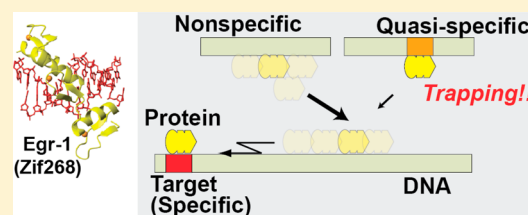


Influence of Quasi-Specific Sites on Kinetics of Target DNA Search by a Sequence-Specific DNA-Binding Protein

Catherine A. Kemme, Alexandre Esadze,[†] and Junji Iwahara*

Department of Biochemistry & Molecular Biology, Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, Galveston, Texas 77555, United States

ABSTRACT: Functions of transcription factors require formation of specific complexes at particular sites in cis-regulatory elements of genes. However, chromosomal DNA contains numerous sites that are similar to the target sequences recognized by transcription factors. The influence of such “quasi-specific” sites on functions of the transcription factors is not well understood at present by experimental means. In this work, using fluorescence methods, we have investigated the influence of quasi-specific DNA sites on the efficiency of target location by the zinc finger DNA-binding domain of the inducible transcription factor Egr-1, which recognizes a 9 bp sequence. By stopped-flow assays, we measured the kinetics of Egr-1’s association with a target site on 143 bp DNA in the presence of various competitor DNAs, including nonspecific and quasi-specific sites. The presence of quasi-specific sites on competitor DNA significantly decelerated the target association by the Egr-1 protein. The impact of the quasi-specific sites depended strongly on their affinity, their concentration, and the degree of their binding to the protein. To quantitatively describe the kinetic impact of the quasi-specific sites, we derived an analytical form of the apparent kinetic rate constant for the target association and used it for fitting to the experimental data. Our kinetic data with calf thymus DNA as a competitor suggested that there are millions of high-affinity quasi-specific sites for Egr-1 among the 3 billion bp of genomic DNA. This study quantitatively demonstrates that naturally abundant quasi-specific sites on DNA can considerably impede the target search processes of sequence-specific DNA-binding proteins.



Many transcription factors and DNA-repair/modifying enzymes perform their function by recognizing particular sequences or structural signatures as targets in DNA. In eukaryotes, this must be accomplished in the presence of billions of base pairs of genomic DNA containing numerous nonspecific sites that are structurally similar to the targets. While scanning DNA, these proteins should encounter numerous sites on DNA, which positively and negatively impact the kinetics of the protein’s target search. Nonspecific sites near targets can accelerate the target association process by creating an antenna that directs the protein to its target through one-dimensional diffusion along DNA (“sliding”).^{1–6} In contrast, nonspecific sites far outside the antenna on the same DNA or sites on different DNA molecules can effectively trap proteins because sliding or hopping from such sites does not directly lead to target association.^{4,7}

Since the discovery of amazingly rapid association of *Escherichia coli* lac repressor with operator DNA in 1970,⁸ many studies have focused on the mechanisms that accelerate target DNA search by proteins. Translocation processes such as sliding, hopping, and intersegment transfer were proposed as the mechanisms for efficient target location, initially based on indirect evidence from various biochemical experiments.^{1,7,9} Mainly in the 21st Century, these translocation processes were directly confirmed by biophysical methods such as nuclear magnetic resonance (NMR) and single-molecule techniques.^{10–13} Meanwhile, studies that focus on factors that decelerate the search process remain rare.⁵ Trapping of proteins at nonfunctional sites on DNA could be prevalent in

the nucleus because of extremely high DNA density (~100 mg/mL).¹⁴ Even though 80% of the DNA is covered by histones,¹⁵ the concentration of accessible DNA (i.e., linkers) in the nuclei is estimated to be as high as ~0.5 mM. Furthermore, genomic DNA includes many sites that are similar to the target sequence. Such sites, which we term “quasi-specific” sites, should exhibit relatively high affinities and therefore potentially trap the proteins more effectively and hinder their search for targets.^{16,17} To date, however, the influence of quasi-specific sites on functions of sequence-specific DNA-binding proteins remains to be investigated by experimental means.

We address this problem for the inducible transcription factor Egr-1 (also known as Zif268), which recognizes the 9 bp sequences, GCG(T/G)GGGCG, via three zinc finger domains.^{18,19} In the nervous system, Egr-1 functions as a regulator of synaptic plasticity to promote memory formation.^{20,21} In the cardiovascular system, Egr-1 mediates the formation of scar tissue and intimal thickening in response to damage caused by cardiovascular injury.^{22,23} To activate these responses, Egr-1 must locate its target sequence and initiate the gene response within a short time, because of its limited lifetime in the nucleus (half-life of ~0.5–1 h).²² In our previous studies using NMR and stopped-flow fluorescence spectroscopic methods,^{24–31} we investigated DNA scanning and recognition by the DNA-binding domain (DBD) of Egr-1 at molecular and atomic levels.

Received: August 31, 2015

Revised: October 22, 2015

Published: October 26, 2015

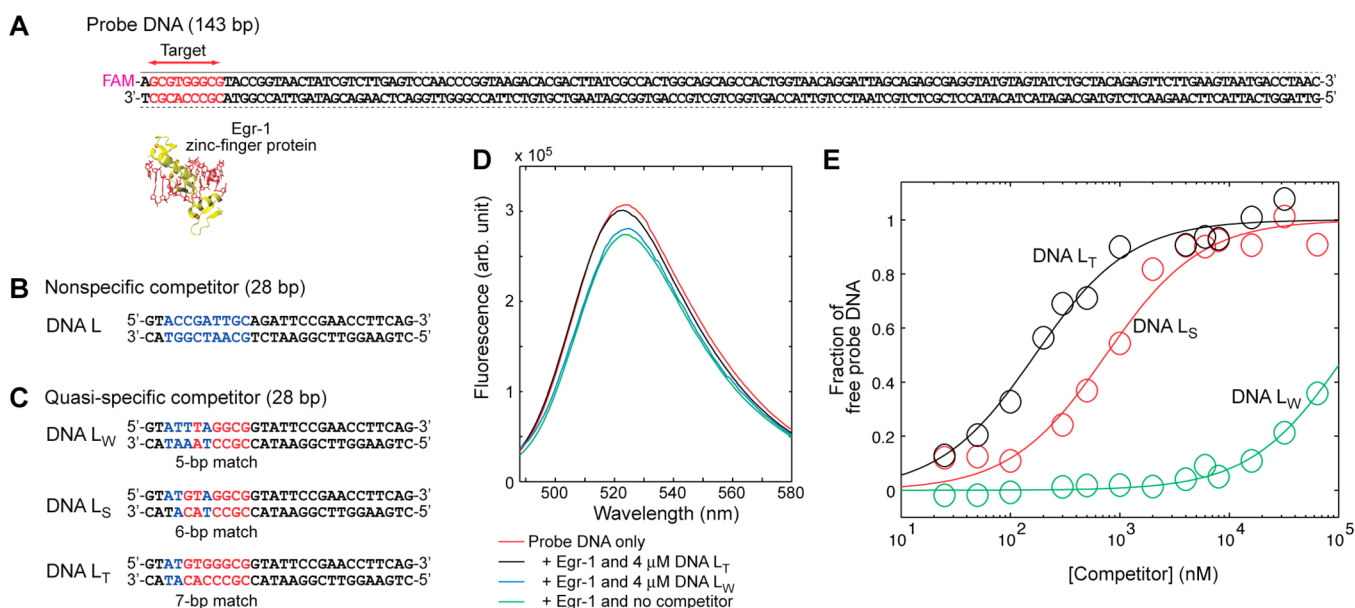


Figure 1. Measurement of relative affinities of quasi-specific DNA duplexes for the Egr-1 zinc finger protein. (A) FAM-labeled 143 bp DNA duplex as the probe DNA. The Egr-1 target site is colored red. The same probe DNA was used in our previous studies.^{25,26} (B) Nonspecific competitor DNA. This 28 bp duplex termed DNA L does not contain any sites similar to Egr-1. This nonspecific DNA was also used in our previous studies.^{25,26,29} (C) Quasi-specific DNA duplexes L_W, L_S, and L_T, which contain a sequence similar to the Egr-1 target. (D) FAM fluorescence emission spectra measured for 2.5 nM probe DNA. (E) Data from the competition assays. FAM fluorescence was measured for the solutions of 2.5 nM probe DNA, 30 nM protein, and competitor DNA at varied concentrations in 10 mM Tris-HCl (pH 7.5), 0.2 μM ZnCl₂, and 150 mM KCl. Fractions of the free probe DNA were measured from FAM fluorescence as a function of the concentration of the quasi-specific 28 bp DNA. Solid lines show the best-fit curves obtained via nonlinear least-squares fitting with eq 3.

This system is suited for research on the target search process, especially because the Egr-1 DBD behaves well in various biochemical and biophysical characterizations.

In this study, using fluorescence methods, we demonstrate the influence of various quasi-specific DNA on the efficiency in the target search by Egr-1. Our work presents a kinetic model for analyzing the effect of quasi-specific sites during the target DNA search process and provides insight into how much this effect impedes Egr-1's search process in the nucleus.

MATERIALS AND METHODS

Protein and DNA. The protein construct used in this study was the Egr-1 DBD, which consists of three zinc fingers (human Egr-1 residues 335–423). For the sake of simplicity, we will refer to this construct as Egr-1 hereafter. This protein was expressed in *E. coli* strain BL21(DE3) and purified as described in our previous papers.^{25,28,29} All fluorescence experiments used a 143 bp probe DNA duplex containing an Egr-1 target sequence, GCGTGGGCG, near a 5'-end to which a fluorescein amidite (FAM) is attached (Figure 1A). The same 143 bp probe DNA was used in our previous studies.^{25,26} This DNA duplex was generated by polymerase chain reaction (PCR) with a FAM-labeled primer, an unlabeled reverse primer, and the pUC19 plasmid (New England BioLabs), and extensively purified with the PCR purification kit (Qiagen), anion exchange chromatography, and polyacrylamide gel electrophoresis, as described previously.²⁵ Four types of unlabeled 28 bp competitor DNA duplexes were used in these experiments. One competitor, termed DNA L, is a completely nonspecific duplex (Figure 1B), which was also used in our previous work.^{25,26,29} The other three competitor 28 bp duplexes are derivatives of DNA L and contain a quasi-specific site with a 5, 6, or 7 bp match with the 9 bp target sequence

GCGTGGGCG (Figure 1C). Each chemically synthesized DNA strand was purchased from Integrated DNA Technologies and purified via Mono-Q anion exchange chromatography (GE Healthcare). After complementary strands had been annealed, the 28 bp DNA duplexes were purified with a second Mono-Q anion exchange chromatography as described. Calf thymus DNA was purchased from Invitrogen and sonicated for fragmentation into an average size of ~500 bp, which was confirmed by 0.9% agarose gel electrophoresis in TBE buffer (Invitrogen).

Competition Assays for the Specific versus Quasi-Specific and Nonspecific DNA Duplexes. Relative affinities of quasi-specific DNA duplexes for the Egr-1 zinc finger protein were measured by fluorescence-based completion assays with an ISS PC1 spectrofluorometer. Using an excitation wavelength of 460 nm and an emission wavelength of 521 nm, the FAM fluorescence was measured for 2 mL solutions of the 143 bp FAM-labeled DNA (2.5 nM), protein (30 nM), and competitor DNA (0–64 μM) in a buffer of 10 mM Tris-HCl (pH 7.5), 0.2 μM ZnCl₂, and 150 mM KCl. FAM fluorescence was also measured in the absence of both protein and competitor DNA, which corresponds to the maximal fluorescence intensity caused by the absence of quenching by macromolecular interactions. The FAM fluorescence in the presence of 30 nM protein but in the absence of competitor DNA corresponds to the minimal intensity because of complete association of the target with the protein under these conditions. The FAM fluorescence was measured as a function of concentrations of competitor DNA and was normalized to the intensity of the free probe with no competitor DNA. A control experiment with no protein but with competitor DNA was also performed under identical conditions. The normalized intensities from the control experiment were subtracted from the intensity data at

individual concentrations of competitor DNA, so that any direct influence of competitor DNA on FAM fluorescence would be removed. The fraction of the free probe DNA (p_{free}) was calculated from these intensities, assuming that each obtained intensity is the population-weighted average of the intensities for the free and protein-bound states of the probe DNA.

When the total concentrations of the probe DNA (D_{tot}), protein (P_{tot}), and competitor DNA (C_{tot}) satisfy the relationship $D_{\text{tot}} \ll P_{\text{tot}} \ll C_{\text{tot}}$, the fraction of the probe DNA in the free state (p_{free}) is given by³⁰

$$p_{\text{free}} = \frac{1 + C_{\text{tot}}/K_{\text{d(compet)}}}{1 + C_{\text{tot}}/K_{\text{d(compet)}} + P_{\text{tot}}/K_{\text{d(probe)}}} \quad (1)$$

where $K_{\text{d(compet)}}$ and $K_{\text{d(probe)}}$ are the dissociation constants for the competitor and probe DNA duplexes, respectively. The observed fluorescence intensity (I_{obs}) should be a function of p_{free} as follows:

$$I_{\text{obs}} = p_{\text{free}} I_{\text{free}} + (1 - p_{\text{free}}) I_{\text{bound}} \quad (2)$$

where I_{free} and I_{bound} are intrinsic fluorescence intensities for free and protein-bound probe DNA duplexes, respectively. If $C_{\text{tot}} \gg K_{\text{d(compet)}}$, eq 2 becomes a simple expression:

$$I_{\text{obs}} = \frac{\Gamma C_{\text{tot}} I_{\text{free}} + P_{\text{tot}} I_{\text{bound}}}{\Gamma C_{\text{tot}} + P_{\text{tot}}} \quad (3)$$

The parameter Γ represents a relative affinity defined as $K_{\text{d(probe)}}/K_{\text{d(compet)}}$. This equation was used to determine the relative affinity Γ of the quasi-specific DNA duplexes via nonlinear least-squares fitting to the experimental I_{obs} data as a function of C_{tot} . Note that reaching the asymptote at high concentrations of the competitor in this titration experiment is not a requisite for determination of Γ , because the asymptote corresponds to I_{free} , the fluorescence intensity of the free state of the probe DNA, which was directly measured. The fitting calculations were performed with the MATLAB software.

Stopped-Flow Fluorescence Kinetic Assays. The target search kinetics of Egr-1 was measured at 20 °C using an ISS PC-1 spectrofluorometer equipped with an Applied Photophysics RX.2000 stopped-flow device. In these experiments, the following two solutions were rapidly mixed in a 1:1 volume (~0.5 mL) ratio by the stopped-flow device: (1) a solution of the Egr-1 zinc finger protein and (2) a DNA solution of FAM-labeled probe DNA and competitor DNA. Both solutions were in a buffer of 10 mM Tris-HCl (pH 7.5), 0.2 μM ZnCl_2 , and 150 mM KCl. Immediately after the flow for mixing had been stopped, the time course data of the fluorescence intensity were collected for 4–35 s with a time interval of 20–50 ms. The FAM fluorophore was excited at 460 nm, and the emission light that passed through a long-pass filter with a cutoff at 515 nm (Edmund Optics) was recorded. For the competitor, we used the synthetic 28 bp duplexes shown in Figure 1 and the sonicated calf thymus DNA. When the mixtures of synthetic 28 bp duplexes were used as competitor DNA, the total concentrations of nonspecific and quasi-specific 28 bp duplexes was kept constant at 2 μM , though the concentrations of quasi-specific duplexes were varied between 0.05 and 0.25 μM . When the sonicated calf thymus DNA was used as the competitor, the experiment was performed at two different “base pair” concentrations, 56 and 112 μM (corresponding to 37 and 74 $\mu\text{g/mL}$, respectively). Each measurement was repeated 8–20 times via multiple injections. In all kinetic measurements, the

concentration of the probe DNA (D_{tot}) was 2.5 nM, whereas the concentrations of the protein (P_{tot}) and competitor (C_{tot}) were varied. To create a pseudo-first-order condition that simplifies the kinetic analysis,³² all binding reactions were conducted under conditions of $D_{\text{tot}} \ll P_{\text{tot}} \ll C_{\text{tot}}$.^{25,26} The apparent pseudo-first-order rate constant (k_{app}) for target association was determined from the time course of fluorescence intensity, $I(t)$, via nonlinear least-squares fitting with

$$I(t) = I_{\infty} + (I_0 - I_{\infty}) \exp(-k_{\text{app}} t) \quad (4)$$

where I_0 and I_{∞} represent the intensities at time zero and infinite time, respectively. Rate constant k_{app} was measured as a function of protein, and the protein concentration dependence data were analyzed with the kinetic model that is described below. MATLAB software was used for nonlinear least-squares fitting.

RESULTS

Relative Affinities of Quasi-Specific DNA Duplexes.

For quantitative characterizations of the quasi-specific sites, we first assessed their relative affinities with respect to the target site. Our previous studies^{25,26,29} on nonspecific interactions between the Egr-1 zinc finger protein and DNA utilized a completely nonspecific 28 bp duplex, which we term DNA L (Figure 1B). This DNA does not contain any sequences similar to the Egr-1 target. For the investigations of quasi-specific sites, we made three variants of DNA L, which were named L_{W} , L_{S} , and L_{T} (Figure 1C). Each contains a quasi-specific sequence involving a 5 bp (L_{W}), 6 bp (L_{S}), or 7 bp (L_{T}) match with the 9 bp target sequence GCGTGGGCG, and the subscripts in the names of these variants stand for weak, strong, and tight, respectively, representing their relative affinity for Egr-1.

Using fluorescence-based competition assays,³⁰ we investigated affinities of these quasi-specific DNA duplexes. In these experiments, the Egr-1 zinc finger protein (30 nM) and the FAM-labeled 143 bp probe DNA (2.5 nM) were mixed with competitor DNA, and the FAM fluorescence at equilibrium was measured as a function of the competitor concentration. A fluorescent FAM moiety is attached covalently to the 5'-end proximal to the target site on the probe DNA. The FAM fluorescence is partially quenched upon Egr-1's association with the target site (Figure 1D). In the absence of competitor DNA, the target site on the probe DNA is virtually 100% bound to the protein because of its high affinity for the target ($K_{\text{d}} < 0.1$ nM) under the current conditions.^{25,26} Addition of high-affinity quasi-specific DNA increased the unbound target due to transfer of protein from the target to the competitor, thereby reducing the fluorescence quenching effect (Figure 1D). From the fluorescence intensity data along with the intensities for the free and protein-bound states, we obtained the fractions of the free state of the target site on the probe DNA at individual concentrations of competitor DNA (Figure 1E). Competitor DNA duplexes at high concentrations outcompeted the target site on the probe DNA, increasing the fraction of its free state. Using these data, we determined the relative affinities of these quasi-specific DNA duplexes with respect to the affinity of the target on the probe DNA via nonlinear least-squares fitting with eq 3. The first two concentration points were excluded from the fitting calculations because these concentrations do not satisfy the inequality $P_{\text{tot}} \ll C_{\text{tot}}$, which is required for eq 3. The best-fit curves are shown in Figure 1E. Values of $\Gamma = K_{\text{d}}^{\text{quasi-specific}}/$

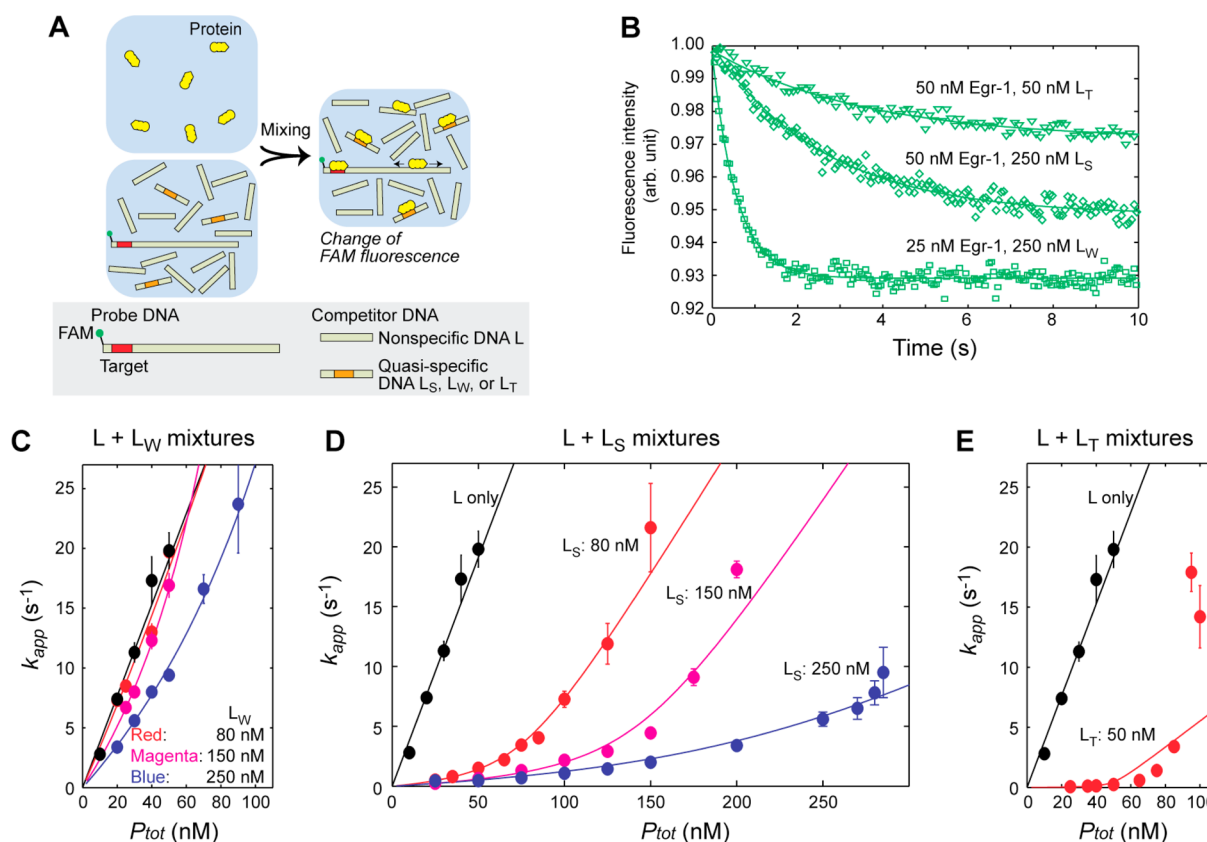


Figure 2. Impact of quasi-specific DNA on the target search kinetics of the Egr-1 zinc finger protein. (A) Schematic of the stopped-flow fluorescence assay for investigating the impact of quasi-specific DNA. In this assay, the change in FAM fluorescence was monitored upon mixing the solution of the Egr-1 zinc finger protein with the solution containing the 143 bp FAM-labeled DNA, nonspecific 28 bp DNA, and quasi-specific 28 bp DNA. The concentration of the probe DNA was 2.5 nM. The total concentration of 28 bp duplexes (nonspecific + quasi-specific) was 2000 nM, and the concentration of the quasi-specific 28 bp DNA was varied. (B) Examples of the fluorescence time course data and monoexponential fittings. (C–E) Protein concentration dependence of the apparent pseudo-first-order rate constant (k_{app}) for target association in the presence of quasi-specific DNA L_W (C), L_S (D), or L_T (E). Circles show the k_{app} constants obtained from monoexponential fitting to the fluorescence time course data. The solid lines represent the best-fit curves obtained via nonlinear least-squaring fitting with eqs 6–9. In these calculations, only two parameters, $K_{d,q}$ and k_{on} , were optimized. The buffer conditions for these experiments were 10 mM Tris-HCl (pH 7.5), 0.2 μ M ZnCl₂, and 150 mM KCl. Note that protein concentration dependence of the target search kinetics becomes biphasic (rather than linear) in the presence of high-affinity quasi-specific sites.

$K_d^{specific}$ for DNA duplexes L_T , L_S , and L_W were determined to be 5.6 ± 0.8 , 25 ± 4 , $(3.9 \pm 1.7) \times 10^3$, respectively. These results qualitatively indicate that a sequence more similar to the target sequence exhibits a stronger affinity, which is quite reasonable. This set of quasi-specific DNA duplexes allowed us to examine the relationship between the affinity and kinetic impact of quasi-specific sites, as described below.

Impact of Quasi-Specific Sites on the Kinetics of Target Search. By stopped-flow fluorescence assays similar to those described in our previous studies,^{25,26} we investigated the influence of the quasi-specific DNA on the target search kinetics of Egr-1. The basic scheme for the kinetic experiment is depicted in Figure 2A. In these assays, a protein solution is mixed with a DNA solution containing the probe DNA (final concentration, 2.5 nM), nonspecific competitor DNA L , and quasi-specific competitor DNA L_W , L_S , or L_T . The final total concentration of the competitor DNA duplexes (i.e., nonspecific + quasi-specific) was kept constant at 2000 nM, whereas the concentration of the quasi-specific competitor was varied. Immediately after the flow of mixing was stopped, the reaction time course for the association of the protein to the target site was recorded by measuring the change in the FAM fluorescence intensity over time. Some of the time course data are shown in Figure 2B. The percent change in fluorescence

intensity was typically 3–7%, depending on the fraction of the protein-bound state of the target site on DNA at equilibrium. The change was relatively small when the target site on the probe DNA (2.5 nM) was outcompeted by the high-affinity quasi-specific site of a substantially higher concentration (e.g., see the data with 50 nM DNA L_T in Figure 2B). Time courses for the fluorescence intensity were found to be monoexponential. The pseudo-first-order rate constants (k_{app}) were determined from the time course data at various concentrations of protein and quasi-specific DNA.

While the total competitor concentration was kept at 2 μ M, the pseudo-first-order rate constants (k_{app}) were measured at various concentrations of the protein in the presence of 80, 150, and 250 nM quasi-specific DNA L_W (Figure 2C) or L_S (Figure 2D). For the quasi-specific DNA L_T , only a single concentration of 50 nM was tested (Figure 2E) because the kinetic measurement at a higher concentration of this duplex was difficult due to the small magnitude of the fluorescence change. In all cases tested, we found that the presence of the quasi-specific DNA made the target search kinetics considerably slower. For each quasi-specific DNA, we measured the rate constants k_{app} using various concentrations of Egr-1, starting at low concentrations (10–25 nM) and increasing until we reached the upper limit of our instrument's measurable range

($\sim 20 \text{ s}^{-1}$). We found that as we increased the protein concentration, the rate of association increased, as well. In the case with only nonspecific DNA L being present as a competitor, the dependence of k_{app} on protein concentration was linear (black in Figure 2C–E), as expected for any second-order process. The data for the cases in the presence of DNA L_W were also almost linear (Figure 2C). However, we found that the protein concentration dependence of k_{app} in the presence of DNA L_S or L_T was clearly biphasic rather than linear (Figure 2D,E). At concentrations below the concentration of quasi-specific DNA, the rate of Egr-1 increased linearly with a shallow slope. However, when the concentration of the Egr-1 zinc finger protein exceeded that of the quasi-specific DNA, the slope increased dramatically and proceeded again in a linear fashion. This tendency was more pronounced for high-affinity quasi-specific DNA.

Kinetic Model for the Target Search in the Presence of Quasi-Specific Sites. To quantitatively understand the kinetic influence of the quasi-specific site, we modified our previous analytical expression for the target search kinetics in the presence of nonspecific competitor DNA. Previously, for a system involving protein, probe DNA, and competitor DNA, we showed that when $D_{\text{tot}} \ll P_{\text{tot}} \ll C_{\text{tot}}$ the apparent second-order rate constant (k_a) for target association is related to the intrinsic association rate constant ($k_{\text{on},n}$) for each nonspecific site as follows:²⁶

$$k_a = \rho \eta S k_{\text{on},n} \quad (5)$$

Parameter ρ represents a scaling factor ($0 < \rho < 1$) due to the trapping of protein at nonspecific sites and corresponds to the fraction of protein molecules that are not trapped by any nonspecific sites during the target search process. Parameter S represents the so-called antenna effect;^{4,26,33} nonspecific sites near the target on the same DNA serve as an antenna that attracts the protein and makes the target association S -fold faster. Parameter η represents an enhancement factor ($\eta > 1$) due to intersegment transfer. On the basis of the discrete stochastic kinetic model of Veksler and Kolomeisky,³⁴ we previously gave explicit forms of parameters η and S as functions of various kinetic rate constants, equilibrium constants, and configurational factors.²⁶ When $D_{\text{tot}} \ll P_{\text{tot}} \ll C_{\text{tot}}$ parameter ρ is given by

$$\rho = \frac{1}{Z} = \frac{1}{1 + N_{\text{tot}}/K_{d,n}} \quad (6)$$

where Z corresponds to a partition function for protein at the pseudoequilibrium during the target search process, N_{tot} is the total concentration of nonspecific sites (on competitor and probe DNA, excluding those in the antenna region), and $K_{d,n}$ is the dissociation constant for each nonspecific site.

For the systems involving quasi-specific sites on competitor DNA, we make the following two assumptions: (1) Parameters S and η are virtually unaffected by the presence of quasi-specific sites on competitor DNA, and (2) interactions of protein with quasi-specific sites and with nonspecific sites reach steady states well before the interaction with the target site reaches equilibrium. The first assumption should be valid in the current case because the quasi-specific sites are located only on the competitor DNA, not on the probe DNA. The second assumption is justified when the concentrations of the quasi-specific and nonspecific sites are far greater than the concentration of the target site. The pseudoequilibrium for

the nonspecific DNA was rigorously validated using exact numerous simulations for the system with only nonspecific competitor DNA in our previous work.²⁵ Under the assumption of the pseudoequilibrium during the target search process, the trapping effect is represented by the following parameter ρ_{nq} :

$$\rho_{\text{nq}} = \frac{1}{Z_{\text{nq}}} = \frac{1}{1 + N_{\text{tot}}/K_{d,n} + [Q]/K_{d,q}} \quad (7)$$

where Z_{nq} represents a partition function in the form of the binding polynomial³⁵ for protein at the pseudoequilibrium in the presence of quasi-specific sites; $[Q]$ is the concentration of the quasi-specific sites in the free state; and $K_{d,q}$ is the dissociation constant for each quasi-specific site.

Equation 7 together with eq 5 can qualitatively explain the biphasic dependence of the apparent pseudo-first-order rate constants (k_{app}) on the total protein concentration (P_{tot}) as seen in Figure 2C–E. A slope of protein concentration dependence corresponds to an apparent second-order rate constant k_a . When the protein concentration is low, a high affinity (i.e., $K_{d,q} \ll K_{d,n}$) of the quasi-specific site and a large fraction of its free state can make the $[Q]/K_{d,q}$ term predominant in partition function Z_{nq} , rendering $\rho_{\text{nq}} \ll \rho$. This corresponds to the first phase of the biphasic dependence, where the slope is far gentler than one in the absence of the quasi-specific sites. When the protein concentration is significantly higher than the total concentration of the quasi-specific site, most quasi-specific sites are bound to the protein and $[Q]$ can become virtually zero, making $\rho_{\text{nq}} \approx \rho$. This corresponds to the second phase of the biphasic dependence, where the slope should be virtually the same as that in the system involving no quasi-specific site.

In the case presented here, the concentration of quasi-specific sites in the free state in the pseudoequilibrium during the target search process is given by

$$[Q] = \frac{1}{2}(Q_{\text{tot}} - P_{\text{tot}} - K_{d,q}Z) + \frac{1}{2}\sqrt{(Q_{\text{tot}} - P_{\text{tot}} - K_{d,q}Z)^2 + 4Q_{\text{tot}}K_{d,q}Z} \quad (8)$$

where $Z = 1 + N_{\text{tot}}/K_{d,n}$ (i.e., the same as Z in eq 6). This expression is derived by solving the equations $K_{d,n} = N_{\text{tot}}[P]/[NP]$, $K_{d,q} = [Q][P]/[QP]$, $P_{\text{tot}} = [P] + [NP] + [QP]$, and $Q_{\text{tot}} = [Q] + [QP]$, where $[P]$, $[NP]$, and $[QP]$ represent the concentrations of free protein, nonspecific sites bound to protein, and the quasi-specific site bound to protein, respectively. The apparent pseudo-first-order rate constant (k_{app}) is given by

$$k_{\text{app}} = (\rho_{\text{nq}}/\rho)k_{a0}P_{\text{tot}} \quad (9)$$

where k_{a0} corresponds to the second-order rate constant when no quasi-specific site is involved in competitor DNA.

For our experimental data in panels C–E of Figure 2, we conducted fitting calculations with eqs 6–9 via optimization of two parameters, k_{a0} and $K_{d,q}$. These calculations require the experimental value of the dissociation constant ($K_{d,n}$) for the affinity of each nonspecific site. In our previous study,²⁶ we determined $K_{d,n}$ to be 16 μM for Egr-1 under the identical buffer conditions with 150 mM KCl. The best-fit curves are shown together with the experimental data in the graphs in Figure 2C–E. The fitting gave good agreement with the experimental data. From these fittings to the kinetic data, $K_{d,q}$

values of DNA duplexes L_T , L_S , and L_W were calculated to be 0.07 ± 0.05 , 1.0 ± 0.3 , and 44 ± 7 nM, respectively. With experimental uncertainties taken into consideration, ratios of these values from the kinetic data are consistent with the relative affinity data from the competition assays. These results suggest that our kinetic model can explain the kinetic influence of quasi-specific sites both qualitatively and quantitatively.

Quasi-Specific Sites in Genomic DNA. To examine the influence of natural quasi-specific sites in genomic DNA on the target search kinetics of Egr-1, we conducted the stopped-flow fluorescence assays using calf thymus DNA as a competitor. In this experiment, sonicated calf thymus DNA (average length, ~ 500 bp) was used instead of synthetic duplexes such as DNA L , L_W , L_S , and L_T . Using base pair concentrations of 56 and 112 μM for the sonicated calf thymus DNA (equivalent to 2 and 4 μM , respectively, for 28 bp DNA) as a competitor, we measured the target search kinetics of Egr-1 at 150 mM KCl. Figure 3 shows the dependence of measured k_{app} constants on

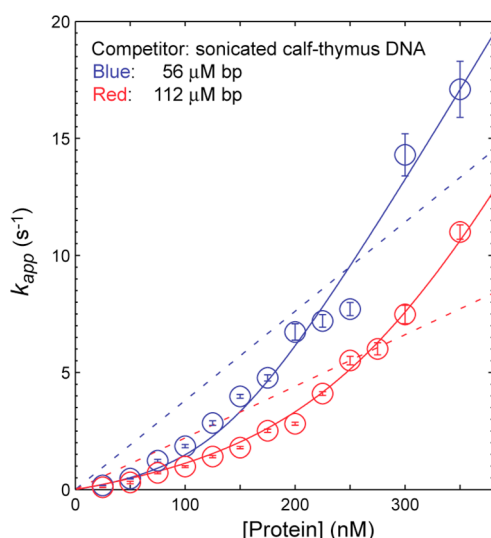


Figure 3. Evidence of the kinetic influence of natural quasi-specific sites on the target search process by Egr-1. The graph shows the protein concentration dependence of k_{app} constants measured with the stopped-flow assay using calf thymus DNA as a competitor. To reduce the viscosity, calf thymus DNA was fragmented into an average length of ~ 500 bp by sonication. The dotted lines represent fitting with proportional functions. The solid lines are the best-fit curves obtained via nonlinear least-squares fitting with eqs 6–9. The fitting calculation was performed for the two data sets simultaneously. In this calculation, four fitting parameters were optimized: two k_{30} parameters at two different overall DNA concentrations, the apparent affinity ($K_{d,q}$), and probability (f_q) of quasi-specific sites. The global fitting calculations gave an apparent probability of the quasi-specific sites among the genomic DNA of $0.28 \pm 0.03\%$, and $K_{d,q} = 3.7 \pm 0.8$ nM. These data suggest that there are $\sim 10^6$ – 10^7 quasi-specific sites with high affinity for Egr-1 in the genomic DNA.

protein concentration. The dependence in these experiments with calf thymus DNA appeared to be nonlinear, as seen in the case with synthetic quasi-specific DNA. In fact, fitting with proportional functions assuming a simple second-order kinetics gave poor agreement with the experimental data as shown in Figure 3 (dotted lines). These results strongly suggest the significant influence of quasi-specific sites in calf thymus DNA.

To gain insight into the quantity and affinity of quasi-specific sites in calf thymus DNA, we used our kinetic model to

conduct global fitting for the 56 and 112 μM base pair data. In this calculation, we defined a probability f_q for quasi-specific sites, with which $Q_{\text{tot}} = f_q N_{\text{tot}}$ in eqs 8 and 9, and optimized four parameters: f_q , $K_{d,q}$, and two k_{30} parameters individually defined for the two data sets. Application of the current kinetic model to the genomic DNA containing various different quasi-specific sites is obviously simplistic, because this model assumes a uniform $K_{d,q}$ for all quasi-specific sites. Therefore, the affinity ($K_{d,q}$) and concentration (Q_{tot}) from these calculations should be regarded merely as apparent parameters. The global fitting calculation with eqs 6–9 showed excellent agreement with both experimental data sets (solid lines, Figure 3) and yielded a coefficient of determination higher than that of the linear model (R^2 values of 0.985 vs 0.849). This calculation gave values for the apparent affinity ($K_{d,q}$) and probability of quasi-specific sites (f_q) of 3.7 ± 0.8 nM and 0.0028 ± 0.0003 , respectively. These results suggest that high-affinity quasi-specific sites number as many as $\sim 10^6$ – 10^7 in 3 billion base pairs of calf thymus genomic DNA.

DISCUSSION

Trapping at Nonfunctional Sites. Recently, methods such as ChIP-on-chip³⁶ and ChIP-seq³⁷ have allowed for genome-wide studies of binding sites of transcription factors *in vivo*. Such genome-wide studies showed that transcription factors bind to many DNA sites that are apparently nonfunctional in the nuclei.^{38,39} As these methods detect only high occupancies of transcription factors at sites with the strongest affinities,⁴⁰ there must be a far greater number of quasi-specific sites with weaker affinities that are similar to the recognition sequences. This should be particularly true for eukaryotes because their genome is large and eukaryotic transcription factors recognize relatively short sequences (typically <10 bp).^{41,42} Because of the large abundance, quasi-specific sites could substantially influence transcription factors *in vivo* in both thermodynamic and kinetic terms, as theoretically considered by Chakrabarti et al.¹⁶

In fact, our current results from the kinetic experiment with calf thymus DNA suggest that target DNA search by Egr-1 can be considerably impeded due to $\sim 10^6$ – 10^7 quasi-specific sites, which substantially increase the mean search time of Egr-1. For a pool of random sequences, the probability of finding m bp match in a window of n bp covered by a transcription factor is given by

$$P_{n,m} = 2(1/4)^m (3/4)^{n-m} {}_nC_m \quad (10)$$

where ${}_nC_m$ represents combinations and the factor of 2 accounts for the sequence match for the complementary strand. Using this, the total number of quasi-specific sites ($m \geq 6$) for Egr-1 ($n = 9$) is estimated to be on the order of 10^7 sites in human genomic DNA comprised of 3×10^9 bp. Thus, our experimental results are roughly consistent with this probabilistic estimate.

Potential Role of Quasi-Specific Sites in the Regulation of Transcription Factors. Another remarkable finding from our data is that the adverse effect of the quasi-specific sites on target association disappears once quasi-specific sites are completely occupied by proteins. This gives two important implications. First, a relatively high expression level of the transcription factors is required for efficient regulation of their target genes, unless other proteins occupy the nonfunctional quasi-specific sites. When the level of the transcription factor

exceeds a threshold at which binding to quasi-specific sites is saturated, target association of the transcription factors will become drastically enhanced. This sharp response is essentially similar to the ultrasensitive response caused by protein sequestration, which was studied for genetic circuits in yeast.^{43,44} Second, functions of the transcription factors would be considerably enhanced if other proteins (e.g., histones and other nuclear proteins) bind to the quasi-specific sites and make them inaccessible for the transcription factors. The quasi-specific sites could also be blocked by other proteins of the same transcription factor family due to similar sequence specificity in DNA binding. DNA methylation could block quasi-specific sites by altering their affinities or by attracting methyl-CpG-binding proteins to quasi-specific sites containing methylated CpG dinucleotides. The latter should be particularly relevant to Egr-1. The 9 bp Egr-1 target sequences contain two CpG dinucleotides, yet their methylation does not weaken association of Egr-1 with target DNA *in vitro*.³⁰ Interestingly, a genome-wide ChIP-on-chip study of Egr-1-binding sites⁴⁵ showed that the functional target sites for Egr-1 are colocalized with CpG islands. Note that DNA methylation is rare (typically <10%) in CpG islands, although the overall CpG methylation level is as high as 80% in mammalian genomic DNA.^{46–49} Because of this distribution, it is likely that methyl-CpG-binding proteins *do not* block the functional target sites for Egr-1 in the CpG islands but *do* block the majority of quasi-specific sites. Western blot and DNA association data for nuclear extracts (e.g., refs 22 and 23) suggest that when induced, the level of nuclear Egr-1 *in vivo* is roughly on the order of 10^{-9} to 10^{-7} M, corresponding up to $\sim 10^4$ copies per nucleus. Considering that this number is smaller than the estimated number of quasi-specific sites in genomic DNA, blocking or releasing of quasi-specific sites may work as an effective mechanism for the regulation of Egr-1 and other transcription factors. Further studies are required to examine this interesting possibility.

CONCLUDING REMARKS

This study demonstrates a quantitative description of the impact of quasi-specific sites on target search kinetics for Egr-1. Depending on the affinities and numbers of quasi-specific sites, they can substantially impede the search process due to trapping of the protein. Because of this effect, the protein concentration dependence of the apparent pseudo-first-order kinetic rate constant for target association in the presence of quasi-specific sites is biphasic (rather than linear) despite the second-order nature of the target association process. When all quasi-specific sites are saturated with proteins, the target association becomes far faster because the strong trapping effect becomes absent. Given this observation, it is reasonable to consider that quasi-specific sites can substantially attenuate functions of transcription factors *in vivo* and that quasi-specific sites might play a role in the regulation of transcription factors via indirect interplay with other nuclear proteins.

AUTHOR INFORMATION

Corresponding Author

*Address: 301 University Blvd., Medical Research Building 5.104C, Galveston, TX 77555-1068. E-mail: j.iwahara@utmb.edu. Phone: 409-747-1403. Fax: 409-772-6334.

Present Address

†A.E.: Department of Pharmacology and Molecular Sciences, Johns Hopkins University School of Medicine, Baltimore, MD 21205.

Funding

This work was supported by Grant R01-GM107590 from the National Institutes of Health (to J.I.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Dr. Anatoly Kolomeisky for useful discussion about theoretical issues.

ABBREVIATIONS

DBD, DNA-binding domain; FAM, fluorescence amidite.

REFERENCES

- (1) von Hippel, P. H., and Berg, O. G. (1989) Facilitated target location in biological systems. *J. Biol. Chem.* 264, 675–678.
- (2) Halford, S. E., and Marko, J. F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* 32, 3040–3052.
- (3) Zhou, H. X. (2005) A model for the mediation of processivity of DNA-targeting proteins by nonspecific binding: dependence on DNA length and presence of obstacles. *Biophys. J.* 88, 1608–1615.
- (4) Mirny, L., Slutsky, M., Wunderlich, Z., Tafvizi, A., Leith, J., and Kosmrlj, A. (2009) How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *J. Phys. A: Math. Theor.* 42, 434013.
- (5) Halford, S. E. (2009) An end to 40 years of mistakes in DNA-protein association kinetics? *Biochem. Soc. Trans.* 37, 343–348.
- (6) Kolomeisky, A. B. (2011) Physics of protein-DNA interactions: mechanisms of facilitated target search. *Phys. Chem. Chem. Phys.* 13, 2088–2095.
- (7) Winter, R. B., Berg, O. G., and von Hippel, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli* lac repressor-operator interaction: kinetic measurements and conclusions. *Biochemistry* 20, 6961–6977.
- (8) Riggs, A. D., Bourgeois, S., and Cohn, M. (1970) The lac repressor-operator interaction. 3. Kinetic studies. *J. Mol. Biol.* 53, 401–417.
- (9) Berg, O. G., Winter, R. B., and von Hippel, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* 20, 6929–6948.
- (10) Clore, G. M. (2011) Exploring translocation of proteins on DNA by NMR. *J. Biomol. NMR* 51, 209–219.
- (11) Clore, G. M., and Iwahara, J. (2009) Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem. Rev.* 109, 4108–4139.
- (12) Gorman, J., and Greene, E. C. (2008) Visualizing one-dimensional diffusion of proteins along DNA. *Nat. Struct. Mol. Biol.* 15, 768–774.
- (13) Tafvizi, A., Mirny, L. A., and van Oijen, A. M. (2011) Dancing on DNA: kinetic aspects of search processes on DNA. *ChemPhysChem* 12, 1481–1489.
- (14) Lewin, B. (2000) *Genes VII*, Oxford University Press, Oxford, U.K.
- (15) Valouev, A., Johnson, S. M., Boyd, S. D., Smith, C. L., Fire, A. Z., and Sidow, A. (2011) Determinants of nucleosome organization in primary human cells. *Nature* 474, 516–520.
- (16) Chakrabarti, J., Chandra, N., Raha, P., and Roy, S. (2011) High-affinity quasi-specific sites in the genome: how the DNA-binding proteins cope with them. *Biophys. J.* 101, 1123–1129.

- (17) Marcovitz, A., and Levy, Y. (2011) Frustration in protein-DNA binding influences conformational switching and target search kinetics. *Proc. Natl. Acad. Sci. U. S. A.* 108, 17957–17962.
- (18) Elrod-Erickson, M., Rould, M. A., Neklodova, L., and Pabo, C. O. (1996) Zif268 protein-DNA complex refined at 1.6 Å: a model system for understanding zinc finger-DNA interactions. *Structure* 4, 1171–1180.
- (19) Pavletich, N. P., and Pabo, C. O. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 252, 809–817.
- (20) Bozon, B., Davis, S., and Laroche, S. (2003) A requirement for the immediate early gene zif268 in reconsolidation of recognition memory after retrieval. *Neuron* 40, 695–701.
- (21) Lee, J. L., Everitt, B. J., and Thomas, K. L. (2004) Independent cellular processes for hippocampal memory consolidation and reconsolidation. *Science* 304, 839–843.
- (22) Khachigian, L. M., Lindner, V., Williams, A. J., and Collins, T. (1996) Egr-1-induced endothelial gene expression: a common theme in vascular injury. *Science* 271, 1427–1431.
- (23) Yan, S. F., Fujita, T., Lu, J., Okada, K., Shan Zou, Y., Mackman, N., Pinsky, D. J., and Stern, D. M. (2000) Egr-1, a master switch coordinating upregulation of divergent gene families underlying ischemic stress. *Nat. Med.* 6, 1355–1361.
- (24) Chen, C., Esadze, A., Zandarashvili, L., Nguyen, D., Pettitt, B. M., and Iwahara, J. (2015) Dynamic Equilibria of Short-Range Electrostatic Interactions at Molecular Interfaces of Protein-DNA Complexes. *J. Phys. Chem. Lett.* 6, 2733–2737.
- (25) Esadze, A., and Iwahara, J. (2014) Stopped-flow fluorescence kinetic study of protein sliding and intersegment transfer in the target DNA search process. *J. Mol. Biol.* 426, 230–244.
- (26) Esadze, A., Kemme, C. A., Kolomeisky, A. B., and Iwahara, J. (2014) Positive and negative impacts of nonspecific sites during target location by a sequence-specific DNA-binding protein: origin of the optimal search at physiological ionic strength. *Nucleic Acids Res.* 42, 7039–7046.
- (27) Iwahara, J., and Levy, Y. (2013) Speed-stability paradox in DNA-scanning by zinc-finger proteins. *Transcription* 4, 58–61.
- (28) Takayama, Y., Sahu, D., and Iwahara, J. (2010) NMR studies of translocation of the Zif268 protein between its target DNA Sites. *Biochemistry* 49, 7998–8005.
- (29) Zandarashvili, L., Vuzman, D., Esadze, A., Takayama, Y., Sahu, D., Levy, Y., and Iwahara, J. (2012) Asymmetrical roles of zinc fingers in dynamic DNA-scanning process by the inducible transcription factor Egr-1. *Proc. Natl. Acad. Sci. U. S. A.* 109, E1724–1732.
- (30) Zandarashvili, L., White, M. A., Esadze, A., and Iwahara, J. (2015) Structural impact of complete CpG methylation within target DNA on specific complex formation of the inducible transcription factor Egr-1. *FEBS Lett.* 589, 1748–1753.
- (31) Zandarashvili, L., Esadze, A., Vuzman, D., Kemme, C. A., Levy, Y., and Iwahara, J. (2015) Balancing between affinity and speed in target DNA search by zinc-finger proteins via modulation of dynamic conformational ensemble. *Proc. Natl. Acad. Sci. U. S. A.* 112, E5142–5149.
- (32) Pollard, T. D., and De La Cruz, E. M. (2013) Take advantage of time in your experiments: a guide to simple, informative kinetics assays. *Mol. Biol. Cell* 24, 1103–1110.
- (33) Hu, T., Grosberg, A. Y., and Shklovskii, B. I. (2006) How proteins search for their specific sites on DNA: the role of DNA conformation. *Biophys. J.* 90, 2731–2744.
- (34) Veksler, A., and Kolomeisky, A. B. (2013) Speed-Selectivity Paradox in the Protein Search for Targets on DNA: Is It Real or Not? *J. Phys. Chem. B* 117, 12695–12701.
- (35) Dill, K. A., and Bromberg, S. (2011) Multi-site & cooperative ligand binding. In *Molecular Driving Forces: Statistical thermodynamics in biology, chemistry, physics, and nanoscience*, pp 559–584, Garland Science, New York.
- (36) Harbison, C. T., Gordon, D. B., Lee, T. I., Rinaldi, N. J., Macisaac, K. D., Danford, T. W., Hannett, N. M., Tagne, J. B., Reynolds, D. B., Yoo, J., Jennings, E. G., Zeitlinger, J., Pokholok, D. K., Kellis, M., Rolfe, P. A., Takusagawa, K. T., Lander, E. S., Gifford, D. K., Fraenkel, E., and Young, R. A. (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature* 431, 99–104.
- (37) Johnson, D. S., Mortazavi, A., Myers, R. M., and Wold, B. (2007) Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316, 1497–1502.
- (38) Fisher, W. W., Li, J. J., Hammonds, A. S., Brown, J. B., Pfeiffer, B. D., Weizmann, R., MacArthur, S., Thomas, S., Stamatoyannopoulos, J. A., Eisen, M. B., Bickel, P. J., Biggin, M. D., and Celniker, S. E. (2012) DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 109, 21330–21335.
- (39) Li, X. Y., MacArthur, S., Bourgon, R., Nix, D., Pollard, D. A., Iyer, V. N., Hechmer, A., Simirenko, L., Stapleton, M., Luengo Hendriks, C. L., Chu, H. C., Ogawa, N., Inwood, W., Sementchenko, V., Beaton, A., Weizmann, R., Celniker, S. E., Knowles, D. W., Gingeras, T., Speed, T. P., Eisen, M. B., and Biggin, M. D. (2008) Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol.* 6, e27.
- (40) Todeschini, A. L., Georges, A., and Veitia, R. A. (2014) Transcription factors: specific DNA binding and specific gene regulation. *Trends Genet.* 30, 211–219.
- (41) Badis, G., Berger, M. F., Philippakis, A. A., Talukder, S., Gehrke, A. R., Jaeger, S. A., Chan, E. T., Metzler, G., Vedenko, A., Chen, X., Kuznetsov, H., Wang, C. F., Coburn, D., Newburger, D. E., Morris, Q., Hughes, T. R., and Bulyk, M. L. (2009) Diversity and complexity in DNA recognition by transcription factors. *Science* 324, 1720–1723.
- (42) Wunderlich, Z., and Mirny, L. A. (2009) Different gene regulation strategies revealed by analysis of binding motifs. *Trends Genet.* 25, 434–440.
- (43) Buchler, N. E., and Cross, F. R. (2009) Protein sequestration generates a flexible ultrasensitive response in a genetic network. *Mol. Syst. Biol.* 5, 272.
- (44) Lee, T. H., and Maheshri, N. (2012) A regulatory role for repeated decoy transcription factor binding sites in target gene expression. *Mol. Syst. Biol.* 8, 576.
- (45) Kubosaki, A., Tomaru, Y., Tagami, M., Arner, E., Miura, H., Suzuki, T., Suzuki, M., Suzuki, H., and Hayashizaki, Y. (2009) Genome-wide investigation of in vivo EGR-1 binding sites in monocytic differentiation. *Genome Biol.* 10, R41.
- (46) Bergman, Y., and Cedar, H. (2013) DNA methylation dynamics in health and disease. *Nat. Struct. Mol. Biol.* 20, 274–281.
- (47) Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21.
- (48) Jones, P. A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* 13, 484–492.
- (49) Schubeler, D. (2015) Function and information content of DNA methylation. *Nature* 517, 321–326.