Using single molecule FRET to study the mechanisms of DNA bending by TBP and HMGB1

by

REBECCA BLAIR

B.S., Bucknell University, Lewisburg, PA 2005

A Thesis Submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Chemistry and Biochemistry

2013

This thesis entitled:

Using single molecule FRET to study the mechanisms of DNA bending by TBP and HMGB1

Written by:

Rebecca Blair

has been approved for the Department of Chemistry and Biochemistry

James A. Goodrich

Jennifer F. Kugel

December 16, 2013

The final copy of this thesis has been examined by the
Signatories, and we find that both the content and the form meet
Acceptable presentation standards of scholarly work in the
Above mentioned discipline

ABSTRACT

Blair, Rebecca (Ph.D., Biochemistry)

Using single molecule FRET to study the mechanisms of DNA bending by TBP and HMGB1

Thesis directed by Professors James A. Goodrich and Jennifer F. Kugel

Defining mechanisms of transcriptional regulation is important for understanding how gene expression is controlled, which is essential to cellular viability. Outlined in this thesis are studies that characterize DNA-protein interactions involved in transcriptional regulation. Specifically, we have used single-molecule FRET (smFRET) to investigate how proteins bend DNA. A homebuilt TIRF microscope was assembled for single molecule fluorescence studies, which is described in Chapter 2 of this thesis.

We used smFRET to study the extent and kinetics of DNA bending by the transcription factor TBP (TATA binding protein) with consensus and nonconsensus TATA DNA. TBP bent different TATA sequences to the same, homogenous, bent population. Further, TFIIA did not change the extent of DNA bending by TBP, but increased the stability of the TBP-DNA complex. We found that TBP bent consensus TATA DNA to two different kinetic populations bent to the same extent, but only one kinetic bent population existed when TBP was bound to the nonconsensus sequence. The uniform bending of DNA by TBP was not predicted by previous ensemble studies and provided insight into the mechanism of DNA binding by TBP.

HMGB1 is a nuclear protein that binds and bends DNA to facilitate transcription. It has 3 domains: the A and B box, which bend DNA independently, and an acidic unstructured C-terminal tail. We studied how full length HMGB1 and its different domains bend DNA using smFRET. We determined that the full length HMGB1 protein bent DNA to a similar extent as its individual domains, the A box and the B box. However, removal of the C-terminal tail caused the protein to bend DNA to a greater extent and with more heterogeneity than the full length protein. Further, a truncation that contains the B box and C-terminal tail interacted with DNA less efficiently than the B box alone; this truncation could bend DNA at a high concentration to the same extent as the B box alone. Taken together we propose that the full length HMGB1 protein bends DNA primarily through its A box, while the B box interacts with the C-terminal tail, which prevents its interaction with DNA.

ACKNOWLEDGEMENTS

I would like to thank the people who have encouraged me throughout my scientific career. Primarily, Dr. Jim Goodrich and Dr. Jen Kugel, who have advised me in graduate school. They have been instrumental in teaching me scientific skills, scientific communication, and project management. I would like to thank my committee, Dr. Amy Palmer, Dr. Thomas Perkins, Dr. Arthur Pardi, and Dr. Robert Kuchta. They have offered great advice and direction on my projects. I would like to thank Dr. Kathy Rowlen, Dr. Laura Kuck, and Dr. Erica Dawson who introduced me to the exciting field of biotechnology and mentored me in my first job after college. I would like to thank my high school science teacher, Doc, who encouraged me to study biochemistry in college. The NIH funded my project through the Mostly Molecular Biology Training Grant. I would like to thank those members of the lab who have guided me in experiments, general scientific knowledge, how to use the pH meter (Petro), and fun happy hours, ski trips, and hikes. Thanks Ryan, Stacey, Ben, Petro, Nicole, Tereza, Linda, Jess, Steve, and McSwiggles. To all those people who helped me stay sane through my first year graduate school and beyond - Michelle, Kevin, Katarina, Ryan, and Virginia.

My amazing friends and family have supported me during graduate school and in life. I can count on them for being there for the amazing moments in life and for the difficult ones. I am so lucky to know all of these fantastic people and I am glad to surround myself with such great people. I would specifically like to thank my parents, grandparents, brothers, sisters, brother-in-laws, sister-in-laws, parent-in-laws, nieces, and nephews. They are always there for a laugh, a cry, and most definitely, a good bottle of wine and some stinky cheese. Finally, I would like to thank my amazing “boyfriend” of 10 years. I couldn’t have done it without you.

TABLE OF CONTENTS

The regulation of mammalian mRNA transcription

Building an objective-based total internal reflection fluorescence (TIRF) microscope

Single molecule FRET shows uniformity in TBP-induced DNA bending and heterogeneity in bending kinetics

The bending of linear DNA by full length HMGB1 is mediated through one HMG domain

Summary and future directions

TABLES

FIGURES

CHAPTER 1

The regulation of mammalian mRNA transcription

Activation and transcription of protein encoding genes

The mammalian eukaryotic genome is transcribed by 3 RNA polymerases. RNA polymerase I (Pol I) synthesizes the noncoding ribosomal RNA precursors (rRNA) (1), and RNA polymerase III (Pol III) transcribes one rRNA, transfer RNAs (tRNAs), and other noncoding RNAs, including SINEs and U6 RNA (2). RNA polymerase II (Pol II) transcribes messenger RNA (mRNA) for all protein-encoding genes and noncoding RNAs, for example, microRNAs (3, 4). Transcription and its regulation in mammalian cells is extremely complex. For protein-encoding genes transcribed by Pol II, there are many DNA elements, proteins, and RNAs that function together to transcribe each specific gene.

A general overview of activator-dependent transcription is that protein activators bind to enhancer regions on DNA far from the transcription start site, which recruits co-activators and chromatin remodelers. The co-activators can interact with general transcription factors (GTFs) that are part of the basal transcriptional machinery and Pol II to form the pre-initiation complex (PIC) at the core promoter that ultimately initiates transcription at that specific gene (Figure 1.1) (5). All of these stages in the transcriptional process are regulated, both positively and negatively, to inhibit or activate transcription at specific genes. Accurate transcriptional regulation is essential to healthy cellular metabolism because unregulated or misregulated transcription can lead to disease. The first section of this chapter will describe in detail the stages of transcription before initiation.

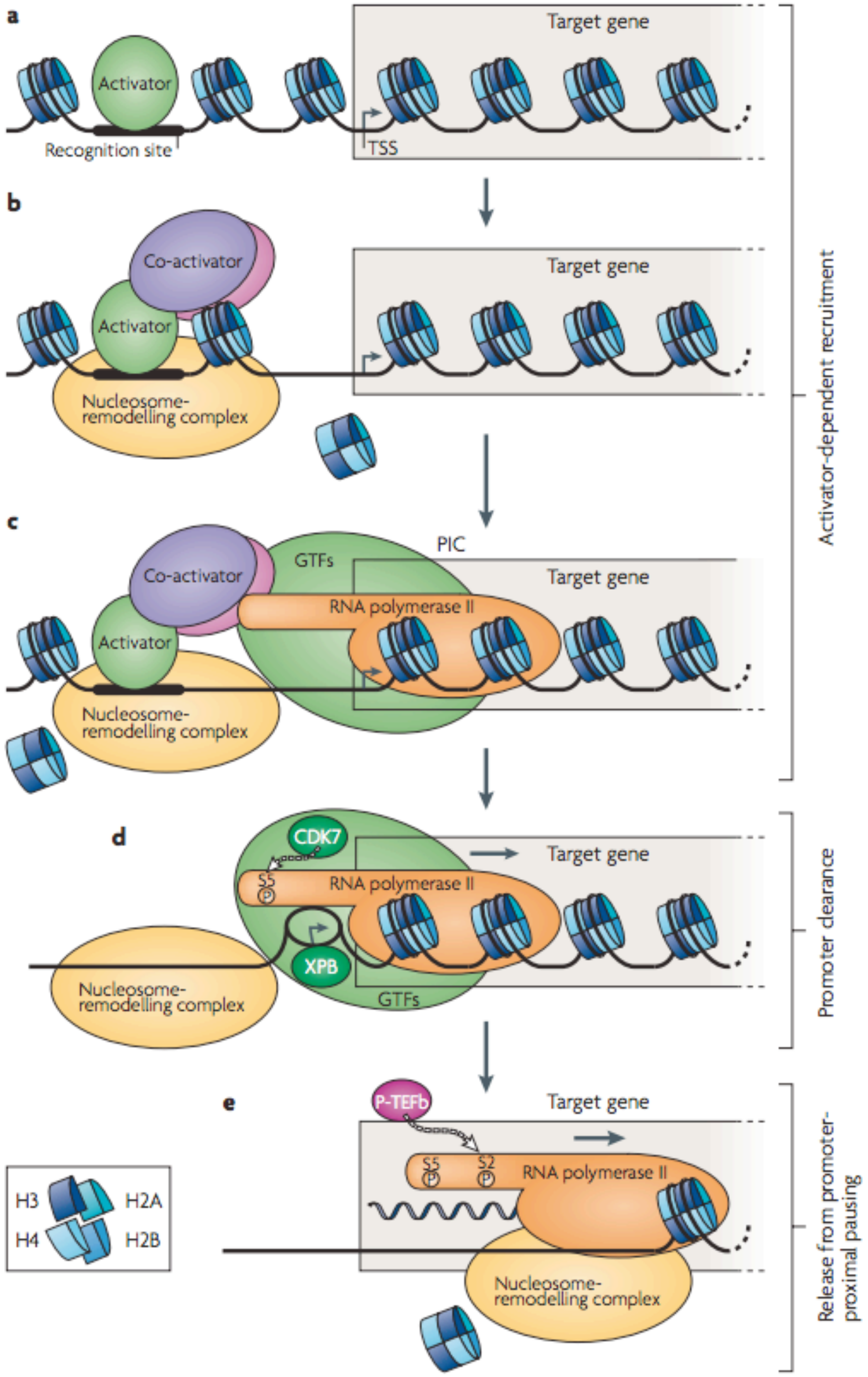


Figure 1.1. Activator-induced transcription. **(A)** An activator binds to its recognition sequence. **(B)** The activator recruits co-activators and nucleosome remodeling complexes. **(C)** Histones are ejected from the DNA and/or modified, and general transcription factors and Pol II can be recruited and bind the promoter. **(D)** Pol II is phosphorylated and DNA is melted to create an open complex. **(E)** Transcription of that specific gene occurs. The Figure is from (5).

Promoter DNA elements

Promoter DNA contains many elements to facilitate transcription start site (TSS) selection and activation of a specific gene (Figure 1.2). The TSS is surrounded by core promoter elements, which can include the TATA box and the initiator sequence, among others, that are 40 bp upstream and downstream from the promoter start site. These elements are important for GTF recognition of the promoter and to recruit Pol II to the gene (4). The proximal promoter elements, which are a couple hundred base pairs upstream of the core promoter, are where activators bind (6). Enhancer elements are cis-acting elements that can be kilobases away from the TSS. Activators that bind the enhancer can interact with specific promoter-bound proteins or can interact with the core promoter factors through intermediary proteins (6). Similar to enhancers, silencers are cis-acting DNA elements which repressors bind to mediate inhibition of transcription through condensation of the chromatin or otherwise blocking activation from occurring. Boundary elements, such as insulators, ensure that a specific gene is activated by a specific enhancer and protect a gene from being activated by enhancers from neighboring genes (7). All of these DNA elements are important for start site selection and to ensure that each gene is activated at the correct time. DNA elements can also be modified through methylation or other modifications to control protein binding to the DNA (8).

Chromatin dynamics

The DNA in eukaryotic genomes is packaged into chromatin, a specifically arranged nucleoprotein complex with histone proteins at the core. Histones are ubiquitous proteins that interact with DNA and allow for DNA condensation and gene silencing or DNA relaxation and

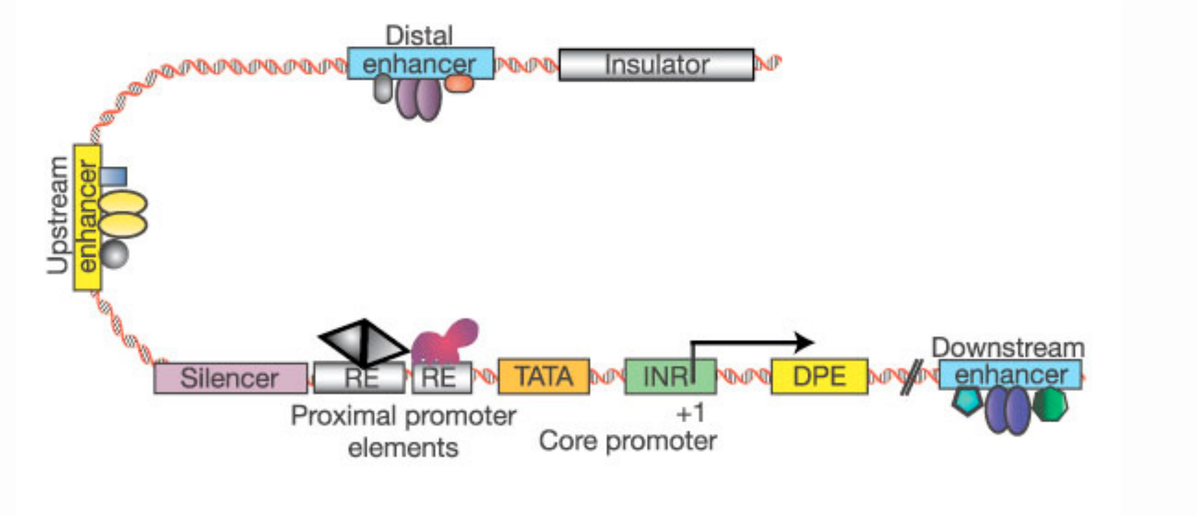


Figure 1.2. A model of a mammalian mRNA promoter. DPE is the downstream promoter element, INR is the initiator sequence, TATA is the TATA Box, and RE are regulatory elements. This Figure is from (25).

gene activation. A nucleosome is composed of an octamer of histone proteins with 2 copies each of 4 histone proteins, H2A, H2B, H3, H4, and 147 base pairs of DNA wrapped around the histone octamer (9). The nucleosome is capped with histone H1, and there are approximately 29-43 bp of DNA between nucleosomes, forming a structure similar to beads on a string. Nucleosome arrays are further packaged into higher ordered structures in the nucleus of cells.

DNA packaging generally inactivates genes by blocking DNA binding sites for transcription factors and transcription initiation by Pol II (10). Hence, nucleosomes have to be dynamic for transcription to occur. Some activators can bind nucleosomal DNA and recruit two different types of enzymes known to alter chromatin structure, histone modifying enzymes and chromatin remodeling enzymes (11). Histone modifying enzymes make post-translational modifications to the N-terminal tails of histones, which signal whether the chromatin will condense, open up, or the histones will be removed or slid from that piece of DNA for gene activation. Also, other proteins are known to recognize specific histone modifications, which can further recruit factors for gene activation or inhibition. Chromatin remodeling enzymes are ATPases that change histone-DNA contacts to facilitate histone sliding and disassembly (10, 12). These signals for nucleosome alteration and chromatin structural changes are important for activation or repression of a gene. Specifically, for complete PIC formation that includes Pol II at the promoter region, nucleosomes have to be removed, but they can be present in partially formed PIC complexes (13). Pol II also has to transcribe through nucleosomes, which can occur by histones being removed downstream of the transcription complex and replaced upstream of the complex, which is facilitated by chromatin remodeling factors (10).

Activators, repressors, and co-regulatory proteins

Activators are proteins that bind to enhancer elements or promoter proximal regulatory regions upstream from the TSS. These are sequence-specific sites that activators recognize and are important for downstream signaling and recruitment (5). Activators can be regulated through cellular localization, being associated with a subunit that inhibits activator activity, and/or different post translational modifications, among other ways. For example, NF-κB is bound to an inhibitor in the cytosol that is released upon upstream signaling, which allows NF-κB to move into the nucleus and bind specific DNA sequences (14). Upon activator binding to DNA, co-activators and chromatin remodelers are recruited to the upstream enhancer site. Activators can also directly contact GTFs, for example, specific subunits of transcription factor IID (TFIID) are known to interact with activators such as p53, Sp1, and c-Jun (15), which facilitates the formation of the PIC. Repressors can bind silencer elements in promoters to repress transcription of specific genes. Inhibition can occur through condensing chromatin by recruiting co-repressors, such as histone deacetylases or DNA methylases (16, 17), or inactivating the PIC complex through blocking other factors from binding or removing important GTFs from the promoter (18).

Activators often work through co-activators. One such co-activator is mediator, which cannot bind DNA, but instead binds specifically to the unphosphorylated C-terminal domain (CTD) of Pol II and TFIID (19). Mediator is also known to interact with the activation domains of activators, linking the activators to the PIC. Mediator can also be involved in activator-independent transcription and can stimulate basal transcription in vitro through faster recruitment of transcription factor IIB (TFIIB) (19). GCN5/PCAF is another co-activator that has chromatin modifying activity. Once the histones are acetylated by GCN5/PCAF, CBP recognizes enhanceosome and Pol II and SWI/SNF are recruited to the promoter (20). In general, many proteins need to be coordinated for activation-dependent transcription to occur.

PIC formation

PIC formation is pivotal for transcription of protein-encoding genes. Formation of PICs is nucleated by TFIID, which contains TATA box binding protein (TBP) and TBP-associated factors (TAFs). TBP recognizes and bends TATA box DNA, an A/T-rich sequence upstream from the transcription start site. Several of the TAF components of TFIID bind to other sequence-specific DNA elements in the promoter region, including the initiator region, the downstream promoter element, and the downstream core element (Figures 1.2, 1.3) (4, 21). Since not all promoters contain a TATA box, these elements help stabilize TFIID on TATA-less promoters.

TFIIB and transcription factor IIA (TFIIA) are recruited to the gene following TFIID (Figure 1.3). TFIIB recognizes the BRE elements in the DNA that flank the TATA box (4). Both TFIIA and TFIIB increase the stability of TBP bound to the promoter (22). TFIIB recruits Pol II and transcription factor IIF (TFIIF). Both TFIIB and TFIIF aid in TSS selection, but TFIIF increases the affinity of Pol II for the other factors at the promoter and recruits transcription factors IIE and IIH (TFIIE, TFIIH) (4). TFIIE and TFIIH complete formation of the core PIC. TFIIF, TFIIE, and TFIIH are all important in the transition from initiation to promoter escape. Specifically, TFIIH has helicase and ATPase activities which melt the DNA, and a kinase activity which phosphorylates Pol II to signal transcription initiation.

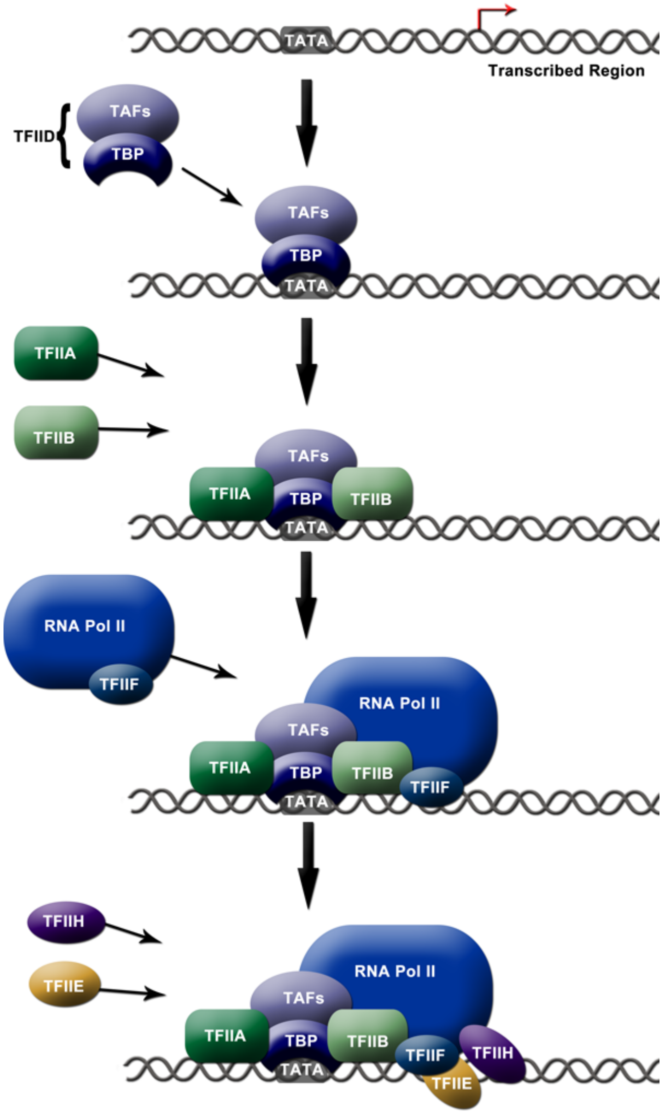


Figure 1.3. PIC formation. The sequential model of PIC formation is outlined above. The promoter contains a TATA box. TBP and its associated factors (TAFs) are recruited to the TATA box and bind DNA. This recruits TFIIA and TFIIB, followed by RNA Pol II and TFIIF. Once these transcription factors are recruited, TFIIE and TFIIH bind to the promoter, which can modify Pol II, melt the promoter, and stimulate open complex formation for mRNA transcription to occur.

There are two different models for how it is thought that the PIC forms. One is the sequential model and the other is the holoenzyme model. The traditional sequential model is discussed above, where one GTF recruits the next factor and so on. There is evidence that this model may not occur at all genes in the exact order given above. The holoenzyme model is based on the observation that there are complexes containing Pol II and GTFs not bound to promoter DNA that can be purified from cells. These complexes have different components but are generally comprised of Pol II, TFIIF, TFIIE, TFIIH, and chromatin remodeling and mRNA processing enzymes (4, 21). These proteins are recruited to genes as a large complex. Both models of PIC formation could occur in cells (23, 24).

TATA binding protein (TBP)

As mentioned above, TBP is an important factor in PIC formation, which is required for transcription initiation. TBP binds to the TATA box, which is an A/T-rich 8-bp sequence about 25-bp upstream from the transcription start site. TBP binds to the consensus sequence, TATA(A/T)AA(G/A), with high affinity (4). About 30% of genes have a TATA box, and those without TATA boxes generally fall into the housekeeping genes, oncogene/growth factor and transcription factors categories (4). TBP recognizes this sequence through the inherent flexibility of the DNA. When TBP binds DNA, it simultaneously bends the DNA through intercalating phenylalanines between basepairs 1 and 2, and basepairs 7 and 8 in the minor groove of DNA, creating “2 kinks” within the bent structure (26, 27). This bends the DNA and slightly unwinds the DNA, which can be seen in the crystal structure of TBP bound to DNA (Figure 1.4A). This sharp bend is one signal for recruiting the GTFs to the promoter region.

Altering the TATA box sequence affects the binding affinity (26, 28-30), where the binding affinity and kinetic stability decrease as the DNA sequence moves away from consensus. Further, the bend angle of a nonconsensus TATA DNA sequence bound by TBP has been observed to be lower compared to the consensus sequence (26, 28, 29, 31, 32). Changing the TATA sequence has also been shown to affect the transcriptional activity of a specific gene in vitro and in yeast cells, with the consensus sequence having the highest transcriptional activity (26, 33). All of these data indicate that the TATA sequence is biologically important because the bend angle, binding affinity, and transcriptional activity depend on the sequence.

As mentioned above, other transcription factors, such as TFIIB and TFIIA, affect the DNA-TBP complex. TFIIA is known to stabilize the DNA-TBP interaction on both consensus

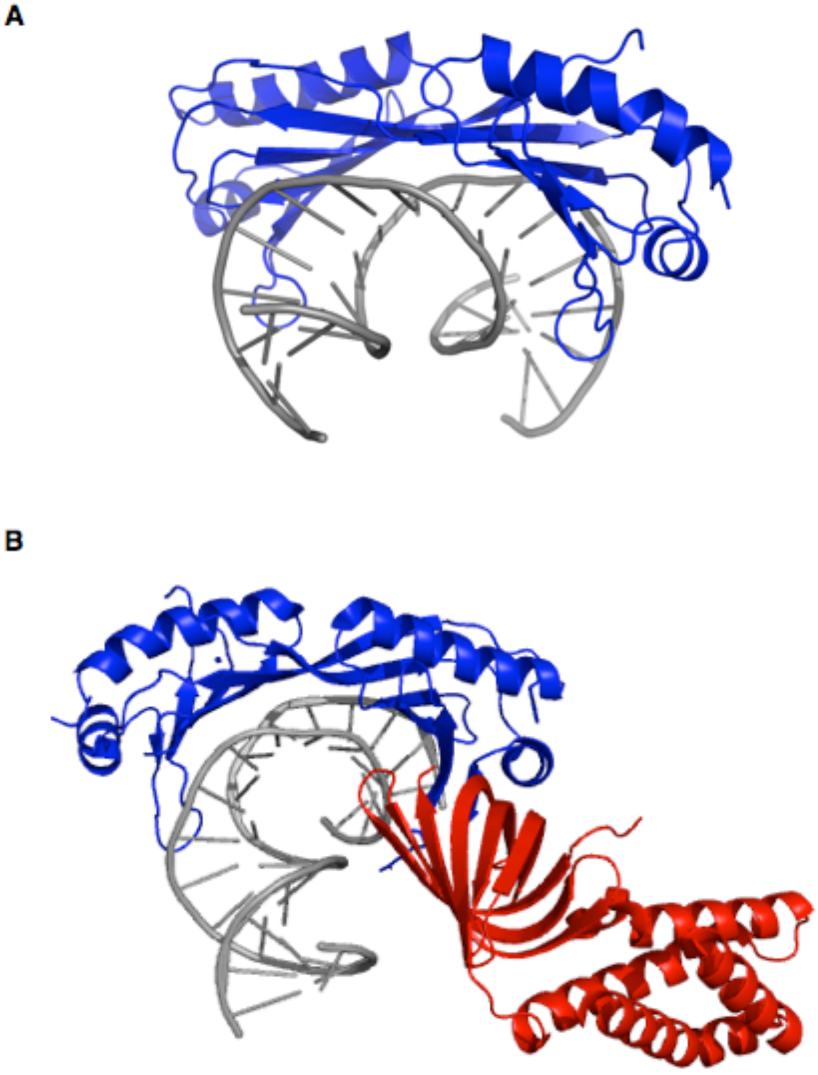


Figure 1.4. TBP bends TATA DNA. **(A)** TBP (blue) interacting and bending double stranded DNA (gray). **(B)** TBP and TFIIA (red) complexed on dsDNA. These images are from the crystal structures uploaded in PDB (37-39).

and nonconsensus TATA boxes (22, 29, 34-36). The crystal structure of TFIIA-TBP-DNA indicates that the DNA is still bent when TFIIA joins the TBP-DNA complex (Figure 1.4B). The extent to which TBP bends DNA in the presence of TFIIA is debatable, when comparing solution measurements to the crystal structure (29, 37).

In Chapter 3, I study TBP bending promoter DNA sequences in vitro using single molecule fluorescence studies. Specifically, I examine the extent to which TBP bends sequences that have different binding affinities and further investigate the effect of TFIIA on TBP-induced DNA bending.

HMG proteins

High mobility group (HMG) proteins are small, DNA-binding proteins mainly found in the nucleus of cells (40). They are involved in transcriptional regulation through altering transcription factor interactions with DNA and are also involved in chromatin rearrangement and regulation. Overexpression of HMG proteins has been found to cause tumors and is found in most cancers (40). There are different classes of HMG proteins. HMGA proteins contain AT hooks that bind A/T-rich sequences, HMGB proteins contain HMG boxes, which bind and bend DNA, and HMGN proteins bind nucleosomes in a sequence-independent manner (41).

Generally, all families of HMG proteins are involved in the activation of transcription. Specifically, HMGN proteins alter chromatin structure by changing protein-protein interactions that would otherwise lead to DNA condensation. The presence of HMGNs are also known to alter post-translational modifications of histones, but modifications to an HMGN protein itself, reduces its affinity for nucleosomes, which in turn allows for DNA condensation (40, 42).

HMGA proteins are found in embryos and tumors, and bind in the minor groove of DNA at specific sequences, bending the DNA slightly (40, 43). They are known to bind to enhancers, which facilitates the binding of other transcription factors. HMGA can positively and negatively regulate gene expression because acetylation of HMGA causes disassembly of proteins from enhancers and phosphorylation reduces its affinity for DNA (40).

There are different types of HMGB proteins, but each HMG box is approximately 80 aa and binds in the minor groove to bend the DNA (44, 45). HMGB proteins can bind DNA with sequence specificity, such as SRY or LEF, or nonspecifically, such as HMGB1-4 (40, 45). HMGB proteins are known to increase transcription of genes through chromatin activation and facilitating transcription factor binding, but they also function in the DNA repair pathways and when secreted can act as a cytokine to signal for inflammation (45).

HMGB1

Introduction

HMGB1 is 80% homologous to HMGB2-3. All three of these proteins are found in vertebrates (45), but their tissue-specific expression varies; HMGB1 is expressed in almost all cell types, and HMGB2 and HMGB3 are expressed only in the testis and lymphoids and during embryogenesis (40). HMGB1 is an essential protein since the knockout is lethal in mice after birth due to misregulation of the glucocorticoid receptor (41, 46). HMGB1 is also found to be upregulated in most cancers (47). HMGB1 is found mostly in the nucleus of cells, and is the second most abundant nuclear protein after histones (41). The protein has homologs that are found in vertebrates, drosophila, and yeast (48). HMGB1 has loose association with DNA within the cell, cycling between bound and unbound (49), and it interacts with many factors within the nucleus. HMGB1 has many functions within the cell, which are centered around it interacting with and bending DNA to signal other proteins to bind DNA.

Biological functions of HMGB1

Along with binding DNA directly, HMGB1 uses multiple mechanisms to facilitate opening chromatin, thus stimulating transcription. First, HMGB1 can bind to linker DNA between the nucleosomes. When HMGB1 binds, it bends the DNA, and loosens the DNA that is wrapped around the nucleosomes (40). Second, HMGB1 interacts with the tail of histone H3 in nucleosomes, which decreases the contacts between H3 and DNA, thus allowing HMGB1 to bind (50, 51). Lastly, histone H1 interacts with the C-terminal tail on HMGB1, allowing HMGB1 to bind DNA. Histone H1 and HMGB1 both bind linker DNA, leading to conflicting roles in nucleosome structure and chromatin stability, where histone H1 causes chromatin condensation and HMGB1 loosens chromatin. The dynamics between histone H1 and HMGB1 is transient and causes local nucleosome regulation (52, 53). Through these different mechanisms, HMGB1 can interact with chromatin and it can also bind and bend linker DNA, increasing the transcriptional activity of a particular gene.

HMGB1 is also an architectural protein that can facilitate DNA binding by sequence-specific transcription factors. The mechanism is unclear, but HMGB1 could prebend the DNA before the transcription factor binds or form a complex with the transcription factor and bring it to the DNA (Figure 1.5) (48). HMGB1 increases the affinity with which some proteins bind DNA, including transcription factors and DNA repair proteins, by interacting with the specific protein (45). The list of HMGB1 interacting factors is considerable, but includes OCT1, p53, SREBP-1, TBP, and NF-κB (45). One transcription factor, p53, has been shown to interact unstably with HMGB1. It is hypothesized that the interaction occurs through one of the HMG boxes in HMGB1, while the second HMG box binds to the DNA, creating a ternary complex (54). Although HMGB1 is known to interact with proteins to increase their affinity for binding DNA, very few ternary complexes (transcription factor/HMGB1/DNA) have been visualized (45), indicating that the ternary interactions are very transient.

HMGB1 is also involved in DNA repair. It has high affinity for damaged and distorted DNA, which signals to proteins in different repair pathways (55). Specifically, HMGB1 can identify and bind to damage that disrupts the helical structure of DNA. The high affinity of HMGB1 for these structures may be the first step of DNA damage recognition in the nucleotide excision repair pathway (55). Likewise, HMGB1 has been implicated in recognizing DNA

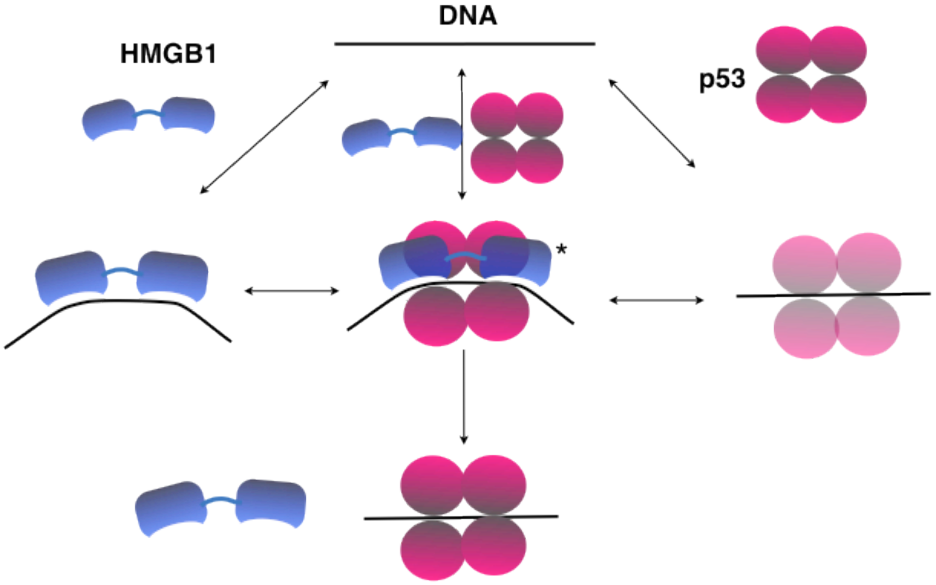


Figure 1.5. Models for how HMGB1 could facilitate transcription factor binding, using p53 as an example. The left most pathway depicts HMGB1 prebending DNA prior to the p53 binding to the DNA. The central pathway indicates HMGB1 and p53 interacting off of DNA before binding DNA. The right pathway illustrates p53 unstably binding to DNA first, but HMGB1 coming in and stabilizing the p53-DNA interaction. In all cases, HMGB1 can then release from the complex, which is consistent with not observing stable ternary complexes.

damage in both the mismatch repair and base excision repair pathways, to the extent that it has been shown to immunoprecipitate with other proteins in these pathways (55). Further, HMGB1 can bind to double stranded DNA breaks, which activates the repair pathway by placing the ends close together so that ligation is more efficient (55).

Beyond nuclear function, HMGB1 can also function as a cytokine and be secreted into the extracellular space (47). HMGB1 is secreted by macrophages when the inflammation pathway is activated. Acetylation of multiple lysines on HMGB1 causes the protein to be shuttled from the nucleus to the cytoplasm. HMGB1 can then be loaded into secretory vesicle and released into the extracellular space; it can also passively diffuse from the cell. HMGB1 binding to RAGE or TLR receptors activates signaling pathways, which ultimately results in transcription of inflammatory genes. Alternatively, HMGB1 can be involved in the wound-healing response, but its downstream effect is reliant on other cytokines in its environment to determine which pathway is activated. Since HMGB1 has an important role in overall inflammation and disease (including sepsis, arthritis, infection, and cancer) (47), studying the role of HMGB1 in the immune response is currently an active area of research.

HMGB1 structure and mechanism of binding

HMGB1 contains 2 separate and independent DNA binding domains, an A box and a B box, and it has a C-terminal acidic tail (Figure 1.6) (40, 45). It binds DNA without sequence specificity in the minor groove and bends the DNA through insertions of hydrophobic amino acids into the DNA (44, 45). HMGB1 has a higher affinity for distorted DNA, such as 4-way

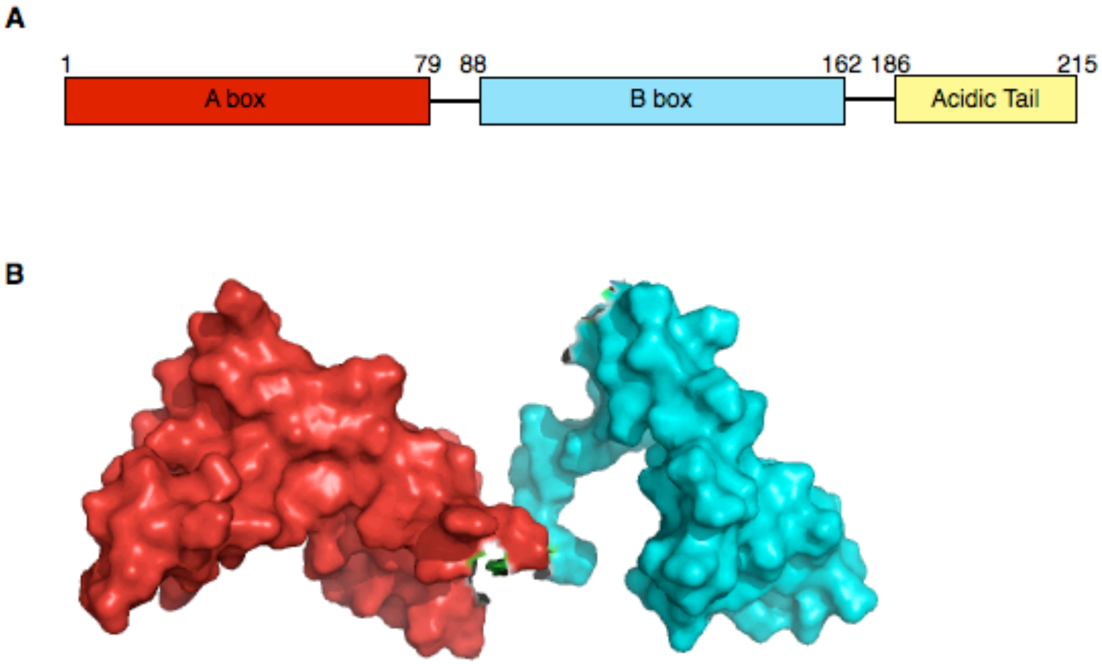


Figure 1.6. HMGB1 domain structure. **(A)** A schematic of human HMGB1 where the A box is at the N-terminus of the protein, the B box is a central domain and the acidic tail is on the C-terminus of HMGB1. Adapted from (47). **(B)** The structure of the A and B boxes of HMGB1, where the DNA binding domains are at the bottom. Data from PDB 2YRQ.

junctions and cisplatin-modified DNA, than linear DNA, but it bends all DNA (45). This bending can decrease the persistence length of double stranded DNA (56).

Post translational modifications are one important mechanism for regulation of HMGB1 activity. Specifically, Lys2 and Lys11 are known to become acetylated. In vitro experiments have shown that acetylation by CBP decreases the affinity of the protein for linear DNA by abolishing lysine-phosphate contacts, but increases A box and full length HMGB1 affinity for distorted DNA (57, 58). Hyper acetylation, phosphorylation, and methylation all regulate the cellular location of HMGB1 by either signaling the protein to leave the nucleus or move back into the nucleus (45).

DNA binding by HMGB1 is extremely dynamic in cells; it is thought to only stay bound to DNA for short periods of time (less than 10 s), which was discovered through FRAP and FLIP studies (49). Biochemical studies have measured the affinity of the HMGB1/DNA interaction to be anywhere from the low nanomolar to micromolar range (56, 59-63), indicating that there is wide variation regarding how this protein interacts with DNA both in cells and in vitro. Due to its many important biological functions, it is important to understand the underlying mechanism of how HMGB1 binds and bends DNA, which is the focus of Chapter 4.

Fluorescence measurements and their benefits for understanding biology

Properties of fluorophores

Fluorescence occurs when an electron in a molecule absorbs energy and is excited to a higher electronic state and then drops down to the ground state while releasing a photon. It needs to occur on the timescale of less than ~10 ns or the excited state energy will be lost as nonradiative decay, which could include thermal exchange into the environment (64). A Jablonski diagram illustrates this phenomenon (Figure 1.7); when an electron is excited it goes from the ground electronic state (S_0) to a high vibrational state in an excited electronic state (S_1). Internal conversion or vibrational relaxation occurs immediately, and the electron drops to the lowest vibrational state in the S_1 electronic state (64, 65). When the electron drops back down to the ground state, the photon released is a lower energy species than that of the excited state due to the internal conversion. This phenomenon can partially account for the wavelength shift in the peak excitation and emission spectra that occurs during fluorescence. In other words, a molecule is excited at a higher energy wavelength and emission occurs at a lower energy wavelength, which is referred to as the Stokes shift.

Upon excitation, fluorophores can also go into a triplet or dark state (Figure 1.7) through intersystem crossing. This occurs when the electron spin switches from up to down or down to up after being excited; the electron stays in the excited state and a photon is not released immediately (64). The conversion out of the dark state can take on the order of milliseconds to seconds and is referred to as photoblinking (66).

The photophysical properties of a fluorophore are important in determining which fluorophore works best for a given experimental situation. First, the quantum yield (ϕ_F) of a

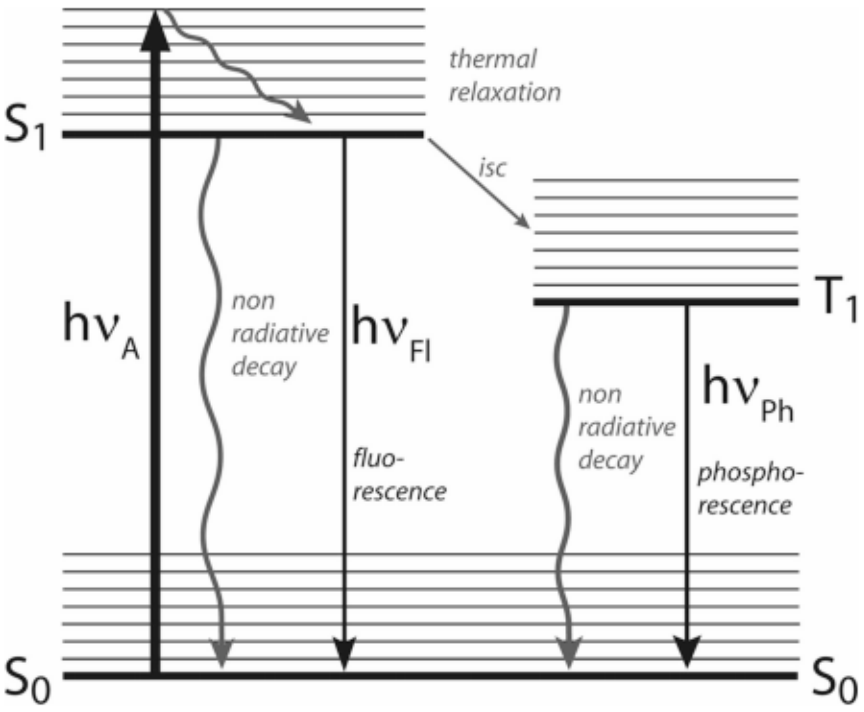


Figure 1.7. A Jablonski Diagram illustrating light absorption and emission by electrons.
T is the triplet state, S are the electronic states, and the energy states within the electronic states are vibrational states (grey lines). Absorption occurs when the molecule is excited by a photon of light ($h\nu_a$). Internal conversion (thermal relaxation) drops the electron to the lowest vibrational state in the higher electronic state. Emission can occur when the electron drops to the ground state. The shift in wavelength between excitation and emission for a given fluorophore is partially due to the differential energy of absorption and dropping to the ground state because of internal conversion ($h\nu_a$ compared to $h\nu_f$). Intersystem crossing (ISC) occurs when the spin of the electron is flipped, and the molecule goes into the triplet or dark state. This figure is from (76)

fluorophore indicates how efficiently it emits a photon after it is absorbed. The quantum yield is related to the rate of decay for the excited state.

$$\phi_F = k_{rad}/(k_{rad}+k_{nr}) = \tau_F/\tau_{rad}$$

$$\tau_F = (k_{rad}+k_{nr})^{-1}$$

$$\tau_{rad} = 1/k_{rad}$$

where k_{rad} is the fluorescence, or radiative decay rate, k_{nr} is the nonradiative decay rate, such as triplet state conversion, τ_F is the fluorescence lifetime (i.e. the inverse sum of the decay rates), and τ_{rad} is the radiative lifetime (65). A weak fluorophore has a small quantum yield, and only a fraction of the photons absorbed are emitted because energy is lost due to nonradiative decay processes.

The fluorescence lifetime is another important property of a dye that describes the amount of time spent in the excited state before going to the ground state. A molecule that returns to the ground state rather quickly, in less than a few of nanoseconds, can reabsorb and reemit photons immediately, leading to a higher emission rate. The maximum emission rate of a fluorophore is equal to the inverse of τ_{rad} (65).

The extinction coefficient is another property used to describe a dye and refers to how well a fluorophore or molecule absorbs light. Finally, the photostability or maximum number of photons emitted from a fluorophore before irreversible destruction is an important parameter as well, which is about 10^6 photons for most fluorophores. In general, a high quantum yield, a short fluorescence lifetime without dark state transitions, and high photostability are important characteristics of a good fluorophore.

Different types of fluorophores

Fluorophores are usually planar molecules that have a high resonance and the excitation wavelength is based on the length of delocalized π electrons in the molecule, where increased conjugation allows the energy to be absorbed at higher wavelengths (66). Some small organic fluorophores, like Alexa and Cy dyes, are fairly bright and photostable compared to other small dyes (67). These dyes are easily solubilized in water and can be conjugated to proteins or nucleic acids fairly easily (67). Genetically encodable fluorescent proteins are another class of fluorophores. They are known for their beta barrel structure, but contain a planar structure derived from 3 amino acids inside the barrel that allows the proteins to fluoresce. These fluorophores are useful for intracellular imaging and easy for tagging proteins of interest, however, they can have a smaller quantum yield and photobleach faster than fluorescent dyes and the spectra for absorption and emission are fairly broad (68). Other forms of fluorophores are long lifetime probes and quantum dots. Quantum dots have a very high extinction coefficient, a narrow excitation peak, and do not photobleach quickly, but they have a tendency to blink a lot, are fairly large, and are cytotoxic (69, 70). Overall, there are many options for choosing a fluorophore, depending on the application.

Fluorescence resonance energy transfer (FRET)

Fluorescence resonance energy transfer (FRET) is a useful tool for studying biological molecules because a large change in FRET can be observed for distance changes in the low nanometer range, which occur frequently in biology. FRET is a dipole-dipole interaction between two fluorophores, a donor and an acceptor, that spectrally overlap. After the donor dye is excited, the energy is transferred to the acceptor dye and light is given off at the acceptor emission wavelength (Figure 1.8A). The amount of energy transfer is inversely related to distance over a range of 2 to 10 nm; when the donor and acceptor are closer, there is more energy transfer (Figure 1.8B). The Förster radius, R_0, is equal to the distance at which the FRET efficiency is 0.5 and is unique to each dye pair (64). Many factors influence the Förster radius, which is illustrated in the equation:

$$R_0^6 = \frac{9000\ (\ln 10)\ \kappa^2 Q_D}{128\pi^5 N n^4} \int_0^\infty F_D(\lambda)\varepsilon_A(\lambda)\ \lambda^4\ d\lambda$$

where the $F_D(\lambda)$ is the corrected fluorescence intensity of the donor at the wavelength which it emits, n is the refractive index of aqueous solution (1.4) , $\varepsilon_A(\lambda)$ is the extinction coefficient of the acceptor at a certain wavelength, Q_D is the quantum yield of the donor, N is Avogadro's number, and κ^2 is the relative orientation factor of the dipoles of the donor and acceptor dyes. The orientation factor is hard to determine and is assumed to be 2/3 when both dyes are freely rotating in solution (64), however, this is not always the case (71). After taking into account all of these factors, the FRET efficiency (E_{FRET}) is described by the following equation:

$$E_{FRET} = R_0^6/(R_0^6 + r^6)$$

where r is the distance between the donor and acceptor dyes. The FRET efficiency between two dyes can be experimentally determined using the following equation:

$$E_{FRET} = A/(A+D)$$

where A is the acceptor emission signal after excitation at the donor wavelength, and D is the donor emission signal (64). The FRET efficiency may need to be corrected due to the effects of

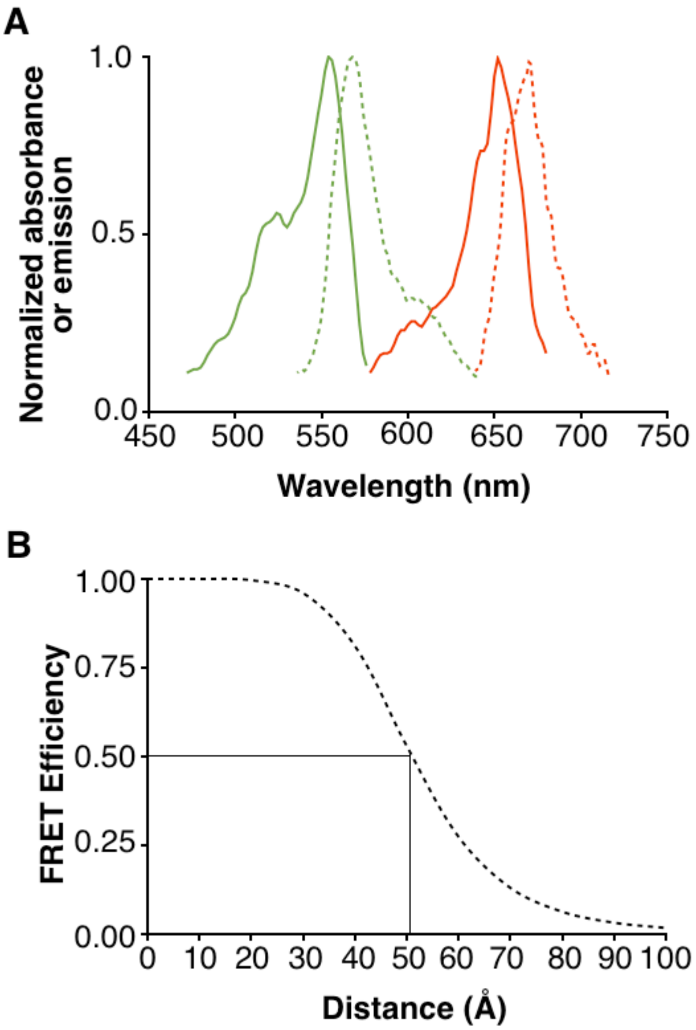


Figure 1.8. Spectral overlap of fluorophores **(A)** Spectral overlap of Alexa 555 and Alexa 647. The excitation spectra (solid) and emission spectra (dashed) of Alexa 555 (green) and Alexa 647 (red) are shown. Peak excitation of Alexa 555 is 555 nm and peak emission is 565 nm. Peak excitation of Alexa 647 is 650 nm and peak emission is 670 nm (77). The overlap between the Alexa 555 emission spectra and the Alexa 647 excitation spectra allows for energy transfer to occur between the two molecules. **(B)** FRET efficiency is plotted versus distance for the Alexa 555 and Alexa 647 FRET pair, where the R_0 is 51 Å.

other factors, including the influence of other molecules in solution, such as a protein, and instrument properties such as donor emission bleedthrough onto the acceptor channel.

Single molecule FRET studies

Using FRET to study a system in bulk can be very informative, but some bulk techniques have issues with averaging, where there may be two populations in solution, but the results only show the average of the two. The data from single molecule FRET studies can be histogrammed, which can reveal the presence of distinct subpopulations with different FRET values. Moreover, the histograms can be fit with Gaussians to provide quantitative information about the shape and range of the distributions (65). For example, in some ensemble FRET experiments that monitor a binding interaction, all molecules in solution may need to be saturated or synchronized to make useful measurements, so lower affinity interactions and less specific or heterogeneous reactions cannot be studied easily. In single molecule FRET all molecules in the solution do not have to be saturated or synchronized because the dynamics of individual molecules can be monitored. Also, transient and otherwise unknown intermediates can often be extracted from single molecule FRET data (72). For these reasons, we decided to probe two different protein-induced DNA bending systems using single-molecule FRET.

Photoblinking and dark state conversions are a problem in single molecule studies. In bulk studies several molecules going into the dark state for a second will be averaged out of the FRET data. However, in single molecule studies, single fluorophores are monitored, so blinking can strongly affect the data. For this reason, anti-fading chemicals are often added to single molecule solutions to reduce blinking and increase the stability of the dyes. Since oxygen in solution causes faster photobleaching of the fluorophore, an oxygen scavenging system is used to reduce the level of oxidation of the dyes (73, 74). The GODCAT system uses two enzymes, glucose oxidase and catalase (Figure 1.9A), and glucose to reduce the oxygen in the solution (73). The other system uses protocatechuic acid (PCA) and protocatechuate-3,4-dioxygenase (PCD) to decrease the concentration of oxygen in solution and are also used to increase fluorophore stability (Figure 1.9B) (73). Trolox is an antioxidant that is used to reduce dark state conversions, as indicated by decreased blinking (Figure 1.10) (75). All of these solution additives together aid in improved fluorescence through reducing photobleaching and photoblinking.

The laser intensity is also very important for increasing the signal output of the fluorophore. The laser intensity must be a high enough that the signal from the dye can be distinguished from background, but the laser cannot be so high that the molecule absorption rate is saturated, because then the fluorescence output is not optimal (65). A laser power that is too high also causes increased fluorophore blinking (74).

In Chapter 2, I discuss assembling a TIRF microscope, including all components we used and how we put the microscope together. In Chapter 3 and 4, I describe studies that use single molecule FRET and TIRF microscopy to investigate bending of DNA by two different proteins.

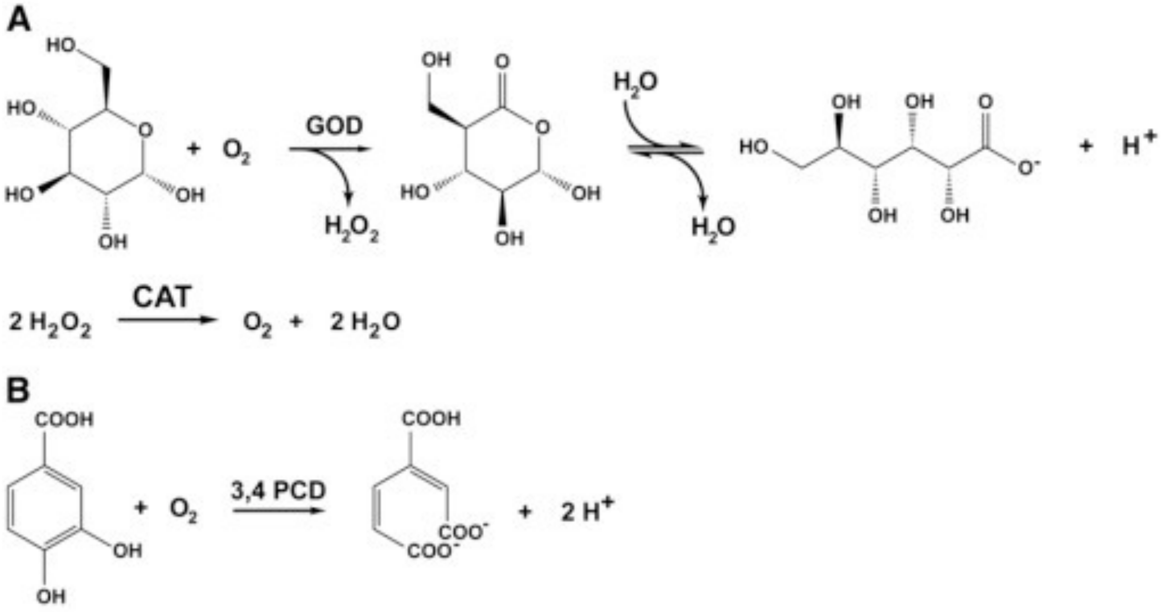


Figure 1.9. Oxygen scavenging systems regularly used in single molecule experiments. **(A)** Glucose oxidase can convert glucose and water into hydrogen peroxide then catalase converts 2 hydrogen peroxides into oxygen and water, reducing the overall amount of oxygen in solution. **(B)** The PCA/PCD system. PCD converts oxygen and PCA into a product and 2 protons (73).

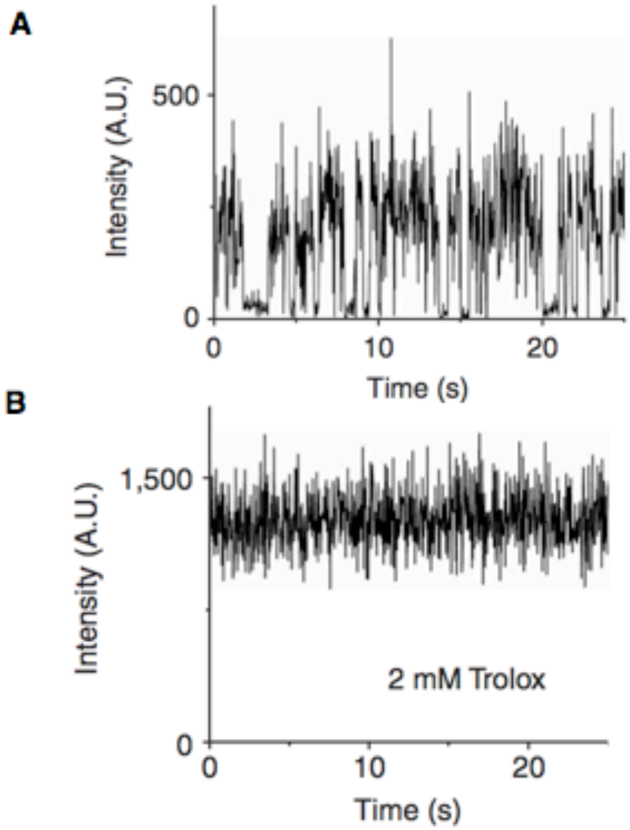


Figure 1.10. The effect of trolox on dye blinking. **(A)** A single Cy5 dye, excited in the absence of Trolox. **(B)** Cy5 fluorescence in the presence of trolox. This figure is from (75).

CHAPTER 2

Building an objective-based total internal reflection fluorescence (TIRF) microscope

Summary

An objective-based TIRF microscope was originally built by former members of the Pardi Lab, Vasiliy Fomenko and Haemi Lee, and housed in the Cristol Chemistry building. When the Biochemistry Department moved to the Jennie Smoly Caruthers Biotechnology Building, the microscope had to be disassembled and reassembled. The Goodrich, Pardi, and Falke laboratories were involved in rebuilding this instrument. We altered many components of this microscope system. On the excitation side of the microscope, we: (1) changed the lens sets to expand the sizes of the two laser beams to obtain a larger illumination area on our sample and more even excitation, (2) removed the fiber optic cable, and (3) altered the TIRF lens and the manner in which we get the lasers beams into TIRF; now we translate the TIRF lens as opposed to the mirror. On the emission side of the microscope, we: (1) added a second camera to increase the data output during every run, (2) rerouted the green emission channel to the new camera, (3) altered the magnification of the sample onto the camera, and (4) ensured the emission beam was not being cut off by other components in the dark box. These changes resulted in a 8-fold increase in data collection per frame. The improved microscope was used for the majority of the smFRET experiments in Chapter 4. The original microscope in the Cristol building was used for the experiments in Chapter 3.

Introduction

Properties of TIRF

Epifluorescence is a traditional fluorescence technique that excites all molecules in solution by shining a laser beam through a sample (Figure 2.1A). Total internal reflection fluorescence (TIRF) is an alternative technique that allows fluorescent molecules near the surface of a cover slip to be observed in the absence of unwanted fluorescence background from the solution (Figure 2.1B). TIRF occurs when a laser excites a sample at an angle greater than the "critical angle" (described below), which causes the laser beam to be reflected back through the objective and an evanescent wave to go through the sample at the same wavelength as the incoming laser light. For this to occur, the laser beam has to travel first through a surface with a higher refractive index (e.g. glass) then the sample (e.g. water) (78). This reduces the number of fluorophores excited in solution because the electromagnetic wave excites only molecules close to the surface instead of the entire solution (Figure 2.1C). The critical angle of the laser beam can be determined by Snell's law (in radians):

$$\theta_c=\sin^{-1}(n_w/n_g) \qquad \text{Equation 2.1}$$

where $n_g>n_w$, n_w is the refractive index of the sample, which is normally water (1.3), n_g is the refractive index of glass (1.5), and θ_c is the critical angle as compared to a completely perpendicular beam, which is defined as hitting the surface of the cover slip with an angle of 0° (64, 78). For water based solutions, the critical angle, θ_c , is approximately 60°.

The depth that the evanescent wave penetrates the sample is dependent on the incident angle of the laser beam. This penetrating distance decreases exponentially over the glass surface. The maximum distance into the solution is defined by the following equation:

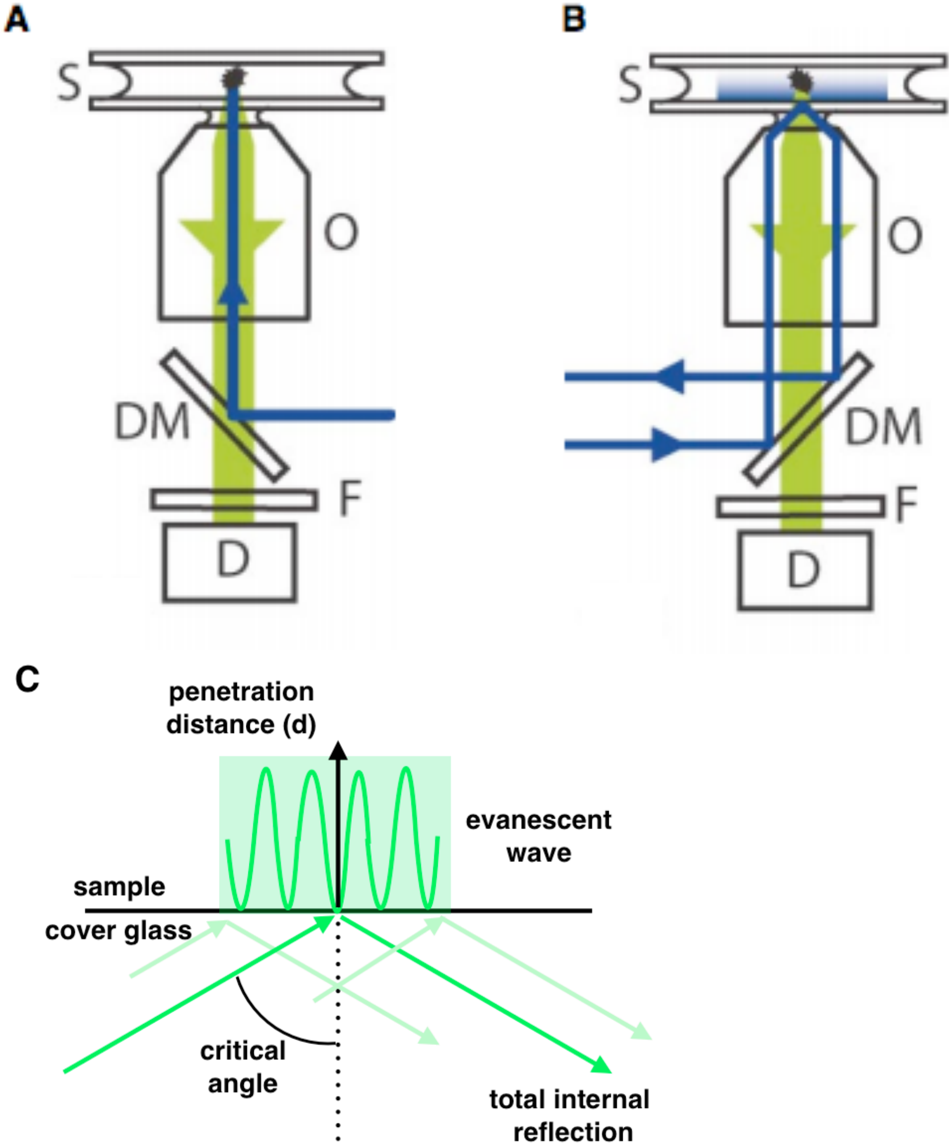


Figure 2.1. Epifluorescence versus TIRF illumination. The laser beam (blue line) enters the sample (S), and the emitted light (green line) exits through the objective (O) and hits the detector (D). **(A)** An epifluorescence system, where the light is reflected into the center of the objective and emitted light is collected through the same path. **(B)** An objective-based TIRF setup, where the laser beam enters the microscope and the light is reflected off a dichroic mirror (DM) onto the edge of an objective. Figure is from (81). **(C)** A diagram depicting the critical angle for TIRF to occur.

$$d = \lambda/4\pi \, (n_g^2 \sin^2 \theta_I - n_w^2)^{-1/2} \qquad \text{Equation 2.2}$$

where d is distance, λ is the wavelength of the incident light, and θ_I is the angle of the incident TIRF beam (79). For an incident angle 62° (just above the critical angle), the penetration distance will be 168 nm for 532 nm light and 202 nm for 640 nm light. If a higher incident angle of 80° is used, the distance the wave goes into the sample will be 60 nm for 532 nm light and 72 nm for 640 nm light. Hence, increasing the angle reduces the penetration depth of the light into the sample, which limits the excitation of fluorophores in the bulk solution above the surface.

An objective lens with a high numerical aperture is needed for TIRF because the edge of the lens in the objective causes the light to bend above the critical angle (80). To obtain TIRF, the numerical aperture (NA) of the objective is determined by:

$$NA = n_g \sin(\theta_I) \qquad \text{Equation 2.3}$$

For a TIRF beam just above the critical angle (62°), the numerical aperture has to be at least 1.33 to bend and collect the light (78). The numerical aperture of the objective is also important for resolving two fluorophores that are close together. The Rayleigh diffraction limit of 2 spots is defined by the following equation:

$$R = 0.61\lambda / NA \qquad \text{Equation 2.4}$$

where R is the resolution of the objective lens, and NA is the numerical aperature. In our system, the resolution is 218 nm for the 532 nm laser and 262 nm for the 640 nm laser for Rayleigh resolution. Hence, two fluorophores have to be about 200 to 250 nm apart to resolve them as two distinct spots. Importantly, the objective lens only collects light at the focal point; molecules emitting above and below the focal point are not collected, which reduces fluorescent background and noise from the solution.

Background sources of fluorescence

To ensure that single fluorescent molecules can be observed, several sources of background from the solution are important to consider. First, Raman scattering increases the emission background in a sample. This occurs because the excitation light is scattered by the sample medium (water), which causes the initial wavelength to be altered such that light appears at a higher wavelength, thereby increasing the apparent fluorescence emission of the solution (78). There are ways to reduce Raman scattering, such as reducing the amount of medium that is excited, but this effect cannot be fully removed from the emission spectra. Raman scattering appears in the emission channel at two different wavelengths, the excitation wavelength and a longer wavelength closer to the emission of a given fluorophore, both of which are based on the wavelength of the excited light. The following equation can be used to determine the higher wavelength emission from Raman scattering based on the excitation wavelength.

$$1/\lambda_R = 1/\lambda_E - \delta_W \qquad \text{Equation 2.5}$$

Where a property of water, δ_W, is 0.00033 nm^{-1}, λ_R and λ_E are the wavelengths of Raman scattering and excitation, respectively (67, 78). Another source of background from the solution is Rayleigh scattering of the laser light, which occurs when scattered excitation light reaches the detector (65, 78). One important way to reduce such scattering is to reduce the number of molecules that are excited, which is achieved in TIRF by only exciting molecules near the surface.

Optical Components

Filters are important for many aspects of the optical system used for TIRF microscopy. Emission and excitation filters block unwanted wavelengths of light and reduce the types of scattering described above. To ensure that the laser light excites in a narrow range of wavelengths, 10 nm bandpass filters are used before the objective. To reduce unwanted wavelengths when collecting emission data, bandpass and longpass filters are used to block the excitation light from the lasers and background emissions outside the fluorophore range. Dichroics are important for separating or combining two different wavelengths of light. Neutral density filters are used in the laser excitation line to systematically reduce the laser power coming into the microscope.

Lenses are another important component of the optics system that are used to expand and collimate the laser beam. The size of the laser beam exiting the light source can be described as a point of light. For useful measurements at the single molecule level, the beam needs to be expanded to excite a larger area of the sample. The beam expansion through two lenses can be determined through the following equation:

$$D_{out} = D_{in} * (f_2/f_1) \qquad \text{Equation 2.6}$$

Where D_{out} is the diameter of the beam after the lens set, D_{in} is the diameter of the beam before the lens set, and f_1 and f_2, are the focal lengths of the first and second lenses in the lens set, respectively. To expand the beam, the smaller focal length lens should be placed first in the beam path and the larger focal length lens second. Next, the laser beam should be collimated exiting the lens set. To be collimated, all light after the lens set should be exiting parallel, so that the beam should not change in size over an infinite distance. The approximate distance the lenses should be separated for collimation is described by:

$$d = f_1 + f_2 \qquad \text{Equation 2.7}$$

Ideally, all wavelengths of light should tolerate the lens sets similarly, however, chromatic aberrations in the lenses can cause two different wavelengths to exit the same lens set differently. Achromatic lens coatings can reduce this effect. Finally, as the beam is expanded through the lens sets, the collimated beam is brighter in the center than at the edge. The laser beam intensity has a Gaussian shape, where the highest intensity is in the center of the beam. To decrease the effect this has on sample illumination, the edge of the lenses cut off the expanded beam so that only the center portion of the beam illuminates the sample, resulting in a relatively even intensity across the sample.

Results (System Design)

Overview of the optical system and improvements

We designed an inverted objective-based TIRF microscope system that has two lasers, a 532 nm green laser and a 640 nm red laser (Table 2.1). The schematic in Figure 2.2 shows the layout of the components on the optical table. The laser beams each go through two beam expander sets, followed by a TIRF lens, which focuses the laser beam on the back focal plane of the objective lens. The objective lens has a high numerical aperture of 1.49, so it can collect emission data from an incident excitation TIRF angle of 61° to 83° (Equation 2.3). After the molecules are excited at the surface of the cover slip, the emission light travels down the objective lens, through a dichroic and a tube lens into the dark box (Figure 2.3). The dark box contains filters to remove light with extraneous wavelengths that could reduce the signal to noise of the output data. It also contains lenses that allow the emission light in each of the two channels to be focused onto a camera chip (Figure 2.4). The microscope was designed to accommodate multiple types of experiments, including FRET, 2-color co-localization studies, and particle tracking. We also planned for future modifications, including adding a third laser, setting up alternating laser excitation (ALEX), and adding a piezo stage for autofocusing.

Upon reassembly, the optical system upstream of the objective lens was changed in multiple ways to improve performance. We increased the size of the laser beams to illuminate a larger area of the sample and to get a more even laser intensity across the image. This increased the amount of data collected per image, compared to the previous microscope setup. We also removed the fiber optic cable during the reassembly process. The microscope was originally built with only a green laser and we retrofitted a red laser to it in the old building. Upon rebuilding the

	Excitation Side of the Microscope	
	Green Channel	
Lasers		532 nm laser diode, 50 mW
Neutral Density Filters	Thorlab absorptive	OD 2.0, 1.0, 0.5, 03
Lens set 1	Lens #1	30 mm
	Lens #2	150 mm
Laser Line Filter		532/10 nm bandpass
	Red Channel	
Lasers	Dragon Laser	640/10 nm laser diode, 150 mW
Neutral Density Filters	Thorlab absorptive	OD 1.5, 1.0, 0.5
Lens set 1	Lens #1	50 mm
	Lens #2	200 mm
Laser Line Filter		640/10 nm bandpass
	Both channels	
Dichroic 1	Green transmission, red reflection	
Lens set 2	Lens #1	50 mm
	Lens #2	150 mm
TIRF Lens		400 mm
	Microscope	
Microscope	Nikon	Eclipse TE2000-U
Dichroic 2	Cy3/Cy5 dichroic	Reflects: 320-350, 450, 515-555, 630-655nm
Objective	Nikon CFI APO TIRF 60X oil; NA=1.49, WD=0.12, 3 mm lens	
Objective oil	Nikon NF nd=1.515 cat# mxa 220124	
Tube Lens		200 mm
Beam height (from table)		209 mm
Beam center on the back of the microscope	Microscope width is 17.25 cm, center is 8.75 cm	
	Dark Box	
Filter 1		HQ 545/LP
Lens #1		175 mm
Dichroic 3	Red+ transmission, orange reflection	
	Green Channel	
Lens #2		200 mm
Filter 2		HQ 585/70
Camera	Photometrix	Evolve (512x512, 16μmx16 μm)
	Red Channel	
Lens #2		150 mm
Filter 2		HQ 725/120
Camera	Photometrix	Cascade II 512 (512x512, 16μmx16 μm)
	Software	
Software		NIS Elements

Table 2.1. The components on the optical table.

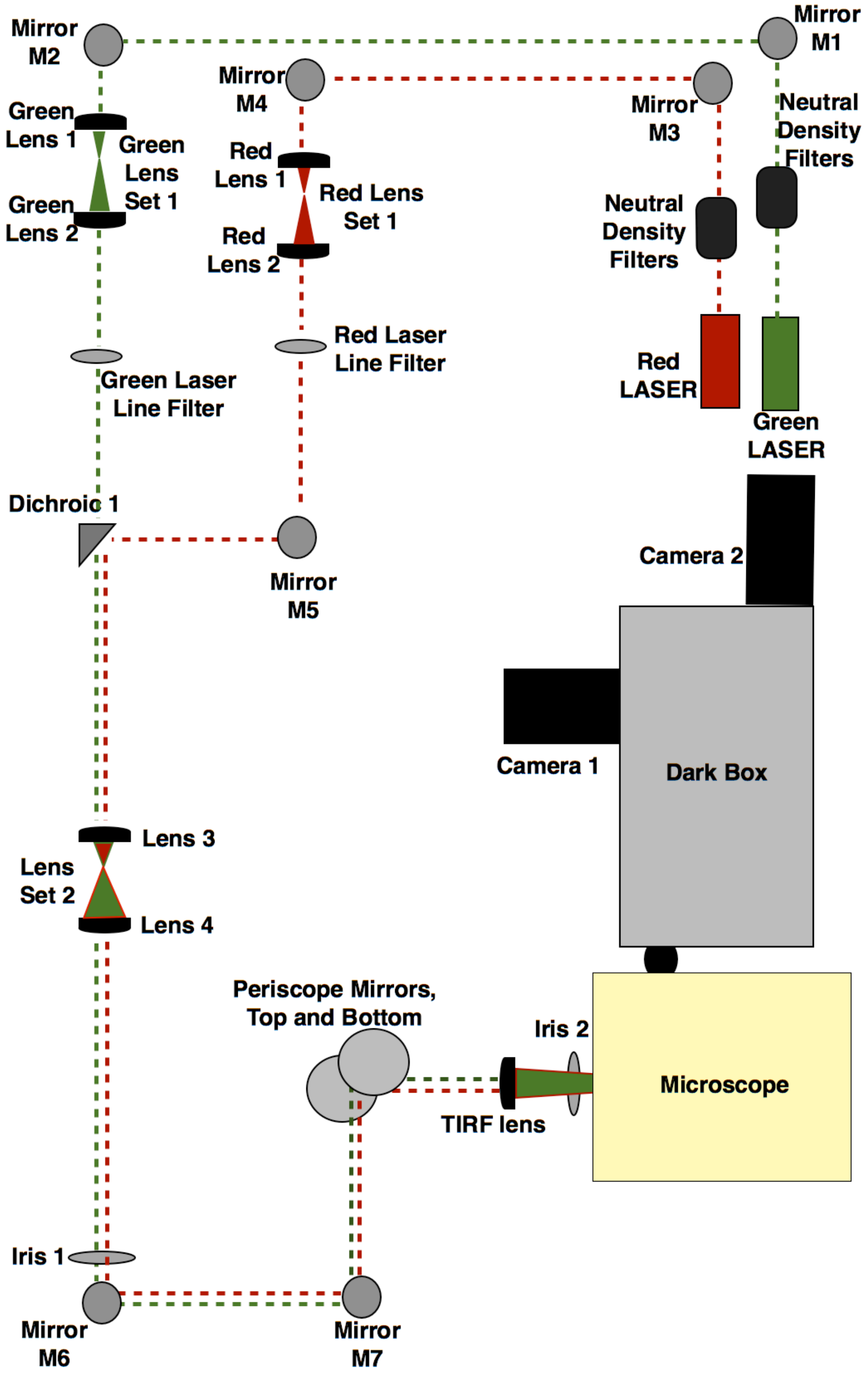


Figure 2.2. The components on the optical table. This diagram depicts the excitation side of the TIRF setup, where the beam is expanded, collimated, passes through the TIRF lens, and enters the microscope. The microscope and dark box components are depicted in Figures 2.3 and 2.4, respectively.

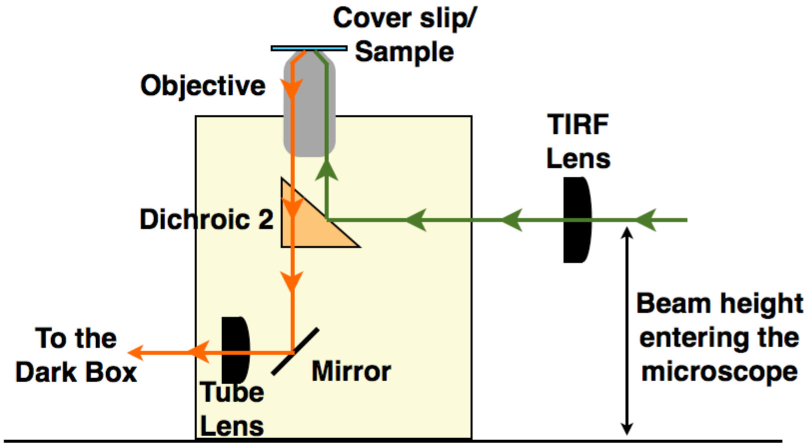


Figure 2.3. The setup of the microscope. The laser beam is collimated before entering the TIRF lens. The TIRF lens is set so that the distance of the TIRF lens from the back focal plane of the objective is the focal length of the lens, 400 mm. The emission light exiting the objective is collimated and goes through the tube lens. The dichroic and tube lens are described in Table 2.1. The emitted signal hits a mirror to direct the light leaving the microscope horizontally into the dark box.

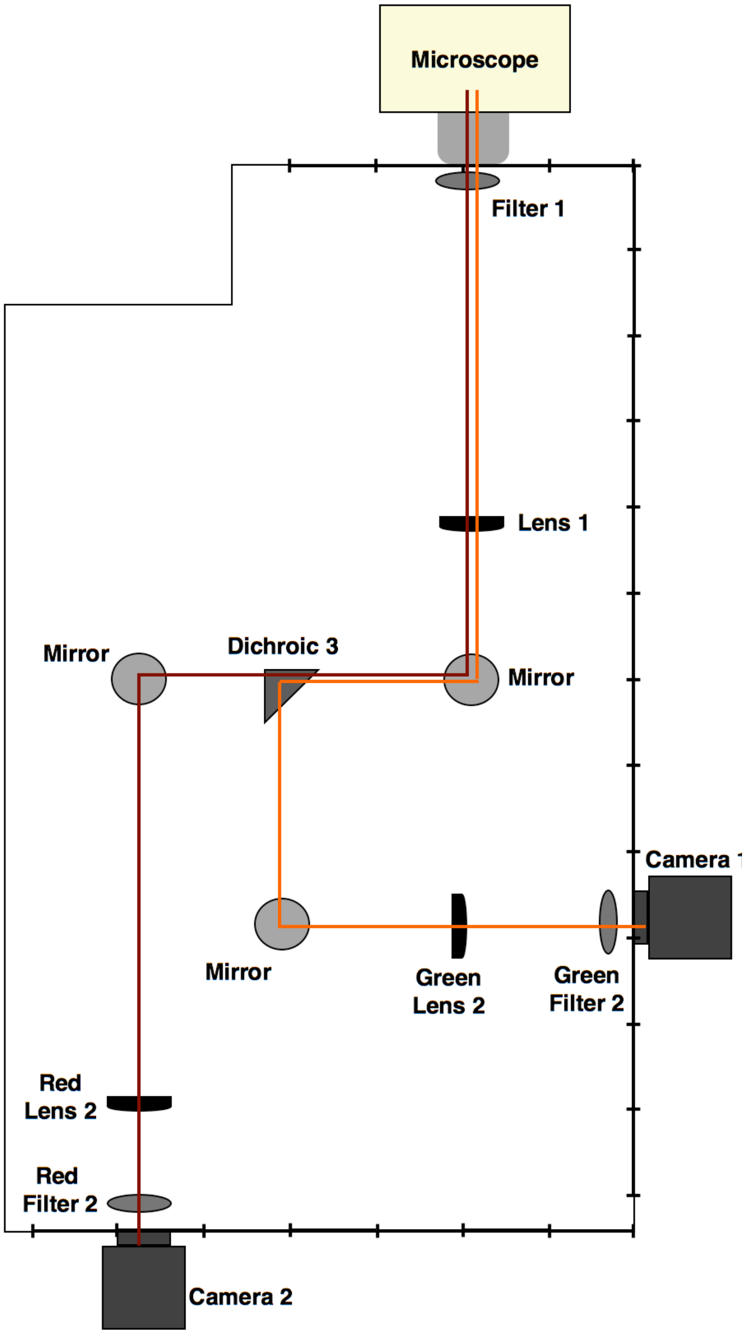


Figure 2.4. The dark box setup. The emission light enters the dark box through the port on the right side of the microscope. It then goes through a long pass filter and a lens before hitting a mirror. The light is separated by a dichroic into the green channel (orange wavelength) and the red channel (dark red wavelength). Light in each channel goes through another lens to focus it onto the camera CCD chip.

instrument, we more accurately installed the red laser to ensure that it excites the same region of the sample as the green laser, which is important for co-localization studies. We changed the TIRF lens to one with a smaller focal length to (1) increase the power at the surface and (2) ease assembly based on space on the table. Finally, we now translate the TIRF lens to bring the laser beam in and out of TIRF, instead of translating a mirror, as had been done previously.

We also made substantial changes and improvements in the dark box and camera system. Instead of viewing the emission of both the green and red channels on one camera, as we did in the past, we now collect data in the green and red channels on two separate cameras. To do so, we needed to significantly change the setup in the dark box. We also altered the lens set up in the dark box to reduce the magnification of the emitted light before it reaches the cameras, further increasing the efficiency of data acquisition.

Excitation filters

We added filters in the excitation optical path to limit the laser power and optimize the wavelengths that reach the sample. Specifically, neutral density filters were placed in front of the two lasers to reduce the amount of power reaching the objective lens. Neutral density filters cut the laser power on the log scale, as described in the following equation:

$$d= -\log(I/I_0) \qquad \text{Equation 2.8}$$

Where d is the optical density, I is the intensity after the filter, and I_0 is the initial laser power. In the green path, the filters have optical densities of 2.0, 1.0, 0.5, and 0.3 (Table 2.1). Therefore, when the neutral density filters are in the optical path, independently, the output from the 50 mW green laser (actually 41.5 mW, as measured with a power meter) is decreased to approximately 0.39, 3.1, 8.7, and 12.9 mW, respectively. The output from the red laser is reduced to 9.1, 29, and 85 mW when the neutral density filters with optical densities of 1.5, 1.0, and 0.5 are in its optical path, respectively. (The red laser with full power saturated the power meter. We back calculated the laser intensity after the neutral density filters using Equation 2.8, and found the red laser power is between 270 and 285 mW.) Bandpass filters are also in the red and green excitation lines to remove any wavelength of light generated by the laser that is not within a specific 10 nm range. For example, the bandpass filter in the 532/10 nm line eliminates the majority of light below 527 nm and above 537 nm going into the microscope (Table 2.1). Likewise, the 640/10 nm filter in the red laser line limited the red excitation wavelength to the range from 635 nm to 645 nm.

Emission Filters

Filters were added to the dark box to reduce unwanted scatter and background fluorescence from being detected by the cameras. We placed a longpass filter in the emission light path prior to splitting it into two channels to remove light below 545 nm, from entering the dark box. After splitting the emission light into two channels, we inserted a bandpass filter into each of the green and red emission channels, to reduce light emitted due to laser light scatter (Table 2.1) and to eliminate the higher edge of the raman scattering for the 532 nm and 640 nm lasers, which are calculated to be 645 nm and 811 nm, respectively (Equation 2.5). Although, we used a dichroic to split the red and green emission signals, the bandpass filters also served to reduce signal bleedthrough.

Beam size into the microscope and area of sample illumination

The microscope has a dual camera system to simultaneously image the green and the red emission channels on two cameras, one for each channel. Both cameras have an imaging chip that contains 512 x 512 pixels, with each pixel measuring 16 µm x 16 µm. To ensure that we obtain the maximum amount of information from our experiments we want 1 pixel to equal a diffraction limited spot, or 200 nm (Equation 2.4). The following equation is used to determine the area of the sample that needs to be illuminated by the laser (A_{ex}) to ensure that when the circular emission signal just fills the maximal area of the square CCD chip that the image will be at the limit of diffraction.

$$A_{ex} = n_{px} * M_{em} \qquad \text{Equation 2.9}$$

Where n_{px} is the number of pixels on the chip (512), and M_{em} is the magnification of the emission data onto the camera (200 nm/pixel). Under these conditions, the size of the laser beam exciting the sample needs to be 102.4 µm for the data to fill the entire camera chip. The area of illumination on the sample is determined by the diameter of the collimated laser beam when it hits the TIRF lens, the focal length of the TIRF lens (400 mm), and the focal length of the objective lens (3 mm). Using Equation 2.6 we determined that the laser beam needed to have a diameter of 13.3 mm when it hits the TIRF lens to illuminate a circular region of the sample with a diameter of 100 µm.

For each laser beam, we use 4 lenses (2 lens sets) to enlarge and collimate the beam (Table 2.1, Figure 2.2, Equation 2.6). The green laser has a 2 mm beam that goes into lens set #1, which enlarges it 5-fold to 10 mm. That green laser beam then goes into the second beam expander set, which enlarges the beam 3-fold to 30 mm. The diameter of our lenses cut off the beam, so the beam coming out of that lens is actually 23 mm. The red laser has a beam size of 2x3mm (oblong), which is increased 4-fold to about 9 mm, and further enlarged 3-fold to 28 mm, but again, the beam is cut off to 23 mm by the lens. We further decrease the beam size to 13.3 mm after the last lens set using an iris (Figure 2.2, Iris 1) before the periscope.

We imaged fluorescent beads to optimize the microscope setup and data collection. Green and red emission data obtained with a laser beam of approximately 13-14 mm entering the TIRF lens is shown in Figures 2.5A and 2.5B, respectively. Although the images are not exactly centered on the camera, the beam diameter of 13-14 mm resulted in illumination of a circular region of the sample that just fit within the 512 x 512 pixel CCD chip, yielding a pixel size of 200 nm. Further, the red and green lasers excited the same region of the sample, as shown by the overlay of the images in Figure 2.5C. The red and green images were further processed using an in-house written IDL program that corrects for subtle aberrations in the images to obtain the best overlap of spots between the images (Figure 2.5D).

The laser power outputs for the green and red lasers, as last measured, were 42 mW and 270 mW, respectively. The green and red laser powers drop to 4.6 mW and 12 mW, respectively, at the last iris before the microscope. We think most of this loss is due to the beams being cutoff during expansion and collimation. Moreover, the intensity of the red laser beam is not evenly distributed across the entire region of illumination, which can be seen in the image shown in Figure 2.6A (right panel). We think this might be due to the actual laser itself, which is not an expensive HeNe laser, but instead a laser diode that could have variable intensity across the laser beam.

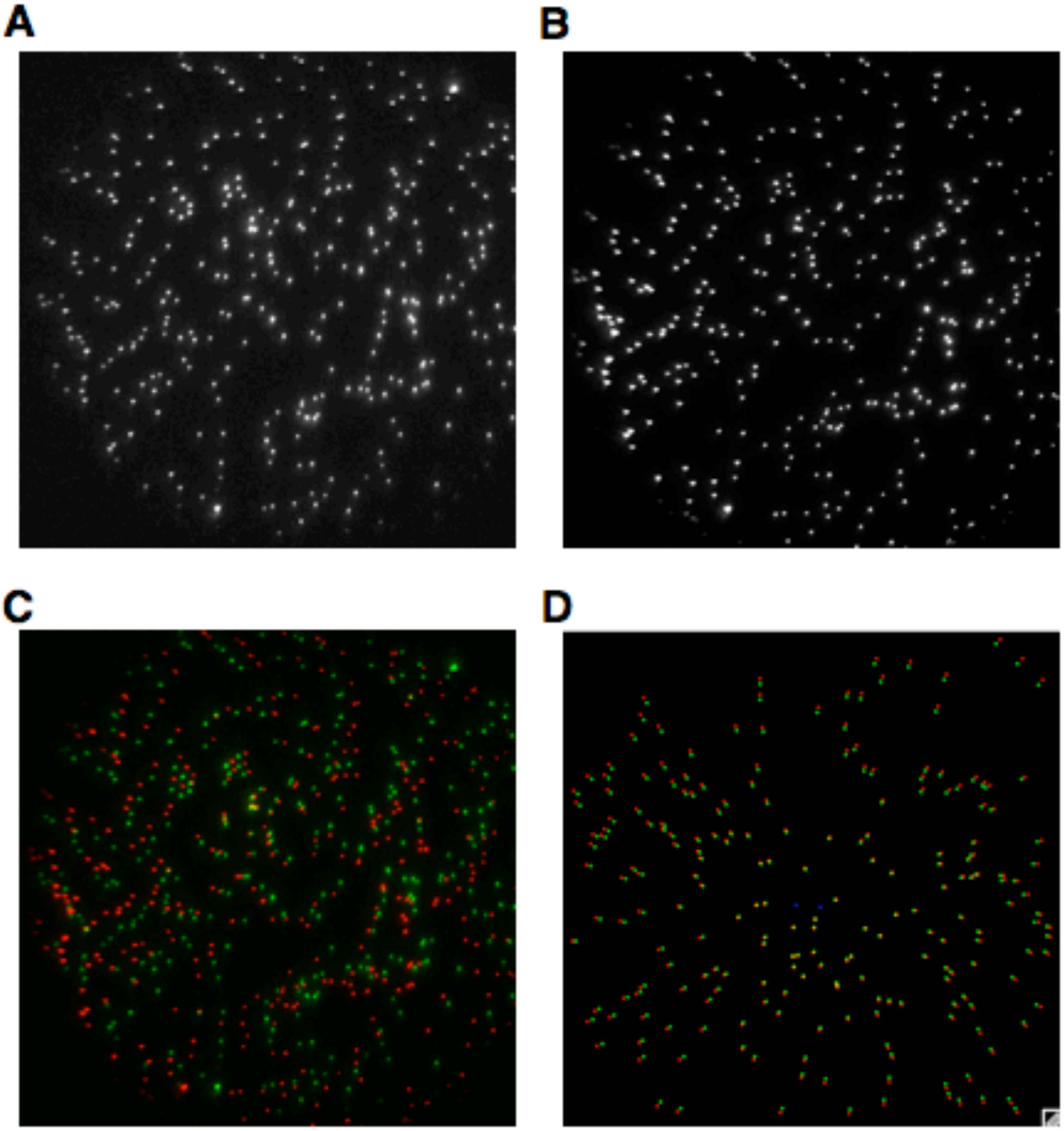


Figure 2.5. An overlay of the green and red emission channels. **(A)** An image of 100 nm Tetraspeck beads (Invitrogen) excited at 532 nm and imaged on the green channel. **(B)** 100 nm Tetraspeck beads excited at 640 nm and imaged on the red channel. The red image was rotated 90 degrees counterclockwise and transposed with respect to the green image to get the two images in the same orientation. **(C)** An overlay of the green and red spots. **(D)** Overlay of the green and red images after computationally modifying the red image to obtain the best overall overlap of spots.

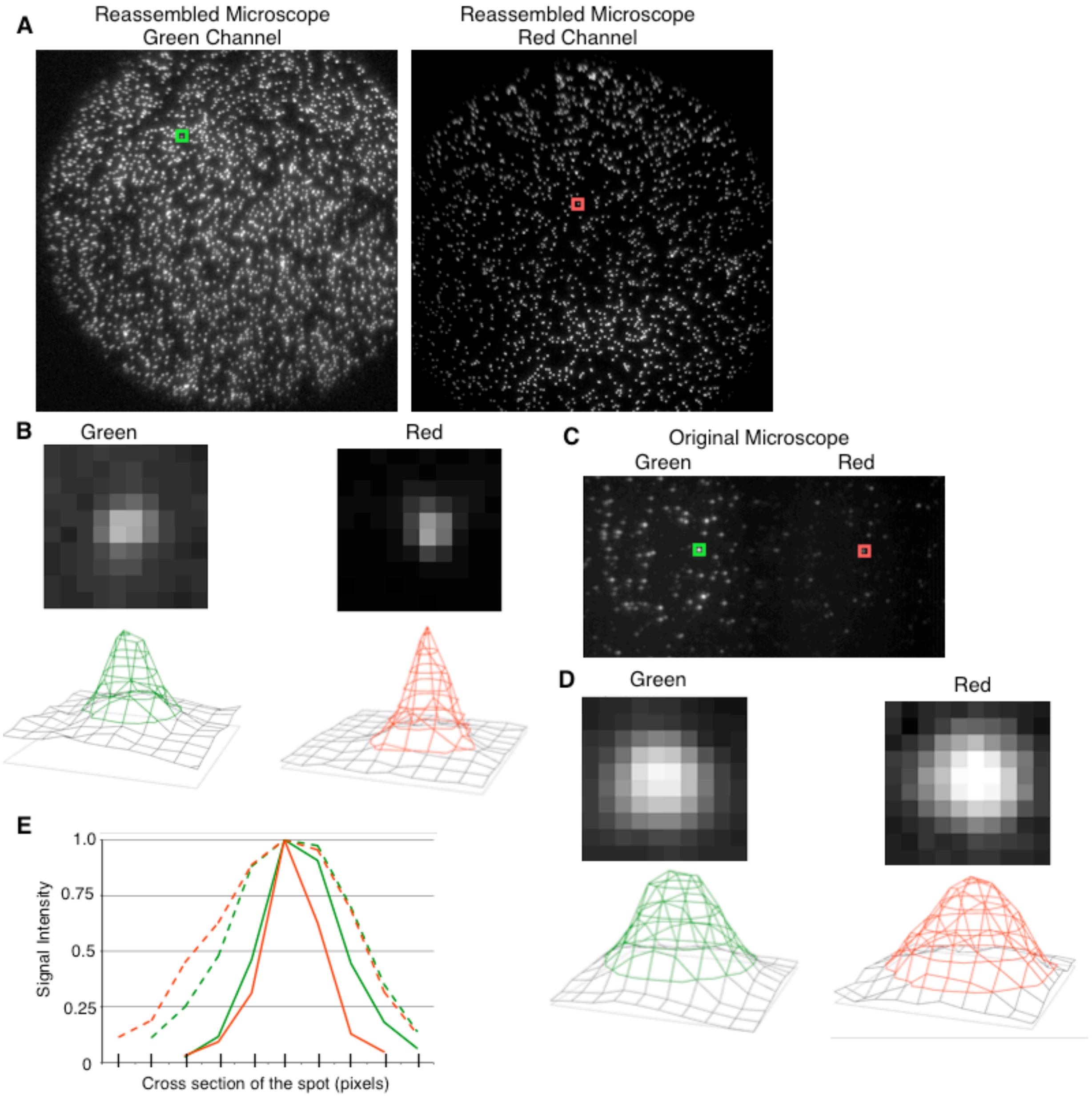


Figure 2.6. Single spots for the reassembled microscope versus the original microscope housed in Cristol Chemistry. **(A)** The images from the reassembled microscope are of 200 nm fluorescent beads collected with the green and red emission cameras that contain 512 x 512 pixels each. The laser beams entering the objective were approximately 14 mm in diameter. Raw images are shown for the green laser exciting 200 nm orange (540 nm excitation) fluosphere beads (Invitrogen) on the green emission channel (left) and the red laser exciting 200 nm dark red (660 nm excitation) fluosphere beads (Invitrogen) emitting in red (right). **(B)** A spot in each of the images in panel A (designated with a box) was expanded to allow visualization of the pixel size of each of the spots. Below each spot image is a 3D intensity plot of the spot across 10 x 10

pixels. **(C)** A split 512 x 256 image collected on the original Cristol microscope, showing fluorescent DNA excited by a 532 nm laser, where the left side is the green emission and the right side of the image is the red emission. The DNA contains donor and acceptor dyes undergoing approximately 25% FRET. **(D)** A spot in each of the images in panel C (designated with a box) was expanded to allow visualization of the pixel size of each of the spots. Below each spot image is a 3D intensity plot of the spot across 10 x10 pixels. **(E)** A cross section of each of the four expanded spot images (reassembled instrument, solid lines; original instrument, dashed lines) to compare the resolution of the two different instruments.

Magnification of the emitted light

Since we want each 16 μm pixel on the CCD chip to correspond to a 200 nm region of the illuminated sample, the magnification for the emission signal has to be 80-fold (16 μm/200 nm). Since the magnification from the objective is 60X, we further increased the size of the image 1.3-fold before the CCD chip. The tube lens in the microscope for the outgoing emissions beam has a positive focal length of 200 mm, but we do not know the actual position of this lens in the microscope. The emission beam is not collimated after passing through the tube lens. We placed a 175 mm focal length lens 22 cm into the dark box. Our beam path is long enough that the emission data would be too large and get cut off by our filters and mirrors if we only had one lens in the optical path, therefore, we put another lens closer to the cameras to focus the red and green beams onto the CCD chips (Figure 2.4, lens #2).

Figure 2.6A shows 200 nm beads excited by the green laser with green emission (left image) and excited by the red laser with red emission (right image). Figure 2.6B shows a 10 x10 pixel square surrounding a single 200 nm bead from each image. The 3D surface plot of the signal intensity for each spot is Gaussian-shaped. We compared these data to data obtained using the Cristol TIRF system, which consisted of a 512 x 256 pixel area in which the left half of the image was the green channel and the right half of the image was the red channel (Figure 2.6C). These data were collected using a green laser and a DNA construct with an Alexa555 dye and an Alexa647 dye 18 bp apart undergoing FRET. The right image is weaker since it is about 25% of the full strength signal. If we zoom in on a spot in each of the green and red images and again consider a 10 x 10 pixel area (Figure 2.6D), it can be seen that the spot size is larger than that obtained with the JSCBB TIRF system. To further illustrate this point, for each of the 4 spots considered in panels B and D, I normalized the intensities across the middle of the 10 x 10 pixel region (vertical pixel number 6) to the maximal pixel intensity and plotted the values versus pixel number (Figure 2.6E). The full widths at half max (FWHM) of these four spots are approximately 3 pixels for the red and green channels on the JSCBB microscope and approximately 6 pixels for the image obtained using the Cristol microscope. Hence, we effectively decreased the spot size for the red and green channels by 2-fold. We also increased the amount of data we collect on each channel by ~4-fold, thereby increasing the total amount of data collected in a given image by 8-fold.

We wanted to determine the FWHM of the emission intensities of beads across an entire image in both the green and red channels on the JSCBB microscope. The program that Jim Goodrich wrote to find and co-localize spots also outputs the FWHM for the x and y direction for each spot. Using this program to analyze 100 nm Tetraspek beads, the FWHM values for 152 spots in the green and red images were separately binned in the x and y directions (Figure 2.7A). The mean Gaussian of the FWHM for the green x and y are 2.4 ± 0.21 and 2.6 ± 0.22 pixels, respectively, and the mean for the red x and y are 2.7 ± 0.75 and 2.5 ± 0.43 pixels, respectively. Notice that the binned spots have a long tail on the higher end, indicating an inconsistent Gaussian fit to each spot. If the FWHM versus the pixel number in the x (Figure 2.7B) direction is plotted, there is more size variation on the left side of the image for both the red and green channels, which is apparent in the image of Figure 2.5A and B. The FWHM versus pixel number in the y direction (Figure 2.7C) does not show as much variation, indicating that the aberration is on the left side of the image. We have tried to figure out what the spot elongation is due to and have ruled out dirty filters and lenses, a scratch on the objective, distortions in lens or filter

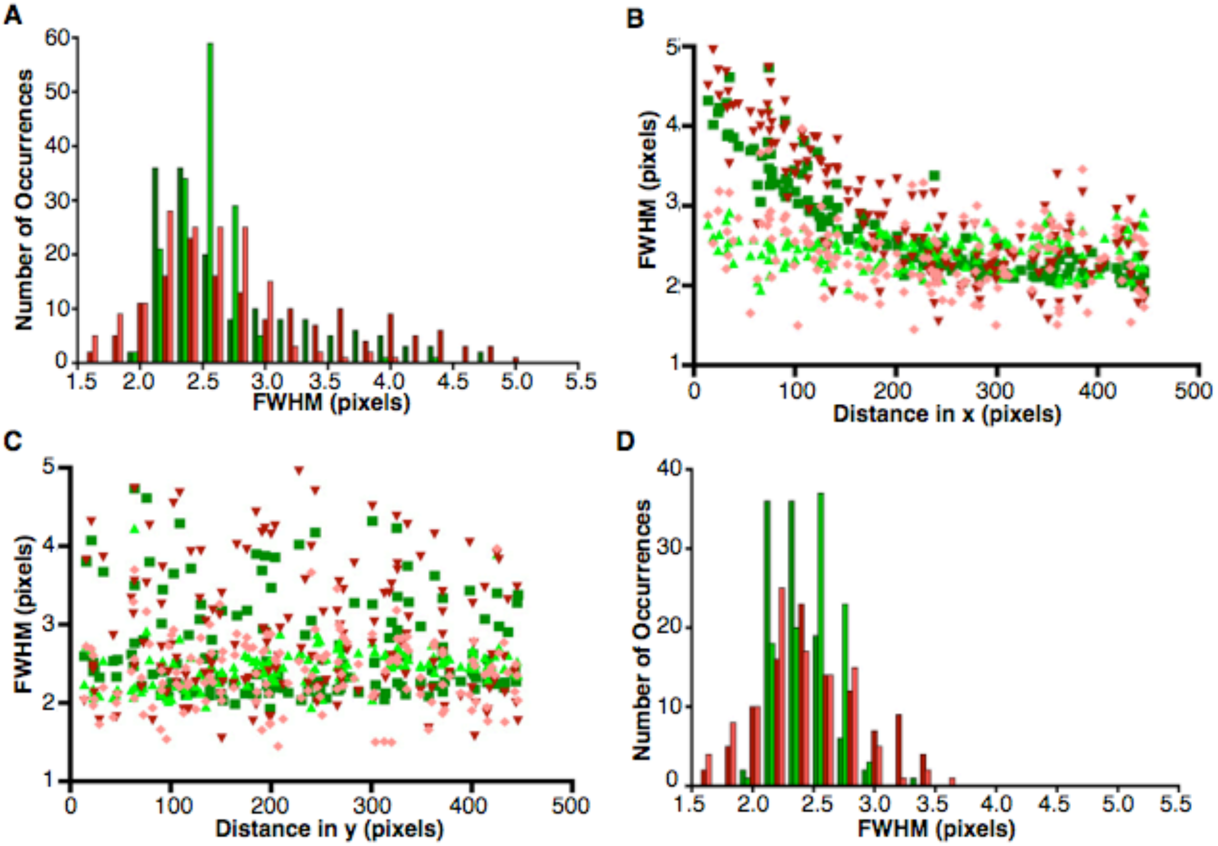


Figure 2.7. The full width at half maximal height (FWHM) measurements of green and red emission spots collected on the reassembled instrument. These measurements are shown for green spots in the x (dark green) and y directions (light green), and for the red spots in the x (dark red) and y directions (light red). **(A)** Binned FWHM data from the red and green channels in the x and y directions. **(B)** FWHM plotted versus distance in the x direction from left to right across the camera images. **(C)** FWHM plotted versus distance in the y direction from top to the bottom of the camera images. **(D)** Binned FWHM data after omitting the left-most 150 pixels in x direction.

position, and have not come up with anything conclusive. We think that the slide assembly with double-sided tape may be leading to this problem. If the data from the left most 150 pixels are removed and are binned, the tail on the Gaussian is no longer observed and the FWHM are 2.3 ± 0.19 and 2.6 ± 0.24 pixels for the green x and y, and 2.5 ± 0.44 and 2.4 ± 0.40 pixels in the red x and red y direction.

Experimental Procedures

There are important considerations for the alignment of the laser beams when building a TIRF microscope. First, having the correct beam path into the microscope is crucial for being able to translate the beam into TIRF. Prior to translating the TIRF lens into the TIRF position, the laser beam must be centered on the back of the objective and must enter the objective such that the beam is perpendicular with the objective lens. Second, the beam has to be collimated at the interface of the cover glass and the objective to get a good signal when using TIRF. Here, we discuss tools we used to obtain the correct line for the laser beams into the microscope and to get good collimation of the laser beams.

Aligning the path without lenses

To establish a straight and parallel beam on the table and going into the microscope, we placed the lasers, mirrors, periscope mirrors, and the microscope on the table first. We used a bullseye with a pinhole in the center, (Thor labs #LMR1AP), placed on a lens mount to make sure that our beam remained at the same height and along the lines we chose on the table. Using 2 irises of the same height, or an iris and the pinhole bullseye we ensured that the beam path was correct. We planned out the microscope path before attaching anything to the table. In doing so, placed two mirrors immediately after the lasers so that we could adjust the path of the beam coming out of the laser into where the first lens set would be placed (Figure 2.2, M1, M2 (green), M3, M4 (red)). The red laser line contained a mirror after the first lens set as well, (Figure 2.2, M5) so that the beam could merge with the green line through a dichroic (Figure 2.2, dichroic 1). In the future, it might be a good idea to put mirrors in the green line after the first lens set and before the second one, so that the path of the green laser can be adjusted after collimating the first set without adjusting the lenses themselves. At the periscope, the center of the bottom mirror was at the height of the beam to angle the beam onto the top mirror, which then directed the beam into the microscope. After all the mirrors were placed on the table and beam paths were correct, we roughly adjusted the periscope mirrors to get the beam into the center of the microscope.

Placing the lenses in the beam path

We used positive focal length, plano convex lenses. The direction of the lenses are important for reducing unwanted distortions in the laser beam. For the first lens in the lens set, the bulging end of the lens should face the collimated laser beam. For the second lens in the lens set, the flat end should face the first lens and the bulging end should face the beam path as it exits the lens set and goes toward the microscope. To collimate the beam exiting a lens set, the two lenses should be at the approximate distance of the sum of the two focal lengths (see Equation 2.7).

We placed the lens sets closest to the laser into the optical path first (Figure 2.2, green lens set 1 and red lens set 1), and centered the beam onto the lenses. The two lenses are approximately 180 mm apart for the green laser and 270 mm apart for the red laser. We made sure that the lenses did not change the beam path we previously defined and that they were not twisted, which would distort the beam. We used a pinhole, an iris, and a power meter to verify that the center of the Gaussian of the laser beam was in the center of the beam path. We next collimated the lens set by finely adjusting the distance between the two lenses so that the beam size remained unchanged over a large distance. We placed the final lens set immediately after where the green and the red beam lines converge (Figure 2.2, lens set 2). We collimated the set and they are about 200 mm apart.

Aligning the laser beams into the microscope

After expanding, collimating, and co-aligning the green and red laser beams, we only used the two periscope mirrors to direct the beams into the microscope. This was done in the absence of the TIRF lens and the objective lens. We used the bottom periscope mirror to adjust the beam path onto the back of the objective (or the center of the brass pinhole) and used the top mirror to make sure the angle going up to the ceiling was correct (the center of the mirror placed face down on the open objective holder). When adjusting the two mirrors, we overcompensated with one mirror and brought the beam back with the other mirror. For example, to adjust the bottom mirror to center the beams on the brass piece, we brought the beam to center and then past center with the bottom mirror and then brought the beam back to center with the top mirror. The iterative process is described in detail below.

We did not have the objective in the holder when we aligned the beams into the microscope. We used the combination of a mirror and a brass piece that screws into the objective holder and has a 5 mm pinhole in the center of it to determine the beam path. The brass pinhole gave us the center point of the bottom of the objective; we used the bottom periscope mirror to center the beam in the pinhole. To ensure that the beams were perpendicular with the microscope stage we placed a mirror face-down onto the open hole in the microscope. The mirror had to be completely flat against the surface of the objective holder so that the reflected beam path could be used for alignment. We then used a pinhole on an iris immediately after the second lens set (Figure 2.2, Iris 1) to decrease the size of the incoming laser beam and then adjusted the beam path with the periscope so that the position of the outgoing (reflected) beam was perfectly centered on the pinhole of the iris. We iterated between using the brass pinhole and the mirror on the objective holder until the incoming and reflected laser beams were aligned. We then marked a spot on the ceiling where the beam hit, which could be used in future alignment. (Note: This ceiling spot will only work for future alignment if the table legs stay pumped to the same position at all times.)

After determining the beam path into the microscope without the objective and without the TIRF lens, we wanted to make sure that the beam was still collimated over a large distance. To do this, we placed the mirror face down on the open objective holder and we opened Iris #1 up to 1 to 2 cm. We made sure that the reflected beam was the same size as the incoming beam at the lower periscope mirror and for each laser beam we adjusted the distance between the first lens set (Figure 2.2, green lens set 1, and red lens set 1) to get the beam collimated.

Alignment and collimation of the TIRF lens

We next placed the TIRF lens into the optical path at the position of epi-fluorescence. The placement of the TIRF lens in the optical path is important. It needs to focus the beam on the back focal plane of the objective and then the beam will travel to the 3 mm lens in the objective to collimate the light onto the surface of the coverslip. Hence, the TIRF lens should be placed the approximate distance of its focal length from the back of the objective (the focal length of the objective lens is negligible compared to the TIRF lens focal length). After initially positioning the TIRF lens we wanted to be positive that the beam path was the same with and without the TIRF lens in the optical path. We used the mirror and the brass pinhole to get the incoming and reflected beams to be the same, but instead of adjusting the periscope mirrors, we adjusted the TIRF lens to make sure that the beam was centered on the brass pinhole and that the reflection was the same going into and out of the microscope. It was also helpful to look at the mark on the ceiling and adjust the TIRF lens accordingly. If the beam path was slightly off with the TIRF lens, minor periscope mirror adjustments were performed.

To perfectly position the TIRF lens with respect to the objective lens, we put the objective lens in the microscope and kept the focus of the objective at the height close to where the beam would be focused on the coverslip. We then adjusted the TIRF lens distance from the microscope. If the TIRF alignment above was correct, the spot would get bigger and smaller but not move from the marking on the ceiling as the TIRF lens was moved. We determined the position of the TIRF lens where the beam on the ceiling was the smallest, which was the position where the light was best collimated coming out of the objective. In addition, the beam should form concentric circles (not ellipses) when hitting the ceiling, otherwise the placement of the TIRF lens is off.

All alignment done up to this point was to ensure the beam was properly angled into the objective lens in epifluorescence mode. There are two methods that can be used to put a microscope into TIRF (80). The first method is to translate the laser beam to the left or right as it enters the microscope. This is how the original Cristol microscope was put into TIRF. The second method is to translate the TIRF lens left or right, which is the method we use for the

JSCBB scope. The circular reflected beam comes out of the back of the microscope on a different path than the incoming beam when the microscope is in TIRF.

Dark box alignment

We set up the dark box in a rough manner to see if we could observe TIRF spots on the cameras while we were aligning the laser beams into the objective (above); it was helpful to watch the images from the cameras on the computer screens while attempting to get the microscope to enter TIRF. Specifically, when the microscope is in epifluorescence mode, the objective can focus on different planes in the solution. If you cannot focus on different planes in the solution and can only focus on the surface, then the microscope is in TIRF position. We did not set up the dark box permanently until we determined the exact beam path into the microscope, because the emission angles of the light entering the dark box would change.

To set up the dark box, we used very bright 200 nm fluorescent beads in PBS on a coverslip (200 nm Fluospheres from Invitrogen), focused the objective, and put the microscope into TIRF by translating the TIRF lens. In the dark, we could see the emitted light coming out of the microscope using a kimwipe in the optical path to follow the path of the light. We placed the mirrors and filters into the optical path(s) using this bright emission. Then we placed the lenses in both the red and green emission channels (Figure 2.4. Lens 1, Red Lens 2, Green Lens 2). We made sure that the lenses were at a position where the data was focused onto the camera without being cut off, and the magnification of the data was as intended (see “Magnification of the sample onto the camera” section).

Jim developed software in the IDL programming language that allowed us to analyze time course data from the green and red channels. The green and red emission data had to be aligned onto the cameras accurately so that the software could overlay the data from the green and red channels. When running the program, the user manually inputs the offset from the center and the degree of rotation of the red image. For instance, on May 30, 2013, when we reinstalled a red laser onto the microscope table, we collected data using fluorescent beads and found that the red channel offset from the center was +4 pixels in the x direction, and -3 pixels in the y direction, and the red image needed to be rotated 7 degrees. This optimization of x shift, y shift and rotation needs to be redone each time the optical table is changed, either intentionally or due to drift over time.

CHAPTER 3

Single molecule FRET shows uniformity in TBP-induced DNA bending and heterogeneity in bending kinetics[1]

---

[1] This chapter is adapted from **Blair RH**, Goodrich JA, and Kugel JF. (2012). Single molecule FRET shows uniformity in TBP-induced DNA bending and heterogeneity in bending kinetics. *Biochemistry* **51**, 7444-55.

Summary

TATA binding protein (TBP) is a key component of the eukaryotic RNA polymerase II (Pol II) transcription machinery that binds to TATA boxes located in the core promoter regions of many genes. Structural and biochemical studies have shown that when TBP binds DNA, it sharply bends the DNA. We used single-molecule FRET (smFRET) to study DNA bending by human TBP on consensus and mutant TATA boxes in the absence and presence of TFIIA. We found that the state of the bent DNA within populations of TBP/DNA complexes is homogeneous; partially bent intermediates were not observed. In contrast to previous ensemble studies, TBP was found to bend a mutant TATA box to the same extent as the consensus TATA box. Moreover, in the presence of TFIIA the extent of DNA bending was not significantly changed, although TFIIA did increase the fraction of DNA molecules bound by TBP. Analysis of the kinetics of DNA bending and unbending revealed that on the consensus TATA box two kinetically distinct populations of TBP/DNA complexes exist, however, the bent state of the DNA is the same in the two populations. Our smFRET studies reveal that human TBP bends DNA in a largely uniform manner under a variety of different conditions, which was unexpected given previous ensemble biochemical studies. Our new observations lead to us to revise the model for the mechanism of DNA binding by TBP and for how DNA bending is affected by TATA sequence and TFIIA.

Introduction

Transcription of protein-encoding genes is a primary step in gene expression during which mRNA is synthesized from a DNA template. In eukaryotic cells, Pol II synthesizes mRNA, and requires numerous accessory factors to transcribe genes in a regulated manner (4). One such factor is the TFIID (transcription factor IID) complex, which consists of the TATA binding protein (TBP) and TBP associated factors (TAFs). The TFIID complex nucleates preinitiation complex formation at the promoters of genes by binding core promoter elements and recruiting the remaining general transcription factors and Pol II (4). The TBP subunit of TFIID binds to TATA boxes (consensus sequence TATA(A/T)AA(G/A)) that are located approximately 25 basepairs upstream from transcription start sites in a subset of eukaryotic protein-encoding genes (4).

Biochemical experiments and X-ray crystallography have shown that yeast and human TBP are saddle-shaped proteins that bind the minor groove of TATA DNA (82-86). Phenylalanine residues in TBP insert between the bases at positions 1/2 and 7/8 of the TATA box, which disrupts base stacking interactions and bends the DNA away from the protein. TBP-induced DNA bending has previously been described using a two-kink model, which allows the bend angle to be calculated assuming the bent DNA is planar (26, 27). Applying this model to measurements made using fluorescence lifetime and fluorescence resonance energy transfer (FRET) revealed that human and yeast TBP bend the consensus TATA sequence to angles of ~100 and ~80 degrees, respectively (26, 29, 31, 87). A more complex model has also been described in which the DNA is both bent and twisted, such that the ends are no longer in the same plane (29); using this model the aforementioned ensemble FRET data would be consistent with a wide range of bend and twist angle combinations.

Several different models have been proposed for the mechanism of TBP binding and bending of TATA DNA using human and yeast TBP. In one model, TBP first interacts with DNA, and then subsequently bends the DNA (88, 89). Using rigorous ensemble lifetime measurements and other biochemical experiments Parkhurst and colleagues concluded that TBP simultaneously binds and bends the DNA (31, 90, 91). In their model, two bound intermediates with unique kinetic profiles exist en route to the final complex; however, each of the intermediates has the same degree of bending as the final complex. Optical tweezer experiments led another group to conclude that a long-lived partially bent intermediate forms between the unbound and fully bent states, suggesting a stepwise bending model. They proposed that this intermediate occurs when one side of the TBP saddle intercalates into the DNA, resulting in a partially bent state (92). Additional studies using other approaches would provide further insight into the mechanism of TBP-induced DNA bending.

The TATA sequence itself affects multiple properties of the TBP/DNA complex. Altering the consensus sequence at one or two positions decreases the affinity of TBP for DNA and the kinetic stability of the TBP/DNA complex (26, 28-30). In addition, ensemble biochemical studies have concluded that the angle at which human and yeast TBP bend DNA is affected by the TATA sequence; mutant sequences that TBP binds with lower affinity appear less bent compared to the consensus sequence (26, 28, 29, 31, 32). Moreover, for yeast TBP, weaker TATA boxes that appear to be bent to a lesser degree in vitro (26) are less transcriptionally active in cells (26, 33).

The transcription factor IIA (TFIIA) has been shown to regulate the TBP/DNA interaction; it increases the kinetic stability of TBP/DNA complexes at consensus as well as nonconsensus TATA sequences (22, 29, 34-36) and alters the sequence specificity with which TBP binds DNA (34). Moreover, TFIIA has been found to change the extent to which DNA is bent by human TBP. When the planar two-kink model was applied to ensemble FRET data, TFIIA appeared to decrease the bend angle of a consensus TATA sequence and increase the bend angle of a mutant sequence, resulting in both sequences having a similar bend angle (29).

Here, we used single molecule FRET (smFRET) to monitor individual bent and unbent DNA molecules in the presence of human TBP. This system was used to study the extent of heterogeneity in DNA bending by human TBP, and how the bent conformation is affected both by TFIIA and a mutation in the TATA sequence. We found that the DNA molecules within TBP/DNA complexes exist in a homogeneous bent state; we did not observe a population of DNA molecules partially bent by TBP. Moreover, we found that the extent to which the DNA is bent by TBP is the same with consensus and mutant DNA, and in the presence and absence of TFIIA. We also analyzed the kinetics of bending and unbending on consensus and mutant TATA DNA. We found that DNA bending by human TBP displays heterogeneous kinetics on the consensus TATA box, but not a mutant TATA box. Our smFRET results provide unique insights that lead to a re-evaluation of existing models for TBP DNA bending that were derived from ensemble experiments.

Results

Consensus TATA DNA molecules are uniformly bent by TBP

To study DNA bending by human TBP using smFRET, we designed a DNA construct with donor and acceptor fluorophores positioned such that when TBP induced a bend, the distance between the donor and the acceptor fluorophores would decrease, and an increase in FRET efficiency would be observed (Figure 3.1A). The donor (Alexa 555) and acceptor (Alexa 647) dyes were attached to the 5'-ends of two oligos, that when annealed, formed an 18 bp DNA with a consensus TATA box at its center (sequence from the Adenovirus major late promoter, AdMLP, shown in Table 3.1) and a 20 nt single stranded 3' overhang. This single stranded overhang was complementary to the sequence of a third oligo that had a biotin on its 5'-end. After annealing the three-oligo construct, it was attached to a glass surface through a biotin-streptavidin linkage. The sample chamber was placed on a microscope and imaged using total internal reflection fluorescence (TIRF) microscopy. In all experiments the donor was excited with a 532 nm laser and the donor and acceptor emissions were monitored over time with a CCD camera. Apparent FRET efficiencies were calculated as described in the methods.

We first imaged DNA in the absence of TBP. An example of an unbound DNA trace is shown in Figure 3.1B; the donor (green), acceptor (red), and total fluorescence (gold) remained constant over time until the donor dye photobleached at 53 s. The blue trace (lower panel) shows the FRET efficiency calculated from the donor and acceptor emissions, with the black line showing the FRET state fit to the data. In all time courses of DNA imaged in the absence of TBP we observed no noticeable change in FRET efficiency prior to fluorophore photobleaching.

When human TBP was flowed into the sample chamber, we observed changes in the

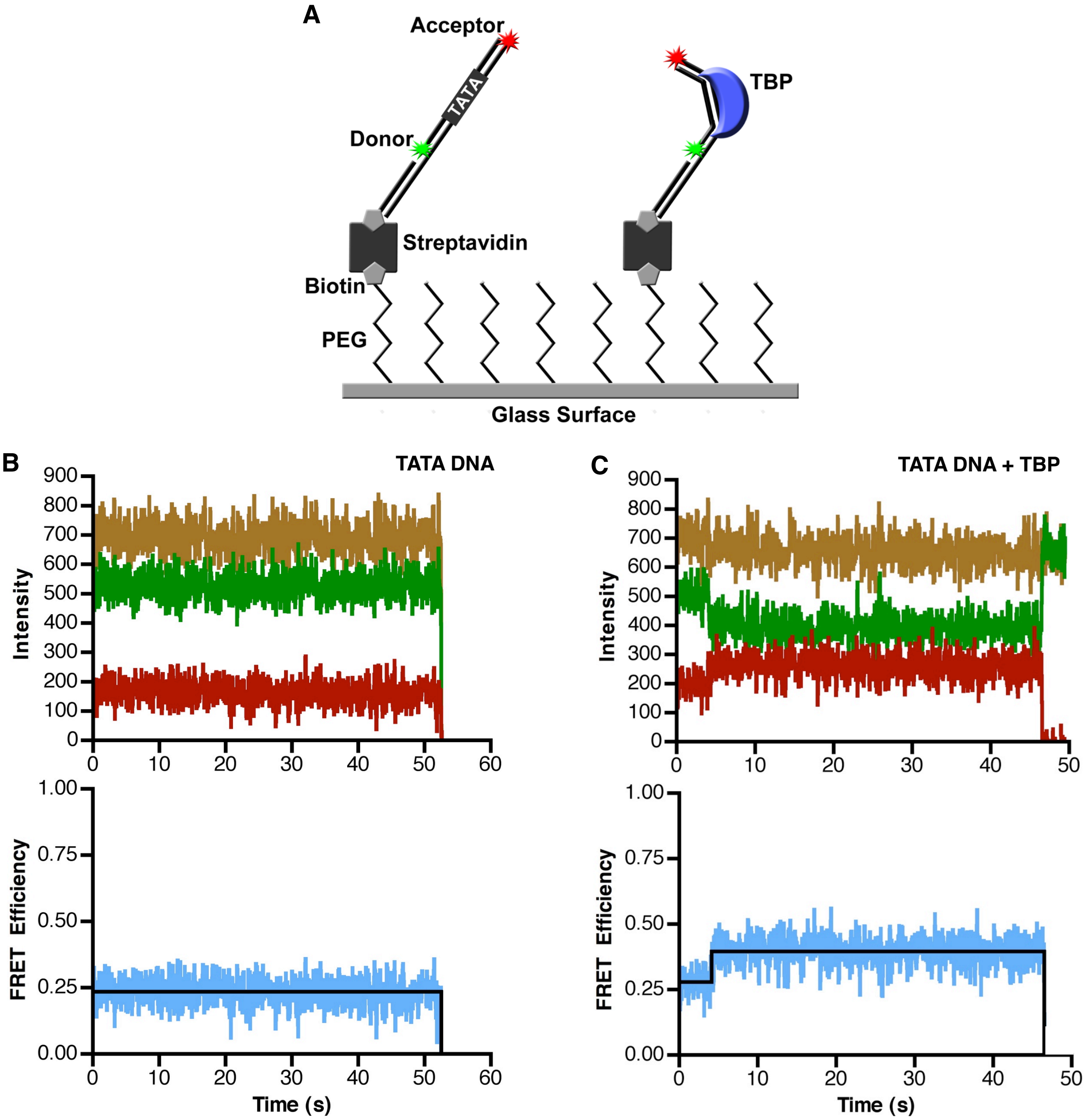


Figure 3.1. TBP causes an increase in the FRET observed from single immobilized consensus TATA DNA molecules. **(A)** The single-molecule surface configuration for monitoring TBP-induced DNA bending. When TBP bends the immobilized DNA it causes an increase in FRET efficiency. **(B)** The consensus TATA construct exhibits constant low FRET in the absence of TBP. The donor dye (green), the acceptor dye (red), and sum of the two dyes (gold) did not change over the course of imaging (upper panel). The calculated FRET at each time point is shown in blue and the derived FRET state of 0.24 is shown in black (lower panel). For the purpose of display, these data were smoothed by averaging the values at 3 time points. **(C)** TBP causes an increase in FRET consistent with DNA bending. The donor and acceptor dyes show anti-correlated changes in signal intensity (upper panel) and an increase in calculated FRET efficiency (lower panel) at 4 seconds, indicating TBP bent the DNA. At 47 s, the acceptor dye photobleached, causing emission from the donor dye to increase. For the purpose of display, these data were smoothed by averaging the values at 3 time points.

Oligo	**Sequence (5' to 3')**
Consensus TATA donor	CTATAAAAG
Consensus TATA acceptor	CTTTTATAG
TATA(A3) donor	CTAAAAAAG
TATA(A3) acceptor	CTTTTTTAG
TATA-less donor	TAGAGTCGG
TATA-less acceptor	CCGACTCTA

Table 3.1. Sequences unique to each of the constructs.

FRET efficiency. A time trace for a single DNA is shown in Figure 3.1C. Here, the donor and acceptor signals (upper panel) and FRET efficiency (lower panel) remained constant until approximately 4 s, at which time the acceptor intensity increased, and the donor intensity simultaneously decreased. This anti-correlated change in dye intensities resulted in an increase in FRET efficiency from 0.28 to 0.4, indicative of TBP bending the DNA (lower panel of Figure 3.1C). The FRET efficiency remained steady until the acceptor dye photobleached at 47 s. The sum of the donor and acceptor intensities (gold trace in the upper panel) did not appreciably change throughout the time course, indicating that TBP does not affect the fluorophores themselves.

The FRET efficiencies for 301 immobilized consensus TATA DNA molecules in the absence of TBP were plotted on a histogram and fit to a Gaussian, yielding a single peak with a FRET efficiency centered at 0.24 (Figure 3.2A and Table 3.2). Analysis of time traces obtained for 477 DNA molecules in the presence of TBP, yielded 1727 distinct FRET occurrences that were histogrammed and fit to Gaussians, yielding two separate peaks with mean FRET efficiencies of 0.26 and 0.39 (Figure 3.2B). We conclude that the lower FRET state is unbent DNA and the upper FRET state, which depends on the addition of TBP, is the TBP-bound and bent DNA. Table 3.2 shows the mean FRET values obtained from all Gaussian fits, and Figure 3.3 shows the standard deviations for the Gaussian fits, as a representation of the widths of the Gaussians.

In order to determine whether the change in FRET efficiency upon addition of TBP was dependent on a TATA box, a TATA-less construct (sequence shown in Table 3.1) was immobilized and used for smFRET experiments in the absence and presence of TBP. In the

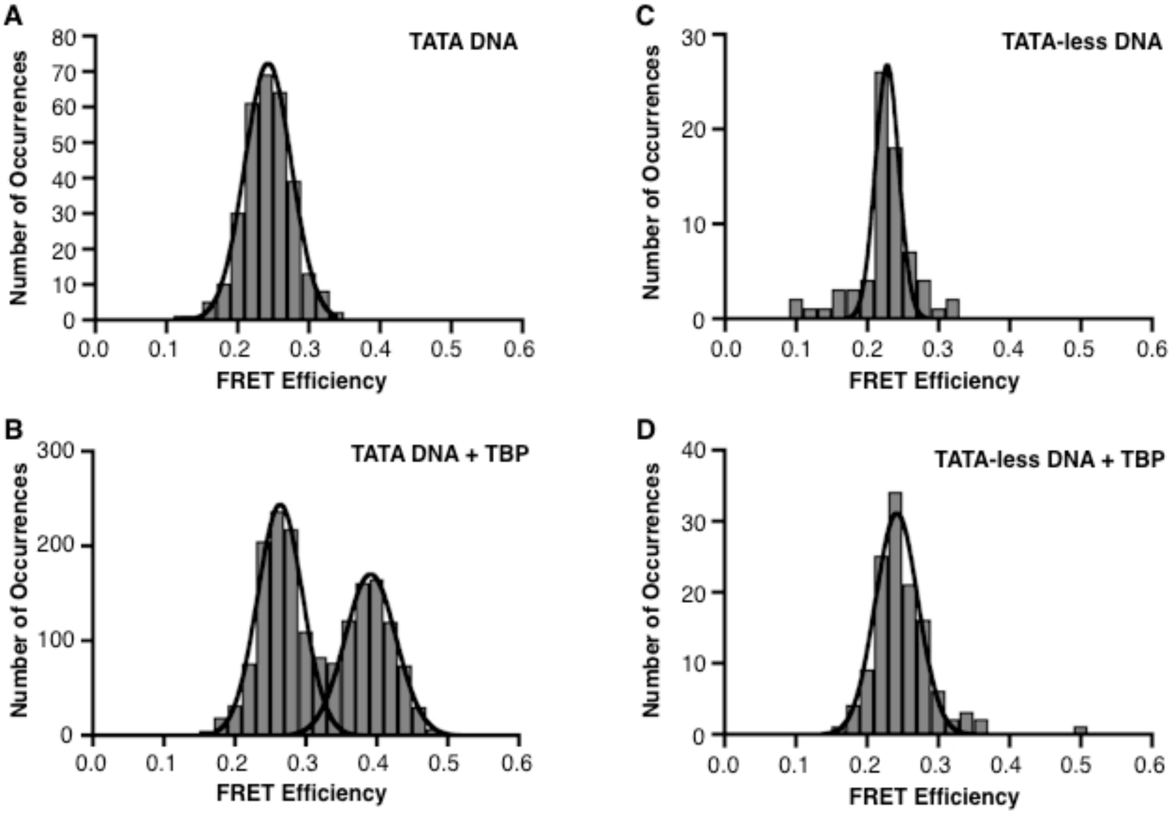


Figure 3.2. TBP/DNA complexes exist in a uniform FRET state. **(A)** Consensus TATA DNA exists in a homogenous population in a relatively low FRET state. FRET efficiencies for 301 DNA molecules were plotted as a histogram and fit with a Gaussian. The mean FRET efficiency is 0.24. **(B)** In the presence of TBP, the DNA molecules exist in a two-state population (unbent and bent). Time traces for 477 DNA molecules were analyzed and 1727 distinct FRET occurrences were plotted as a histogram and fit with Gaussians. The low and high FRET states have mean FRET efficiencies of 0.26 and 0.39, respectively. **(C)** Shown is the histogram and Gaussian fit of FRET efficiencies determined for 72 TATA-less DNA molecules in the absence of TBP. The mean FRET efficiency is 0.23. **(D)** TBP does not alter the FRET efficiency of a TATA-less DNA construct. The histogram and Gaussian fit of 124 distinct FRET occurrences observed in time traces from 121 TATA-less DNA molecules in the presence of TBP is shown. The mean FRET efficiency is 0.24.

DNA construct	[KCl] mM	TBP	TFIIA	Unbent mean FRET	Bent mean FRET	Unbent 95%CI	Bent 95%CI
Cons. TATA	150	–	–	0.24	N/A	0.241-0.244	N/A
Cons. TATA	150	+	–	0.26	0.39	0.261-0.266	0.388-0.394
Cons. TATA	150	+	+	ND	0.39	ND	0.386-0.398
Cons. TATA	50	0.4x	–	0.26	0.37	0.256-0.271	0.363-0.380
TATA(A3)	150	–	–	0.25	N/A	0.249-0.253	N/A
TATA(A3)	150	+	–	0.26	ND	0.259-0.262	ND
TATA(A3)	150	+	+	0.27	0.41	0.271-0.277	0.402-0.411
TATA(A3)	50	–	–	0.25	N/A	0.250-0.258	N/A
TATA(A3)	50	+	–	0.27	0.36	0.262-0.274	0.353-0.369
TATA(A3)	50	4x	–	0.26	0.37	0.253-0.271	0.361-0.379
TATA-less	150	–	–	0.23	N/A	0.224-0.231	N/A
TATA-less	150	+	–	0.24	ND	0.239-0.244	ND

Table 3.2. Mean FRET efficiencies of DNA under each condition.

N/A – not applicable
ND – not detected
95% CI – 95% confidence interval for the mean FRET

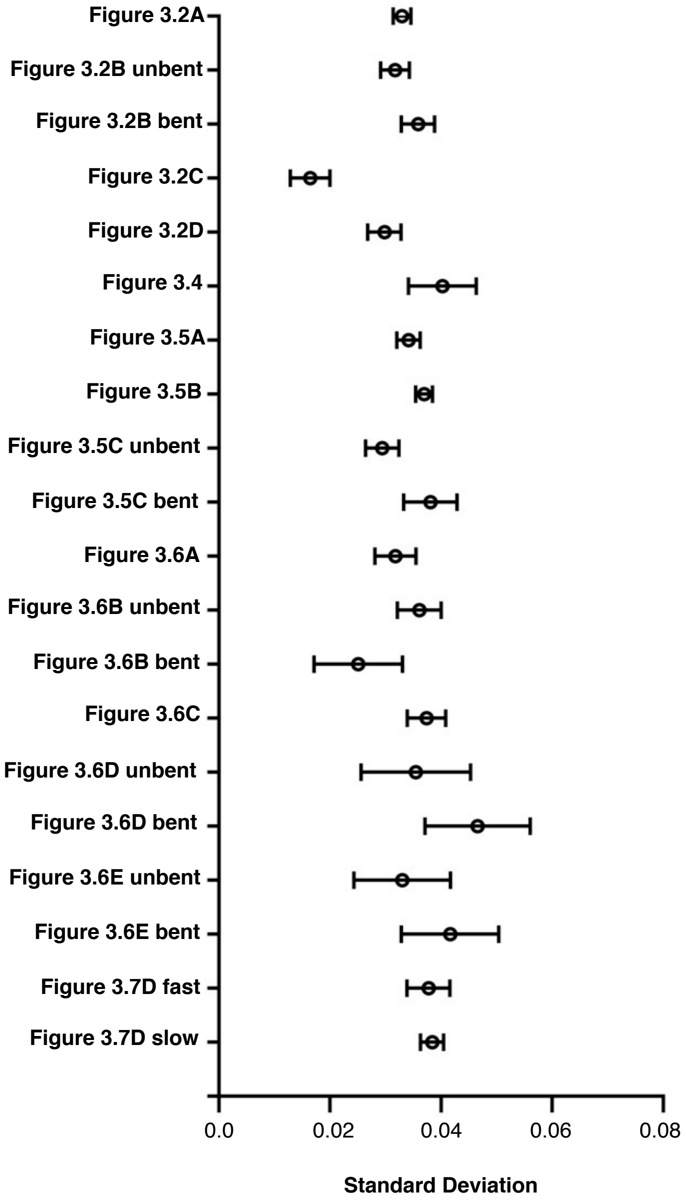


Figure 3.3. Plot of the standard deviations of the Gaussian fits as a measurement of their widths. The error bars represent the 95% confidence intervals. The free TATA-less DNA (Figure 3.2C) fits with a slightly narrower Gaussian; the widths of the remainder of the Gaussian fits are largely overlapping within their 95% confidence intervals.

absence of TBP, this construct had a FRET efficiency of 0.23 (Figure 3.2C). A histogram of the TATA-less construct in the presence of TBP resulted in a single peak with a FRET efficiency of 0.24 (Figure 3.2D). Therefore, we conclude that the higher FRET state observed with the consensus TATA construct (Figure 3.2B) represents sequence-specific DNA bending by TBP. We further conclude that the TBP/DNA complexes contain TATA DNA in a uniformly bent state.

TFIIA shifts the population of DNA molecules to the bound state but does not change the extent of DNA bending

We next asked whether TFIIA could affect either the population distribution or the FRET efficiency of the bent state of TATA DNA molecules in our smFRET system. When TBP and TFIIA were flowed together into a chamber containing the immobilized consensus TATA construct, the DNA population strongly shifted to a higher FRET state (Figure 3.4). These data fit to a single Gaussian distribution with a peak FRET state of 0.39. We can draw two primary conclusions from these results: 1) TFIIA shifted the population of TATA DNA toward the TBP bound and bent state, consistent with its ability to increase TBP's affinity for DNA (22, 29, 34), and 2) TFIIA did not change the FRET state of the TBP bound/bent DNA.

TBP bends a mutant TATA DNA similarly to the consensus TATA DNA

We investigated how mutating the TATA sequence changed TBP-induced DNA bending at the single molecule level. We used the TATA(A3) sequence, which contains a single T to A point mutation in the consensus sequence (Table 3.1) and has previously been shown to bind TBP with significantly lower affinity and to be bent to a lesser extent by TBP in ensemble studies

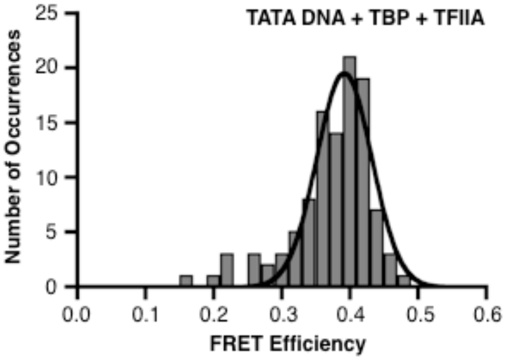


Figure 3.4. When included with TBP, TFIIA shifts the consensus TATA DNA to a homogenous population in the higher FRET state. Shown is the histogram and Gaussian fit of 107 distinct FRET occurrences observed in time traces from 91 consensus DNA molecules in the presence of TBP and TFIIA; the mean FRET efficiency is 0.39.

(26, 28, 29). Immobilized mutant TATA(A3) DNA in the absence of TBP exhibited a FRET efficiency similar to that of the consensus TATA DNA (Figure 3.5A), indicating that, as expected, the point mutation had little, if any, effect on the FRET observed with unbound DNA. After flowing TBP into the sample chamber, we did not observe a higher FRET state in the histogram, indicating that the DNA molecules were unbent (Figure 3.5B), which is consistent with the weak affinity of TBP for this mutant DNA compared to that of the consensus TATA DNA.

We anticipated that TFIIA would increase the extent of TBP binding to the mutant TATA sequence because TFIIA has been shown to stabilize the TBP/DNA complex with both consensus and nonconsensus sequences (22, 29, 34). Indeed, when TFIIA and TBP were flowed together into a chamber containing the immobilized TATA(A3) construct, we observed many molecules that occupied a higher FRET state (Figure 3.5C) with a mean FRET efficiency of 0.41, which is similar to that observed with TBP/TFIIA on the consensus TATA DNA.

To study TBP bending the TATA(A3) construct in the absence of TFIIA, we reduced the KCl concentration from 150 mM to 50 mM; the lower salt concentration did not affect the FRET of the TATA(A3) construct in the absence of TBP (Figure 3.6A). In the presence of TBP, a higher FRET state indicative of bending emerged on the TATA(A3) DNA with a mean efficiency of 0.36 (Figure 3.6B). To ensure that the unbent states in the histogram did not affect the Gaussian fit of the bound/bent states, we histogrammed just the bent states. These bound/bent states were obtained only from those traces in which a distinct transition from a low FRET state to high FRET state, and back to a low FRET state was observed. The Gaussian fit of the histogram yielded a mean FRET efficiency of 0.36 (Figure 3.6C), identical to that observed when the unbent and bent states were histogrammed together (Figure 3.6B). We also performed the

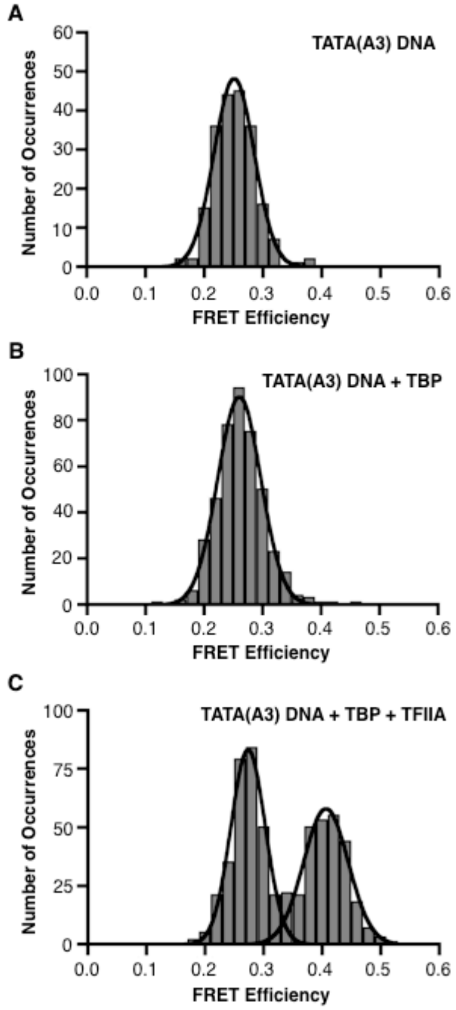


Figure 3.5. TFIIA facilitates TBP bending the mutant TATA(A3) sequence. **(A)** In the absence of TBP, the mutant TATA sequence exists in a uniform FRET state similar to that observed with the consensus TATA sequence. The mean FRET efficiency for the Gaussian fit of data from 206 DNA molecules is 0.25. **(B)** TBP does not shift the TATA(A3) DNA into a higher FRET state. Shown is the histogram and Gaussian fit of 427 distinct FRET occurrences observed in time traces from 414 TATA(A3) DNA molecules in the presence of TBP; the mean FRET efficiency is 0.26. **(C)** TFIIA facilitates the bending of the TATA(A3) DNA by TBP. The mean FRET efficiencies of the low (319 occurrences) and high (274 occurrences) states obtained from the Gaussian fits of distinct FRET occurrences observed in time traces from 400 TATA(A3) molecules are 0.27 and 0.41, respectively.

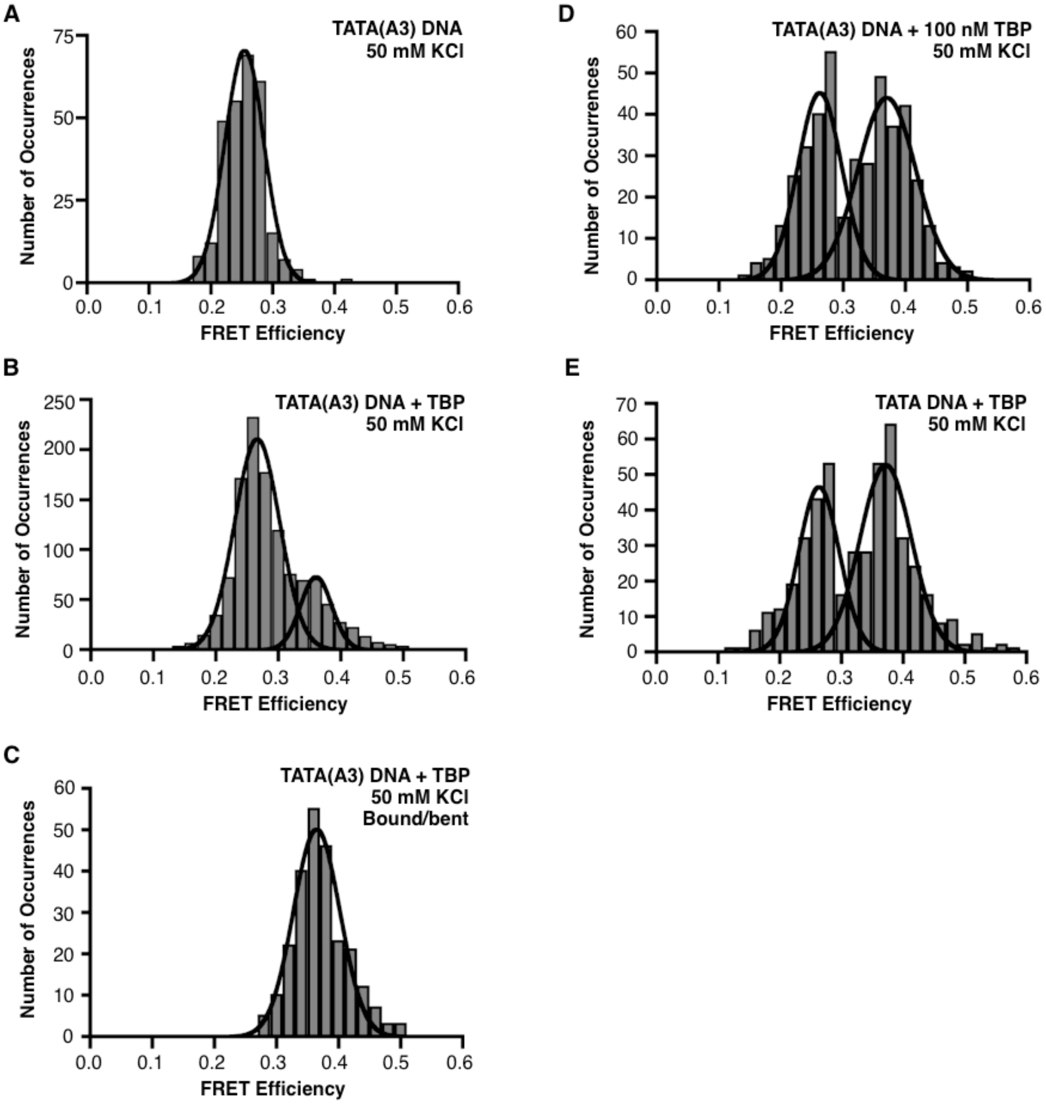


Figure 3.6. TBP bends the TATA(A3) sequence to a similar extent as the consensus TATA box. **(A)** Shown is the histogram of FRET efficiencies for 282 TATA(A3) molecules in the absence of TBP at 50 mM KCl; the mean FRET is 0.25. **(B)** TBP shifts a portion of the TATA(A3) DNA into a higher FRET state at 50 mM KCl. 1165 distinct FRET occurrences observed in time traces from 642 TATA(A3) molecules were histogrammed; the low and high FRET states have mean FRET efficiencies of 0.27 and 0.36, respectively. **(C)** Bent states on the TATA(A3) DNA were histogrammed and fit to a single Gaussian; the mean FRET efficiency is 0.36. **(D)** Raising the concentration of TBP to 100 nM increases the number of bound/bent TATA(A3) DNA molecules. 421 distinct FRET occurrences observed in time traces from 141 TATA(A3) molecules were histogrammed; the low and high FRET states have mean FRET efficiencies of 0.26 and 0.37, respectively. **(E)** TBP bending the consensus TATA box at 50 mM KCl. The experiment was performed with 10 nM TBP. 467 distinct FRET occurrences observed in time traces from 244 TATA molecules were histogrammed; the low and high FRET states have mean FRET efficiencies of 0.26 and 0.37, respectively.

experiment at a 4-fold higher concentration of TBP, which we expected would increase the fraction of TATA(A3) DNA molecules occupied by TBP. As shown in Figure 3.6D, this resulted in a shift of the population toward the bound/bent state, which exhibited a mean FRET efficiency of 0.37. To enable a more direct comparison to the consensus TATA box, we monitored TBP bending the consensus TATA construct at 50 mM KCl, which yielded a mean FRET efficiency for the bound/bent states of 0.37 (Figure 3.6E). At 50 mM KCl, the mean FRET efficiency values for the bent state on the TATA(A3) construct (0.36 and 0.37) and the consensus TATA DNA (0.37) are indistinguishable. Taken together, our results indicate that the extent of bending of the TATA(A3) construct is the same as that of the consensus TATA element, although it remains possible that the two constructs exist in different bent conformations that result in similar FRET values.

TBP bends consensus TATA DNA, but not the TATA(A3) mutant DNA, in two kinetically distinct complexes bent to the same extent

We observed that many DNA molecules that switched between unbent and bent FRET states, which enabled us to investigate the kinetics with which TBP bends TATA DNA as well as the kinetics with which bent complexes revert back to unbent DNA. Figure 3.7A shows a representative time trace of a DNA molecule switching between low FRET (unbent) and higher FRET (bent) states; the donor (green), acceptor (red), and total (gold) fluorescence are shown on the top plot and the FRET efficiency (blue) and FRET state fit (black line) are shown in the bottom plot. To evaluate the kinetics of TBP-DNA bending, we histogrammed the dwell times for unbent states that were flanked by bent states on the consensus TATA box in the presence of

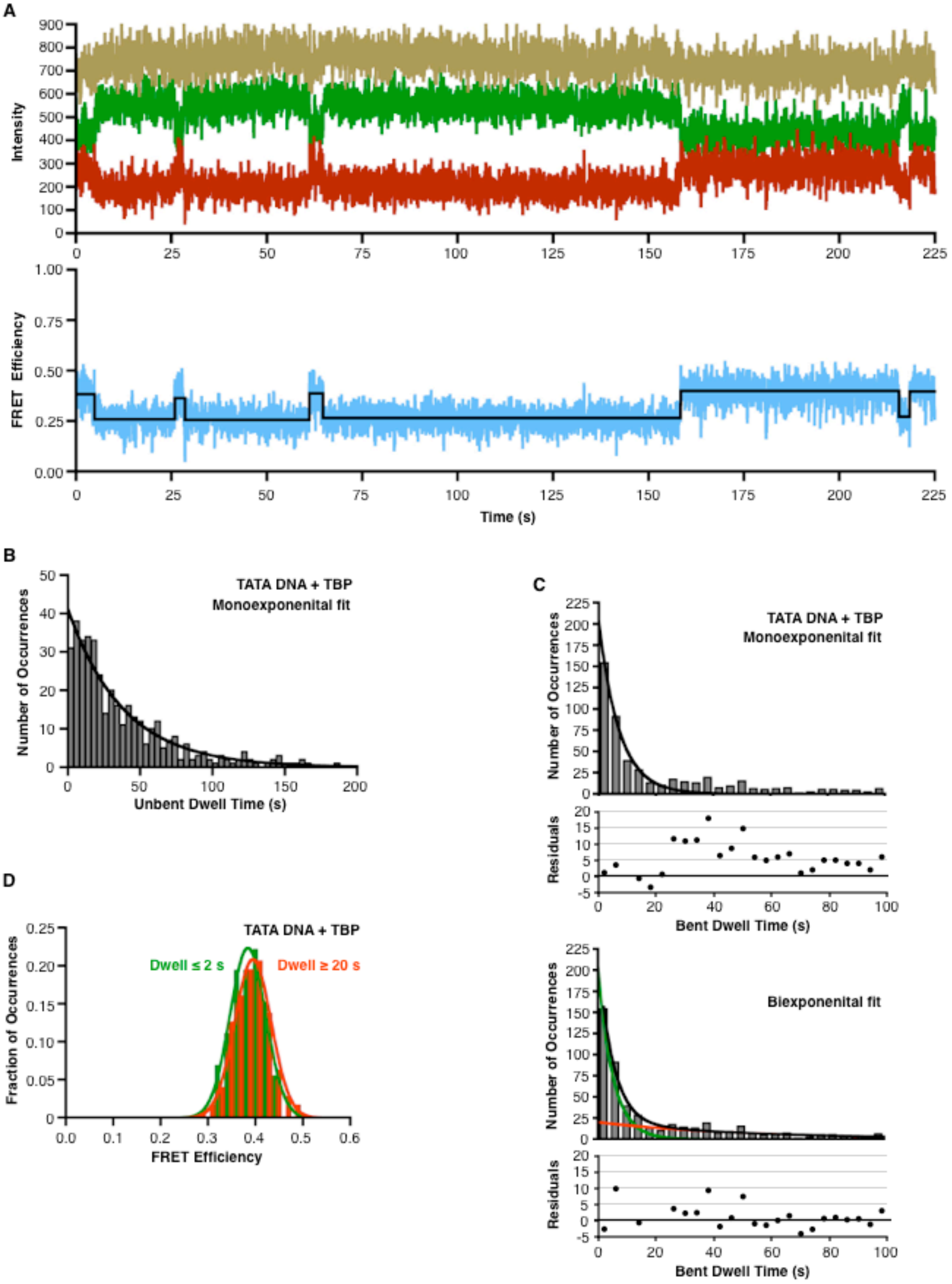

A
Intensity
FRET Efficiency
Time (s)
B
TATA DNA + TBP
Monoexponenital fit
Number of Occurrences
Unbent Dwell Time (s)
C
TATA DNA + TBP
Monoexponenital fit
Number of Occurrences
Residuals
Bent Dwell Time (s)
Biexponenital fit
Number of Occurrences
Residuals
Bent Dwell Time (s)
D
TATA DNA + TBP
Dwell ≤ 2 s
Dwell ≥ 20 s
Fraction of Occurrences
FRET Efficiency

Figure 3.7. Unbending consensus TATA DNA is biphasic. **(A)** Representative time trace of a single DNA molecule switching between low FRET (unbent) and higher FRET (bent) states. The top plot shows the donor emission in green, the acceptor in red, and the sum of the donor and acceptor in gold. The bottom plot shows the corresponding FRET in blue, with the black line indicating the FRET state fit of the data. For the purpose of display, these data were smoothed by averaging the values at 3 time points. **(B)** Analysis of the rate at which TBP bends the consensus TATA box. Dwell times for the unbent occurrences observed at 150 mM KCl were histogrammed and fit to a monoexponential. The observed rate constant is 0.026 ± 0.003 s^{-1}, where the error represents the 95% confidence interval. **(C)** Unbending of the consensus TATA box occurs in two distinct kinetic phases. Dwell times for the bent occurrences observed at 150 mM KCl were histogrammed and fit to a monoexponential (upper panel) or a biexponential (lower panel). Residuals are plotted for each of the fits. The two rate constants obtained from the biexponential fit are 0.19 ± 0.03 s^{-1} and 0.019 ± 0.008 s^{-1}, where the error represents the 95% confidence interval **(D)** TATA DNA molecules in complexes with different kinetic stabilities are bent to the same extent. FRET efficiencies were separately histogrammed for bent occurrences having dwell times ≤ 2 s (green) or ≥ 20 s (red). Gaussian fits showed mean FRET efficiencies of 0.38 and 0.39, respectively.

TBP at 150 mM KCl (for example, the four low FRET states in Figure 3.7A). A monoexponential fit of the data yielded an observed rate constant of 0.026 s^{-1} (Figure 3.7B). We compare this value, as well as the values of the other rate constants we measure by smFRET, to published data in the Discussion.

To investigate the kinetics of TBP/DNA complexes that revert from the bent state to unbent DNA, we histogrammed the dwell times for bent states that were flanked by unbent states. The histogram was fit to a monoexponential, however all of the bins with dwell times longer than 25 s were above the curve (Figure 3.7C). This is apparent in the plot of residuals below the histogram. For this reason we fit the histogram of bent state dwell times to a biexponential (Figure 3.7C). The black curve shows the biexponential fit, the green and red curves show the respective single exponential decay curves for the fast and slow populations generated using the rate constants obtained from the biexponential fit (0.19 s^{-1} and 0.019 s^{-1}, respectively). The biexponential fit of the data was much better as indicated by the lack of a directional trend in the residuals (plotted below the histogram). This indicates at least two kinetically distinct bent TBP/TATA complexes exist that revert back to unbent DNA with different rates. Moreover, single DNA molecules exhibited both the faster and the slower rates of unbending (see Figure 3.7A). To investigate the extent of bending in these kinetically distinct populations, we independently histogrammed the FRET efficiencies of less stable bent complexes (dwell time for the bent state $\leq$ 2 s) and more stable bent complexes (dwell time for the bent state $\geq$ 20 s). As shown in Figure 3.7D, the two histograms have very similar mean FRET efficiencies, 0.38 for the less stable complexes and 0.39 for the more stable complexes. Therefore, the DNA in these kinetically different complexes is bent to the same extent.

We also analyzed the kinetics of bending and unbending for complexes formed on the mutant TATA(A3) DNA. The rate of DNA bending was assessed by histogramming the unbent state dwell times and fitting to a monoexponential, which yielded an observed rate constant of 0.12 s^{-1} (Figure 3.8A). To assess the kinetics with which the bent state decays on the TATA(A3) construct, the bent state dwell times were histogrammed. Surprisingly, the bent state dwell times for the TBP/TATA(A3) DNA complex fit well to a monoexponential (Figure 3.8B), yielding a rate constant of 0.31 s^{-1}. Therefore, bent complexes on the mutant TATA DNA decay with a different kinetic profile compared to bent complexes on the consensus TATA DNA.

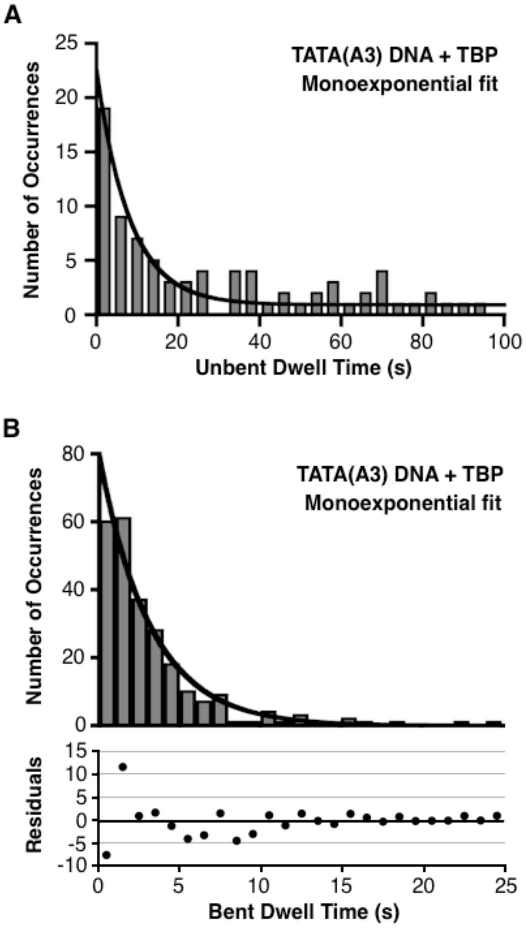


Figure 3.8. Unbending on the TATA(A3) DNA in monophasic. **(A)** Analysis of the rate at which TBP bends the TATA(A3) construct. Dwell times for the unbent occurrences observed at 50 mM KCl were histogrammed and fit to a monoexponential. The observed rate constant is 0.12 ± 0.03 s^{-1}, where the error represents the 95% confidence interval **(B)** Unbending of the TATA(A3) construct. Dwell times for the bent occurrences observed at 50 mM KCl were histogrammed and fit to a monoexponential. Residuals are plotted below the rate plot. The rate constant obtained is 0.31 ± 0.04 s^{-1}, where the error represents the 95% confidence interval**.**

Discussion

Here we used smFRET to study the heterogeneity and kinetics of DNA bending by human TBP. We found that with either a consensus or mutant TATA box the populations of DNA molecules bound by TBP were bent to a uniform FRET state in both the absence and presence of TFIIA. Throughout our studies, we did not observe populations of TBP/DNA complexes with intermediate bent states, consistent with a mechanism for TBP-DNA binding that lacks partially bent intermediates. Our kinetic analyses suggest that the mechanism of TBP-DNA bending differs on the consensus and mutant TATA boxes. Together our single molecule studies led to several unexpected observations: 1) mutating the TATA box did not significantly decrease the extent of bending, as was previously concluded from ensemble experiments, 2) although TFIIA profoundly increased the affinity of TBP for DNA, it did not change the extent of DNA bending, as was previously concluded from ensemble experiments, 3) for the consensus TATA box, the transition from bent DNA to unbent DNA was biphasic, whereas that from the mutant TATA(A3) construct was monophasic. Together our smFRET studies provide new insight into the mechanism of DNA bending by TBP that is distinct from earlier ensemble experiments and lead us to a revised model for TBP-induced DNA bending. Hence, our work contributes to an ever-growing list of systems for which single-molecule techniques have resolved heterogeneous kinetics and conformational states (93-95).

Populations of TBP/DNA complexes exist in a uniform bent state

Within the time and FRET resolution of our experiments, our data show that TBP bends consensus and non-consensus TATA DNA uniformly in the presence and absence of TFIIA.

Under each of the experimental conditions containing TBP, we observed only one subpopulation with a FRET state consistent with bent DNA. This suggests a high degree of homogeneity in the conformation of TBP/TATA complexes regardless of DNA sequence or the presence of TFIIA. It is possible that different DNA conformations within TBP-bound complexes give rise to the same FRET state if the twist in the DNA is considered. We have modeled and discussed this concept in a previous study that used ensemble FRET to monitor TBP-induced DNA bending, and found that many combinations of bend and twist angles can give rise to a given FRET value (29). We also used a crystal structure of the human TBP/TATA DNA complex to create a model of an 18 bp TBP-bent DNA with a centered TATA box (85). This showed that the distance between the 5' carbons on the ends of the bent DNA is 54.7 Å. The distance between the fluorophores in the bent conformation calculated from our smFRET data is 55.3 Å. Hence, our data suggest that the bend we observe in smFRET is similar to that observed in the crystal structure; however, the distance calculated from our smFRET values is only an estimate due to underlying assumptions in the orientation and environment of the dyes. It is also possible that TBP can bind DNA prior to bending the TATA box (88, 89), which would be resolved in future smFRET experiments involving 3 fluorophores, two labeling the DNA (to monitor bending by FRET) and one on TBP (to monitor DNA binding).

In a general sense, our results are consistent with a model developed by Parkhurst, Brenowitz, and colleagues in which TBP simultaneously binds and bends TATA DNA. In their experiments using yeast TBP, complexes containing partially bent DNA were not observed (90, 91). By contrast, single molecule optical tweezer experiments of DNA bending by yeast TBP led to the conclusion that a partially bent, long-lived intermediate forms en route to the fully bent state (92). We do not observe a distinct third FRET state between the unbent and bent states in the histograms of our smFRET data with human TBP as would be predicted from the optical tweezer study. Hence, our data best support a model in which TBP uniformly bends DNA via a mechanism that does not involve a partially bent intermediate.

Mutating the TATA box strongly reduces the affinity of TBP DNA binding, but does not affect the extent of DNA bending

Our smFRET data show that TBP binds and bends TATA(A3) DNA with a much reduced affinity compared to the consensus sequence, which corroborates data from ensemble biochemical experiments (26, 29). Indeed, we did not observe TATA(A3) DNA molecules in the higher FRET state (i.e. bent) in the presence of TBP under our standard conditions (Figure 3.5B). We did, however, observe TATA(A3) molecules in the higher FRET state when either TFIIA was included with TBP (Figure 3.5C) or the KCl concentration was reduced (Figures 3.6B, 3.6C, and 3.6D). Both of these observations are consistent with previous ensemble experiments showing that TBP/DNA complexes are more stable at lower salt concentrations and/or in the presence of TFIIA (29).

Our smFRET results lead us to conclude that TBP bends the TATA(A3) mutant DNA to the same extent as the consensus sequence (smFRET states of 0.36-0.37 and 0.37 at 50 mM KCl, respectively). This similarity was not expected based on previous ensemble studies that compared DNA bending by human and yeast TBP on these two sequences (26, 28, 29). Given the change in bending reported in the ensemble studies and our measured smFRET efficiency of 0.37 for the bent consensus TATA box, we expected to observe an smFRET state of approximately

0.29 for the TATA(A3) mutant. Instead we clearly observed a bent population with a FRET state of 0.36-0.37, therefore, the mutant TATA(A3) DNA can be far more bent than would be predicted from ensemble studies. We cannot, however, rule out that a weakly populated 0.29 FRET state exists since this would be difficult to resolve from the unbent DNA. The inconsistency between smFRET and ensemble studies with human TBP is likely due to the weak binding affinity of TBP for the TATA(A3) sequence. In the previous ensemble FRET studies with human TBP (29), if the TATA(A3) DNA was not fully occupied, then the measured FRET would have been an average of the FRET from free DNA and bent DNA in the population, resulting in a lower calculated bend angle for the population. Ensemble studies with yeast TBP have also concluded that TBP bends TATA(A3) DNA significantly less than consensus TATA DNA (26, 28). It will be interesting to further investigate DNA binding by yeast TBP on consensus and mutant TATA boxes using single molecule approaches.

Many studies have shown that mutating the TATA sequence away from consensus results in decreased affinity for binding TBP as well as a reduction in transcriptional activity both in vitro and in cells (26, 28, 33). Some ensemble studies have concluded that decreased DNA bending by TBP correlates with decreased affinity and decreased transcriptional activity (26, 31). Our smFRET results show that there in no discernible difference in bending between the consensus and TATA(A3) constructs although there is a substantial difference in affinity, hence the degree of bending does not strongly correlate with binding affinity. This is consistent with crystal structures in which TBP induced the same conformational change in 10 different naturally occurring TATA box variants, most of which reduced transcriptional activity (96).

TFIIA does not substantially change the extent to which TATA DNA is bent

Our smFRET experiments showed that the fraction of DNA molecules occupying the TBP-induced higher FRET state increased in the presence of TFIIA for both the consensus and mutant TATA(A3) constructs, indicating that TFIIA facilitates TBP binding to and bending DNA. This is consistent with previous data showing that TFIIA increases the affinity and kinetic stability with which TBP interacts with DNA (22, 29, 34, 35). Although TFIIA increased the portion of DNA molecules in the higher FRET state, it did not appreciably change the mean FRET efficiency of the TBP-bent state on either the consensus or TATA(A3) constructs. Therefore, in this system TFIIA does not alter the extent, or angle, at which DNA is bent by TBP, which is consistent with crystal structures of the human and yeast TFIIA/TBP/TATA and TBP/TATA complexes (39, 97, 98). Our previous ensemble FRET studies indicated that TFIIA decreased the FRET efficiency of TBP-bent consensus TATA DNA, which reflected either a change in the angle at which TBP bent the DNA within the TATA box, the angle at which the DNA emerging from the TBP/DNA complex twisted, or a combination of changes in both bend and twist angles (29). We performed ensemble experiments using smFRET conditions, and observed no effect of TFIIA on the FRET efficiency of TBP-bent DNA. It is possible that contacts between TFIIA and upstream DNA in the 38 bp smFRET construct restrict changes in the bend or twist angle that could occur with the 14 bp DNA construct used in the original study.

smFRET studies of the kinetics of bending and unbending reveal new aspects of the mechanism of DNA binding by human TBP

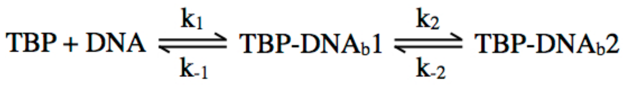
Our smFRET experiments allowed us to investigate the kinetics with which TBP bends DNA as well as the kinetics with which bent DNA in complexes reverts back to unbent DNA. For the consensus TATA box in the presence of 25 nM TBP, a histogram of the dwell times of unbent DNA occurrences fit nicely with a monoexponential, yielding a pseudo first order rate constant of 0.026 s^{-1}. Taking into account the TBP concentration provides a calculated second order rate constant of 1 x 10^6 $M^{-1}s^{-1}$, which is consistent with previous studies that have measured rates of TBP/DNA association that are well below the diffusion limit (30, 87, 89, 99, 100). A histogram of the dwell times of bent DNA occurrences did not fit nicely with a monoexponential, but instead was fit with a biexponential yielding rate constants of 0.19 s^{-1} and 0.019 s^{-1}. The second of these two rate constants compares well with our previous measurement of the rate of unbending using ensemble FRET (0.013 s^{-1}, (29)). Observing biphasic decay in our smFRET data shows that two populations of TBP/consensus TATA complexes formed, having distinct kinetic stabilities. The DNA in these complexes was bent to the same extent. We used the bent and unbent dwell times to calculate an apparent K_D for the TBP/DNA interaction, presuming binding and bending are concerted. For the data collected at 50 mM KCl (Figure 3.6E) we obtain an ${}_{app}K_D$ of 6 nM, and for the data collected at 150 mM KCl (Figure 3.2B) we obtain an ${}_{app}K_D$ of 78 nM. Both the ${}_{app}K_D$ values and the 13-fold difference in affinity due to salt concentration are consistent with previous measurements for yeast TBP (30).

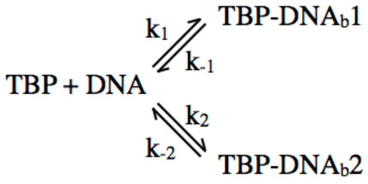
When we analyzed the kinetics on the TATA(A3) sequence we found that histograms of both the dwell times of unbent occurrences and bent occurrences were nicely fit with monoexponentials, yielding rate constants of 0.12 s^{-1} and 0.31 s^{-1}, respectively. The affinity of TBP binding to the TATA(A3) DNA is much lower than binding to the consensus TATA box (26, 28). Our data suggest that this is largely due to the mutation causing a dramatic decrease in the kinetic stability of the bent DNA complexes, as opposed to an effect on the rate of bending. The observed rate of bending on the TATA(A3) DNA was faster than what we observed on the consensus TATA DNA. We believe that the difference in the rates of bending likely reflects salt-dependent differences in the rates of binding; the unbent dwell times on the mutant DNA were measured at 50 mM KCl and on the consensus DNA at 150 mM KCl. Previous kinetic studies of the TBP/DNA interaction have shown that the rate of association increases as the salt concentration decreases (30).

We were unable to determine rate constants from data obtained in the presence of TFIIA on both the consensus and TATA(A3) constructs due to the fact that: 1) we did not observe a population with a low FRET state, hence we were unable to obtain dwell times for unbent molecules in the presence of TBP, and 2) high FRET occurrences photobleached prior to transitioning back to a low FRET state, hence we could not obtain dwell times for bent molecules. The rate constant for photobleaching in our system is approximately 0.004 s^{-1}, which provides an upper limit for the rate constant for unbending in the presence of TFIIA on the consensus and TATA(A3) constructs, which is consistent with previous measurements (29).

Our kinetic analyses lead us to reconsider models for DNA binding and bending by TBP. Previous studies concluded from global analysis of ensemble biochemical data that human TBP binds and bends the AdMLP TATA box via a three step linear mechanism (31). In this model, the two intermediates and the final bent complex all contain DNA bent to the same extent. We observe two kinetically distinct populations of complexes containing similarly bent DNA; however, a third kinetically distinct population was not observed in our smFRET studies. In the published model, the first bent intermediate is predicted to form within seconds and dissociate 7 times more rapidly than the rate at which it moves forward to the second intermediate. In a smFRET experiment this would predict that a given DNA molecule would undergo multiple relatively rapid bending/unbending transitions (dwell time of the bent state would be less than 1 s) prior to transitioning into a stable bent state. We do not observe this in our smFRET traces with the AdMLP consensus TATA box, even though our time resolution (30 ms) would allow us to observe these rapid transitions if they occurred.

Our kinetic data with the AdMLP consensus TATA box can be approximated by either of the following two bent state models.





In both the linear model and branched model the two DNA bound complexes (TBP-DNA$_b$1 and TBP-DNA$_b$2) contain DNA bent to the same extent, but the formation and dissociation of the two complexes are governed by distinct rate constants. It remains possible that other steps exist that occur too rapidly to observe in our system (e.g. states with dwell times less than 30 ms). Because the FRET states for both kinetically distinct complexes are the same, our studies do not allow us to distinguish between these two models. The measurement of two rate constants for unbending does not preclude one model over the other, and we cannot say with certainty that either of the apparent rate constants we measure is k_{-1} or k_{-2}. Previous studies allow for speculation on the structural or mechanistic basis for observing two kinetic populations. For example, TBP binds the consensus TATA box in two orientations (101); it is possible that the two kinetic populations represent TBP bound in two different orientations with the bend induced in each orientation being the same. It is also possible multiple populations allow for increased specificity in the binding reaction, with one reflecting a less-stable "check-point" and the other reflecting a more-stable final complex. Although in this scenario, the less-stable complex would be bent to the same extent as the final complex.

With the TATA(A3) construct we did not observe biphasic decay of bent complexes. It is possible that k_2 is sharply decreased on the mutant construct, therefore in the linear model the transition from the first bent complex to the second would not occur, and in the branched model the second complex would not form to a significant extent. Alternatively, a second population of complexes might be present on the TATA(A3) construct, but the formation and dissociation of these complexes occurs too rapidly to resolve in our system. Finally, for both the consensus and TATA(A3) constructs in the presence of TFIIA the complexes we observe were kinetically very stable. If two kinetically distinct populations exist in the presence of TFIIA they both must have average dwell times that are much longer than the rate of photobleaching in our system (half time for decay of over 180 s). Alternatively, TFIIA could force complexes into a single kinetic state by pushing the reaction to the second complex in the linear pathway (e.g. decreasing k_{-1} or k_{-2}) or to the more stable complex in the branched pathway.

Experimental Procedures

Oligonucleotides, proteins, and other reagents

Three oligos labeled at their 5’ ends with Alexa657 (IDT), Alexa555 (Invitrogen), or biotin (IDT) were HPLC purified. The sequences (5' to 3') of the acceptor, donor, and biotin linker oligos were Alexa647-CGTCCNNNNNNNNNCCTGCACCACCACCACCACCA, Alexa555-CAGGNNNNNNNNNGGACG, and Biotin(dT)-TGGTGGTGGTGGTGGTGGT, respectively, where the specific sequences of the variable regions (indicated with N) are shown in Table 3.1. Constructs containing a donor oligo, an acceptor oligo, and the biotin-linker oligo were assembled by heating to 95°C for 3 min, then incubating the DNA at 60°C for 45 min and cooling 0.1°C/s to room temperature prior to use. Recombinant human TBP and human TFIIA were expressed and purified from *E. coli* as previously described (102-104). Fluorescence anisotropy values were measured for DNA constructs containing Alexa555 or Alexa647 in the absence and presence of TBP under the experimental conditions used for smFRET. In all cases the anisotropy values were between 0.21 and 0.24.

Catalase (Sigma) and glucose oxidase (Sigma) were resuspended to 6.8 mg/ml and 44 mg/ml, respectively, in storage buffer (50% (v/v) glycerol, 100 mM Tris (pH 7.9), and 50 mM KCl), and stored at -20°C. Streptavidin (1 mg/ml, Sigma) was stored at -20°C in 10 mM Tris (pH 7.9), 10% (v/v) glycerol, 50 mM KCl, and 5 mM $MgCl_2$. D-glucose 10% (w/v) was dissolved in 10 mM Tris (pH 7.9) and 50 mM NaCl, and stored at -20°C. Trolox (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid, Sigma) was dissolved in 18 MΩ water to a concentration of 100 mM by adding 4 M NaOH. Trolox was stored at 4°C for no longer than one week.

Preparation of surfaces for smFRET experiments

Glass slides and coverslips (VWR) were cleaned by the following procedure with thorough rinsing in 18 MΩ water between each step. A 1% alconox solution was heated to 90°C and added to a container of glass slides and coverslips and sonicated for 20 min. The slides and coverslips were then sonicated in ethanol for one hour, followed by a 20 min sonication in 1 M KOH. The slides and coverslips were dried by rinsing in ethanol and microwaving for 90 s. Slides and coverslips were flamed over a Bunsen burner and after cooling were silanated by placing them in methanol containing 2% aminosilane (N-(2-Aminoethyl) 3-aminopropyltrimethoxysilane, UCT) for 21 min, with a 1 min sonication after 10 min. Slides and coverslips were rinsed in methanol and baked at 110°C for 5-10 min. A solution of 0.38% (w/v) biotin-PEG-SC (MW 5000, Laysan Bio, Inc.) and 20% (w/v) mPEG-SVA (MW 5000, Laysan Bio, Inc.) in 0.1 M sodium bicarbonate was sandwiched between a coverslip and slide and allowed to react with the silanated glass for 3.5-4 h in a humidified environment (105). The surfaces were rinsed thoroughly with water and dried with nitrogen gas. Reaction chambers were assembled by attaching a coverslip to a slide with an overhang using double-sided tape. Sample chambers were stored in a dark, dry environment until use.

smFRET reaction conditions

A sample chamber was incubated with 0.02 mg/ml streptavidin for 5 min in buffer A (10% (v/v) glycerol, 10 mM Tris (pH 7.9), 50 mM KCl, 1 mM DTT, 0.05 mg/ml BSA, 10 mM HEPES (pH 7.9), 5 mM $MgCl_2$, and 0.1% nonidet P-40). The chamber was washed with excess buffer A prior to flowing in a solution containing 20-40 pM of a 3-oligo construct in buffer A.

After 10 min, the chamber was washed with excess buffer A. For all fluorescent measurements with or without proteins, imaging solution was flowed into a chamber, which contained 12% (v/v) glycerol, 14 mM Tris (pH 7.9), 1 mM DTT, 0.05 mg/ml BSA, 10 mM HEPES (pH 7.9), 5 mM $MgCl_2$, 9 mM NaCl, 1 mg/ml glucose oxidase, 11 μg/ml catalase, 0.8% (w/v) D-glucose, 2 mM trolox, 0.1% nonidet P-40, and 150 mM KCl, unless stated otherwise. All smFRET experiments were performed at 22°C. Unless otherwise noted in the text, TBP was at 25 nM and TFIIA was at 15 nM.

Data collection and analysis

FRET data were collected on an objective-based total internal reflection fluorescence microscope (Nikon TE-2000U) equipped with a Cascade II CCD camera, a 1.49 NA immersion objective, and a 532 nm CW laser. Acceptor and donor emission data were collected with MetaMorph software (Molecular Devices) using a 30 ms timescale and a bin of 2.

Movies were analyzed using in-house developed software to identify spot pairs in the donor and acceptor channels. To obtain time traces of donor and acceptor emissions for individual spots, background was subtracted from each frame using an area neighboring the spot. The following equation was used to correct for bleed through of the donor dye into the acceptor channel.

$$A_{corrected} = A - \alpha * D$$

Where A is the background-subtracted signal from the acceptor channel, D is the background-subtracted signal from the donor channel, and α is a factor used to correct for bleedthrough of the donor emission into the acceptor channel; α was 0.14. The ratio of the donor and acceptor detection efficiencies (γ) was one, and the acceptor emission resulting from direct excitation of the acceptor by the donor laser was approximately 3%; hence, data were not corrected for these two factors. The Alexa555/Alexa647 has a calculated R_0 of 51 Å, assuming $\kappa^2 = 2/3$. Apparent FRET efficiencies (E) were calculated using the equation:

$$E = A_{corrected}/(A_{corrected} + D)$$

Unsmoothed data were further analyzed using in-house developed software to determine the FRET states and dwell times of each state present throughout individual time traces. Blocks of frames exhibiting photoblinking were removed from the data. FRET states lasting fewer than 3 frames were removed from further analysis. Under our experimental conditions, a minimum change in FRET of 0.05 was required to detect a change between two neighboring states in a time trace. We continuously monitored the sum of the donor and acceptor emissions as a means to assess the effect of TBP and TFIIA on the donor and acceptor fluorophores; this sum remained relatively constant during bending and unbending events.

FRET states were extracted from all data collected under a single condition and binned at 0.02 FRET units to generate the histograms. Gaussian curves were fit to histograms using Prism to obtain mean FRET efficiencies. Dwell times were extracted from the time traces. Time traces that showed multiple state changes on a single DNA molecule were separated into the unbent (lower FRET) and bent (higher FRET) states, and dwell times were extracted for each state. Dwell times were removed for states at the beginning and end of time traces (which included molecules that did not change states throughout the entire time trace), and when a dye blinked during a state. Dwell times for unbent and bent states were binned separately, histogrammed, and fit to monoexponential or biexponential equations using Prism, as noted in the Results section.

CHAPTER 4

The bending of linear DNA by full length HMGB1 is mediated through one HMG domain

Summary

HMGB1 is a highly abundant, dynamic protein that interacts with chromatin to help regulate transcription. HMGB1 has 2 DNA binding domains, an A box and a B box, and a C-terminal acidic tail. Using single molecule FRET, we studied full length HMGB1 and truncations of the protein to determine how its domains bend DNA. We found that the full length HMGB1 protein bends DNA to a similar extent as the A box and B box individually. An HMGB1 construct lacking the C-terminal tail (AB protein), however, bends the DNA to a greater extent than the full length protein or the individual DNA binding domains. This suggests that the AB protein can simultaneously use both the A and B boxes to bend DNA, whereas the full length protein uses only one DNA binding domain. Our data support a model in which the C-terminal tail blocks simultaneous DNA bending by the A and B boxes in the context of the full length protein.

Introduction

High mobility group box protein 1 (HMGB1) has broad functions intra- and extra-cellularly, where it is involved in DNA repair, the immune response to inflammation, and transcriptional regulation (40, 45, 47, 55). HMGB1 can upregulate transcription through nucleosome remodeling by interacting with histone H1 and H3, causing the chromatin to become less compact (50-52). It is very abundant with about 1 million molecules in the nucleus, equating to approximately 1 HMGB1 molecule for every 10 nucleosomes (41). HMGB1 also activates transcription by interacting with specific transcription factors to facilitate binding to their DNA recognition sites (54). HMGB1, itself, binds DNA in a sequence-independent manner, bending the DNA as it binds. This DNA binding and bending is thought to be very dynamic, where the protein can rapidly bind and release from DNA (106-108). Within cells, HMGB1 has been shown to cross the nucleus in a few seconds (49). How exactly HMGB1 binds and bends DNA and interacts with other proteins is not fully understood.

HMGB1 is composed of 3 domains, an A box and a B box, both of which can bind and bend DNA independently, and a C-terminal acidic tail. The A and B boxes are similar in overall structure; both have 3 α-helices arranged in an L-shape that is bent about 70-80 degrees (109-111). However, the A and B boxes interact with DNA differently since their tertiary structures have slightly different shapes and orientations, and the residues that intercalate into DNA to induce a bend are different (112). The A box has a phenylalanine that intercalates into DNA, and deletion of this residue eliminates bending (44). By contrast, the B box has 2 intercalating residues, a different phenylalanine and an isoleucine (45), creating a different mechanism of bending by the two boxes. The A box has a higher binding affinity than the B box for distorted DNA structures, such as cisplatinated DNA and 4-way junctions (113), but the B box has been found to bend the DNA to a greater extent than the A box (114). ITC and NMR studies showed that the A and B domains do not interact with each other in the context of the full length protein, so they are completely independent domains (115, 116). Although both boxes can bind and bend DNA, one model for how HMGB1 functions to control transcription is that one of the boxes interacts with a transcription factor, and the other box interacts with DNA to facilitate transcription factor binding (54).

The C-terminal acidic tail is very important for regulating HMGB1 function. The C-terminal tail can interact with the A and B box DNA binding domains, therefore limiting the binding of HMGB1 to DNA (62, 116-118). One NMR study showed that the A and B boxes can sandwich the C-terminal tail such that the DNA binding domains of the boxes are only transiently exposed this conformation is dynamic so the boxes can become available to interact with DNA (117). When truncations are made in the tail region such that the number of acidic residues in the C-terminal domain decreases, the protein has a higher affinity for DNA (116), presumably because less of the tail can interact with the DNA binding sites in the A and B box.

In vitro binding studies reveal a wide range of affinities for HMGB1 binding to DNA (56, 59, 60, 62, 113, 119-123). Stros et. al. found that the affinity of full length HMGB1 for DNA is in the low nanomolar range (123). Removal of the C-terminal tail can yield a higher binding affinity for linear DNA compared to the full length protein; the magnitude of the effect varies from slight to a 10-fold change, depending on experimental conditions (113, 120-123) The A box and B box individually have lower binding affinities for DNA than the full length and tailless proteins (123). Moreover, there are two basic linker regions in HMGB1 (one between the A and

B box and one between the B box and the C-terminal tail) that appear to increase the binding affinity of HMGB1 for DNA because they interact with the phosphate backbone (54, 62, 121, 124, 125). For example, the binding affinity of the A box for DNA increased approximately 70-fold with addition of the basic linker region that is C-terminal of the A box. Likewise, incorporating the two basic regions that flank the B box increased the binding affinity approximately 20-fold (120, 121).

The mechanism by which full length HMGB1 binds and bends DNA remains unclear. Different studies report that HMGB1 bends DNA using different boxes. For example, one group reported that only the A box is used to bend DNA when full length HMGB1 binds (62); another group hypothesized that the B box is responsible for DNA bending (54). Moreover, a crystal structure of a tailless dual box protein shows that both boxes interact with DNA simultaneously (126). Although HMGB1 binding DNA has been studied with various biochemical techniques, there is still no clear mechanistic picture of how it bends DNA. To provide insight into this process, we used single molecule FRET (smFRET) to study DNA bending by full length HMGB1 and truncated proteins containing various combinations of its domains.

Results

Expression and purification of full length HMGB1 and truncations

Since we wanted to dissect how HMGB1 bends and interacts with DNA, we cloned, expressed, and purified the full length (FL) human protein, as well as various truncations. A schematic of the different truncations used in these studies is shown in Figure 4.1. All cloning was done using the pET19b vector, and each construct contained a His tag. The natural NdeI site in the HMGB1 cDNA sequence was removed by site-directed mutatgenesis, but the amino acid sequence was not altered. Proteins were expressed and purified as described in the Experimental Procedures. Importantly, recombinant HMGB1 had to be expressed in the presence of glucose and lysozyme to prevent leaky expression, because the full length protein appeared toxic to the *E. coli*.

Figure 4.2A shows an SDS gel with fractions from a representative purification of the full length construct, in which the soluble cell lysate was purified over a nickel resin. The expressed protein was depleted, as evident by the flow through sample. HMGB1 was eluted from the nickel column using 250 mM imidazole. For the full length protein, the eluate was passed over a Q column, and the flow through from the Q column was further purified using a dsDNA cellulose column. For the AB, A, and B constructs, the eluate from the nickel column was dialyzed, and placed onto the dsDNA cellulose column, and the protein was eluted with high salt. The BC construct was only purified over the nickel column. The experimental details for all purifications are described in the Experimental Procedures. As seen in Figure 4.2B, each protein was highly purified.

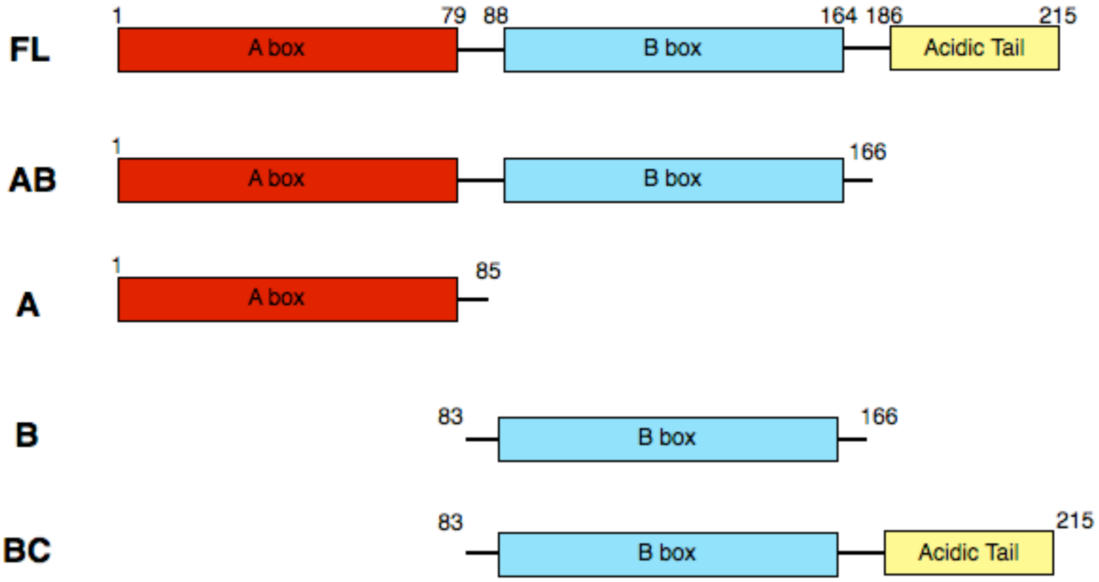


Figure 4.1. Full length HMGB1 (FL) and truncations of human HMGB1. This includes the tailless AB truncation (AB), the A box (A), the B box (B), and the B box with the C-terminal tail (BC). The amino acid numbers associated with each truncation are indicated.

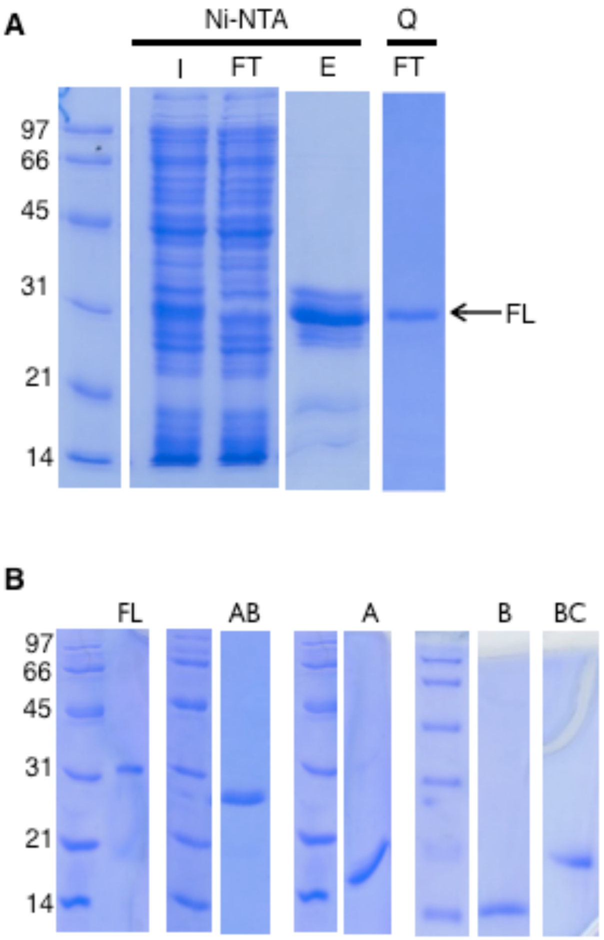


Figure 4.2. Purification of HMGB1 constructs. **(A)** SDS-PAGE analysis of the purification strategy for FL HMGB1. The arrow to the right designates where the full length protein runs on the gel. The soluble cell lysate was flowed over a Ni-NTA column. The input is designated I. The flow through (FT) is shown to be depleted of the expressed protein. The protein was eluted (E) off the nickel column using imidazole. The eluate was then placed over a Q column, and the FT was collected. This was then placed over a dsDNA cellulose column for further purification. The final purified protein is shown in (B). **(B)** SDS-PAGE showing the final purities of each protein construct. The molecular weights of the expressed proteins are FL (28 kDa), AB (22 kDa), A (13 kDa), B (12 kDa), BC (18 kDa).

The DNA bending activities of each of the protein constructs were assessed using ensemble FRET. An 18-bp DNA construct with a donor and acceptor dye on either 5'-end was used. FRET between the dyes was monitored as HMGB1 was titrated into solution; an increase in FRET was indicative of protein-induced DNA bending. The change in FRET was plotted versus HMGB1 concentration and the data were fit with a binding equation to determine the K_d These dissociation constants are apparent K_d values since we measured binding activity through bending of DNA, as opposed to monitoring DNA binding directly. First, we determined the affinity of FL HMGB1 using DNA constructs with different sequences, DNA4 and DNA5 (Figure 4.3). For DNA4 and DNA5, the titration curves are shown in Figure 4.4A and apparent equilibrium dissociation constants (${}_{app}K_d$) for FL protein were determined to be 1.7 and 1.9 nM, respectively (Table 4.1), indicating very little sequence specificity for DNA bending by the full length protein. The apparent dissociation constant for the AB protein bending DNA was found to be 1.5 nM. The A box and the B box were found to have ${}_{app}K_d$ values of 24 and 37 nM, respectively. These latter measurements have a significant amount of error, however, their affinities for bending DNA are notably lower than the full length protein or AB construct even when error is considered, which is consistent with published data (123). Finally, the BC construct has an apparent dissociation constant of 99 nM, well above the B box without the C-terminal tail, which is consistent with previous studies showing the C-terminal acidic tail interacts with the HMG domains, decreasing the association with DNA (62, 116-118).

To assess kinetic stability, off-rate experiments were performed using a fluorimeter to monitor the decrease in FRET over time after adding and manual mixing 2 different amounts of nonspecific competitor DNA to HMGB1/DNA complexes. We found that the off-rate constant

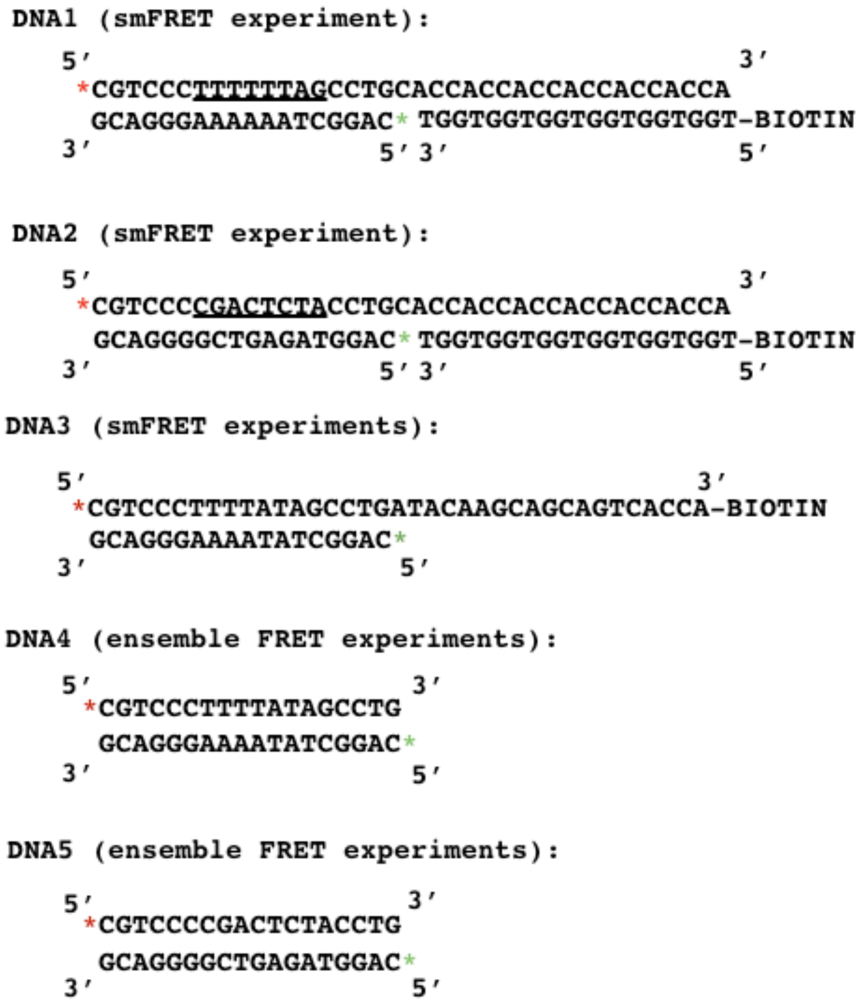


Figure 4.3. Double stranded DNA constructs used in experiments in this chapter. The acceptor fluorophore (*), Alexa 647, is on the 5' end of one oligo and the donor fluorophore (*), Alexa 555, is on the 5' end of the other oligo in each double-stranded construct. DNA1 and DNA2 were used in the sequence dependence smFRET experiments in Figure 4.4. The differences in sequence between these two constructs are underlined. DNA3 was used in all other smFRET experiments. DNA4 and DNA5 were used for ensemble FRET experiments; DNA4 is similar to DNA3, without the linker.

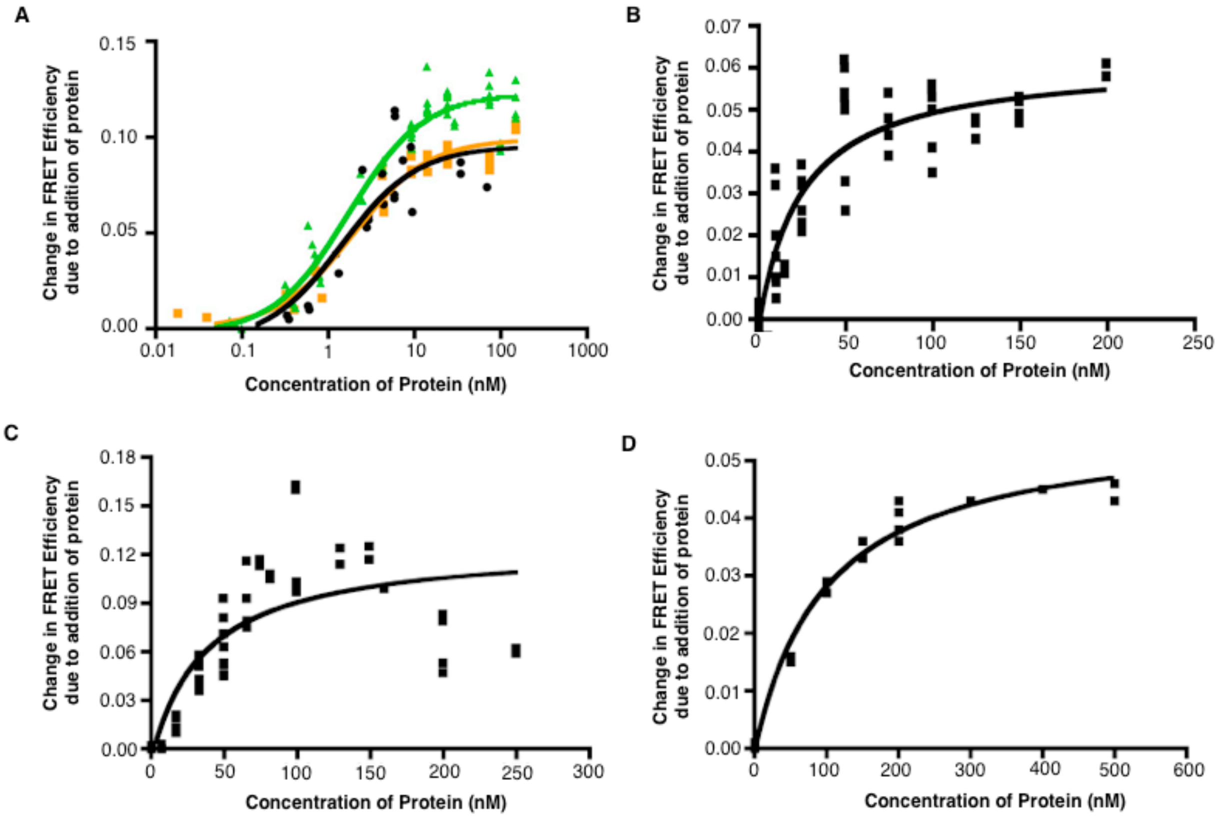


Figure 4.4. Titration curves of HMGB1 bending 18 bp DNA determined by bulk FRET efficiencies. **(A)** FL HMGB1 bending DNA4 (green), FL HMGB1 bending DNA5 (orange), and AB bending DNA4 (black). The dissociation constants determined by these curves are shown in Table 4.1. **(B)** A box bending DNA4. **(C)** B box bending DNA4. **(D)** BC bending DNA4. Each experiment had duplicate concentrations and more than two data points at a single concentration were from independent experiments.

	$_{app}K_d$ (nM)
FL (DNA4)	1.7 ± 0.4
FL (DNA5)	1.9 ± 0.7
AB (DNA4)	1.5 ± 1.0
A (DNA4)	24 ± 12
B (DNA4)	37 ± 25
BC (DNA4)	99 ± 30

Table 4.1. Apparent affinities of HMGB1 constructs interacting with and bending DNA determined using ensemble FRET experiments. The protein constructs are indicated, with the DNA construct in parentheses. The errors in each measurement are the 95% confidence interval of the fit.

changed as the competitor was titrated up, suggesting that excess competitor was actively causing dissociation. Although this could be interesting for future study, this observation did not impact the smFRET experiments since the DNA was immobilized on the surface and not free to facilitate dissociation of other HMGB1/DNA complexes.

Full length HMGB1 shows no apparent sequence dependence for bending DNA under smFRET conditions

It is widely known that HMGB1 shows very little sequence dependence for DNA binding (41), which was the case in our ensemble FRET studies. To further test this and assess the potential for using smFRET to study DNA bending we used two DNA constructs from the experiments in Chapter 3 to test bending by FL HMGB1. We separately attached these DNAs to coverslips through a biotin/streptavidin linkage. We excited the coverslip containing DNA with a 532 nm laser, and collected the emission on the donor emission channel (585/70nm) and the acceptor emission channel (725/120nm), as outlined in Chapter 3. We then determined the FRET efficiency of single DNA molecules at each time point and put the time traces through a state finding program that fits the FRET efficiency of each time trace. The output FRET efficiencies were binned and histogrammed to study populations of DNA molecules.

The two different DNA constructs used in these initial experiments, DNA1 and DNA2, are shown in Figure 4.3. The histogram of the binned FRET efficiencies obtained with DNA1 is shown in Figure 4.5A. The data show a population of 139 DNA molecules that can be fit with a single Gaussian with a mean FRET efficiency of 0.25. We then flowed FL HMGB1 into the chamber, imaged the DNA molecules, and binned the FRET efficiencies. In the presence of FL

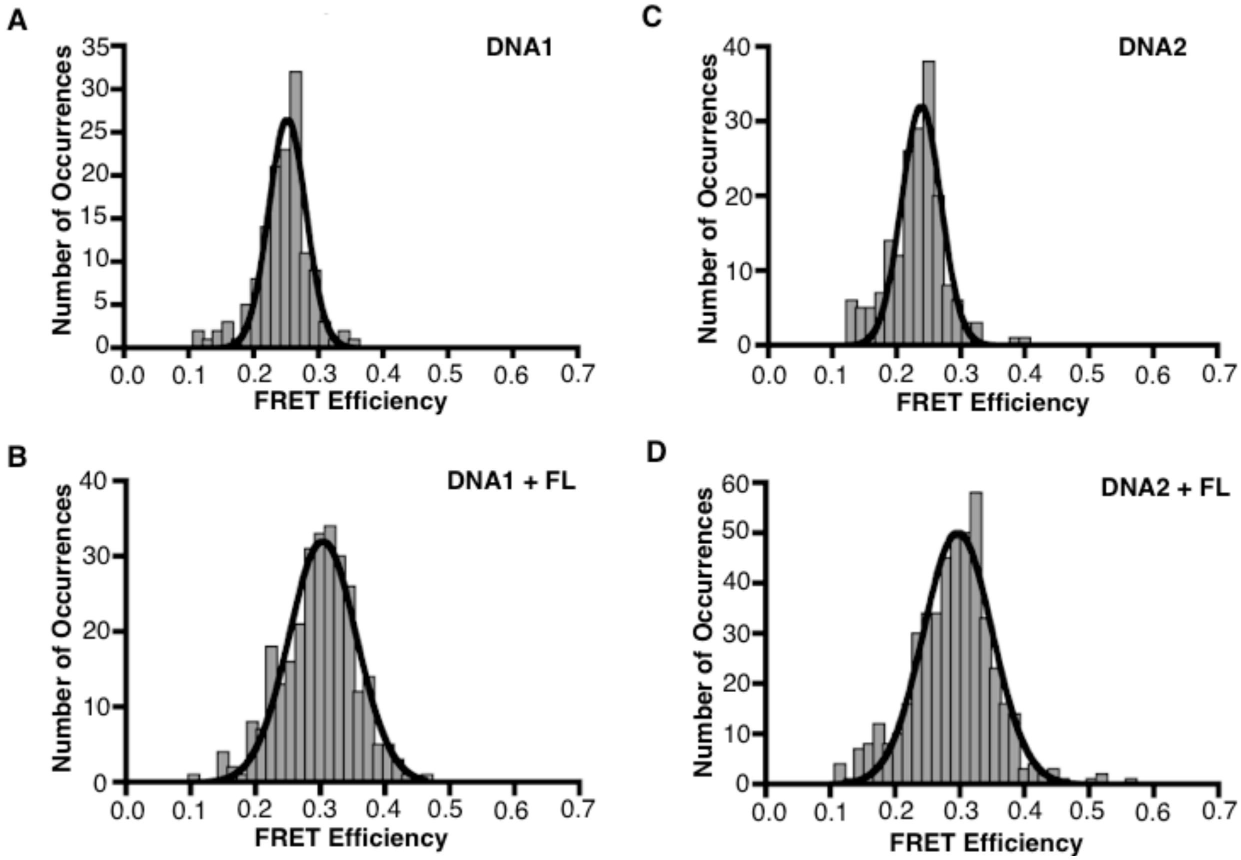


Figure 4.5. Sequence independent DNA bending by FL HMGB1 measured using smFRET. **(A)** DNA1 smFRET in the absence of HMGB1. 139 DNA molecules were imaged and 139 FRET occurrences were plotted. The average FRET is 0.25. **(B)** DNA1 smFRET in the presence of FL HMGB1. 286 FRET occurrences were plotted, giving a mean FRET efficiency of 0.30. The number of DNA molecules imaged was 252. **(C)** DNA2 smFRET in the absence of HMGB1. 184 molecules were imaged and 184 occurrences were plotted. The mean FRET efficiency of the Gaussian is 0.24. **(D)** DNA2 smFRET in the presence of FL HMGB1. 422 DNA molecules were imaged and 469 FRET occurrences were plotted. The mean FRET efficiency is 0.30.

HMGB1 the mean FRET efficiency from the Gaussian shifted to 0.30 (Figure 4.5B). The increase in FRET efficiency due to adding FL HMGB1 indicates that the FL protein bends the DNA. To determine whether FL HMGB1 bends two different DNA sequences similarly, we imaged a coverslip containing the DNA2 sequence. The DNA-only mean FRET efficiency was 0.24, which is similar to that of DNA1 (Figure 4.5C), as expected since the donor and acceptor dyes are spaced identically in both constructs. We next added FL HMGB1. Figure 4.5D shows that the population of DNA2 molecules shifted to a FRET state of 0.30 when FL was added to the chamber. Together these results indicate that HMGB1 bends both DNA sequences similarly, and that the smFRET could effectively monitor DNA bending by HMGB1.

Full length HMGB1 stably bends DNA

We designed a different DNA construct to use in subsequent single molecule studies (Figure 4.3, DNA3). Similar to the previous DNA constructs, the new construct had 18 bp of dsDNA between the donor and acceptor dyes; however, the new DNA construct had a 19 nt ssDNA linker that attached the construct to the surface through a biotin-streptavidin linkage. A schematic of the surface used for the remainder of the studies in this chapter is shown in Figure 4.6. The glass surface was PEGylated, with some PEG molecules functionalized with biotin. Streptavidin was used to link the PEG to the biotinylated DNA construct. The DNA was imaged and the intensities of the donor and acceptor dyes were monitored over time using a 60 ms exposure time (Figure 4.7A). The intensities were converted into FRET efficiencies (Figure 4.7B, blue). The FRET efficiency time courses were analyzed using an in-house state finding

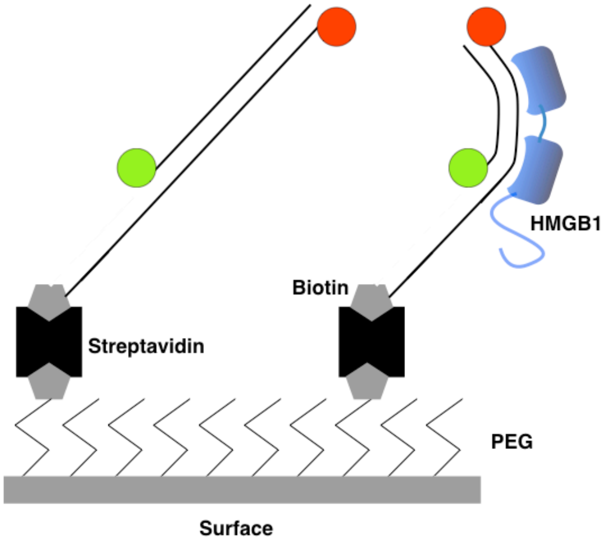


Figure 4.6. The surface used for the single molecule FRET studies. The glass coverslip is coated with polyethylene glycol (PEG) to prevent nonspecific interactions of the proteins and DNA with the surface. Some PEG molecules are functionalized with biotin, which allows immobilization of streptavidin. Biotinylated DNA binds to the streptavidin. The left side of the figure shows DNA attached to the surface. When HMGB1 is flowed in, the protein bends the DNA (right side), resulting in a higher FRET efficiency than the unbent DNA.

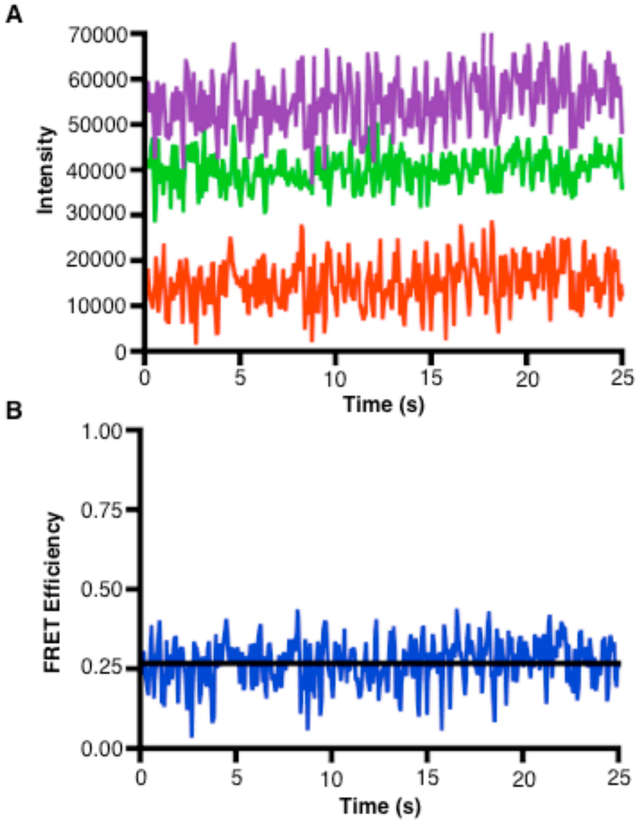


Figure 4.7. A sample timetrace of signals from a single molecule experiment. **(A)** The donor (green) and the acceptor (red) intensities of a single DNA molecule are shown over time. The sum of the donor and acceptor intensities (purple) is also shown. **(B)** The FRET efficiency shown over time (blue). The fit of the FRET efficiency is shown in black. The signals were smoothed by averaging 2 adjacent frames only for the purpose of this image.

program (Figure 4.6B, black line). In this example, the single observed state fit with a FRET efficiency of 0.26.

The FRET efficiencies of a population of single DNA molecules were binned and fit with a Gaussian, yielding a mean efficiency of 0.27 (Figure 4.8A). The mean and confidence interval of the Gaussian fit for this and all subsequent figures is provided in Table 4.2. We then flowed FL HMGB1 into the chamber, imaged the DNA molecules, determined the FRET occurrences, and histogrammed the data. In the presence of protein, the Gaussian distribution had a mean FRET efficiency of 0.33 (Figure 4.8B), indicating that FL HMGB1 bent the DNA. Out of 248 molecules examined in the presence of FL HMGB1, we observed only 3 changes in FRET states between bending and unbending. This lack of dynamic bending and unbending appears inconsistent with the suggestion that HMGB1 comes on and off DNA rapidly (106-108). Our experiments, however, would not allow us to observe binding and release independent of DNA bending.

Individual HMG boxes bend DNA to the same extent as full length HMGB1

We next wanted to determine how the individual A and B boxes bend DNA compared to the FL protein. We imaged DNA in the presence of purified A box. The resulting FRET data fit to a single Gaussian that showed a mean FRET efficiency of 0.32 (Figure 4.9A). Like full length protein, bending by the A box did not appear to be dynamic since no state changes were observed. The A box increased the noise in the FRET signal compared to all other truncations tested (data not shown), for reasons that we do not yet understand.

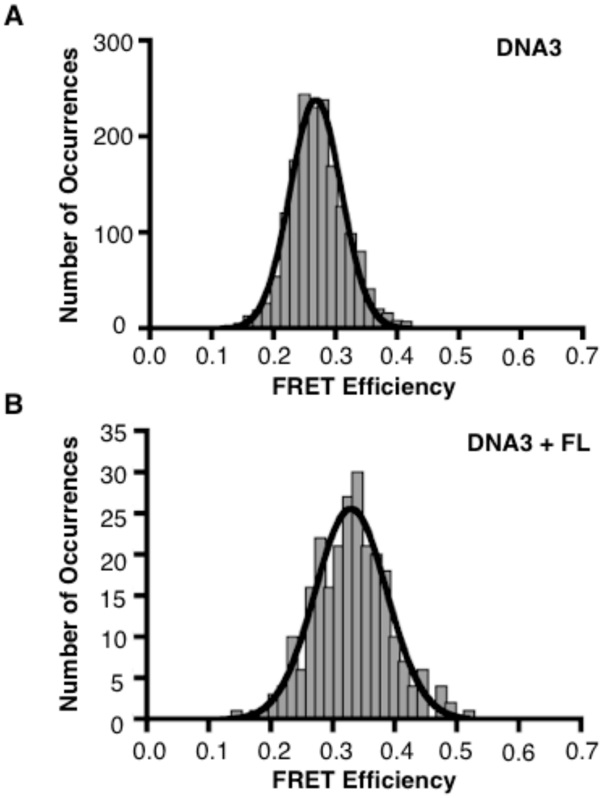


Figure 4.8. FL HMGB1 bends DNA. **(A)** The FRET efficiency of DNA3 in the absence of protein. 1690 DNA molecules were analyzed and 1701 FRET occurrences were plotted. The mean FRET state is 0.27. **(B)** FL HMGB1 bends DNA as observed by smFRET. 20 nM FL HMGB1 was flowed into the sample chamber. 248 DNA molecules analyzed with 251 occurrences plotted. The mean FRET efficiency is 0.33.

DNA	**Protein**	**mean FRET lower Gaussian**	**mean FRET upper Gaussian**	**95% CI**
DNA3	-	0.27	-	0.266-0.271
DNA3	FL	0.33	-	0.323-0.335
DNA3	A	0.32	-	0.309-0.325
DNA3	B	0.33	-	0.327-0.335
DNA3	AB	0.32	0.43	0.312-0.324, 0.418-0.433
DNA3	A + B	0.34	-	0.334-0.344
DNA3	BC (50 nM)	0.27	-	0.272-0.278
DNA3	BC (200 nM)	0.28	0.34	0.277-0.282, 0.332-0.348

Table 4.2. Summary of smFRET data.

lower Gaussian - lower FRET population or only FRET population in histogram
upper Gaussian - higher FRET population in histogram (if there is more than one)
95% CI – 95% confidence interval for the mean FRET

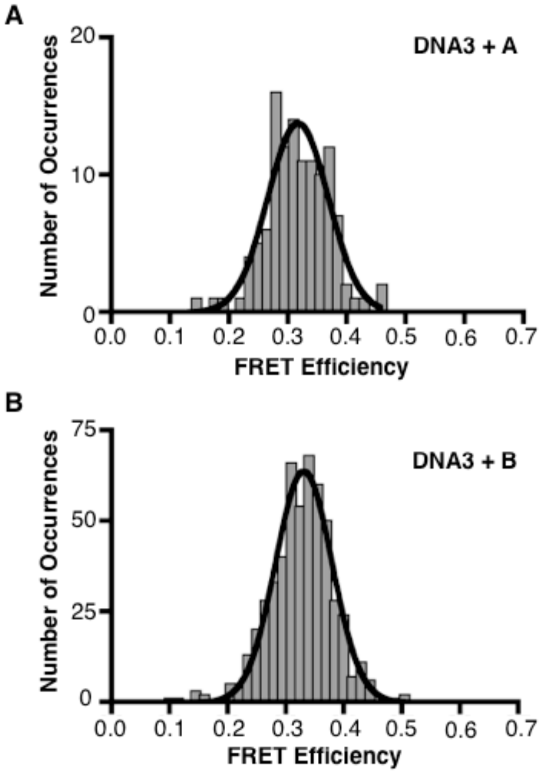


Figure 4.9. The HMGB1 A box and B box individually bend DNA. **(A)** 50 nM A box was flowed into the sample chamber with DNA3 immobilized on the surface. 118 DNA molecules were analyzed and 118 occurrences were plotted. The mean of the Gaussian fit is 0.32. **(B)** 50 nM B box was flowed into the sample chamber containing immobilized DNA3. 522 DNA molecules were analyzed and 529 occurrences were plotted. The mean FRET efficiency is 0.33.

We monitored DNA bending by the B box on the same DNA construct. We observed a mean FRET efficiency of 0.33 (Figure 4.9B). Again, we did not observe many state changes during imaging. Together our data show that the individual boxes bend the 18 bp dsDNA to the same extent as one another and FL HMGB1. This observation differs from some previous experiments, where the B box bent DNA to a greater extent than the A box (114, 123, 125), however, indirect assessments of DNA bending were used.

The AB construct lacking the C-terminal tail bends DNA differently than FL HMGB1

We next tested the contribution of the C-terminal tail to DNA bending by removing it and asking how the AB protein bends DNA compared to FL HMGB1. The smFRET results with the AB construct showed a very broad FRET distribution that appeared to be composed of two different bent populations. Accordingly, we fit the data with two Gaussians. The mean FRET states of the two peaks are 0.32 and 0.43 (Figure 4.10). The population with the lower FRET state of 0.32 is similar to the mean FRET efficiency observed in the presence of the FL, A, and B constructs. These data suggest that the AB protein bends DNA heterogeneously, with a portion of the AB protein bending DNA using only one HMG box, and second population simultaneously used both HMG boxes to bend DNA, resulting in a higher mean FRET of 0.43. Interestingly, a higher number of state changes were observed with the AB construct than any of the other protein truncations. The number of state changes, however, was not significant enough to allow us to analyze the kinetics of the state change, but qualitatively some molecules appeared to go between unbent, partially bent, and fully bent states. This could be interesting to study in the future.

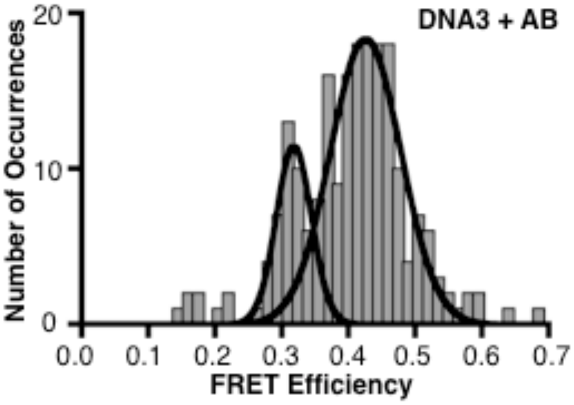


Figure 4.10. HMGB1 with the C-terminal tail truncated (AB) bends DNA heterogeneously. 7nM AB was flowed into the sample chamber with DNA3 attached to the surface. 191 DNA molecules were analyzed and 209 FRET occurrences were plotted. The data were fit with 2 Gaussians and the mean FRET efficiencies are 0.32 and 0.43.

We next asked whether the unique DNA bending by the AB protein could be mimicked by flowing the individual A and B boxes into the chamber together. When we flowed the A box and the B box into the chamber at the same time, we determined a mean FRET state of 0.34 (Figure 4.11), which is similar to that of the the A box and the B box alone. Therefore, we conclude the A box and the B box need to be covalently connected to exhibit bending resulting in the high FRET state of 0.43. This result is consistent with previous data showing that the individual boxes do not interact in solution (115, 116).

Attaching the C-terminal tail to the B box does not change the extent of DNA bending, but lowers the apparent binding affinity

Our data are consistent with a model in which only one of the two HMG boxes in the full length protein bends the DNA at any given time. It is known that the C-terminal tail can interact with both HMG boxes (62, 116-118). To determine whether the C-terminal tail affects DNA bending of single DNA molecules, we expressed and purified a construct that contains the B box connected to the C-terminal tail (Figure 4.1, BC). We flowed the BC construct into a chamber at a concentration equal to that we previously used to study DNA bending by the B box (i.e. 50 nM). Figure 4.12A shows that under these conditions the mean FRET efficiency was 0.27, which is indistinguishable from free, unbent DNA. This is consistent with the C-terminal tail limiting the amount of DNA bending (and likely DNA binding) by the B box (117). We next tested the BC construct at a concentration of 200 nM and observed two FRET populations with mean FRET efficiencies of 0.28, and 0.34 (Figure 4.12B). The lower FRET state is consistent with

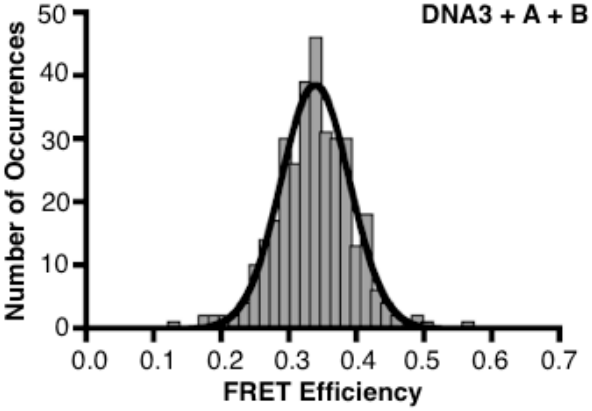


Figure 4.11. When combined, the A box and B box bend DNA to the same extent as either alone. 75 nM A box and 50 nM B box were flowed into the chamber together. 327 DNA molecules were analyzed and 333 occurrences were plotted. The mean FRET efficiency is 0.34.

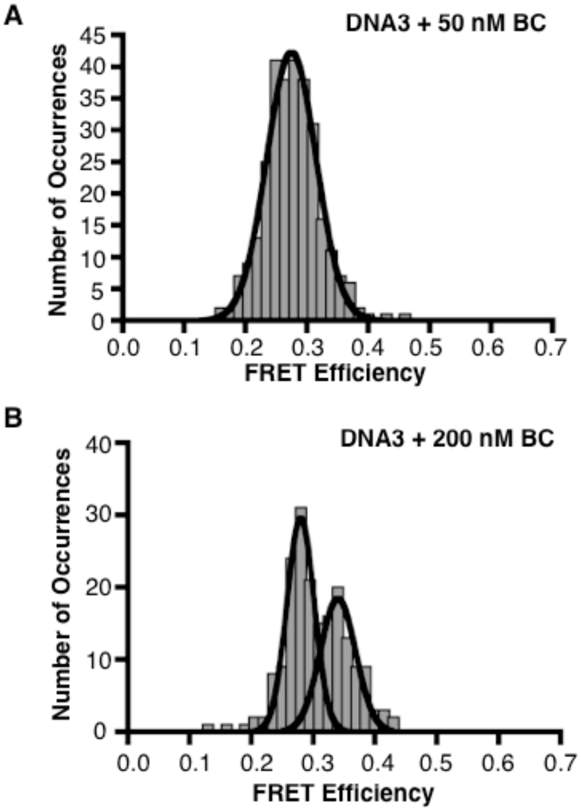


Figure 4.12. An HMGB1 construct lacking the A box (BC) does not bend DNA as efficiently as the B box alone. **(A)** The BC truncation was flowed into the sample chamber at a concentration of 50 nM. 292 DNA molecules were analyzed and 292 occurrences were plotted. The mean FRET efficiency of the Gaussian is 0.27. **(B)** 200 nM BC was flowed into the chamber. 189 DNA molecules were analyzed and 190 FRET occurrences were plotted. The data were fit with 2 Gaussians yielding mean FRET efficiencies of 0.28 and 0.34.

unbent DNA, whereas the higher FRET state is similar to that observed for DNA bent by the B box.

Discussion

We studied how HMGB1 bends DNA using smFRET. This technique allows us to observe individual bent DNA-protein complexes and characterize different populations of bent DNA in a single reaction. We found that the individual A and B boxes bend DNA to the same extent as full length HMGB1. We observed heterogeneity in bending with an AB construct that lacked the C-terminal region. In this case, two states were observed, one that was bent to the same extent as the FL HMGB1 and a second that was bent to a greater extent, which suggests that both HMG boxes can bind and bend DNA simultaneously in the absence of the C-terminal region. We also studied a BC construct, which lacks the A box, and found that a significantly higher protein concentration was needed to observe DNA bending, but the DNA, when bent, was bent to the same extent as observed with the B box alone. We propose a model that in the context of the full length protein the A box is free to bind and bend DNA, whereas the B box is not due to its interaction with the acidic C-terminal tail. In the tailless construct, AB, both boxes are available to bend DNA, resulting in DNA being bent to a greater extent.

Two HMG boxes can simultaneously bend DNA in the absence of the C-terminal tail

We found that the FL HMGB1 protein bends DNA to a single state that is similar to that observed with the individual A and B box domains. By contrast, the AB construct bends DNA to two different states, with one population of DNA bent similarly to FL and a second that appears to bend DNA to a greater extent. We propose that in the context of the AB construct, the A and the B boxes can simultaneously bend a single molecule of DNA, resulting in the higher FRET state. Furthermore, we propose that in the context of FL HMGB1, the A box bends DNA, and the

B box interacts with the C-terminal tail, limiting its ability to bend DNA. This is supported by our experiments showing that the BC construct did not efficiently bend DNA.

Previously published studies support the model that full length HMGB1 interacts with DNA differently than the tailless AB truncation. For example, the AB tailless protein was found to have a higher affinity for DNA-cellulose compared to full length HMGB1 (116), although we did not see this in the ensemble FRET experiments shown here. Also, it was found that when the A box was acetylated, the FL protein could not bind DNA, whereas the AB protein could bind DNA (57), suggesting that the B box was free to interact with DNA in the AB protein but not the FL protein. A structural study supports our model that the AB construct bends DNA using the two HMG boxes simultaneously. A crystal structure of a hybrid AB box composed of the Sry HMG box attached to the HMGB1 B box showed bending of DNA by both boxes using a similar sized piece of dsDNA (126). Future studies need to be performed to more directly test our hypothesis that the A and the B box can both bend DNA in the context of the AB construct. For example, we could mutate the A box in the AB and FL proteins. The point mutation F38A in the A box removes its ability to bend DNA and reduces its affinity for binding DNA (44). If our overall model is correct, then the A box mutant FL protein will not bend DNA, whereas the A box mutant AB truncation will act like a single box (the B box) in bending the DNA.

The A box and B box bend DNA to the same extent

We found that the A box and the B box bend DNA to the same extent. These results are different from those of previous studies in which the B box was found to bend DNA to a greater extent than the A box. The amount of bending by individual boxes has only been studied by a few groups and the ensemble assays used provided indirect assessments of bending. One group performed recombination experiments, in which a protein needs to bend DNA for recombination to occur. They found that FL HMGB1 and the B box bent DNA enough for recombination to occur, but the A box and the BC construct did not (114). Other groups assessed bending through ligation-mediated ring closure, where short (less than 150 bp) linear DNA cannot be circularized by T4 ligase unless a protein bends the DNA. The AB protein and the FL protein were able to mediate ring closure (124, 127). The A box (1-83) and the B box (85-164) and various truncations of these boxes could not mediate ring closure. However, when the basic linker was added to the B box (84-180), this longer peptide could bend DNA enough to allow for circularization of DNA (123, 125). One interpretation of these results is that there is more extensive bending of DNA by the B box containing the basic linker extension (aa 167-184), than the B box truncation (83-166) used in our studies. Another group showed that an acetylated FL HMGB1 lost the ability for ligase mediated-circularization, but did not lose the ability to bend DNA based on AFM studies (58). This indicates that circularization may not directly correlate with the amount of DNA bending induced by a protein.

The role of the C-terminal tail in controlling HMGB1 DNA binding and bending

Our data are consistent with the model that the C-terminal acidic tail affects how HMGB1 interacts with DNA. The smFRET studies presented here show that the tail affects the ability of the B box to bend DNA, but does not affect the extent to which the B box bends DNA. It took a much greater concentration of the BC construct to observe bending, however, the mean FRET efficiency of the bent state was 0.34, which is similar to the B box alone. Our observations are consistent with previous studies. Thermal denaturation studies have shown that the AB truncation and FL HMGB1 denature differently, indicating that the C-terminal tail interacts with at least one domain (115). Further, the tail has been shown to interact with both the A and B boxes (51, 62, 116-118); although some studies suggest that it interacts primarily with the B box (62, 114), while others suggest that it interacts with the A box more (51, 118). In the context of the full length protein, the C-terminal tail interacts with the B box with a higher affinity than the A box (117). We found that the tail limits DNA bending when fused to the B box, which is consistent with observations from other assays that used recombination as a readout for DNA bending. Recombination could be facilitated by B box, but not by BC (114).

HMGB1 is known to interact with many other factors within the nucleus, including nucleosomes. It can interact with nucleosomes through its C-terminal acidic tail (50-52). Our data suggest that this interaction with nucleosomes could change how HMGB1 bends DNA; if the tail is bound to a nucleosome, the A and B boxes would be free to bend DNA simultaneously.

Experimental Procedures

Oligonucleotides and other reagents

The oligonucleotide sequences used in these experiments are shown in Figure 4.3. The sequences were annealed as described in Chapter 3. Catalase, glucose oxidase, and glucose were prepared as described in Chapter 3. Trolox was prepared by dissolving 6 mM Trolox into 30% (v/v) glycerol, 150 mM Tris, pH 7.9, 150 mM KCl, and 15 mM $MgCl_2$ nutating overnight in the dark, then 3 mM DTT, 0.15 mg/mL BSA, and 0.3% NP-40 were added. Single molecule surfaces were prepared as described in Chapter 3.

Protein preparation and purification

The cDNA encoding human FL HMGB1 (1-215) was cloned into a pET19b vector after mutating out the native NdeI site in the nucleotide sequence of HMGB1. The following HMGB1 truncations were made: AB(1-166), A(1-85), B(83-166), and BC(83-215). In each construct, we mutated all the natural cysteine codons to serine codons and then mutated in a single cysteine codon so future modifications could be made to the purified proteins, such as labeling with a fluorophore. The proteins contained the following mutations: FL HMGB1 (C23S, C45S, C106S, E204C), AB (C23S, C45S, C106S, S35C), A box (C23S, C45S, S35C), B box (C106S, A137C), and BC (C106S).

The HMGB1 expression plasmids were transformed into *E. coli* BL-21 DE3 pLysS cells and grown overnight at 37°C on LB plates containing 100 µg/mL ampicillin (AMP) and 50 µg/mL chloramphenicol (CAM). For each construct, one colony was picked and grown in a 5 mL culture containing LB, 2 g/L D-glucose, AMP, and CAM at concentrations described above with overnight shaking at 200 rpm at 37°C. 2 mL of the 5 mL starter culture was used to inoculate 400-500 mL of LB containing glucose, AMP, and CAM and cells were grown to OD 0.4-0.5. The cultures were then induced with 0.5 mM IPTG for 2 hours, and cells were pelleted by centrifugation.

The pellets were resuspended in 10 mL lysis buffer (20 mM Tris (pH7.9), 5 mM 2-mercaptoethanol, 500 mM NaCl, 10% (w/v) glycerol, 1X EDTA-free protease inhibitors (Roche), and 10 mM imidazole) and sonicated. The sonicated mix was centrifuged at 15K rpm for 15 min and the supernatant was added in batch to a HisPur Ni-NTA column (Thermofisher) and nutated at 4°C for 1 h. The resin was placed in a column, and protein was eluted from the column with lysis buffer containing 250 mM imidazole. The eluate was dialyzed in buffer I (20 mM HEPES (pH7.0), 50 mM KCl, 10% glycerol, 5 mM $MgCl_2$) at 4°C. The full length construct was then placed over a Q column and the flow through was collected and put over the dsDNA cellulose column described below. For the AB, A, and B truncations, the eluates from the Ni-NTA column were dialyzed and incubated with dsDNA cellulose in batch. The dsDNA resin then placed into a column and protein was eluted at 500 mM KCl. Eluates were dialyzed in buffer II (20 mM HEPES (pH7.9), 50 mM KCl, 10% glycerol, 5 mM $MgCl_2$, 1 mM DTT) at 4°C. We were not sure whether the BC truncation would bind to the dsDNA cellulose, so that truncation was dialyzed after the nickel column and not placed over the dsDNA column. All proteins were flash frozen in liquid nitrogen and stored at -80°C.

SmFRET reaction conditions

0.2 mg/mL streptavidin was flowed into the sample chamber in Buffer A (10% (v/v) glycerol, 25 mM Tris (pH 7.9), 50 mM KCl, 1 mM DTT, 0.05 mg/mL BSA, 5 mM $MgCl_2$, 0.1% NP-40) and incubated for 5 min. Unbound streptavidin was flushed from the chamber with excess buffer A. 200-400 pM DNA was flowed into the chamber in Buffer A and incubated for 10 min. Unbound DNA was flushed from the chamber with excess Buffer A. For the sequence-dependence experiments using DNA1 and DNA2, the concentration of KCl in solution was 150 mM in Buffer A and the concentration of full length HMGB1 flowed into the chamber was 50 nM. These experiments were done on the Cristol microscope and data were analyzed as described in Chapter 3. All experiments done with DNA3 were performed on the reassembled microscope in JSCBB. These data were collected using 50 mM KCl and the concentrations of HMGB1 proteins indicated in the figure legends.

smFRET data collection and analysis

The TIRF microscope used to collect most of the data is described in Chapter 2. In short, each sample was excited with a 532-nm laser using an objective-based TIRF system and data were collected on EM-CCD cameras. Data were analyzed using in-house software as described in Chapter 3. Data in Figure 4.5 were collected on the microscope in Cristol Chemistry, so Chapter 3 describes any correction factors used for that data. For all other single molecule data in this chapter, we used an alpha of 0.1 and a gamma of 0.68, for correction factors on the new microscope. After analyzing the data, we histogrammed the data using a bin of 0.015 and fit the data with Gaussians to extract mean FRET efficiencies.

Ensemble FRET reactions

Reactions were incubated for 20 min at room temperature in Reaction Buffer (10% (v/v) glycerol, 10 mM Tris (pH 7.9), 50 mM KCl, 1 mM DTT, 0.05 mg/mL BSA, 0.1% NP-40, 5 mM $MgCl_2$, and 12.5 mM HEPES (pH 7.9)) with 1 nM dsDNA. Varying amounts of protein were added to each reaction. To determine the affinity for FL HMGB1 bending DNA4 and DNA5, 0-150 nM protein was titrated into solution. For determining the affinity for truncations on DNA4, 0-70 nM AB, 0-200 nM A, 0-250 nM B, and 0-500 nM BC were titrated into binding reactions. 18 μL of each reaction were placed in a well of a 384-well SensoPlate plate (Greiner). Plates were scanned using a Typhoon Scanner (GE Healthcare) at the following wavelengths: 532 nm excitation with 580/30 nm emission, 633 nm excitation with 670/30 nm emission, and 532 nm excitation with 670/30 nm emission. Corrections were made for bleedthrough of the donor into the acceptor channel and direct excitation of the acceptor by the donor laser, as described in (128). In each case, the FRET efficiency of a sample without any protein were subtracted from each sample with protein, giving a change in FRET. The change in FRET was plotted versus protein concentration and fit with a binding curve.

$$\begin{matrix}\text{Bound}\\\text{Complex}\end{matrix} = (\text{FRET}_{\text{max}} - \text{FRET}_{\text{min}})\left(\frac{[P]}{[P] + K_d}\right) + \text{FRET}_{\text{min}}$$

where P is the concentration of free protein in solution. Since the DNA concentration was 1 nM and some affinities are close to this, we had to estimate the free protein concentration. The free protein concentration at each point was calculated by subtracting the concentration of protein bound from the total concentration assuming 1 protein molecule bound per DNA molecule in the bent conformation. The change in FRET versus free protein concentration was plotted and fit with the above equation, giving the binding/bending affinity. The errors in each measurement are the 95% confidence interval of the fit.

CHAPTER 5

Summary and Future Directions

Summary

In this thesis I investigated how proteins involved in transcription bend DNA using single molecule FRET. In Chapter 2, I discussed the assembly and alignment of a TIRF microscope used to study these DNA-protein interactions. I also introduced the physical requirements for building a TIRF microscope, including why we chose specific lenses and filters for the microscope, and the requirements of collimation and focal point location for alignment of the beam through the objective to achieve total internal reflection.

In Chapter 3, we studied the bending of consensus and nonconsensus TATA DNA by TBP. Previous publications had shown that TBP bends different TATA sequences to varying extents, dependent on the affinity of TBP for that particular sequence. Using single molecule FRET, we found that TBP bends two different TATA sequences for which it has varying affinities to the same extent. Further, analysis of the kinetics of TBP bending consensus DNA showed that the TBP/bent DNA complex exits in two kinetically distinct populations that contain DNA bent to the same extent.

In Chapter 4, I studied HMGB1 bending DNA. HMGB1 binds and bends without sequence specificity, which was verified using smFRET with two different DNA sequences. We then elucidated how HMGB1 bends DNA. We found that full length HMGB1 bends DNA to the same extent as the A box alone and the B box alone. The AB bi-domain, lacking the C terminal tail, bent DNA to two distinct populations, with one population having the same extent of bending as a single domain and a second population with DNA bent to a greater extent. We hypothesized that the AB construct can either bend DNA using one box or using two boxes simultaneously. Further, we hypothesized that the C-terminal tail would prevent the full length protein from bending DNA with two boxes simultaneously. Accordingly, we made a construct that contained the B box with the C-terminal acidic tail and studied how it bent DNA. We found that the BC construct bent DNA to the same extent as the B box alone, but the fraction of the DNA that was bent was much lower than with the B box alone, indicating that the C-terminal tail hinders the B box from interacting with DNA.

Future Directions

In the studies described here we directly examined DNA bending by TBP and HMGB1 at the single molecule level. This allowed us to determine the population distributions, which cannot be done using ensemble studies. Although we have a clearer picture of how these proteins bend DNA, many new questions arise from this research.

Chapter 2

The TIRF microscope in the basement of the JSCBB building was reassembled so that additional modifications can be made to the table in the future. First, we have the hardware to perform alternating laser excitation (ALEX) experiments; we would need to place the modulators into the laser lines and control them with software. Currently, we do not have the capability for autofocusing during imaging. We have the piezo stage and the rest of the hardware connections to do so, we just need to incorporate software capable of autofocusing. We also have room to add a 3rd (and 4th) laser. All of these modifications are rather straightforward to make on the optics table, the more difficult part would be to interface the hardware with the required software.

Chapter 3

In Chapter 3, we studied TBP bending DNA. These studies were not able to differentiate between the two kinetic models of binding/bending in which two bent populations with different kinetic stabilities existed in solution. In the branched model, it is possible that TBP bound in two different orientations with respect to the TATA box, which is not palindromic. This could give rise to complexes with different kinetic stabilities. The second model we proposed is a linear pathway where TBP/DNA exists an intermediate, unstable complex that is bent to the same extent as the stable, final complex. To test the branched model, we can label one side of TBP and then label one end of the TATA DNA. If TBP binds and bends DNA in two different orientations with different kinetic stabilities, we would see two different FRET states between labeled TBP and DNA and the two FRET populations would have different kinetic stabilities. Moreover, we could study this with the consensus and nonconsensus sequences to elucidate whether TBP binds nonconsensus sequences in both directions as well.

This minimalist TBP-DNA system lays the groundwork for studying basic factors involved in transcription. We could add other general transcription factors, one by one, into the system to study the assembly of preinitiation complexes and the initiation of transcription. We could also study other TBP-containing complexes, such as complexes involved in Pol I and Pol III transcription (129).

Chapter 4

To further test our model that only the A box bends DNA in the context of the full length protein, we intend to use a point mutation in the A box that eliminates bending and reduces binding. We can also engineer a B box point mutation. The effect of the A box and B box mutations on DNA bending can be studied in the context of multiple HMGB1 constructs, including FL, AB, A, and B. If our model is correct, we would expect the point mutation in the A box in the full length construct to greatly reduce the population of bent DNA. By contrast, mutation of the A box in the AB construct would result in a single population of bent DNA. The A box construct containing the A box mutation would not be able to bend DNA. The mutation in the B box would not affect bending by the full length protein, would lead to a single bent population with the AB construct, and would eliminate bending by the B box construct.

Our smFRET results with HMGB1 would be further strengthened by performing additional controls of the type that were performed with TBP. For example, we will determine whether the dyes freely rotate on the DNA in the absence and presence of the different HMGB1 protein constructs using anisotropy experiments. Further, we need to ensure that the proteins do not affect the fluorescence output of the dyes. To do this, we need to determine whether the intensity of the emission signal from each fluorophore on DNA changes when protein is added. There is higher noise in the signal after adding certain HMGB1 constructs, potentially indicating that there are protein-dye interactions.

Future experiments could address whether HMGB1 binds and bends DNA dynamically. In anticipation of testing this, I engineered each of the HMGB1 constructs to contain a single, surface-exposed cysteine. Each of the proteins can be labeled with a fluorophore, which will allow the dynamics of DNA binding to be measured directly using fluorescence co-localization. There are limitations on co-localization studies, for example, the labeled protein has to be at a low enough concentration so not to saturate the detector, which may be a problem with constructs that bind DNA with lower affinity. To facilitate the measurement of DNA bending dynamics, we could change the dye pair on the DNA to get a larger FRET response to bending.

Binding site size is an important factor for HMGB1 binding and bending DNA because of its nonsequence specific interaction with DNA. The full length protein binds 16-19 bp, AB binds 12 bp, and individual boxes bind 5-6 bp (113). This begs the question of how the A box and B box constructs show relatively uniform bending in the smFRET experiments when they could theoretically bind in a total of 13 different registers in the forward and backward orientations. To test this we could shorten the DNA length and to determine if the extent of bending remains consistent as the length of DNA, and therefore the number of binding registers, changes.

Finally, within cells there are many proteins known to interact with HMGB1. We could add other proteins to the single molecule system and examine how these proteins affect DNA binding and bending by full length HMGB1. An example of this would be adding a histone tail, which is known to interact with the C-terminal tail of HMGB1. Potentially, this interaction could remove the C-terminal tail from the B box DNA binding domain, allowing both the A and B boxes to bend DNA simultaneously.

BIBLIOGRAPHY

1. Schneider, D. A. (2012) RNA polymerase I activity is regulated at multiple steps in the transcription cycle: Recent insights into factors that influence transcription elongation, *Gene 493*, 176-184.

2. White, R. J. (2008) RNA polymerases I and III, non-coding RNAs and cancer, *Trends Genet 24*, 622-629.

3. Brookes, E., and Pombo, A. (2009) Modifications of RNA polymerase II are pivotal in regulating gene expression states, *Embo Rep 10*, 1213-1219.

4. Thomas, M. C., and Chiang, C. M. (2006) The general transcription machinery and general cofactors, *Crit Rev Biochem Mol Biol 41*, 105-178.

5. Weake, V. M., and Workman, J. L. (2010) Inducible gene expression: diverse regulatory mechanisms, *Nat Rev Genet 11*, 426-437.

6. Blackwood, E. M., and Kadonaga, J. T. (1998) Going the distance: A current view of enhancer action, *Science 281*, 60-63.

7. West, A. G., Gaszner, M., and Felsenfeld, G. (2002) Insulators: many functions, many mechanisms, *Genes Dev 16*, 271-288.

8. Jaenisch, R., and Bird, A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals, *Nat Genet 33*, 245-254.

9. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution, *Nature 389*, 251-260.
10. Armstrong, J. A. (2007) Negotiating the nucleosome: factors that allow RNA polymerase II to elongate through chromatin, *Biochem Cell Biol 85*, 426-434.

11. Mellor, J. (2005) The dynamics of chromatin remodeling at promoters, *Mol Cell 19*, 147-157.

12. Li, B., Carey, M., and Workman, J. L. (2007) The role of chromatin during transcription, *Cell 128*, 707-719.

13. Zanton, S. J., and Pugh, B. F. (2006) Full and partial genome-wide assembly and disassembly of the yeast transcription machinery in response to heat shock, *Genes Dev 20*, 2250-2265.

14. Pahl, H. L. (1999) Activators and target genes of Rel/NF-kappaB transcription factors, *Oncogene 18*, 6853-6866.

15. Liu, W. L., Coleman, R. A., Ma, E., Grob, P., Yang, J. L., Zhang, Y., Dailey, G., Nogales, E., and Tjian, R. (2009) Structures of three distinct activator-TFIID complexes, *Genes Dev 23*, 1510-1521.

16. Wright, E., Bain, M., Teague, L., Murphy, J., and Sinclair, J. (2005) Ets-2 repressor factor recruits histone deacetylase to silence human cytomegalovirus immediate-early gene expression in non-permissive cells, *J Gen Virol 86*, 535-544.

17. Shapiro, M. J., and Shapiro, V. S. (2011) Transcriptional repressors, corepressors and chromatin modifying enzymes in T cell development, *Cytokine 53*, 271-281.

18. Auble, D. T., Hansen, K. E., Mueller, C. G., Lane, W. S., Thorner, J., and Hahn, S. (1994) Mot1, a global repressor of RNA polymerase II transcription, inhibits TBP binding to DNA by an ATP-dependent mechanism, *Genes Dev 8*, 1920-1934.

19. Casamassimi, A., and Napoli, C. (2007) Mediator complexes and eukaryotic transcription regulation: an overview, *Biochimie 89*, 1439-1446.

20. Agalioti, T., Lomvardas, S., Parekh, B., Yie, J., Maniatis, T., and Thanos, D. (2000) Ordered recruitment of chromatin modifying and general transcription factors to the IFN-beta promoter, *Cell 103*, 667-678.

21. Baumann, M., Pontiller, J., and Ernst, W. (2010) Structure and basal transcription complex of RNA Polymerase II core promoters in the mammalian genome: An Overview, *Mol Biotechnol 45*, 241-247.

22. Imbalzano, A. N., Zaret, K. S., and Kingston, R. E. (1994) Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA, *J Biol Chem 269*, 8280-8286.

23. Parvin, J. D., and Young, R. A. (1998) Regulatory targets in the RNA polymerase II holoenzyme, *Curr Opin Genet Dev 8*, 565-570.

24. Lemon, B., and Tjian, R. (2000) Orchestrated response: a symphony of transcription factors for gene control, *Genes Dev 14*, 2551-2569.

25. Levine, M., and Tjian, R. (2003) Transcription regulation and animal diversity, *Nature 424*, 147-151.

26. Wu, J., Parkhurst, K. M., Powell, R. M., Brenowitz, M., and Parkhurst, L. J. (2001) DNA bends in TATA-binding protein-TATA complexes in solution are DNA sequence-dependent, *J Biol Chem 276*, 14614-14622.

27. Wu, J., Parkhurst, K. M., Powell, R. M., and Parkhurst, L. J. (2001) DNA sequence-dependent differences in TATA-binding protein-induced DNA bending in solution are highly sensitive to osmolytes, *J Biol Chem 276*, 14623-14627.

28. Starr, D. B., Hoopes, B. C., and Hawley, D. K. (1995) DNA bending is an important component of site-specific recognition by the TATA binding protein, *J Mol Biol 250*, 434-446.

29. Hieb, A. R., Halsey, W. A., Betterton, M. D., Perkins, T. T., Kugel, J. F., and Goodrich, J. A. (2007) TFIIA changes the conformation of the DNA in TBP/TATA complexes and increases their kinetic stability, *J Mol Biol 372*, 619-632.

30. Petri, V., Hsieh, M., Jamison, E., and Brenowitz, M. (1998) DNA sequence-specific recognition by the Saccharomyces cerevisiae "TATA" binding protein: promoter-dependent differences in the thermodynamics and kinetics of binding, *Biochemistry 37*, 15842-15849.

31. Whittington, J. E., Delgadillo, R. F., Attebury, T. J., Parkhurst, L. K., Daugherty, M. A., and Parkhurst, L. J. (2008) TATA-binding protein recognition and bending of a consensus promoter are protein species dependent, *Biochemistry 47*, 7264-7273.

32. Bareket-Samish, A., Cohen, I., and Haran, T. E. (2000) Signals for TBP/TATA box recognition, *J Mol Biol 299*, 965-977.

33. Wobbe, C. R., and Struhl, K. (1990) Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription in vitro, *Mol Cell Biol 10*, 3859-3867.

34. Bonham, A. J., Neumann, T., Tirrell, M., and Reich, N. O. (2009) Tracking transcription factor complexes on DNA using total internal reflectance fluorescence protein binding microarrays, *Nucleic Acids Res 37*, e94.

35. Weideman, C. A., Netter, R. C., Benjamin, L. R., McAllister, J. J., Schmiedekamp, L. A., Coleman, R. A., and Pugh, B. F. (1997) Dynamic interplay of TFIIA, TBP and TATA DNA, *J Mol Biol 271*, 61-75.

36. Pugh, B. F. (2000) Control of gene expression through regulation of the TATA-binding protein, *Gene 255*, 1-14.

37. Nikolov, D. B., Chen, H., Halay, E. D., Usheva, A. A., Hisatake, K., Lee, D. K., Roeder, R. G., and Burley, S. K. (1995) Crystal structure of a TFIIB-TBP-TATA-element ternary complex, *Nature 377*, 119-128.

38. Juo, Z. S., Chiu, T. K., Leiberman, P. M., Baikalov, I., Berk, A. J., and Dickerson, R. E. (1996) How proteins recognize the TATA box, *J Mol Biol 261*, 239-254.

39. Bleichenbacher, M., Tan, S., and Richmond, T. J. (2003) Novel interactions between the components of human and yeast TFIIA/TBP/DNA complexes, *J Mol Biol 332*, 783-793.

40. Bianchi, M. E., and Agresti, A. (2005) HMG proteins: dynamic players in gene regulation and differentiation, *Curr Opin Genet Dev 15*, 496-506.

41. Agresti, A., and Bianchi, M. E. (2003) HMGB proteins and gene expression, *Curr Opin Genet Dev 13*, 170-178.

42. Kugler, J. E., Deng, T., and Bustin, M. (2012) The HMGN family of chromatin-binding proteins: dynamic modulators of epigenetic processes, *Biochim Biophys Acta 1819*, 652-656.

43. Cleynen, I., and Van de Ven, W. J. (2008) The HMGA proteins: a myriad of functions (Review), *Int J Oncol 32*, 289-305.

44. Ohndorf, U. M., Rould, M. A., He, Q., Pabo, C. O., and Lippard, S. J. (1999) Basis for recognition of cisplatin-modified DNA by high-mobility-group proteins, *Nature 399*, 708-712.

45. Stros, M. (2010) HMGB proteins: interactions with DNA and chromatin, *Biochim Biophys Acta 1799*, 101-113.

46. Calogero, S., Grassi, F., Aguzzi, A., Voigtlander, T., Ferrier, P., Ferrari, S., and Bianchi, M. E. (1999) The lack of chromosomal protein HMG1 does not disrupt cell growth but causes lethal hypoglycaemia in newborn mice, *Nat Genet 22*, 276-280.

47. Lotze, M. T., and Tracey, K. J. (2005) High-mobility group box 1 protein (HMGB1): nuclear weapon in the immune arsenal, *Nat Rev Immunol 5*, 331-342.

48. Thomas, J. O. (2001) HMG1 and 2: architectural DNA-binding proteins, *Biochem Soc Trans 29*, 395-401.

49. Scaffidi, P., Misteli, T., and Bianchi, M. E. (2002) Release of chromatin protein HMGB1 by necrotic cells triggers inflammation, *Nature 418*, 191-195.

50. Ueda, T., Chou, H., Kawase, T., Shirakawa, H., and Yoshida, M. (2004) Acidic C-tail of HMGB1 is required for its target binding to nucleosome linker DNA and transcription stimulation, *Biochemistry 43*, 9901-9908.

51. Kawase, T., Sato, K., Ueda, T., and Yoshida, M. (2008) Distinct domains in HMGB1 are involved in specific intramolecular and nucleosomal interactions, *Biochemistry 47*, 13991-13996.

52. Cato, L., Stott, K., Watson, M., and Thomas, J. O. (2008) The interaction of HMGB1 and linker histones occurs through their acidic and basic tails, *J Mol Biol 384*, 1262-1272.

53. Catez, F., Yang, H., Tracey, K. J., Reeves, R., Misteli, T., and Bustin, M. (2004) Network of dynamic interactions between histone H1 and high-mobility-group proteins in chromatin, *Mol Cell Biol 24*, 4321-4328.

54. Rowell, J. P., Simpson, K. L., Stott, K., Watson, M., and Thomas, J. O. (2012) HMGB1-facilitated p53 DNA binding occurs via HMG-Box/p53 transactivation domain interaction, regulated by the acidic tail, *Structure 20*, 2014-2024.

55. Lange, S. S., and Vasquez, K. M. (2009) HMGB1: the jack-of-all-trades protein is a master DNA repair mechanic, *Mol Carcinog 48*, 571-580.

56. Czapla, L., Peters, J. P., Rueter, E. M., Olson, W. K., and Maher, L. J. (2011) Understanding apparent DNA flexibility enhancement by HU and HMGB architectural proteins, *J Mol Biol 409*, 278-289.

57. Assenberg, R., Webb, M., Connolly, E., Stott, K., Watson, M., Hobbs, J., and Thomas, J. O. (2008) A critical role in structure-specific DNA binding for the acetylatable lysine residues in HMGB1, *Biochem J 411*, 553-561.

58. Ugrinova, I., Mitkova, E., Moskalenko, C., Pashev, I., and Pasheva, E. (2007) DNA bending versus DNA end joining activity of HMGB1 protein is modulated in vitro by acetylation, *Biochemistry 46*, 2111-2117.

59. Dragan, A. I., Read, C. M., Makeyeva, E. N., Milgotina, E. I., Churchill, M. E., Crane-Robinson, C., and Privalov, P. L. (2004) DNA binding and bending by HMG boxes: energetic determinants of specificity, *J Mol Biol 343*, 371-393.

60. McCauley, M. J., Zimmerman, J., Maher, L. J., and Williams, M. C. (2007) HMGB binding to DNA: single and double box motifs, *J Mol Biol 374*, 993-1004.

61. Jamieson, E. R., Jacobson, M. P., Barnes, C. M., Chow, C. S., and Lippard, S. J. (1999) Structural and kinetic studies of a cisplatin-modified DNA icosamer binding to HMG1 domain B, *J Biol Chem 274*, 12346-12354.

62. Jung, Y., and Lippard, S. J. (2003) Nature of full-length HMGB1 binding to cisplatin-modified DNA, *Biochemistry 42*, 2664-2671.

63. Pil, P. M., and Lippard, S. J. (1992) Specific binding of chromosomal protein HMG1 to DNA damaged by the anticancer drug cisplatin, *Science 256*, 234-237.

64. Lakowicz, J. R. (2006) Principles of fluorescence spectroscopy, 1-954.

65. Moerner, W. E., and Fromm, D. P. (2003) Methods of single-molecule fluorescence spectroscopy and microscopy, *Rev Sci Instrum 74*, 3597.

66. Levitus, M., and Ranjit, S. (2011) Cyanine dyes in biophysical research: the photophysics of polymethine fluorescent dyes in biomolecular environments, *Q Rev Biophys 44*, 123-151.

67. Tsien, R. Y., Ernst, L., and Waggoner, A. (1995) Fluorophores for confocal microscopy: photophysics and photochemistry, in *Handbook of Biological Confocal Microscopy* (Pawley, J. B., Ed.) Plenum Press, New York.

68. Shaner, N. C., Steinbach, P. A., and Tsien, R. Y. (2005) A guide to choosing fluorescent proteins, *Nat Methods 2*, 905-909.

69. Clapp, A. R., Medintz, I. L., and Mattoussi, H. (2006) Forster resonance energy transfer investigations using quantum-dot fluorophores, *Chemphyschem 7*, 47-57.

70. Barroso, M. M. (2011) Quantum dots in cell biology, *J Histochem Cytochem 59*, 237-251.

71. Iqbal, A., Arslan, S., Okumus, B., Wilson, T. J., Giraud, G., Norman, D. G., Ha, T., and Lilley, D. M. (2008) Orientation dependence in fluorescent energy transfer between Cy3 and Cy5 terminally attached to double-stranded nucleic acids, *Proc Natl Acad Sci U S A 105*, 11176-11181.

72. Walter, N. G., Huang, C. Y., Manzo, A. J., and Sobhy, M. A. (2008) Do-it-yourself guide: how to use the modern single-molecule toolkit, *Nat Methods 5*, 475-489.

73. Aitken, C. E., Marshall, R. A., and Puglisi, J. D. (2008) An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments, *Biophys J 94*, 1826-1835.

74. Dave, R., Terry, D. S., Munro, J. B., and Blanchard, S. C. (2009) Mitigating unwanted photophysical processes for improved single-molecule fluorescence imaging, *Biophys J 96*, 2371-2381.

75. Rasnik, I., McKinney, S. A., and Ha, T. (2006) Nonblinking and long-lasting single-molecule fluorescence imaging, *Nat Methods 3*, 891-893.

76. Sinkeldam, R. W., Greco, N. J., and Tor, Y. (2010) Fluorescent analogs of biomolecular building blocks: design, properties, and applications, *Chem Rev 110*, 2579-2619.

77. Johnson, I., and Spence, M. T. Z. (2010) Molecular Probes Handbook, A Guide to Fluorescent Probes and Labeling Technologies, 11th Edition,

78. Wazawa, T., and Ueda, M. (2005) Total internal reflection fluorescence microscopy in single molecule nanobioscience, *Adv Biochem Eng Biot 95*, 77-106.

79. Axelrod, D., Burghardt, T. P., and Thompson, N. L. (1984) Total Internal Reflection Fluorescence, *Annu. Rev. Biophys. Bioeng. 13*, 247-268.

80. Mashanov, G. I., Tacon, D., Knight, A. E., Peckham, M., and Molloy, J. E. (2003) Visualizing single molecules inside living cells using total internal reflection fluorescence microscopy, *Methods 29*, 142-152.

81. Peterman, E. J., Sosa, H., and Moerner, W. E. (2004) Single-molecule fluorescence spectroscopy and microscopy of biomolecular motors, *Annu Rev Phys Chem 55*, 79-96.

82. Starr, D. B., and Hawley, D. K. (1991) TFIID binds the minor groove of the TATA box, *Cell 67*, 1231-1240.

83. Lee, D. K., Horikoshi, M., and Roeder, R. G. (1991) Interaction of TFIID in the minor groove of the TATA element, *Cell 67*, 1241-1250.

84. Kim, Y., Geiger, J. H., Hahn, S., and Sigler, P. B. (1993) Crystal structure of a yeast TBP/TATA-box complex, *Nature 365*, 512-520.

85. Nikolov, D. B., Chen, H., Halay, E. D., Hoffman, A., Roeder, R. G., and Burley, S. K. (1996) Crystal structure of a human TATA box-binding protein/TATA element complex, *Proc Natl Acad Sci USA 93*, 4862-4867.

86. Kim, J. L., Nikolov, D. B., and Burley, S. K. (1993) Co-crystal structure of TBP recognizing the minor groove of a TATA element, *Nature 365*, 520-527.

87. Masters, K. M., Parkhurst, K. M., Daugherty, M. A., and Parkhurst, L. J. (2003) Native human TATA-binding protein simultaneously binds and bends promoter DNA without a slow isomerization step or TFIIB requirement, *J Biol Chem 278*, 31685-31690.

88. Zhao, X., and Herr, W. (2002) A regulated two-step mechanism of TBP binding to DNA: a solvent-exposed surface of TBP inhibits TATA box recognition, *Cell 108*, 615-627.

89. Hoopes, B. C., LeBlanc, J. F., and Hawley, D. K. (1992) Kinetic analysis of yeast TFIID-TATA box complex formation suggests a multi-step pathway, *J Biol Chem 267*, 11539-11547.

90. Parkhurst, K. M., Brenowitz, M., and Parkhurst, L. J. (1996) Simultaneous binding and bending of promoter DNA by the TATA binding protein: real time kinetic measurements, *Biochemistry 35*, 7459-7465.

91. Parkhurst, K. M., Richards, R. M., Brenowitz, M., and Parkhurst, L. J. (1999) Intermediate species possessing bent DNA are present along the pathway to formation of a final TBP-TATA complex, *J Mol Biol 289*, 1327-1341.

92. Tolic-Norrelykke, S. F., Rasmussen, M. B., Pavone, F. S., Berg-Sorensen, K., and Oddershede, L. B. (2006) Stepwise bending of DNA by a single TATA-box binding protein, *Biophys J 90*, 3694-3703.

93. Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C., and Ha, T. (2008) Advances in single-molecule fluorescence methods for molecular biology, *Annu Rev Biochem 77*, 51-76.

94. Shi, J., Dertouzos, J., Gafni, A., and Steel, D. (2008) Application of single-molecule spectroscopy in studying enzyme kinetics and mechanism, *Methods Enzymol 450*, 129-157.

95. Tinoco, I. J., and Gonzalez, R. L. J. (2011) Biological mechanisms, one molecule at a time, *Genes Dev 25*, 1205-1231.

96. Patikoglou, G. A., Kim, J. L., Sun, L., Yang, S. H., Kodadek, T., and Burley, S. K. (1999) TATA element recognition by the TATA box-binding protein has been conserved throughout evolution, *Genes Dev 13*, 3217-3230.

97. Tan, S., Hunziker, Y., Sargent, D. F., and Richmond, T. J. (1996) Crystal structure of a yeast TFIIA/TBP/DNA complex, *Nature 381*, 127-134.

98. Geiger, J. H., Hahn, S., Lee, S., and Sigler, P. B. (1996) Crystal structure of the yeast TFIIA/TBP/DNA complex, *Science 272*, 830-836.

99. Grove, A., Galeone, A., Yu, E., Mayol, L., and Geiduschek, E. P. (1998) Affinity, stability and polarity of binding of the TATA binding protein governed by flexure at the TATA Box, *J Mol Biol 282*, 731-739.

100. Hoopes, B. C., LeBlanc, J. F., and Hawley, D. K. (1998) Contributions of the TATA box sequence to rate-limiting steps in transcription initiation by RNA polymerase II, *J Mol Biol 277*, 1015-1031.

101. Cox, J. M., Hayward, M. M., Sanchez, J. F., Gegnas, L. D., van der Zee, S., Dennis, J. H., Sigler, P. B., and Schepartz, A. (1997) Bidirectional binding of the TATA box binding protein to the TATA box, *Proc Natl Acad Sci U S A 94*, 13475-13480.

102. Weaver, J. R., Kugel, J. F., and Goodrich, J. A. (2005) The sequence at specific positions in the early transcribed region sets the rate of transcript synthesis by RNA polymerase II in vitro, *J Biol Chem 280*, 39860-39869.

103. Ozer, J., Moore, P. A., Bolden, A. H., Lee, A., Rosen, C. A., and Lieberman, P. M. (1994) Molecular cloning of the small (gamma) subunit of human TFIIA reveals functions critical for activated transcription, *Genes Dev 8*, 2324-2335.

104. Sun, X., Ma, D., Sheldon, M., Yeung, K., and Reinberg, D. (1994) Reconstitution of human TFIIA activity from recombinant polypeptides: a role in TFIID-mediated transcription, *Genes Dev 8*, 2336-2348.

105. Joo, C., and Ha, T. (2008) Single-molecule FRET with total internal reflection microscopy, in *Single-Molecule Techniques: A Laboratory Manual* (Selvin, P. R., and Ha, T., Eds.) pp 3-36, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

106. Zimmerman, J., and Maher, L. J. (2008) Transient HMGB protein interactions with B-DNA duplexes and complexes, *Biochem Biophys Res Commun 371*, 79-84.

107. Skoko, D., Wong, B., Johnson, R. C., and Marko, J. F. (2004) Micromechanical analysis of the binding of DNA-bending proteins HMGB1, NHP6A, and HU reveals their ability to form highly stable DNA-protein complexes, *Biochemistry 43*, 13867-13874.

108. Pallier, C., Scaffidi, P., Chopineau-Proust, S., Agresti, A., Nordmann, P., Bianchi, M. E., and Marechal, V. (2003) Association of chromatin proteins high mobility group box (HMGB) 1 and HMGB2 with mitotic chromosomes, *Mol Biol Cell 14*, 3414-3426.

109. Weir, H. M., Kraulis, P. J., Hill, C. S., Raine, A. R., Laue, E. D., and Thomas, J. O. (1993) Structure of the HMG box motif in the B-domain of HMG1, *EMBO J 12*, 1311-1319.

110. Read, C. M., Cary, P. D., Crane-Robinson, C., Driscoll, P. C., and Norman, D. G. (1993) Solution structure of a DNA-binding domain from HMG1, *Nucleic Acids Res 21*, 3427-3436.

111. Hardman, C. H., Broadhurst, R. W., Raine, A. R., Grasser, K. D., Thomas, J. O., and Laue, E. D. (1995) Structure of the A-domain of HMG1 and its interaction with DNA as studied by heteronuclear three- and four-dimensional NMR spectroscopy, *Biochemistry 34*, 16596-16607.

112. Thomas, J. O., and Travers, A. A. (2001) HMG1 and 2, and related ‘architectural’ DNA-binding proteins, *Trends Biochem Sci 26*, 167-174.

113. Muller, S., Bianchi, M. E., and Knapp, S. (2001) Thermodynamics of HMGB1 interaction with duplex DNA, *Biochemistry 40*, 10254-10261.

114. Paull, T. T., Haykinson, M. J., and Johnson, R. C. (1993) The nonspecific DNA-binding and -bending proteins HMG1 and HMG2 promote the assembly of complex nucleoprotein structures, *Genes Dev 7*, 1521-1534.

115. Ramstein, J., Locker, D., Bianchi, M. E., and Leng, M. (1999) Domain-domain interactions in high mobility group 1 protein (HMG1), *Eur J Biochem 260*, 692-700.

116. Watson, M., Stott, K., and Thomas, J. O. (2007) Mapping intramolecular interactions between domains in HMGB1 using a tail-truncation approach, *J Mol Biol 374*, 1286-1297.

117. Stott, K., Watson, M., Howe, F. S., Grossmann, J. G., and Thomas, J. O. (2010) Tail-mediated collapse of HMGB1 is dynamic and occurs via differential binding of the acidic tail to the A and B domains, *J Mol Biol 403*, 706-722.

118. Knapp, S., Muller, S., Digilio, G., Bonaldi, T., Bianchi, M. E., and Musco, G. (2004) The long acidic tail of high mobility group box 1 (HMGB1) protein forms an extended and flexible structure that interacts with specific residues within and between the HMG boxes, *Biochemistry 43*, 11992-11997.

119. Wagner, J. P., Quill, D. M., and Pettijohn, D. E. (1995) Increased DNA-bending activity and higher affinity DNA binding of high mobility group protein HMG-1 prepared without acids, *J Biol Chem 270*, 7394-7398.

120. Yamamoto, A., Ando, Y., Yoshioka, K., Saito, K., Tanabe, T., Shirakawa, H., and Yoshida, M. (1997) Difference in affinity for DNA between HMG proteins 1 and 2 determined by surface plasmon resonance measurements, *J Biochem 122*, 586-594.

121. Saito, K., Kikuchi, T., Shirakawa, H., and Yoshida, M. (1999) The stabilized structural array of two HMG1/2-boxes endowed by a linker sequence between them is requisite for the effective binding of HMG1 with DNA, *J Biochem 125*, 399-405.

122. Belgrano, F. S., de Abreu da Silva, I. C., Bastos de Oliveira, F. M., Fantappie, M. R., and Mohana-Borges, R. (2013) Role of the Acidic Tail of High Mobility Group Protein B1 (HMGB1) in Protein Stability and DNA Bending, *PLoS One 8*, e79572.

123. Stros, M. (1998) DNA bending by the chromosomal protein HMG1 and its high mobility group box domains. Effect of flanking sequences, *J Biol Chem 273*, 10355-10361.

124. Grasser, K. D., Teo, S. H., Lee, K. B., Broadhurst, R. W., Rees, C., Hardman, C. H., and Thomas, J. O. (1998) DNA-binding properties of the tandem HMG boxes of high-mobility-group protein 1 (HMG1), *Eur J Biochem 253*, 787-795.

125. Teo, S. H., Grasser, K. D., and Thomas, J. O. (1995) Differences in the DNA-binding properties of the HMG-box domains of HMG1 and the sex-determining factor SRY, *Eur J Biochem 230*, 943-950.

126. Stott, K., Tang, G. S., Lee, K. B., and Thomas, J. O. (2006) Structure of a complex of tandem HMG boxes and DNA, *J Mol Biol 360*, 90-104.

127. Pil, P. M., Chow, C. S., and Lippard, S. J. (1993) High-mobility-group 1 protein mediates DNA bending as determined by ring closures, *Proc Natl Acad Sci USA 90*, 9465-9469.

128. Blair, R. H., Goodrich, J. A., and Kugel, J. F. (2013) Using FRET to monitor protein-induced DNA bending: the TBP-TATA complex as a model system, *Methods Mol Biol 977*, 203-215.

129. Meyers, R. E., and Sharp, P. A. (1993) TATA-binding protein and associated factors in polymerase II and polymerase III transcription, *Mol Cell Biol 13*, 7953-7960.

130. Biggs, J. R., Ahn, N. G., and Kraft, A. S. (1998) Activation of the mitogen-activated protein kinase pathway in U937 leukemic cells induces phosphorylation of the amino terminus of the TATA-binding protein, *Cell Growth Differ 9*, 667-676.

131. Chibazakura, T., Watanabe, F., Kitajima, S., Tsukada, K., Yasukochi, Y., and Teraoka, H. (1997) Phosphorylation of human general transcription factors TATA-binding protein and transcription factor IIB by DNA-dependent protein kinase--synergistic stimulation of RNA polymerase II basal transcription in vitro, *Eur J Biochem 247*, 1166-1173.

132. Solow, S. P., Lezina, L., and Lieberman, P. M. (1999) Phosphorylation of TFIIA stimulates TATA binding protein-TATA interaction and contributes to maximal transcription and viability in yeast, *Mol Cell Biol 19*, 2846-2852.

APPENDIX

The phosphorylation state of TBP does not affect the amount of DNA bending

TBP can be phosphorylated in cells, potentially by ERK (130) and DNA-dependent Protein Kinase (DNA-PK) (131). The phosphorylation of TBP has been shown to increase transcriptional activity in vitro (131), potentially by stimulating formation of the PIC. The goal of this project was to determine whether phosphorylated TBP bent DNA differently or had a different affinity for consensus TATA DNA.

Experimental Procedures

Phosphorylation of TBP

To ensure that ERK could phosphorylate TBP, duplicates of 4 phosphorylation reactions were set up. All reactions contained 0.5 mM ATP, 10% glycerol, 10 mM Tris (pH 7.9), 50 mM KCl, 1 mM DTT, 0.5 mg/mL BSA, 10 mM HEPES (pH 7.9), and 5 mM $MgCl_2$ in 10 μL. When applicable, 530 nM TBP, and 1.6 μM ERK were added. The reactions were incubated at 30°C for 15 or 30 min. Reactions were stopped by adding 10 μL 2X SDS loading buffer and boiled for 3 min at 95°C. Each sample was loaded onto a 10% SDS-PAGE gel. Half the gel was silver stained, while the other half of the gel was used in a Western blot with an anti-TBP antibody (abcam 818).

SmFRET experiment

We followed the same procedure as described in Chapter 3, but using 134 mM KCl, 6.5 mM NaCl, and 0.3 mM DTT. Phosphorylated TBP (25 nM), using the reaction conditions described above (so it contained 75 nM ERK), was flowed into the chamber and imaged. To determine the rate constants for unbending, we used all bent dwell times, even those that photobleached.

Results and Discussion

We first wanted to determine whether ERK would phosphorylate TBP. We incubated TBP with ERK and ATP for 15 or 30 min. A control for this experiment was to leave the TBP out of the reaction to see if ERK phosphorylated itself. The silver stained gel of these phosphorylation reactions is shown in Figure A.1A, where lane 1 contains unphosphorylated TBP, compared to the phosphorylated TBP in lanes 3 and 4. The unphosphorylated band disappears and higher molecular weight phosphorylated TBP emerges. To ensure that those higher molecular weight bands were TBP, we used an anti-TBP antibody to Western blot for the protein (Figure A.1B), where unphosphorylated TBP is shown in lane 1 and phosphorylated TBP is shown in lanes 3 and 4. This assay indicated that ERK phosphorylates TBP.

We studied how phosphorylated TBP behaved in our single-molecule system and whether it interacts with the DNA differently than the unphosphorylated form. We flowed phosphorylated TBP into the chamber with the consensus TATA DNA used in Chapter 3. After imaging, we determined that the mean FRET peak for unbent consensus DNA was 0.26, and bent DNA was 0.38 in the presence of phosphorylated TBP (Figure A.2). These results are consistent with unphosphorylated TBP, where the mean FRET peaks were 0.26 and 0.39, respectively (Figure 3.2B). The bent states are similar between unphosphorylated and phosphorylated TBP, meaning that the phosphorylation did not change the extent of bending.

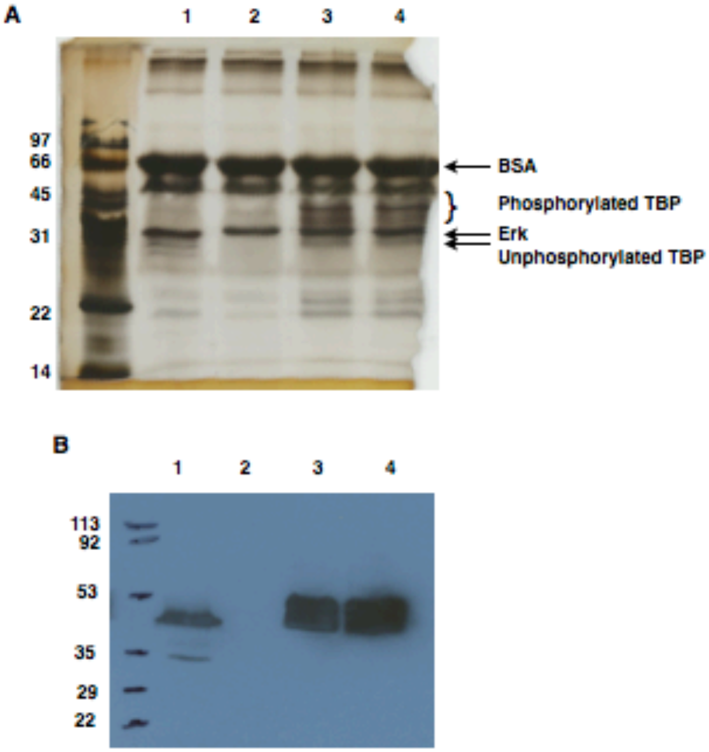


Figure A.1. ERK phosphorylates TBP. **(A)** A silver stained SDS gel of different phosphorylation states of TBP. **(B)** A Western blot for TBP. The contents of the lanes in both gels are: unphosphorylated TBP without ERK (lane 1), ERK protein without TBP (lane 2), TBP incubated with ERK for 15 min (lane 3), and TBP incubated with ERK for 30 min (lane 4).

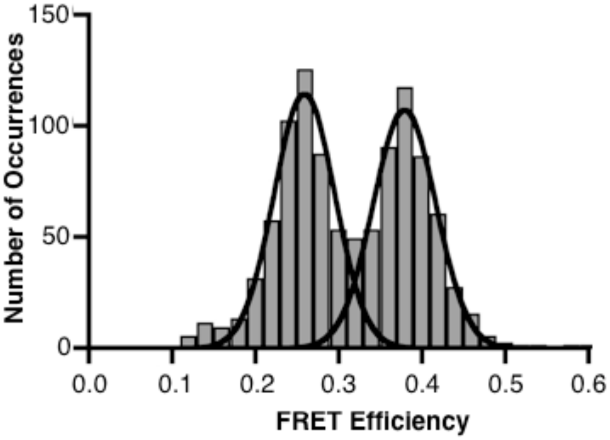


Figure A.2. Phosphorylated TBP/DNA complexes are bent to the same extent as unphosphorylated TBP/DNA complexes. In the presence of phosphorylated TBP, the DNA molecules still exist in a two-state population (unbent and bent). Time traces for 341 DNA molecules were analyzed and 1000 distinct FRET occurrences were plotted as a histogram and fit with Gaussians. The low and high FRET states have mean FRET efficiencies of 0.26 and 0.38, respectively.

Previous experiments have shown that the amino terminus of TBP is phosphorylated by ERK, where there are 8 potential phosphorylation sites, but it appears that only 4 sites at the N-terminus are phosphorylated by ERK (130). Here, we see that TBP bending promotor DNA is not affected by phosphorylation, indicating that its interaction with promoter DNA may be unaffected by phosphorylation. The phosphorylation of TBP by DNA-PK does not affect TBP and TFIIB binding to promoter DNA. In these experiments, phosphorylation stimulated the formation of basal transcription complexes and transcription from the AdML promoter (131), which we did not test.

We were next interested in whether the bending kinetics were altered with phosphorylated TBP. The unbent dwell times were binned, histogrammed, and analyzed with a monoexponential decay. The pseudo-first order rate constant of phosphorylated TBP bending DNA was 0.083 ± 0.02 s^{-1} (Figure A3.A). We did the same with the bent dwell times, fitting the data with both monoexponential and biexponential decay cureves. The latter determined two unbending rate constants of 0.33 ± 0.1 s^{-1} and 0.032 ± 0.02 s^{-1} for phosphorylated TBP (Figure A3.B). We used the biexponential fit because there are data with really long dwell times that would not otherwise fit in the monoexponential decay (data not shown).

The bending and unbending rates for both forms of TBP are slightly different. As in Chapter 3, we accounted for the concentration of TBP to give us an estimate of a second order on rate constant for phosphorylated TBP binding and bending DNA: 3.3×10^6 $M^{-1}s^{-1}$. The second order rate constant for unphosphorylated TBP was 1×10^6 $M^{-1}s^{-1}$ (determined in Chapter 3), and the unbending constants were 0.19 ± 0.03 s^{-1} and 0.019 ± 0.008 s^{-1}. The apparent on rate constant for phosphorylated TBP is 3-fold faster than unphosphorylated, and the apparent off rate constant

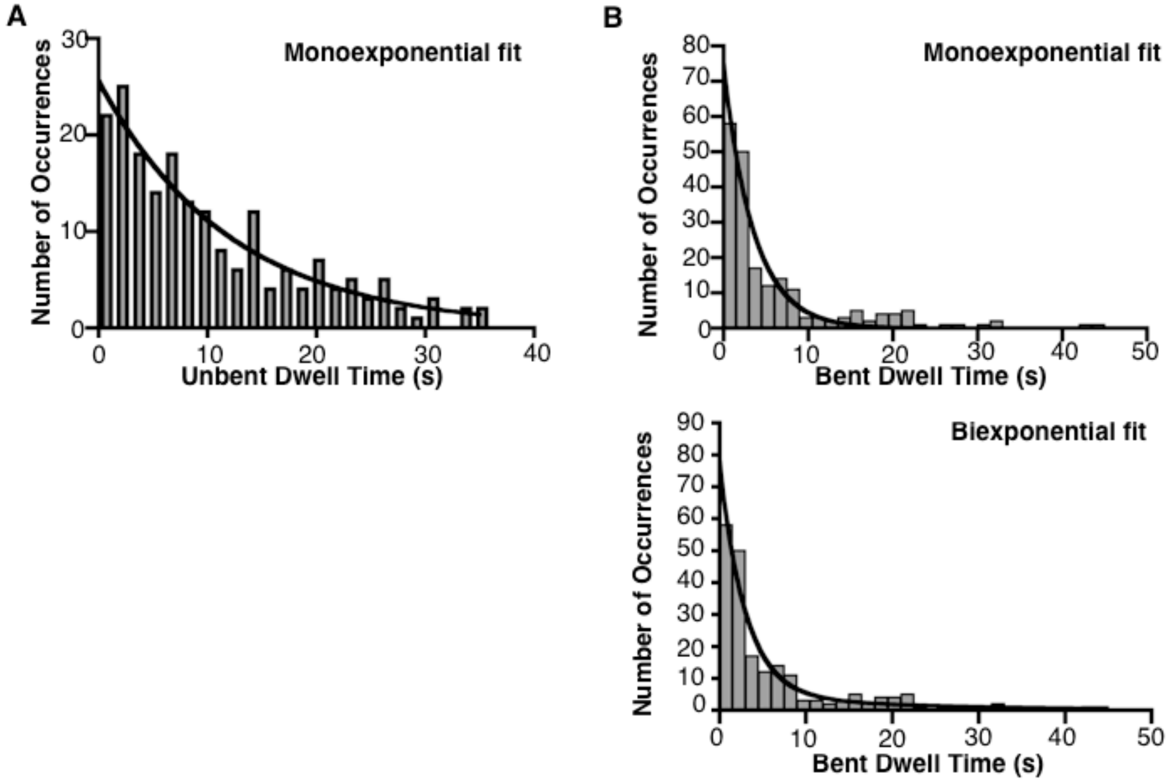


Figure A.3. Unbending consensus TATA DNA is biphasic. **(A)** Analysis of the rate at which phosphorylated TBP bends the consensus TATA box. Dwell times for the unbent occurrences observed at 134 mM KCl were histogrammed and fit to a monoexponential. The observed rate constant is 0.083 ± 0.02 s^{-1}, where the error represents the 95% confidence interval. **(B)** Unbending of the consensus TATA box occurs in two distinct kinetic phases. Dwell times for the bent occurrences observed at 134 mM KCl were histogrammed and fit to a monoexponential (upper panel) or a biexponential (lower panel). The two rate constants obtained from the biexponential fit are 0.33 ± 0.1 s^{-1} and 0.032 ± 0.02 s^{-1}, where the error represents the 95% confidence interval for phosphorylated TBP is less than 2-fold slower than unphosphorylated TBP. Whether or not these differences are significant needs to be determined. The N-terminus of TBP is known to interact with the DNA binding domain in the C-terminus of TBP, resulting in slower binding of TBP to DNA. Potentially, phosphorylation of the N-terminus reduces its interaction with the DNA binding domain, allowing for faster binding by phosphorylated TBP to DNA. A longer off-rate could play into the idea that phosphorylated TBP stimlates in vitro transcription by being at the promoter longer (131).

A good follow up study would be to look at the phosphorylation of TFIIA and the stability of the TBP-DNA complex. It has been shown that the phosphorylation of TFIIA increases the stability of the TBP-DNA complex, which may be important in cells for increased transcriptional activity (132). To ensure that TBP-DNA stability is in a range suitable for single molecule studies in the presence of TFIIA, these studies would need to be performed on nonconsensus TATA sequences.

TFIIA does not bind DNA by itself

It is known that TFIIA does not to bind DNA independently. TFIID or TBP has to be present for TFIIA to stably interact with a DNA/protein complex. To ensure that we did not see DNA bending by TFIIA, independently, a control of TFIIA-only was performed.

Experimental Procedure

The single molecule procedures found in Chapter 3 were performed, but the NaCl concentration was 300-400 mM instead of 150 mM, because of how the trolox was dissolved. TBP concentration was 230 nM, and 67 nM TFIIA was used.

Results and Discussion

Since the salt and protein concentrations were different from the conditions in Chapter 3, we wanted to ensure that TBP binds under these conditions or the TFIIA control would be useless. The mean FRET efficiency of TATA DNA without protein in solution was 0.23 (data not shown), which was as expected from Chapter 3. When TBP was flowed into the sample chamber (Figure A.4A.), the mean FRET efficiency of the unbent peak was 0.23, and the bent DNA peak had a mean FRET efficiency of 0.37. This is consistent with previous results in Chapter 3, where TBP bends DNA. Further, to show that TFIIA stabilizes the TBP-DNA interaction and increases the amount of DNA molecules in the bent state, TBP and TFIIA were flowed into the chamber together under these conditions (Figure A.4B). Here, the number of TBP-dependent bent DNA complexes increased in the presence of TFIIA, as seen in Chapter 3 as well. To ensure that TFIIA did not interact with and bend the DNA by itself, TFIIA alone was flowed into the sample

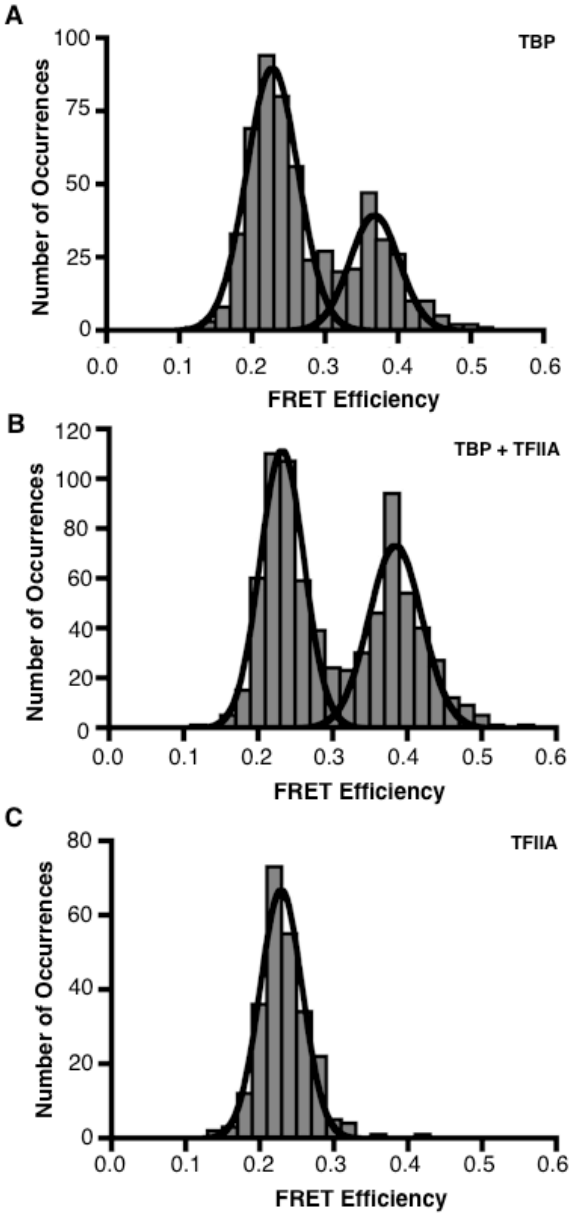


Figure A.4. TFIIA does not bind TATA DNA without TBP present. **(A)** TBP binds TATA DNA, where the lower Gaussian gives a mean FRET peak at 0.23 and the higher mean FRET efficiency is 0.37. There are 570 states plotted. **(B)** TFIIA is active and aids in TBP binding and bending TATA DNA. The lower, unbent DNA peak gives a FRET efficiency of 0.24 and the higher peak has a FRET efficiency of 0.38. There are 762 states binned in this graph. **(C)** TFIIA only in solution does not bend the DNA. 248 DNA molecules give a single Gaussian with a mean FRET efficiency of 0.23.

chamber (Figure A.4C). There was no evidence of TFIIA bending the TATA DNA alone, since the only population in solution appeared to be unbent DNA with a mean FRET efficiency of 0.23. This control indicates that TFIIA does not bend TATA DNA, without another factor present.