatted 4HCA.pdb file was used as the input structure to obtain the result representing the S_p -R5 diastereomer. Then, a modified pdb file was generated by switching the designations of “OP1” and “OP2” in the target nucleotide while keeping the coordinates of the atoms unchanged. This modified pdb file was used as the input structure to obtain the result representing the R_p -R5 diastereomer. Inter-spin distances obtained from the S_p - and R_p -R5 were then combined together as the ensemble of ALLNOX predicted distances.

Results and Discussions

Validating ALLNOX predictions using a GATA3 protein-DNA complex crystal structure

To validate ALLNOX, spin labeling measurements were carried out on a complex with a reported crystal structure (pdb id 4HCA.pdb,²³ Figure 1C), which has the GATA3 double-finger fragment (i.e., “DF”, Figure 1D) bound to a DNA duplex containing palindromic GATA recognition sites separated by one base-pair (pGATA, with individual strands designated respectively as pX and pY) (Figure 1E). Four labeling sites were investigated in this work: T280 (N-finger) and I362 (C-tail) on the protein (Figures 1D and S1); and pX15 and pY15 on the DNA (Figure 1E). Based on the reported 4HCA crystal structure,²³ these sites are solvent accessible and located away from the protein/DNA interface, and therefore allow efficient spin labeling and minimize potential disruption of the protein-DNA interactions. Indeed, control studies using gel shift assays (SI, S2.1) and continuous-wave EPR spectroscopy (SI, S2.2) both confirmed that mutagenesis and spin-labeling did not interfere with formation of the protein-DNA complex, indicating that spin labeling did not significantly perturb folding and function of the GATA3 DF fragment.

A complex was first assembled using unlabeled pGATA DNA duplex and double-labeled GATA3 fragment, with R5p attached to T280 and I362. The inter-spin distance was measured

using DEER, with the observed echo decay profile clearly showing oscillations (Figure 2A). Analyses of the echo decay profile using the Tikhonov regularization approach yielded a distance distribution profile with the major population having an average distance $\langle R_{\text{deer}} \rangle = 3.07$ nm, with a width of distribution $\sigma_{\text{deer}} = 0.31$ nm (Figure 2B). Using the reported 4HCA crystal structure, R5p labels were modeled at T280 and I362, and the sterically allowed conformers were identified (Figure 2C). The analyses gave an averaged distance $\langle R_{\text{nox}} \rangle = 3.22$ nm and a width of distribution (σ_{nox}) of 0.25 nm (Figures 2D). The $\langle R_{\text{nox}} \rangle$ value differs from $\langle R_{\text{deer}} \rangle$ by only 0.15 nm (Table 1), which is below the 0.3 nm (3 Å) range of variation in the measured and predicted averaged distance.¹⁷ The predicted width of distribution is also comparable with the measured distribution (Table 2). As such, the T280-I362 measurement indicates that ALLNOX accurately models the distance between a pair of R5p labels attached to the protein within a protein/DNA complex. In ALLNOX, the protein prediction module is derived from the previously developed PRONOX program.¹³ For the R1 nitroxide label (linker “C-S-S-C β ”) used most frequently for proteins,¹ ALLNOX and PRONOX give similar results – control studies showed that for the T280-I362 pair, the predicted inter-R1 distance is 3.28 nm with ALLNOX and 3.26 nm with PRONOX. However, R1 cannot be used in the present experiments as the presence of the reducing agent breaks the “S-S” linker, while PRONOX cannot model the R5p label (linker C-S-C β , Figure 1B) because it only accepts R1. ALLNOX allows custom label definition, thus overcoming this limitation in PRONOX.

In addition to the protein module, ALLNOX has a module to handle labels attached to nucleic acids, which was built upon the previously validated NASNOX program.^{10, 11} To test the nucleic acid prediction functionality in ALLNOX, we attached a pair of R5 labels at the DNA (pX15 and pY15, Figure 1E), and measured the inter-spin distance within the DF/pGATA complex. The DEER measurement yield a major population with $\langle R_{\text{deer}} \rangle = 2.80$ nm (Figure 3A). ALLNOX gave a predicted distance $\langle R_{\text{nox}} \rangle = 2.74$ nm (Table 1). The difference of 0.06 nm is well within the 0.3 nm range of variability, and indicates that ALLNOX correctly predicted the DNA-DNA distance.

A critical step forward in the ALLNOX program is to predict inter-spin distances between protein and DNA, which is beyond the ability of either PRONOX or NASNOX. To test this functionality, we measured distances in the DF/pGATA complex with a R5p label attached at the protein fragment and a R5 label attached at the DNA duplex. Three DEER data sets were obtained, I362-pX15, I362-pY15, and T280-pY15, with the major populations giving $\langle R_{\text{deer}} \rangle$ of 2.90 nm,

3.60 nm, and 3.32 nm, respectively (Figure 3B). The corresponding ALLNOX predicted $\langle R_{\text{nox}} \rangle$ values were 2.95 nm, 3.64 nm, and 3.23 nm, respectively (Table 1). The differences between $\langle R_{\text{deer}} \rangle$ and $\langle R_{\text{nox}} \rangle$ range between 0.03 – 0.09 nm, well within the 0.3 nm range of variation. In addition, the predicted widths of distribution (σ_{nox}) are comparable with the measured distribution (Table 2). As such, the analyses demonstrated that ALLNOX is able to accurately predict inter-spin distances between protein and DNA.

We also note that the measured distance distribution profiles presented multiple populations (Figures 2 and 3). For each dataset, the “major” population with the highest occupancy (i.e., highest probability) correlated nicely with the values predicted by ALLNOX (Table 1). Additional analyses were carried out on the minor population(s) using the DEERconstruct program developed by the Fanucci group³⁴ to determine whether the population was “insignificant” (i.e., representing artifacts arising from DEER data analyses as constrained by the experimental setup and signal-to-noise ratio) or “significant” (i.e., representing additional distances presented in the sample) (SI, section S3). The DEERconstruct analyses indicated that the minor populations presented in the protein-protein measurement (Figure 2) and protein-DNA measurements (Figure 3B) were “significant” (SI, Table S1), indicating additional distances in the samples measured. These minor distance populations likely represent conformations of the DF/pGATA complex that differ from the 4HCA crystal structure, although the details of these conformations are not known, and it is not clear whether these conformations are related to other DF structures obtained in crystallography studies.²³

GATA 3 C-finger binds to a single DNA site in a similar fashion to that with pGATA

As an application of ALLNOX, we examined the solution conformation of the DF fragment bound to a DNA duplex containing only one “GATA” sequence (designated as sGATA, Figure 4A). Previous studies indicate that the GATA3 C-finger prefers a WGATAR (W=A/T; R=A/G) sequence while the N-finger favors GATC.^{19,21} Thus, it is expected that the sGATA construct will engage the C-finger of the DF fragment. In addition, multiple crystal or NMR structures have been published on the local binding mode between C-finger and GATA DNA.²³⁻²⁷ Although the DNA sequence and the protein construct used in those structures vary, the interface remains highly conserved. As such, one would expect that the conformation of C-finger bound to one GATA site would be highly similar to that observed in the crystal structures of the double-finger protein.

Two sites at sGATA were labeled, one at a time, with R5 (sX12 and sY18, Figure 4A). Complexes were assembled with the single R5-labeled DNA and a DF fragment with an R5p attached at I362, and the inter-spin interaction was measured by DEER. The data set for I362-sX12 showed a clear oscillation in the echo evolution trace, and the resulting distance distribution showed one dominant distance population with $\langle R_{\text{deer}} \rangle = 2.89$ nm (Figure 4B). sX12, which is located between T and A on the GATA site, is the equivalent site of pX15 in the 4HCA crystal structure (Figures 1E, 4C). This measured I362-sX12 distance is the same as that of I362-pX15 measured with the pGATA DNA (Figure 3B), and differs by 0.06 nm from the ALLNOX predicted value obtained from equivalent sites within the 4HCA crystal structure (“I362-pX15”, 2.95 nm, Table 1). This supports the notion that the C-finger binds to the single GATA construct with a similar conformation as that observed in the crystal structure of the complex with palindromic GATA sites.

A second DEER data set was obtained between I362 and sY18, which yielded a major population with $\langle R_{\text{deer}} \rangle = 3.57$ nm (Figure 4B). sY18, located eight base-pairs away from sX12 on the complementary strand, is equivalent to pY17 in the 4HCA crystal structure (Figures 1E, 4C). Using the 4HCA structure as input, ALLNOX predicted $\langle R_{\text{nox}} \rangle = 3.66$ nm for I362-sY18. This very nice agreement between $\langle R_{\text{deer}} \rangle$ and $\langle R_{\text{nox}} \rangle$ further supports the conclusion that the C-finger binds with a similar conformation to DNA containing single or composite GATA sites.

We also note that crystal structures have been reported that show the C-finger of the DF fragment binds to a GATA site at one DNA while the N-finger binds to another site at a separate DNA.²³ In the sGATA DNA experiments, the DNA:protein ratio was 1:1, and the C-finger is expected to bind to the single GATA site, consequently excluding the N-finger from specific binding to the DNA. DEER measurements in complexes assembled with R5p labeled at T280 of the DF fragment (located at the N-finger) and R5-labeled sGATA DNA revealed dipolar interactions between the spin labels, but the resulting distance distribution profile did not reveal a well-defined distance population (SI, section S4). This indicates that the N-finger is located in the proximity of the sGATA DNA, but does not engage the DNA in a specific conformation.

Conclusions

Studies of the GATA3 protein/DNA complex demonstrated that ALLNOX predicts average inter-spin distances that match with values measured between sites on protein, on DNA, and

between protein and DNA. This establishes a solid foundation for using SDSL in conjunction with ALLNOX to investigate complexes without high-resolution structures, such as that of the double-finger GATA3 protein bound to a single GATA site. This is highly beneficial for SDSL studies on structure-function relationships in protein-nucleic acid complexes, such as the CRISPR-Cas systems being developed for genome editing.^{28, 35} Furthermore, while work reported here focused on spin labeling, ALLNOX allows great flexibility for addressing various labels, linkers, and targets, and should be valuable in integrative structural modeling based on information derived from spectroscopic labels (e.g., fluorophores, spin labels).

Supporting Information: Supporting Information is available for additional Methods and Data.

Acknowledgement:

We thank Dr. Lin Chen for advice on the GATA3 system, and Brian Sutch and Jaime Avila for assistance on the web-based ALLNOX modeling. Work reported was supported in part by grants from NSF (CHE-1213673; MCB-1818107; PZQ) and NIH (R01-GM124413; PZQ).

Table 1: Comparison of averaged distances.

Data Set		DEER Measured ^(a) (nm)	ALLNOX (nm)	Difference (nm)
Protein-Protein	T280 – I362	3.07	3.22	-0.15
DNA-DNA	pX15 – pY15	2.80	2.74	0.06
Protein-DNA	I362 – pX15	2.90	2.95	-0.05
	I362 – pY15	3.60	3.64	-0.04
	T280 – pY15	3.32	3.23	0.09

(a) Major population in the DEER measured distance distribution profile.

Table 2: Comparison of the widths of distance distributions.

Data Set		DEER Measured ^(a) (nm)	ALLNOX (nm)
Protein-Protein	T280 – I362	0.31	0.25
DNA-DNA	pX15 – pY15	0.35	0.25
Protein-DNA	I362 – pX15	0.28	0.20
	I362 – pY15	0.28	0.25
	T280 – pY15	0.24	0.34

(a) Major population in the DEER measured distance distribution profile.

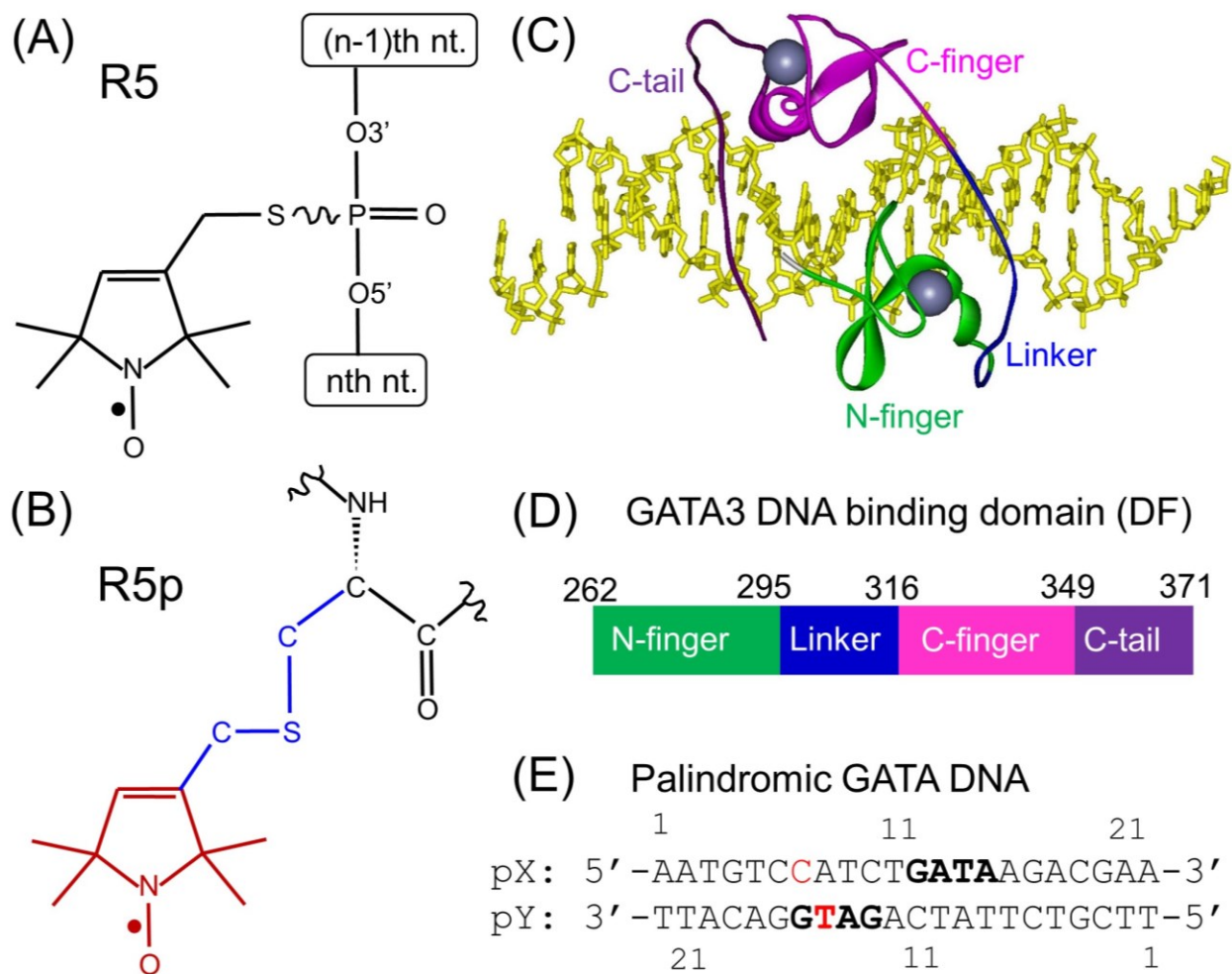


Figure 1: Constructs for spin-labeling studies on GATA3 protein/DNA complex. (A) Chemical structure of the R5 label on DNA. (B) Chemical structure of the R5p label on protein, with the “label” shown in red, “linker” in blue, and “target” in black. (C) A crystal structure (4HCA.pdb) with the “double-finger” fragment of GATA3 bound to a DNA with palindromic GATA sites. The DNA is shown as stick, protein is shown as ribbon with the domains colored, and the two zinc ions are shown in grey. (D) Schematic of the double-finger GATA3 DNA-binding domain fragment (i.e., “DF”). The two zinc-fingers and the adjacent basic regions are color-coded as that in panel (C). Detailed protein sequence is shown in Figure S1. (E) The pGATA DNA duplex. The “GATA” motifs are shown in bold, and the pX15 and pY15 nucleotides used for R5 attachment are shown in red.

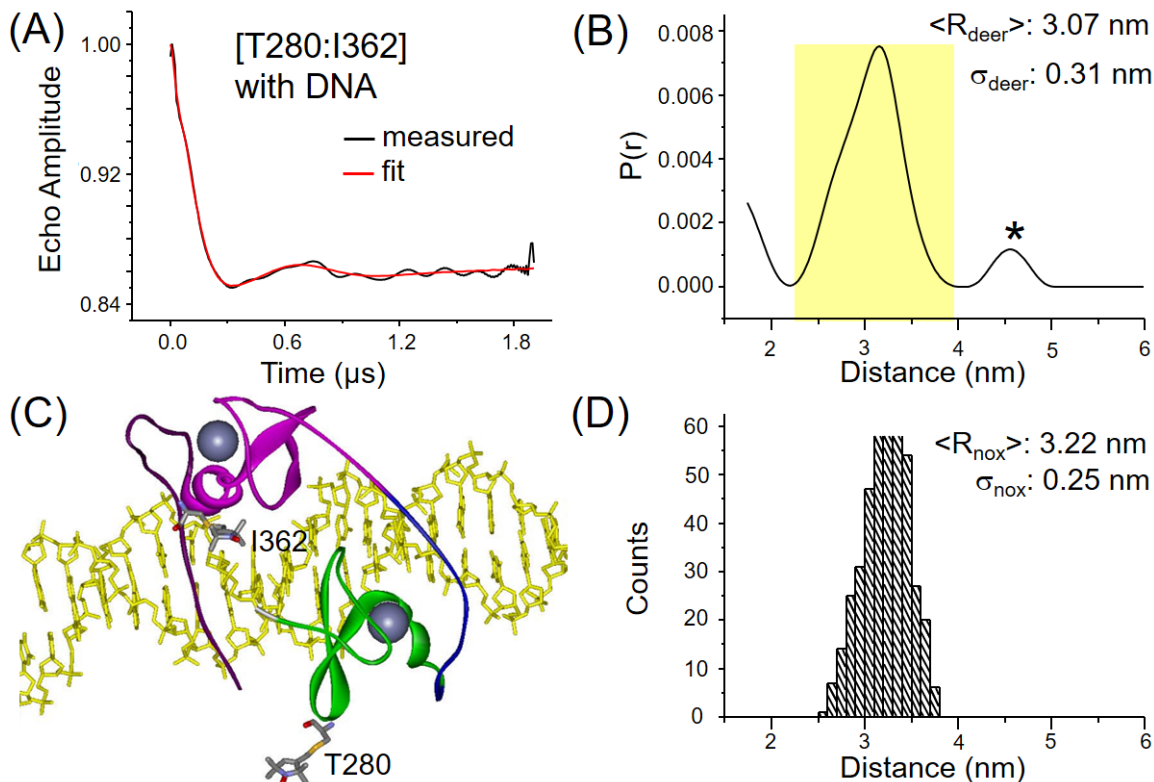


Figure 2: Assessment of the inter-spin distance between a pair of R5p attached at T280 and I362 of the protein. (A) DEER measured background corrected echo evolution data on the DF/pGATA complex assembled with a protein/DNA ratio of 1:1.1. (B) Distance distribution profile computed from data shown in (A). The major population is shaded in yellow, and “*” indicates a minor population deemed “significant” based on DEERconstruct analyses (see SI section S3). (C) ALLNOX generated structure of the spin-labeled DF/pGATA complex. The structure is depicted with the same scheme as described in Figure 1C. At the T280 and I362 site, one of the allowable R5p conformers is shown in stick. (D) Histogram of ALLNOX predicted distance distribution.

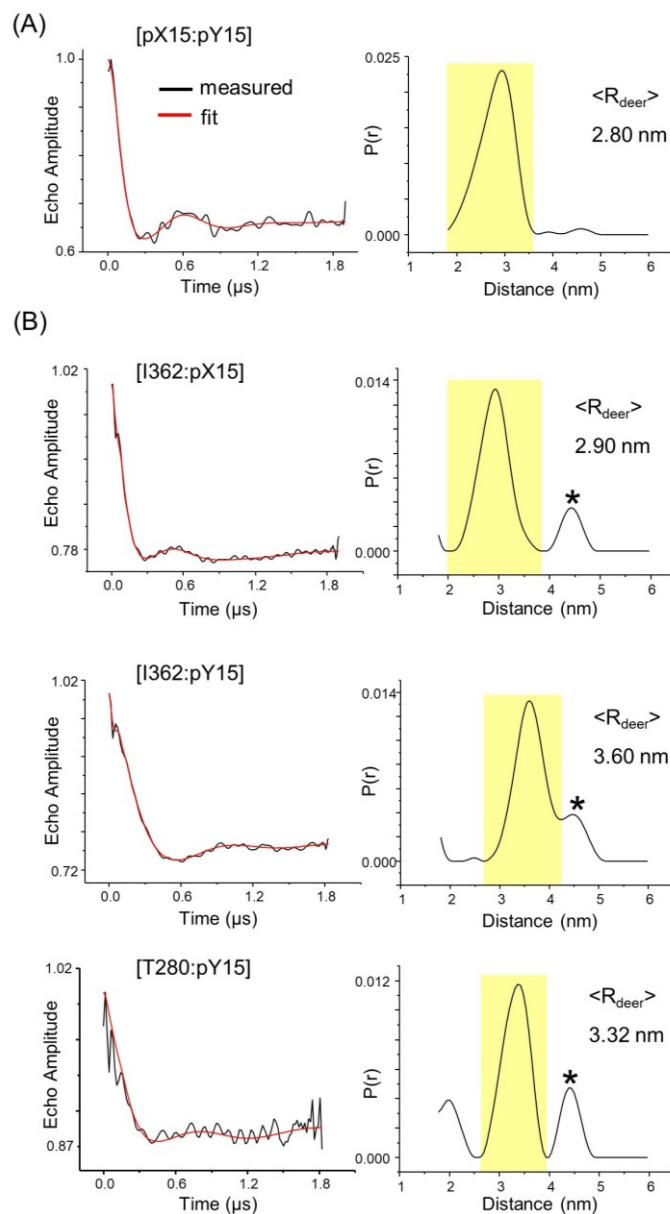


Figure 3: DEER data obtained in the DF/pGATA complex. (A) A pair of R5 labels at DNA pX15 and pY15 sites. To ensure all spin-labeled DNA was bound, the complex was assembled with a protein/DNA ratio of 1.1:1. (B) A R5p label at the protein and a R5 label at the DNA. Complexes were assembled with a protein/DNA ratio of 1:1. In each panel, shown on the left is the background correct echo evolution (black) with the fit trace (red) overlaid; and on the right is the distance distribution profile with the major population shaded in yellow, and “*” indicates a minor population deemed “significant” based on DEERconstruct analyses (see SI section S3).

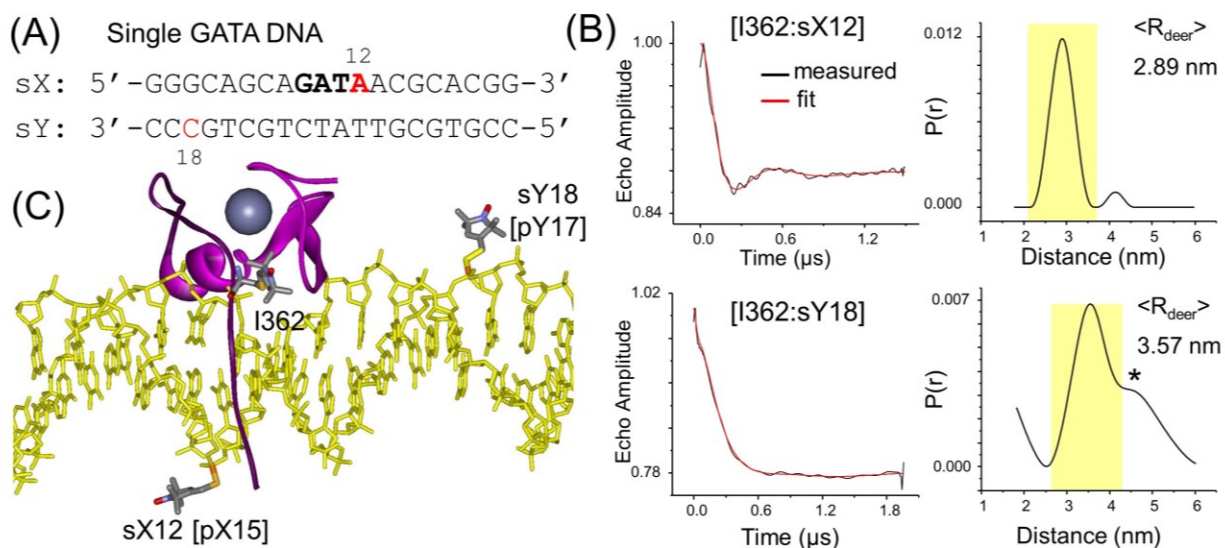
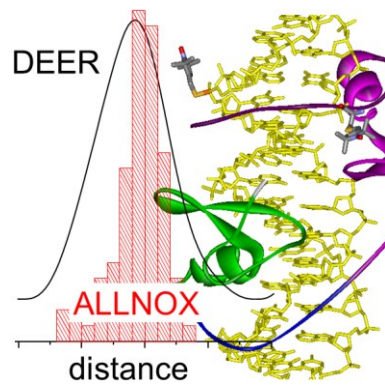


Figure 4: GATA3 double-finger fragment binding to a DNA with a single GATA site. (A) The sGATA DNA duplex. The “GATA” motif is in bold, and the R5 labeling sites are in red. (B) DEER result on I362-sX12 (top) and I362-sY18 (bottom). The complexes were assembled with a protein/DNA ratio of 1:1. In each dataset, shown on the left is the background correct echo evolution (black) with the fit trace (red) overlaid; and on the right is the distance distribution profile with the major population shaded in yellow, and “*” indicates a minor population deemed “significant” based on DEERconstruct analyses (see SI section S3). (C) ALLNOX generated spin-labeled complex structure. The structure is depicted with the same scheme as described in Figure 1C. One each of the allowable spin label conformers is shown at I362, sX12, and sY18. To mimic C-finger binding to one GATA site, the N-finger and the linker of GATA3 are not shown.

TOC graphic



References

1. Hubbell, W. L.; Altenbach, C., Investigation of structure and dynamics in membrane proteins using site-directed spin labeling. *Curr. Opin. Struct. Biol.* **1994**, *4*, 566-573.
2. Ding, Y.; Nguyen, P.; Tangprasertchai, N. S.; Reyes, C. V.; Zhang, X.; Qin, P. Z., Nucleic acid structure and dynamics: perspectives from site-directed spin labeling. In *Electron Paramagnetic Resonance*, The Royal Society of Chemistry: 2015; Vol. 24, 122-147.
3. Sahu, I. D.; Lorigan, G. A., Site-directed spin labeling EPR for studying membrane proteins. *Biomed Res Int.* **2018**, *2018*, 13.
4. Bagryanskaya, E. G.; Krumkacheva, O. A.; Fedin, M. V.; Marque, S. R. A., Development and application of spin traps, spin probes, and spin labels. In *Methods Enzymol.*, Qin, P. Z.; Warncke, K., Eds. Academic Press: 2015; Vol. 563, 365-396.
5. Feintuch, A.; Otting, G.; Goldfarb, D., Gd³⁺ spin labeling for measuring distances in biomacromolecules: why and how? In *Methods Enzymol.*, Qin, P. Z.; Warncke, K., Eds. Academic Press: 2015; Vol. 563, 415-457.
6. Yang, Z.; Ji, M.; Cunningham, T. F.; Saxena, S., Cu²⁺ as an ESR probe of protein structure and function. In *Methods Enzymol.*, Qin, P. Z.; Warncke, K., Eds. Academic Press: 2015; Vol. 563, 459-481.
7. Altenbach, C.; Oh, K. J.; Trabanino, R. J.; Hideg, K.; Hubbell, W. L., Estimation of inter-residue distances in spin labeled proteins at physiological temperatures: experimental strategies and practical limitations. *Biochemistry* **2001**, *40*, 15471-82.
8. Pannier, M.; Veit, S.; Godt, A.; Jeschke, G.; Spiess, H. W., Dead-time free measurement of dipole-dipole interactions between electron spins. *J Magn Reson* **2000**, *142*, 331-40.
9. Schmidt, T.; Wälti, M. A.; Baber, J. L.; Hustedt, E. J.; Clore, G. M., Long distance measurements up to 160 Å in the GroEL tetradecamer using Q-band DEER EPR spectroscopy. *Angew. Chem. Int. Ed.* **2016**, *128*, 16137-16141.
10. Cai, Q.; Kusnetzow, A. K.; Hubbell, W. L.; Haworth, I. S.; Gacho, G. P.; Van Eps, N.; Hideg, K.; Chambers, E. J.; Qin, P. Z., Site-directed spin labeling measurements of nanometer distances in nucleic acids using a sequence-independent nitroxide probe. *Nucl. Acids Res.* **2006**, *34*, 4722-30.
11. Price, E. A.; Sutch, B. T.; Cai, Q.; Qin, P. Z.; Haworth, I. S., Computation of nitroxide-nitroxide distances for spin-labeled DNA duplexes. *Biopolymers* **2007**, *87*, 40-50.

12. Polyhach, Y.; Bordignon, E.; Jeschke, G., Rotamer libraries of spin labelled cysteines for protein studies. *PCCP* **2011**, *13*, 2356-2366.
13. Hatmal, M. m. M.; Li, Y.; Hegde, B. G.; Hegde, P. B.; Jao, C. C.; Langen, R.; Haworth, I. S., Computer modeling of nitroxide spin labels on proteins. *Biopolymers* **2012**, *97*, 35-44.
14. Hagelueken, G.; Ward, R.; Naismith, J. H.; Schiemann, O., MtsslWizard: In Silico spin-labeling and generation of distance distributions in PyMOL. *Appl. Magn. Reson.* **2012**, *42*, 377-391.
15. Zhang, X.; Tung, C.-S.; Sowa, G. Z.; Hatmal, M. m. M.; Haworth, I. S.; Qin, P. Z., Global structure of a three-way junction in a phi29 packaging RNA dimer determined using site-directed spin labeling. *J. Am. Chem. Soc.* **2012**, *134*, 2644–2652.
16. Zhang, X.; Dantas Machado, A. C.; Ding, Y.; Chen, Y.; Lu, Y.; Duan, Y.; Tham, K. W.; Chen, L.; Rohs, R.; Qin, P. Z., Conformations of p53 response elements in solution deduced using site-directed spin labeling and Monte Carlo sampling. *Nucl. Acids Res.* **2014**, *42*, 2789-2797.
17. Beasley, K. N.; Sutch, B. T.; Hatmal, M. m. M.; Langen, R.; Qin, P. Z.; Haworth, I. S., Computer modeling of spin labels: NASNOX, PRONOX, and ALLNOX. In *Methods Enzymol.*, Qin, P. Z.; Warneke, K., Eds. Academic Press: 2015; Vol. 563, 569-593.
18. Ko, L. J.; Engel, J. D., DNA-binding specificities of the GATA transcription factor family. *Mol Cell Biol* **1993**, *13*, 4011-22.
19. Merika, M.; Orkin, S. H., DNA-binding specificity of GATA family transcription factors. *Mol Cell Biol* **1993**, *13*, 3999-4010.
20. Pedone, P. V.; Omichinski, J. G.; Nony, P.; Trainor, C.; Gronenborn, A. M.; Clore, G. M.; Felsenfeld, G., The N-terminal fingers of chicken GATA-2 and GATA-3 are independent sequence-specific DNA binding domains. *EMBO J* **1997**, *16*, 2874-82.
21. Newton, A.; Mackay, J.; Crossley, M., The N-terminal zinc finger of the erythroid transcription factor GATA-1 binds GATC motifs in DNA. *J Biol Chem* **2001**, *276*, 35794-801.
22. Trainor, C. D.; Omichinski, J. G.; Vandergon, T. L.; Gronenborn, A. M.; Clore, G. M.; Felsenfeld, G., A palindromic regulatory site within vertebrate GATA-1 promoters requires both zinc fingers of the GATA-1 DNA-binding domain for high-affinity interaction. *Mol Cell Biol* **1996**, *16*, 2238-47.

23. Chen, Y.; Bates, D. L.; Dey, R.; Chen, P. H.; Machado, A. C.; Laird-Offringa, I. A.; Rohs, R.; Chen, L., DNA binding by GATA transcription factor suggests mechanisms of DNA looping and long-range gene regulation. *Cell Rep* **2012**, *2*, 1197-206.
24. Bates, D. L.; Chen, Y.; Kim, G.; Guo, L.; Chen, L., Crystal structures of multiple GATA zinc fingers bound to DNA reveal new insights into DNA recognition and self-association by GATA. *J Mol Biol* **2008**, *381*, 1292-306.
25. Omichinski, J. G.; Clore, G. M.; Schaad, O.; Felsenfeld, G.; Trainor, C.; Appella, E.; Stahl, S. J.; Gronenborn, A. M., NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science* **1993**, *261*, 438-46.
26. Starich, M. R.; Wikstrom, M.; Arst, H. N., Jr.; Clore, G. M.; Gronenborn, A. M., The solution structure of a fungal AREA protein-DNA complex: an alternative binding mode for the basic carboxyl tail of GATA factors. *J Mol Biol* **1998**, *277*, 605-20.
27. Liew, C. K.; Simpson, R. J.; Kwan, A. H.; Crofts, L. A.; Loughlin, F. E.; Matthews, J. M.; Crossley, M.; Mackay, J. P., Zinc fingers as protein recognition motifs: structural basis for the GATA-1/friend of GATA interaction. *Proc Natl Acad Sci U S A* **2005**, *102*, 583-8.
28. Vazquez Reyes, C.; Tangprasertchai, N. S.; Yogesha, S. D.; Nguyen, R. H.; Zhang, X.; Rajan, R.; Qin, P. Z., Nucleic acid-dependent conformational changes in CRISPR-Cas9 revealed by site-directed spin labeling. *Cell Biochem Biophys* **2017**, *75*, 203-210.
29. Qin, P. Z.; Haworth, I. S.; Cai, Q.; Kusnetzow, A. K.; Grant, G. P.; Price, E. A.; Sowa, G. Z.; Popova, A.; Herreros, B.; He, H., Measuring nanometer distances in nucleic acids using a sequence-independent nitroxide probe. *Nat Protoc* **2007**, *2*, 2354-65.
30. Tangprasertchai, N. S.; Zhang, X.; Ding, Y.; Tham, K.; Rohs, R.; Haworth, I. S.; Qin, P. Z., An integrated spin-labeling/computational-modeling approach for mapping global structures of nucleic acids. In *Methods Enzymol*, Qin, P. Z.; Warncke, K., Eds. Academic Press: 2015; Vol. 564, 427-453.
31. Zhang, X.; Cekan, P.; Sigurdsson, S. T.; Qin, P. Z., Studying RNA using site-directed spin-labeling and continuous-wave electron paramagnetic resonance spectroscopy. In *Methods Enzymol.*, Herschlag, D., Ed. Academic Press: 2009; Vol. 469, 303-328.
32. Ding, Y.; Zhang, X.; Tham, K. W.; Qin, P. Z., Experimental mapping of DNA duplex shape enabled by global lineshape analyses of a nucleotide-independent nitroxide probe. *Nucl. Acids Res.* **2014**, *42*, e140.

33. Jeschke, G.; Chechik, V.; Ionita, P.; Godt, A.; Zimmermann, H.; Banham, J.; Timmel, C. R.; Hilger, D.; Jung, H., DeerAnalysis2006 - a comprehensive software package for analyzing pulsed ELDOR data. *Appl. Magn. Reson.* **2006**, *30*, 473-498.
34. Casey, T. M.; Fanucci, G. E., Spin labeling and Double Electron-Electron Resonance (DEER) to deconstruct conformational ensembles of HIV protease. In *Methods Enzymol.*, Qin, P. Z.; Warncke, K., Eds. Academic Press: 2015; Vol. 564, 153-187.
35. Tangprasertchai, N. S.; Di Felice, R.; Zhang, X.; Slaymaker, I. M.; Vazquez Reyes, C.; Jiang, W.; Rohs, R.; Qin, P. Z., CRISPR–Cas9 mediated DNA unwinding detected using site-directed spin labeling. *ACS Chem. Biol.* **2017**, *12*, 1489-1493.