Bioorganic & Medicinal Chemistry Letters 19 (2009) 3779–3782

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

CSI–FID: High throughput label-free detection of DNA binding molecules

Karl E. Hauschild [a], James S. Stover [b], Dale L. Boger [b,*], Aseem Z. Ansari [a,*]

[a] The Genome Center of Wisconsin and the Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706, USA
[b] Department of Chemistry and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

ARTICLE INFO

Article history:
Received 20 February 2009
Revised 16 April 2009
Accepted 20 April 2009
Available online 3 May 2009

Keywords:
CSI microarray
FID
Small-molecule specificity
Label-free detection

ABSTRACT

Determining the sequence specifity of DNA binding molecules is a non-trivial task. Here we describe the development of a platform for assaying the sequence specificity of DNA ligands using label free detection on high density DNA microarrays. This is achieved by combining Cognate Site Identification (CSI) with Fluorescence Intercalation Displacement (FID) to create CSI–FID. We use the well-studied small molecule DNA ligand netropsin to develop this high throughput platform. Analysis of the DNA binding properties of protein- and small molecule-based libraries with CSI–FID will advance the development of genome-anchored molecules for therapeutic purposes.

A critical challenge at the interface of biology, chemistry, and molecular medicine is developing highly specific small molecules that target the genome to regulate its function.[1–10] A greater understanding of the principles that govern specificity will enhance our ability to predict their biological action on genomes, advancing the development of genome-anchored therapeutics. Similarly, understanding natural DNA binding proteins will help elucidate their regulatory function in cells. Given its importance, many methods have been developed to assess the binding of small molecules and proteins to DNA, including low throughput methods, such as nuclease protection[11,12], affinity cleavage[13], and electrophoretic mobility shift assays[14,15] (EMSAs), mid throughput assays including fluorescence anisotropy[16] and fluorescence resonance energy transfer[17], label free methods such as surface plasmon resonance (SPR)[18] and photonics based approaches[19], and high throughput assays including SELEX[20] and DNA microarrays.[21–24] Among these, two high throughput methods which can determine DNA binding specificity of biomolecules or synthetic ligands in a rapid and unbiased manner are the Cognate Site Identifier (CSI) arrays and the Fluorescent Intercalator Displacement (FID) assay (Fig 1A).[23,25,26] CSI arrays determine the specificity and affinity of DNA ligands using a microarray displaying double-stranded DNA (dsDNA) hairpin oligonucleotides containing all permutations of up to 12 positional variants (~2 million sequences). For CSI, a fluorescently labeled DNA ligand of interest is applied to the array to provide a distribution of intensities related to DNA binding affinity. Sequences with the highest intensities are evaluated to identify a consensus motif. The array data, therefore, provides the full-spectrum of binding specificities across the entire sequence space of a 12mer. The FID assay is a plate-based technique that measures the amount of ligand-induced displacement of an intercalated dye (commonly ethidium bromide, EtBr) from a DNA hairpin to determine the sequence preference of an unlabeled DNA binding ligand. The assay measures affinities in solution and provides a rapid means of measuring binding affinities (Fig 1A).

Both CSI and FID assays have been used successfully to determine the specificity and affinity of several DNA binding small molecules, as well as triplex forming oligonucleotides, proteins, and polyamides.[23,26–30] These methods offer complementary strengths toward the goal of understanding DNA ligand specificity. CSI can be used to interrogate the entire sequence space of at least 12mer DNA binding sites with a high dynamic range, and FID offers the benefit of label-free detection. Both approaches can also be used to determine DNA dissociation constants for the DNA ligand of interest.[23,25–28] However, with CSI the detection of DNA binding factors to the dsDNA array is dependent on either direct fluorescent labeling or indirect detection methods. This hinders its ability to analyze the DNA binding of unlabeled proteins and small molecules. Also, the labeling of small molecules or proteins with fluorescent tags may perturb their DNA binding properties. With FID, the major limitation is the DNA library size, with most assays being performed on all permutations of 5mer DNA (512 sequences). Larger DNA binding sites are possible; however, these typically only use a subset of the total library members to avoid an exponential increase in the cost of synthesis and purification of complex (>5mer) libraries. By combining CSI and FID we can overcome

* Corresponding authors. Tel.: +1 858 784 7522; fax: +1 858 784 7550 (D.L.B.); tel.: +1 608 265 4690; fax: +1 608 262 3453 (A.Z.A.).
E-mail addresses: boger@scripps.edu (D.L. Boger), ansari@biochem.wisc.edu (A.Z. Ansari).

0960-894X/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2009.04.097

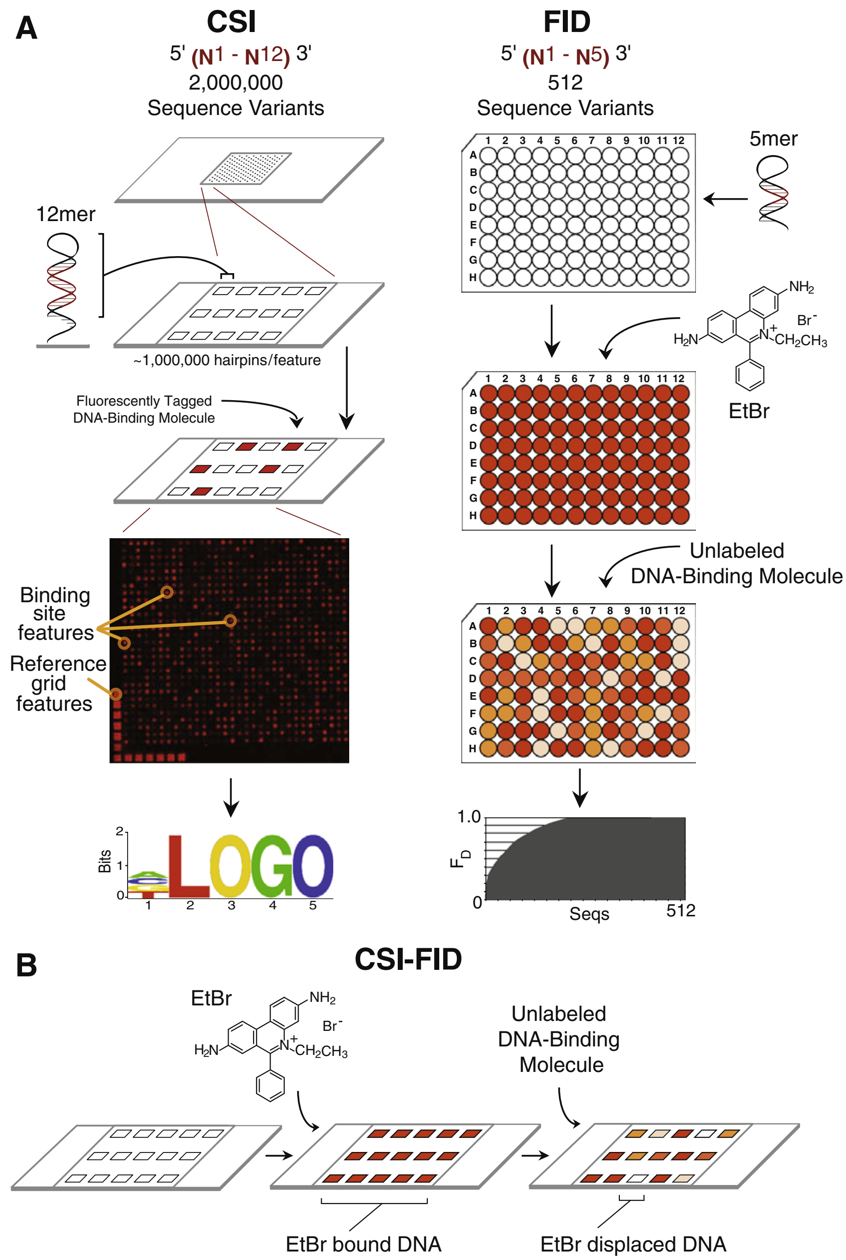

Figure 1. CSI–FID combines CSI and FID for label free detection on highly complex DNA libraries. (A) The left panel illustrates the CSI array technology. First a high density oligonucleotide array (up to two million sequence variants) is synthesized using maskless array synthesis. The oligos are hairpinned to form dsDNA, and a fluorescently tagged DNA-binding molecule is applied to the array. A readout of fluorescent intensities is then used to generate a sequence motif of the highest bound DNA sequences. The right panel shows the procedure for the FID assay. First a library of hairpinned DNA oligos are individually arrayed in 96 well plates, with one DNA hairpin sequence per well. EtBr is then added to each well and the fluorescence is measured. Next an unlabeled DNA molecule of interest is applied to the plate and the displacement of EtBr from the DNA is determined. The sequences with the highest EtBr displacement bind the molecule of interest with the highest affinity and these sites are used to generate a consensus motif for the molecule. (B) Depiction of FID being adapted for use on CSI arrays to generate the CSI–FID assay. First EtBr is incubated with the CSI array, then an unlabeled molecule of interest is applied to the array. The ratios of the EtBr intensity for the displaced versus undisplaced values for each DNA feature provides a measurement of the affinity of the small molecule for each DNA sequence. This technology measures the sequence specific binding for unlabeled DNA-binding molecules on CSI microarrays bearing highly complex DNA libraries.

the limitations of both by utilizing the label-free detection of ligand–DNA interactions on arrays with a 3–4 order of magnitude increase in DNA library complexity, which is the genesis of CSI–FID (Fig 1B).

Using CSI–FID, we examined the comprehensive binding profile of netropsin, a minor groove DNA binding small molecule, that exhibits antiviral and antitumor properties (Fig 2A).[31] Netropsin has served as a model for a class of sequence specific minor groove DNA binding agents[32,33], and previous FID assays have shown that it is an excellent candidate for assay validation of CSI–FID.[26]

Development and optimization of CSI–FID. Successful implementation of CSI–FID arrays required adapting the dye displacement ability of FID for a CSI microarray platform. The intercalating dye EtBr was chosen as it has several desirable properties; it increases in fluorescence upon DNA binding thereby decreasing background signal, equilibrates rapidly with DNA, and has low sequence specificity.[26,34] For the CSI–FID assay we designed an array which contained all possible nine base pair (bp) DNA permutations. This makes the low sequence specificity of EtBr an especially important property for the optimal performance of the assay as the EtBr dye

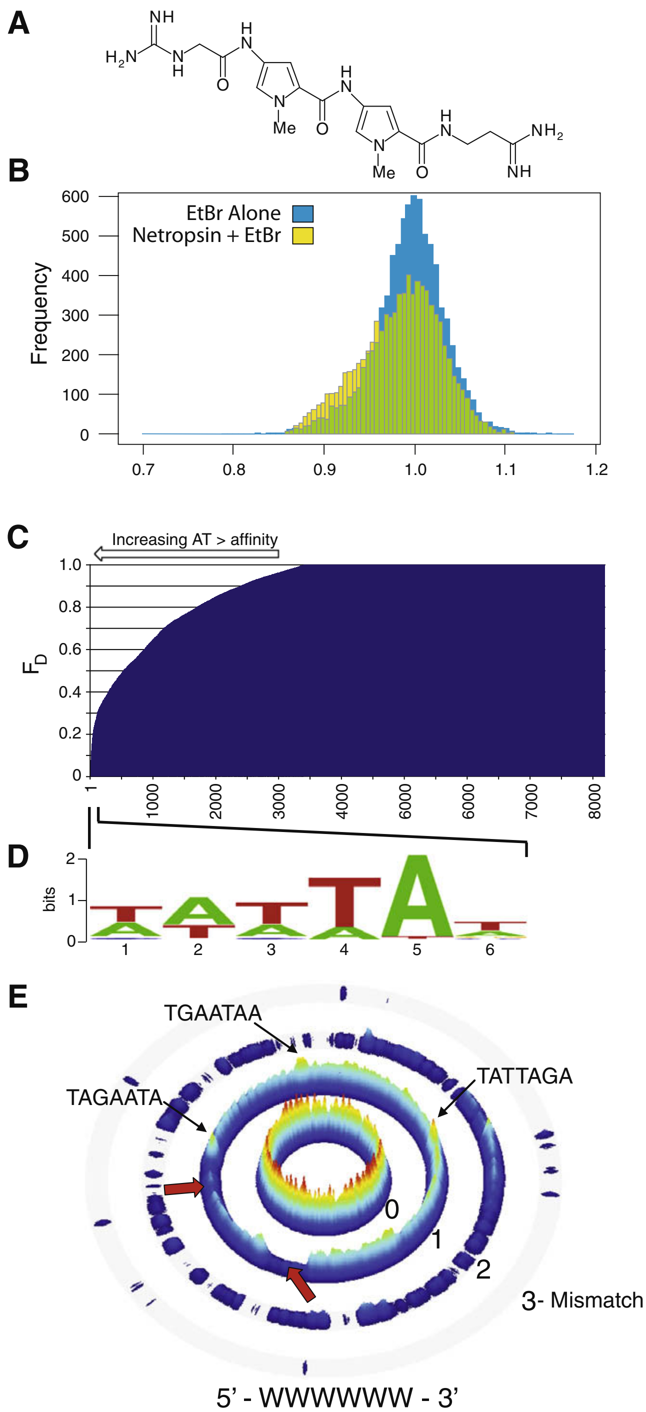

Figure 2. CSI–FID analysis of the small molecule DNA ligand netropsin. (A) The structure of netropsin. (B) Histograms of the intensities derived from CSI–FID with EtBr (blue) and EtBr + netropsin (yellow). On the left side of the histogram the yellow region shows the displacement of EtBr by netropsin, green shows overlap between both arrays. (C) A displacement display of all 7mer probes (using $F_D = 1 - (1 - R_{seq})/(1 - R_{max})$, where R_{seq} is determined using the ratio derived from the displaced and undisplaced EtBr intensities of the sequence, and R_{max} is the ratio of the sequence with the highest fluorescent displacement). This display shows a clear trend of a preference of netropsin for increasingly AT-rich DNA. (D) A logo diagram constructed from top netropsin DNA binding sequences (bracket). (E) A sequence specificity landscape (SSL) display generated for the motif 5′-WWWWWWW-3′ (W = A/T). SSLs display sequences that match the consensus on the inner ring, with mismatches to the consensus on the outer rings. The height of each sequence is calculated using $1 - F_D$. Netropsin shows a clear preference for regions of DNA with greater than 4 bp AT stretches of DNA (center ring red and yellow peaks, 1 mismatch ring ridgeline and peaks). The valleys indicated by the red arrows in the 1 mismatch ring are regions of weak netropsin DNA binding and are dominated by sequences which contain 2–3 bp AT stretches (WWSWWW or WWWSWW).

should be bound to each dsDNA probe. Initial titrations of EtBr concentrations in binding buffer (100 mM NaCl, 100 mM Tris pH 8.0) with the CSI–FID arrays indicated an optimal range of 3–6 μM EtBr, above its micromolar K_D.[35] This EtBr amount generated a 10-fold signal-to-noise (S/N) ratio for the intensity of EtBr binding to dsDNA over the array surface. Larger concentrations of EtBr increased surface binding and decreased the S/N ratio, while lower concentrations showed minimal dsDNA binding.

CSI–FID for the analysis of netropsin DNA binding. To assess assay performance netropsin DNA binding was examined using a CSI–FID array. First, EtBr at 6 μM was incubated with the array for 1 h. Subsequently, 3 μM of netropsin (K_D of 1–100nM[26,36]) was added to the EtBr solution, and the array was incubated for another hour to allow the binding reaction to achieve equilibrium. To account for any sequence bias of EtBr, and as a control for displacement, a second array was run with EtBr alone. The subsequent imaging of the CSI–FID arrays was performed with a readily available 5 μm microarray scanner using a standard 532 nm (Cy3 compatible) laser.

For both arrays we obtained a distribution of intensities indicating EtBr bound to the dsDNA probes. A comparison of the histograms for the EtBr array shows a gaussian distribution centered on 100% EtBr binding (Fig 2B), whereas the histogram of the netropsin array shows a subset of sequences with a distinct decrease in EtBr binding (Fig 2B, yellow bars on left of center), demonstrating the sequence dependent displacement of EtBr by netropsin. Based on the full displacement of EtBr from the CSI–FID array by netropsin, a fluorescent decrease of 23–29% was expected based on a 4–5 bp binding site of netropsin from 17 bp total for each DNA hairpin (9 bp of variable region plus 4 bp of constant flanking sequence on either side). The data indicate that a displacement of 20% was obtained for the best netropsin probes, close to the maximal allowable for this array design. Further analysis of the netropsin CSI–FID data indicated that a library of 7mer sequences (16,384 members) was sufficient to represent the full binding profile of netropsin. Using 9mer arrays therefore had the added benefit of increasing the number of internal replicates and adding greater sequence context for each 7mer probe. When the netropsin displacement data is plotted, there is a clear preference for netropsin binding to AT-rich DNA (Fig 2C). A sequence motif obtained from the strongest netropsin binding sites further confirms this result (Fig 2D).

Sequence specificity landscapes reveal insight into netropsin binding specificity. To further distill the sequence binding preferences of netropsin, we displayed all binding intensities in a sequence specificity landscape (SSL) format (Fig 2E, Warren et al., submitted). SSLs display all sequences that are a perfect match for a chosen motif in the innermost ring, and subsequent rings display those sequences which contain mismatches to the chosen motif. The height of each peak corresponds to the fluorescent displacement (F_D) of sequences on the CSI–FID array. The SSL display allows an unbiased and comprehensive analysis of the entire binding data, which is particularly beneficial for netropsin as most motif finding algorithms are unable to identify motifs using input sequences less than 8 bp in length.[37,38]

For netropsin, the highest peaks (Fig 2E, red to yellow) are present in the innermost ring indicating that netropsin prefers DNA regions with a high AT content (perfect match to consensus 5′-WWWWWW-3′). Interestingly, in the 1 mismatch ring, the majority of the ridgeline (Fig 2D, light blue) contains DNA sequence stretches with at least four or more AT bp (WSWWWW or SWWWWW), with some higher peaks interspersed on the ridge (Fig 2E). However, the sequences present in the valley regions (dark blue, red arrows in Fig 2E) are dominated by DNA sequences with only 2–3 bp AT stretches (WWSWWW or WWWSWW). These results indicate a strong preference of netropsin for DNA with at least four or more AT bp and agree well with previous studies on the sequence specificity of netropsin.[13,26,31]

CSI–FID versus solution-based assays. Binding to solid-surface immobilized oligonucleotides, as in CSI–FID, can be affected by mass transfer, probe density, surface characteristics, and washing steps. The data obtained by CSI–FID was therefore compared to previously obtained solution-based netropsin FID data.[26] For this analysis we parsed both datasets to represent all 4mer binding sites (136 members). Comparisons indicate that a clear correlation exists between both datasets (R^2 of 0.76, Fig 3A). Of note is that all 10 of the possible 4mer AT rich sequences are represented in the top 10 binders for both datasets. There is also a distinct step (decrease in affinity) when moving from the top CSI–FID sequences to those containing even one GC bp (Fig 3B). This comparison indicates that surface-tethered probes yield similar results as solution-based methodologies. Taken together these results represent a strong validation of the specificity data obtained by the CSI–FID assay.

CSI–FID: complex dsDNA libraries and label free detection. While there are many assays available for the study of small molecule–DNA and protein–DNA interactions, CSI–FID surmounts several inherent shortcomings of these techniques. CSI–FID can overcome the throughput and library complexity limitations inherent with other label-free detection assays by providing a high throughput assay capable of assessing ligand binding to large DNA libraries. CSI–FID is a rapid, technically non-challenging, cost effective, and adaptable assay for the label-free detection of DNA binding by natural or engineered DNA binding molecules.

In the future, CSI–FID will be applied to other additional DNA targets, including complex mixtures of proteins and small molecule ligands. Therefore, CSI–FID will greatly enhance our ability to determine DNA binding motifs for unlabeled proteins and small molecules, which has direct applications for proteomic approaches and small molecule screening. CSI–FID will contribute dramatically to the understanding of ligand–DNA binding toward the development of genome-anchored therapeutics.

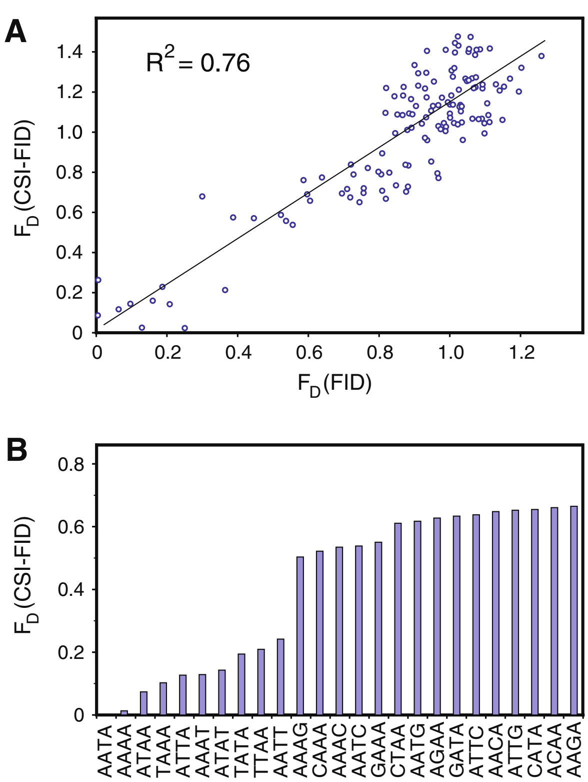

Figure 3. A comparison of netropsin data from the solution-based FID assay to the CSI–FID assay. (A) A correlation plot of all 4mer DNA sequences for both CSI–FID versus FID. The plot shows a clear correlation (R^2 = 0.76) between the two netropsin datasets and demonstrates that the data obtained from CSI–FID are not affected by any possible surface based effects. (B) A displacement display for the 4mer CSI–FID data showing a distinct step when moving from sites with 4 AT stretches (10 sites) to sites with 3 and 2 bp AT stretches.

Acknowledgments

We thank Mary Ozers and Christopher Warren for helpful discussions of the Letter, and Clayton Carlson for assistance with specificity landscapes. We gratefully acknowledge the support of the NIH (AZA: GM069420, DLB: CA041986, CA078045), March of Dimes (AZA: FY07-511), USDA (AZA: HATCH) as well as Shaw Scholar, W. M. Keck Foundation and Vilas Associate awards (A.Z.A.). K.E.H. was supported by an NSERC PGS fellowship.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2009.04.097.

References and notes

1. Geierstanger, B. H.; Wemmer, D. E. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 463.
2. Chaires, J. B. *Curr. Opin. Struct. Biol.* **1998**, *8*, 314.
3. Yang, X. L.; Wang, A. H. J. *Pharmacol. Ther.* **1999**, *83*, 181.
4. Gottesfeld, J. M.; Turner, J. M.; Dervan, P. B. *Gene Expr.* **2000**, *9*, 77.
5. Neidle, S. *Nat. Prod. Rep.* **2001**, *18*, 291.
6. Ptashne, M.; Gann, A. *Genes Signals*; Cold Springs Harbor Laboratory Press: New York, 2002.
7. Darnell, J. E., Jr. *Nat. Rev. Cancer* **2002**, *2*, 740.
8. Ansari, A. Z.; Mapp, A. K. *Curr. Opin. Chem. Biol.* **2002**, *6*, 765.
9. Waring, M. J. *Sequence-Specific DNA Binding Agents*; Royal Society of Chemistry: Cambridge, 2006.
10. Hauschild, K. E.; Carlson, C. D.; Donato, L. J.; Moretti, R.; Ansari, A. Z. In *Wiley Encyclopedia of Chemical Biology*; Begley, T., Ed.; John Wiley & Sons: New York, 2008.
11. Galas, D. J.; Schmitz, A. *Nucleic Acids Res.* **1978**, *5*, 3157.
12. Trauger, J. W.; Dervan, P. B. *Methods Enzymol.* **2001**, *340*, 450.
13. Dyke, M. W. V.; Hertzberg, R. P.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 5470.
14. Fried, M.; Crothers, D. M. *Nucleic Acids Res.* **1981**, *9*, 6505.
15. Garner, M. M.; Revzin, A. *Nucleic Acids Res.* **1981**, *9*, 3047.
16. Heyduk, T.; Ma, Y.; Tang, H.; Ebright, R. H. *Methods Enzymol.* **1996**, *274*, 492.
17. Heyduk, T.; Heyduk, E. *Nat. Biotechnol.* **2002**, *20*, 171.
18. Brockman, J. M.; Frutos, A. G.; Corn, R. M. *Anal. Biochem.* **1993**, *214*, 251.
19. Chan, L. L.; Pineda, M.; Heeres, J. T.; Hergenrother, P. J.; Cunningham, B. T. *ACS Chem. Biol.* **2008**, *3*, 437.
20. Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505.
21. Ren, B.; Robert, F.; Wyrick, J. J.; Aparicio, O.; Jennings, E. G.; Simon, I.; Zeitlinger, J.; Schreiber, J.; Hannett, N.; Kanin, E.; Volkert, T. L.; Wilson, W. J.; Bell, S. P.; Young, R. A. *Science* **2000**, *290*, 2306.
22. Wang, J. K.; Li, T. X.; Lu, Z. H. *J. Biochem. Biophys. Methods* **2005**, *63*, 100.
23. Warren, C. L.; Kratochvil, N. C.; Hauschild, K. E.; Foister, S.; Brezinski, M. L.; Dervan, P. B.; Phillips, G. N., Jr.; Ansari, A. Z. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 867.
24. Mukherjee, S.; Berger, M. F.; Jona, G.; Wang, X. S.; Muzzey, D.; Snyder, M.; Young, R. A.; Bulyk, M. L. *Nat. Genet.* **2004**, *36*, 1331.
25. Tse, W. C.; Boger, D. L. *Acc. Chem. Res.* **2004**, *37*, 61.
26. Boger, D. L.; Fink, B. E.; Brunette, S. R.; Tse, W. C.; Hedrick, M. P. *J. Am. Chem. Soc.* **2001**, *123*, 5878.
27. Puckett, J. W.; Muzikar, K. A.; Tietjen, J.; Warren, C. L.; Ansari, A. Z.; Dervan, P. B. *J. Am. Chem. Soc.* **2007**, *129*, 12310.
28. Tse, W. C.; Ishii, T.; Boger, D. L. *Bioorg. Med. Chem.* **2003**, *11*, 4479.
29. Keles, S.; Warren, C. L.; Carlson, C. D.; Ansari, A. Z. *Nucleic Acids Res.* **2008**, *36*, 3171.
30. Yeung, B. K. S.; Tse, W. C.; Boger, D. L. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3801.
31. Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 1376.
32. Dervan, P. B.; Edelson, B. S. *Curr. Opin. Struct. Biol.* **2003**, *13*, 284.
33. White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468.
34. Waring, M. J. *J. Mol. Biol.* **1965**, *13*, 269.
35. LePecq, J. B.; Paoletti, C. *J. Mol. Biol.* **1967**, *27*, 87.
36. Rentzeperis, D.; Marky, L. A.; Dwyer, T. J.; Geierstanger, B. H.; Pelton, J. G.; Wemmer, D. E. *Biochemistry* **1995**, *34*, 2937.
37. Das, M. K.; Dai, H. K. *BMC Bioinformatics* **2007**, *8*, S21.
38. Tompa, M.; Li, N.; Bailey, T. L.; Church, G. M.; De Moor, B.; Eskin, E.; Favorov, A. V.; Frith, M. C.; Fu, Y.; Kent, W. J. *Nat. Biotechnol.* **2005**, *23*, 137.