The Dissertation Committee for Catherine A. Kemme
certifies that this is the approved version of the following dissertation:

Complexity of transcription factor target search: effect of DNA natural decoys and DNA methylation through interplay with methyl-CpG-binding proteins on transcription factor DNA association

Committee:

Junji Iwahara, Supervisor

Montgomery Pettitt

Wlodek Bujalowski

Anatoly Kolomeisky

Kay Choi

Complexity of transcription factor target search: effect of DNA natural decoys and DNA methylation through interplay with methyl-CpG-binding proteins on transcription factor DNA association

by

Catherine A. Kemme, B.S.

DISSERTATION

Presented to the Faculty of the Graduate School of
The University of Texas Medical Branch at Galveston
in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS MEDICAL BRANCH AT GALVESTON
December 2017

With love and gratitude to my parents, Paul and Angela Kemme; and to my sister, Caroline Kemmme

Acknowledgments

I would like thank my mentor Junji Iwahara, for assisting me in my PhD work. His extensive knowledge as a scientist and his mentorship have helped me tremendously in the past four years as I have studied and researched under his watch.

I would also like to thank my committee members, Montgomery Pettitt, Wlodek Bujalowski, Kay Choi, and Anatoly Kolomeisky for supporting me in my training, and giving me advice when I needed it.

I would like to thank previous lab members Alex Esadze and Levani Zandarashvili, who helped me extensively in my first two years a graduate student, and taught me many of the experimental techniques I know today. I also would like to thank my current lab members, Dan Nguyen for encouraging me as we both worked on our PhD together, and Esmeralda Martinez, for the support and the opportunity to help train her as a new student.

I would especially like to thank Dr. James Lee; who, without his support and the potential he saw in me, I would not be here today.

I would also like to thank my parents Paul and Angela Kemme, from whom I have had never ending support. Without them, none of this would be possible. I also would like to thank my siblings, Caroline, Meredith, Aaron, and Olivia, who have been there to listen and provide encouragement when

things seemed difficult. I would especially like to mention my sister Caroline, who in spite of her own busy schedule as a PhD student, took out of her time to help me proofread my articles and this dissertation, even if it was late into the night, and we could hardly concentrate any longer.

Above all, I would like to thank the Lord and my Savior Jesus Christ, who is a help far more than any human can provide. He is there through all my struggle and through my happiness, and without Him, my life would be meaningless. I owe Him my life, my family, and my career, to which I am extremely grateful. I am privileged to be able to study the details of His Creation, and am amazed at its seemingly incomprehensible complexity.

Complexity of transcription factor target search: effect of DNA natural decoys and DNA methylation through interplay with methyl-CpG-binding proteins on transcription factor DNA association

Publication No. _____

Catherine A. Kemme, Ph.D.
The University of Texas Medical Branch at Galveston, 2017

Supervisor: Junji Iwahara

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. These sequences potentially serve as natural decoys that sequester transcription factors. The impact that these decoys have on transcription factor target search is largely unknown, and various factors such as methylation, other DNA binding proteins, and side chain dynamics of individual protein residues may play a role in transcription factor target search during binding and scanning.

First, we quantitatively demonstrate the impact of quasi-specific DNA on the kinetics of the inducible transcription factor Egr-1 using stopped flow

assays on a fluorescence spectrometer. We show that the impact of quasi-specific sites depends strongly on their affinity to Egr-1 and on their concentration, and that the presence of these sites are highly probable within genomic DNA. We show that naturally abundant quasi-specific sites(natural decoys) on DNA can considerably impede the target search processes of sequence-specific DNA-binding proteins.

Second, using a stopped-flow fluorescence method, we examined the kinetic impact of DNA methylation of decoys on the search process of the Egr-1 zinc-finger protein. We analyzed its association with an unmethylated target site on fluorescence-labeled DNA in the presence of competitor DNA duplexes, including Egr-1 decoys. The results suggest that methylated decoys attract methyl binding proteins, effectively blocking them and thereby allowing Egr-1 to avoid sequestration in non-functional locations. This effect may occur *in vivo* for DNA methylation outside CpG islands (CGIs) and facilitate localization of some transcription factors within regulatory CGIs, where DNA methylation is rare.

Lastly, we address the conformational dynamics of Egr-1's arginine side chains' guanidino N_ϵ-H_ϵ moieties when Egr-1 is bound to completely nonspecific and quasi-specific DNA. We compare Egr-1 in complex with both non-specific and quasi-specific DNA. The resulting data gives a more detailed look at the molecular interactions and dynamics during the scanning process. We demonstrate that there is a significant decrease in mobility of some of the Arg side chains when Egr-1 is non-specifically bound. However, there is a crucial

residue which is mobile that may contribute to a conformational switch of the protein from when it is scanning (scanning mode) to when it is bound (recognition mode).

As a whole, these studies suggest that there are significant underlying influences affecting a protein's interaction with DNA. It is possible that DNA is used as a regulatory mechanism of protein interactions through its quasi-specific sequences, methylation, and interaction with other DNA binding proteins.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

RNAP	RNA polymerase
ENCODE	Encyclopedia of DNA Elements
C/EPBα	CCAAT/enhancer binding protein α
Egr-1	Early growth response protein 1
NMR	Nuclear magetic resonance
MeCP2	Methyl CpG binding protein 2
ND	Natural decoy
CGI	CpG islands
MBD	Methyl binding domain
DBD	DNA binding domain
ZF	Zinc finger
FAM	Fluorescein amitide
HISQC	Heteronuclear in-phase single-quantum coherence

Chapter 1

Introduction

1.1 Gene regulation by transcription factors

Gene activation in eukaryotic organisms is a multifaceted process involving numerous types of DNA binding proteins which regulate and transcribe genes to control metabolic activity and development.[1,2] Eukaryotic organisms have a high concentration of DNA in the cell nucleus, which is organized and compacted in nucleosomal arrays. The nucleosomes, made of a complex of proteins called histones, are designed to compact and cover inactive portions of the genome.[3,4] A gene consists of several elements, the core element for general transcription factor binding, the regulatory element for silencing and enhancing transcription factors, and the initiation start site for RNAPII binding and transcription initiation. While general transcription factors are crucial for RNAPII binding and transcription initiation, what often determines the activation or silencing of gene is the binding of transcription factors to sequences within the regulatory element.[5,6] These transcription factors recognize a specific sequence of DNA, which is typically between 6 - 10 base pairs, with more base pairs often indicating a protein with a strong binding affinity because of a high number of electrostatic interactions.[7,8] Once bound to its target sequence, transcription factors recruits a protein complex called mediator. Mediator

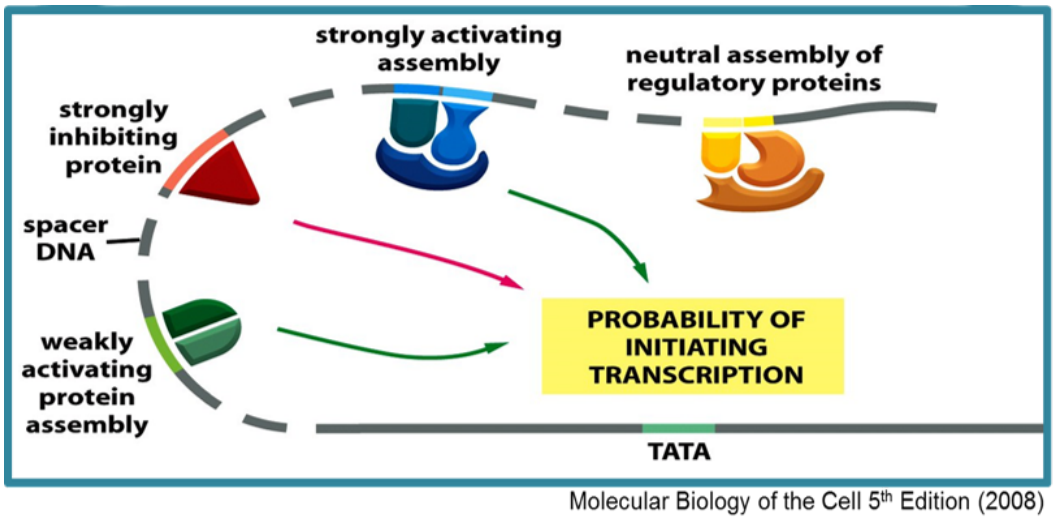

Figure 1.1: Regulatory elements contain a number of specific sites for various transcription factors. The probability of initiating transcription is determined the binding of the appropriate transcription factor to its target site

works with the RNAPII complex to appropriately regulate the gene.[9] While this seems simple, the regulatory element of a gene contains a number of DNA binding sequences which are specific to a particular transcription factor, either for activating, silencing, or enhancing the gene (Figure 1.1). Each transcription factor must appropriately recognize its specific sequence to regulate gene activation. Considering that the DNA landscape is hardly structural diverse, being similar in shape and charge distribution throughout, the possibility of a transcription factor recognizing such a short sequence in a large portion of DNA seems to be an impossible task (Figure 1.2). However, numerous *in vitro* studies have shown that DNA binding proteins can find their specific sequences with great efficiency, although the complexity of the genome *in vivo* is neglected in these studies.[10–13] In fact, the Encyclopedia of DNA Elements

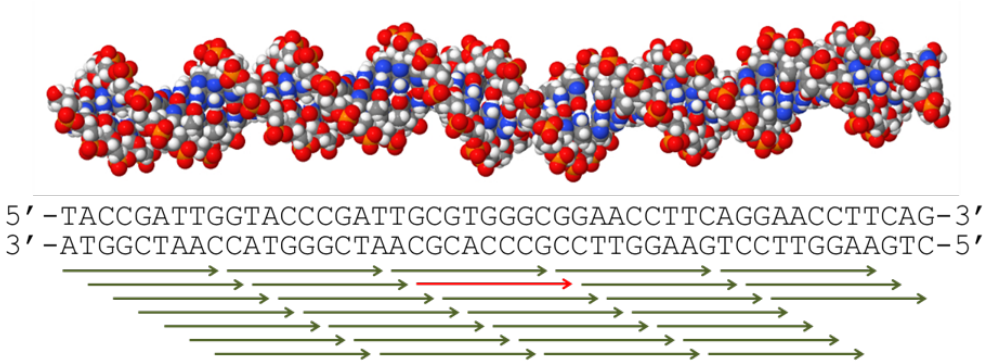

Figure 1.2: DNAs overall uniform structure and electrostatic charge decreases a transcription factors overall probability of reaching its target site. For example, this means that a 9 base pair target (indicated by the arrows) is overall similar to the next 9 base pair site. However, a protein can recognize and bind to a specific set of 9 base pairs (Egr-1s target site indicated by the red arrow) while ignoring the nonspecific DNA.

(ENCODE) project provides extensive information about transcription factor association on genomic DNA.[14] ChIP-on-chip and ChIP-seq methods have allowed for genome-wide studies on the binding sites of eukaryotic transcription factors *in vivo*.[15] Such genome-wide studies showed that transcription factors bind to not only functional sites in the cis-regulatory elements of the genes but also many other apparently non-functional sites,[15,16] though definitions of functional and non-functional are somewhat arbitrary and controversial.[17,18] Because of the high concentration of non-functional or off target sites, it is reasonable to think that these sites significantly affect a transcription factors ability to locate its target site. Even when considering that much of the genome is compacted in nucleosomal arrays, still at least 10% of the genome is free of nucleosomes, leaving a high concentration of DNA available for transcription

factor binding.[4] The question how these transcription factors can still locate their target even in the presence of high concentration of non-functional or off target sites remains to be answered.[1]

1.2 Classical mechanisms for efficient scanning of DNA

Both eukaryotic and prokaryotic organisms require rapid gene activation in response to environmental stimuli. This allows the organism to quickly change and adapt to its external environment. While it was known that DNA binding proteins were involved in this rapid adaptation of the organisms by recognizing specific target sequences found in the genome,[19,20] the mechanisms behind this process was unclear.

Before 1970, it was uncertain how proteins located their target in the genome. However, in 1971, Riggs found that there was a discrepancy between this diffusion rate limit and the second-order binding rate that he calculated from his studies on the lac repressor.[13] In 1981, OG Berg and von Hippel introduced the theory on protein-DNA search, which recognized that transcription factor target search by 3-dimensional diffusion alone was not sufficient for protein kinetics. They realized that the DNA concentration in the nucleus is too high for a protein to find its target in a 3-dimensional search alone, and hypothesized that there were several mechanisms by which a protein can

[1]Paragraph adapted with permission from Kemme, Catherine A., et al. "Regulation of transcription factors via natural decoys in genomic DNA." Transcription 7.4 (2016): 115-120.

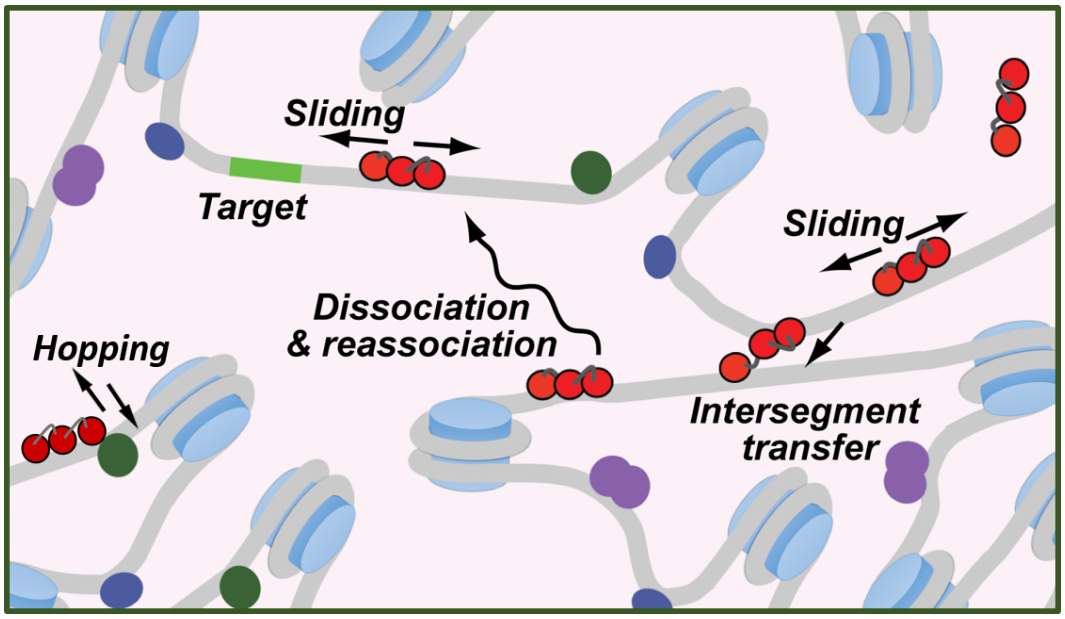

Figure 1.3: Four classical mechanism of target search on DNA by transcription factors. Hopping is also known as micro dissociation and , and sliding are also known as one-dimensional diffusion.

search for its target.[11,12] It was this research that first suggested four different mechanisms for specific sequence recognition by DNA binding proteins (Figure 1.3).

1.2.1 Macroscopic dissociation/reassociation.

This process occurs when the protein completely dissociates from the DNA. Then through 3-dimensional Brownian motion, the protein can interact with a completely uncorrelated location on the DNA. This process is limited by the diffusion constant, which is shown to be less than the measured kinetic rates of some DNA binding proteins.[11,21] Therefore, simple 3-dimensional

diffusion cannot completely explain the efficiency of protein target search.

1.2.2 Microscopic dissociation/reassociation.

This occurs when a protein samples a series of correlated DNA base pairs within the same region. The protein will associate nonspecifically to a DNA sequence, and then dissociate but remain in the adjoining location. Therefore, the protein is still under the influence of the DNAs electrostatic field, causing the protein to reassociate at a location very near its original binding site.[11,12] This allows the protein to be technically still bound to DNA, but enables the protein to test adjacent sites on the DNA sequence. There are a number of MD[22–24] and theoretical studies[25,26] addressing "hopping," but experimental data has yet to prove this mechanism.

1.2.3 One-dimensional diffusion.

This occurs when a protein binds non-specifically to DNA, but relocates to another adjoining place on the DNA strand without completely dissociating from the DNA sequence. The protein breaks some of its hydrogen bonds, while the remainder of the protein non-specifically binds to the DNA sequence. In this fashion, it can perform 1-dimensional Brownian diffusion for a certain length of n base pairs along the DNA sequence, which on average is 35 bp and independent of molecular size.[12] This sliding length, as it is called, is different from hopping in that the protein remains associated to DNA even during translocation. The sliding process ends once the protein completely

dissociates from DNA, or binds specifically to its target site. While sliding prevents a protein from unnecessarily searching in 3 dimensional space, strong interactions slow the sliding process such that this mechanism becomes inefficient.[27] Therefore, this process is most likely used in combination with other mechanisms. Much research has calculated 1D coefficients of a number of proteins,[28–30] and visualized the sliding process in single molecule techniques.[31–35] However, these techniques are not detailed enough to distinguish atomic interactions.

1.2.4 Intersegment transfer.

Considering the high concentration of DNA in the nucleus, it is possible that completely uncorrelated DNA strands are adjacent to one another. A protein can use this to its advantage by transitioning between two uncorrelated DNA strands. A protein can bind simultaneously to both strands of DNA, but because binding to both is unfavorable the protein would either remain on the current strand or move to the uncorrelated strand. This process would allow the protein to explore uncorrelated locations far from where it began its search, and improve the efficiency of target location.

While these classical mechanisms give a good description of how a protein searches for its target DNA, it only demonstrates a protein interacting with its target in the presence of nonspecific DNA. Most studies' main focus demonstrates the efficiency and rapidness of the target search process and few address factors that would decelerate target search.[28,36–39] Classical models

demonstrate a linear relationship between the concentration of protein and the acceleration of search speed, that is to say that as the protein concentration increases, there is a complimentary increase in the target search speed. However, other studies have demonstrated that proteins also bind to off target sequences, and that these off target sequences (natural decoys) affect the concentration of the transcription factor.[40,41] A genes' response to a change in concentration of the transcription factor becomes non-linear and more like a binary on/off switch. Lee and Maheshri demonstrated this non-linear response in budding yeasts by quantitively analyzing the effect of decoys in tandem repeats on target gene expression.[40] In this work, we show kinetic data on the impact of natural decoys(NDs) on the target search kinetics for Egr-1. While NDs would appear to decelerate target search, gene regulation by a transcription factor would be greatly enhanced when its NDs are blocked by other proteins, which weakens functional sequestration. Because most NDs are slightly different from recognition sequences, a subset of NDS for a particular transcription factor may overlap with NDs for other transcription factors. Such overlaps would allow for indirect crosstalk between these transcription factors via NDs. For example, sequestration of C/EBPα in the centromeres is reduced when another transcription factor, Pit-1, is co-expressed, causing an increase in binding of C/EBPα to its functionally important sites to activate its target genes.[42,43] Both the effect of natural decoys and interplay with proteins can be regarded as novel mechanisms whereby transcription factors are regulated in the genome. In this work we discuss both of these novel

mechanisms in chapters 2 and 3.[2]

1.3 Model system for research

To study these mechanisms it is important to have an efficient model system to use for experimentation. For our purposes, a model system must behave well in biochemical experimentation, and be easily prepared for experimentation. It also should demonstrate rapid biological function, in other words, efficient and quick gene activation. Egr-1 is a protein with such characteristics and has been well studied and used for numerous biochemical studies from nuclear magnetic resonance, isothermal calorimetry, to stopped flow fluorescence spectrometry.[21,44–50] The protein was also the basis for zinc finger technology a DNA recognition based gene editing tool designed to manipulate an organism's genome.[51–53] Therefore, for our current studies, Egr-1 is an ideal protein of study.

Early growth response protein 1 (Egr-1) is an inducible transcription factor, which is involved in early development processes and functions as a stress response protein in a number of body systems.[54] In the cardiovascular system it is recruited in response to cardiovascular distress where it assists in repair and inflammatory pathways.[54,55] In the nervous system, Egr-1 activates genes in response to external stimuli for creating memory pathways and main-

[2]Paragraph adapted with permission from Kemme, Catherine A., et al. "Regulation of transcription factors via natural decoys in genomic DNA." Transcription 7.4 (2016): 115-120.

taining synaptic plasticity in the brain.[56,57] Egr-1 has also been reported to be involved in a number of tumor activating pathways and regulates 100 genes overall within the human genome.[58] Many of its functions are time-based responses so Egr1 must quickly locate a specific 9 base pair (GCGTGGCGC) target sequence. Egr-1 has a short half-life (0.5-1 hr), and is expressed in relatively low concentrations (50-100 nM) when induced.[20,54] While low concentrations and a short half-life ensure that the protein does not over activate stress response genes, it is imperative that Egr-1 is efficient in locating in its DNA target. This suggestion seems impossible considering the high density of DNA (100 mg/ml)[14,55] in the nucleus, as well as a myriad of other proteins and small molecules that could affect protein binding. However, previous research in our lab indicates that Egr-1 is quite efficient at locating its target even when presented with high concentrations of nonspecific DNA.[21,45] Egr-1 has several domains; however, only one of these domains is known to bind to DNA.[58] The zinc finger DNA binding domain of Egr-1 has three zinc finger motifs which are connected by a flexible linker region. Each zinc finger recognizes three of the nine base pairs in the recognition sequence, and flexible linker regions allow for incomplete binding during the protein's search mode for its target site.[46] In our previous research we showed that Egr-1 can maintain a balance between its affinity to DNA and the rate at which is locates its target site by altering its conformation between two different modes when bound to DNA. These two modes, termed recognition and search modes (Figure 1.4), differ in that the protein is bound with greater affinity in the recognition mode. We

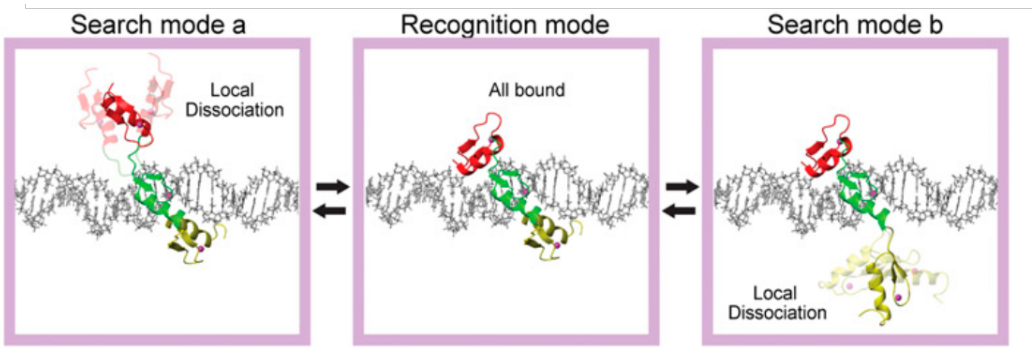

Figure 1.4: There are two possible conformations of zinc finger protein domains when scanning DNA. The search mode in the first and last panels where part of the protein is dissociated from DNA, and the recognition mode in the middle panel, where are zinc finger domains are bound to DNA.(Picture adapted from Zandarashvili, Levani, et al. "Balancing between affinity and speed in target DNA search by zinc-finger proteins via modulation of dynamic conformational ensemble." Proceedings of the National Academy of Sciences 112.37 (2015): E5142-E5149.)

explored this phenomenon of conformational switch by studying the kinetics of Egr-1 after mutating residues in the DNA binding domain which affected electrostatic interactions within in a particular zinc finger domain. This data revealed how changes in electrostatic interactions affected the kinetics and thermodynamics of Egr-1, causing it to switch confirmations when scanning nonspecific DNA or binding to its specific target. When the protein was mutated so that it remained in search mode, its affinity to DNA decreased and the efficiency of target search was improved through faster kinetics. However, when the protein was mutated so that it remained in recognition mode, the opposite effect was seen.[50]

Included in the residues which are significant for specific sequence recog-

nition are basic residues arginine and lysine. Basic side chains are extremely important for recognizing nucleic acids, as they are responsible for electrostatic interactions with the DNA phosphate backbone and for hydrogen binding with nucleic acids. Furthermore, research indicates that basic side chains drive many protein-DNA associations by releasing condensed counterions from DNA.[59–61] To understand more how these basic sidechains interact with DNA upon sequence recognition, Esadze et al. looked at conformational dynamics and mobility of Egr-1s' arginine and lysine residues that interact with DNA using NMR spectroscopy.[62] When the conformational dynamics of the residues were compared between free and bound states of Egr-1, they found that there are significant differences between the mobility of the side chains. They also found notable differences between the dynamics of arginine and lysine side chains when the sidechains are bound to DNA. While many of the arginine sidechains showed much less mobility when bound to DNA, the lysine sidechains and two of the arginine sidechains showed little change in their mobility.[62] Overall, this data indicated that DNA recognition requires both mobile and rigid interactions of sidechains with nucleic acids and the DNA phosphate backbone. These studies give some atomic level information on Egr-1 scanning making it probably the best characterized scanning system so far.

In this current project, we focus on just the DNA binding domain which is responsible for binding to the 9 basepair sequence (GCGTGGCGC) (Figure 1.5). The binding domain is easily purified making it a suitable model system

for studying transcription factor kinetics with NMR and fluorescent spectrometer assays. This project explores Egr-1's activity on a molecular and atomic level, using nuclear magnetic resonance spectroscopy, biochemical assays, and fluorescence spectroscopy. These techniques allowed us to observe Egr-1 activity on a detailed scale. Our previously developed kinetic assays[21,45] using fluorescence spectroscopy, along with a theoretical method to quantify the different molecular mechanisms involved in DNA scanning, including sliding and intersegment transfer are also used for a major portion of this project. Using this methodology and improving it will help us reach our overall objective.

1.4 Overall objective

While the aforementioned mechanisms of transcription factor target search have been experimentally tested by several research scientists, how proteins scan DNA and locate their target sites is not well understood at an atomic level, despite the long history of the field.[11–13] Many studies have shown how different DNA binding proteins can successfully and efficiently navigate a high concentration of nonspecific DNA, but few have addressed how obstacles would enhance or deter protein scanning on DNA. However, it is plausible that target location by proteins under cellular conditions involve a multitude of distinct effects, including protein-protein interactions, kinetic traps at DNA sites analogous to the target (quasi-specific DNA), DNA modifications, and competition of proteins binding to the same target site. In the nucleus, the presence of quasi-specific DNA sites and other DNA binding proteins may

cause conflicting impacts on the protein's activity. The overall objective of this study seeks to understand how Egr-1 is able to scan DNA efficiently in the presence of obstacles such as quasi-specific sites, and potential molecular blockers. Knowledge of how this scanning process occurs in the nucleus would contribute to the development of artificial proteins to be used in gene editing technologies for genetic diseases for which Egr-1 has been a template.[51–53] Previous work in our lab has demonstrated that Egr-1 activity is heavily dependent on the protein's dynamics and affinity to DNA. Studies have shown both theoretically and experimentally how different DNA binding proteins can efficiently navigate a high concentration of nonspecific DNA,[11,12,63–65] but few have addressed how obstacles would alter protein scanning on DNA. However, more attention has been drawn to the scanning process of these proteins, since research suggests that scanning is a crucial part of the proteins ability to bind to its target site properly.[63–65] Our focus in this project is to address issues beyond DNA sequence recognition and sequence specific binding, and to explore additional *in vivo* factors that would affect efficient protein binding to DNA.

1.5 Questions to be addressed

While many researchers have affirmed Bergs theory of binding mechanisms to DNA,[11,12] most of this research has been done under very simplified conditions *in vitro*, or has looked at the scanning process as a whole in single cell studies.[28,36–38] The impact of cellular conditions such as quasi-specific sites

on DNA, DNA methylation, and potential molecular blockers such as MeCP2 on DNA has not been explored. Our project seeks to quantify this impact. We know a protein interacts with its binding sequence with high affinity, but what types of interaction are occurring when the protein binds to nonspecific DNA or DNA with similar affinity to the target? How are a protein's kinetics affected when exposed to a variation in DNA sequence? DNA is constantly being modified (i.e. methylation) to regulate the expression of genes. How does methylation affect the binding and scanning of proteins on DNA? The nucleus not only contains a high concentration of DNA, but also a high number of proteins for maintenance, regulation, and gene activation. How would these proteins influence transcription factor binding? These are just a few of the questions that remain unanswered. Exploration of these factors have allowed us to reveal another layer of complexity in understanding the process of DNA scanning as it would occur *in vivo*. Our project addresses this gap of knowledge through atomic and molecular assays by asking the following questions.

1.5.1 1. How is transcription factor target search affected by the presence of natural decoy DNA? Is there any trapping effect that prevents the protein from binding to its target?

Transcription factors may exhibit significantly strong affinities for quasi-specific sequences (natural decoys NDs) that are similar but not identical to target sequences. In fact, such non-cognate sites are known to play several roles in development.[66] Simple probabilistic estimation implies that NDs are

highly abundant in the nuclei. If a transcription factor recognizes an n-bp sequence of DNA, the transcription factor may also exhibit strong affinity for sequences with a m-bp match ($m < n$) to an n-bp recognition sequence. Given a pool of random sequences, the probability of finding an m-bp match in a window of n bps is given by $2(1/4)^m(3/4)^{n-m}\ {}_n\mathrm{C}_m$ where ${}_n\mathrm{C}_m$ represents the combinatorial, and a factor of 2 accounts for the complementary sequence match. Although somewhat simplistic, this calculation estimates that the total number of NDs ($m = 7$) is $\sim 10^7$ in 3×10^9 base pairs of the human genome for a transcription factor that recognizes 9-bp targets (n = 9). While 90% of the genome is inaccessible due to histones in chromatin, the estimated number of accessible NDs is still 10^6 sites on genomic DNA.[67] In fact, biophysical studies on the inducible transcription factor Egr-1 (for which n = 9) suggested that genomic DNA contain $\sim 10^6$-10^7 NDs, considerably impeding the Egr-1 target search on DNA.[21,41] Compared to these numbers, functionally important target sites are far fewer (Figure 1.6). A typical transcription factor only target $\sim 10^2$-10^3 genes. Therefore, the total number of functional target sites for each transcription factor is approximately 10^2-10^4 sites in the genome, as cis-regulatory elements of each target gene typically involve only several binding sites of the same transcription factor. Thus, NDs overwhelmingly exceed target sites in number. Because of their abundance, NDs should substantially influence transcription factors in vivo. If a nucleus is a sphere with a diameter of 6 μM,[67] a quantity of 10^6 NDs corresponds to a concentration of 10^{-4} M in the nucleus. Even if only 1-10% of NDs are accessible, their con-

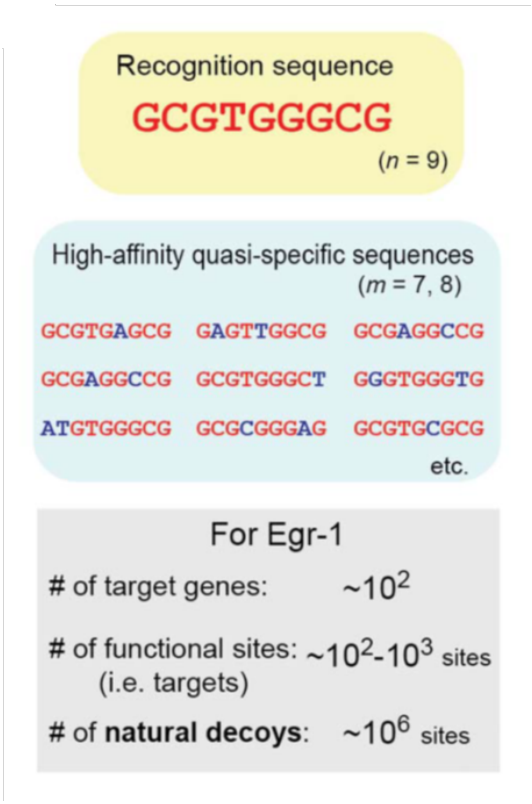

Figure 1.6: Examples of natural decoys (NDs) for a transcription factor. The recognition sequence and high-affinity quasi-specific sequences for Egr-1.

centration is as high as 10^{-5} to 10^{-6} M, which is far greater than the typical dissociation constants (10^{-10}- 10^{-7} M) for specific or quasi-specific DNA complexes of transcription factors. Consequently, binding to NDs would effectively sequester transcription factors and preclude them from binding to functional target sites. A well-characterized example is the ND for the transcription factor CCAAT/enhancer-binding protein α(C/EBPα). Because many NDs for this protein exist in tandem repeats (typically, thousands of copies) of 171-bp α-satellite DNA in the centromere regions of mammalian chromosomes, the C/EBPα molecules are effectively sequestered in the centromeres. This sequestration reduces the transcriptional capability of C/EBPα.[42,43] An altered specificity mutation of C/EBPα, which reduces binding to α-satellite DNA but permits binding to the functional target sites, causes an elevation in binding of C/EBPα to a promoter and an increase in transcriptional output from the promoter.[42] This phenomenon suggests that NDs play a role in gene regulation via the functional sequestration of the transcription factor. Sequestration in NDs may also have a positive impact on transcription factors. Burger et al. conducted a theoretical study on the potential role of NDs as protectant for transcription factors.[68] If their DNA-bound states are less susceptible to proteolysis, NDs may prolong the mean lifetimes of transcription factors, partially offsetting the negative consequences of functional sequestration in non-functional regions.[3]

[3]Section adapted with permission with permission from Kemme, Catherine A., et al. "Regulation of transcription factors via natural decoys in genomic DNA." Transcription 7.4 (2016): 115-120.

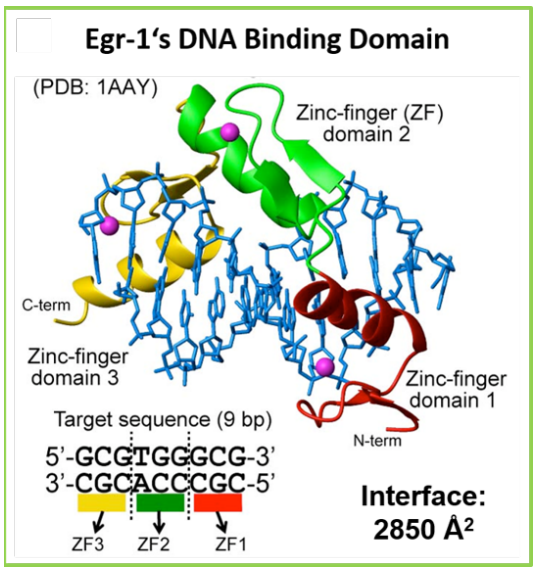

Figure 1.5: The DNA binding domain of Egr-1 bound to its target site. Each zinc finger binds to 3 base pairs, shown the yellow, green, and red bars below the recognition sequence.

In Chapter 2, we address our questions using fluorescence methods and have investigate 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. By stopped-flow assays, we measure 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 decelerates the target association of Egr-1. The impact of quasi-specific sites depends strongly on their affinity, their concentration, and the degree of their binding to the protein. To demonstrate the presence of natural quasi-specific sites, we repeat our experiments with calf thymus DNA as a competitor. To quantitatively describe the kinetic impact of quasi-specific sites, we derive an analytical form of the apparent kinetic rate constant for the target association and use it to fit the experimental data.[4]

[4]Paragraph adapted with permission from Kemme, Catherine A., et al. "Influence of quasi-specific sites on kinetics of target DNA search by a sequence-specific DNA-binding protein." Biochemistry 54.44 (2015):

1.5.2 2. How can TFs reach functionally important sites despite the enormous presence of NDs?

DNA methylation may also control the sequestration of transcription factors in NDs. The methylation of CpG dinucleotides in DNA attracts methyl-CpG-binding proteins, such MBD1, MBD2, and MeCP2.[69] These proteins may block NDs for some transcription factors (eg., ATF2, Egr-1, Elf1, E2F4, HIF1α, Nrf1, Sp1, and USF1)[70] that recognize sequences containing CpG dinucleotides. For example, the 9-bp recognition sequence of Egr-1 contains two CpG dinucleotides (see Figure 1.5), but their methylation does not affect the intrinsic affinity of Egr-1 for its target DNA.[49] Many functionally important target sites of these transcription factors are located in CpG islands (CGIs). Because the total length of all CGIs is less than 1% of the genome size,[69] the vast majority of NDs should be located outside CGIs. Interestingly, CpG methylation is rare ($< 10\%$) in CGIs, whereas the overall level of CpG methylation in genomic DNA is $\sim$85%.[71] Therefore, the target site in CGIs are likely unmethylated, whereas CpG dinucleotides within or near NDs are methylated. This distribution may enable methyl-CpG-binding proteins to selectively block NDs, assisting transcriptional activators in binding to functionally important sites within CGIs.[5] (Figure 1.7)

In chapter 3, we examine the kinetic impact of DNA methylation of

6684.http://pubs.acs.org/doi/abs/10.1021%2Facs.biochem.5b00967

[5]Paragraph adapted with permission from Kemme, Catherine A., et al. "Regulation of transcription factors via natural decoys in genomic DNA." Transcription 7.4 (2016): 115-120.

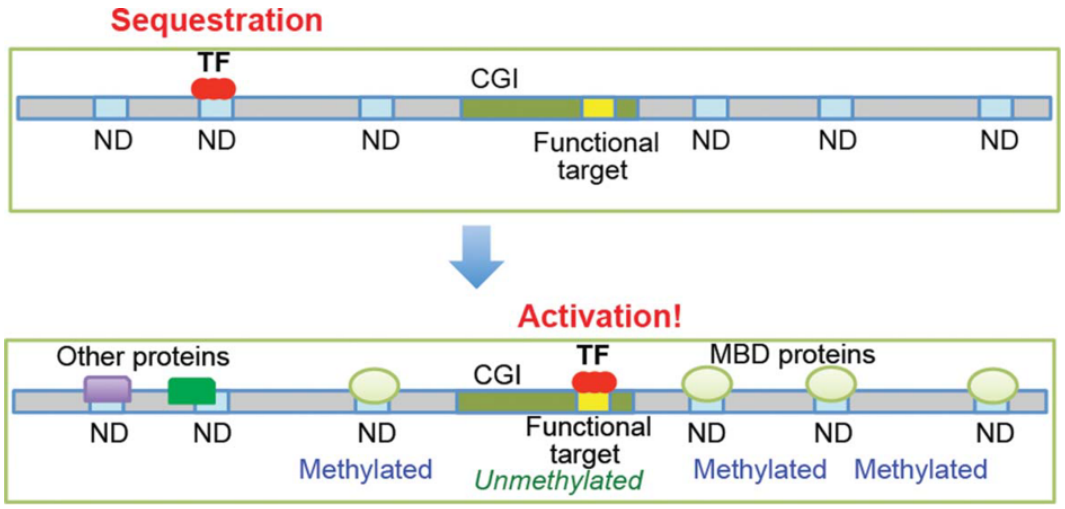

Figure 1.7: Top panel) Transcription factors, without the presence of other proteins, could be sequestered by off target sequences, and consequently never reach their target site within CGIs. Bottom panel) With the presence of other proteins, transcription factors can activate their functional target because the natural decoys are bound.

decoys on the search process of Egr-1. We analyze its association with an unmethylated target site on fluorescence-labeled DNA in the presence of competitor DNA duplexes, including Egr-1 decoys. We show that for Egr-1, DNA methylation alone does not affect target search kinetics. In the presence of MeCP2's methyl-CpG binding domain (MBD), however, DNA methylation of decoys substantially (∼10-30 fold) accelerated the target search process of Egr-1. This acceleration does not occur when the target is methylated. These results suggest that decoy methylation binds MBD proteins, which block methylated decoy and allow Egr-1 to avoid sequestration in non-functional locations. This effect may occur *in vivo* for DNA methylation outside of CpG islands (CGIs) and could facilitate localization of transcription factors within CGIs. [6]

1.5.3 3. How do transcription factors scan DNA? How do interfacial side chains scan DNA bases?

Transcription factors locate their DNA target site through the process of scanning, which consists of nonspecific interactions with DNA. Nonspecific interactions are thought to be significant in protein scanning because they act as a "guide" for protein translocation through one-dimensional sliding or intersegment transfer.[72] Many of transcription factors contain basic amino acids (arginine and lysine) which are used to recognize nucleic acids when binding to

[6]Paragraph adapted from Kemme, Catherine A., et al. "Potential role of DNA methylation as a facilitator of target search processes for transcription factors through interplay with methyl-CpG-binding proteins." Nucleic Acids Research (2017). under the Creative Commons License (http://creativecommons.org/license/by-nc/4.0/)

DNA. The side chains are extremely important for hydrogen bonding to DNA base pairs or for electrostatic interactions between side chains and the DNA backbone.[73–75] Their importance has also been discussed in thermodynamic studies, as they play a role in condensed counterion release when a transcription factor binds to DNA.[59–61] Previous studies by (Esadze et al.) have shown the dynamic difference of Egr-1's basic sidechains when it is free or bound to its target, and indicate that there are significant changes in mobility of both arginine and lysine sidechains. While basic sidechains which contacted the phosphate group on the DNA backbone remained significantly mobile, the basic sidechains which hydrogen bond to DNA bases become strongly immobilized.[62] In 2006, Iwahara et al showed that arginine sidechains involved in HoxD9 binding to DNA are indeed more mobile when bound nonspecifically to DNA sequences than when bound to its specific sequence.[72]

In Chapter 4, we use various NMR assays to calculate the S^2 order parameters which quantitatively describes the dynamics/mobility of a given molecule. We discuss the dynamics of Egr-1's arginine sidechains when Egr-1 is bound to completely nonspecific DNA compared to when it is bound to quasi-specific or specific DNA. We found that Egr-1 in complex with quasi-specific DNA causes a decrease in mobility for many of the arginine sidechains, while side chains are much more dynamic when Egr-1 is in complex with non-specific DNA.

1.6 Significance

Our current work provides a new perspective on target DNA search mechanisms for transcription factors. Over the past four decades, mechanisms allowing transcription factors to rapidly locate their specific targets in the genome have been the subject of considerable interest in biophysics and biochemistry.[10,63–65] While recent studies using nuclear magnetic resonance (NMR) and single-molecule techniques revealed great details on how proteins scan DNA efficiently,[28,36–38] studies that focus on factors that impede the search process have been rare. Numerous nonfunctional high-affinity sites in the genome could serve as decoys that trap transcription factors and affect their functions. Our study shows that the target search process for transcription factors can be greatly accelerated when other proteins block such decoys. This mechanism does not require any direct protein-protein interactions, although the current paradigm for synergy between transcription factors typically assumes their direct interactions.[76,77] In addition, our work shows that MeCP2 MBD facilitates the association of the Egr-1 zinc-finger protein with unmethylated target by blocking methylated decoys. In principle, other proteins could also provide the same acceleration mechanism, as long as they selectively block decoys but not targets. For example, highly expressed transcription factors with similar but different sequence specificity may enhance the function of other transcription factors by selectively blocking their decoys. Natural decoys and their DNA methylation may play key roles in regulation of transcription factors. By studying these characteristics, we will develop a

better understanding of transcription factor interactions *in vivo*.[7]

Studying the mechanisms behind transcription factor target search also allows researchers to apply this knowledge to possible gene regulatory therapeutics through the engineering of artificial proteins for DNA binding. For example, Egr-1, an inducible transcription factor, has been used as a template for zinc finger technology, a transcription factor-based therapeutic designed for genome editing. This therapeutic applies the knowledge of DNA sequence recognition to remove or add portions of the genome.[51–53] It has been applied to an HIV therapeutic in clinical trials, as well as a gene editing tool for animal models in scientific experiments.[78] By improving our understanding of the mechanisms behind protein DNA recognition, engineered proteins can be perfected to fit the physiological criteria for protein DNA binding.

[7]Paragraph adapted from Kemme, Catherine A., et al. "Potential role of DNA methylation as a facilitator of target search processes for transcription factors through interplay with methyl-CpG-binding proteins." Nucleic Acids Research (2017) under the Creative Commons Attribution License.(http://creativecommons.org/license/by-nc/4.0/

Chapter 2

Effects of natural decoy DNA on the target search kinetics of Egr1 [1]

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 numbers of sites on DNA, which positively and negatively impact the kinetics of the proteins target search. Nonspecific sites near targets can accelerate the target association process by creating an antenna that directs the proteins to its target through one-dimensional diffusion along DNA ("sliding").[10,39,63–65,79] 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.[11,64]

Since the discovery of amazing rapid association of Escherichia coli lac

[1] Chapter adapted with permission from Kemme, Catherine A., et al. "Influence of quasi-specific sites on kinetics of target DNA search by a sequence-specific DNA-binding protein." Biochemistry 54.44 (2015): 6684. http://pubs.acs.org/doi/abs/10.1021%2Facs.biochem.5b00967

repressor with operator DNA in 1970,[13] 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.[10–12] Mainly in the 21st Century, these translocation processes were directly confirmed by biophysical methods such as nuclear magnetic resonance (NMR) and single-molecule techniques.[28,36–38] Meanwhile, studies that focus on factors that decelerate the search process remain rare.[39] Trapping of proteins at nonfunctional sites on DNA could be prevalent in the nucleus because of extremely high DNA density($\sim$100 mg/mL).[3] Even though 80% of the DNA is covered by histones,[4] the concentration of accessible DNA (i.e., linkers) in the nuclei is estimated to be as high as $\sim$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[80,81](Figure 2.1). 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.[82,83] In the nervous system, Egr-1 functions as a regulator of synaptic plasticity to promote memory formation.[56,57] In the cardiovascular system, Egr-1 mediates the formation of scar tissue and inti-

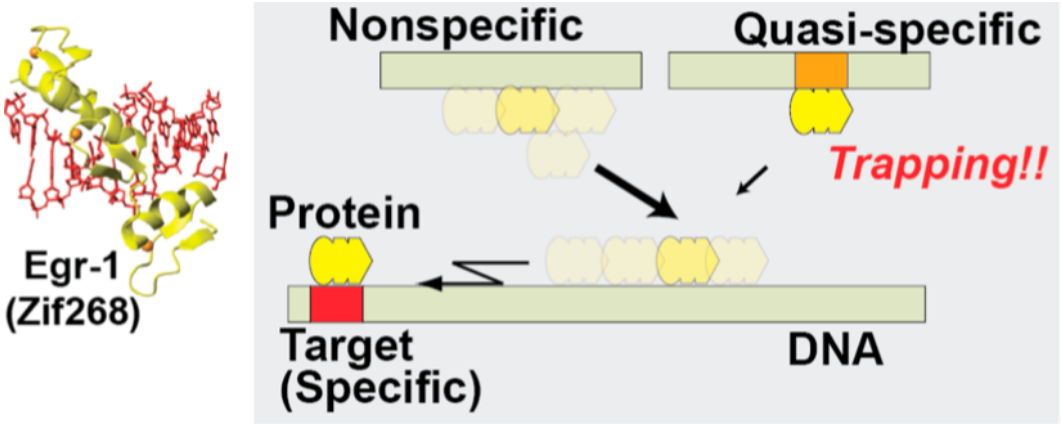

Figure 2.1: Hypothetical model showing the trapping effect caused by the presence of quasi-specific sites. Decreased dissociation from the high affinity quasi-specific site would reduce the available concentration of Egr-1, preventing it from binding to the actual target site.

mal thickening in response to damage cause by cardiovascular injury.[54,55] 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 hr.)[54] 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, and show that there is a decrease in target search efficiency due to the presence of quasi-specific DNA (Figure 2.1). This chapter 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.

2.1 Materials and Methods

2.1.1 Egr-1 preparation and purification

The protein construct used in this study was the Egr-1 DNA binding domain (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. A chemically synthesized gene from DNA oligonucleotides which encodes the zinc-finger domains of Egr-1 (residues 335 - 432) was amplified by PCR and inserted in a pET-49b vector (Novagen). After transfecting BL21(DE3) Escherichia coli cells with the plasmids, we grew a 4-liter culture at 37°C in minimal media to express Egr-1 zinc finger domain fused with glutathione-S-transferase. To isotopically label the protein, 13C glucose and 15N ammonium chloride were used in the minimal media. Once the O.D.600 of the culture reached 1.0 -1.2, we induced protein expression with 0.4 mM isopropyl-β-D-1-thiogalactopyranoside, and added 50 μM $ZnCl_2$. Upon induction, the culture temperature was reduced to 24°C and cultivated for ∼18 hours. The cells were then collected and centrifuged at 4000 × g and resuspended in a buffer containing, 50 mM Tris-HCl (pH 7.5), 500 mM NaCl, and 5% glycerol, and 1% Triton X-100 and 2 tablets of EDTA-free protease inhibitor cocktail (Roche). The cells were disrupted through sonication and then centrifuged for 20 min at 30000×g and 4°C. The supernatant was loaded onto a GSTPrep FF column (GE Healthcare) and washed with 50 mM Tris-HCl (pH 7.5), 400 mM NaCl, and 1% Triton X-100. We then eluted the GST-Egr-1 zinc finger domain fusion protein with 50 mM Tris-HCl, 400 mM NaCl, 10 mM glutathione, pH 7.5. The

fusion protein was cleaved with 100 units of HRV-3C protease (GE Healthcare) overnight at 4°C and confirmed by SDS-PAGE. We then concentrated the protein to 10 mL with an Amicon Ultra-15 device (Millipore) and purified with Sephacryl S100 26/60 column (size-exclusion chromatography from GE Healthcare) using a buffer of 50 mM Tris-HCl(pH7.5), 1 M NaCl, 2mM β-mercaptoethanol (BME), and 0.2 mM $ZnCl_2$. Fractions containing Egr-1 were confirmed using SDS-PAGE, then collected and buffer exchanged to 50 mM Tris-HCl (pH 7.0), 200 mM NaCl, 2 mM β-mercaptoethanol, 5% glycerol, and 100 mM $ZnCl_2$. The protein was purified using a Mono-S 5/5 cation exchange column (GE Healthcare) equilibrated with the same buffer and eluted with a gradient from 200- 1 M NaCl.

2.1.2 DNA preparation.

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 2.2A.) The same 143 bp probe DNA was used in previous studies. A 5' FAM labeled 33-mer single-stranded DNA primer was purchased from Integrated DNA Technology (IDT), along with a reverse primer, and a pUC-19 plasmid as a template. This DNA duplex was generated by polymerase chain reaction (PCR) with the FAM-labeled 33-mer primer, an unlabeled reverse primer, and the pUC19 plasmid (New England BioLabs) using Vent DNA polymerase (New England Biolabs). After the reaction, we purified it with a Resource-Q anion exchange column eluting the

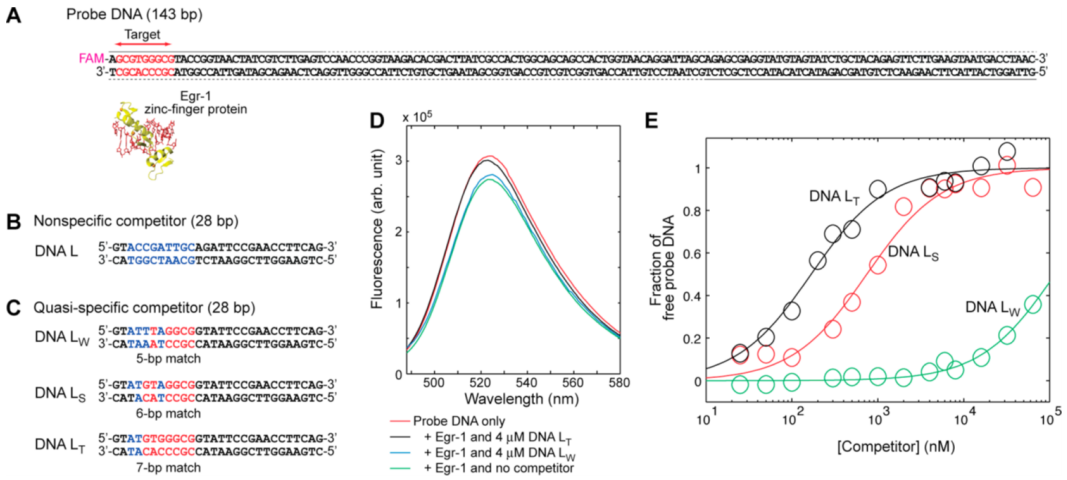

Figure 2.2: Measurement of relative affinities of quasi-specific DNA duplexes for the Egr-1 zinc finger protein. (A) FAM-labeled 143 bp DNA duplex as the probe DNA. The Egr-1 target site is colored red. The same probe DNA was used in our previous studies. (B) Nonspecific competitor DNA. This 28 bp duplex termed DNA L does not contain any sites similar to Egr-1. This nonspecific DNA was also used in our previous studies. (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 meausured for 2.5 nM probe DNA. (E) Data from the competition assays. FAM fluorescence was measured for the solutions fo 2.5 nM probe DNA, 30 nM protein, and competitor DNA at varied concentrations in 10 mM Tris-HCl (pH 7.5), 0.2 muM $ZnCl_2$, 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.2.3.

DNA with a gradient of 0 to 1.5 M NaCl, with a buffer of 50 mM Tris-HCl (pH 7.5) and 1 mM ethylenediaminetetraacetic acid. The fractions from the purified reaction mixture were collected and concentrated down, then purified via polyacrylamide gel electrophoresis (PAGE) using 4-20% gradient polyacrylamide/TBE gels (Invitrogen). The PCR product was then cut out from the gel, crushed, and shaken at room temperature in 10 mM Tris-HCl (pH 7.5) and 40 mM KCl for 24-36 hours. Once the PCR product sufficiently diffused into solution, it was purified with the PCR purification kit (Qiagen). 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 2.2B), was purchased as single stranded DNA from IDT and purified via Mono-Q anion exchange chromatography. The complementary strands were quantified, annealed, and repurified on the Mono-Q after annealing. The other three 28 bp competitor 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 2.2C). Each chemically synthesized DNA strand was purchased from Integrated DNA Technologies and purified in the same manner as DNA L. Calf thymus DNA was purchased from Invitrogen and sonicated for fragmentation into an average size of $\sim$500 bp, which was confirmed by 0.9% agarose gel electrophoresis in TBE buffer (Invitrogen).

2.1.3 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 competition 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 $ZnCl_2$, 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 macro-molecular 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 DNAs (D_{tot}), protein (P_{tot}), and competitor DNA (C_{tot}) satisfy the relationship $D_{\mathrm{tot}} << P_{\mathrm{tot}} << C_{\mathrm{tot}}$, the fraction of the probe DNA in the free states (p_{free}) is given by[49]

$$p_{\mathrm{free}} = \frac{1 + C_{\mathrm{tot}}/K_{\mathrm{d(comp)}}}{1 + C_{\mathrm{tot}}/K_{\mathrm{d(comp)}} + P_{\mathrm{tot}}/K_{\mathrm{d(probe)}}} \tag{2.1}$$

where $\mathrm{K_{d(comp)}}$ and $\mathrm{K_{d(probe)}}$ 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}

$$I_{\mathrm{obs}} = p_{\mathrm{free}} I_{\mathrm{free}} + (1 - p_{\mathrm{free}})\, I_{\mathrm{bound}} \tag{2.2}$$

where I_{free} and I_{bound} are intrinsic fluorescence intensities for free and protein-bound probe DNA duplexes, respectively. If $C_{\mathrm{tot}} >> K_{\mathrm{d(comp)}}$, eq 2.2 becomes a simple expression:

$$I_{\mathrm{obs}} = \frac{p_{\mathrm{free}} I_{\mathrm{free}} + P_{\mathrm{tot}} I_{\mathrm{bound}}}{\Gamma C_{\mathrm{tot}} + P_{\mathrm{tot}}} \tag{2.3}$$

The parameter Γ represents a relative affinity defined as $K_{\mathrm{d(probe)}}/K_{\mathrm{d(comp)}}$. This equation was used to determine the relative affinity Γ of the quasi-specific DNA duplexes via nonlinear least-squares fitting to the experimental $\mathrm{I_{obs}}$ data as a function of $\mathrm{C_{tot}}$. 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 $\mathrm{I_{free}}$, the fluorescence intensity of the free state of the probe DNAs, which was directly measured. The fitting calculations were performed with MATLAB software.

2.1.4 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 (∼0.5ml) 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 use the synthetic 28 bp duplexes show in Figure 2.2 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 μg/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, all binding reactions were conducted under conditions of $D_{\text{tot}} << P_{\text{tot}} << C_{\text{tot}}$. 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 lest-squares fitting with

$$I(t) = I_\infty + (I_0 - I_\infty)exp(-k_{\text{app}}t) \tag{2.4}$$

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.

2.2 Results

2.2.1 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 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 2.2B). 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 2.2C). 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 CGCTGGGCG, 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 competitions assays,[49] 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 2.2D). 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_d <0.1 nM) under the current conditions.[21,45] Addition of high affinity quasi-specific DNA increased the unbound targets due to transfer of protein from the target to the competitor, thereby reducing the fluorescence quenching affect (Figure 2.2D). 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 2.2E). Competitor DNA duplexes at high concentrations out competed 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 non-linear least-

squares fitting with eq 2.3. The first two concentration points were excluded from the fitting calculations because these concentrations do not satisfy the inequality $P_{tot} << C_{tot}$, which is required for eq 2.3. The best fit curves are shown in Figure 2.2E. Value of $\Gamma = K_{\mathrm{d}}^{\text{quasi-specific}}/K_{\mathrm{d}}^{\text{specific}}$ for DNA duplexes L_T, L_S, 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 kinetics impact of quasi-specific sites, as described below.

2.2.2 Impact of quasi-specific sites on the kinetics of target search.

By stopped flow-fluorescence assays similar to those described in our previous studies, we investigated the influence of the quasi-specific DNA on the target search kinetics of Egr-1.[21,45] The basic scheme for the kinetic experiment is depicted in Figure2.3A. 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

Fig 2.3B. 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 out-competed 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 2.3B). Time courses for the fluorescence intensity were found to be mono-exponential. 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 μM, 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 2.3C, or L_S (Figure 2.3D). For the quasi-specific DNA L_T, only a single concentration of 50 nM was tested (Figure 2.3E) 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 ($\sim 20\ s^{-1}$). 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 con-

centration was linear (black in Figure 2.3C-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 2.3C). 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 2.3D, 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.

2.2.3 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 $D_{\text{tot}} << P_{\text{tot}} << C_{\text{tot}}$ 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:[21]

$$k_a = \rho\eta S k_{\text{on,n}} \tag{2.5}$$

d

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;[21,64,84] 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 the enhancement factor ($\eta > 1$) due to intersegment transfer. On the basis of the discrete stochastic kinetic model of Veksler and Kolomeisky,[85] we previously gave explicit forms of parameters η and S as functions of various kinetic rate constants, equilibrium constants, and configurational factors.[21] When $D_{\mathrm{tot}} << P_{\mathrm{tot}} << C_{\mathrm{tot}}$ parameter ρ is given by

$$\rho = \frac{1}{Z} = \frac{1}{1 + N_{\mathrm{tot}}/K_{\mathrm{d,n}}} \tag{2.6}$$

Where Z corresponds to a partition function for protein at the pseudoequilibrium during the target search processs, $\mathrm{N_{tot}}$ 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 ν 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 previous work.[45] Under the assumptions of the pseudoequilibrium during the target search process, the trapping effect is represented by the following parameter ρ_{nq}:

$$\rho_{\mathrm{nq}} = \frac{1}{1 + N_{\mathrm{tot}}/K_{\mathrm{d,n}} + [Q]/K_{\mathrm{d,q}}} \tag{2.7}$$

where Z_{nq} represents a partition function in the form of the binding polynomial 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 $K_{\mathrm{d,q}}$ is the dissociation constant for each quasi-specific site.

Eq. 2.7 together with eq. 2.5 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 2.3C-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., $K_{\mathrm{d,q}} << K_{\mathrm{d,n}}$) of the quasi-specific site and a large fraction of its free state can make the $[Q]/K_{\mathrm{d,q}}$ term predominant in partition function Z_{nq}, rendering $\rho_{\mathrm{nq}} << \rho$.

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 $\rho_{\mathrm{nq}} << \rho$. 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

$$[Q] = \frac{1}{2}(Q_{\mathrm{tot}} + P_{\mathrm{tot}} + K_{\mathrm{d,q}}Z) + \frac{1}{2} \tag{2.8}$$

where $Z = 1 + N_{\mathrm{tot}}/K_{\mathrm{d,n}}$(i.e., the same as Z in eq.2.6). This expression is derived by solving the equations $K_{\mathrm{d,n}} = N_{\mathrm{tot}}[P]/[NP]$, $K_{\mathrm{d,q}} = [Q][P]/[QP]$, $P_{\mathrm{tot}} = [P] + [NP] + [QP]$, and $Q_{\mathrm{tot}} = [Q] + [QP]$, 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

$$k_{\mathrm{app}} = (\rho_{\mathrm{nq}}/\rho)k_{\mathrm{a0}}P_{\mathrm{tot}} \tag{2.9}$$

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 2.3, we conducted fitting calculations with eqs 2.6-2.9 via optimization of two parameters,k_{a0}

and $K_{\mathrm{d,q}}$. These calculations require the experimental value of the dissociation constant ($K_{\mathrm{d,n}}$) for the affinity of each nonspecific site. In a previous study,[21] we determined $K_{\mathrm{d,n}}$ 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 2.3C-E. The fitting gave good agreement with the experimental data. From these fittings to the kinetic data, $K_{\mathrm{d,q}}$ values of DNA duplexes L_T, L_S, 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.

2.2.4 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 $\sim$ 500bp) was used instead of synthetic duplexes such as DNA L, $\mathrm{L_T}$, $\mathrm{L_S}$, $\mathrm{L_W}$. 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 2.4 shows the dependence of measured k_{app} constants on 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 2.4 (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 probablity f_{q} for quasi-specific sites, with which $Q_{\mathrm{tot}} = f_{\mathrm{q}} N_{\mathrm{tot}}$ in eqs 82.8 and 92.9, and optimized four parameters: f_{q}, $K_{\mathrm{d,q}}$ and two k_{a0} 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 as uniform $K_{\mathrm{d,q}}$ for all quasi-specific sites. Therefore, the affinity($K_{\mathrm{d,q}}$) and concentration (Q_{tot}) from these calculations should be regarded merely as apparent parameters. The global fitting calculations with eqs 6-9 showed excellent agreement with both experimental data sets (solid lines, Figure 2.4) 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 value for the apparent affinity ($K_{\mathrm{d,q}}$) 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.

2.3 Discussion

2.3.1 Trapping at nonfunctional sites

Recently, methods such as ChIP-on-chip[86] and ChIP-seq[14] 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.[16,87] As these methods detect only high occupancy of transcription factors at sites with the strongest affinities,[88] 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).[7,8] 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.[80]

Our current results from these kinetics experiments 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 k sequences, the probability of finding m bp match in a window of n bp covered by a transcription factor is given by

$$P_{\mathrm{n,m}} = 2(1/4)^m(3/4)^{n-m}{}_nC_m \tag{2.10}$$

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.

2.3.2 Switch like response via natural decoys

When NDs sequester a transcription factor, the regulation of its target genes requires a higher concentration of the transcription factor. More importantly, the genes response to a change in the concentration of the transcription factor becomes non-linear and more like a binary on/off switch. Until the binding to the decoys is saturated, target association is not significant because the transcription factors are trapped at the NDs before reaching the targets (Figure 2.5). The concentration at which the saturation occurs corresponds to the threshold for the "on" state of the switch. At this point, the inhibitory effects of NDs are eliminated. When the level of a transcription factor exceeds this threshold, the target association of the transcription factor is drastically enhanced.

If the threshold exist between normal (100%) and 50% concentrations of a transcription factor, the heterozygous $(+/-)$ and homozygous $(-/-)$ knockouts of this transcription factor should result in almost equal changes in expression levels of this target genes. In fact, for example, a study $Dmp1^{+/+}$,$Dmp1^{+/-}$, $Dmp1^{-/-}$ mice showed such results for the transcription factor Dmp1 (Dmtf1). The on-off switch-like response via NDs may also be relevant to a sharp spatial boundary of expression in response to the gradient of the transcription factor

Bicoid, a morphogen in *Drosophila* development. In the switch model, as long as a transition between the "on" and "off" phases is involved, even a relatively moderate up or down regulation of a transcription factor may result in drastic changes in expression levels of this target genes. This might be relevant to some diseases, where protein levels are often affected by internal or genetic mutations.[2]

2.3.3 Potential of role of other proteins in the regulation of transcription factors.

The adverse effect of the quasi-specific sites on target association given two important implications. First, a relatively high expression level of a transcription factor is required for efficient regulation of their target genes, unless other proteins occupy the nonfunctional quasi-specific sites. 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 inacessible 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

[2]Section adapted with permission from Kemme, Catherine A., et al. "Regulation of transcription factors via natural decoys in genomic DNA." Transcription 7.4 (2016): 115-120.

Egr-1 target sequences contain two CpG dinucleotides, yet their methylation does not weaken association of Egr-1 with target DNA *in vitro*. Interestingly, a genome wide ChIP-on-chip study of Egr-1 binding sites showed that 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. Because of this distribution, it is likely that methyl-CpG-binding proteins do not block functional target sites for Egr-1 in the CpG islands but blcok the majority of quasi-specific sites. Western blot and DNA association data for nuclear extracts (e.g., refs) 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.

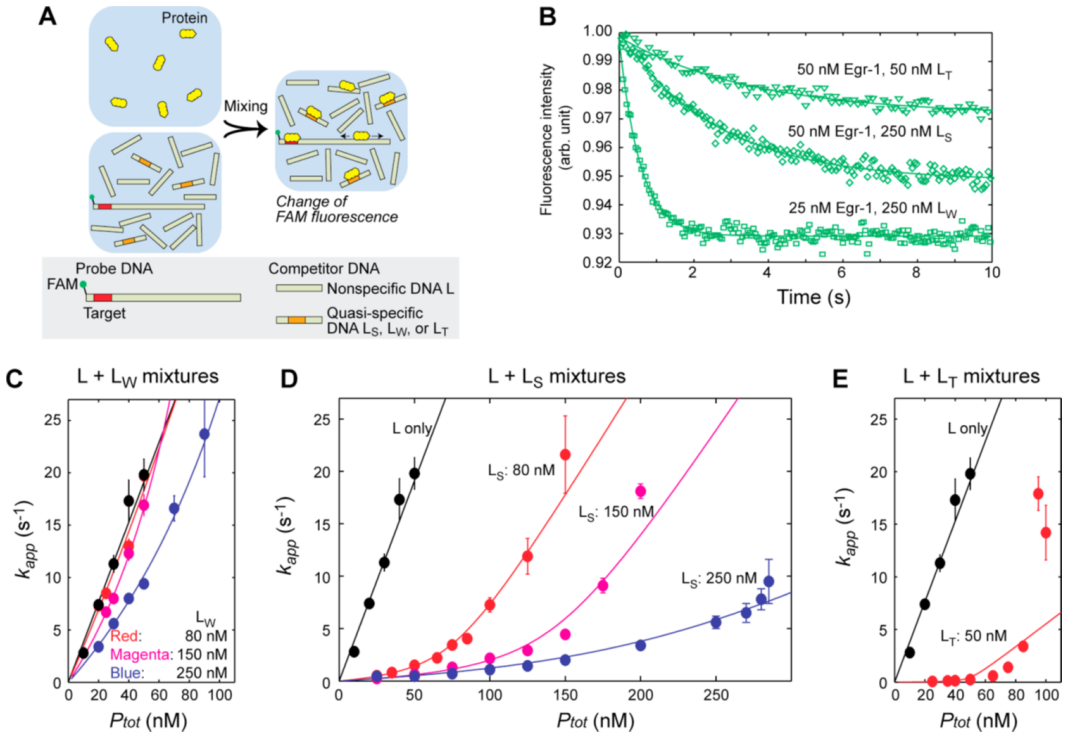

Figure 2.3: Impact of quasi-specific DNA on the target search kinetics of the Egr-1 zinc finger protein. (A) Schematic of the stopped flow fluorescence assay for investigating the impact of quasi-specific DNA. In this assay, the change in FAM fluorescence was monitored upon mixing the solution of the 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 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), and 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 non-linear least-squares fitting with eq.2.6, eq.2.7, eq.2.8, and eq.2.9. In these calculations, only two parameters, $K_{\mathrm{d,q}}$ and k_{a0}, were optimized. The buffer conditions for these experiments were 10 mM Tris-HCl (pH 7.5), 0.2 μM $ZnCl_2$, 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.

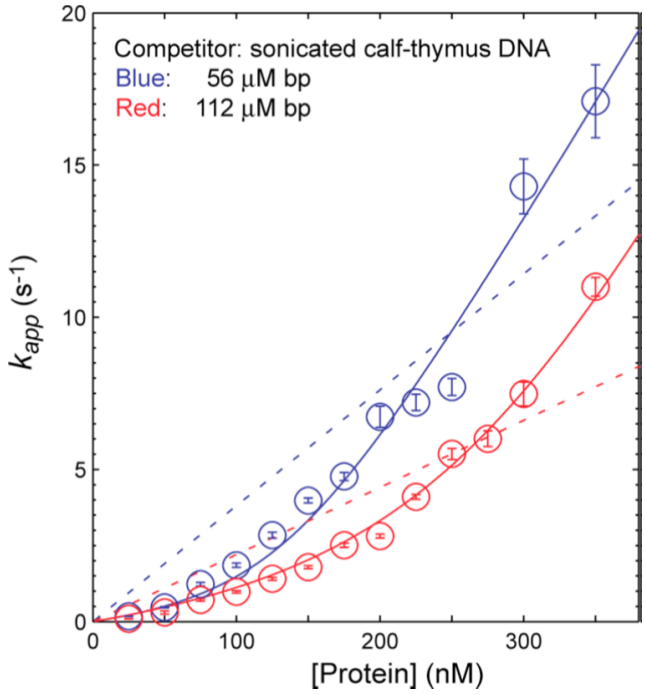

Figure 2.4: Evidence of the kinetic influence of natural quasi-specific sites on the target search process by Egr-1. The graph shows the protein concentration dependence of k_{app} constants measured with the stopped flow assay using calf thymus DNA as a competitor. To reduce the viscosity, calf thymus 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 2.6-2.9. The fitting calculation was performed for the two data sets simultaneously. In this calculation, four fitting parameters were optimized: two k_{a0} parameters at two different overall DNA concentrations, the apparent affinity ($\mathrm{K_{d,q}}$), and the 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 ± 0.03%, and $\mathrm{K_{d,q}}$ = 3.7± 0.8 nM. These data suggest that there are ∼10^6-10^7 quasi-specific sites with high affinity for Egr-1 in the genomic DNA.

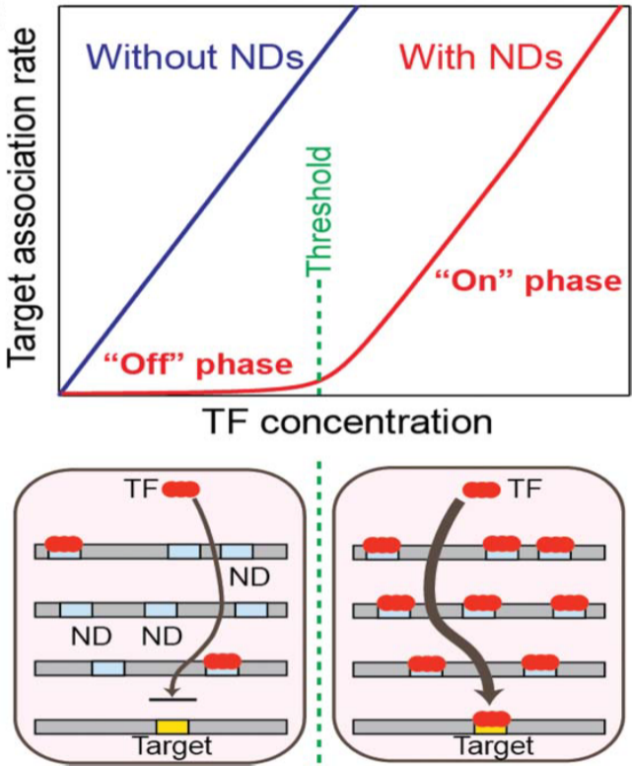

Figure 2.5: The influence of NDs on the target association kinetics of a transcription factor (TF). NDs preclude TFs from their targets. When all NDs are occupied by proteins, target association is drastically increased because the inhibitory effects of the NDs are eliminated. This phenomenon creates as on-off switch-like response.

Chapter 3

Influence of methylation in the presence of MeCP2 on the target search kinetics of Egr-1[1]

In chapter 2, we demonstrated 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 much faster because the strong trapping effect becomes absent . Given this observation, it is reasonable to consider that a large number of decoys makes it difficult for Egr-1 to adequately occupy each functional target because sequestration of Egr-1 molecules may occur in off target locations.

Despite these circumstances, how can Egr-1 reach functionally impor-

[1]Chapter adapted from Kemme, Catherine A., et al. "Potential role of DNA methylation as a facilitator of target search processes for transcription factors through interplay with methyl-CpG-binding proteins." Nucleic Acids Research (2017) under the Creative Commons License (http://creativecommons.org/licenses/by-nc/4.0)

tant targets? This is probably relevant to Egr-1's co-localization with CpG islands (CGIs) *in vivo*, which has been shown in some cell types in several genome-wide ChIP-on-chip and ChIP-seq studies.[15,89,90] CGIs are regions of DNA with a high density of CpG dinucleotides (CpGs) present within 200-3000 bp. The human genome contains ∼25,000 CGIs, most of which are sites of transcription initiation.[91] In fact, the majority (∼ 70%) of human gene promoters are associated with CGIs, although CGIs represent only 0.8% of the human genome.[69] Perhaps surprisingly, CGIs are typically unmethylated for active genes, although CpGs are the substrates of major DNA methyltransferases and ∼85% of CpGs in the human genome are methylated.[71] Because the Egr-1 consensus sequence, GCGTGGGCG, contains two CpGs, it is reasonable to consider that functionally important Egr-1 sites within CGIs are unmethylated, whereas Egr-1 decoys outside CGIs are methylated.

Given this information, one may think that DNA methylation may diminish affinity of decoys, allowing Egr-1 to locate unmethylated targets within CGIs. CpG methylation moderately changes structural properties of DNA and thereby affects some transcription factors.[92] However, our recent biochemical study showed that the affinity of Egr-1 for its recognition sequence is unaffected by CpG methylation.[49] The crystal structures of Egr-1 complexes with unmethylated and methylated targets.[49,82,83,93] Thus, DNA methylation alone cannot drive Egr-1 to unmethylated recognition sequences within CGIs.

Based on these considerations, we formulated a hypothetical model to explain how Egr-1 locates unmethylated targets within CGIs. This model is

schematically depicted in Figure 3.1 and would work in the following manner.

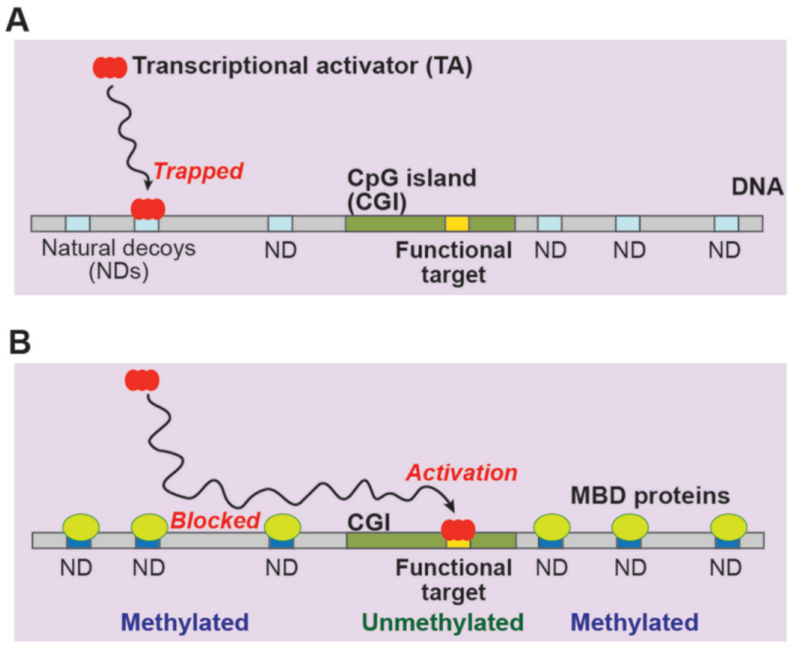

Figure 3.1: A) DNA methylation as a potential mechanism that facilitates the association of transcriptional activators with functional targets in CpG islands (CGIs) through competitive interplay with methyl-CpG-binding domain (MBD) proteins. The genome contains numerous non-functional high-affinity sites that are identical or similar to the target sequences of transcription activators. When these sites are accessible, they can serve as natural decoys that sequester transcriptional activators in non-functional locations. B) In vertebrate genomes, CpGs in CGI promoters of active genes are unmethylated, whereas ∼85% of CpGs outside CGIs are methylated. For activators that recognize sequence containing CpG (e.g. GCGTGGGCG for Egr-1: see also Figure 6) it is likely that their functional targets in CGIs are unmethylated, whereas decoys outside CGIs are methylated. DNA methylation outside CGIs may enhance the association of transcriptional activators with unmethylated targets in CGIs, because highly expressed MBD proteins block sequestration of the transcription activators in methylated decoys.

Because CGIs represent only 0.8% of the genome, the vast majority of natural

decoys for Egr-1 should be located outside CGIs and therefore methylated at CpGs. Due to this methylation, methyl-CpG-binding domain (MBD) proteins occupy these high-affinity non-functional sites, allowing Egr-1 to avoid being trapped there. In this manner, MBD proteins would indirectly guide Egr-1 to unmethylated targets within CGIs. Supporting this model, genome-wide analysis of bisulfite sequencing and ChIP-seq data show that DNA regions ±300 bp from Egr-1 ChIP-seq peak centers are unmethylated.[70] Our model differs from previous model for competition between MBD proteins and other transcription factors.[91,94] Our competitive model focuses on decoys outside CGIs, whereas the previous models focus on competition for functional sites within CGIs. Due to the large number of decoys in the genome, our model requires a high expression level of MBD proteins. The methyl-CpG-binding protein 2 (MeCP2) satisfies this requirement in some cell types: for instance, neuronal cells are estimated to contain as many as 10^7 MeCP2 molecules per nucleus.[95] In cell types where MeCP2 expression is lower, other MBD proteins, such Mbd1 and Mbd2, are often more abundant, according to the protein abundance database PaxDb (http://pax-db.org). Thus, it seems that the overall expression level of MBD proteins is considerably high in many cell types, satisfying the condition required for our model.

In this chapter, we present our biochemical study as the first step to assess this model. We used a simplified system involving the DNA-binding domains of Egr-1 and MeCP2 together with synthetic DNA duplexes. The Egr-1 DNA-binding domain comprised three zinc fingers is well suited for bio-

chemical and biophysical on the target search process.[21,41,45,47,48,50,96] Taking advantage of this system, we have performed kinetic studies on the Egr-1 zinc-finger protein in the presence and absence of the MeCP2 MBD, when it is challenged with methylated CpG and unmethylated CpG sites within Egr-1 decoys. Our data demonstrate that DNA methylation of decoys accelerates Egr-1 target association through competitive interplay with the MeCP2 MBD, supporting our hypothetical model.

3.1 Materials and Methods

3.1.1 Preparation of Egr-1 zinc finger proteins

The DNA-binding domain of the human Egr-1 protein (residues 335-423) comprised three zinc fingers was expressed in *Escherichia coli* and purified using affinity, size exclusion and cation exchange columns, as described in chapter 2. This construct was used in nuclear magnetic resonance (NMR),[?,48,50,96,97] X-ray,[49] and fluorescence studies.[21,41,45,98] For simplicity's sake, this construct is referred to as Egr-1 hereafter. Egr-1 was quantified using the Pierce bicinchoninic acid (BCA) protein assay kit (Thermo Fisher).

3.1.2 Preparations of MeCP2 MBD

A synthetic gene encoding the MBD of the human MeCP2 (residues 77-167) was sub-cloned into the Xmal/XhoI sites of the pET-49b vector (Novagen). *Escherichia coli* strain BL21(DE3) transformed with this plasmid was cultured at 37°C in 4 L of M9 media containing kanamycin (30 μg/ml).

At $OD_{600} \approx 0.8$, protein expression was induced by 0.6 mM isopropyl β-D-thiogalactopyranoside, and the culture was continued at 37°C for an additional 2 hours. The *E. coli* cells were harvested and disrupted by sonication in a buffer containing 50 mM Tris-HCl (pH 7.5), 1 mM ethylenediaminetetraacetic acid (EDTA), 500 mM NaCl, 2 mM dithiothreitol, 5% glycerol, 1% Triton X-100 (Sigma-Aldrich) and a Roche protease inhibitor cocktail (one tablet per 50 ml). After centrifugation at 30,000 × g and 4°C for 20 min, the supernatant of the lysate was loaded on a glutathione S-transferase (GST)-Prep FF 16/10 column (GE Healthcare) equilibrated with 50 mM Tris-HCl (pH 7.5), 400 mM NaCl, and 10 mM glutathione. The fusion protein was cleaved with 100 unites of HRV-3C protease (GenWayBiotech). After confirming the cleavage by polyacrylamide gel electrophoresis, the reaction mixture was concentrated to ∼10 ml with an Amicon Ultra-15 device, and loaded onto a Sephacryl S100 size exclusion column (GE-Healthcare) equilibrated with 50 mM Tris-HCl (pH 7.5), 1 M NaCL, 2 mM EDTA for separation. Fractions containing the MeCP2 MBD proteins were pooled, and the buffer was exchanged to 50 mM Tris-HCl (pH 7.0), 200 mM NaCl and 5% glycerol and then eluted with a gradient of 200-500 mM NaCl. The MeCP2 MBD was quantified by UV absorbance at 280 nm based upon an extinction coefficient of 11,460 $M^{-1}cm^{-1}$ (http://web.expasy.org/protparam/).

3.1.3 Preparation of DNA

Chemically synthesized DNA strands were purchased from Integrated DNA Technologies, Inc. Each strand was purified with a Mono-Q anion-exchange column installed on an AKTA Purifier system (GE Healthcare). The sequences of the DNA duplexes used in the current study are shown in Figure 3.2A. The sequence of the decoy DNA duplexes (L_S and mL_S) was chosen based in chapter 2. This 28-bp sequence contains a quasi-specific site matching 6 bp out of the 9 bp of the Egr-1 recognition sequence and exhibits ~400-fold higher affinity than completely non-specific 28-bp DNA L.[41] The duplexes L_S and mL_S are identical in sequence, but differ in that the CpG sequence in the quasi-specific site is methylated in mL_S. All 5-methylcytosine for CpG methylation were introduced when DNA was chemically synthesized. A fluorescein amidite (FAM) labeled 143-bp DNA duplex containing an Egr-1 target sequence was generated by polymerase chain reaction and purified as described previously.[45] This FAM-labeled 143-bp probe DNA was used in our previous stopped-flow kinetic studies.[21,41,45,50] All other DNA duplexes were prepared through annealing of complementary strands and removal of excess single-stranded DNA as described previously.[45]

3.1.4 Stopped-flow fluorescence experiments

The target search kinetics of Egr-1 were measured at 20°C using as Applied Photophysics SX20-LED stopped flow spectrofluorometer. In these experiments, the following two solutions were rapidly mixed in 1:1 volume

(∼0.5 ml) ratio by the stopped flow device: a solution of the Egr-1 zinc finger protein, and a DNA solution of FAM-labeled probe DNA (143-bp) and competitor DNA, with the addition of the MeCP2 MDB in some experiments. Both solutions were in a buffer containing 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 fluorescence intensity were collected for a period of 4-50s with time intervals ranging from 0.02 to 0.05 s. A light-emitting diode with maximum intensity at 470 nm was used for excitation of the FAM fluorophore. The emission light that passed through a long-pass filter with a cutoff at 515 nm was recorded. This configuration with no monochromator involved increases sensitivity in fluorescence detection. However, compared to observations with monochromators, it also increases sensitivity in fluorescence detection. However, compared to observation with monochromators, it also increases non-fluorescent background and reduces the percentage change in the FAM fluorescence intensity upon the target association. The total concentration of the competitor 28 bp duplexes was kept constant at 2 μM, though the concentration of the quasi-specific duplexes as Egr-1 decoys was either 0 to 150 nM. The concentration of the probe DNA 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 to simplify the kinetic analysis,[99] all binding reactions were conducted under conditions of $D_{\mathrm{tot}} >> P_{\mathrm{tot}} >> C_{\mathrm{tot}}$. The apparent pseudo-first-order kinetic rate constant for target association was determined from the time course of fluorescence intensity. *I(t)*, by non-linear least-squares

$I(t) = (I_0 - I_\infty)exp(-k_{app}t) + I_\infty$, where I_0 and I_∞ represent the intensities at time zero and infinite time, respectively. Rate constants k_{app} were measured at various concentrations for Egr-1. For each kinetic rate constant, the measurement was replicated 8-10 times. MATLAB software (MathWorks) was used for non-linear least-squares fitting.

3.1.5 Competition assays for methylated and non-methylated DNA duplexes

Affinities of the methylated and unmethylated 28-bp DNA duplexes (the sequence shown in Figure 3.2A) to Egr-1 and the MeCP2 MBD were measured using fluorescence based competitive binding assays with an ISS PC1 spectrofluorometer. To measure the affinities of Egr-1, the 143-bp FAM labeled DNA probe was used with one of the 28-bp DNA duplexes as competitor DNA. For measuring the affinities of the MeCP2 MBD, a FAM-labeled 19-bp DNA duplex, FAM-CTGGAACGGAATTCTTCTA in which the underlined CpG is methylated, was used as a probe. Fam fluorescence anistotropy was measured as a function of the concentration of the competitor DNA. Excitation and emission wavelengths were 490 and 521 nm, respectively. The concentrations of the fluorescent probe and the protein were 2.5 and 50 nM in the assays for the Egr-1 zinc-finger protein and 10 and 100 nM in those for the MeCP2 MBD, respectively. FAM fluorescence anisotropy was also measured in the absence of protein. Each anisotropy measurement was conduced for a 2 mL solution in a buffer containing 10 mM Tris-HCl (pH 7.5), 0.2 μM $ZnCl_2$ and 150 mM KCl. For each competitor DNA, the dissociation constants k_d were

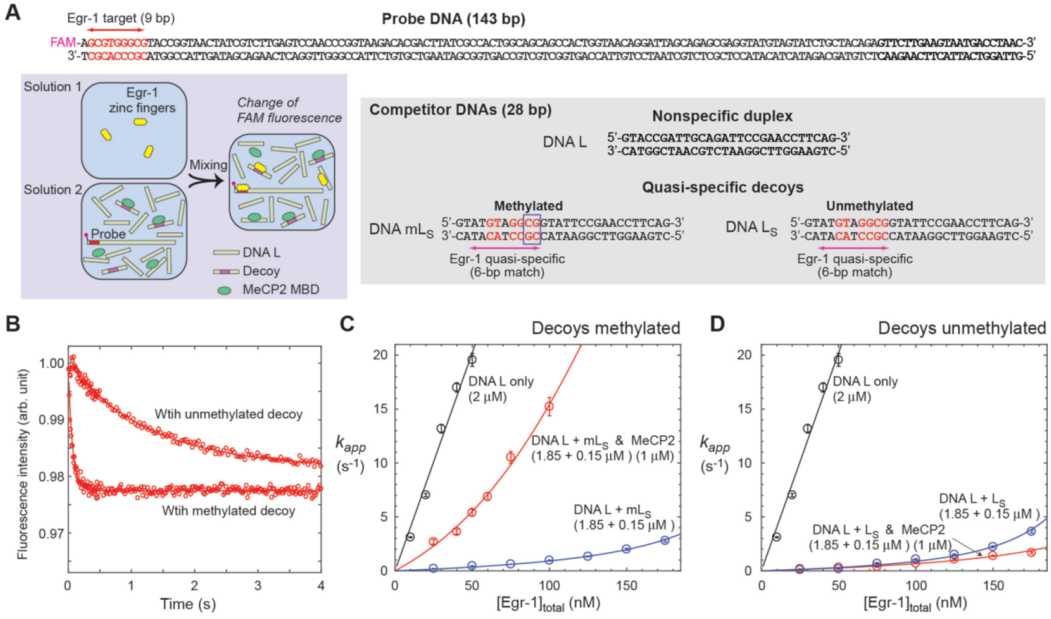

Figure 3.2: Stopped-flow kinetic assays for investigating how DNA methylation of decoys affects the target search kinetics for Egr-1 in the presence and absence of the MeCP2 MBD. (A) Macromolecular components mixed in the stopped-flow experiment. The FAM-labeled probe DNA (143 bp), non-specific DNA L (28 bp) and quasi-specific decoy DNA L_S were used in chapter 2. The decoy DNA mL_S is identical to DNA L_S in sequence, but is methylated at the quasi-specific site. (B) The time-course data of FAM fluorescence intensity recorded for the binding reactions using 2.5 nM probe DNA, 100 nM Egr-1, 1850 nM non-specific DNA L, 150 nM unmethylated (DNA L_S) or methylated (DNA mL_S) decoy and 1,000 nM MeCP2 MBD. (C and D) Apparent pseudo-first-order kinetic rate constants for the target association of Egr-1 in the presence of competitor DNAs. The Egr-1 quasi-specific sites in the decoys were methylated in panel C and unmethylated in panel D. The concentration of the probe DNA was 2.5 nM. The buffer was 10 mM Tris-HCl (pH 7.5), 150 mM KCl and 0.2 μ $ZnCl_2$ for all experiments. The presence of the MeCP2 MBD substantially accelerated the target search kinetics of Egr-1 only when the decoys were methylated.

determined from the anisotropy data by non-linear least squares fitting using MATLAB software as described.[100]

3.2 Results

In chapter 2, using a stopped-flow fluorescence method, we demonstrated that the presence of quasi-specific sequences, which are similar to the target sequence, impedes target DNA association of the Egr-1 zinc-finger protein. In this chapter, using a similar system together with the MeCP2 MBD, and methylated DNA, we examine the above-mentioned hypothetical model on the association of Egr-1 with its unmethylated targets in the presence of decoys (Figure 3.1). The components we used in these stopped-flow experiments are shown in Fig 3.2A. To measure the Egr-1-target association kinetics, we collected time course data for FAM fluorescence intensity upon mixing a solution of the Egr-1 zinc-finger protein with a solution of FAM-labeled probe DNA, non-specific DNA, decoy DNA and in some experiments, the MeCP2 MBD. Typical examples of the time-course fluorescence data are shown in Fig 3.2B. As observed in chapter 2, the association of the Egr-1 zinc-finger protein with the target on the probes caused a decrease in fluorescence intensity. As shown in Fig 3.3 the MeCP2 MBD does not interact with Egr-1 directly. However, indirect interplay between these proteins through competition for methylated decoys enhances the apparent activity of Egr-1 for unmethylated targets as described below.

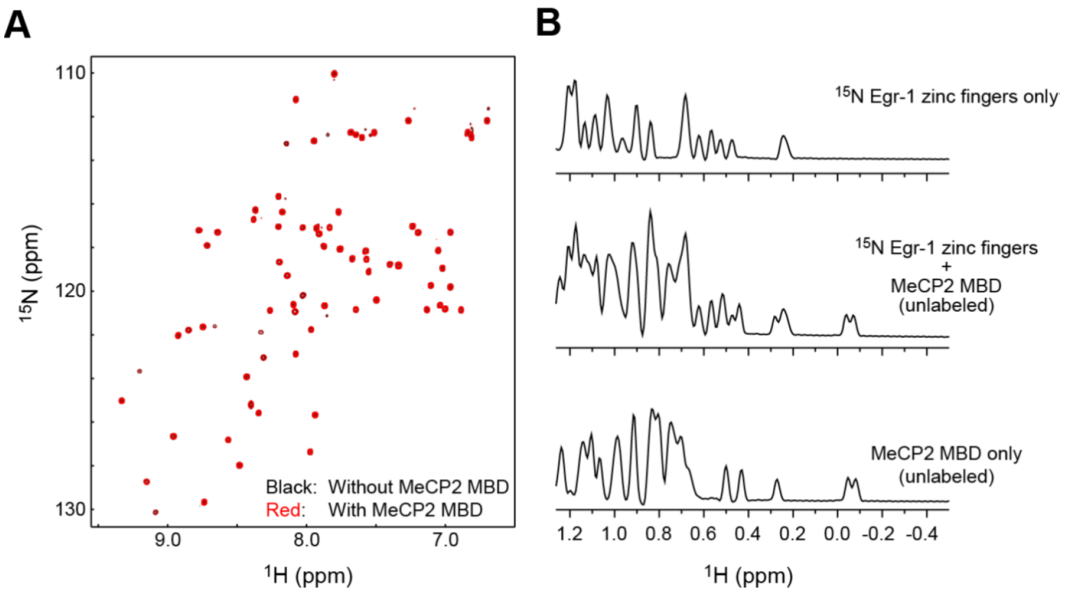

Figure 3.3: NMR data suggesting that the Egr-1 does not interact with the MeCP2 MBD. (A) Overlaid 2D ^{1}H- ^{15}N HSQC spectra recorded for the 0.2 mM ^{15}N-labeled Egr-1(black) and for a mixture of 0.2 mM ^{15}N labeled Egr-1 and 0.2 mM unlabeled MeCP2 MBD (red). (B) 1D ^{1}H spectra (methyl regions) recorded for 0.2 mM ^{15}N-labeled Egr-1 DBD (top), for a mixture of 0.2 mM ^{15}N-labeled Egr-1 and 0.2 mM unlabeled MeCP2 MBD (middle), and for 0.2 mM MeCP2 MBD (bottom). These spectra were recorded at 25°C and a ^{1}H frequency of 750 MHz. The buffer was 10 mM Tris-HCl (pH 7.5), 150 mM KCl, 0.2μM $ZnCl_2$, and 5% D_2O. The data in panel B clearly indicate the presence of Egr-1 and MeCP2 MBD proteins in the mixture. If these two proteins interacted, the ^{1}H-^{15}N spectrum of this mixture should differ from that of the free Egr-1. However, the ^{1}H-^{15}N HSQC spectrum of the ^{15}N-labeled Egr-1 in this mixture did not exhibit any perturbations due to the presence of the MeCP2 MBD. These results suggest that these two proteins do not interact.

3.2.1 Unmethylated and methylated quasi-specific impede target search of Egr-1

We first examined whether methylation of the quasi-specific decoy DNA affects the target search by the Egr-1 zinc-finger protein in the absence of any other proteins. Fig 3.2C and D shows the results of kinetic measurements using methylated and unmethylated quasi-specific DNA duplexes, respectively. The data points shown in blue are the results from the experiments with 2.5 nM probe DNA, 1850 nM non-specific 28-bp DNA, 150 nM 28-bp decoy DNA containing a quasi-specific sequence and various concentrations of the Egr-1 zinc finger protein. Note that the concentration of the decoy was set to be 60-fold greater than that of the target. To clarify the impact of the decoys, these figures also show the results from the experiments with 2000 nM non-specific 28-bp DNA and no decoy(shown in black). For both methylated and unmethylated quasi-specific sites, the presence of decoy DNA was found to substantially impede the target association of the Egr-1 zinc-finger protein. this effect occurred to a similar degree regardless of the decoy DNA's methylation state.

These results suggest that DNA methylation itself does not diminish the ability of a quasi-specific site to trap the Egr-1 zinc-finger protein. We verified this by measuring the relative affinities of the unmethylated and methylated quasi-specific DNA duplexes using competitive binding assays (Figure 3.4A) In this assay, anisotropy of FAM fluorescence from the 143-bp probe DNA was measured to determine the equilibrium population of the protein-bound tar-

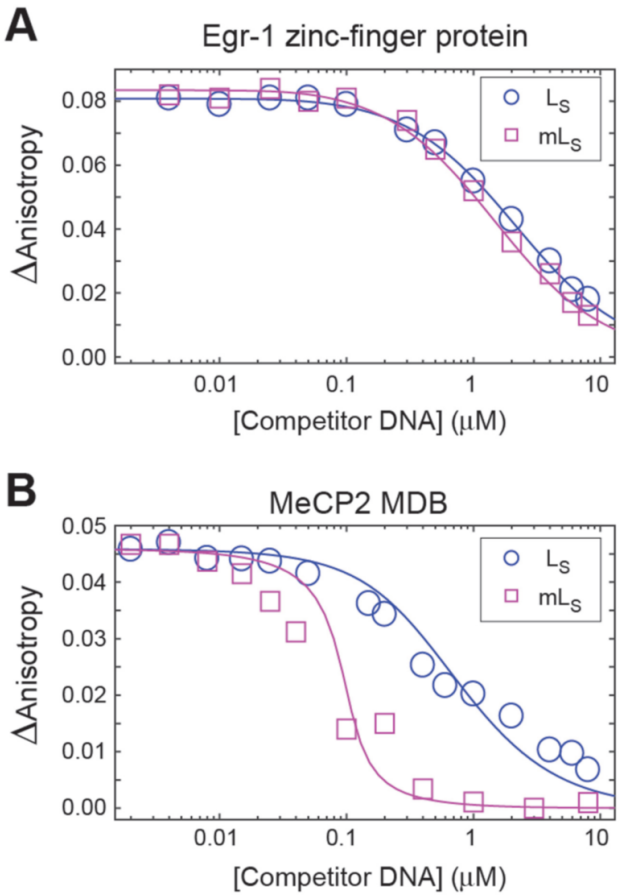

Figure 3.4: Equilibrium competitive binding assays to investigate the impact of CpG methylation on the affinity of the decoy DNA for the Egr-1 zinc-finger protein (A) and the MeCP2 MBD (B). In panel A, FAM fluorescence anisotropy was measured for solutions containing 2.5 nM FAM-labeled 143-bp DNA containing the Egr-1 target (see Figure 3.2A), 50 nM Egr-1 zinc-finger protein and various concentrations of DNA L_S or mL_S. In panel B, fluorescence anisotropy for a FAM-labeled 19-bp DNA containing a methyl-CpG was measured for solutions containing 10 nM probe, 100 nM MeCP2 MBD and various concentrations of DNA L_S or mL_S. The vertical axes show the difference between fluorescence anisotropy values for each sample and free probe DNA solutions.

get at various concentrations of the competitor DNA. Fluorescence anisotropy is related to the fluorophore's effective rotational correlation time, which can change upon molecular association.[101] The FAM fluorescence anisotropy was measured to be 0.077 for the free probe and 0.161 for the protein-bound probe. The anisotropy measurements were conducted at various concentrations of the 28-bp competitor DNA duplexes with (mL_S) and without (L_S) the CpG methylation of the quasi-specific sites (magenta and blue data points, respectively, in Figure 3.4 A). The anisotropy data showed that at relatively low concentrations of the competitor, the target on the fluorescence probe was predominantly in the protein-bound state. However, an increase in competitor concentration caused a decrease in the population of the protein-bound target due to transfer of the protein to the competitor. The concentration dependence data were very similar for methylated and unmethylated competitors, indicating that CpG methylation does not diminish the Egr-1's affinity for quasi-specific sites.

3.2.2 MeCP2 accelerates target search of Egr-1 when quasi-specific DNA is methylated

To test our hypothetical model (Figure 3.1), we examined whether the MeCP2 MBD would influence the association of the Egr-1 with unmethylated target when decoys were methylated. For this purpose, we conducted the same stopped-flow kinetic assays in the presence of the MeCP2 MBD in solution with the 143-bp DNA probe and methylated or unmethylated competitor DNA. Data points shown in red in Figure 3.2C and D show apparent pseudo-first-

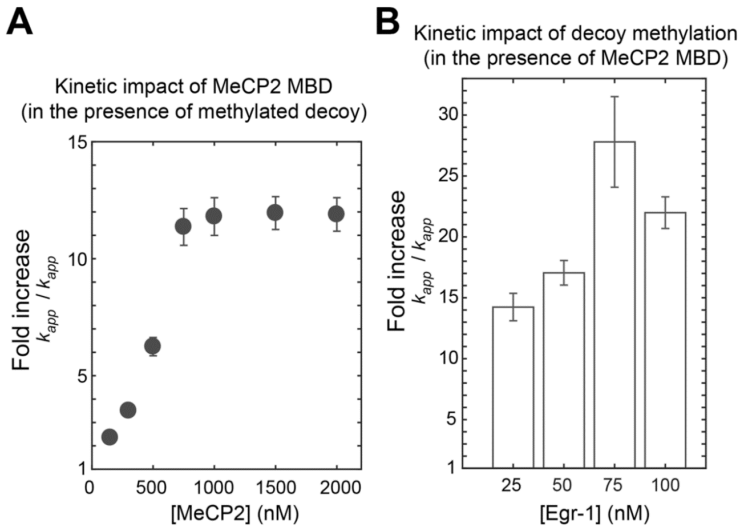

Figure 3.5: Acceleration of target association of the Egr-1 by the MeCP2 MBD and decoy methylation. (A) Dependence on the concentration of the MeCP2 MBD when the Egr-1 decoys were methylated. (B) Impact of the decoy DNA methylation when 1 μM MeCP2 MBD was present. Fold increases in the apparent pseudo-first-order rate constant k_{app} for the target association are plotted in each panel.

order rate constants measured for Egr-1 target association in the presence of 1 μM MeCP2 MBD.

When the MeCP2 MBD was present and the decoy was methylated in the system, the Egr-1 exhibited substantially faster association with the unmethylated target on the probe DNA (Figure 3.2C). At 1 μM MeCP2 MBD, there was a 12-fold increase in the rate constant for target association for Egr-1, compared to that in the absence of the MeCP2 MBD. As shown in Figure 3.5A we found that this acceleration was dependent on the concentration of the MeCP2 MBD but became saturated at concentration far higher than the

decoy concentration. These results strongly suggest that the MeCP2 MBD accelerates the association of the Egr-1 with unmethylated target by blocking the methylated decoys as predicted by our hypothetical model (Figure 3.1).

In contrast, the acceleration of Egr-1's target association by the MeCP2 MBD was not observed when the decoy DNA was unmethylated (Figure 3.2D). In fact, the presence of MeCP2 MBD slightly slowed down the target association of Egr-1 in the experiments with the unmethylated decoy. The slower target search of Egr-1 in this situation may be due to additional binding of the MeCP2 MBD to the probe DNA. To illustrate the impact of decoy DNA methylation on Egr-1, Figure 3.5B shows fold increases in target association rate constants upon the decoy methylation. In the presence of 1 μM MeCP2 MBD, the decoy methylation resulted in 14- to 28- fold acceleration of the target association of Egr-1.

Although unmethylated and methylated decoys could almost equally impede the target search of Egr-1, the MeCP2 MBD markedly accelerated Egr-1 target association only when the decoys were methylated. This is likely due to specific binding of the MeCP2 MBD to methyl-CpGs at Egr-1 quasi-specific sites. We verified this by conducting equilibrium competitive binding assays and comparing the relative affinities of the MeCP2 MBD for these 28-bp quasi-specific DNA duplexes (Figure 3.4B). The affinity of the MeCP2 MBD for the methylated DNA (the dissociation constant K_d = 2.3 nM) was 54-fold higher than that for the unmethylated DNA (K_d = 124 nM), which is consistent with a previous study.[102]

3.2.3 No acceleration by the MeCP2 MBD when the target is also methylated

If blocking methylated decoys by the MeCP2 MBD directly causes acceleration of Egr-1 target association as shown in Figure 3.2, it is expected that such acceleration will not occur when the target sequence on the probe DNA is also methylated. This is because the MeCP2 DBD should block the methylated target sequence on the probe DNA as well. We examined whether this is indeed the case by conducting the same experiment using another probe DNA (33 bp) in which the target sequence is methylated (Figure 3.6) Although

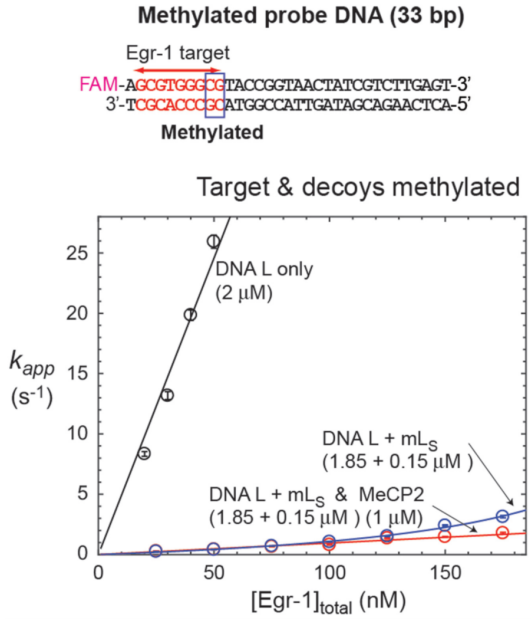

Figure 3.6: The MeCP2 MBD does not accelerate target association of the Egr-1 zinc-finger protein when the target is also methylated. The graph shows the apparent kinetic rate constants measured for the target association for the Egr-1 zinc-finger protein using the same conditions as those in Figure 3.2C except that the Egr-1 target in the 33-bp probe DNA was methylated.

CpG methylation of the same 143-bp DNA would be preferable, we used this DNA because site specific incorporation of methyl-CpG was difficult to achieve with the FAM-labeled 143-bp probe DNA. Using the methylated 33-bp probe DNA together with the methylated decoy DNA mL_S, we measured kinetic rate constants for target association of the Egr-1 zinc-finger protein in the presence and absence of the MeCP2 MBD. As expected, the MeCP2 MBD did not accelerate Egr-1 target association in this case (compare red and blue data points in Figure 3.6) but rather slowed it. In contrast, when the same 33-bp probe was unmethylated, the MeCP2 MBD significantly accelerated Egr-1 target association (Figure 3.7) , as seen for the 143-bp probe DNA containing an unmethylated target. These results suggest that when Egr-1 targets are methylated, the MeCP2 MBD binds to them and thereby blocks Egr-1's association with the methylated targets.

3.3 Discussion

In vertebrates, DNA methylation is an important epigenetic mechanism that is associated with various biological processes such as development, genomic imprinting, and X-chromosome inactivation.[103] Abnormalities in DNA methylation status are associated with cancer and many other diseases.[71,104] While the regulatory mechanisms of DNA methylation and demethylation remain under active investigation,[70,105] it has been well established that DNA methylation within CGIs is directly associated with gene silencing[69,103,106] . Based on our biochemical data, we argue that DNA methylation outside CGIs

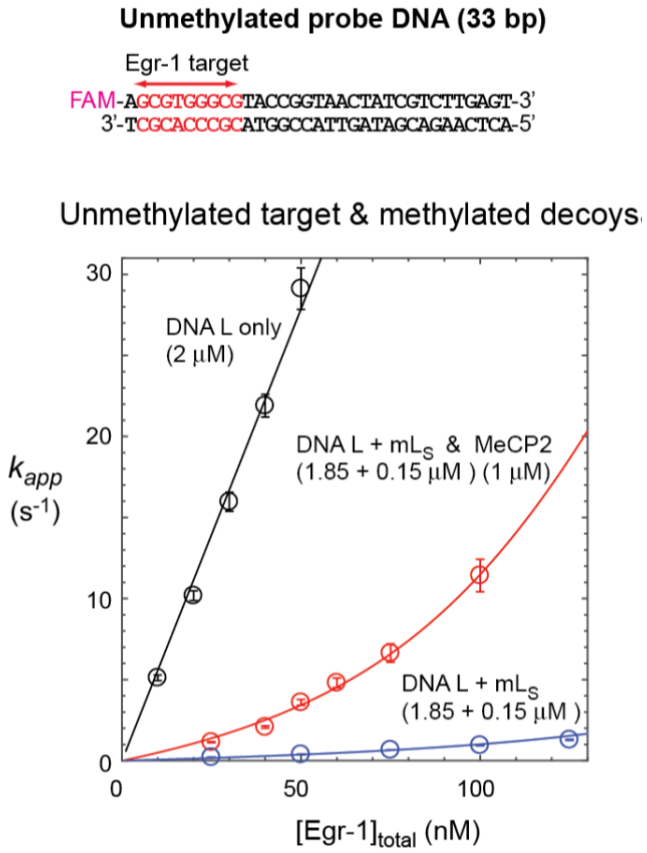

Figure 3.7: Kinetic data of Egr-1 target association measured using the FAM-labeled 33 bp probe DNA containing an unmethylated target. The graph shows the apparent kinetic rate constants measured for the target association of the Egr-1 zinc-finger protein using the same conditions as those in Figure 5 except that the Egr-1 target in this 33-bp probe DNA was unmethylated. In this case, the MeCP2 MBD substantially accelerated the target association of the Egr-1 zinc-finger protein in the presence of the methylated decoy DNA mL_S.

may directly enhance activities of transcriptional activators at CGIs through competitive interplay with MBD proteins.

3.3.1 DNA methylation as potential facilitator of transcription factor-CGI interactions

Our current biochemical data show that DNA methylation of decoys can greatly facilitate association of the Egr-1 zinc finger protein with its unmethylated target through competitive interplay with the MeCP2 MBD. In this manner, MBD proteins may indirectly guide Egr-1 to unmethylated targets within CGIs by blocking methylated decoys outside CGIs. We speculate that this model may be applicable for some other transcriptional activators that bind to CGI promoters. In fact, there are many transcriptional activators that recognize CpG-containing sequences (some examples are shown in Figure 3.8)[70] The binding of these transcriptional activators to their unmethylated targets in CGI promoters may be facilitated by highly expressed MBD proteins that block numerous methylated decoys outside CGIs, as shown in Figure 3.1.

3.3.2 Potential mechanism for transactivation by MeCP2

Our hypothetical model (Figure 3.1) provides insight into the transcriptional activation by MeCP2. With an MBD and a transcriptional repression domain, this protein was originally regarded as a methyl CpG-dependent transcription repressor.[107] However, despite the lack of any known transcriptional activation domains, MeCP2 activates more than 2,000 genes in neurons under the control of CGI promoters where CpG methylation is rare.[108] In fact, MeCP2 activates ~6 times more genes than it represses. Chahrour et al. pro-

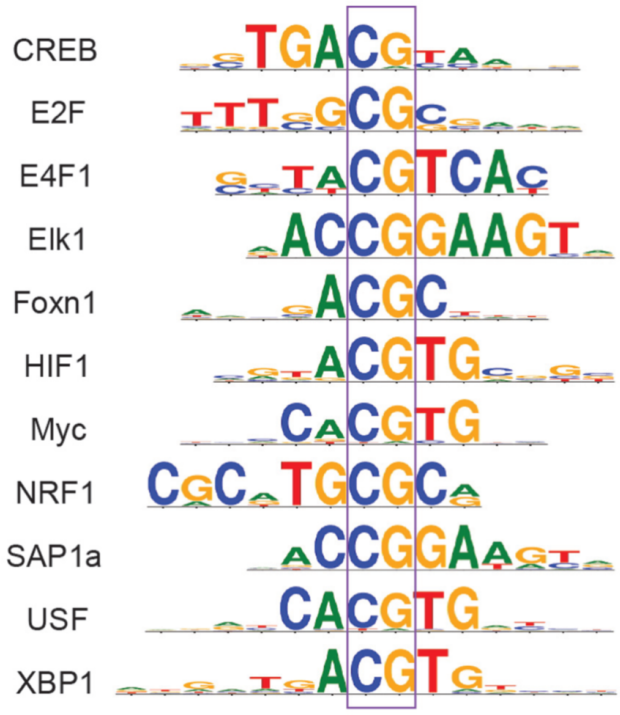

Figure 3.8: Some examples of human transcriptional activators that recognize CpG-containing sequences. Height of each letter represents likelihood of base type at each position within the recognition sequence. Adapted from the MotifMap database (http://motifmap.ics.uci.edu)

posed that protein-protein interactions with the transcription factor CREB1 could be responsible for transactivation by MeCP2.[108] While this may explain the activation of some genes, it remains unclear why MeCP2 can activate so many genes. Our model (Figure 3.1) may provide an explanation. Because CpGs are highly methylated outside CGIs, highly expressed MeCP2 can effectively block non-functional high affinity sites for transcriptional activators that recognize CpG-containing sequences. This will indirectly guide transcriptional activators to unmethylated targets within CpGs and thereby activate

downstream genes. Our model does not require any direct protein-protein interactions with particular transcriptional activators and could explain why MeCP2 activates so many genes. This model is directly applicable to CREB1 as well because the CREB1 recognition sequence, TGACGTCA, contains a CpG. Our model is also consistent with the observation that the promoters of genes activated by MeCP2 are not methylated.[108]

3.3.3 Relevance to 'DNA methylation paradox'

Our model could also explain the so called 'DNA methylation paradox',[106,109] which concerns DNA methylation yielding opposite effects depending on genomic contexts. In vertebrates, although DNA methylation within CGI promoters is typically associated with gene silencing, DNA methylation in the gene bodies is positively correlated with gene expression level.[106,109] In our model (Figure 3.1), a high level of DNA methylation outside CGIs should increase CGI-binding of transcriptional activators that recognize CpG-containing sequences. Because the MBD proteins block methylated decoys oustide CGIs, the transcriptional activators can occupy their functional sites within CGI promoters more easily, thereby increasing gene expression level. On the other hand, CpG methylation within CGIs should cause binding of MBD proteins to CGI promoters and reduce the expression of downstream genes. This repression could occur through excluding transcriptional activators from CGI promoters and recruiting co-repressor proteins via the transcriptional repressor domain of MBD proteins.

However, it should be noted that our model is not applicable for transcriptional activators such as Klf4 and CEBPB, which preferentially bind to CpG-methylated DNA *in vitro*.[70] Genome-wide ChIP-seq and bisulfite sequencing studies for Klf4 and CEBPB show that ∼20-40% of the genomic regions bound by these proteins are methylated *in vivo*.[110,111] Protein-protein interactions with other protein might increase occupancies of Klf4 and CEBPB in these regions. It remains to be addressed why these proteins can occupy the CpG-methylated regions despite high abundance of MBD proteins in the cells.

3.3.4 New perspective on target DNA search

Our current work provides a new perspective on the target DNA search mechanisms for transcription factors. Over the past four decades, the mechanisms allowing transcription factors to rapidly locate their specific targets in the genome have been the subject of considerable interest in biophysics and biochemistry.[10,63–65] While recent studies using NMR and single-molecule techniques revealed great details on how proteins scan DNA efficiently,[28,36–38] studies that focus on factors that impede the search process have been rare.[39] Numerous non-functional high-affinity sites in the genome could serve as decoys that trap transcription factors and affect their functions.[40–42,68,112–114] Our study shows that the target search process for transcription factors can be greatly accelerated when other proteins block such decoys. This mechanism does not require any direct protein-protein interactions, although the current

paradigm for synergy between transcription factors typically assumes their direct interactions.[76,77] In our current case, the MeCP2 MBD facilitates the association of the Egr-1 zinc-finger protein with the unmethylated target by blocking methylated decoys. In principle, other proteins could also provide the same acceleration mechanism, as long as they selectively block decoys but not targets. For example, highly expressed transcription factors with similar but different sequence specificity may enhance the function of other transcription factors by selectively blocking their decoys. Natural decoys and their DNA methylation may play key roles in regulation of transcription factors.

Chapter 4

Investigating conformational dynamics of Egr-1 arginine side chains during DNA scanning and sequestration by quasi-specific sites.

DNA in the nucleus exhibits structural similarity and exists in the nucleus in high concentrations.[3] It is likely that this would affect protein scanning in some way. Previous work has shown how nonspecific DNA plays an important role in protein scanning;[50] we also demonstrated that quasi-specific DNA decelerates this process, effectively trapping protein on off-target sequences, as shown in chapter 2. This trapping would be most likely stronger *in vivo* due to the high density of DNA in the nucleus. Yet, in spite of the trapping effect from off-target locations, proteins still efficiently locate and bind to their target site[10].[28,36,39,63,64,79,115] While this may be due to a change in the protein's conformational dynamics during the scanning process,[48,50] the specific interactions during a transcription factor's (TF) scanning process are largely unknown.

Transcription factors recognize DNA sequences using atomic interactions within individual amino acids to bind to DNA nucleotides. TFs exhibit a number of different interactions including Van der Waals, water-mediated

interactions, and hydrogen bonding when binding to DNA.[44,73,74,74,75] Hydrogen bonding is often used to stabilize TFs on DNA, with many hydrogen bonds forming between guanine and either polar or charged residues of the protein.[73,116] The side chains of charged residues, such as lysine and arginine, make up a large portion of the electrostatic interactions between the protein and DNA. Not only do these residues interact with nucleotide bases, they also interact with the DNA phosphate backbones by forming ion pairs and releasing condensed counterions from DNA.[73–75] This release of counterions has been shown to be important for protein-DNA association.[59–61] When a protein scans DNA, it must continuously change its interactions with the DNA bases to translocate on the DNA strand. The number and strength of interactions between the protein and DNA are thought to be relevant to a protein's sliding efficiency.[48,50,62] Egr-1 has a total of 15 arginine residues (Figure 4.1A) 5 of which recognize a guanine base in the 9 bp target sequence of Egr-1 (GCGTGGGCG). Two residues R24 and R80 solely interact with guanines (G8 and G2 respectively), while residue R18, R46, and R74 interact with both guanine (G10, G7, and G4 respectively) and aspartic acid (D20, D48, and D76 respectively) (Figure 4.1B and C).[21,44,45,50,117] Interacting with just the guanine bases forms hydrogen bonds with only the N_{η}-$H_{\eta 2}$, while the guanine/aspartic acid combination binds both the N_{η}-$H_{\eta 2}$ moieties and the N_{ϵ}-H_{ϵ} moieties for an additional 2 hydrogen bonds.[49,82,93] These interactions of the arginine residues make up over half of Egr-1's binding sequence, and therefore loss of these interactions cause a significant increase in entropy and

dynamics of the arginine sidechains, and subsequently a decrease in affinity for DNA. Since these interactions during protein scanning would be entropically unfavorable, higher affinity sequences display a trapping effect on a protein during target search, as shown in chapter 2. In addition, it has been shown that a conformational switch can allow for improving efficiency during sliding, although it is unclear what kinds of bonds form during the sliding process.[62] Much research has focused on the conformational dynamics of the NH backbone and the CH_3 side chain groups of amino acids residues,[118–120] but there is little knowledge on the atomic detail of the side chain dynamics and their role in overall protein DNA interactions and scanning. Previous research on the Egr-1 zinc finger domains has shown that there are significant differences in side chain dynamics of arginine and lysine when Egr-1 is bound to its target compared to when it is free. While many of the arginine side chains become highly immobile upon hydrogen binding with DNA bases, lysine side chains remain highly mobile even when bound.[62] This data shows that interactions for DNA binding are not necessarily all rigid interfaces, but contain a combination of both dynamic and rigid interfaces. In this chapter, we address the conformational dynamics of Egr-1's arginine side chain's guanidino N_ϵ-H_ϵ moieties when Egr-1 is bound to completely nonspecific and quasi-specific DNA (discussed in chapter 2). Using nuclear magnetic spectrometry (NMR), we compare Egr-1 in complex with both nonspecific and quasi-specific DNA. The resulting data gives a more detailed look at the molecular interactions and dynamics during the scanning process. It also allows us to compare the

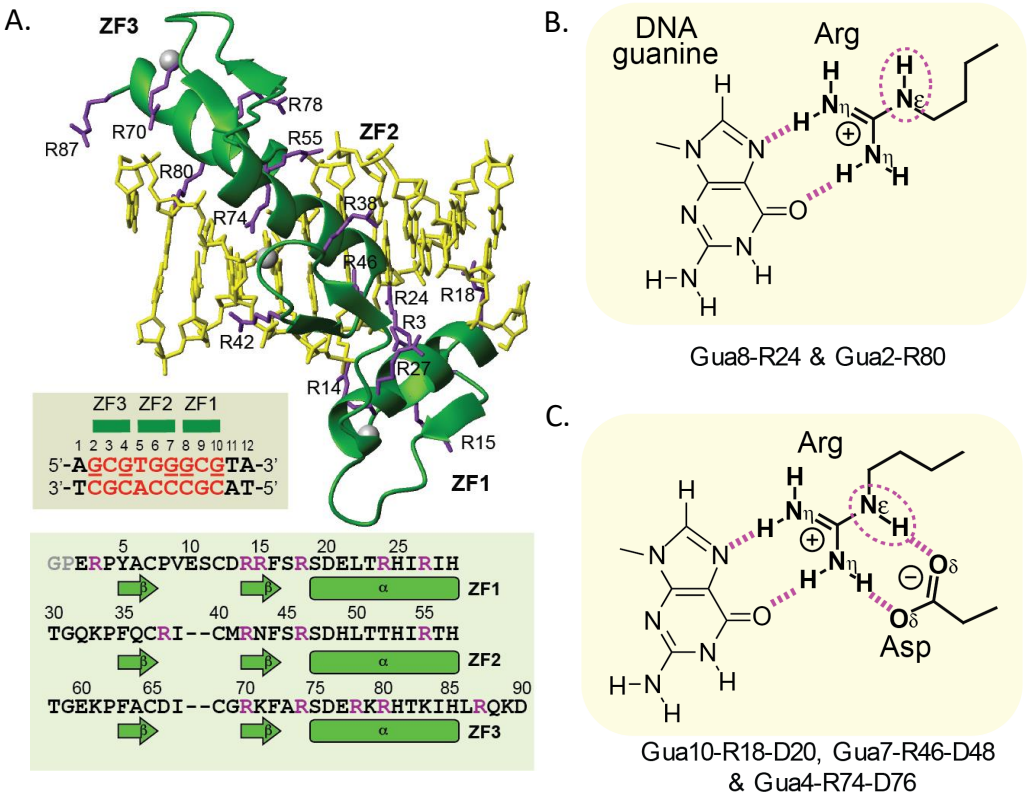

Figure 4.1: A)A total of 15 arginine residues are found within Egr-1 structure. B) Five of the arginine side chains recognize 5 of the 9 base pairs within the specific sequence for Egr-1. Two of these arginines form two hydrogen bonds with guanine bases, C) while the other three form both hydrogen bonds with guanine bases and neighboring aspartic acid residues. This additional bonding further immobilizes the arginine residue and provides more stability to the complex.(Figure adapted from Esadze, Alexandre, et al. "Changes in conformational dynamics of basic side chains upon proteinDNA association." Nucleic acids research 44.14 (2016): 6961-6970.)

dynamic properties and interactions of the bound protein in its recognition mode[62] versus the bound protein in its scanning mode.

4.1 Materials and Methods

4.1.1 Protein Preparation

The zinc finger DNA binding domain of Egr-1 was prepared and purified as previously described in chapter 2. ^{15}N and/or ^{13}C labeled protein was grown in minimal media using ^{15}N labeled ammonium and ^{13}C labeled glucose.

4.1.2 DNA preparation

Two different DNA constructs were used for NMR experiments. Both were 28 bp DNA, one completely nonspecific DNA (DNA L) and another quasi-specific DNA (DNA L_S) with a 6 bp match to Egr-1's target site. The DNA was prepared as previously described in chapter 2.

4.1.3 NMR sample preparation

NMR samples of Egr-1 and Egr-1/DNA complexes were made of 270 μL solutions with 0.4 mM Egr-1 and 0.8 mM DNA L, in a buffer containing 20 mM KCl, 20 mM succinate pH 5.8, and 0.1 mM $ZnCl_2$. With these conditions most of the protein should be bound to DNA. Free samples of Egr-1 (0.4 mM) were prepared using the same buffer. Samples were sealed in either Norell co-axial tubes or Shigemi co-axial tubes. Norell co-axial tubes contain an inner tube which allows for a separate compartment for D_2O which is used as an NMR lock. By separating the D_2O and the sample, it reduces line broadening and isotope shifts of ^{15}N resonances caused by exchanging deuterium and hydrogen protons. The Shigemi co-axial tubes also have an inner sleeve, but the sample

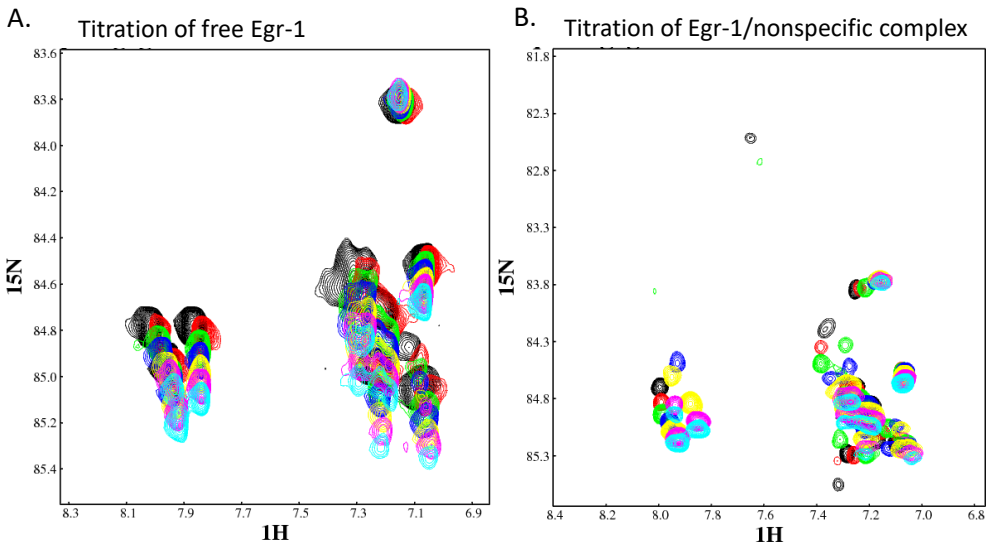

Figure 4.2: A)Salt titration of free Egr-1. Spectrums were recorded at 7 different salt concentrations. The color of the peaks corresponds to a spectrum at a different salt concentration as follows: black - 20 mM, red - 100 mM, green - 200 mM, blue - 300 mM, yellow - 450 mM, magenta - 600 mM, cyan - 750 mM. B)Salt titration of Egr-1 in complex with nonspecific DNA (DNA L). Color coding for each titration is the same as for the free Egr-1 sample.

is placed in the inner sleeve instead of the outer sleeve. The outer sleeve is used for the D_2O NMR lock.[121]

4.1.4 NMR experiments

All NMR experiments were performed on Bruker Avance III NMR spectrometers on a ^1H frequency of 600, 750, or 800 MHz. While all spectrometers were equipped with a cryogenic probe, due to probe failure many of the experiments performed on the 600 MHz were excepted. NMR data were

processed with the program NMR-Pipe[122] and visualized and analyzed with NMR-View.[123] For the Egr-1/DNA L_S complex, ^{1}H, ^{13}C, and ^{15}N resonances were assigned to the protein's residues using 3D HN(CO)CA, HNCA, HCACO, HNCO, HNCACB, CBCA(CO)NH spectra.[124] Arginine side chains $^{15}N\epsilon$-$^{1}H\epsilon$ resonances were assigned using 3D ^{15}N-edited NOESY spectra.[125] For the Egr-1/DNA L complex, the backbone resonance assignments were assigned based on data collected and published previously.[62] The previously published assignment was done in the same manner as the Egr-1/DNA L_S complex assignment. To assign arginine side $^{15}N_\epsilon$ and $^{1}H_\epsilon$ resonances for the Egr-1/DNAL complex, an Arg N_ϵ-H_ϵ selective ^{1}H-^{15}N heteronuclear in-phase single-quantum coherence (HISQC) of free Egr-1 at 25°C and of the Egr-1/DNA L complex at 35°C were performed at 7 different salt concentrations (20, 100, 200, 300, 450, 600, 750 mM KCl)(Figure 4.2) and a 1D-^{1}H spectra and arginine HISQC spectrum were run at each. At 750 mM KCl, the spectrum of protein in complex with DNA L matched the spectrum of free protein and the peaks could be back traced to 20 mM KCl and then assigned. To correct the baseline position of the ^{1}H frequency (6.676 ppm), sodium 2,2-dimethyl 2-silapentane-5-sulfonate (DSS) was used to calculate the shift of the spectrum from the baseline. This correction was then subtracted from the ^{1}H frequency and the data was processed with NMR-View using the corrected parameter. All other experiments were performed at 35°C.

4.1.5 Fitting calculation for determination of K_d of Egr-1 to non-specific DNA at difference salt concentrations

To calculate the dissociation constant for Egr-1 at different salt concentrations, the NMR salt titration data was fitted based on the counterion condensation theory developed by Manning et. al, which describes the attraction of counterions to polyelectrolytes such as DNA.[126–129] Counterions condense a polyelectrolyte surface to reduce axial charge density. According to Manning, this charge density should be less than 0.14 net charge per Å. A polyelectrolyte's initial charge density will change the condensation population. To determine if the counterion condensation occurs for a particular polyelectrolyte, Manning used the following equation:

$$\xi = \frac{l_B}{b} = \frac{e^2}{4\pi\epsilon_0 D k_B T b} \tag{4.1}$$

where l_B represents the solvent's Bjerrum length characteristic, b is the charge spacing on a polyelectrolyte's axis, e is the electronic charge magnitude, ϵ_0 is vacuum permittivity, D is the dielectric constant, k_B is the Bolztmann constant, and T is the temperature. If the value ξ is greater than one, the probability of counterions condensing on the polyelectrolyte surface is high. In our case, the ξ value for B-form DNA is calculated to be 4.2,[130] which indicates that there will be condensation of counterions around the DNA. Therefore, in order for a protein to interact with DNA, the counterions must be released to form intermolecular ion pairs. Releasing counterions is entropically favorable, therefore an analytical expression can be used to describe the entropic effect

of counterion release:[131]

$$logK_d = alog[M^+] + b \tag{4.2}$$

Where

$$b = logK_{d,1M} \tag{4.3}$$

and

$$a = z\varphi \tag{4.4}$$

in which $K_{d,1M}$ represents the dissociation constant of 1M of $[M^+]$ (monovalent cation concentration, in our case KCl), z represents the number of DNA/ligand interactions intermolecular ion pairs and φ is the number of counterions bound per DNA phosphate in terms of thermodynamics. In our salt titration experiment, we used the equation to fit the salt titration curves for each arginine residue, which are obtained from the ^{15}N chemical shift change from the peaks in each spectrum at the different salt concentrations (Figure 4.4A). The titration curves are fitted with eq. 4.2, and both the a and b values are calculated from the fitting. The a and b values can be plotted with the increasing KCl concentration to obtain the corresponding K_d for each salt concentration (Figure 4.4B).

4.1.6 Rotational diffusion parameter experiments

To determine rotational correlation time for the backbone of both complexes, two experiments for the backbone ^{15}N R_1 and R_2 relaxation rates were

done at 35°C at a ^{1}H frequency of 800 MHz. To determine S^2 order parameters of the arginine residues for the Egr-1/DNA L complex and Egr-1/DNA L_S complex, data for ^{15}N R_1 and heteronuclear NOE data were recorded at a ^{1}H frequency of 600 and 750 MHz and ^{15}N R_2 was recorded at 750 MHz. The pulse sequences for NH groups of the Arg N_ϵ-H_ϵ moieties were INEPT schemes which included ^{15}N rSNOB 180° pulses,[132] while the ^{15}N carrier position was set to 81 ppm to allow for visualization of the Arginine side chain residues.

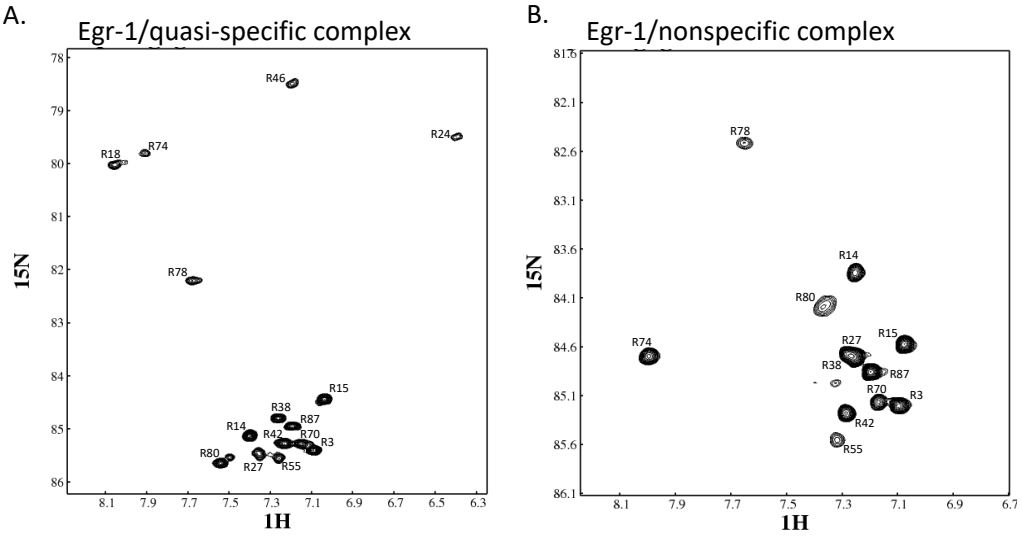

Figure 4.3: A)Arg Nϵ - Hϵ selective ^{1}H- ^{15}N heteronuclear in-phase single quantum coherence (HISQC) of Egr-1 in complex with quasi-specific DNA B)Arg Nϵ - Hϵ selective ^{1}H- ^{15}N HISQC for Egr-1 in commplex with nonspecific DNA.

4.1.7 ^{15}N relaxation data analysis

After recording R_1, R_2 and NOE data from 600 MHz and 750 MHz, a C program with a GNU scientific library[133] was used to determine the rotational diffusion parameters of the axial symmetric diffusion model ($D_\parallel$ and $D_\perp$ and 2 polar angles for the main principal axis) from the backbone relaxation data. The rotational correlation time $\tau_{r,eff}$ and the rotational diffusional anisotropy r are calculated by $(2D_\parallel + 4D_\perp)^{-1}$ and $D_\parallel / D_\perp$, respectively.[134] The zinc finger domains tumble independently when Egr-1 is free as shown in other studies;[48,50] however, only the first zinc finger (ZF) domain tumbles independently when the protein is bound to nonspecific DNA. For the Egr-1 DNA L complex, $\tau_{r,eff}$ was calculated separately for ZF domain 1 and ZF domains 2 and 3. From the relaxation data obtained at ^{1}H frequencies of 600 and 750 MHz, MATLAB software was used to calculate the order parameters for Arg N_ϵ H_ϵ groups. Two constants ^{15}N chemical shift anisotropy ($\sigma_\parallel - \sigma_\perp$) and N_ϵ-H_ϵ distance for arginine side chains were set at -114 ppm and 1.04 Å. This was based on previous work by Trbovic et al.[135] We used four different spectral density functions for each Arg N_ϵ-H_ϵ group and selected the best fit based on Akaike's information criterion.[136] The first two spectral density functions were model-free functions developed by Lipari and Szabo,[137] and the second two were extended model-free functions from Clore et al.[138]

4.2 Results

In this chapter, we observe the dynamics of Arg side chains when the protein is bound to nonspecific DNA L or quasi-specific DNA L_S and compare them to previous data of the dynamics of Arg side chains when Egr-1 is free or bound to its target site. The arginine residues are labeled according to a numbering scheme previously given by Pabo et al.[82,83]

4.2.1 NMR spectra of Egr-1 in complex with nonspecific and quasi-specific DNA

For Egr-1 in complex with quasi-specific DNA (DNA L_S), the NMR spectra of the side chain cationic groups were well dispersed and all 15 arginine residues were able to be assigned. The distribution of the peaks is very wide in comparison to the NMR spectra for free Egr-1 probably due to hydrogen bond or ion pair formation with DNA. In addition, the spectra for Egr-1 in complex with non-specific DNA was not as dispersed as the complex with quasi-specific DNA and only 12 of the 15 residues were able to be assigned from the spectra shown in Figure4.3. In order to assign these residues we performed a salt titration for both the free protein Egr-1 and the Egr-1/DNA L complex. Each titration started at 20 mM KCl and was increased over 7 salt concentration points (20, 100, 200, 300, 450, 600, 750) to 750 mM KCl. At a salt concentration of 750 mM KCl, both the spectra for free Egr-1 and Egr-1/DNA L complex are comparable, likely due to the loss of DNA interaction caused high salt concentrations. As shown in Figure 4.4C, comparing the

change in chemical shifts of the ^{15}N groups at high salt concentrations, there is no significant difference between the change in chemical shifts between free and bound protein. By calculating the difference in chemical shift ($\Delta\delta$, (Figure 4.4A)), we can obtain the relationship of the proteins affinity to DNA at different salt concentration. In Figure 4.4B, we show that there is a linear relationship between the dissociation constant and the salt concentration. As the salt concentration increases, there is a proportional decrease in the proteins affinity for DNA. By then back tracing the spectrum to the original salt concentration, the peaks for the Egr-1/DNA L complex were assigned. However, a 13th residue appears in the relaxation data and was able to be assigned and analyzed from the spectra for relaxation data. Since all of the signals for the Egr-1/DNA L_S and Egr-1/DNA L complex were well isolated or mostly isolated, we can study the change in dynamics from Egr-1's free state to its bound state on quasi-specific or non-specific DNA.

4.2.2 Basic side chain ^{15}N relaxation data for Egr-1/DNA complexes

We conducted relaxation experiments for ^{15}N relaxation of Arg N_ϵ-$^1H_\epsilon$ at ^{1}H frequencies of 750 and 600 MHz. The arginine relaxation data can be seen in Table 4.1 and 4.2 for both complexes of Egr-1 with non-specific and quasi-specific DNA. The relaxation parameters for both the free protein and the specific complex of Egr-1 have been reported in a previous work.[62] The current data and the previous data were collected under the same conditions, making it possible to compare these data to each other. While there are clearly significant

Table 4.1: Relaxation data for Arg side-chain N_ϵ-H_ϵ groups of Egr-1 in complex with quasi-specific DNA L_S

Arg N_ε-H_ε	$^{15}N_\varepsilon$ R_1 (s^{-1}) 600 MHz	$^{15}N_\varepsilon$ R_1 (s^{-1}) 750 MHz	$^{15}N_\varepsilon$ R_2 (s^{-1}) 750 MHz	{^{1}H-} $^{15}N_\varepsilon$ NOE 600 MHz	{^{1}H-} $^{15}N_\varepsilon$ NOE 750 MHz
R3	0.709 ± 0.021	0.559 ± 0.007	7.734 ± 0.106	-0.479 ± 0.065	-0.237 ± 0.033
R14	0.816 ± 0.012	0.674 ± 0.005	7.243 ± 0.067	-0.447 ± 0.044	-0.259 ± 0.022
R15	0.808 ± 0.014	0.638 ± 0.004	6.928 ± 0.059	-0.102 ± 0.033	-0.045 ± 0.019
R18	0.650 ± 0.064	0.483 ± 0.016	18.904 ± 0.590	0.651 ± 0.162	0.518 ± 0.080
R24	0.874 ± 0.106	0.507 ± 0.021	19.939 ± 0.811	0.945 ± 0.298	0.742 ± 0.104
R27	0.717 ± 0.060	0.502 ± 0.013	17.998 ± 0.415	0.446 ± 0.161	0.193 ± 0.047
R38	0.443 ± 0.008	0.358 ± 0.004	3.089 ± 0.046	-1.282 ± 0.041	-0.948 ± 0.033
R42	0.793 ± 0.018	0.650 ± 0.006	10.208 ± 0.087	-0.030 ± 0.054	0.031 ± 0.027
R46	0.608 ± 0.101	0.569 ± 0.022	22.209 ± 0.838	0.459 ± 0.268	0.855 ± 0.092
R55	0.633 ± 0.050	0.506 ± 0.012	16.180 ± 0.285	0.174 ± 0.177	0.459 ± 0.054
R70	0.703 ± 0.026	0.650 ± 0.012	8.327 ± 0.121	-0.371 ± 0.081	-0.287 ± 0.042
R74	0.726 ± 0.100	0.548 ± 0.024	21.517 ± 0.675	-0.305 ± 0.253	0.814 ± 0.108
R78	0.910 ± 0.064	0.687 ± 0.022	12.484 ± 0.289	0.349 ± 0.140	0.386 ± 0.086
R80	1.000 ± 0.034	0.759 ± 0.008	7.785 ± 0.099	0.022 ± 0.071	0.111 ± 0.033
R87	0.415 ± 0.014	0.356 ± 0.021	1.652 ± 0.120	-1.921 ± 0.129	-1.411 ± 0.122

Table 4.2: Relaxation data for Arg side chain N_ϵ-H_ϵ groups of Egr-1 in complex with quasi-specific DNA L

Arg N_ε-H_ε	$^{15}N_\varepsilon$ R_1 (s^{-1}) 600 MHz	$^{15}N_\varepsilon$ R_1 (s^{-1}) 750 MHz	$^{15}N_\varepsilon$ R_2 (s^{-1}) 750 MHz	{^{1}H-} $^{15}N_\varepsilon$ NOE 600 MHz	{^{1}H-} $^{15}N_\varepsilon$ NOE 750 MHz
R3	0.784 ± 0.025	0.572 ± 0.006	4.433 ± 0.061	-0.813 ± 0.023	-0.588 ± 0.025
R14	0.929 ± 0.043	0.673 ± 0.012	9.005 ± 0.119	-0.172 ± 0.034	-0.048 ± 0.032
R15	0.922 ± 0.017	0.658 ± 0.004	6.042 ± 0.049	-0.169 ± 0.016	-0.149 ± 0.019
R24	1.300 ± 0.130	0.775 ± 0.038	22.495 ± 1.210	0.100 ± 0.076	0.213 ± 0.075
R27	0.750 ± 0.018	0.550 ± 0.005	5.174 ± 0.054	-0.353 ± 0.024	-0.437 ± 0.019
R38	0.649 ± 0.015	0.384 ± 0.004	3.751 ± 0.053	-1.427 ± 0.030	-0.822 ± 0.019
R42	0.904 ± 0.028	0.669 ± 0.007	9.009 ± 0.078	-0.327 ± 0.025	-0.180 ± 0.025
R55	0.709 ± 0.089	0.548 ± 0.020	16.269 ± 0.421	0.207 ± 0.058	0.184 ± 0.047
R70	0.855 ± 0.045	0.626 ± 0.010	9.480 ± 0.143	-0.207 ± 0.030	-0.035 ± 0.027
R74	0.895 ± 0.030	0.639 ± 0.007	7.038 ± 0.106	-0.299 ± 0.026	-0.198 ± 0.030
R78	1.007 ± 0.130	0.694 ± 0.027	11.622 ± 0.311	0.037 ± 0.067	0.185 ± 0.048
R80	1.170 ± 0.148	0.691 ± 0.033	15.591 ± 0.409	-0.237 ± 0.062	-0.220 ± 0.064
R87	0.661 ± 0.012	0.467 ± 0.003	2.695 ± 0.030	-1.289 ± 0.020	-0.848 ± 0.015

differences in ^{15}N relaxation data between free protein and protein in complex based on rotational correlation times and changes in internal motions, the differences between the two complexes (nonspecific and quasi-specific DNA) are more subtle with only a few key residues showing changes from each other.

4.2.3 Calculation of order parameters for the arginine side chain groups from the relaxation data of protein in complex with DNA

Table 4.3: S^2 order parameters for the free protein, protein in complex with nonspecific DNA, protein in complex with quasi-specific DNA, and protein in complex with specific DNA. The residues are segregated according to the domain with they are associated. The exceptions are R27 and R55, which are located in linker regions between zinc finger 1 and 2 and zinc finger 2 and 3 respectively.

Side chains	S^2 (Free protein)	S^2 (Complex Nonspecific)	S^2 (Complex Quasispecific)	S^2 (Complex Specific)
Zinc Finger Domain 1				
3	0.215 ± 0.021	0.185 ± 0.005	0.294 ± 0.007	0.393 ± 0.009
14	0.370 ± 0.018	0.386 ± 0.015	0.263 ± 0.005	0.292 ± 0.004
15	0.462 ± 0.027	0.233 ± 0.032	0.247 ± 0.047	0.351 ± 0.004
18	0.295 ± 0.010	NA	0.723 ± 0.030	0.968 ± 0.02
24	0.245 ± 0.018	0.765 ± 0.049	0.860 ± 0.035	0.968 ± 0.033
27	0.220 ± 0.010	0.212 ± 0.025	0.694 ± 0.042	0.883 ± 0.016
Zinc Finger Domain 2				
38	0.228 ± 0.010	0.148 ± 0.002	0.115 ± 0.002	0.108 ± 0.002
42	0.386 ± 0.023	0.388 ± 0.003	0.407 ± 0.062	0.666 ± 0.010
46	0.454 ± 0.022	NA	0.807 ± 0.071	0.962 ± 0.023
55	0.241 ± 0.013	0.693 ± 0.076	0.719 ± 0.018	0.899 ± 0.021
Zinc Finger Domain 3				
70	0.339 ± 0.027	0.415 ± 0.009	0.318 ± 0.049	0.302 ± 0.006
74	0.306 ± 0.012	0.291 ± 0.008	0.859 ± 0.038	0.908 ± 0.039
78	0.399 ± 0.195	0.515 ± 0.028	0.478 ± 0.036	0.630 ± 0.025
80	0.238 ± 0.034	0.690 ± 0.021	0.273 ± 0.008	0.897 ± 0.015
87	0.132 ± 0.024	0.097 ± 0.002	0.050 ± 0.005	0.085 ± 0.003

The relaxation data was used to calculate the S^2 order parameters for Arg N_ϵ-H_ϵ for both complexes. Because the rotational correlation time is

required to calculate these order parameters, ^{15}N backbone relaxation rates R_1 and R_2 were measured at a ^{1}H frequency of 800 MHz and were used to determine the rotational diffusion parameters with a 1.6 Åcrystal structure of the complex (PDBID:1AAY). For the nonspecific complex, the rotational diffusion parameter for zinc finger domain 1 was calculated separately from zinc finger domains 2 and 3. This was done in light of a previous study which showed independent tumbling for ZF 1 on nonspecific DNA, but not for ZF 2 and 3.[48,50] The rotational diffusion parameters for the quasi-specific complex zinc finger domains were calculated together. The diffusion parameters for the complexes are in Table 4.4. Both the rotational diffusion data and the side chain ^{15}N relaxation data were used to calculate the order parameters of the Arg N_ϵ-H_ϵ, which are listed in Table 4.3 along with data previously calculated for the free and specifically bound protein.[62] The S^2 order parameter defines the mobility of the arginine sidechains, with a value of 0 being most mobile, and a value of 1 being completely immobilized.[59] Here we describe the differences of arginine side chain mobility between the proteins bound to two different types of DNA, non-specific and quasi-specific DNA. We then compare it to both the free protein and the protein bound to its specific complex. The relaxation data for both complexes, although similar, have significant differences at residues Arg 27, which is important for the protein's conformational switch,[50] and residue Arg 80, important for a hydrogen bonding of the second guanine within the specific target sequence for Egr-1.

Table 4.4: Rotational correlational times for the Egr-1/nonspecific complex and the Egr-1/quasi-specific complex.

	$\tau_{r,eff}$ (ns)
Nonspecific complex at 35°	
ZF1	13.12 ± 0.23
ZF2-ZF3	14.37 ± 0.10
Quasi-specific complex at 35°	
ZF1-ZF2-ZF3	15.19 ± 0.06

4.2.4 Comparison of free protein with quasi-specifically bound protein

There are 15 known Arg interactions when Egr-1 complexes with its specific target sequence.[62] When measuring the change in mobility of cationic groups of Egr-1 bound to its target, the majority of arginine side chains showed significant perturbations of the S^2 order parameters. This indicates a notable decrease in mobility of the side chains because of their interactions with DNA. However, when comparing the N_ϵ-H_ϵ order parameters of free Egr-1 and the Egr-1 quasi-specific complex (Figure 4.5A), we found significant differences in only 6 of the 15 residues. Four of the residues with decreased mobility (R18, R24, R46, R74) were those with direct interactions with DNA bases. Arginine 24 interacts with just a guanine DNA base in the specific complex, while residues R18, R46, and R74 form not only hydrogen bonds with guanine bases, but also two hydrogen bonds with neighboring aspartate side chains (D20, D48, and D76). The two other residues, R27 and R55, are located in the linker regions between ZF 1 and 2 and ZF 2 and 3 respectively. They also show a large decrease in mobility. This loss in mobility for R27 is likely due

to interaction with Q32 and the DNA phosphate backbone, while R55 interacts with E60. When comparing the quasi-specific complex with the specific complex of Egr-1(see Figure 4.5B.), most of the residues showed a comparable decrease in mobility, with the exception of residue 80. This is expected since the quasi-specific site is missing an important guanine at position 8 of the specific sequence. Loss of guanine 8 reduces affinity of the protein to the specific complex as demonstrated in chapter 2. Without this interaction, Egr-1's affinity to the specific site is significantly decreased.

4.2.5 Conformational changes of Egr-1 bound to nonspecific DNA

The NMR data for Egr-1 in complex with nonspecific DNA indicated decreased mobility for 3 residues: R24, R55, and R80 (Figure 4.6A and B). Two residues, R18 and R46, could not be observed, possibly due to extremely slow chemical exchange between free and bound conformations. Interestingly, one residue, R15, showed a significant increase in mobility in the complex, compared to its free state, but it is unclear why this would occur. It is possible that conformational changes upon binding cause this mobility enhancement as has been suggested for a few other proteins.[120,139,140] Because immobilizing other side chains causes an entropic loss, it is possible that the increased mobility of this noninteracting residue compensates for the entropy loss. Both residues R24 and R80 showed decreased mobility, probably due to interactions with guanine bases on the nonspecific DNA. Residue 55 in the linker region between ZF 2 and 3 showed decreased mobility, which most likely serves to

maintain a conformational change and stabilize the binding of ZF 2 and 3. Interestingly, R27 showed no decrease in mobility, suggesting that ZF 1 is relatively mobile because of the flexible linker region. This is different from the quasi-specific complex, where decreased mobility was observed for both linker residues R27 and R55 (Figure 4.7 on page 103). While there is significant loss of mobility from Egr-1's free state while bound to nonspecific DNA, there is still more mobility when compared to its specifically bound state (Figure 4.6C on page 4.6). It is especially significant in residues R27 and R74, which are highly mobile in comparison to the complex. In fact, the mobility of these residues is the same as the free state of the protein (Figure 4.6A on page 4.6).

4.3 Discussion

4.3.1 Conformational modes of Egr-1 during target search

Previous work on Egr-1 indicates that it has two conformational modes, a scanning mode with ZF 1 free from DNA while ZF 2 and 3 are bound, and recognition mode where all zinc finger domains are bound to DNA. Once completely bound to DNA, all but 2 arginine residues become rigid.[48,50] R27, although in a linker region, also loses its mobility and interacts with Q32 and the DNA phosphate backbone when the protein is in recognition mode. When the protein interacts with nonspecific DNA, R27 in the linker regions between ZF 1 and 2 was still highly mobile, just as in the free state. This suggests that in scanning mode, R27's mobility allows ZF 1 to be free from DNA. In fact, previous experiments tested mutant Egr-1 zinc finger domains

by using site-directed mutagenesis to replace crucial residues, including R27, to affect conformational changes of the zinc finger domains. Zandarashvili et al found that the flexibility of the linker regions affected the kinetic rate Egr-1's search for its target. Forcing the protein into recognition mode caused a decrease in the apparent kinetic rate, while keeping the protein in scanning mode significantly increased the apparent kinetic rate.[50] This observation is supported in our current work, where we found the difference in mobility between protein bound to nonspecific and quasi-specific DNA mainly lay in the mobility of the residues found in the linker regions of the zinc fingers. Residues 27 and 55, between domains 1 and 2, and domains 2 and 3 respectively lend rigidity to the protein domains once specifically bound to DNA. In the quasi-specific sequence, both residues 27 and 55 showed decreased mobility indicating that inter-domain motion is reduced in this state. However, in the nonspecific complex, only residue 55 showed decreased mobility, while residue 27 remained highly mobile. This suggests that due to the mobility of this residue, the first domain of the zinc finger remains dynamic. This would allow for domain 1 to be more mobile than the other two domains, contributing to the scanning conformation of the protein. This is also supported by rotational correlation time data, which shows a lower correlation time rate for the first domain of the zinc finger in complex with DNA than for the other two domains (see Table 4.4). This is not the case for protein bound to quasi-specific DNA, which shows a larger correlational time for all domains. Overall, the data indicates two different conformational modes of the protein during the target search.

4.3.2 Third domain reduced interactions with quasi-specific DNA

When the protein is bound to the quasi-specific sequence, most of the arginine side chains showed decreased mobility with the exception of residue R80. When bound to the specific DNA sequence, R80 form two hydrogen bonds to guanine 8, which has been replaced in the quasi-specific sequence with an adenine. Without the the guanine at position 8, R80's mobility is almost identical to free protein, indicating that no interaction is occurring to keep it stable. Recent work has shown that the binding of ZF 3 to DNA lends a significant amount of stability to the complex,[141] therefore complete loss of binding for R80 would cause a significant in decrease in stability and affinity for DNA. This was qualitatively demonstrated in chapter 2 where we showed Egr-1's reduced affinity to DNA L_S compared to its specific site. While for the specific site loss of R80 interaction would be detrimental to proper binding, it is beneficial for off-target search, since it would reduce the trapping affect causing by binding to off-target locations.

4.4 Concluding Remarks

This chapter demonstrates the differences in arginine side chains' mobility when Egr-1 is bound to quasi-specific and nonspecific sequences. We show that there is a significant decrease in mobility for both types of DNA, but the decrease is not as strong as that for Egr-1 binding to specific DNA. The differences in mobility between quasi-specific and non-specific DNA are subtle, with only 3 residues significantly different in mobility. Of importance

is the linker residue R27, whose mobility in the nonspecific complex, support the idea of two conformational modes, scanning and recognition. This work provides atomic level data describing the nature of of Egr-1's arginine side chain dynamics during the scanning process. Further experiments are needed to explore the properties of other residues such as lysine, which also interact at the protein-DNA interface.

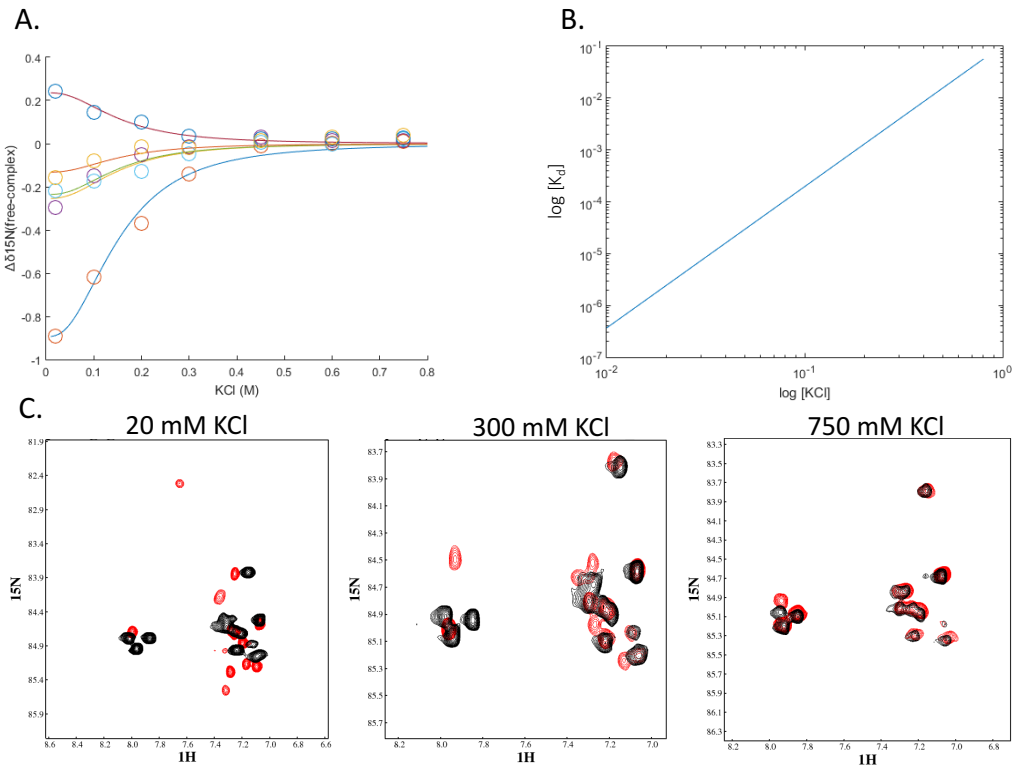

Figure 4.4: A)The difference in ^{15}N chemical shift from the Arginine residues from free Egr-1 versus Egr-1 bound to nonspecific DNA. As salt concentration increases, the chemical shift difference decreases until they reach approximately zero. Since the chemical shifts for both the complex and free are equivalent at 750 mM KCl, the spectra can be compared and the complex can be assigned from the currently known assignments from the free Egr-1 spectrum. B)Values a and b from equation 4.2 were plotted as a function of salt concentration to determine the K_d over increasing salt concentrations. C) An overlay of an HISQC spectrum for free Egr-1 (black spectrum) and Egr-1/DNA L complex (red spectrum) over three different salt concentrations. The difference in the spectra at 20 mM is significant, but as salt concentration is increased, the spectra become nearly identical as seen in the spectrum at 750 mM KCl.

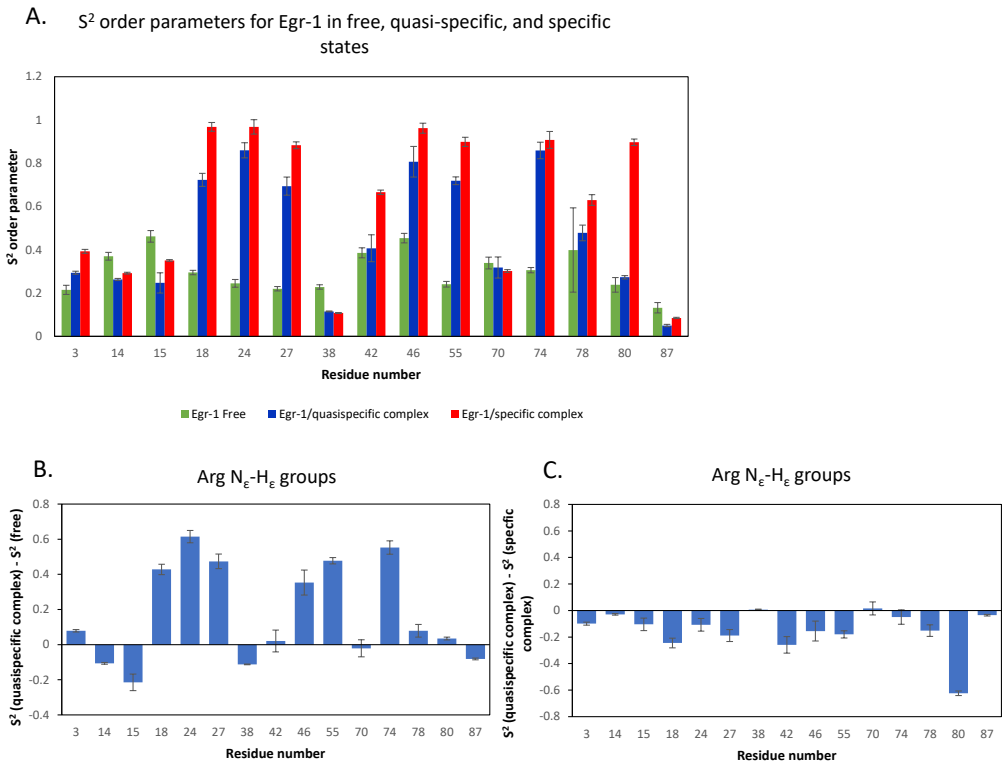

Figure 4.5: A) S^2 order parameters for free Egr-1 (green), quasi-specific complex (blue), and specific complex (red). A larger value indicates decreased mobility. B) Difference of the S^2 order parameters between Egr-1 bound to quasi-specific DNA and free Egr-1. The larger value indicates a decrease in mobility from the free state. C)Difference in S^2 order parameter between Egr-1 bound to quasi-specific DNA and Egr-1 bound to specific DNA. The negative difference indicates an increase in mobility in the quasi-specific bound protein compared to the specifically bound state.

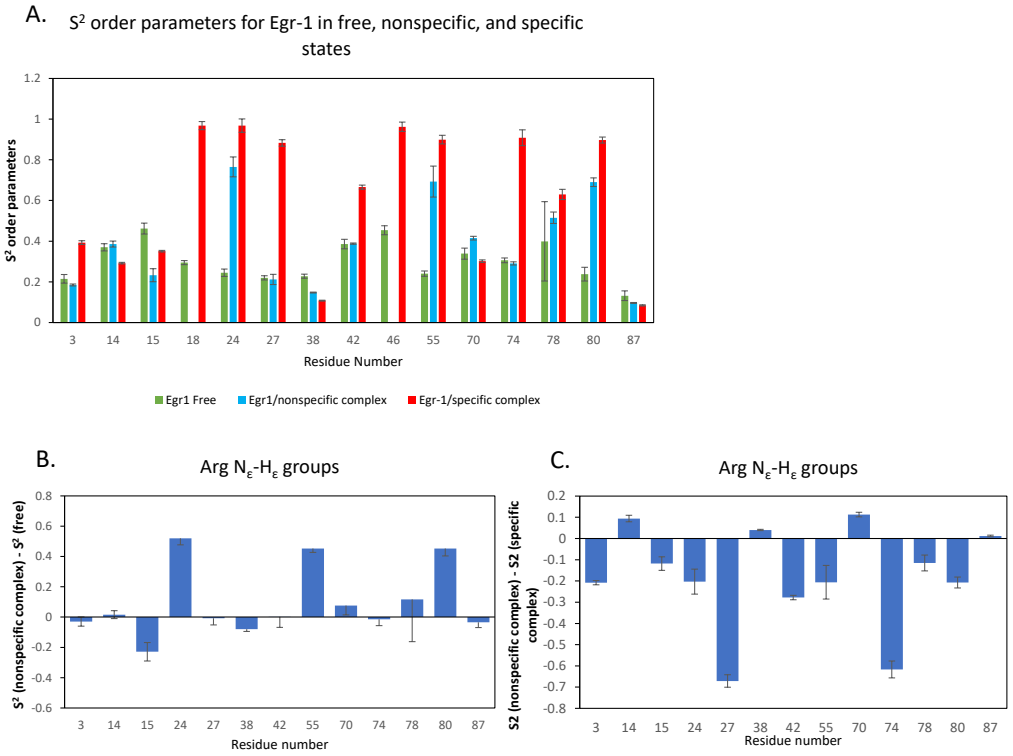

Figure 4.6: A) S^2 order parameters for free Egr-1 (green), nonspecific complex (cyan), and specific complex. The nonspecific complex is missing two residues (R18 and R46), since they were not able to be resolved in the spectrum. B) Difference is S^2 order parameters between Egr-1 bound to nonspecific DNA and free Egr-1. The larger value indicates a decrease in mobility from the free state. C)Difference in S^2 order parameter between Egr-1 bound to non-specific DNA and Egr-1 bound to specific DNA. The negative difference indicates an increase in mobility in the nonspecific bound protein compared to the specifically bound state.

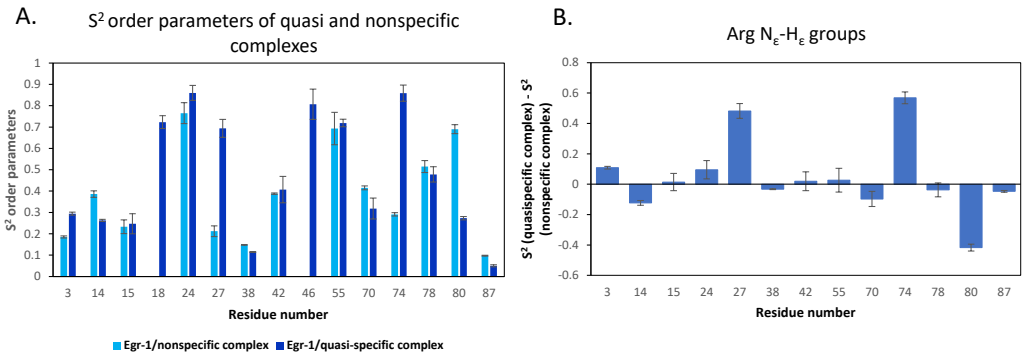

Figure 4.7: A) S^2 order parameters of nonspecific complex and the quasi-specific complex. B) Difference of the S^2 order parameters between Egr-1 bound to quasi-specific DNA and nonspecific DNA. The larger value indicates a decrease in mobility from the nonspecific bound state.

Chapter 5

Conclusions

Our current work provides a new perspective on target DNA search mechanisms for transcription factors. In chapter 2, our study demonstrated a quantitative description of the impact of quasi-specific sites on target search kinetics of 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 bi-phasic (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 is much faster because the strong trapping effect largely disappears. 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. [1]

[1] Adapted with permission from Kemme, Catherine A., et al. "Influence of quasi-specific sites on kinetics of target DNA search by a sequence-specific DNA-binding protein." Biochemistry 54.44 (2015): 6684. http://pubs.acs.org/doi/abs/10.1021%2Facs.biochem.5b00967

The idea of quasi-specific sites affecting regulation of transcription factors via indirect interplay with other nuclear proteins lays the foundation for chapter 3. Here we showed that the target search process for transcription factors can be greatly accelerated when other proteins block such quasi-specific sites(decoys). This mechanism does not require any direct protein-protein interactions, although the current paradigm for synergy between transcription factors typically assume their direct interactions.[76,77] In chapter 3, we demonstrate that MeCP2 MBD facilitates the association of the Egr-1 zinc-finger protein with the unmethylated target by blocking methylated decoys. In principle, other proteins could also provide the same acceleration mechanisms, as long as they selectively block decoys but not targets. For example, highly expressed transcription factors with similar but different sequence specificity may enhance the function of other transcription factors by selectively blocking their decoys. Natural decoys (NDs) and their DNA methylation may play key roles in regulation of transcription factors.[2]

To delve further into transcription factor's interactions with DNA during the scanning process, in chapter 4 we analyze Egr-1's arginine side chain interactions with DNA bases when it is bound to non-specific DNA and quasi-specific DNA. We demonstrate that while there is a significant decrease in mobility of some of the Arg side chains when it is non-specifically bound, there is

[2]Adapted from Kemme, Catherine A., et al. "Potential role of DNA methylation as a facilitator of target search processes for transcription factors through interplay with methyl-CpG-binding proteins." Nucleic Acids Research (2017) under the Creative Commons License (http://creativecommons.org/licenses/by-nc/4.0/)

a crucial residue which is mobile. This residue was previously suggested to be involved in Egr-1's conformational switch between scanning and recognition mode.[50] The scanning mode allows the protein to be partially bound to DNA, possibly reducing the trapping effect *in vivo* caused by a significant range of NDs.

5.1 Future directions

5.1.1 Chromatin structure and natural decoys

The sequestration of transcription factors via NDs possibly depends on chromatin structure. Janssen et al. studied drug-induced chromatin openings of DNA satellite V involving GAGAA repeats in *Drosophila*.[142] They found that chromatin opening led to increased sequestration of the GAGA factor and reduced expression of its target genes.[142] This finding suggests that the locations of NDs in the genome are important for functional sequestration of transcription factors. When NDs are located in accessible regions, they may sequester transcription factors to a greater degree. Because acetylation and methylation of histone tails are associated with the regulation of chromatin structure,[143] the accessibility of NDs may, in principle, be assessed through the bioinformatics analysis of databases on nucleosome positions and histone modifications. Such investigations might allow for the prediction of the efficacy of NDs for each transcription factor.[3]

[3]Adapted with permission from Kemme, Catherine A., et al. "Regulation of transcription factors via natural decoys in genomic DNA." Transcription 7.4 (2016): 115-120.

5.1.2 Synthetic versus natural decoys for transcription factors

Understanding the mechanisms underlying how transcription factors recognize DNA, will improve development of genetic therapeutics such as synthetic DNA decoys. Abundant NDs may adversely impact the effectiveness of synthetic DNA decoys, which are short duplexes designed to inhibit particular transcription factors for therapeutic purposes. Applications of this approach for various transcription factors have been examined. For example, STAT3, which is constitutively active in cancerous cells, has been targeted to abrogate the growth of head and neck cancer cells;[144] Egr-1 to inhibit neointimal hyperplasia;[145] and NF-κB to prevent myocardial infarction.[146] Though this approach was successful in some applications, the synthetic decoys typically produced only modest inhibitory effects compared to other oligonucleotide-based gene suppression methods.[147]

This inadequate inhibition may be partially due to the presence of NDs in genomic DNA, although the short life span and poor delivery efficiency of the synthetic decoys may also be responsible. The synthetic decoys must compete with these NDs for transcription factors, and as mentioned above, the net concentration of the accessible NDs in the nuclei could be as high as $\sim 10^{-5}$-10^{-6} M. Because an uptake of the synthetic decoys at more than 10^{-6} M, in the nuclei of living cells, is very unlikely in practice, it is difficult for the synthetic decoys to competitively overcome the NDs unless the synthetic decoys exhibit a much higher affinity than the NDs. To achieve such conditions, the oxygen-to-sulfur substitution in the phosphate groups of DNA backbone that interact

with the protein may be useful.[100] Additionally, the inhibition of transcription factors by synthetic decoys should occur more effectively in the cytoplasm (i.e., pre-nuclear localization), due to the absence of NDs. For the successful therapeutic application of natural decoys, it may be necessary to consider competition with NDs in genomic DNA. [4]

5.2 Final remarks

NDs can be regarded as a novel class of regulatory DNA that controls the activities of sequence-specific transcription factors by precluding them from binding to their functional target sites on DNA. In the nuclei, NDs always exist in large quantities without the need for expression. However, NDs' inhibitory effects on transcription factors depend on various factors, such as chromatin structure, CpG methylation, and competition with other proteins. The functional activity of a transcription factor may be greatly enhanced through blocking of its NDs by other proteins. When most NDs become inacessible, the transition in target association behavior may resemble the behavior of an on-off switch. Thus, the sequestration of transcription factors in NDs could serve as a controllable mechanism of gene regulation. Since the ENCODE project stated that 80% of the human genome is functional,[15] the role of so-called "junk DNA" has been controversial. Each ND may be junk in terms of primary sequence and non-functional compared to the target sites, but ensembles

[4] Adapted with permission from Kemme, Catherine A., et al. "Regulation of transcription factors via natural decoys in genomic DNA." Transcription 7.4 (2016): 115-120.

of abundant NDs in the genome may have profound effects on the functions of transcription factors. In addition, these sites, depending on their sequence can change how the protein binds to DNA, through the alteration of residual interactions. Alteration of residual interactions causes a conformational change in the protein and affects how the protein scans DNA. Currently, very little is known about NDs, and further characterizations, including analysis of ND distributions in the genome, are necessary. Integration of cell biology, biochemistry, biophysics, and bioinformatics is required to delineate the roles of NDs in the transcriptional regulation of genes. [5]

[5] Adapted with permission from Kemme, Catherine A., et al. "Regulation of transcription factors via natural decoys in genomic DNA." Transcription 7.4 (2016): 115-120.

Bibliography

1. Stephen Buratowski. Mechanisms of gene activation. *Science*, 270(5243):1773, 1995.

2. Harley H McAdams and Adam Arkin. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences*, 94(3):814–819, 1997.

3. B. Lewin. *Genes VIII*. Oxford University Press, Oxford, U.K., 2001.

4. Anton Valouev, Steven M Johnson, Scott D Boyd, Cheryl L Smith, Andrew Z Fire, and Arend Sidow. Determinants of nucleosome organization in primary human cells. *Nature*, 474(7352):516, 2011.

5. Hinrich Boeger, David A Bushnell, Ralph Davis, Joachim Griesenbeck, Yahli Lorch, J Seth Strattan, Kenneth D Westover, and Roger D Kornberg. Structural basis of eukaryotic gene transcription. *FEBS letters*, 579(4):899–903, 2005.

6. Edgar Serfling, Maria Jasin, and Walter Schaffner. Enhancers and eukaryotic gene transcription. *Trends in Genetics*, 1:224–230, 1985.

7. Gwenael Badis, Michael F Berger, Anthony A Philippakis, Shaheynoor Talukder, Andrew R Gehrke, Savina A Jaeger, Esther T Chan, Genita Metzler, Anastasia Vedenko, Xiaoyu Chen, et al. Diversity and complexity in dna recognition by transcription factors. *Science*, 324(5935):1720–1723, 2009.

8. Zeba Wunderlich and Leonid A Mirny. Different gene regulation strategies revealed by analysis of binding motifs. *Trends in genetics*, 25(10):434–440, 2009.

9. Ronald C Conaway and Joan Weliky Conaway. Function and regulation of the mediator complex. *Current opinion in genetics & development*, 21(2):225–230, 2011.

10. Peter H von Hippel and OG Berg. Facilitated target location in biological systems. *Journal of Biological Chemistry*, 264(2):675–678, 1989.

11. Robert B Winter, Otto G Berg, and Peter H Von Hippel. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. the escherichia coli lac repressor-operator interaction: kinetic measurements and conclusions. *Biochemistry*, 20(24):6961–6977, 1981.

12. Otto G Berg, Robert B Winter, and Peter H Von Hippel. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. models and theory. *Biochemistry*, 20(24):6929–6948, 1981.

13. Arthur D Riggs, Suzanne Bourgeois, and Melvin Cohn. The lac represser-operator interaction: Iii. kinetic studies. *Journal of molecular biology*, 53(3):401–417, 1970.

14. David S Johnson, Ali Mortazavi, Richard M Myers, and Barbara Wold. Genome-wide mapping of in vivo protein-dna interactions. *Science*, 316(5830):1497–1502, 2007.

15. ENCODE Project Consortium et al. An integrated encyclopedia of dna elements in the human genome. *Nature*, 489(7414):57, 2012.

16. Xiao-yong Li, Stewart MacArthur, Richard Bourgon, David Nix, Daniel A Pollard, Venky N Iyer, Aaron Hechmer, Lisa Simirenko, Mark Stapleton, Cris L Luengo Hendriks, et al. Transcription factors bind thousands of active and inactive regions in the drosophila blastoderm. *PLoS biology*, 6(2):e27, 2008.

17. W Ford Doolittle. Is junk dna bunk? a critique of encode. *Proceedings of the National Academy of Sciences*, 110(14):5294–5300, 2013.

18. Dan Graur, Yichen Zheng, Nicholas Price, Ricardo BR Azevedo, Rebecca A Zufall, and Eran Elhaik. On the immortality of television sets:function in the human genome according to the evolution-free gospel of encode. *Genome biology and evolution*, 5(3):578–590, 2013.

19. Tertia De Jager, Theo Pelzer, Stephan Müller-Botz, Asiya Imam, Jenny Muck, and Ludwig Neyses. Mechanisms of estrogen receptor action in the myocardium rapid gene activation via the erk1/2 pathway and serum response elements. *Journal of Biological Chemistry*, 276(30):27873–27880, 2001.

20. Thomas Eulgem, Paul J Rushton, Elmon Schmelzer, Klaus Hahlbrock, and Imre E Somssich. Early nuclear events in plant defence signalling: rapid gene activation by wrky transcription factors. *The EMBO Journal*, 18(17):4689–4699, 1999.

21. Alexandre Esadze, Catherine A Kemme, Anatoly B Kolomeisky, and Junji Iwahara. 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 research*, 42(11):7039–7046, 2014.

22. Dana Vuzman, Ariel Azia, and Yaakov Levy. Searching dna via a monkey bar mechanism: the significance of disordered tails. *Journal of molecular biology*, 396(3):674–684, 2010.

23. Amir Marcovitz and Yaakov Levy. Obstacles may facilitate and direct dna search by proteins. *Biophysical journal*, 104(9):2042–2050, 2013.

24. Marijn TJ van Loenhout, MV de Grunt, and Cees Dekker. Dynamics of dna supercoils. *Science*, 338(6103):94–97, 2012.

25. C Loverdo, Olivier Benichou, Raphael Voituriez, A Biebricher, I Bonnet, and P Desbiolles. Quantifying hopping and jumping in facilitated diffusion of dna-binding proteins. *Physical review letters*, 102(18):188101, 2009.

26. Joseph D Schonhoft and James T Stivers. Timing facilitated site transfer of an enzyme on dna. *Nature chemical biology*, 8(2):205–210, 2012.

27. Otto G Berg and Clas Blomberg. Association kinetics with coupled diffusional flows: special application to the lac repressor-operator system. *Biophysical chemistry*, 4(4):367–381, 1976.

28. Jason Gorman and Eric C Greene. Visualizing one-dimensional diffusion of proteins along dna. *Nature structural & molecular biology*, 15(8):768–774, 2008.

29. Jason Gorman, Aaron J Plys, Mari-Liis Visnapuu, Eric Alani, and Eric C Greene. Visualizing one-dimensional diffusion of eukaryotic dna repair factors along a chromatin lattice. *Nature structural & molecular biology*, 17(8):932–938, 2010.

30. Johan Elf, Gene-Wei Li, and X Sunney Xie. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*, 316(5828):1191–1194, 2007.

31. Kaushik Ragunathan, Cheng Liu, and Taekjip Ha. Reca filament sliding on dna facilitates homology search. *Elife*, 1:e00067, 2012.

32. Jason C Bell, Jody L Plank, Christopher C Dombrowski, and Stephen C Kowalczykowski. Direct imaging of reca nucleation and growth on single molecules of ssb-coated ssdna. *Nature*, 491(7423):274, 2012.

33. Jichuan Zhang, Ruobo Zhou, Jin Inoue, Tsutomu Mikawa, and Taekjip Ha. Single molecule analysis of thermus thermophilus ssb protein dynamics on single-stranded dna. *Nucleic acids research*, 42(6):3821–3832, 2013.

34. Géraldine Farge, Niels Laurens, Onno D Broekmans, Siet Mjl Van Den Wildenberg, Linda CM Dekker, Martina Gaspari, Claes M Gustafsson, Erwin JG Peterman, Maria Falkenberg, and Gijs JL Wuite. Protein sliding and dna denaturation are essential for dna organization by human mitochondrial transcription factor a. *Nature communications*, 3:1013, 2012.

35. Luda S Shlyakhtenko, Alexander Y Lushnikov, Atsushi Miyagi, Ming Li, Reuben S Harris, and Yuri L Lyubchenko. Nanoscale structure and dynamics of abobec3g complexes with single-stranded dna. *Biochemistry*, 51(32):6432–6440, 2012.

36. G Marius Clore. Exploring translocation of proteins on dna by nmr. *Journal of biomolecular NMR*, 51(3):209, 2011.

37. G Marius Clore and Junji Iwahara. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chemical reviews*, 109(9):4108–4139, 2009.

38. Anahita Tafvizi, Leonid A Mirny, and Antoine M van Oijen. Dancing on dna: kinetic aspects of search processes on dna. *Chemphyschem*, 12(8):1481–1489, 2011.

39. Stephen E Halford. An end to 40 years of mistakes in dna–protein association kinetics?, 2009.

40. Tek-Hyung Lee and Narendra Maheshri. A regulatory role for repeated decoy transcription factor binding sites in target gene expression. *Molecular systems biology*, 8(1):576, 2012.

41. Catherine A Kemme, Alexandre Esadze, and Junji Iwahara. Influence of quasi-specific sites on kinetics of target dna search by a sequence-specific dna-binding protein. *Biochemistry*, 54(44):6684–6691, 2015.

42. Xiaowei Liu, Bo Wu, Jaroslaw Szary, Eric M Kofoed, and Fred Schaufele. Functional sequestration of transcription factor activity by repetitive dna. *Journal of Biological Chemistry*, 282(29):20868–20876, 2007.

43. Qi-Qun Tang and M Daniel Lane. Activation and centromeric localization of ccaat/enhancer-binding proteins during the mitotic clonal expansion of adipocyte differentiation. *Genes & development*, 13(17):2231–2241, 1999.

44. Chuanying Chen, Alexandre Esadze, Levani Zandarashvili, Dan Nguyen, B Montgomery Pettitt, and Junji Iwahara. Dynamic equilibria of short-range electrostatic interactions at molecular interfaces of protein–dna complexes. *The journal of physical chemistry letters*, 6(14):2733–2737, 2015.

45. Alexandre Esadze and Junji Iwahara. Stopped-flow fluorescence kinetic study of protein sliding and intersegment transfer in the target dna search process. *Journal of molecular biology*, 426(1):230–244, 2014.

46. Junji Iwahara and Yaakov Levy. Speed-stability paradox in dna-scanning by zinc-finger proteins. *Transcription*, 4(2):58–61, 2013.

47. Yuki Takayama, Debashish Sahu, and Junji Iwahara. Nmr studies of translocation of the zif268 protein between its target dna sites. *Biochemistry*, 49(37):7998–8005, 2010.

48. Levani Zandarashvili, Dana Vuzman, Alexandre Esadze, Yuki Takayama, Debashish Sahu, Yaakov Levy, and Junji Iwahara. Asymmetrical roles of zinc fingers in dynamic dna-scanning process by the inducible transcription factor egr-1. *Proceedings of the National Academy of Sciences*, 109(26):E1724–E1732, 2012.

49. Levani Zandarashvili, Mark A White, Alexandre Esadze, and Junji Iwahara. Structural impact of complete cpg methylation within target dna on specific complex formation of the inducible transcription factor egr-1. *FEBS letters*, 589(15):1748–1753, 2015.

50. Levani Zandarashvili, Alexandre Esadze, Dana Vuzman, Catherine A Kemme, Yaakov Levy, and Junji Iwahara. Balancing between affinity and speed in target dna search by zinc-finger proteins via modulation of dynamic conformational ensemble. *Proceedings of the National Academy of Sciences*, 112(37):E5142–E5149, 2015.

51. Andrew C Jamieson, Jeffrey C Miller, and Carl O Pabo. Drug discovery with engineered zinc-finger proteins. *Nature reviews. Drug discovery*, 2(5):361, 2003.

52. Fyodor D Urnov, Edward J Rebar, Michael C Holmes, H Steve Zhang, and Philip D Gregory. Genome editing with engineered zinc finger nucleases. *Nature reviews. Genetics*, 11(9):636, 2010.

53. David J Segal, Roger R Beerli, Pilar Blancafort, Birgit Dreier, Karin Effertz, Adrian Huber, Beate Koksch, Caren V Lund, Laurent Magnenat,

David Valente, et al. Evaluation of a modular strategy for the construction of novel polydactyl zinc finger dna-binding proteins. *Biochemistry*, 42(7):2137–2148, 2003.

54. Levon M Khachigian, Volkhard Lindner, Amy J Williams, Tucker Collins, et al. Egr-1-induced endothelial gene expression: a common theme in vascular injury. *SCIENCE-NEW YORK THEN WASHINGTON-*, pages 1427–1430, 1996.

55. Shi-Fang Yan, Tomoyuki Fujita, Jiesheng Lu, Kenji Okada, Yu Shan Zou, Nigel Mackman, David J Pinsky, and David M Stern. Egr-1, a master switch coordinating upregulation of divergent gene families underlying ischemic stress. *Nature medicine*, 6(12):1355–1362, 2000.

56. Jonathan LC Lee, Barry J Everitt, and Kerrie L Thomas. Independent cellular processes for hippocampal memory consolidation and reconsolidation. *Science*, 304(5672):839–843, 2004.

57. Bruno Bozon, Sabrina Davis, and Serge Laroche. A requirement for the immediate early gene zif268 in reconsolidation of recognition memory after retrieval. *Neuron*, 40(4):695–701, 2003.

58. XM Cao, RA Koski, A Gashler, M McKiernan, CF Morris, R Gaffney, RV Hay, and VP Sukhatme. Identification and characterization of the egr-1 gene product, a dna-binding zinc finger protein induced by differentiation and growth signals. *Molecular and cellular biology*, 10(5):1931–1939, 1990.

59. Junji Iwahara, Alexandre Esadze, and Levani Zandarashvili. Physicochemical properties of ion pairs of biological macromolecules. *Biomolecules*, 5(4):2435–2463, 2015.

60. Peter L Privalov, Anatoly I Dragan, and Colyn Crane-Robinson. Interpreting protein/dna interactions: distinguishing specific from nonspecific and electrostatic from non-electrostatic components. *Nucleic acids research*, 39(7):2483–2491, 2010.

61. M Thomas Record, Jeung-Hoi Ha, and Matthew A Fisher. [16] analysis of equilibrium and kinetic measurements to determine thermodynamic origins of stability and specificity and mechanism of formation of site-specific complexes between proteins and helical dna. *Methods in enzymology*, 208:291–343, 1991.

62. Alexandre Esadze, Chuanying Chen, Levani Zandarashvili, Sourav Roy, B Montgometry Pettitt, and Junji Iwahara. Changes in conformational dynamics of basic side chains upon protein–dna association. *Nucleic acids research*, 44(14):6961–6970, 2016.

63. Stephen E Halford and John F Marko. How do site-specific dna-binding proteins find their targets? *Nucleic acids research*, 32(10):3040–3052, 2004.

64. Leonid Mirny, Michael Slutsky, Zeba Wunderlich, Anahita Tafvizi, Jason Leith, and Andrej Kosmrlj. How a protein searches for its site on dna: the mechanism of facilitated diffusion. *Journal of Physics A: Mathematical and Theoretical*, 42(43):434013, 2009.

65. Anatoly B Kolomeisky. Physics of protein–dna interactions: mechanisms of facilitated target search. *Physical Chemistry Chemical Physics*, 13(6):2088–2095, 2011.

66. Andrea I Ramos and Scott Barolo. Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. *Phil. Trans. R. Soc. B*, 368(1632):20130018, 2013.

67. Bruce Alberts. *Molecular biology of the cell*. Garland science, 2017.

68. Anat Burger, Aleksandra M Walczak, and Peter G Wolynes. Abduction and asylum in the lives of transcription factors. *Proceedings of the National Academy of Sciences*, 107(9):4016–4021, 2010.

69. Robert J Klose and Adrian P Bird. Genomic dna methylation: the mark and its mediators. *Trends in biochemical sciences*, 31(2):89–97, 2006.

70. Adam Blattler and Peggy J Farnham. Cross-talk between site-specific transcription factors and dna methylation states. *Journal of Biological Chemistry*, 288(48):34287–34294, 2013.

71. Yehudit Bergman and Howard Cedar. Dna methylation dynamics in health and disease. *Nature structural & molecular biology*, 20(3):274–281, 2013.

72. Junji Iwahara, Markus Zweckstetter, and G Marius Clore. Nmr structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific dna. *Proceedings of the National Academy of Sciences*, 103(41):15062–15067, 2006.

73. Nicholas M Luscombe, Roman A Laskowski, and Janet M Thornton. Amino acid–base interactions: a three-dimensional analysis of protein–dna interactions at an atomic level. *Nucleic acids research*, 29(13):2860–2874, 2001.

74. Katalin Nadassy, Shoshana J Wodak, and Joël Janin. Structural features of protein- nucleic acid recognition sites. *Biochemistry*, 38(7):1999–2017, 1999.

75. Remo Rohs, Xiangshu Jin, Sean M West, Rohit Joshi, Barry Honig, and Richard S Mann. Origins of specificity in protein-dna recognition. *Annual review of biochemistry*, 79:233–269, 2010.

76. Ngaio C Smith and Jacqueline M Matthews. Mechanisms of dna-binding specificity and functional gene regulation by transcription factors. *Current opinion in structural biology*, 38:68–74, 2016.

77. Cynthia Wolberger. Multiprotein-dna complexes in transcriptional regulation. *Annual review of biophysics and biomolecular structure*, 28(1):29–56, 1999.

78. Pablo Tebas, David Stein, Winson W Tang, Ian Frank, Shelley Q Wang, Gary Lee, S Kaye Spratt, Richard T Surosky, Martin A Giedlin, Geoff Nichol, et al. Gene editing of ccr5 in autologous cd4 t cells of persons infected with hiv. *New England Journal of Medicine*, 370(10):901–910, 2014.

79. Huan-Xiang Zhou. A model for the mediation of processivity of dna-targeting proteins by nonspecific binding: dependence on dna length and presence of obstacles. *Biophysical journal*, 88(3):1608–1615, 2005.

80. J Chakrabarti, Navin Chandra, Paromita Raha, and Siddhartha Roy. High-affinity quasi-specific sites in the genome: how the dna-binding proteins cope with them. *Biophysical journal*, 101(5):1123–1129, 2011.

81. Amir Marcovitz and Yaakov Levy. Frustration in protein–dna binding influences conformational switching and target search kinetics. *Proceedings of the National Academy of Sciences*, 108(44):17957–17962, 2011.

82. Monicia Elrod-Erickson, Mark A Rould, Lena Nekludova, and Carl O Pabo. Zif268 protein–dna complex refined at 1.6 å: a model system for understanding zinc finger–dna interactions. *Structure*, 4(10):1171–1180, 1996.

83. Nikola P Pavletich and Carl O Pabo. Zinc finger-dna recognition: crystal structure of a zif268-dna complex at 2.1 a. *Science*, 252(5007):809, 1991.

84. Tao Hu, A Yu Grosberg, and BI Shklovskii. How proteins search for their specific sites on dna: the role of dna conformation. *Biophysical journal*, 90(8):2731–2744, 2006.

85. Alex Veksler and Anatoly B Kolomeisky. Speed-selectivity paradox in the protein search for targets on dna: is it real or not? *The journal of physical chemistry B*, 117(42):12695–12701, 2013.

86. Christopher T Harbison, D Benjamin Gordon, Tong Ihn Lee, Nicola J Rinaldi, Kenzie D Macisaac, Timothy W Danford, Nancy M Hannett, Jean-Bosco Tagne, David B Reynolds, Jane Yoo, et al. Transcriptional regulatory code of a eukaryotic genome. *Nature*, 431(7004):99, 2004.

87. William W Fisher, Jingyi Jessica Li, Ann S Hammonds, James B Brown, Barret D Pfeiffer, Richard Weiszmann, Stewart MacArthur, Sean Thomas, John A Stamatoyannopoulos, Michael B Eisen, et al. Dna regions bound at low occupancy by transcription factors do not drive patterned reporter

gene expression in drosophila. *Proceedings of the National Academy of Sciences*, 109(52):21330–21335, 2012.

88. Anne-Laure Todeschini, Adrien Georges, and Reiner A Veitia. Transcription factors: specific dna binding and specific gene regulation. *Trends in genetics*, 30(6):211–219, 2014.

89. Radosveta Koldamova, Jonathan Schug, Martina Lefterova, Andrea A Cronican, Nicholas F Fitz, Faith A Davenport, Alexis Carter, Emilie L Castranio, and Iliya Lefterov. Genome-wide approaches reveal egr1-controlled regulatory networks associated with neurodegeneration. *Neurobiology of disease*, 63:107–114, 2014.

90. Atsutaka Kubosaki, Yasuhiro Tomaru, Michihira Tagami, Erik Arner, Hisashi Miura, Takahiro Suzuki, Masanori Suzuki, Harukazu Suzuki, and Yoshihide Hayashizaki. Genome-wide investigation of in vivo egr-1 binding sites in monocytic differentiation. *Genome biology*, 10(4):R41, 2009.

91. Aimée M Deaton and Adrian Bird. Cpg islands and the regulation of transcription. *Genes & development*, 25(10):1010–1022, 2011.

92. Ana Carolina Dantas Machado, Tianyin Zhou, Satyanarayan Rao, Pragya Goel, Chaitanya Rastogi, Allan Lazarovici, Harmen J Bussemaker, and Remo Rohs. Evolving insights on how cytosine methylation affects protein–dna binding. *Briefings in functional genomics*, 14(1):61–73, 2014.

93. Hideharu Hashimoto, Yusuf Olatunde Olanrewaju, Yu Zheng, Geoffrey G Wilson, Xing Zhang, and Xiaodong Cheng. Wilms tumor protein recognizes 5-carboxylcytosine within a specific dna sequence. *Genes & development*, 28(20):2304–2313, 2014.

94. Dirk Schübeler. Function and information content of dna methylation. *Nature*, 517(7534):321, 2015.

95. Peter J Skene, Robert S Illingworth, Shaun Webb, Alastair RW Kerr, Keith D James, Daniel J Turner, Rob Andrews, and Adrian P Bird.

Neuronal mecp2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Molecular cell*, 37(4):457–468, 2010.

96. Levani Zandarashvili, Alexandre Esadze, Catherine A Kemme, Abhijnan Chattopadhyay, Dan Nguyen, and Junji Iwahara. Residence times of molecular complexes in solution from nmr data of intermolecular hydrogen-bond scalar coupling. *The journal of physical chemistry letters*, 7(5):820–824, 2016.

97. Abhijnan Chattopadhyay, Alexandre Esadze, Sourav Roy, and Junji Iwahara. Nmr scalar couplings across intermolecular hydrogen bonds between zinc-finger histidine side chains and dna phosphate groups. *The Journal of Physical Chemistry B*, 120(41):10679–10685, 2016.

98. Abhijnan Chattopadhyay, Levani Zandarashvili, Ross H Luu, and Junji Iwahara. Thermodynamic additivity for impacts of base-pair substitutions on association of the egr-1 zinc-finger protein with dna. *Biochemistry*, 55(47):6467–6474, 2016.

99. Thomas D Pollard and M Enrique. Take advantage of time in your experiments: a guide to simple, informative kinetics assays. *Molecular biology of the cell*, 24(8):1103–1110, 2013.

100. Levani Zandarashvili, Dan Nguyen, Kurtis M Anderson, Mark A White, David G Gorenstein, and Junji Iwahara. Entropic enhancement of protein-dna affinity by oxygen-to-sulfur substitution in dna phosphate. *Biophysical journal*, 109(5):1026–1037, 2015.

101. J.R.i Lakowicz. *Principles of Fluorescence Spectroscopy, 3rd ed.* Springer, NY, 2006.

102. Victoria Valinluck Lao, Agus Darwanto, and Lawrence C Sowers. Impact of base analogues within a cpg dinucleotide on the binding of dna by the methyl-binding domain of mecp2 and methylation by dnmt1. *Biochemistry*, 49(47):10228–10236, 2010.

103. Adrian Bird. Dna methylation patterns and epigenetic memory. *Genes & development*, 16(1):6–21, 2002.

104. Keith D Robertson. Dna methylation and human disease. *Nature reviews. Genetics*, 6(8):597, 2005.

105. Hideharu Hashimoto, Xing Zhang, Paula M Vertino, and Xiaodong Cheng. The mechanisms of generation, recognition, and erasure of dna 5-methylcytosine and thymine oxidations. *Journal of Biological Chemistry*, 290(34):20723–20733, 2015.

106. Peter A Jones. Functions of dna methylation: islands, start sites, gene bodies and beyond. *Nature reviews. Genetics*, 13(7):484, 2012.

107. Xinsheng Nan, F Javier Campoy, and Adrian Bird. Mecp2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell*, 88(4):471–481, 1997.

108. Maria Chahrour, Sung Yun Jung, Chad Shaw, Xiaobo Zhou, Stephen TC Wong, Jun Qin, and Huda Y Zoghbi. Mecp2, a key contributor to neurological disease, activates and represses transcription. *Science*, 320(5880):1224–1229, 2008.

109. Peter A Jones. The dna methylation paradox. *Trends in Genetics*, 15(1):34–37, 1999.

110. Shaohui Hu, Jun Wan, Yijing Su, Qifeng Song, Yaxue Zeng, Ha Nam Nguyen, Jaehoon Shin, Eric Cox, Hee Sool Rho, Crystal Woodard, et al. Dna methylation presents distinct binding sites for human transcription factors. *Elife*, 2:e00726, 2013.

111. Ishminder K Mann, Raghunath Chatterjee, Jianfei Zhao, Ximiao He, Matthew T Weirauch, Timothy R Hughes, and Charles Vinson. Cg methylated microarrays identify a novel methylated sequence bound by the cebpb— atf4 heterodimer that is active in vivo. *Genome research*, 23(6):988–997, 2013.

112. Catherine A Kemme, Dan Nguyen, Abhijnan Chattopadhyay, and Junji Iwahara. Regulation of transcription factors via natural decoys in genomic dna. *Transcription*, 7(4):115–120, 2016.

113. Anat Burger, Aleksandra M Walczak, and Peter G Wolynes. Influence of decoys on the noise and dynamics of gene expression. *Physical Review E*, 86(4):041920, 2012.

114. Zhipeng Wang, Davit A Potoyan, and Peter G Wolynes. Molecular stripping, targets and decoys as modulators of oscillations in the nf-κb/iκbα/dna genetic network. *Journal of The Royal Society Interface*, 13(122):20160606, 2016.

115. Joshua I Friedman and James T Stivers. Detection of damaged dna bases by dna glycosylase enzymes. *Biochemistry*, 49(24):4957, 2010.

116. Nadrian C Seeman, John M Rosenberg, and Alexander Rich. Sequence-specific recognition of double helical nucleic acids by proteins. *Proceedings of the National Academy of Sciences*, 73(3):804–808, 1976.

117. Tatyana B Hamilton, Franck Borel, and Paul J Romaniuk. Comparison of the dna binding characteristics of the related zinc finger proteins wt1 and egr1. *Biochemistry*, 37(7):2051–2058, 1998.

118. Kendra King Frederick, Michael S Marlow, Kathleen G Valentine, and A Joshua Wand. Conformational entropy in molecular recognition by proteins: A first example from the calmodulin system. *Nature*, 448(7151):325, 2007.

119. Michael S Marlow, Jakob Dogan, Kendra K Frederick, Kathleen G Valentine, and A Joshua Wand. The role of conformational entropy in molecular recognition by calmodulin. *Nature chemical biology*, 6(5):352–358, 2010.

120. Shiou-Ru Tzeng, Charalampos G Kalodimos, et al. Protein activity regulation by conformational entropy. *Nature*, 488(7410):236–240, 2012.

121. Junji Iwahara, Young-Sang Jung, and G Marius Clore. Heteronuclear nmr spectroscopy for lysine nh3 groups in proteins: unique effect of water exchange on 15n transverse relaxation. *Journal of the American Chemical Society*, 129(10):2971–2980, 2007.

122. Frank Delaglio, Stephan Grzesiek, Geerten W Vuister, Guang Zhu, John Pfeifer, and AD Bax. Nmrpipe: a multidimensional spectral processing system based on unix pipes. *Journal of biomolecular NMR*, 6(3):277–293, 1995.

123. Bruce A Johnson and Richard A Blevins. Nmr view: A computer program for the visualization and analysis of nmr data. *Journal of biomolecular NMR*, 4(5):603–614, 1994.

124. G Marius Clore and Angela M Gronenborn. Determining the structures of large proteins and protein complexes by nmr. *Trends in biotechnology*, 16(1):22–34, 1998.

125. Junji Iwahara and G Marius Clore. Sensitivity improvement for correlations involving arginine side-chain nε/hε resonances in multi-dimensional nmr experiments using broadband 15 n 180 pulses. *Journal of biomolecular NMR*, 36(4):251–257, 2006.

126. Gerald S Manning. Limiting laws and counterion condensation in polyelectrolyte solutions ii. self-diffusion of the small ions. *The Journal of Chemical Physics*, 51(3):934–938, 1969.

127. Gerald S Manning. Limiting laws and counterion condensation in polyelectrolyte solutions. iii. an analysis based on the mayer ionic solution theory. *The Journal of Chemical Physics*, 51(8):3249–3252, 1969.

128. Gerald S Manning. Limiting laws and counterion condensation in polyelectrolyte solutions: Iv. the approach to the limit and the extraordinary stability of the charge fraction. *Biophysical chemistry*, 7(2):95–102, 1977.

129. Gerald S Manning. Limiting laws and counterion condensation in polyelectrolyte solutions i. colligative properties. *The journal of chemical Physics*, 51(3):924–933, 1969.

130. M Thomas Record, Timothy M Lohman, and Pieter De Haseth. Ion effects on ligand-nucleic acid interactions. *Journal of molecular biology*, 107(2):145–158, 1976.

131. M Thomas Record, Charles F Anderson, and Timothy M Lohman. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. *Quarterly reviews of biophysics*, 11(2):103–178, 1978.

132. Eriks Kupce, Jonathan Boyd, and Iain D Campbell. Short selective pulses for biochemical applications. *Journal of Magnetic Resonance, Series B*, 106(3):300–303, 1995.

133. Nico Tjandra, Scott E Feller, Richard W Pastor, and Ad Bax. Rotational diffusion anisotropy of human ubiquitin from 15n nmr relaxation. *Journal of the American Chemical Society*, 117(50):12562–12566, 1995.

134. DE Woessner. Nuclear spin relaxation in ellipsoids undergoing rotational brownian motion. *The Journal of Chemical Physics*, 37(3):647–654, 1962.

135. Nikola Trbovic, Jae-Hyun Cho, Robert Abel, Richard A Friesner, Mark Rance, and Arthur G Palmer III. Protein side-chain dynamics and residual conformational entropy. *Journal of the American Chemical Society*, 131(2):615, 2009.

136. Edward J d'Auvergne and Paul R Gooley. The use of model selection in the model-free analysis of protein dynamics. *Journal of biomolecular NMR*, 25(1):25–39, 2003.

137. Giovanni Lipari and Attila Szabo. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. theory and range of validity. *Journal of the American Chemical Society*, 104(17):4546–4559, 1982.

138. G Marius Clore, Attila Szabo, Ad Bax, Lewis E Kay, Paul C Driscoll, and Angela M Gronenborn. Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *Journal of the American Chemical Society*, 112(12):4989–4991, 1990.

139. Junji Iwahara, Robert D Peterson, and Robert T Clubb. Compensating increases in protein backbone flexibility occur when the dead ringer at-rich interaction domain (arid) binds dna: A nitrogen-15 relaxation study. *Protein science*, 14(5):1140–1150, 2005.

140. Andrew L Lee, Sandra A Kinnear, and A Joshua Wand. Redistribution and loss of side chain entropy upon formation of a calmodulin? peptide complex. *Nature structural biology*, 7(1), 2000.

141. Sergei Rudnizky, Hadeel Khamis, Omri Malik, Allison Squires, Amit Meller, Philippa Melamed, and Ariel Kaplan. Single-molecule dna unzipping reveals asymmetric modulation of the transcription factor egr-1 by its binding site sequence and context. *bioRxiv*, page 188458, 2017.

142. Sam Janssen, Olivier Cuvier, Martin Müller, and Ulrich K Laemmli. Specific gain-and loss-of-function phenotypes induced by satellite-specific dna-binding drugs fed to drosophila melanogaster. *Molecular cell*, 6(5):1013–1024, 2000.

143. Tony Kouzarides. Chromatin modifications and their function. *Cell*, 128(4):693–705, 2007.

144. Paul L Leong, Genevieve A Andrews, Daniel E Johnson, Kevin F Dyer, Sichuan Xi, Jeffrey C Mai, Paul D Robbins, Seshu Gadiparthi, Nancy A Burke, Simon F Watkins, et al. Targeted inhibition of stat3 with a decoy oligonucleotide abrogates head and neck cancer cell growth. *Proceedings of the National Academy of Sciences*, 100(7):4138–4143, 2003.

145. K Ohtani, K Egashira, M Usui, M Ishibashi, K-I Hiasa, Q Zhao, M Aoki, Y Kaneda, R Morishita, and A Takeshita. Inhibition of neointimal hyperplasia after balloon injury by cis-element decoyof early growth response gene-1 in hypercholesterolemic rabbits. *Gene therapy*, 11(2), 2004.

146. Ryuichi Morishita, Toshiko Sugimoto, Motokuni Aoki, Iwao Kida, Naruya Tomita, Atsushi Moriguchi, Kazuhiro Maeda, Yoshiki Sawa, Yasufumi Kaneda, Jitsuo Higaki, et al. In vivo transfection of cis element decoy

against nuclear factor-κb binding site prevents myocardial infarction. *Nature medicine*, 3(8):894–899, 1997.

147. John Goodchild. *Therapeutic oligonucleotides*. Springer, 2011.

Kemme, Catherine A.

From: US Journal Permissions <USJournalPermissions@taylorandfrancis.com>
Sent: Tuesday, September 19, 2017 9:24 AM
To: Kemme, Catherine A.
Subject: RE: ktrn20:Regulation of transcription factors via natural decoys in genomic DNA

Dear Ms. Kemme,
You are free to use your Open Access article in your dissertation with full acknowledgment of your original work.
Thank you for contacting us.

Mary Ann Muller – Permissions Coordinator, US Journals Division

My Work Schedule is Tuesday, Wednesday, and Friday.

Find digital versions of our articles on: www.tandfonline.com to use RightsLink, our online permissions web page, for immediate processing of your permission request.

Please Note: Permissions requests for US journals may take up to three weeks for processing due to demand.

Taylor & Francis Group
an informa business

530 Walnut Street – 8th floor
Philadelphia | PA | 19106 | United States of America
Direct line: 215-606-4334
Main Office: 215 625 8900, ext. 14334

maryann.muller@taylorandfrancis.com
www.taylorandfrancisgroup.com
www.tandfonline.com

From: Academic UK Non Rightslink
Sent: Monday, September 18, 2017 3:52 AM
To: US Journal Permissions <USJournalPermissions@taylorandfrancis.com>
Subject: FW: ktrn20:Regulation of transcription factors via natural decoys in genomic DNA

Hi Mary-Ann,

Here is one for you- KTRN

Best wishes,
Lisa

From: Catherine Kemme [mailto:cakemme@utmb.edu]
Sent: 15 September 2017 17:30
To: Academic UK Non Rightslink <permissionrequest@tandf.co.uk>
Subject: ktrn20:Regulation of transcription factors via natural decoys in genomic DNA

Permissions Request
Contact name: Catherine Kemme
Street address: 3205B TX-3
Town: TEXAS CITY
Postcode/ZIP code: 77591
Country: United States
Contact telephone number: 5137875348
Contact email address: cakemme@utmb.edu
Article title: Regulation of transcription factors via natural decoys in genomic DNA
Article DOI: 10.1080/21541264.2016.1188873
Author name: Catherine A. Kemme, Dan Nguyen, Abhijnan Chattopadhyay & Junji Iwahara
Journal title: Transcription
Volume number: 7
Issue number: 4
Year of publication: 2016
Page number(s): 1-6
Are you the sole author/editor of the new publication?: No
Are you requesting the full article?: Yes
If no, please supply extract and include number of word:
If no, please supply details of figure/table:
Name of publisher of new publication: Dissertation work University Of Texas Medical Branch
Title of new publication: Complexity of transcription factor target search: effect of DNA natural decoys and DNA methylation through interplay with methyl-CpG-binding proteins
Course pack: No
Number of Students:
Is print: Yes
Electronic: Yes
E-reserve: No
Period of use: Perpetual
Short loan library?: No
Thesis: Yes
To be reprinted in a new publication?: No
In print format: Yes
In ebook format?: No
ISBN:
Languages:
Distribution quantity:
Retail price:
Additional comments:

Kemme, Catherine A.

From: support@services.acs.org
Sent: Monday, September 25, 2017 10:54 AM
To: Kemme, Catherine A.
Subject: Request #9131-7777942 Resolved (Request for permissions)

Click here if you have difficulty reading this email >>

Dear Dr. Kemme,

Thank you for contacting ACS Publications Help Desk. Your permission requested is granted and there is no fee for this reuse. In your planned reuse, you must cite the ACS article as the source,

Please add this direct link http://pubs.acs.org/doi/abs/10.1021%2Facs.biochem.5b00967

for further permissions related to the material excerpted should be directed to the ACS.

Should you have any questions or need further assistance, please feel free to contact me.

Sincerely,

Deleep Kumar

ACS Customer Services & Information

https://help.acs.orgn

Your help request has been resolved. If you have further issues regarding this matter, please let us know by responding to this email. Please note that this request will auto-close in 14 days. If you need to contact us after 14 days regarding this matter please submit a new help request and refer to this help request number.

How are we doing? Let us know!
Please click on your selection below to begin our two-question survey.

Based on this support interaction, how satisfied were you with the help provided?

Not at all satisfied *Completely satisfied*

E-mail Information:

Attachments	
cc	

Request Information:

Request #	9131-7777942
Date Created	9/15/2017 12:52 PM EDT
Summary	Request for permissions
Details	Good afternoon, I am writing to request permission to reuse an entire article (of which I am first author) in my dissertation titled “Complexity of transcription factor target search: effect of DNA natural decoys and DNA methylation through interplay with methyl-CpG-binding proteins. ” Anticipated completion date for my dissertation is November 2017. The article is the following: Kemme, Catherine A., Alexandre Esadze, and Junji Iwahara. "Influence of quasi-specific sites on kinetics of target DNA search by a sequence-specific DNA-binding protein." Biochemistry 54.44 (2015): 6684-6691. http://pubs.acs.org/doi/abs/10.1021/acs.biochem.5b00967 Information regarding the requestor is the following: Catherine Kemme 301 University Blvd. Medical Research Building 5.112 University Texas Medical Branch Galveston, TX 77550 Phone number: 513 787 5348 E-mail address: mailto:cakemme@utmb.edu Thank you for your time, Catherine Kemme

To update or check the status of this request:

Click here to access this request online.

Creative Commons Legal Code

Attribution-NonCommercial 4.0 International

Official translations of this license are available in other languages.

Creative Commons Corporation (“Creative Commons”) is not a law firm and does not provide legal services or legal advice. Distribution of Creative Commons public licenses does not create a lawyer-client or other relationship. Creative Commons makes its licenses and related information available on an “as-is” basis. Creative Commons gives no warranties regarding its licenses, any material licensed under their terms and conditions, or any related information. Creative Commons disclaims all liability for damages resulting from their use to the fullest extent possible.

Using Creative Commons Public Licenses

Creative Commons public licenses provide a standard set of terms and conditions that creators and other rights holders may use to share original works of authorship and other material subject to copyright and certain other rights specified in the public license below. The following considerations are for informational purposes only, are not exhaustive, and do not form part of our licenses.

> ***Considerations for licensors:*** *Our public licenses are intended for use by those authorized to give the public permission to use material in ways otherwise restricted by copyright and certain other rights. Our licenses are irrevocable. Licensors should read and understand the terms and conditions of the license they choose before applying it. Licensors should also secure all rights necessary before applying our licenses so that the public can reuse the material as expected. Licensors should clearly mark any material not subject to the license. This includes other CC-licensed material, or material used under an exception or limitation to copyright. More considerations for licensors.*

> ***Considerations for the public:*** *By using one of our public licenses, a licensor grants the public permission to use the licensed material under specified terms and conditions. If the licensor’s permission is not necessary for any reason–for example, because of any applicable exception or limitation to copyright–then that use is not regulated by the license. Our licenses grant only permissions under copyright and certain other rights that a licensor has authority to grant. Use of the licensed material may still be restricted for other reasons, including because others have copyright or other rights in the material. A licensor may make special requests, such as asking that all changes be marked or described. Although not required by our licenses, you are encouraged to respect those requests where reasonable. More considerations for the public.*

Creative Commons Attribution-NonCommercial 4.0 International Public License

By exercising the Licensed Rights (defined below), You accept and agree to be bound by the terms and conditions of this Creative Commons Attribution-NonCommercial 4.0 International Public License ("Public License"). To the extent this Public License may be interpreted as a contract, You are granted the Licensed Rights in consideration of Your acceptance of these terms and conditions, and the Licensor grants You such rights in consideration of benefits the Licensor receives from making the Licensed Material available under these terms and conditions.

Section 1 – Definitions.

a. **Adapted Material** means material subject to Copyright and Similar Rights that is derived from or based upon the Licensed Material and in which the Licensed Material is translated, altered, arranged, transformed, or otherwise modified in a manner requiring permission under the Copyright and Similar Rights held by the Licensor. For purposes of this Public License, where the Licensed Material is a musical work, performance, or sound recording, Adapted Material is always produced where the Licensed Material is synched in timed relation with a moving image.
b. **Adapter's License** means the license You apply to Your Copyright and Similar Rights in Your contributions to Adapted Material in accordance with the terms and conditions of this Public License.
c. **Copyright and Similar Rights** means copyright and/or similar rights closely related to copyright including, without limitation, performance, broadcast, sound recording, and Sui Generis Database Rights, without regard to how the rights are labeled or categorized. For purposes of this Public License, the rights specified in Section 2(b)(1)-(2) are not Copyright and Similar Rights.
d. **Effective Technological Measures** means those measures that, in the absence of proper authority, may not be circumvented under laws fulfilling obligations under Article 11 of the WIPO Copyright Treaty adopted on December 20, 1996, and/or similar international agreements.
e. **Exceptions and Limitations** means fair use, fair dealing, and/or any other exception or limitation to Copyright and Similar Rights that applies to Your use of the Licensed Material.
f. **Licensed Material** means the artistic or literary work, database, or other material to which the Licensor applied this Public License.
g. **Licensed Rights** means the rights granted to You subject to the terms and conditions of this Public License, which are limited to all Copyright and Similar Rights that apply to Your use of the Licensed Material and that the Licensor has authority to license.
h. **Licensor** means the individual(s) or entity(ies) granting rights under this Public License.
i. **NonCommercial** means not primarily intended for or directed towards commercial advantage or monetary compensation. For purposes of this Public License, the exchange of the Licensed Material for other material subject to Copyright and Similar Rights by digital file-sharing or similar means is NonCommercial provided there is no payment of monetary compensation in connection with the exchange.
j. **Share** means to provide material to the public by any means or process that requires permission under the Licensed Rights, such as reproduction, public display, public performance, distribution, dissemination, communication, or importation, and to make material available to the public including in ways that members of the public may access the material from a place and at a time individually chosen by them.
k. **Sui Generis Database Rights** means rights other than copyright resulting from Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, as amended and/or succeeded, as well as other essentially equivalent rights anywhere in the world.
l. **You** means the individual or entity exercising the Licensed Rights under this Public License. **Your** has a corresponding meaning.

Section 2 – Scope.

a. **License grant**.
    1. Subject to the terms and conditions of this Public License, the Licensor hereby grants You a worldwide, royalty-free, non-sublicensable, non-exclusive, irrevocable license to exercise the Licensed Rights in the Licensed Material to:
        A. reproduce and Share the Licensed Material, in whole or in part, for NonCommercial purposes only; and
        B. produce, reproduce, and Share Adapted Material for NonCommercial purposes only.
    2. Exceptions and Limitations. For the avoidance of doubt, where Exceptions and Limitations apply to Your use, this Public License does not apply, and You do not need to comply with its terms and conditions.
    3. Term. The term of this Public License is specified in Section 6(a).
    4. Media and formats; technical modifications allowed. The Licensor authorizes You to exercise the Licensed Rights in all media and formats whether now known or hereafter created, and to make technical modifications necessary to do so. The Licensor waives and/or agrees not to assert any right or authority to forbid You from making technical modifications necessary to exercise the Licensed Rights, including technical modifications

necessary to circumvent Effective Technological Measures. For purposes of this Public License, simply making modifications authorized by this Section 2(a)(4) never produces Adapted Material.

5. Downstream recipients.
   A. Offer from the Licensor – Licensed Material. Every recipient of the Licensed Material automatically receives an offer from the Licensor to exercise the Licensed Rights under the terms and conditions of this Public License.
   B. No downstream restrictions. You may not offer or impose any additional or different terms or conditions on, or apply any Effective Technological Measures to, the Licensed Material if doing so restricts exercise of the Licensed Rights by any recipient of the Licensed Material.
6. No endorsement. Nothing in this Public License constitutes or may be construed as permission to assert or imply that You are, or that Your use of the Licensed Material is, connected with, or sponsored, endorsed, or granted official status by, the Licensor or others designated to receive attribution as provided in Section 3(a)(1)(A)(i).

b. **Other rights**.

1. Moral rights, such as the right of integrity, are not licensed under this Public License, nor are publicity, privacy, and/or other similar personality rights; however, to the extent possible, the Licensor waives and/or agrees not to assert any such rights held by the Licensor to the limited extent necessary to allow You to exercise the Licensed Rights, but not otherwise.
2. Patent and trademark rights are not licensed under this Public License.
3. To the extent possible, the Licensor waives any right to collect royalties from You for the exercise of the Licensed Rights, whether directly or through a collecting society under any voluntary or waivable statutory or compulsory licensing scheme. In all other cases the Licensor expressly reserves any right to collect such royalties, including when the Licensed Material is used other than for NonCommercial purposes.

Section 3 – License Conditions.

Your exercise of the Licensed Rights is expressly made subject to the following conditions.

a. **Attribution**.

1. If You Share the Licensed Material (including in modified form), You must:
   A. retain the following if it is supplied by the Licensor with the Licensed Material:
      i. identification of the creator(s) of the Licensed Material and any others designated to receive attribution, in any reasonable manner requested by the Licensor (including by pseudonym if designated);
      ii. a copyright notice;
      iii. a notice that refers to this Public License;
      iv. a notice that refers to the disclaimer of warranties;
      v. a URI or hyperlink to the Licensed Material to the extent reasonably practicable;
   B. indicate if You modified the Licensed Material and retain an indication of any previous modifications; and
   C. indicate the Licensed Material is licensed under this Public License, and include the text of, or the URI or hyperlink to, this Public License.
2. You may satisfy the conditions in Section 3(a)(1) in any reasonable manner based on the medium, means, and context in which You Share the Licensed Material. For example, it may be reasonable to satisfy the conditions by providing a URI or hyperlink to a resource that includes the required information.
3. If requested by the Licensor, You must remove any of the information required by Section 3(a)(1)(A) to the extent reasonably practicable.
4. If You Share Adapted Material You produce, the Adapter's License You apply must not prevent recipients of the Adapted Material from complying with this Public License.

Section 4 – Sui Generis Database Rights.

Where the Licensed Rights include Sui Generis Database Rights that apply to Your use of the Licensed Material:

a. for the avoidance of doubt, Section 2(a)(1) grants You the right to extract, reuse, reproduce, and Share all or a substantial portion of the contents of the database for NonCommercial purposes only;
b. if You include all or a substantial portion of the database contents in a database in which You have Sui Generis Database Rights, then the database in which You have Sui Generis Database Rights (but not its individual contents) is Adapted Material; and
c. You must comply with the conditions in Section 3(a) if You Share all or a substantial portion of the contents of the database.

For the avoidance of doubt, this Section 4 supplements and does not replace Your obligations under this Public License where the Licensed Rights include other Copyright and Similar Rights.

Section 5 – Disclaimer of Warranties and Limitation of Liability.

a. Unless otherwise separately undertaken by the Licensor, to the extent possible, the Licensor offers the Licensed Material as-is and as-available, and makes no representations or warranties of any kind concerning the Licensed Material, whether express, implied, statutory, or other. This includes, without limitation, warranties of title, merchantability, fitness for a particular purpose, non-infringement, absence of latent or other defects, accuracy, or the presence or absence of errors, whether or not known or discoverable. Where disclaimers of warranties are not allowed in full or in part, this disclaimer may not apply to You.

b. To the extent possible, in no event will the Licensor be liable to You on any legal theory (including, without limitation, negligence) or otherwise for any direct, special, indirect, incidental, consequential, punitive, exemplary, or other losses, costs, expenses, or damages arising out of this Public License or use of the Licensed Material, even if the Licensor has been advised of the possibility of such losses, costs, expenses, or damages. Where a limitation of liability is not allowed in full or in part, this limitation may not apply to You.

c. The disclaimer of warranties and limitation of liability provided above shall be interpreted in a manner that, to the extent possible, most closely approximates an absolute disclaimer and waiver of all liability.

Section 6 – Term and Termination.

a. This Public License applies for the term of the Copyright and Similar Rights licensed here. However, if You fail to comply with this Public License, then Your rights under this Public License terminate automatically.

b. Where Your right to use the Licensed Material has terminated under Section 6(a), it reinstates:

    1. automatically as of the date the violation is cured, provided it is cured within 30 days of Your discovery of the violation; or
    2. upon express reinstatement by the Licensor.

    For the avoidance of doubt, this Section 6(b) does not affect any right the Licensor may have to seek remedies for Your violations of this Public License.
c. For the avoidance of doubt, the Licensor may also offer the Licensed Material under separate terms or conditions or stop distributing the Licensed Material at any time; however, doing so will not terminate this Public License.
d. Sections 1, 5, 6, 7, and 8 survive termination of this Public License.

Section 7 – Other Terms and Conditions.

a. The Licensor shall not be bound by any additional or different terms or conditions communicated by You unless expressly agreed.
b. Any arrangements, understandings, or agreements regarding the Licensed Material not stated herein are separate from and independent of the terms and conditions of this Public License.

Section 8 – Interpretation.

a. For the avoidance of doubt, this Public License does not, and shall not be interpreted to, reduce, limit, restrict, or impose conditions on any use of the Licensed Material that could lawfully be made without permission under this Public License.
b. To the extent possible, if any provision of this Public License is deemed unenforceable, it shall be automatically reformed to the minimum extent necessary to make it enforceable. If the provision cannot be reformed, it shall be severed from this Public License without affecting the enforceability of the remaining terms and conditions.
c. No term or condition of this Public License will be waived and no failure to comply consented to unless expressly agreed to by the Licensor.
d. Nothing in this Public License constitutes or may be interpreted as a limitation upon, or waiver of, any privileges and immunities that apply to the Licensor or You, including from the legal processes of any jurisdiction or authority.

Creative Commons is not a party to its public licenses. Notwithstanding, Creative Commons may elect to apply one of its public licenses to material it publishes and in those instances will be considered the “Licensor.” The text of the Creative Commons public licenses is dedicated to the public domain under the CC0 Public Domain Dedication. Except for the limited purpose of indicating that material is shared under a Creative Commons public license or as otherwise permitted by the Creative Commons policies published at creativecommons.org/policies, Creative Commons does not authorize the use of the trademark “Creative Commons” or any other trademark or logo of Creative Commons without its prior written consent including, without limitation, in connection with any unauthorized modifications to any of its public licenses or any other arrangements, understandings, or agreements concerning use of licensed material. For the avoidance of doubt, this paragraph does not form part of the public licenses.

Creative Commons may be contacted at creativecommons.org.

Additional languages available: Bahasa Indonesia, Deutsch, français, hrvatski, italiano, Nederlands, norsk, polski, suomeksi, svenska, te reo Māori, Türkçe, українська, العربية, 日本語. Please read the FAQ for more information about official translations.

Vita

Catherine Kemme was born on June 30th, 1990 to Paul and Mary Ann Kemme in Cincinnati, Ohio. Catherine earned a Bachelor of Science (May 2013) in Natural Sciences with a concentration in biology from Pensacola Christian College in Pensacola, Florida. After her undergraduate degree, she entered the graduate program in Biochemistry and Molecular Biology at the University of Texas Medical Branch in August 2013. In 2014, she began her thesis work under the supervision of Dr. Junji Iwahara in the Biochemistry and Molecular Biology Department studying the kinetics and thermodynamics of protein-DNA interactions. During her stay at UTMB, Catherine received the following awards: Bohdan Nechay Endowment, Barbara Bowman Memorial Award for Research Excellence, and Michael Tacheeni Scott Endowed Scholarship. In addition to her research responsibilities, she served on the Biochemistry and Molecular Biology's curriculum committee, and served as the vice president for the biological chemistry student organization in 2015.

Permanent address: 7399 Walnut Creek Dr.
West Chester, Ohio 45069

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.

EDUCATION

B.S., May 2013, Pensacola Christian College, Pensacola, Florida

PUBLICATIONS

A. Articles in Peer-Reviewed Journals

Esadze, Alexandre, **Catherine A. Kemme**, Anatoly B. Kolomeisky, and Junji Iwahara. "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 Research 42, no. 11 (2014): 7039-7046.

Zandarashvili, Levani, Alexandre Esadze, Dana Vuzman, **Catherine A. Kemme**, Yaakov Levy, and Junji Iwahara. "Balancing between affinity and speed in target DNA search by zinc-finger proteins via modulation of dynamic conformational ensemble." Proceedings of the National Academy of Sciences 112, no. 37 (2015): E5142-E5149.

Esadze, Alexandre, **Catherine A. Kemme**, Anatoly B. Kolomeisky, and Junji Iwahara. "Kinetic Mechanisms of Target Location by Inducible Transcription Factor Egr-1." Biophysical Journal 108, no. 2 (2015): 76a.

Kemme, Catherine A., Alexandre Esadze, and Junji Iwahara. "Influence of quasi-specific sites on kinetics of target DNA search by a sequence-specific DNA-binding protein." Biochemistry 54, no. 44 (2015): 6684.

Zandarashvili, Levani, Alexandre Esadze, **Catherine A. Kemme**, Abhijnan Chattopadhyay, Dan Nguyen, and Junji Iwahara. "Residence times of molecular complexes in solution from NMR data of intermolecular hydrogen-bond scalar coupling." The Journal of Physical Chemistry Letters 7, no. 5 (2016): 820.

Kemme, Catherine A., Dan Nguyen, Abhijnan Chattopadhyay, and Junji Iwahara. "Regulation of transcription factors via natural decoys in genomic DNA." Transcription 7, no. 4 (2016): 115-120.

Kemme, Catherine A., Rolando Marquez, Ross H. Luu, and Junji Iwahara. "Potential role of DNA methylation as a facilitator of target search processes for transcription factors through interplay with methyl-CpG-binding proteins." Nucleic Acids Research (2017).

B. Abstracts in Proceedings of Conferences/Symposia:

Catherine Kemme, Sandy Sass, Dawn Verleye, Henry Vu, and John G. Duman, "HeLa cell cryopreservation with antifreeze proteins 1 and 2 from Dendroides Canadensis," Emerging Researchers National Conference (ERN) in STEM, February, 2011, Atlanta, GA.

Catherine Kemme, Alexandre Esadze, Anatoly Kolomeisky, and Junji Iwahara, "Next Step in elucidation of rapid gene activation by Egr-1," 19th Annual Sealy Structural Biology Symposium, 2013, University of Texas Medical Branch, Galveston, TX.

Catherine Kemme, Alexandre Esadze, Anatoly Kolomeisky, and Junji Iwahara, "Kinetic impact of semi-specific DNA sites during target location by a sequence-specific DNA-binding protein," 20th Annual Sealy Structural Biology Symposium, 2014, University of Texas Medical Branch, Galveston, TX

Catherine Kemme, Alexandre Esadze, Anatoly Kolomeisky, and Junji Iwahara. "Quantitative Experimental Analysis of the Influence of Quasi-Specific Sites on Kinetics of DNA Scanning by the Zinc-Finger Protein Egr-1." Biophysical Journal 110.3 (2016): 237a., Biophysical Society Meeting, Los Angeles, CA.

Catherine Kemme, Rolando Marquez, Ross H. Luu, and Junji Iwahara. "Potential role of methylation as a facilitator of target search processes for transcription factors through interplay with methyl CpG binding proteins." 21st Annual Sealy Structural Biology Symposium, 2017, University of Texas Medical Branch, Galveston, TX.