Protein–DNA Interactions

DOI: 10.1002/anie.200803588

Electron Spin Resonance Shows Common Structural Features for Different Classes of *Eco*RI–DNA Complexes**

Katherine M. Stone, Jacqueline E. Townsend, Jessica Sarver, Paul J. Sapienza, Sunil Saxena, and Linda Jen-Jacobson**

Herein, we show that the *Eco*RI restriction endonuclease binds different classes of DNA sites in the same binding cleft. *Eco*RI generates widespread interest because it has an extraordinary sequence selectivity to carry out its function of cleaving incoming foreign DNA without causing potentially lethal cleavage of cellular DNA. For example, *Eco*RI binds to its correct recognition site GAATTC up to 90000-fold better than miscognate sites that have one incorrect base pair.[1,2] The circa 650 specific sites in the E. coli genome are protected from cleavage by double-strand methylation. About 21000 miscognate sites that are not methylated are still cleaved by the *Eco*RI with a second-order rate constant that is about 10^9-fold lower.[1,2] *Eco*RI forms only non-specific complexes, with no cleavage at sites that differ from GAATTC by two or more base pairs.[1,2]

To understand the source of such high specificity, it is necessary to determine how the structures of *Eco*RI complexes differ at specific, miscognate (5/6 bp match), and non-specific ($\leq$4/6 bp match) DNA sites. This effort is timely, given the extensive genetic, biochemical, and biophysical data on *Eco*RI.[1–9] Footprinting results[1] suggest that the three classes of complexes are "structurally" distinct, and thermodynamic profiles ($\Delta G°$, $\Delta H°$, $\Delta S°$, ΔC_p°)[3,4] suggest that the specific complex has more restricted conformational–vibrational mobility of the protein and the DNA. There are crystal structures of the free protein,[6] and the metal-free specific protein–DNA complex.[6,7] Miscognate and non-specific complexes, however, have not been readily accessible to crystallographic analysis. Indeed, for the circa 3600 known restriction endonucleases, there are currently 73 crystal structures of 38 distinct enzymes in complex with specific DNA. However, there are only 4 structures of miscognate or non-specific complexes in the protein data bank.[10] Herein, we demonstrate the utility of pulsed ESR distance measurements to shed light on miscognate and non-specific complexes.

Figure 1 shows the structure of the *Eco*RI-specific complex.[6,7] The protein contains a large, relatively rigid, and structured globular "main" domain and a smaller "arm" region. The protein arms are invisible in the free protein[6] but become ordered and enfold the DNA in the specific complex, where they play a role in modulating specificity.[2,4] Mutations R131C, S180C, and K249C-S180C were chosen based on the crystal structure.[6,7] These sites are solvent-accessible, and are therefore likely to spin-label with minimal perturbation to protein structure. Residues R131 and S180 lie in the inner and outer arms, respectively. The residue K249 is in the main domain, which has very restricted movement[6] and acts as a reference point. As the *Eco*RI is a 62 kDa homodimer, single cysteine mutations provide two sites for spin labeling, and double mutations provide four sites.

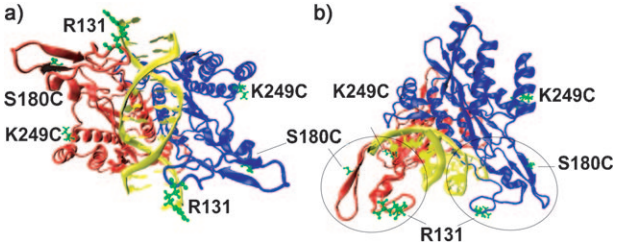

Figure 1. X-ray structure of the EcoRI specific complex. a) "Bottom" view, b) "side" view. Monomers are indicated in red and blue, the arm domain by circles, the DNA sequence in yellow, and the residues mutated to cysteine in green. Coordinates are determined from a highly refined version[6,7] of the protein data base (PDB) entry 1CKQ.

The proteins were spin-labeled at the cysteines with the methanethiosulfonate spin label (MTSSL). There is an intrinsic cysteine at position 218, but it is buried, leading to less than 10% labeling even with a 100-fold molar excess of the spin label. The mutant proteins and their spin-labeled derivatives catalyze DNA cleavage and have DNA binding affinities similar to that of the wild-type *Eco*RI, indicating that they are functionally active (see the Supporting Information).

The double electron–electron resonance (DEER) experiments[11] were performed on spin-labeled S180C specific and non-specific complexes, and on R131C and K249C-S180C specific, miscognate, and non-specific complexes. The DEER experiment is now well established for measuring distance constraints in membrane proteins,[12,13] soluble proteins,[14,15]

[*] K. M. Stone,[+] J. Sarver, Prof. S. Saxena
Department of Chemistry, University of Pittsburgh
219 Parkman Ave, Pittsburgh, PA 15260 (USA)
Fax: (+1) 412-624-8611
E-mail: sksaxena@pitt.edu
Homepage: http://www.chem.pitt.edu/saxena
J. E. Townsend,[+] P. J. Sapienza, Prof. L. Jen-Jacobson
Department of Biological Sciences, University of Pittsburgh
320 Clapp Hall, Pittsburgh, PA 15260 (USA)
Fax: (+1) 412-624-4759
E-mail: ljen@pitt.edu
[+] These authors contributed equally to this work.

[**] This work was supported by an NSF CAREER grant (MCB 0346898) to S.S. and an NIH MERIT 5R37M029207 grant to L.J.-J.
Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200803588.

peptides,[16] oligonucleotides,[17,18] and synthetic oligomers.[19–21] Recently, the DEER experiment has been used to probe structural rearrangements upon metal binding in the anthracis repressor, a DNA binding protein.[22]

The four-pulse DEER data for each of the mutant complexes are shown in Figure 2a. The time-domain signals

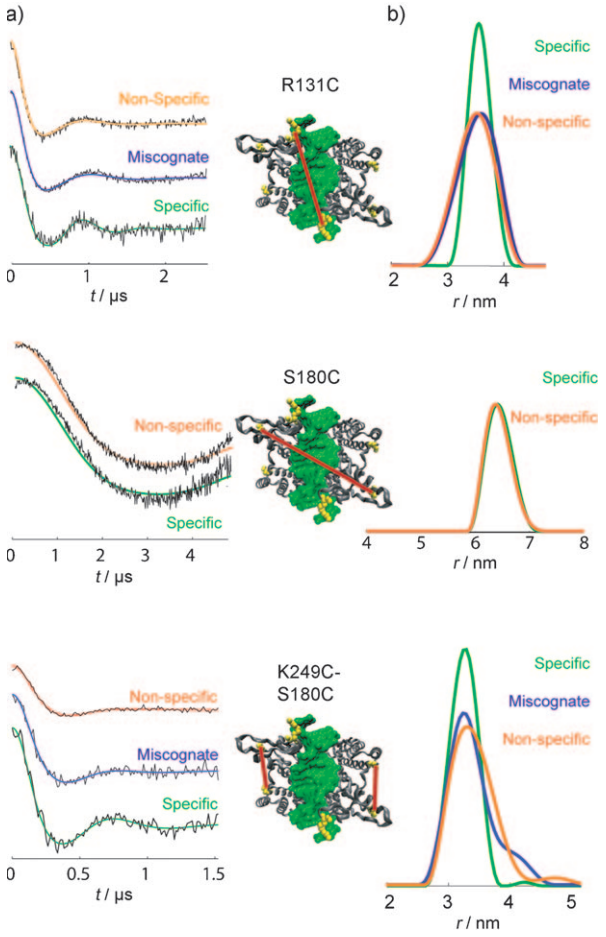

Figure 2. a) The double electron–electron resonance (DEER) data for DNA complexes with R131C, S180C, and K249C-S180C EcoRI mutants. Simulated traces based on the distance distributions shown on the right are overlaid on the experimental data. b) Normalized distance distribution functions. Red lines in the crystal structure indicate the distance measured. DNA sequences are: TCGCGAATTCGC (specific), TCGCAAATTCGC (miscognate), and GTGCCTTAAGCGCG (non-specific).

were inverted to obtain the distance distribution functions, using a Tikhonov regularization method in the DEERAnalysis2006 program.[23] The resulting distance distribution functions are shown in Figure 2b.

The most probable distances between the spin labels for the R131C *Eco*RI specific, miscognate, and non-specific complexes are 35, 36, and 35 Å, respectively. The R131 C_β–C_β distance in the crystal structure of the specific complex is 32 Å.[6,7] The interspin distance measured by ESR is expected to differ because of the added length of the spin label. The most probable distance for the S180C mutant in the specific and non-specific complexes is 64 Å. To enable measurement of such a large distance, a large volume of S180C in 30% deuterated glycerol, 65% deuterated water, and 5% protonated water was used, and the temperature was lowered to 40 K. With the enhanced signal and increased phase memory time (3 μs), a sufficiently long dipolar evolution time could be collected (Figure 2a, middle panel).

For specific, miscognate, and non-specific complexes of the K249C-S180C mutant protein, the most probable experimental distance was 33 Å in all cases (Figure 2b, lower). In principle, multiple distances corresponding to S180C-S180C, K249C-K249C, and S180C-K249C are anticipated for the K249-S180C double mutant. The corresponding C_β–C_β distances in the specific complex crystal structure are 27 Å (S180C-K249C intra-monomer), 59 Å (S180C-S180C), 60 Å (K249-K249), and 57 Å (S180-K249 inter-monomer).[6,7] It is likely that the larger distances were not detected in this series of experiments given that only about 1.5 μs of the data could be collected owing to short phase memory times. The 33 Å peak for the double mutant can thus be assigned to the S180C-K249C intra-monomer distance.

Strikingly, the experimental point-to-point distances are very similar for specific, and for non-cognate (i.e. miscognate and non-specific) EcoRI-DNA complexes. The data show preservation of the distances between the inner arms (R131C data), the outer arms (S180C data), and from the outer arm (S180C) to a fixed reference point (K249C) in the main domain. For both the R131C and the K249C-S180C mutant proteins, the distance distribution is narrower for the specific complex than for the corresponding non-cognate complexes. This might indicate a greater flexibility of the arms in the EcoRI complex with non-cognate DNA. Further ESR experiments that probe dynamics are underway to confirm this hypothesis. In addition, the distributions for both the R131C inter-arm distance and the K249C-S180C distance show asymmetries in the non-cognate complexes. However, it is unclear if this represents an asymmetric set of accessible conformations of the arms or different orientations accessible to the spin labels.

Taken together, the data suggest that on average, the EcoRI arms envelop the DNA and are similarly oriented in non-cognate and specific DNA complexes. This implies that the DNA in the specific and non-specific complexes occupies roughly the same binding cleft of the EcoRI dimer. In addition, slopes of the salt dependence for formation of specific and non-specific complexes are the same ($d \log K_A / d \log[\mathrm{NaCl}] \approx -11$)[24] and are consistent with the number of Coulombic interactions observed in the specific complex. This provides additional strong evidence that the arms enfold the DNA in the non-specific complex. This enfolding may contribute to processivity as the protein slides along non-specific DNA[8,9,25,26] to locate its specific recognition site. Our results on a DNA–protein complex by pulsed ESR establish a methodology that can measure the solution structure and range of conformational states for complexes with different classes of DNA sites for which there is little or no prior structural information.

Received: July 23, 2008
Published online: November 19, 2008

Keywords: DNA cleavage · EPR spectroscopy · non-cognate complexes · proteins · structure elucidation

[1] D. Lesser, M. Kurpiewski, L. Jen-Jacobson, *Science* **1990**, *250*, 776 – 786.
[2] P. Sapienza, C. dela Torre, W. McCoy, S. Jana, L. Jen-Jacobson, *J. Mol. Biol.* **2005**, *348*, 307 – 324.
[3] L. Jen-Jacobson, L. Engler, J. Ames, M. Kurplewski, A. Grigorescu, *Supramol. Chem.* **2000**, *12*, 143 – 160.
[4] P. Sapienza, J. Rosenberg, L. Jen-Jacobson, *Structure* **2007**, *15*, 1368 – 1382.
[5] M. Kurpiewski, L. Engler, L. Wozniak, A. Kobylanska, M. Koziolkiewicz, W. Stec, L. Jen-Jacobson, *Structure* **2004**, *12*, 1775 – 1788.
[6] A. Grigorescu, M. Horvath, P. Wilkosz, K. Chandrasekhar, J. Rosenberg, *Restriction Endonucleases*, Springer-Verlag, Heidelberg, **2004**.
[7] Y. Kim, J. Grable, R. Love, P. Greene, J. Rosenberg, *Science* **1990**, *249*, 1307 – 1309.
[8] A. Pingoud, M. Fuxreiter, V. Pingoud, W. Wende, *Cell. Mol. Life Sci.* **2005**, *62*, 685 – 707.
[9] D. Wright, W. Jack, P. Modrich, *J. Biol. Chem.* **1999**, *274*, 31896 – 31902.
[10] H. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. Shindyalov, P. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235.
[11] Rainer E. Martin, M. Pannier, F. Diederich, V. Gramlich, M. Hubrich, H. Spiess, *Angew. Chem.* **1998**, *110*, 2993 – 2998; *Angew. Chem. Int. Ed. Engl.* **1998**, *37*, 2833 – 2837.
[12] D. Hilger, Y. Polyhach, E. Padan, H. Jung, G. Jeschke, *Biophys. J.* **2007**, *93*, 3675 – 3683.
[13] I. Smirnova, V. Kasho, J.-Y. Choe, C. Altenbach, W. L. Hubbell, H. R. Kaback, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 16504 – 16509.
[14] M. Bennati, J. Robblee, V. Mugnaini, J. Stubbe, J. Freed, P. Borbat, *J. Am. Chem. Soc.* **2005**, *127*, 15014 – 15015.
[15] L. Galiano, M. Bonora, G. Fanucci, *J. Am. Chem. Soc.* **2007**, *129*, 11004 – 11005.
[16] A. Milov, R. I. Samoilova, Y. D. Tsvetkov, F. Formaggio, C. Toniolo, J. Raap, *J. Am. Chem. Soc.* **2007**, *129*, 9260 – 9261.
[17] O. Schiemann, N. Piton, Y. Mu, G. Stock, J. W. Engels, T. F. Prisner, *J. Am. Chem. Soc.* **2004**, *126*, 5722 – 5729.
[18] N. Piton, Y. Mu, G. Stock, T. Prisner, O. Schiemann, J. Engels, *Nucleic Acids Res.* **2007**, *35*, 3128 – 3143.
[19] S. Pornsuwan, G. Bird, C. Schafmeister, S. Saxena, *J. Am. Chem. Soc.* **2006**, *128*, 3876 – 3877.
[20] A. Godt, M. Schulte, H. Zimmermann, G. Jeschke, *Angew. Chem.* **2006**, *118*, 7722 – 7726; *Angew. Chem. Int. Ed.* **2006**, *45*, 7560 – 7564.
[21] B. Bode, D. Margraf, J. Plackmeyer, G. Durner, T. Prisner, O. Schiemann, *J. Am. Chem. Soc.* **2007**, *129*, 6736 – 6745.
[22] K. I. Sen, T. Logan, P. G. Fajer, *Biochemistry* **2007**, *46*, 11639 – 11649.
[23] G. Jeschke, A. Koch, U. Jonas, A. Godt, *J. Magn. Reson.* **2002**, *155*, 72 – 82.
[24] L. Jen-Jacobson, *Biopolymers* **1997**, *44*, 153 – 180.
[25] B. Terry, W. Jack, P. Modrich, *J. Biol. Chem.* **1985**, *260*, 13130 – 13137.
[26] S. Halford, J. Marko, *Nucleic Acids Res.* **2004**, *32*, 3040 – 3052.

