DUAL-SPECTRAL INTERFEROMETRIC SENSOR
FOR QUANTITATIVE STUDY OF PROTEIN-DNA INTERACTIONS

by

XIRUI ZHANG
B.S., Tianjin University, 2009
M.S., Boston University, 2012

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
2014
Approved by

First Reader
M. Selim Ünlü, Ph.D.
Professor of Electrical and Computer Engineering
Professor of Physics
Professor of Biomedical Engineering

Second Reader
Bennett B Goldberg, Ph.D.
Professor of Physics
Professor of Electrical and Computer Engineering
Professor of Biomedical Engineering
Professor of Science Education

Third Reader
Michael L. Smith, Ph.D.
Associate Professor of Biomedical Engineering

Fourth Reader
Allison M. Dennis, Ph.D.
Assistant Professor of Biomedical Engineering
Assistant Professor of Materials Science and Engineering
“Life ... is a relationship between molecules.”

–Linus Pauling
To Mom and Dad,

for teaching me to learn and dream.

To Yihui,

for all the love and support.
ACKNOWLEDGMENTS

When I first came to Boston to pursue my PhD degree five years ago, I did not know anyone in this city. Since the very first night when I arrived at this foreign land, I have been blessed with great advisors, mentors, colleagues, and friends, who wonderfully shaped my life here and brought me to this moment. First, I am deeply grateful to my advisor, Professor Selim Ünlü, for recognizing my potentials after my first year rotation in the lab, giving me a research home, and letting me choose the project I was interested in. His supervision and support during my years in the lab helped me become a better scientist and engineer and grow as a mature woman.

I am also grateful to my committee members, Professor Bennett Goldberg, Professor Michael Smith, and Professor Allison Dennis for their guidance and support during my PhD candidacy and dissertation preparation. I owe a great deal of thanks to Professor Irving Bigio for his recognition of my potentials in research and for offering me a great research and learning experience in the Biomedical Optics Lab in my first year.

Many thanks to my collaborators, Professor Marcella Chiari and Dr. Laura Sola, for their fruitful discussions and help in applying the polymeric coating used for functionalizing the sensor substrate in this dissertation. I would also like to thank Professor Peter Dröge for generously providing the protein for my study and the many helpful discussions related to my experiments and potential applications of the biosensing platform presented in this dissertation.

I would like to give my sincere thanks to my mentors and colleagues in the group for their help, support, and friendship. I had the privilege to work in an extremely
interdisciplinary group consisted of many talented, creative, and diligent scientists and engineers. I was very fortunate to be supervised by Dr. Philipp Spuhler. The combination of Phil’s expertise in engineering, patience in teaching, and consistent encouragement helped me quickly learn and work on the different components of my dissertation project that he initiated. Phil also looks after me like a little sister. He keeps checking in on me and giving me valuable advice when I lose directions. My sincere gratitude also goes to Dr. Sunmin Ahn and Dr. David Freedman for their always-available guidance, help, and support in both research and life. I would also like to thank Dr. George Daaboul for his ideas and contributions to this dissertation. My gratitude also goes to Dr. Abdulkadir Yurt for his help and advice in the optical setup, and to Jacob Trueb and Steve Scherr for their help in designing the flow cell used in this dissertation.

My sincere gratitude also goes to my mentors and friends Dr. Yingjie Sun, Dr. Hao Wang, and Dr. Christen Mulvey for their endless help, guidance, and friendship since the very beginning of my PhD candidacy. I would also like to extend my gratitude to all my friends in the Boston area for making my life in this foreign city colorful and enjoyable.

Finally, I thank my parents for believing in me and bringing up as an independent woman possessing a PhD in engineering. I thank my boyfriend, Yihui, for helping me go through the most difficult times during my PhD candidacy and shaping me to a better person.
DUAL-SPECTRAL INTERFEROMETRIC SENSOR FOR QUANTITATIVE STUDY OF PROTEIN-DNA INTERACTIONS

XIRUI ZHANG

Boston University, College of Engineering, 2014

Major Professor: M. Selim Ünlü, Ph.D., Professor of Electrical and Computer Engineering, Professor of Physics, Professor of Biomedical Engineering

ABSTRACT

The maintenance and functions of the genome are facilitated by DNA-binding proteins, whose specific binding mechanisms are not yet fully understood. Recently, it was discovered that the recognition and capture of DNA conformational flexibility and deformation by DNA-binding proteins serve as an indirect readout mechanism for specific recognition and facilitate important cellular functions. Various biophysical techniques have been employed to elucidate this conformational specificity of protein-DNA interactions. These techniques are not sufficiently high-throughput to perform systematic investigation of various protein-DNA complexes and their functions. Microarray-based high-throughput methods enable large-scale and comprehensive evaluation of the binding affinities of protein-DNA interactions, but do not provide conformational information.

In this dissertation, we developed a tool that enables high-throughput quantification of both conformational specificity and binding affinity of protein-DNA interactions. Our approach is to combine quantitative detection of DNA conformational change and protein-DNA binding in a DNA microarray format. The DNA conformational
change is measured by spectral self-interference fluorescence microscopy that determines surface-immobilized DNA conformation by measuring axial height of fluorophores tagged to specific nucleotides. The amount of bound protein and DNA are measured by white light reflectance spectroscopy that quantifies molecular surface densities by measuring biomolecule layer thicknesses. By implementing a dual-spectral imaging configuration, we can perform the two independent interferometric measurements in parallel using two separate spectral bandwidths.

We used the \textit{E. coli} integration host factor protein, an architectural protein that induces substantial DNA bending, as the model system to demonstrate parallel quantification of DNA conformational change and protein-DNA binding. First, we characterized the conformation of surface-immobilized DNA to establish a foundation for DNA conformational change detection. We also proposed a quantitative model to resolve conformational specific binding and nonspecific binding. Based on the model, we evaluated factors affecting conformational specific protein-DNA interactions on a surface and demonstrated distinguished and parallel detection of conformational specific and nonspecific binding.

The technology demonstrated in this dissertation can be developed as a rapid and convenient method for quantitative and large-scale screening of conformational specific protein-DNA complexes. When combined with computational methods, this technology can facilitate mechanistic and systematic study of protein-DNA interactions and their functions.
# Table of Contents

## Chapter 1

Introduction and motivation

1. The importance of studying protein-DNA interactions
   - 1.1 Overview of protein-DNA interactions
   - 1.2 DNA conformation and functional protein binding
   - 1.3 A good model: Integration Host Factor (IHF)

2. Biophysical techniques for studying protein-DNA interactions
   - 2.1 Biophysical techniques
   - 2.2 High-throughput methods
   - 2.3 Computational methods

3. Motivation: high-throughput biophysical study of protein-DNA interactions

4. Overview of this dissertation

## Chapter 2

Spectral self-interference fluorescence microscopy (SSFM)

1. Fundamentals of spectral self-interference fluorescence microscopy
   - 1.1 Background and development
   - 1.2 Physical model
   - 1.3 Combine SSFM with white light reflectance spectroscopy
   - 1.4 Spectrum analysis

2. Previous applications
   - 2.1 Estimate the conformation of surface-immobilized DNA
2.2.2 Geometric model for DNA orientation calculation.......................... 40
2.2.3 Quantification and control of surface-immobilized DNA orientation ...... 44
2.3 Dual-color SSFM.............................................................................. 54
  2.3.1 Experimental approach ................................................................. 56
  2.3.2 SiO$_2$ thickness design................................................................. 65
  2.3.3 Measure nanometer scale steps .................................................. 68
  2.3.4 Characterization of DNA immobilized on a 3-D polymeric surface ........ 71
  2.3.5 Real-time quantification of DNA hybridization and conformational change.. 75
  2.3.6 Advantages, limitations, and future developments .......................... 79

CHAPTER 3 ..................................................................................... 83

Dual-spectral imaging for the study of protein-DNA interactions............ 83
  3.1 One-dimensional spectral imaging configuration............................. 83
  3.2 Experimental approach .................................................................. 86
    3.2.1 Optical setup ............................................................................. 86
    3.2.2 Spectrum analysis ..................................................................... 90
    3.2.3 Materials.................................................................................. 92
    3.2.4 Experiment design................................................................. 95
  3.3 Quantification of IHF-DNA interaction......................................... 99
    3.3.1 Detection of IHF-DNA interaction at equilibrium..................... 99
    3.3.2 Quantitative analysis............................................................... 112
  3.4 Conclusions.................................................................................. 115

CHAPTER 4 ..................................................................................... 116
Characterization of protein-DNA interactions on a surface.......................... 116

4.1 Motivation........................................................................................................ 116

4.2 Factors affecting binding and conformational change................................. 118
  4.2.1 DNA orientation......................................................................................... 118
  4.2.2 DNA surface density................................................................................ 121
  4.2.3 DNA length and binding site location....................................................... 125
  4.2.4 Ionic strength and short spacer DNA...................................................... 131
  4.2.5 Conclusion ............................................................................................... 135

4.3 Effect of extra macromolecule layer on fluorophore height determination.... 136

CHAPTER 5 ............................................................................................................ 143

Study of specific and nonspecific protein-DNA interactions ......................... 143

5.1 Motivation........................................................................................................ 143

5.2 Parallel detection of specific and nonspecific binding of IHF ..................... 143
  5.2.1 Examining salt-dependence of IHF-DNA interactions............................. 145
  5.2.2 Quantifying DNA bending angle induced by IHF specific binding .......... 150
  5.2.3 Discriminating consensus sequence with a single mutation ................. 154

5.3 Specific and nonspecific protein-DNA binding and functional regulation .... 159

CHAPTER 6 ............................................................................................................ 161

Conclusions and future directions................................................................. 161

6.1 Conclusions.................................................................................................... 161

6.2 Future platform developments ..................................................................... 163
6.2.1 Advantages and limitations ................................................................. 163
6.2.2 Improve throughput and time resolution by implementing spectral filtering 165
6.2.3 Quantum Dots as substitutes for organic fluorophores ....................... 168
6.2.4 Improving fluidic system design .......................................................... 169

6.3 Future studies in protein-DNA interactions ........................................... 170
6.3.1 High-throughput systematic screening of regulatory binding sites ........ 170
6.3.2 Single-strand binding protein-DNA nonspecific binding ...................... 171
6.3.3 A proposed workflow ........................................................................... 172

BIBLIOGRAPHY .......................................................................................... 174

CURRICULUM VITAE .................................................................................. 190
LIST OF TABLES

Table 1 Average height of the fluorophores above the surface measured by SSFM and the optical thickness of the DNA layer on the surface measured by WLRS .................. 39

Table 2 DNA sequences and nomenclature used in dual-color SSFM ............................ 58

Table 3 DNA sequences and nomenclature used for studying IHF-DNA interactions. ... 93

Table 4 Nonlinear LSF fitting results of equilibrium binding isotherms of fluorophore height change and IHF to DNA binding ratio of different DNA sequences at low surface density in two buffers of different salt concentrations. ......................... 109

Table 5 Nonlinear LSF fitting results of equilibrium binding isotherms of fluorophore height change and IHF to DNA binding ratio of different DNA sequences at medium surface density in two buffers of different salt concentrations. ......................... 110

Table 6 Nonlinear LSF fitting results of equilibrium binding isotherms of fluorophore height change and IHF to DNA binding ratio of different DNA sequences at high surface density in two buffers of different salt concentrations. ......................... 111

Table 7 Dissociation time constant of 60-bp IHF binding sequence and Control sequence measured by SSFM and LED-RS. ................................................................. 149
LIST OF FIGURES

Figure 1 The structure of a nucleosome in a eukaryotic cell ........................................ 7
Figure 2 Crystal structure of TATA-binding protein bound to its consensus sequence..... 8
Figure 3 Crystal structure of Tag MutS bound to a mismatched DNA ............................ 9
Figure 4 Crystal structure of the IHF–H’ DNA complex............................................... 11
Figure 5 Compare fluorescence emission on a glass slide and on a reflecting substrate. 28
Figure 6 The classical dipole model of a fluorescence emitter........................................ 29
Figure 7 Dipole emission model showing the direct, incident and reflected waves ....... 30
Figure 8 Compare detection mechanism between WLRS and SSFM .......................... 34
Figure 9 A dsDNA molecule is modeled as a rigid rod............................................. 42
Figure 10 Approximation of the mean of dsDNA orientation ..................................... 43
Figure 11 Simulated relationship between DNA orientation and DNA surface density 44
Figure 12 Control and quantification of 60-bp dsDNA orientation ............................ 46
Figure 13 Probability Density Functions (PDF) of dsDNA orientation ...................... 49
Figure 14 Average dsDNA orientation θ and calculated dsDNA orientation θ ........ 50
Figure 15 dsDNA orientation over a wide range of salt concentrations ................... 52
Figure 16 Dual-color SSFM detection principle...................................................... 55
Figure 17 SSFM combined with LED based white light reflectance spectroscopy. 56
Figure 18 Uniform and consistent DNA spot morphology on the polymer surface ........ 60
Figure 19 Customized flow cell assembly ............................................................. 61
Figure 20 Schematic illustration of the dual-color SSFM optical setup .................. 63
Figure 21 Data acquisition software application developed using MATLAB ......... 64
Figure 22 LED-RS measurement of DNA spot thicknesses .................................. 65
Figure 23 The emission spectrums of the fluorophores and the LED. ...................... 66
Figure 24 Determine SiO$_2$ layer thickness from simulated analysis ....................... 68
Figure 25 Measurements of calibration chips with nanometer scale steps in air .......... 69
Figure 26 Measurements of calibration chips with nanometer scale steps in solution .... 70
Figure 27 Compare etching rates estimated by IRIS, dual-color SSFM, and LED-RS .... 71
Figure 28 Characterization of DNA molecules immobilized on a polymeric surface ..... 73
Figure 29 Real-time detection of DNA conformation and hybridization .................. 76
Figure 30 Spectral imaging configuration of SSFM .............................................. 85
Figure 31 Schematic illustration of spectral imaging configuration optical setup ......... 88
Figure 32 Dual-spectral imaging configuration combining SSFM and LED-RS .......... 90
Figure 33 Denaturing SDS-PAGE experiment to check integrity of IHF protein ........ 94
Figure 34 Determine the concentration of IHF stock solution by Bradford assay. ........ 95
Figure 35 Schematic illustration of IHF-DNA binding detection mechanism ............. 96
Figure 36 The idea of distinguishing IHF binding site location on 60-bp dsDNA .......... 97
Figure 37 Equilibrium binding isotherms of IHF to DNA sequence H’(39) ......... 101
Figure 38 Equilibrium binding isotherms of fluorophore height changes of different DNA sequences immobilized at low surface density measured by SSFM ............... 103
Figure 39 Equilibrium binding isotherms of fluorophore height changes of different DNA sequences immobilized at medium surface density measured by SSFM .......... 104
Figure 40 Equilibrium binding isotherms of fluorophore height changes of different DNA sequences immobilized at high surface density measured by SSFM ............... 105
Figure 41 Equilibrium binding isotherms of IHF to DNA binding ratio of different DNA sequences immobilized at low surface density measured by LED-RS.................. 106
Figure 42 Equilibrium binding isotherms of IHF to DNA binding ratio of different DNA sequences immobilized at medium surface density measured by LED-RS.......... 107
Figure 43 Equilibrium binding isotherms of IHF to DNA binding ratio of different DNA sequences immobilized at high surface density measured by LED-RS............. 108
Figure 44 Analysis of equilibrium isotherm of IHF to DNA binding ratio.............. 113
Figure 45 Compare adjusted $R^2$ of single Langmuir binding isotherm fit and additive Langmuir binding isotherm model fit............................................................. 114
Figure 46 Initial DNA orientation affects fluorophore height change measurements.... 119
Figure 47 Increasing DNA orientation by increasing DNA surface density does not increase the measured fluorophore height change by SSFM.............................. 120
Figure 48 DNA surface density effects on specific and nonspecific binding ratio ...... 123
Figure 49 The effect of DNA surface density on IHF-DNA dissociation constants ...... 123
Figure 50 The effect of surface density on fluorophore height changes..................... 124
Figure 51 Schematic illustration of the physical limits for binding site location......... 127
Figure 52 The lower limit of the binding site center location decreases as DNA orientation increases........................................................................................................ 127
Figure 53 SSFM can distinguish shifted H’ binding site location on 60-bp DNA by measuring IHF binding induced average fluorophore height changes of DNA ..... 130
Figure 54 The difference of fluorophore height changes of DNA sequences with shifted H’ binding sites is more evident if normalized by the specific binding ratio. ....... 131
Figure 55 The effect 20-bp spacer DNA molecules on the DNA orientation and IHF-induced fluorophore height changes in buffer containing 150 mM NaCl. .......... 133
Figure 56 The effect 20-bp spacer DNA molecules on the DNA orientation and IHF-induced fluorophore height changes in buffer containing 50 mM NaCl .......... 133
Figure 57 Schematic illustration of the effect of additional protein layer on the measurement of fluorophore height change............................................. 139
Figure 58 The estimation error of fluorophore height change resulted from IHF-induced DNA bending increases with accumulated IHF layer thickness...................... 140
Figure 59 The effect of IHF layer on the measured fluorophore height change depends on IHF locations on the DNA molecules.................................................. 142
Figure 60 IHF binds nonspecifically to arbitrary Control sequences ....................... 146
Figure 61 Real-time measurement of IHF binding to consensus H' binding site and Control sequence................................................................. 147
Figure 62 Dissociation kinetics of IHF from DNA spots after binding at equilibrium .. 149
Figure 63 The geometric model for quantification of DNA bending angle .............. 153
Figure 64 Estimation of the DNA bending angle induced by specific IHF binding...... 154
Figure 65 Crystal structure of the IHF–H' DNA complex..................................... 155
Figure 66 Detect single nucleotide mutation of IHF binding sequence..................... 157
Figure 67 SSFM distinguishes IHF binding sequence containing a single nucleotide mutation from IHF consensus binding sequence.............................. 157
Figure 68 Differentiation and comparison of IHF binding sequence and IHF binding sequence containing a single nucleotide mutation.............................. 158
Figure 69 An example of the spectral resolution of acousto-optic tunable filter as function of the square of the wavelength ........................................................................................................................................................................ 166

Figure 70 Schematic illustration of SSFM and LED-RS combined with a spectral tunable filter to improve acquisition throughput. ........................................................................................................................................................................ 168

Figure 71 SSBP binding during DNA replication to avoid degradation and base-pairing of the exposed ssDNA. ........................................................................................................................................................................................................ 172

Figure 72 A proposed workflow for effective and efficient study of conformational specific protein-DNA interactions and their functions. .................................................................................................................. 173
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Adenine</td>
</tr>
<tr>
<td>AFM</td>
<td>Atomic force microscopy</td>
</tr>
<tr>
<td>AIC</td>
<td>Akaike information criterion</td>
</tr>
<tr>
<td>AOTF</td>
<td>Acousto-optic tunable filter</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>BIC</td>
<td>Bayesian information criterion</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>BOE</td>
<td>Buffered oxide etch</td>
</tr>
<tr>
<td>C</td>
<td>Cystine</td>
</tr>
<tr>
<td>CAP</td>
<td>Catabolite gene activator protein</td>
</tr>
<tr>
<td>CCD</td>
<td>Charged coupled device</td>
</tr>
<tr>
<td>DI</td>
<td>Deionized</td>
</tr>
<tr>
<td>DIC</td>
<td>Differential interference contrast</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>dsDNA</td>
<td>Double-stranded DNA</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme-linked immunosorbent assay</td>
</tr>
<tr>
<td>FRET</td>
<td>Förster resonance energy transfer</td>
</tr>
<tr>
<td>G</td>
<td>Guanine</td>
</tr>
<tr>
<td>HPLC</td>
<td>High-performance liquid chromatography</td>
</tr>
<tr>
<td>IgG</td>
<td>Immunoglobulin G</td>
</tr>
<tr>
<td>ITO</td>
<td>Indium Tin Oxide</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>IHF</td>
<td>Integration host factor</td>
</tr>
<tr>
<td>IRIS</td>
<td>Interferometric Reflectance Imaging Sensor</td>
</tr>
<tr>
<td>LED</td>
<td>Light emitting diode</td>
</tr>
<tr>
<td>LED-RS</td>
<td>WLRS using LED as illumination source</td>
</tr>
<tr>
<td>LSF</td>
<td>Least square fitting</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular dynamics</td>
</tr>
<tr>
<td>NaCl</td>
<td>Sodium chloride</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein data bank</td>
</tr>
<tr>
<td>QDs</td>
<td>Quantum dots</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulfate</td>
</tr>
<tr>
<td>Si</td>
<td>Silicon</td>
</tr>
<tr>
<td>SiO₂</td>
<td>Silicon dioxide</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-noise ratio</td>
</tr>
<tr>
<td>scIHF</td>
<td>Single-chain integration host factor</td>
</tr>
<tr>
<td>ssDNA</td>
<td>Single-stranded DNA</td>
</tr>
<tr>
<td>SSBP</td>
<td>Single-strand binding protein</td>
</tr>
<tr>
<td>SSFM</td>
<td>Spectral self-interference fluorescence microscopy</td>
</tr>
<tr>
<td>SPR</td>
<td>Surface plasmon resonance</td>
</tr>
<tr>
<td>T</td>
<td>Thymine</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>TBP</td>
<td>TATA-binding protein</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>TIRF</td>
<td>Total internal reflectance fluorescence</td>
</tr>
<tr>
<td>TPM</td>
<td>Tethered particle motion</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris (hydroxymethyl) aminomethane</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>WLRS</td>
<td>White light reflectance spectroscopy</td>
</tr>
<tr>
<td>wt</td>
<td>wild-type</td>
</tr>
<tr>
<td>wtIHF</td>
<td>Wild-type integration host factor</td>
</tr>
<tr>
<td>1-D</td>
<td>One-dimensional</td>
</tr>
<tr>
<td>2-D</td>
<td>Two-dimensional</td>
</tr>
<tr>
<td>3-D</td>
<td>Three-dimensional</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION AND MOTIVATION

1.1 The importance of studying protein-DNA interactions

Our complete genetic information is stored in the human genome and encoded as the sequence of over three billion DNA base pairs. The complete sequencing of the human genome was one of the most notable scientific and collaborative achievements in the last decade. The analysis of DNA sequences has helped us understand cancer and inherited diseases, and led to the development of new therapies\textsuperscript{1-5}. However, decoding the DNA sequences and deciphering the genetic code by which DNA sequences are translated into proteins did not fully reveal the functions of the genome and the molecular mechanisms of diseases, which necessitate another principle macromolecule in the living cell: the protein.

The maintenance of intact genetic information and the functions of the genome critically rely on protein-DNA interactions, which trigger essential molecular processes in the cell. Examples of such processes include DNA replication, DNA repair, gene expression and its regulation, genome rearrangement by DNA recombination and transposition, DNA restriction by endonucleases, and DNA modification by methyltransferases\textsuperscript{6-11}. How proteins specifically recognize their binding sites amongst billions of DNA base pairs and subsequently carry out their unique functions remains to be fully explained. Now that next generation sequencing technologies have made many complete genome sequences available at remarkably reduced cost, more endeavors are
needed to study the mechanisms of protein-DNA interactions to completely understand the human genome and its functions.

1.1.1 Overview of protein-DNA interactions

After deciphering the genetic code, we tend to think of DNA as a linear string of the four letters G, A, T and C, representing the four nucleobases (guanine, adenine, thymine, and cytosine) of the four DNA nucleotides that make up the DNA polymer. The genetic code consists of sets of three DNA nucleotides that specify which one of the 20 amino acids to be added during protein synthesis. Naturally one might wonder if a protein could recognize its specific binding site in the genome by “reading” these letters under a rule similar to the genetic code. But scientists realized many years ago that such a simple code that maps one DNA sequence to one protein sequence does not exist. Depending on the individual interaction, proteins use multiple readout mechanisms at various extents to recognize the specific DNA sequences.

Before we review some of the proposed protein-DNA recognition mechanisms, let’s recall how fascinating and puzzling it is for a protein to recognize a specific DNA sequence in a living cell from biological and biophysical perspectives. First, how do the proteins gain access to the DNA sequences that are confined in the nucleus and formed into chromatin structures? Furthermore, how do the proteins find specific binding sequences and perform their functions in the crowded nucleus containing other proteins and chromosomal DNA? Moreover, how do some low concentration proteins, such as some transcription factors, achieve this task? And how do some regulatory proteins
recognize their binding site, activate or repress gene expressions according to the cellular environment in a timely manner? Finally, how do the proteins possessing similar DNA-binding domains recognize different specific DNA sequences and carry out respective functions?

For over 30 years, significant studies have been carried out to characterize protein-DNA interactions to unravel how proteins recognize their binding site and perform their activities. The first X-ray crystal structures of protein-DNA complexes were the catabolite gene activator protein (CAP), the Cro repressor, and the λ repressor bound to their binding sites. As of the time of this dissertation, more than 4400 X-ray crystal structures of protein-DNA complexes have been deposited in the Protein Data Bank (PDB). Although X-ray crystallography provides atomic scale resolution of the structure of protein-DNA complexes, it is difficult to crystallize these complexes. Nuclear magnetic resonance (NMR) spectroscopy can help infer structures of protein-DNA complexes, but is limited to small proteins that are soluble at high concentrations. Therefore, to study the biophysical properties of protein-DNA interactions a variety of new techniques have been developed and employed. Furthermore, with more genomic sequences and protein-DNA complex structures available from databases, computational methods have been developed to provide predictable and mechanistic insights based on the thermodynamics of protein-DNA interactions. In the last decade, high-throughput methods have revolutionized our ability to obtain protein-DNA binding data and generated comprehensive information of binding affinities.
Sixty years ago, Watson and Crick taught us that DNA adopts the structure of an anti-parallel double helix held together by hydrogen bonds under a specific base-pairing rule between the four nucleobases of the four nucleotides. The unique chemical and structural signatures of the nucleobases along the DNA sequence provide chemical and structural variations for specific recognition of proteins, which also have their own chemical features and three-dimensional (3-D) conformations\textsuperscript{10,28}. In recent years, it is suggested that proteins achieve DNA binding specificity by combining direct readout or “base readout” in the major groove of the DNA double helix and indirect readout or “shape readout” in the minor groove of the DNA double helix at different levels\textsuperscript{14,15,28-31}. In the base readout mechanism, proteins recognize the DNA sequence by forming hydrogen bonds between their amino acid side chains and the nucleobases in the DNA helix major groove. The hydrogen bond donors and acceptors of the amino acid side chains are specific to the unique set of donors and acceptors presented by the nucleobase sequence. In the shape readout mechanism, proteins recognize their binding sites by specifically accommodating, stabilizing or enhancing local DNA conformation variations (e.g. narrowed minor groove widths and DNA kinks) or global DNA conformation variations (e.g. DNA bending, A DNA, Z DNA). The varied DNA conformation also has a biophysical consequence besides structural. The electrostatic potential of the negatively charged DNA phosphate backbone is either enhanced or reduced as the DNA conformation varies, breaking its uniformity along the DNA sequence. Enhanced negative potentials of the DNA helix will attract positively charged amino acid side
chains of the proteins, suggesting a mechanism for protein recognition of sequence-specific DNA shapes.

Over 2000 DNA-binding proteins are executing critical functions and instructions of the genome in the cell. Protein-DNA interactions depend on the recognition and accommodation of two characteristic 3-D structures to achieve binding specificity and functional activities. We continue to examine the relationship between DNA shape or conformation and specific protein binding and the resulting biological functions in the next subsection.

1.1.2 DNA conformation and functional protein binding

DNA molecules are intrinsically flexible. The conformation of DNA in solution is typically modeled as an elastic rod characterized by the persistence length. The deformability of DNA conformation is inversely related to the persistence length, which is influenced by the DNA sequence and the base pair stacking energy. However, in many protein-DNA complexes, DNA undergoes substantial bending or conformational change on scales shorter than the persistence length, which impacts binding affinity and enables specific recognition in the shape readout mechanism.

We shall not focus on the physical mechanisms of how DNA adopts a variety of protein-binding induced conformational changes, but we will give some examples of its biological consequences in the cell. In fact, substantially deformed DNA plays critical roles in many biological functions. First, we take the packaging of genomic DNA for an example. During viral assembly, the double-stranded DNA viruses pack their entire
genomic DNA that is longer than 10 μm into a protein shell, the capsid, with typical dimensions of 30 to 100 nm\textsuperscript{30}. This means the radius of the capsid is smaller than the persistence length and DNA is tightly bent in the virus capsid. In the bacteriophages, it may be bent with a radius of curvature of roughly 3 nm, very unlikely for free DNA. Similarly, in human cells, about 2 meters long chromosomal DNA ((0.34 × 10^{-9} m/bp) × (6 × 10^9 bp)) is packaged into the nucleus with a diameter of 10 to 20 μm at three hierarchical levels. At the lowest level each 147 bp DNA segments is wrapped around a protein complex (the histone octamer) for roughly 1.7 turns to form the repeating structure unit of chromatin, the nucleosome\textsuperscript{33} (Figure 1). The DNA in the nucleosome is tightly bent with the surface of some parts of the DNA sequences buried. The formation of the nucleosome structure is a result of electrostatic interactions between the positively charged side chains of the histone proteins and the DNA backbone and causes our genome to be packaged in a much smaller volume than the DNA alone. It is worth mentioning that although most genomic DNA is wrapped in nucleosomes, nucleosomes are not randomly distributed along genomic DNA. DNA sequences that are more deformable tend to form more stable nucleosomes, which may strongly affect accessibility of DNA-binding proteins and functionality of certain DNA sequences\textsuperscript{34}.
DNA is also bent or looped by DNA-binding proteins that regulate gene expression levels. For example, TATA-binding protein (TBP) plays a major role in transcription initiation by binding to the TATA-box sequence in many prokaryotic and eukaryotic promoters. TBP specifically binds the TATA-box sequence by recognizing the higher flexibility of the TATA-box sequence, which is further bent to about 80° in the stable TBP-DNA complex. In another example of transcription regulation, lac operon controls the production of the lactose digestion enzyme. In the lac regulation, the lac repressor protein binds simultaneously to two operators on the DNA sequence, bringing the two distant operators adjacent to each other and causing a looped DNA segment on the scale of the persistence length. These two examples imply that the gene expression
levels may be controlled by changes of the protein-binding sites that induce DNA looping or bending\textsuperscript{30}. 

![Figure 2 X-ray crystal structure of TATA-binding protein (TBP) bound to its consensus sequence, the TATA-box DNA.]

In humans, tens of thousands of DNA damages occur naturally every day in every normal cell\textsuperscript{36}. The conformation of damaged DNA is destabilized and becomes more flexible at the lesion. DNA repair proteins recognize the increased local flexibility and further induce distortion of DNA lesions\textsuperscript{37-39}. For example, in DNA gap repair, human polymerase λ carries out its role of filling short gaps longer than 1 nucleotide by specifically scrunching the DNA template\textsuperscript{40}. The scrunching places the un-copied template nucleotide to an extrahelical position within a binding pocket that comprises three conserved amino acids of polymerase λ. In DNA mismatch repair, protein MutS distinguishes mismatched and paired bases by recognizing the weakened base stacking
and increased flexibility of the mismatched DNA. MutS further induces a 60° kink at the mismatched location in the MutS-DNA-ATP complex, which then recruit other downstream repair proteins. This conformational driven mechanism allows MutS to bind a mismatch up to 1000 times better than perfectly paired DNA.

Figure 3 Crystal structure of Tag MutS bound to a mismatched DNA at 2.2 Å resolution. DNA is kinked by 60° at the mismatch (Source: Protein Data Bank, ID: 1EWQ).

In the above examples, we aim to show that protein-induced DNA conformational change is a common feature in living organisms and has important biological functions. It is not possible to exhaustively describe all the discovered protein-DNA complexes whose specific recognition and functions rely on protein-induced DNA conformational changes. Also, given that there are more than 3 billion base pair long genomic DNA and over 2000 DNA-binding proteins, a large number of protein-DNA interactions remain unknown. More systematic and quantitative studies and techniques are needed to study the biophysical mechanisms and to unravel the biological consequences of protein-DNA
interactions. Before we review some of the recent technological developments in studying protein-DNA interactions, we introduce another architectural protein selected as the molecular model to demonstrate the capabilities of the interferometric sensing platform developed in this dissertation: the integration host factor protein.

1.1.3 A good model: Integration Host Factor (IHF)

The integration host factor protein from *E. coli* is an architectural protein that relies on indirect readout or uses DNA conformation deformability to achieve its binding specificity\(^42\). IHF is a ~22 kDa heterodimer protein that can induce a sharp bending of a 34-bp long specific DNA sequence. The overall bending angle of the DNA induced by IHF was estimated to be over 140° by electrophoretic mobility shift assays or over 160° by X-ray crystal structure analyses\(^43,44\). The crystal structure of the specific complex of IHF and the 34-bp DNA sequence of the H’ site from bacteriophage \(\lambda\), one of the best characterized IHF-binding site, is available in the Protein Data Bank (PDB ID: 1IHF).

The crystal structural reveals that the H’ binding site is sharply bent and wrapped around IHF, reversing the direction of the helix axis within a very short distance (Figure 4). In the 34-bp IHF binding sequence, two segments indicated by green bases in Figure 4 are highly conserved and directly contact the IHF side chains via salt bridges: the 5’-TATCAA-3’ site at the center and a 5’-TTG-3’ site on “right” the side. A 6-bp A-tract sequence is preferred on the “left” side, where its narrowed minor groove fits into the complex but only interacts with IHF via water-mediated contact. From the crystal structure, it was found that IHF causes DNA sharp bending by introducing two very large
kinks. Two proline side chains of each IHF subunit intercalated between the base pairs stabilize the two kinks. The IHF-DNA complex is further stabilized by electrostatic interactions between the positively charged IHF surface and the wrapped negatively charged DNA backbone.

Figure 4 X-ray crystal structure of the IHF–H’ DNA complex with the α subunit of IHF in grey, the β subunit of IHF in pink, the consensus DNA bases in green and the less conserved bases in blue (Source: Thomas W Lynch, Erik K Read, Aras N Mattis, Jeffrey F Gardner, Phoebe A Rice. Integration Host Factor: Putting a Twist on Protein–DNA Recognition, Journal of Molecular Biology 2003, 330(3): 493 – 502).

By inducing the specific sharp bend of its specific binding site, IHF functions as an architectural component to facilitate various prokaryote cellular processes that require the assembly of higher-order protein-DNA complexes. For example, in λ phage site-specific recombination, IHF binds to the att sites to appose two promoters for interaction with the heterobivalent DNA-binding λ integrase. Although the name of IHF came from
its role in λ phage site-specific recombination, it was later found essential in other cellular functions. In bacterial transcription initiation, IHF binding and bending of a DNA site creates optimal promoter geometry to facilitate contacts between distant proteins and assist the recruitment of RNA polymerase to some promoters\textsuperscript{44,45}. In addition, IHF is also known to facilitate DNA replication initiation, transposition, and phage packaging\textsuperscript{46}. IHF also binds to DNA nonspecifically. In fact, the level of IHF in the cell was found to increase from 12 µM at the exponential growth phase to 55µM at the stationary phase. Thus, as an abundant nucleoid-associated protein, IHF also binds nonspecifically as one of the major histone-like proteins to compact the 4.7 million bp circular DNA into the bacterial chromatin\textsuperscript{47,48}.

The unique sharp “U-turn” of the DNA helix induced by IHF binding attracted many studies on its specific binding mechanism or as a model system for the study of the nucleosome and other large protein-DNA complexes. A variety of techniques have been used, such as time-resolved Förster resonance energy transfer (FRET), hydroxyl radical DNA footprinting assays, stopped-flow and laser temperature jump techniques combined with FRET, and isothermal titration calorimetry\textsuperscript{42}. Additionally, in recent years IHF-DNA complex has become a model system for studying the indirect readout mechanism for specific recognition of DNA binding sites and the biophysics of DNA bending. For instance, the investigation of the transition-state ensemble of the IHF-DNA complex suggested that DNA conformation randomly fluctuates in a sequence dependent manner until the nonspecifically bound protein recognizes and captures the temporarily bent or kinked DNA and continues to stabilize the bound complex\textsuperscript{42}. In this dissertation, we use
the IHF-DNA complex as our molecular model system to demonstrate parallel quantification of DNA conformational change and protein-DNA binding.

1.2 Biophysical techniques for studying protein-DNA interactions

Protein-DNA interactions play critical roles in almost all cellular functions in every living organism. The specific formation and functions of many protein-DNA complexes rely on protein-induced DNA conformational changes. For many years, a variety of traditional methods have facilitated numerous breakthroughs in the study of structures of protein-DNA complexes, such as X-ray crystallography, NMR spectroscopy, and electrophoretic mobility shift assay (EMSA). However, deeper understanding of the underlying biophysical mechanisms requires the application of new experimental and computational approaches to precisely and systematically quantify binding affinity as well as conformational specificity. Here we focus on a few examples of the recently evolved techniques to elucidate protein-DNA interactions that cause DNA conformational changes.

1.2.1 Biophysical techniques

Förster resonance energy transfer (FRET) techniques. FRET describes a distance-dependent interaction in which an excited donor fluorophore transfers energy to a ground-state acceptor fluorophore through non-radiative dipole-dipole coupling. The energy transfer efficiency between the donor and acceptor is inversely proportional to the sixth power of the distance between them. The intermolecular or intramolecular distance between two fluorophores specifically labeled on DNA or protein molecules can be
calculated from FRET efficiency determined from fluorescence spectra measurements\textsuperscript{51}. The distance of the two fluorophores determined by FRET ranges from 1 to 10 nm. The DNA conformational change in a protein-DNA complex can be deduced by measuring the change of distance between the donor and acceptor labeled at known locations on the DNA. FRET also allows for dynamic detection of protein-induced DNA conformational changes at the single-molecule level when combined with highly sensitive fluorescence microscopy techniques, such as confocal microscopy or total internal reflection microscopy. In recent years, Quantum Dots (QDs) have been adapted as ideal FRET donors due to their high brightness, long-lasting, size-tunable and narrow photoluminescence comparing to conventional organic dyes\textsuperscript{50,52,53}. The discovery of molecular beacons and aptamers extended applications of FRET in more protein-DNA binding assays by allowing for rational predictions of the DNA conformational changes\textsuperscript{22}. One of the experimental challenges of FRET is to precisely know the F"{o}rster distance, a critical length scale ruler to determine FRET efficiency. The varied quantum yields, relative orientations, and sizes of the donor and acceptors cause uncertainties to the most important F"{o}rster distance, which needs to be carefully determined to perform accurate quantitative distance measurements on the nanometer scale.

*Atomic force microscopy (AFM).* AFM allows for direct visualization of a single DNA molecule on atomically flat biocompatible support surfaces, such as mica\textsuperscript{54-56}. Signal of AFM arises from the innate molecular forces between the tip of the scanning probe and the biomolecule sample on the flat surface. AFM can be used to perform measurements in air and aqueous buffers, permitting the detection of DNA.
conformational changes induced by protein binding. Depositing DNA molecules on the mica surface is relatively simple, but strong interactions between the DNA molecules and the substrate are required to avoid detachment or damage of DNA molecules from the motion of the scanning tip. Although AFM allows for both static and dynamic studies of protein-DNA interactions, the scanning rate is relatively slow, limiting the ability to observe single molecule binding events. Moreover, to obtain statistically meaningful results, a large population of molecular interactions must be observed, but scanning over a large area or a number of small areas can be time-consuming. Improving the scanning speed without compromising resolution is a challenge yet to be resolved.

*Tethered Particle Motion (TPM).* TPM is another biophysical method used for studying interactions between DNA and DNA-binding proteins that cause DNA bending, shortening or looping\textsuperscript{24,57}. One end of a single DNA molecule is attached to a bead of a diameter of tens to hundreds of nanometer, while the other end is tethered to a surface. The bead-tagged DNA molecules are submerged in an aqueous environment, and the diffusion of the bead under Brownian motion is restricted to a hemisphere defined by the DNA length, which is typically from hundreds to thousands base pairs. An optical microscope, such as differential interference contrast (DIC) microscopy, total internal reflectance fluorescence (TIRF) microscopy, or dark-field or bright-field microscopy, can be used to track the range of allowed mobility of the bead in real-time. Thus DNA length changes resulted from DNA conformational changes caused by protein binding can be monitored in real time. The kinetics of DNA looping has been extensively characterized by TPM. The characteristic time the bead takes to explore the hemisphere space allowed
by the DNA tether through diffusion gives a good estimate of its movement amplitude changes associated with different DNA conformations. The interplay of the effects of DNA contour length, DNA persistence length, and the bead size must be taken into account to calibrate the dynamic measurements by TPM\textsuperscript{24,57}.

\textit{Optical tweezers.} Optical tweezers can probe DNA conformation variations and mechanical properties by precisely manipulating nanometer or micrometer-sized dielectric particles tagged to one or two ends of DNA molecules\textsuperscript{25}. A highly focused laser beam by a microscope objective traps or displaces the dielectric particle by exerting forces on the pico newton scale. An accurate calibration between a known force exertion and the displacement of the particle is key for quantitatively measure the external forces on the DNA from the particle displacements. Optical tweezers enable a variety of assays of protein-DNA interactions\textsuperscript{25}. For example, protein-bound DNA will have different flexibility or force-extension properties comparing to bare DNA. Also, DNA conformational changes can be measured in real time by monitoring the changes of DNA-to-tether length in a trapping assay, in which the magnitude and mechanism of conformational change as well as binding strength can be assessed. The many technical precautions such as mechanical noise minimization and precise calibration limit the wide use of optical tweezers. In addition, only one DNA molecule typically of a few kilo base pairs or longer is investigated at a time, limiting the efficiency, throughput, and resolution of optical trapping measurements. In recent years, optical tweezers are combined with fluorescence microscopy, such as FRET and TIRF, which enables characterization of both the global mechanical properties of a protein-DNA complex and direct observation
of the presence and dynamics of the labeled protein. A disadvantage of integrating fluorescence microscopy to optical tweezers is increased photobleaching rates of the fluorophores illuminated simultaneously by laser beams for fluorescence excitation and particle trapping.

Lately, numerous mechanical, optical, and electrical methods have witnessed striking progress in studying the biophysical mechanisms of protein-DNA interactions. These methods include but not limited to electrochemical DNA sensors, metamaterial-based spectroscopy, nanopore arrays, and parallel molecular force-based assays\textsuperscript{58-62}. A comprehensive review of these techniques is out of the scope of this dissertation.

### 1.2.2 High-throughput methods

Most of the biophysical techniques resolve protein-binding induced DNA conformational changes in solution at the single-molecule level. To obtain statistically meaningful results of a protein-DNA complex or to examine the interactions between various proteins and DNA sequences, these single-molecule biophysical techniques are laborious and time-consuming. High-throughput methods have been proven to provide large-scale and comprehensive protein-DNA binding affinity information important for the analysis of genetic regulatory networks and the effects of genetic variations on those networks\textsuperscript{2,4,9}. For example, protein-binding microarrays (PBM) have enabled the analysis of the binding specificities of many transcription factors (TFs) to high-density, custom-designed microarrays of over 44,000 DNA spots containing all 10-mer sequence variants\textsuperscript{7,11,63}. In PBM experiments, a commercially synthesized single-stranded DNA
oligonucleotide microarray is first converted to double-stranded DNA microarray by solid-phase primer extension. One epitope-tagged TF, either synthesized or purified, is bound directly to the microarray, which is then washed to remove nonspecific binding. A fluorophore-tagged antibody is then added to the protein-bound DNA microarray to determine the amount of protein binding to each DNA sequence. Multiple TFs can also be analyzed in parallel in separate chambers on a single microarray. One technical difficulty for PBM is determining the binding specificity of TFs to longer DNA motifs, which requires many fold increase of the microarray size to include all DNA sequence permutations of the binding motif length. Another approach to select protein-binding DNA sequences is the systematic evolution of ligands by exponential enrichment (SELEX)\textsuperscript{9,10,13-17,64}. First, a library of DNA sequences containing potential protein binding sites is created from randomly synthesized DNA or generated from genomic sequences. Primer binding sites are added to both ends of the DNA sequences so that they can be amplified by polymerase chain reaction (PCR). Then, a synthesized or purified TF is added to the DNA sequence library and the protein-bound and unbound DNA sequences are separated. The protein-bound DNA sequences are eluted and amplified by PCR and bind to target TF again. This selection process is performed multiple times to extract the DNA sequences with high-affinity protein-binding sites, which are finally amplified and sequenced. One advantage of SELEX for high-throughput screening of protein-DNA interactions is that there is no inherent limit of the length of the protein-binding site. The specificity of protein and its binding sites obtained from enormous binding data of SELEX needs to be extracted with computational methods especially
after multiple rounds of selection. Last year, automated high-throughput enzyme-linked immunosorbent assay (ELISA) was reported to effectively screen protein-DNA interactions\textsuperscript{18,19,65}. Biotinylated DNA sequences of an optimized library were separately immobilized on streptavidin-coated wells of a microtiter plate. HIS-epitope-tagged proteins were then bound to the DNA probes in each well. The amount of binding was determined by immunological detection and photometric readout. This automated ELISA was used for the discovery of DNA-binding sites of the uncharacterized DNA-binding protein AtTIFY1, which was also confirmed by other in vitro methods.

The primary results of these high-throughput methods are binding affinities of proteins to their DNA binding sites. However, in the cellular environment the affinity of a protein to its binding site is not as crucial as its specificity. Inside the crowded environment of a cell nucleus, DNA is highly compact and the concentration of DNA is very high. It is the binding specificity that enables a protein to distinguish its binding site from the vast sea of genomic DNA sequences. As we discussed previously, the intrinsic flexibility of DNA conformation and its deformation in protein-DNA complexes contribute to protein-DNA recognition specificity. Although these high-throughput methods permits the analysis of the protein binding site preferences when combined with computational methods, they do not offer conformational specificity information needed to elucidate the biophysical mechanisms of specific protein-DNA interactions.
1.2.3 Computational methods

The specific protein–DNA interaction is based on direct atomic contacts between the interfaces of the 3-D protein and DNA molecules and the adaptation of their conformations\textsuperscript{10,26}. Binding and conformational specificities of protein-DNA complexes can be computed by simulating the atomic forces and energy potentials from physical or statistical models\textsuperscript{20-25,66}. The advent of high-capacity data-driven computational technologies and hardware, and protein-DNA interaction databases\textsuperscript{17,26,27,67,68} have facilitated the use of computational analysis for high-throughput screening and systematic study of protein-DNA interactions. Furthermore, computational methods complement experimental studies by helping interpret experimental results, hypothesize binding mechanisms, and design functional protein-DNA complexes\textsuperscript{26,27,69}.

Examples of using computational methods to study specific protein-DNA interactions include the use of a dynamic algorithm that structurally aligns interfaces of protein–DNA complexes to distinguish the binding specificities of proteins within a family\textsuperscript{17,70}. Furthermore, molecular dynamics (MD) simulations and free energy calculations have been employed to obtain the binding specificities of target proteins to different DNA sequences and to reveal the key forces for the formation of protein-DNA complexes\textsuperscript{10,26-28}. Moreover, the proposed indirect or DNA shape readout mechanism was based on comprehensively analyzing DNA minor groove geometry and calculating electrostatic potentials of DNA surfaces of over 1000 crystal structures of protein-DNA complexes\textsuperscript{14,15,28-31}. Another MD simulations study showed that DNA conformational switching results in overcoming the energy barrier for specific protein-DNA
Recently, a statistical method based on MD simulations was used to correctly predict protein-binding site that were suggested to be involved in the indirect readout mechanism\textsuperscript{30,72}. The method also proposed new protein-binding sites where the indirect readout mechanism can make major contributions to the specific recognition.

The examples listed above by no means include all the studies using computational methods to understand the specific recognition mechanisms of protein-DNA complexes. However, these examples demonstrate that computational approaches are becoming more useful and versatile for discriminating DNA-binding proteins, predicting protein binding sites and protein–DNA binding specificity. The development of more accurate computational methods depends on obtaining more experimental results to increase our understanding of protein-DNA complexes. Therefore, computational methods can be paired with high-throughput methods where predictions from computational models guide experimental designs and experimental binding results provide feedback for improving the computational algorithms.

1.3 Motivation: high-throughput biophysical study of protein-DNA interactions

Protein-DNA interactions maintain and carry out functions of the genome by playing essential roles in various cellular processes. Examples of such processes include DNA replication, DNA repair, gene expression and its regulation, and DNA recombination and transposition. Besides directly forming hydrogen bonds with the nucleobases in the DNA major groove, DNA-binding proteins recognize and induce DNA conformational variations to achieve specific binding and activities. Biophysical
measurements that directly measure the DNA conformational changes in protein-DNA complexes have proven to be informative to elucidate the specific protein-DNA interaction mechanisms. High-throughput methods that provide large-scale protein-DNA binding information have shown effective and comprehensive analysis of protein-DNA binding affinities. The purpose of this dissertation is to provide a tool to study the biophysical mechanisms of specific protein-DNA interactions in a quantitative and high-throughput manner. Our approach is to combine measurement of DNA conformational change in the nanometer scale and quantitative label-free detection of protein-DNA binding in a DNA microarray format. The DNA conformational change is measured using spectral self-interference fluorescence microscopy (SSFM) that determines conformation of surface-immobilized DNA molecules by measuring axial height of fluorophores tagged to specific nucleotides. The amount of protein binding to each DNA spot is detected by a modified white light reflectance spectroscopy that enables quantification of the mass density of surface-bound biomolecules. The ultimate goal of the proposed work is to provide a rapid and convenient quantitative method that enables high-throughput parallel screening of conformational specificity and binding affinity of protein-DNA complexes, and by combining with computational methods, to facilitate mechanistic and systematic interpretations of protein-DNA interactions and their regulatory functions in the living cell.
1.4 Overview of this dissertation

In this dissertation, we present the development of a dual-spectral interferometric sensing platform that provides quantified information of DNA conformation and protein binding for the study of protein-DNA interactions in the fashion of DNA microarrays. In addition, we present the application of the developed optical sensing platform for characterizing factors affecting protein-DNA interactions on a surface. We further demonstrate parallel quantification of conformational specific and nonspecific protein-DNA interactions and its use in understanding binding mechanisms and regulatory functions of protein-DNA complexes.

Chapter 2 presents the physical models and data analysis of the interferometric techniques this dissertation built upon, the spectral self-interference fluorescence microscopy (SSFM) and white light reflectance spectroscopy (WLRS). Applications of SSFM and WLRS for the estimation of surface-immobilized DNA conformation are briefly reviewed. In particular, the combination and development of dual-color SSFM and LED-based WLRS are discussed for the characterization of the DNA conformation on a 3-D polymer functionalized surface.

Chapter 3 presents the one-dimensional dual-spectral imaging configuration of the combined SSFM and LED-based WLRS for simultaneous measurement of DNA conformational change and protein binding on multiple DNA spots. A quantitative model is proposed to resolve specific binding and nonspecific binding of IHF to surface-immobilized DNA molecules.
Chapter 4 discusses the application of the integrated system for the assessment of the factors that influence the detection and quantification of protein-DNA interactions on a surface using the IHF-DNA binding model.

Chapter 5 demonstrates the use of parallel independent quantification of specific and nonspecific IHF binding to surface-immobilized DNA to examine salt-dependence of protein-DNA interactions, quantify specific DNA bending angles, and discriminate a single nucleotide mutation of the IHF consensus binding sequence.

Chapter 6 concludes the dissertation and discusses the potentials and future development of the platform, and suggests future applications in the study of protein-DNA interactions.
CHAPTER 2

SPECTRAL SELF-INTERFERENCE FLUORESCENCE MICROSCOPY (SSFM)

For more than a century it has been known that light of fluorescence emission can interfere with its reflected light by a reflecting surface\textsuperscript{1,3,5}. Spectral self-interference fluorescence microscopy (SSFM) uses the self-interference of fluorescence emission above a reflecting surface to localize monolayer of fluorophores in the axial dimension with sub nanometer precision\textsuperscript{6}. By determining the axial location of fluorophores labeled on DNA molecules, SSFM enables nanoscale characterization of the conformation of DNA immobilized on a surface. In this chapter, an overview of the principles and development of SSFM will be given, followed by its applications in characterizing conformations of surface-immobilized DNA. Then, a recent development of a dual-color SSFM that allows for axial co-localization of two fluorophores attached at different nucleotides of surface-immobilized DNA molecules and its application will be presented. The chapter will end with a discussion of the advantages and limitations of traditional SSFM and dual-color SSFM and its future developments.

2.1 Fundamentals of spectral self-interference fluorescence microscopy

2.1.1 Background and development

Fluorescence microscopy is one of the most powerful tools in modern biological research. The potential of fluorescence as a contrast agent for microscopy was realized by August Köhler in 1904 while he was working on improving the resolution of microscopes. He discovered that biological materials fluoresce when illuminated with
ultraviolet (UV) light\textsuperscript{12}. The development of a variety of fluorophores has made it possible to identify labeled sub-cellular components non-invasively and with a high degree of specificity. Like other conventional light microscopy, spatial resolution of fluorescence microscopy is subject to the diffraction limit. The minimum lateral resolution is proportional to wavelength $\lambda$ and inversely proportional to the numerical aperture (NA, $\text{NA} = n \sin (\theta)$), where $n$ is the refractive index in the object space, and $\theta$ is the half-angle of the largest cone of rays that can enter or leave the optical system. The spatial resolution of conventional high resolution optical microscopes is not better than about 200 nm laterally and about 600 nm longitudinally. Several methods have achieved spatial resolution beyond the diffraction limit utilizing the specific nature of fluorescence\textsuperscript{73}, such as increasing the effective numerical aperture such as in 4Pi confocal microscopy\textsuperscript{74,75}, introducing spatial variation in the excitation light to create finer spatial features in the image such as in standing wave microscopy\textsuperscript{76}, using multiple-photon fluorescence absorption or emission mechanisms that cause non-linear effects in the light field, such as in two-photon microscopy\textsuperscript{77}, and selectively quenching the fluorescence from a focal spot to obtain a very small fluorescing volume, such as in stimulated emission depletion microscopy (STED microscopy)\textsuperscript{78,79}. Localization of fluorescent molecules with high accuracy is also of great interest and provides valuable spatial information not accessible even by high-resolution imaging. Two techniques, stochastic optical reconstruction microscopy (STORM)\textsuperscript{80} and photoactivated localization microscopy (PALM)\textsuperscript{81} enabled lateral localization of sparse fluorescence emitters by estimating the center of the point spread function of the emitters.
Fluorescence emitters can also be localized in the axial dimension with nanometer-scale precision by utilizing self-interference of light. In the early sixties, Drexhage determined the fluorescence decay of an organic dye embedded into a lipid layer as a function of its distance above a silver mirror\(^8\). Later, Fromherz and co-workers developed a fluorescent axial localization technique, fluorescence interference contrast microscopy (FLIC microscopy)\(^8\). FLIC microscopy uses the change in the total emission intensity as a function of distance from a nearby reflecting interface and has been used for measuring cellular membranes\(^8\) and molecular motors\(^8\). In FLIC microscopy, the fluorescently labeled object is within the axial distance of a wavelength \(\lambda\) from the reflecting interface, so there is little difference between the standing waves of different wavelengths within the fluorescence emission spectrum. As a result, the entire fluorescence emission spectrum oscillates as the direct and reflected fluorescence emission undergoes constructive or destructive interference depending on the vertical distance. Careful calibration of fluorescence intensity as a function of monolayer fluorophore distance from the reflecting interface is required to achieve nanometer accuracy\(^8\).

When the axial distance between a fluorophore and the reflecting interface is larger than 10 \(\lambda\), even at the same height, the interference between the fluorescent emission of the direct optical path and that of the reflected path results in several oscillations or fringes of interference within the fluorophore emission spectrum. The oscillation in the emission spectrum is a unique signature of the optical distance of the fluorophore to the interface. SSFM utilizes this interference-based modulation of the
fluorescence emission spectrum to determine the precise axial height of fluorophores to the reflecting interface with sub-nanometer accuracy. Traditionally, SSFM uses a SiO$_2$-Si substrate with a $\sim$5 µm thick SiO$_2$ layer as a spacer between the fluorophores and the reflecting interface, which is the SiO$_2$-Si interface in this case (Figure 5). Small height differences of the fluorophores produce shifts in the fringes and change the period of oscillation.

Figure 5 Compare fluorescence emission on a glass slide and a layered reflecting substrate. (a) Emission spectrums of monolayer of fluorescein immobilized on a glass slide and on top of a Si-SiO$_2$ substrate with two different thicknesses of the oxide layer (10 nm difference). (b) Schematic of the Si-SiO$_2$ substrate (not to scale). (Source: L. Moiseev, C. R. Cantor, M. I. Aksun et al. Spectral self-interference fluorescence microscopy. Journal of Applied Physics 2004, 96(9): 5311-5315).
2.1.2 Physical model

We can model an emitting fluorophore as an oscillating dipole with a random orientation above the reflecting interface\(^6\). All three vectors: the emitter dipole moment \(\mu\), the wave vector \(k\), and the electric field vector \(E\), lie in the same plane; this is the plane of polarization of the emitted light (Figure 6(a)). If the environmental factors remain constant, the far-field amplitude of the electric field of a fluorescently emitted wave is proportional to the sine of the angle between the dipole and the wave vector. The emission is therefore non-uniform; a 3-D illustration of the emission profile of a classical dipole shows a donut shape (Figure 6(b)).

![Classical dipole model](image)

Figure 6 The classical dipole model of a fluorescence emitter. (a) Intensity and polarization of electric field emitted by a classical dipole. (b) 3-D emission pattern of a classical electric dipole. (Source: Lorenzo Pavisi and Philippe M. Fauchet, Biophotonics, Springer, 2008, 73)

To describe the emission pattern of a dipole above a reflecting surface, we consider the intensity and polarization of both the direct and reflected waves. First, two coherent waves are radiated from each dipole, one goes directly to the objective, and the other is incident on the mirror and reflected, propagates in parallel with the direct wave in
the far field and arrives at the same spot on the objective. Since the SiO$_2$-Si reflecting interface is far away from the dipole and the observation point is placed in the far field, near-field radiation is ignored and only the electrical fields is used to calculate the fluorescence intensity. Reflections of plane waves at the SiO$_2$-Si interface are sensitive to polarization and Fresnel reflection coefficients are used to describe reflection for each polarization. The direct, incident and reflected waves are decomposed into two orthogonal components, transverse electric field $E_{TE}$ and transverse magnetic field $E_{TM}$. The reflected fields are incident fields modified by Fresnel reflection coefficient $R_{TE}$ and $R_{TM}$ at the SiO$_2$-Si interface.

Figure 7 Dipole emission model showing the direct, incident and reflected waves $k_{inc}$, $k_{dir}$ and $k_{ref}$. The emitting dipole is located in the x-z plane. (Source: Moiseev, L., Cantor, C. R., Aksun, M. I., Dogan, M., Goldberg, B. B., Swan, A. K., & Ünlü, M. S. (2004). Spectral self-interference fluorescence microscopy. Journal of Applied Physics, 96(9), 5311-5315).
Figure 7 illustrates the angle of the dipole $\theta$ as well as the plane of incidence defined by the polar angle $\theta_{em}$ and azimuthal angle $\varphi$ of the observation point and $z$-axis. The propagation directions of the direct, incident, and reflected waves all lie in the plane of incidence. Therefore, the angular dependence of $E_{TE}$ and $E_{TM}$ can be presented as follows:

$$E_{TE}^{dir} = E_{inc}^{dir} \propto \sin \theta \sin \varphi,$$

$$E_{TE}^{refl} = E_{inc}^{TE} R_{TE} e^{i2\phi},$$

$$E_{TM}^{dir} \propto \cos \theta_{em} \sin \theta \cos \varphi - \sin \theta_{em} \cos \theta,$$

$$E_{TM}^{inc} \propto \cos \theta_{em} \sin \theta \cos \varphi + \sin \theta_{em} \cos \theta,$$

$$E_{TM}^{refl} = E_{inc}^{TM} R_{TM} e^{i2\phi},$$

where $\phi = \left(\frac{2\pi n}{\lambda}\right) hc \cos \theta_{em}$, $h$ is the axial position of the dipole. $R_{TE}$ and $R_{TM}$ are Fresnel coefficients for $E_{TE}$ and $E_{TM}$, which take into account of the thickness of the SiO$_2$ spacer layer and the wavelength-dependence of refractive indices of Si and SiO$_2$. For SSFM, since it is the spectral oscillation of fluorescent emission that is important rather than absolute emission intensity, only the angular dependence of the electrical fields needs to be considered, and the proportionality sign ($\propto$) is used. The total intensity can be represented by the absolute square of the total electrical fields at the observation point in the far field. The intensity is given as

$$I = |E_{TE}|^2 + |E_{TM}|^2,$$

where

$$E_{TE} = E_{TE}^{dir} + E_{TE}^{refl},$$
$$E_{TM} = E_{TM}^{dir} + E_{TM}^{refl}.$$  

The above calculation of a dipole emission above a reflecting surface is for a specific direction characterized by $\theta_{em}$ and $\varphi$. In the applications of SSFM, samples contain thin layers of fluorophores that are assumed to be randomly oriented. Thus, the total emission of a monolayer of random oriented dipoles should also be integrated over all possible angles of $\varphi$ and $\theta$. However, the range of polar tilt angles can sometimes be restricted. We should also integrate over $\sin\theta d\theta$ for light collected by a microscope objective with the maximum collection angle $\theta_{em}^{max}$. As a result, the total emission intensity of a monolayer of random dipoles measured with an objective with a maximum collection angle $\theta_{em}^{max}$ is

$$I_{\text{total}} = \int_{\varphi=0}^{\pi} \int_{\theta_{em} = 0}^{\theta_{em}^{max}} I(\theta, \varphi, \theta_{em}) \sin\theta_{em} d\theta_{em} d\varphi d\theta.$$  

The model needs to be modified if there are additional layers between the dipole and the microscope objective. The dipole can be considered as residing in a cavity, affecting both the direct and reflected fields in the same way and the emission of a dipole in such an environment may be modeled in two steps: first, the total fields without the overhead layers is:

$$E_0 = E_{\text{dir}} + E_{\text{refl}}$$  

Then, the resulting field should be multiplied by the “cavity coefficient” caused by multiple reflections from the top and bottom interfaces inside the cavity, and the transmission to the objective. The total field at the objective is

$$E = \frac{E_0 T'}{1 - R'R'^2}, \quad \phi = \frac{4\pi n}{\lambda} D \cos\theta.$$
where $R'$ and $T'$ are the generalized reflection and transmission coefficients for the layers above the fluorophore in the direction towards the objective, $R$ is the generalized reflection coefficient below the fluorophore, and $D$ and $n$ are the thickness and refractive index of the cavity.

The above expression needs to be calculated for each wavelength in the fluorescence spectrum. The final fluorescence interference spectrum is composed of the envelope of the free-space fluorescence emission spectrum and the oscillatory modulation determined by the axial position.

### 2.1.3 Combine SSFM with white light reflectance spectroscopy

SSFM is combined with white light reflectance spectroscopy (WLRS), which offers great precision in determining the thickness of a transparent spacer layer on a reflecting surface. WLRS has been employed for label-free detection of bimolecular interactions through monitoring the thickness of the biological layer (bio-layer)\(^{87,88,89}\). The surface density of bound biomolecules can be quantified through calibrating the thickness of the bio-layer to known deposited biomass and molecular weight\(^{90-94}\). The advantage of SSFM combined with WLRS is that not only the conformations of DNA can be characterized, but also the thickness and surface density of DNA are orthogonally quantified. The ability to quantify both biomolecule density and conformation are not only desired to characterize the complex biochemical or biophysical properties of a biosensing interface, but also are needed for the study of specific molecular binding
mechanisms. This unique and powerful capability will be demonstrated in studying protein-DNA interactions in later chapters.

In WLRS, typically a broadband light source is focused on to the surface and reflected light is collected. The interference fringes in the collected spectrum are created by waves reflecting from the top and bottom interfaces in contrast to the direct and reflected waves in SSFM (Figure 8). The resulting oscillations are based on the total reflectivity or optical thickness of the stack of dielectric layers.

Figure 8 Compare detection mechanism between WLRS and SSFM. (Left) WLRS is based on spectral variations of reflection from thin transparent films. Interference of light reflected from the top surface and a buried reference surface results in periodic oscillations. (Right) SSFM maps the spectral oscillations emitted by a fluorophore located on a layered reflecting substrate into a precise axial position.

Often lower surface density of molecules results in lower refractive index, and in our cases we can assume the refractive index of the biomaterial to be close to that of SiO$_2$\(^{95-97}\). Thus, the bio-layer thickness is obtained by subtracting a reference thickness
measured at a nearby background from the thickness of the DNA or protein spots. However, deposition of DNA and protein molecules on the surface could change the refractive index, and we should regard the thickness measurements with caution. For both SSFM and WLRS, we can only determine optical path length, which is the product of refractive index and physical path length. However, the additional amount of biomaterial on the surface can always be quantified accurately by calibrating the change of optical path length or measured thickness to known deposited biomass\textsuperscript{96}.

It needs to be noted that in SSFM, the amplitude of direct wave and reflected wave from the SiO\textsubscript{2}-Si interface is comparable. However, for WLRS, the reflection from air-SiO\textsubscript{2} or buffer-SiO\textsubscript{2} interface is much less than that of the SiO\textsubscript{2}-Si interface, resulting in reduction of fringe contrasts (Figure 8). Since the white light source is external, we can increase light intensity to achieve sufficient spectral fringe contrast not to affect the precision of thickness determination.

2.1.4 Spectral analysis

Typical SSFM spectra are composed of three parts: the spectral envelope represented by the fluorophore emission profile, the oscillatory interference component, and the noise from the spectrometer camera. Both WLRS data and SSFM spectra are fit with custom algorithms executed with MATLAB. In the SSFM fitting algorithm, an estimated initial axial position of the fluorophore to the reflecting interface is used to generate an oscillatory curve using the described physical model. Then, the spectrum is divided by the generated curve resulting in the envelope curve, which is fit to a low order
polynomial. If the axial position is not accurate, the envelope curve is not decoupled from oscillations. The final position is determined iteratively until the envelope spectrum is smooth, free from oscillations, and can be fit to a low-order polynomial with minimum Least Square Fitting (LSF) error.

The WLRS fitting algorithm follows the same procedure except that the initial fitting parameter is the thickness of the SiO$_2$. Both algorithms take into account of the wavelength-dependence of the refractive index of SiO$_2$ and the complex reflectivity of the underlying Si. Sufficient number of oscillations in the interference spectrum is required to precisely determine the fluorophore axial position. For traditional SSFM using a substrate with of 5 µm thick SiO$_2$, at least two periods of oscillations, corresponding to a bandwidth of about 1000cm$^{-1}$ were needed to precisely determine axial height with deviation smaller than 1 nm$^{98}$.

The fitting procedure uses laws of linear algebra and optimizes only one variable of interest, the axial height of the fluorophore. Thus the fitting algorithm is very fast and permits real-time feedback within a second during an experiment, and allows thousands of spectrum measurements to be fit in a short time. Also, the fitting algorithm is immune to spectral modifications and any potential fluorescence quenching or photobleaching.

Here, we note that in this dissertation the axial locations of the fluorophores are described as the heights of fluorophores above the SiO$_2$ surface. This is because the fluorophore height above the surface is directly related to the geometric conformation of DNA on the surface. In dry measurements, the SSFM substrate resides in air and the fluorophores are assumed to be right on top of the SiO$_2$ surface. The variable of the
fitting algorithm is the SiO$_2$ thickness. In wet measurements, the SSFM substrate resides in aqueous solution and the fluorophores tagged on surface-immobilized DNA is above the SiO$_2$ surface. The variable of the fitting algorithm is the height of the fluorophore to the SiO$_2$ surface using a SiO$_2$ thickness obtained from dry measurements or WLRS. We also note that the fluorophore heights obtained are ensemble average height of all the fluorophores within the diffraction-limited spot. Similarly, WLRS measures the average bio-layer thickness of measured area defined by the specific optical setup.

2.2 Previous applications

The application of SSFM to characterize DNA molecules immobilized on a surface was motivated by the boom of DNA sensors and nanodevices in the last decade. Taking advantage of DNA designable sequence, chemical stability, and automatic complementary base pairing property, researchers have exploited surface-immobilized DNA sensors and nanodevices in wide-ranging biomedical applications$^{58,99-104}$. For instance, DNA microarrays, which allow for massively parallel multiplexed analysis, have become routine for gene-expression profiling in both research and clinical practice$^{63,105,106}$. In recent years, utilizing surface-immobilized DNA, researchers have developed switch-based highly sensitive biological detectors and dynamic programmable DNA motors$^{101,104,107-110}$. To understand and optimize the sensing and actuation performance of these DNA sensors and nanodevices, it is highly desirable to characterize two vital parameters, the DNA conformation and surface density, which dictate the behaviors of the surface-immobilized DNA molecules$^{90,111-115}$.
The following paragraphs of this section summarized two previous applications of SSFM in characterizing the conformation of surface-immobilized DNA. Meanwhile, models for quantifying the orientation of double-stranded DNA (dsDNA) on neutral and charged surfaces will be presented. The use of WLRS for DNA surface density quantification was not used in these two examples and will be shown in the next section.

2.2.1 Estimate the conformation of surface-immobilized DNA

SSFM was used to estimate the conformation of ssDNA and dsDNA of 50- and 21-base-pair (bp) long immobilized on silanized SSFM substrate (~5 µm thick thermal grown SiO₂ layer on Si)⁸⁷. The optical thicknesses of the SiO₂ layer was determined by WLRS and used in SSFM fitting. The 5’ end of the first strand of the DNA was amino-modified and covalently linked to the surface. Fluorophores were tagged either at the 3’ end of the immobilized first strand or at the 3’ or 5’ end of the second complementary strand. A summary of the results is shown in Table 1.

Adding complementary second strands to 50-bp ssDNA resulted in an increase in the DNA film thickness by approximately 1.0 nm corresponding to a hybridization efficiency of approximately 50%. In principle, the maximum heights of the fluorophores are constrained by the contour length of the dsDNA, which are approximately 7 nm and 17 nm for 21-bp and 50-bp dsDNA respectively. The average heights of the distal end fluorophores are 5.5 nm for 21-bp and 10.5 nm for 50-bp dsDNA, whereas those of the proximal end are 1 nm for 21-bp and 2.5 nm for 50-bp dsDNA. Assuming the short dsDNA are rigid, the orientations of the 21-bp and 50-bp long dsDNA to the surface are
approximately 40° and 50°. A more detailed description of the calculation of DNA orientation using a simple geometric model will be described later in the next subsection.

Table 1 Average height of the fluorophores above the surface measured by SSFM and the optical thickness of the DNA layer on the surface measured by WLRS. (Source: Moiseev, L., Unlü, M. S., Swan, A. K., Goldberg, B. B., & Cantor, C. R. DNA conformation on surfaces measured by fluorescence self-interference. Proceedings of the National Academy of Sciences of the United States of America 2006, 103(8): 2623–2628)

<table>
<thead>
<tr>
<th></th>
<th>50-mer</th>
<th>21-mer</th>
</tr>
</thead>
<tbody>
<tr>
<td>experiments</td>
<td>First strand is not labeled and hybridized with a labeled strand</td>
<td>First strand is labeled and hybridized with a non-labeled strand</td>
</tr>
<tr>
<td>SSFM (nm)</td>
<td>–</td>
<td>10.5</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>6.5</td>
<td>7.5</td>
</tr>
<tr>
<td></td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>white light (nm)</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>2.5</td>
</tr>
</tbody>
</table>

To study the conformation of single-stranded DNA (ssDNA), fluorophore heights at the 3’ (distal) end of surface-immobilized ssDNA were measured. Unlike dsDNA, ssDNA is very flexible, often described as random coils, and little is known about its conformation on surfaces. The height of the fluorophore was about 1 nm for 21-bp ssDNA and 5.5 nm for 50-bp ssDNA, which implied a considerably more extended conformation for the 50-bp ssDNA. When a second unlabeled complementary strand was hybridized, the surface DNA molecules are composed of two species, unhybridized
ssDNA and hybridized dsDNA, both labeled with fluorophores at the distal ends. Thus the average fluorophore height was somewhere between the distal fluorophore heights of ssDNA and dsDNA. The hybridization efficiency can also be estimated based on the change of fluorophore average height before and after hybridization, which was around 30% to 50% in this experiment and was consistent with estimation made by WLRS. The results demonstrate that SSFM can be used to estimate the conformation of surface-immobilized ssDNA and dsDNA: single-stranded DNA can be flexibly extended on the surface whereas double-stranded DNA is rigid and angled from the surface.

2.2.2 Geometric model for DNA orientation calculation

SSFM measures the orientation of surface-immobilized dsDNA using a simple trigonometric model. The contour length of short surface-immobilized dsDNA (<60 bp) is much shorter than its persistence length (about 50nm)\textsuperscript{117}. Therefore, each dsDNA molecule is modeled as a rigid rod tethered to the surface on a pivot. The other end of each dsDNA molecule can rotate around the pivot at various orientations to the surface. SSFM measurements render the average heights of the surface-distal and surface-proximal ends of dsDNA and their average height difference (\(\Delta h\)) by subtraction. The length of the dsDNA is defined by the root-mean-square (rms) end-to-end distance (\(\sqrt{\langle r^2 \rangle}\)) of the dsDNA based on the worm-like chain model\textsuperscript{118}. Therefore, the average orientation of dsDNA measured relative to the surface is \(\theta = \arcsin (\langle \Delta h \rangle / \sqrt{\langle r^2 \rangle})\), where \(\sqrt{\langle r^2 \rangle} = \sqrt{2L^2 (l/L - 1 + e^{-L/l})}\).
If we consider each orientation of the dsDNA to the surfaces as an energy state, the hemisphere in Figure 9 illustrates all the states possibly accessible under random thermal fluctuation on a planar surface. Without any constraints, the probability distribution function (PDF) of the orientations of each dsDNA molecule is:

\[ f(\theta) = \frac{\cos(\theta)}{\int_0^{\pi/2} \cos(\theta) \, d\theta} = \cos(\theta), \left(0 \leq \theta \leq \frac{\pi}{2}\right) \]

Thus, the mean orientation of a dsDNA molecule under random thermal fluctuation or the mean orientation of large number of dsDNA molecules is:

\[ \langle \theta \rangle = \int_0^{\pi/2} \theta f(\theta) \, d\theta = 33^\circ \]

The upper limit of the integration represents the maximum polar angle (the complement of the minimum orientation to surface) accessible for the dsDNA molecules on a planar surface. According to Manning’s counterion condensation theory, dsDNA is inherently negatively charged in electrolyte solutions with equally spaced negative point charges. Thus due to electrostatic repulsion and steric hindrance, closely tethered dsDNA molecules may rotate with smaller than 90° maximum polar angles, resulting in average orientations higher than 33°.
Figure 9 A dsDNA molecule is modeled as a rigid rod with an orientation $\theta$ to the surface. $\sqrt{\langle r^2 \rangle}$ is the root-mean-square of the end-to-end distance of dsDNA under experimental conditions. $\Delta h$ is the axial height between the surface-distal end and surface-proximal end of the dsDNA. The hemisphere illustrates the orientations accessible by an unconstrained dsDNA molecule under random thermal rotations around the pivot on a planar surface.

The average orientation of many dsDNA molecules measured by SSFM is an approximation of their statistical mean orientation. As mentioned above, SSFM renders the ensemble average axial height difference ($\langle \Delta h \rangle$). The average orientation of many dsDNA molecules is calculated as $\theta = \arcsin (\langle \Delta h \rangle / \sqrt{\langle r^2 \rangle})$ rather than $\langle \theta \rangle = \arcsin (\Delta h / \sqrt{\langle r^2 \rangle})$, where $\Delta h$ is the axial height difference between the surface-distal and surface-proximal ends of each individual dsDNA. A simple simulation calculating the average orientation measured by SSFM and the statistical mean orientation of dsDNA under the probability density distribution $f (\theta)$ is shown in Figure 10. The simulated result shows that under random thermal fluctuation, this approximation
made by SSFM underestimates the mean orientation ($\langle \theta \rangle$) of surface-immobilized dsDNA by 2-3 degrees.

Figure 10 Approximation of the mean of dsDNA orientation with calculated average orientation of dsDNA using ensemble average height difference $\langle \Delta h \rangle$ measured by SSFM.

Furthermore, we simulated the relationship between the dsDNA molecules surface density and their average orientation on a 2-D planar surface as result of just steric hindrance amongst adjacent dsDNA molecules. We first calculate the average horizontal distance between nearby dsDNA molecules from dsDNA surface density by:

$$\langle d \rangle = 2/ \sqrt{\pi \rho}$$

where $\rho$ denotes the dsDNA surface density. Using the average inter-dsDNA distance and the rms end-to-end distance ($\sqrt{\langle r^2 \rangle}$) of dsDNA, the maximum orientation of dsDNA to the surface is:

$$\theta_{max} = \arccos (\langle d \rangle/ \sqrt{\langle r^2 \rangle})$$

Thus the maximum dsDNA orientation is a function of dsDNA surface density:
\[ \theta_{\text{max}} = \arccos \left( \frac{2}{\sqrt{\pi \rho \cdot (r^2)}} \right) \]

Using the PDF of dsDNA orientation under random thermal fluctuation, the average dsNDA orientation can be calculated as:

\[ \langle \theta \rangle = \int_0^{\theta_{\text{max}}} \theta \cos (\theta) d\theta \]

Figure 11 Simulated relationship between DNA orientation and DNA surface density on a planar surface. (a) Horizontal distance between adjacent DNA molecules as a function of DNA surface density. (b) DNA orientation as a function of DNA surface density as a result of steric hindrance between nearby DNA molecules.

The simulation of average dsDNA orientation at a particular surface density provides a reference for the interpretation of dsDNA orientation on 3-D polymeric surface.

2.2.3 Quantification and control of surface-immobilized DNA orientation

For biosensing and DNA-based molecular devices, sensitive detection and accurate actuation requires ordering and cooperation of the molecules, such as directed movements and simultaneous response of the molecules. Some techniques, such as
“switchable DNA”\textsuperscript{103}, induce ordering of DNA monolayers immobilized on gold electrodes by applying alternating electrical fields. Negatively charged short oligonucleotides are switched between lying and upright positions on oscillations of positive and negative surface potentials.

A similar concept of electrical field induced ordering was adopted on the SSFM platform, but a novel electromechanical approach was used to orient surface-immobilized dsDNA. The SiO\textsubscript{2} surface was functionalized with a highly amphoteric polymer that adopts a net negative or positive charge depending on the buffer pH\textsuperscript{120-122}. The polymer also has N-Hydroxysuccinimide (NHS-ester) groups that covalently bond with amino-modified DNA molecules. The isoelectric point of the polymer was tested to be around pH 6 by electro-osmotic flow in a capillary coated by a polymer with identical composition. Lower buffer pH results in a net positively charged surface, attracting dsDNA to the surface whereas higher buffer pH results in a net negatively charged surface, repelling dsDNA to a higher orientation to the surface\textsuperscript{123}. An overview of the manipulation of dsDNA orientation on a charged polymer surface through adjustment of buffer pH and ionic strength is shown in Figure 12.
Figure 12 Control and quantification of 60-bp dsDNA orientation on a charged polymer surface. (a) SSFM measurements of 60-bp dsDNA on a charged polymer functionalized substrate. Surface-proximal labeled dsDNA indicate polymer swelling: positively charged polymer collapses onto the negatively charged oxide surface and negatively charged polymer is repelled from the oxide surface. Fixed charges within the polymer also repel each other and contribute to polymer swelling. The surface-distal and surface-proximal fluorophore heights allow precise quantification of dsDNA orientation. (b) Average dsDNA orientations at each pH in buffers containing NaCl of concentrations at 40 mM and 300mM. The dsDNA is oriented in a lying position by the positively charged polymer and in a standing position by the negatively charged polymer. A low salt concentration allows the electric field to penetrate far from the charged polymer to more effectively orient immobilized dsDNA.

The controlled orientation of dsDNA on the charged polymer can be interpreted by considering the electrical DNA switching on a gold surface described above, which was studied extensively by Rant and coworkers. A diffuse double layer of counterions accumulates at the interface of a charged surface and an electrolyte solution. An intense electric field (~100 kV/cm) results near the ionic buffer-polymer interface due to the high concentration gradient of mobile charges that accumulate. The characteristic length of this electric field (on the scale of nanometers) is inversely proportional to the square root of buffer salt concentration. At low ionic strength (10 mM), for instance, the thickness of the double layer is around 3 nm whereas at high ionic strength (300 mM) it
is about 0.6 nm. Thus, the electrical field does not cover the entire length of the dsDNA probes and the electrostatic interactions are confined to the base of the probes. As a result, the charged polymer surface orients the immobilized dsDNA more effectively in low ionic buffers because the electrostatic force is applied to a larger proportion of the dsDNA (Figure 12(b)).

To gain further insight into the behavior of the charged polymeric surface on the SSFM platform, the model proposed by Rant and co-workers\textsuperscript{124} was adapted. First, we regard the charged polymer surface as a charged plane, the surface potential of which varies with buffer pH. Since dsDNA is inherently negatively charged in electrolyte solution, a negatively charged surface repels dsDNA and positively charged surface attracts dsDNA. This is the principle of induced ordering of oligonucleotides for both gold and charged polymer surfaces. In electrolyte solution, the surface electrostatic potential is screened by redistributed ions, and according to the Gouy-Chapman theory, reduces as a function of distance to the surface with a characteristic Debye length. The Debye length $l_D$ is defined as
\[
l_D^2 = \frac{\varepsilon \varepsilon_0 kT}{2ne^2},
\]
where $\varepsilon$ is the dielectric constant of ionic solution, $\varepsilon_0$ is permittivity of vacuum, $k$ is the Boltzmann constant, $T$ is temperature, $e$ is elementary charge and $n$ is ion density. The Gouy-Chapman equation describes the diffusive potential distribution along the z-axis:
\[
\Phi(z) = \frac{2kT}{e} \ln \left( \frac{1 + ye^{\frac{z}{l_D}}}{1 - ye^{\frac{-z}{l_D}}} \right), \gamma = \tanh \left( \frac{e\Phi_0}{4kT} \right).
\]
where $\Phi_0$ is the surface potential. Again, dsDNA is modeled as rigid rod rotating around a pivot on the surface. A charge of $-0.24 e^{126}$ is assumed per point charge along dsDNA to account for the counterion condensation effect$^{127}$. The electrostatic energy of charges along the dsDNA depends on their axial location relative to the surface, which is defined as $h = r \sin\theta$, where $\theta$ is the orientation of dsDNA to the surface, $\sqrt{r^2}$ is the root mean square end-to-end length of the DNA, simplified as $r$ for clarity in later equations. Average height of surface-distal end ($h$) is used to be consistent with the model, assuming surface-proximal end is at zero height from the charged surface plane.

Thus, the electrostatic energy of each rod is a function of $\theta$ and is calculated as the sum of the electrostatic energy of all the charges$^{124}$:

$$E(\theta) = E_{(\Phi_0, l_D)}(\theta) = \sum_i q_{\text{eff}}(z_i) \Phi(z_i).$$

The average orientation of dsDNA is a balance between thermal stochastic three-dimensional rotations and the electrostatic force on the dsDNA. In the presence of varying electric potential, the average orientation is calculated by taking the Boltzmann distribution to calculate the probability density of all the energy levels, thus

$$\langle \theta \rangle = \frac{\int_0^\pi \theta \cos(\theta) e^{-\Delta E(\theta) / kT} d\theta}{\int_0^\pi \cos(\theta) e^{-\Delta E(\theta) / kT} d\theta}.$$

$\cos(\theta)$ is the degeneracy factor, representing the stochastic rotations as degenerate microstates, which occupy the same energy level at the same orientation. The potential of a rod for a given orientation is calculated as $\Delta E(\theta) = E(\theta) - E(90^\circ)$, where the standing rod ($\theta = 90^\circ$) is taken as the reference state. Figure 13 shows the PDFs of dsDNA.
orientations for different surface potentials, $\Phi_0$. We can see that when $\Phi_0$ is positive, most of the dsDNA are at lower orientations and when surface potential is above a critical positive value, almost all of the dsDNA will be pulled down within 2 degrees from the surface with the higher orientation energy states less accessible and unpopulated. When $\Phi_0$ goes from 0 to negative potentials, the PDFs shift to higher orientations with narrower distributions with the low orientation energy states being less probable.

Figure 13 Probability Density Functions (PDF) of dsDNA orientation $\theta$ for different surface potentials. Positive potentials pull the dsDNA to the surface, resulting in most dsDNA occupying the lower orientation states. The PDFs are shifted to higher degrees as the surface potential becomes more negative.

Again, in the experimental data of SSFM, we approximate the mean of dsDNA orientation $\langle \theta \rangle$ as $\theta = \arcsin (\langle h \rangle / r)$. To determine the validity of the approximation, the
dsDNA average orientation is calculated using simulated surface-distal end average height ($\langle h \rangle$) obtained by:

$$\langle h \rangle = \frac{\int_0^\pi r \sin(\theta) \cos(\theta) e^{-\frac{\Delta E(\theta)}{kT}} \, d\theta}{\int_0^\pi \cos(\theta) e^{-\frac{\Delta E(\theta)}{kT}} \, d\theta}$$

The analysis again shows that the approximation underestimates $\langle \theta \rangle$ by 2 to 3 degrees (Figure 14).

Figure 14 Average dsDNA orientation $\langle \theta \rangle$ and calculated dsDNA orientation $\theta$ from average height measurement as a function of surface potential for 60-bp dsDNA. When the surface potential equals to 0, $\langle \theta \rangle$ is 33°, corresponding to the mean orientation under random thermal fluctuation. When the positive surface potential exceeds a particular value, most of dsDNA will be pulled down to the surface with the average orientation of less than 2°. As the surface potential goes further negative, the PDF shifts less and the $\langle \theta \rangle$ eventually plateaus.

To get a further insight into the working mechanism of the charged polymer surface, we analyzed the dsDNA orientation over a wide range of salt concentrations. At pH 7.6, when the charged polymer is negatively charged, the measured orientation
matches well with the calculated results at low ionic strength, with the maximum orientation exceeds 70° (Figure 15). At high ionic strength the effects of negative charge on the polymer become negligible, but the average dsDNA orientation approaches 45° while the expected average orientation is 33° as calculated previously. The 45°
orientation of probes at high ionic strength indicates steric repulsion between the polymer and the base of the dsDNA and between the negatively charged dsDNA themselves. At pH 4.4, when the polymer surface is positively charged, a nearly lying position of dsDNA is induced at low ionic strengths, whereas the orientation once again approaches 45° for high ionic buffer strengths. Since DNA surface density can affect their orientation, it is possible that at the particular DNA surface density in this experiment, the DNA molecules sterically constrained each other’s maximum random thermal rotation, which could result in higher average orientations. In the last section of this chapter, the effect of DNA surface density on a polymer surface is evaluated and discussed.
Figure 15 dsDNA orientation over a wide range of salt concentrations on the charged polymer surface. At pH 4.4, the polymer surface is positively charged, attracting the negatively charged dsDNA; at pH 7.6, the polymer surface is negatively charged, repelling the negatively charged dsDNA. (a) Prediction of orientation using theoretical model. (b) Average orientation of 60-bp dsDNA measured by SSFM. *(Source: Spuhler, P. S., Sola, L., Zhang, X., Monroe, M. R., Greenspun, J. T., Chiari, M., and Ünlü, M. S. Precisely Controlled Smart Polymer Scaffold for Nanoscale Manipulation of Biomolecules. Analytical Chemistry 2012, 84(24): 10593–10599).*

In the theoretical model, the steric effects between the polymer scaffold and the dsDNA are not included in the calculations. Thus, the calculated orientations deviate
from experimental results, and approach 0° for a positively charged surface at low ionic strength and 33° at high ionic strengths. However, the model accurately predicts the ionic strength at which the dsDNA sharply transits from a random to a lying orientation when the polymer is positively charged. This relationship between the probe orientation and the ionic strength reflects the different mechanisms by which DNA are attracted to a positively charged surface as opposed to repelled from a negatively charged surface. As discussed previously, the average orientation measured by SSFM is a balance between the electrostatic force and stochastic thermal motion. On the positively charged polymer surface, when buffer ionic strength transits from high to low, the Debye length is increased and the electrostatic energy is higher for all dsDNA orientations. In the beginning, the attracting electrostatic force does not exceed the thermal motions of dsDNA for most of the orientational states. As the buffer ionic strength goes further down, the dsDNA whose orientation passes a threshold where the attracting electrostatic force dominates over the thermal motion are pulled down and captured. For SSFM steady state ensemble measurements, the measured orientation is a weighted average of the orientations of the captured and un-captured probes. As the ionic strength decreases beneath the threshold value, the electrostatic force is large enough to effectively capture a majority of the dsDNA probes. Therefore, the calculated and experimental average orientations are both seen to switch from freely rotating to a horizontal orientation.
2.3 Dual-color SSFM

The oscillation period of SSFM interference spectrum is inversely proportional to the thickness of the SiO\textsubscript{2} (period $\propto D/\lambda$, $D$ is SiO\textsubscript{2} thickness, $\lambda$ is wavelength, see Physical model section). Therefore, it is possible to obtain sufficient oscillations in multiple non-overlapping spectral bands corresponding to different fluorophores at different axial locations using thicker SiO\textsubscript{2} layers. This section presents a recent improvement of SSFM platform under the work of this dissertation. We developed a dual-color SSFM that uses two distinct spectral ranges to precisely determine the axial heights of two different fluorophore layers. Thus, dual-color SSFM enables nanoscale characterization of DNA conformation by determining the average axial height difference between two positions on a single DNA spot on the surface (Figure 16).

Moreover, dual-color SSFM is combined with modified WLRS to quantify the DNA surface density to evaluate ensemble conformational measurements. Instead of a white light source for illumination, a light-emitting diode (LED) is used. This modality of WLRS will be referred as LED-RS. The emission spectrum of the LED is selected to be distinct from that of the red fluorophores. Thus the bio-layer thickness and the red fluorophore height can be simultaneously measured with a single spectrum acquisition as shown in Figure 17. The bio-layer thickness provides information regarding the molecular surface density while the fluorophore heights indicate the conformation of the DNA molecules at the same location on the surface.
In this section, we first present the experimental approaches to implement the combined dual-color SSFM and LED-RS system. Then, using the combined system, we demonstrate characterization of DNA conformation under different conditions on a 3-D polymer-coated surface. Furthermore, we show simultaneous and independent real-time monitoring of DNA conformational change and complementary strand binding during DNA surface hybridization.

Figure 16 Dual-color SSFM detection principle. (a) Amino-modified DNA molecules labeled Atto647N (red fluorophores) and Atto532 (green fluorophores) on opposite ends are covalently immobilized on a polymer-functionalized substrate. Dual-color SSFM measures the height difference ($\Delta h$) between the DNA surface-distal end (green fluorophores) and the surface-proximal end (red fluorophores). (b) The measured fluorescence interference spectrums of Atto647N and Atto532 on the layered substrate. A few oscillations of each spectrum are fit into an average axial height of the fluorophores to the SiO$_2$-Si interface with sub-nanometer accuracy. Dotted lines are the emission spectral profile of the two fluorophores given by the manufacturer.
Figure 17 SSFM combined with LED based white light reflectance spectroscopy. (a) Compare the detection principles of LED-RS and SSFM. Yellow dashed line indicates the thickness of the bio-layer measured by LED-RS while red dotted line indicates the average height of red fluorophores (Atto647N) measured by SSFM. (b) Simultaneous acquisition of the average fluorophore axial height and bio-layer thickness. The black line represents the interference spectrum for the wideband LED and the red fluorophores taken from a single acquisition. Yellow dashed line is the spectral envelope of the LED measured with a silver mirror while the red dotted line is the emission spectral envelope of Atto647N given by the manufacturer.

2.3.1 Experimental approach

Materials: All buffer solutions were prepared with deionized water (DI water) filtered by Barnstead Nanopure Diamond (18.2 MΩ·cm⁻¹, Thermo Scientific, Waltham, MA). Buffer used for wet measurements was NaCl/Tris solution prepared with Tris buffered saline tablets and NaCl (Tris: 10 mM, NaCl: 50 mM, pH 7.6, Sigma-Aldrich) unless specified. DNA sequences were designed with Oligo Analyzer (Integrated DNA Technologies, Inc., Coralville, IA) to minimize secondary structures. Single-stranded 60-bp DNA molecules were synthesized by IBA GmbH (Goettingen, Germany), and single-
stranded 20-bp and 40-bp DNA molecules were synthesized by IDT, Inc. (Coralville, IA).
All single-stranded DNA (ssDNA) molecules were purified with high-performance liquid chromatography (HPLC) after synthesis. For the purpose of immobilization, ssDNA molecules were modified with amine-C6 linkers on the 5’ or 3’ end. Fluorophores Atto647N or Atto532 were labeled on the other end of the amino-modified strand or on the 5’ or 3’ end of the complementary strand. The DNA sequences and modifications used in different experiments are shown in Table 2. The dsDNA molecules were hybridized at 30 µM in 150 mM sodium phosphate buffer (prepared with sodium phosphate monobasic and sodium phosphate dibasic, pH 8.5, Sigma-Aldrich) before spotting. Microarray substrates were prepared from silicon wafers that had 17.5 µm thick thermally grown oxide (Silicon Valley Microelectronics, Santa Clara, CA), which were cut into 15 mm × 15 mm square chips. Before use, chips were rinsed with acetone, methanol, subjected to sonication in acetone for 5 min, and oxygen plasma ashing (M4LTM, PVA TePla America, Corona, CA) for 10 min.
Table 2 DNA sequences and nomenclature used in dual-color SSFM.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>60-bp dsDNAs</strong></td>
<td></td>
</tr>
<tr>
<td>Seq1 (Scheme1)</td>
<td>5'-GCT GTT AGA AGA TAG GGC CAA AAA AGC ATT GCT TAT CAA TTT GTT GCA CCT GAC CGA TGA-3'-C6-NH₂</td>
</tr>
<tr>
<td>Complement</td>
<td>5'-Atto647N-TCA TCG GTC AGG TGG AAC AAA TTG ATA AGC AAT GCT TTT TTG GCC CTA TCT TCT AAC AGC-3'</td>
</tr>
<tr>
<td>Seq2 (Scheme2)</td>
<td>NH₃-C6-5'-GCT TGT AGA AGA TAG GGC CAA AAA AGC ATT GCT TAT CAA TTT GTT GCA CCT GAC CGA TGA-3'-C6-NH₂</td>
</tr>
<tr>
<td>Complement</td>
<td>5'-Atto647N-TCA TCG GTC AGG TGG AAC AAA TTG ATA AGC AAT GCT TTT TTG GCC CTA TCT TCT AAC AGC-3'</td>
</tr>
<tr>
<td>Seq3</td>
<td>5'-GCT GTT AGA AGA TAG GGC CAA AAA AGC ATT GCT TAT CAA TTT GTT GCA CCT GAC CGA TGA-3'-C6-NH₂</td>
</tr>
<tr>
<td>Complement</td>
<td>5'-Atto532-TCA TCG GTC AGG TGC AAC AAA TTG ATA AGC AAT GCT TTT TTG GCC CTA TCT TCT AAC AGC-3'</td>
</tr>
<tr>
<td>Seq4</td>
<td>NH₃-C6-5'-GCT TGT AGA AGA TAG GGC CAA AAA AGC ATT GCT TAT CAA TTT GTT GCA CCT GAC CGA TGA-3'-C6-NH₂</td>
</tr>
<tr>
<td>Complement</td>
<td>5'-Atto532-TCA TCG GTC AGG TGC AAC AAA TTG ATA AGC AAT GCT TTT TTG GCC CTA TCT TCT AAC AGC-3'</td>
</tr>
<tr>
<td>Seq5 (Scheme3)</td>
<td>5'-Atto647N-GCT GTT AGA AGA TAG GGC CAA AAA AGC ATT GCT TAT CAA TTT GTT GCA CCT GAC CGA TGA-3'-C6-NH₂</td>
</tr>
<tr>
<td>Complement</td>
<td>5'-Atto532-TCA TCG GTC AGG TGC AAC AAA TTG ATA AGC AAT GCT TTT TTG GCC CTA TCT TCT AAC AGC-3'</td>
</tr>
<tr>
<td>Seq6 (Scheme4)</td>
<td>5'-Atto532-GCT GTT AGA AGA TAG GGC CAA AAA AGC ATT GCT TAT CAA TTT GTT GCA CCT GAC CGA TGA-3'-C6-NH₂</td>
</tr>
<tr>
<td>Complement</td>
<td>5'-Atto647N-TCA TCG GTC AGG TGC AAC AAA TTG ATA AGC AAT GCT TTT TTG GCC CTA TCT TCT AAC AGC-3'</td>
</tr>
<tr>
<td><strong>20-bp dsDNA</strong></td>
<td>NH₃-C6-5'-Cy5/GCT GTT AGA AGA TAG GGC CA</td>
</tr>
<tr>
<td>Complement</td>
<td>5'-Atto532N-TGG CCC TAT CTT CTA ACA GC</td>
</tr>
<tr>
<td><strong>40-bp dsDNA</strong></td>
<td>NH₃-C6-5'-Cy5/ATC TGA ACC CGC TAT CCG ATT CCG AGG C</td>
</tr>
<tr>
<td>Complement</td>
<td>5'-Atto532N-CCT CGG AAT CAA GTG CAT GGA ATA GCG GTG GTG TCA GAT</td>
</tr>
</tbody>
</table>

**DNA microarray and polymeric coating preparation:** Clean chips were functionalized with a commercial polymer product MCP-2 from Lucident Polymers (Sunnyvale, CA), which will be referred to as the polymer). The polymer is obtained from radical copolymerization of N,N-dimethylacrylamide(DMA), N-acryloyloxsuccinimide(NAS), and 3-(trimethoxysilyl)-propyl methacrylate (MAPS), and covalently binds to amino-modified DNA as described elsewhere. The polymer has been shown to have many advantages for DNA immobilization, such as high stability and reproducibility, fast and cost-effective preparation procedures, and minimal non-specific binding. The chips were immersed in a 1:1 mixture of 1% (w/v) MCP-2 in DI water and ammonium sulfate at 40% saturation concentration for 30 min, subsequently washed thoroughly with DI water, dried with argon gas, and then baked for 15 min at 80°C. DNA molecules were spotted in micro-arrayed fashion in 150 mM
sodium phosphate (pH 8.5) buffer using a robotic spotter (sciFLEXARRAYER, Scienion, Monmouth Junction, NJ), and then immobilized for twelve hours in 65% humidity at room temperature. The chips were then washed on an orbital shaker three times in 2x saline-sodium citrate buffer (SSC) for 10 minutes, three times in 0.2x SSC for 5 minutes, once in 0.1x SSC for 1 minute, dried with argon gas, and kept in vacuum until measurements were performed. The diameter of each individual DNA spot was about 150 μm. Low background and consistent spot morphology were previously shown for DNA microarrays on glass slides functionalized with the polymer. DNA spot morphology was optimized by changing spotting conditions, such as concentration, temperature and humidity, and examined with a fluorescence scanner before measurements. Within the measured area at the center of the DNA spots, all spots possessed satisfactory uniform morphology (Figure 18).
Figure 18 Uniform and consistent DNA spot morphology on the polymer surface. (a) A fluorescence image of example DNA spots immobilized on the polymer surface scanned by GenePix 4000B microarray scanner. The DNA molecules labeled with Atto647N are spotted at two concentrations, 5 µM and 10 µM. (b) Horizontal fluorescence intensity profile of eight DNA spots of two different densities. (c) Vertical fluorescence intensity profile of two DNA spots of different densities. The variability of fluorescence intensity within each spot is less than 10%, showing spotting is consistent from spot to spot. (d) Zoomed in fluorescence image of one DNA spot. Scale bar shows a distance of 100 µm.

**Measurements in a customized flow cell:** Both DNA orientation and hybridization measurements were performed in a customized flow cell, where the chip surface was incubated in buffer (Figure 19). A silicone rubber sheet (McMaster-Carr, Robbinsville, NJ), a glass window (Edmund Optics, Barrington, NJ), and the flow cell bottom formed the flow chamber, which were clamped and sealed using four screws. The custom-made glass window has anti-reflection coatings in the visible and near-infrared
spectrum specific for each side so as not to affect spectral signals. The height of the flow cell is 1 mm, the length of the flow cell is 20 mm, the width of the flow cell is 5 mm, and thus the working volume of the flow cell was 100 µL. Buffer solutions were filtered with 0.1 µm Millex-LG syringe filter (Millipore, Billerica, MA) before introduction to remove impurities. Each buffer solution was driven by a peristaltic pump (Control Company, Friendswood, TX) into and out of flow cell through the inlet and outlet via stainless steel and non-shedding silicon tubing and tubing connecters at a constant flow rate of 240 µL/min. The flow cell was fixed onto a 2-axis positioning micro-stage (Mad City Labs, Madison, WI), and the scanning of the DNA microarray was implemented by moving the stage by the DNA spot center-to-center distance with sub-micrometer accuracy.

Figure 19 Customized flow cell assembly. (a) Four different components of the flow cell are presented and assembled from top to bottom as indicated by the black arrow. (b) Cross section of an assembled flow cell with the SSFM chip (dark blue) fixed in place is shown.
**Optical setup** (Figure 20): The light beams of a red laser (helium-neon, 633 nm, Melles Griot, Carlsbad, CA) and a green laser (diode-pumped solid-state, 532 nm, Laserglow Technologies, Toronto, Canada) were combined with a dichroic beamsplitter. The combined laser beams were collimated, and expanded with two achromatic lenses (Thorlabs Inc., Newtown, NJ). A mechanical shutter that opens and closes via a transit TTL signal (+5V or 0V) controlled the red laser. A current control module regulated the green laser upon receiving a TTL signal (+5V or 0V). A printed circuit board (PCB) was designed to receive digital signals from the computer and send TTL signals to the shutter and the controller. The laser beams were then reflected by a dual-edge dichroic beam splitter (545/650 nm BrightLine, Semrock, Inc. NY), which specifically reflects the wavelengths of both lasers and transmits over 90% of the emission spectrums of the two fluorophores. A Nikon 5x objective with a numerical aperture (NA) of 0.13 focused the laser beams to diffraction-limited spots at the center of each DNA spot on the sensor surface. Emitted fluorescence was collected by the objective and focused onto a spectrometer connected to a CCD camera (SP-2150i, Princeton Instruments, Trenton, NJ). A yellow LED (M565L2, Thorlabs, Newton, NJ) was used as the illumination source for LED-RS measurements. A pinhole with a diameter of 100 µm was placed in front of the LED and a reduced image of the pinhole was focused to the same location as the laser spots. The reflected LED light was then collected and focused onto the spectrometer with the same focus as the laser beams. The spectrometer, the micro-stage, and the lasers were controlled via custom MATLAB software that performs automatic data acquisition (Figure 21). The interference spectrums consisted of three parts: the
fluorophore or the LED emission spectral envelope, the oscillatory interference component, and the shot noise collected from the CCD camera. Both fluorescence and LED interference spectra were fit with custom algorithms to obtain fluorophore heights and bio-layer thicknesses as described elsewhere\textsuperscript{132}.

Figure 20 Schematic illustration of the optical setup combining the dual-color SSFM and WLRS using a LED as illumination source. Components of the system shown are not to scale.
Figure 21 Data acquisition software application developed using MATLAB.

**Quantification of DNA surface density:** The DNA layer thickness measured by LED-RS was compared with that measured by interferometric reflectance imaging biosensor (IRIS). The two thickness measurements correlate linearly with each other and fit to a line through the origin (Figure 22). Thus, the calibration coefficient of IRIS, $1 \text{ng/mm}^2/\text{nm}$, was used to calculate the DNA mass surface density. Furthermore, using the DNA molecular weight and the Avogadro constant, the molecular surface density can be calculated from the mass surface density. For example, a 1nm thick
dsDNA spot (60-bp, molecular weight: 38700g/mol) renders a molecular surface density of $1.56 \times 10^{12}$ molecules/cm$^2$.

Figure 22 LED-RS measurement of DNA spot thicknesses linearly correlates with IRIS measurement. The scatter plots show mean and standard deviation of 10 dsDNA spots.

2.3.2 SiO$_2$ thickness design

Employing the interference spectrum at two different spectral ranges by dual-color SSFM and LED-RS on the same chip requires us to use substrates with a thicker SiO$_2$ layer. Previously for single-color SSFM using a substrate with 5 µm SiO$_2$ layer, at least two periods of oscillations that spanned a bandwidth of about 1000 cm$^{-1}$ within the fluorescence emission spectrum were needed to precisely determine fluorophore axial heights with less than 1 nm deviation$^{98}$. However, for the combined dual-color SSFM and LED-RS, the bandwidths usable for fitting without sacrificing signal-to-noise ratio were
much narrower, constrained by the emission spectrums of the fluorophores and the LED and the passbands of the available dichroic beam splitters and notch filters. Since the interference oscillation period is inversely proportional to the thickness of the SiO$_2$ layer$^6$, within the narrower bandwidths, we need to use thicker SiO$_2$ layer to obtain sufficient oscillation periods for precise height determination.

We simulated the fluorophore axial height deviation from expected values on substrates with different SiO$_2$ thicknesses. We determined the usable bandwidth of each spectrum by the wavenumbers at which the transmission of the filters are larger than 90% or the fluorescence or LED intensity is larger than 50% of its maximum. Thus the usable bandwidth of red fluorophore is 700 cm$^{-1}$, and that of green fluorophore is 450 cm$^{-1}$, and that of LED is 800 cm$^{-1}$. Figure 23 shows the emission spectrums of the fluorophores and the LED and the passbands of the dichroic beam splitters and laserline notch filters.

![Figure 23](image.png)

Figure 23 The emission spectrums of the fluorophores and the LED used in dual-color SSFM. The usable bandwidth of each spectrum is constrained by the passbands of the notch filters and the dichroic beam splitters shown as dashed and dotted lines.
Fluorescence interference spectra were generated by assuming a certain SiO$_2$ thickness and modeling the fluorophores as classical dipoles on a layered reflecting surface as described in 2.1.2. The interference spectrums were scaled by the emission spectral envelops of the fluorophores given by the manufacturer. Shot noise of the CCD camera following a Poisson distribution was also added to the spectrum. The simulated fluorescence interference spectrums were fit to the custom SSFM algorithm and the average axial fluorophore heights were determined to be the values with the Lease Square Fitting errors. Figure 24(a) shows the mean deviation of the fit fluorophore axial heights from initial set values in 10 simulations as a function of the number of oscillations in the spectral bandwidth. The number of oscillations within the limited spectral bandwidth is proportional to the SiO$_2$ layer thicknesses (Figure 24 (b)). The results show that at a SiO$_2$ layer thicknesses of 17 µm, we can obtain at least two periods of oscillations in each available spectral bandwidth of dual-color SSFM, corresponding to less than 0.5 nm deviation of fluorophore height determination. The same analysis applies to the LED interference spectrum. The relationship between number of interference oscillations and SiO$_2$ layer thicknesses remains the same for LED-RS. Since the available bandwidth of the LED spectrum is wider than that of the fluorophores, the determined SiO$_2$ layer thicknesses for dual-color SSFM is also applicable for LED-RS.

Based on the simulation, we used a substrate with 17.5 µm thick SiO$_2$ layer available in the lab, which allowed for at least two periods of oscillations within each spectral bandwidth used by dual-color SSFM and LED-RS.
Figure 24 Determine SiO₂ layer thickness from simulated analysis. (a) The simulated deviation from expected axial heights on SiO₂ layers of different thicknesses using different spectral bandwidths. The mean deviation is obtained from 10 simulated fitting results. (b) Number of interference oscillations within each spectral bandwidth on SSFM substrate of different SiO₂ thicknesses.

2.3.3 Measure nanometer scale steps

We measured nanometer-scale steps etched on the SiO₂ surface by single-color SSFM, dual-color SSFM and LED-RS. The results show that all modalities can determine the nanometer axial height differences with comparable accuracy both in air and in solution. Nanometer-scale steps were fabricated by patterning the chips with standard photolithographic techniques followed by etching in diluted buffered oxide etch (BOE). The etching was carried out for different time periods (5 min, 8 min, and 10 min) to create nanometer scales steps of different heights. The chips were then coated with MCP-2 and spotted at 500 µm pitch with dsDNA tagged with both Atto532 (green fluorophores) and Atto647N (red fluorophores). We scanned a line profile of the surface
of each chip in both dry (air) and wet conditions (Tris buffer: 10 mM Tris, 50 mM NaCl, pH 7.6). The step height was calculated by averaging the difference of the fluorophore heights to the SiO₂-Si interface on the left side from those on the right side. We measured the step height with both the red and green modality of SSFM. We also validated the results with LED-RS measurements. For both dry and wet conditions, the step heights measured by SSFM and LED-RS matched each other to within one nanometer (Figure 25, Figure 26).

Figure 25 Measurements of calibration chips with nanometer scale steps on the surface with dual-color SSFM and LED-RS. (a) An example of the measured surface line profile of a calibration chip. Eighteen spots of dsDNA labeled with both Atto532 and Atto647N pitched 500 µm on the substrate were measured both in dry (shown in this figure) and in solution (Figure 26). The average step height was calculated by subtracting the average fluorophore height of dsDNA spots on the left side from that of the right side of the chip. An illustration of the one-dimensional etching profile of the steps with labeled dsDNA molecules immobilized is shown at the bottom (not to scale). (b) The measured step heights of the calibration chips separately immersed in BOE for 5 min, 8 min and 10 min by dual-color SSFM and LED-RS. The results of each color modality of SSFM match with the step heights measured with LED-RS within one nanometer. The dsDNA immobilized on the chip etched for 10 min did not have green fluorophores.
Figure 26 Measurements of chips with nanometer scale steps on the surface with dual-color SSFM and LED-RS in solution. (a) Same SSFM chip with DNA spots was immersed in buffer. The polymer surface swells upon hydration, elevating and orienting the dsDNA. Thus, the red fluorophore at the distal end is higher than the green fluorophore at the proximal end of the dsDNA. (b) The measured step heights of the calibration chips separately immersed in BOE for 5 min, 8 min and 10 min by dual-color SSFM and LED-RS. The results of each color modality of SSFM match with the step heights measured with LED-RS within one nanometer. The dsDNA immobilized on the chip etched for 10 min did not have green fluorophores.

We also verified the etching rate with IRIS. We patterned IRIS substrates with an array of circular spots of 100 µm in diameter by photolithography. The etched spots were created in the same etching buffer for different length of time (buffered oxide etching (BOE), diluted 60:1 with DI water). The chips were cleaned with the same procedure described in the Materials and Methods section. An IRIS image was taken of each chip and the absolute height of each pixel was obtained from a fitting algorithm. The etched step height was calculated as the average of the height differences between the average absolute height inside the spot and that of an annuli area outside the spot for ten spots. We estimate an approximate etching rate of about 1.5 nm·min⁻¹ on the IRIS substrates, very close to the etching rate of 1.4 nm·min⁻¹ measured on SSFM substrates by LED-RS and SSFM.
Figure 27 Compare etching rates estimated by IRIS, dual-color SSFM, and LED-RS. (a) Average step heights of circular spots measured on IRIS substrates. Patterned IRIS chips were separately etched in 60:1 diluted BOE for 4 min, 6 min, and 8 min. The results are average values for the etched step heights of 10 spots measured by IRIS. (b) Step heights of SSFM calibration chips separately etched in BOE for 5 min, 8 min and 10 min. The results were measured by dual-color SSFM and LED-RS in both air (dry) and solution (wet). The results of each color of SSFM match with the step heights measured with LED-RS. The step height measurements of SSFM in solution have larger variance due to the dependence of dsDNA orientation on varied surface immobilization density.

2.3.4 Characterization of DNA immobilized on a 3-D polymeric surface

We used dual-color SSFM combined with LED-RS to evaluate the effects of surface density, buffer ionic strength, and DNA length on the conformation of dsDNA on the polymer surface immersed in buffer solution. First, we observed that the orientation of dsDNA to the surface positively correlated to its surface density on the polymer surface. We spotted an array of 60-bp fluorophore-labeled dsDNA at different concentrations (8 µM to 10 µM) to generate dsDNA spots of varied surface densities. We determined the surface densities of the dsDNA spots with LED-RS and their ensemble average orientations with dual-color SSFM. Figure 28(a) shows that dsDNA orientation has a
positive correlation with surface density. This correlation was expected because as the density of dsDNA layer increased, the distance between adjacent dsDNA molecules became smaller, resulting in stronger electrostatic repulsion and more steric hindrance between the dsDNA molecules. As illustrate in Figure 28(d), the random rotational freedom for each dsDNA molecule under thermal fluctuation decreased, and thus the ensemble average orientation of the dsDNA from the surface increased.

The measured orientations of dsDNA shown in Figure 28(a) were lower than the values predicted by trigonometric calculations using a rigid rod model of dsDNA on a 2-D surface, which are over 76° to the surface, much higher than our results. We attribute this discrepancy to the 3-D geometry of the polymer surface. The actual distance between the dsDNA molecules were presumably larger than that calculated based on 2-D surface geometry. The polymeric coating swells by about 7 to 20 nm upon hydration and its functional groups for immobilization are distributed in the axial dimension. As a result, the dsDNA molecules may penetrate the polymer scaffold and also be distributed axially depending on the polymer pore size and the dsDNA length. Hence, the axial penetration and distribution of the dsDNA molecules resulted in reduced ensemble average height differences between surface-distal and surface–proximal ends, and consequently reduced average orientations. Moreover, axial distribution of the immobilized dsDNA molecules potentially added more space for random rotation, which could also lead to lower calculated orientations compared to those based on a 2-D surface model.
Figure 28 Characterization of DNA molecules immobilized on a polymeric surface. Error bar shows standard deviation of 10 dsDNA spots. (a) The average orientation of each dsDNA spot positively correlates to its surface density. The red dashed line is a guide to the eye. (b) The orientation of dsDNA negatively correlates to buffer ionic strength. (c) Long dsDNA spots have larger average orientations than those of short dsDNA spots on the polymer surface. (d) Schematic illustration of the effect of surface density on the orientation of dsDNA. (e) Short dsDNA molecules may penetrate the polymeric surface and orient downwards, resulting in lower average orientation.

Next, we examined the effects of buffer ionic strength on the orientation of 60-bp dsDNA molecules by measuring the average orientation of 10 dsDNA spot replicates in buffers of different NaCl concentrations (Figure 28(b)). We observed that average orientation of dsDNA on the polymer surface negatively correlated to the buffer ionic strength, which agrees with Manning’s counterion condensation theory. The theory states that cations in the buffer can condense onto the negatively charged dsDNA backbones. Therefore, increased concentration of cations in the buffer shielded the electrostatic repulsive forces between the dsDNA molecules, which resulted in more random rotational freedom of the dsDNA and therefore lowered the average orientation.
Figure 28(b) shows that when the salt concentration was sufficiently high, the effect of the electrostatic repulsion between dsDNA molecules on their orientation became minimal. The orientation of 60-bp dsDNA molecule approached 33°, the statistical average orientation of dsDNA with full random rotation freedom under thermal fluctuations on a 2-D surface (See 2.2.2). However, the intermolecular distance calculated from the surface density was about 6.8 nm, where the DNA molecules should be sterically restricted from random free rotation. This observation again implied that under the same surface density used in the measurement, the dsDNA molecules had more free rotational space on the 3-D polymer surface than on a 2-D surface.

Finally, by measuring orientations of three dsDNA species of different lengths, we provide evidence that shorter dsDNA molecules have more freedom of random thermal rotations on the polymer surface. 20-bp, 40-bp, and 60-bp dsDNA spots were immobilized of the same surface density and measured in buffer of the same NaCl concentration. Figure 28(c) shows that the average orientation of longer dsDNA molecules was larger than that of shorter dsDNA molecules. This trend was anticipated since shorter dsDNA molecules, when immobilized at a similar density, have less steric hindrance and electrostatic repulsion between them than longer dsDNA molecules. However, the average orientation of 20-bp dsDNA was smaller than the statistical average (33°) of dsDNA freely rotating on a planar surface\textsuperscript{124}. The length of 20-bp dsDNA (\~6.5 nm) is comparable to the diameter of the polymeric pore size (\~5 nm) in hydrated state\textsuperscript{130}. Hence as illustrated in Figure 28(e) the 20-bp dsDNA may orient downwards in the 3-D polymer surface. The result indicates that short dsDNA of length
comparable to the polymeric pore size can have 3-D random rotational freedom when the polymer is hydrated and swells.

2.3.5 Real-time quantification of DNA hybridization and conformational change

We demonstrate the use of dual-color SSFM and LED-RS for simultaneous quantification of DNA conformational change and complementary strand binding during DNA surface hybridization. We immobilized 60-bp ssDNA modified with amine-C6 linker at the 3’ end (surface-proximal end) and the red fluorophore at the 5’ end (surface-distal end) on a polymer-functionalized chip. The chip was fixed in a customized flow cell with its surface immersed in buffer solution, which was driven by a peristaltic pump. After flowing only hybridization buffer for a baseline measurement, we introduced hybridization buffer containing complementary strands at a concentration of 500 nM at 45 °C. The surface-distal red fluorophore height and the thickness of the ssDNA layer were monitored at a time interval of 21 seconds before and during hybridization while the proximal green fluorophore height was measured after hybridization reached equilibrium.
Figure 29 Real-time detection and quantification of DNA conformation and hybridization using dual-color SSFM and LED-RS. (a) Schematic illustration of the DNA hybridization process on the polymer surface. Average fluorophore heights are indicated as dotted lines. (b) Hybridization efficiency calculated from the fraction of additional thicknesses of the DNA spot measured by LED-RS with a time resolution of 21 seconds. (c) Simultaneous detection of the average height increase of the red fluorophores at the surface-distal end of the ssDNA probes during hybridization. A first-order Langmuir kinetic equation was fit to both measurements and both observed effective rate constants around $1 \times 10^4 \text{M}^{-1}\text{S}^{-1}$.
The average height of the distal red fluorophores increased from 9 nm to about 15 nm and the average thickness of the DNA layer increased from 1.5 nm to 2.7 nm as ssDNA became hybridized into dsDNA till binding equilibrium was reached (Figure 29). The initial thickness of the ssDNA layer corresponded to an immobilization density of $3.8 \times 10^{12}$ cm$^{-2}$. The DNA hybridization efficiency was 0.8 at steady state obtained by calculating the fraction of additional thickness on the surface. Based on the studies of DNA hybridization regimes by Levicky and Melosh$^{115,135}$, the surface density under our experimental condition was in the pseudo-Langmuir regime ($2 \times 10^{12}$ cm$^{-2}$ to $6 \times 10^{12}$ cm$^{-2}$). We thus fit each dynamic hybridization data to the first-order Langmuir kinetic equation (Figure 29):

$$x = X_{eq} \left(1 - e^{-k_{eff}C_t t}\right)$$

For LED-RS measurement, $x$ denotes the hybridization efficiency, and $X_{eq}$ is the steady state hybridization efficiency. For SSFM measurement, $x$ denotes the average ensemble red fluorophore height at the surface-distal end of the ssDNA, and $X_{eq}$ is the average ensemble red fluorophore height at steady state. For both measurements, $k_{eff}$ is the effective rate constant, $C_t$ is the concentration of the target complementary strands, and $t$ is time. The $k_{eff}$ of DNA spot mass density change, indicated by the hybridization efficiency change, and DNA conformation change, indicated by the red fluorophore height change, were both around $1 \times 10^4$ M$^{-1}$S$^{-1}$.

The height increase at the surface-distal end of the ssDNA probes signified that the DNA conformation changed from highly flexible single-stranded coils to rigid double-stranded helices (Figure 29(a)). The average fluorophore height increase was
approximately 6 nm (Figure 29(b)), which agrees with the displacement range between 3 and 10 nm determined by a single-molecule micromechanical method. The complementary strands had green fluorophores tagged to the 5’ end, which resided on the surface-proximal end of the immobilized ssDNA molecules upon hybridization. After hybridization reached equilibrium, we replaced the hybridization solution in the flow cell with pure buffer and measured the height of the green fluorophores at about 2 to 3 nm. If we subtract the average height of the green fluorophores from that of the red fluorophores, we then obtained the height difference between the surface-distal and surface-proximal ends of the ssDNA molecules before hybridization, which was about 6 to 7 nm as shown in Figure 29(a). ssDNA is often modeled as random coils in solution, whose flexibility is characterized by a persistence length of about 1 to 4 nm. This larger than persistence length height difference implies that rather than a complete random coil, the immobilized ssDNA molecules had an extended conformation on the polymer surface. This extended conformation was also observed on short ssDNA tethered to a 2-D gold surface.

Figure 29(b) shows that the DNA hybridization efficiency at equilibrium was smaller than one. Many factors can affect the efficiency of DNA surface hybridization. We first attribute to the electrostatic barrier created by the DNA layer negative potential, which was enhanced along the incorporation of more negatively charged complementary strands. Moreover, hybridized dsDNA and the polymeric surface might also sterically hinder the incorporation of the complementary strands. The hybridization efficiency can
be improved by both lowering the surface density and increasing the buffer ionic strength.

2.3.6 Advantages, limitations, and future developments

We demonstrated the implementation of a dual-color SSFM combined with LED-RS for the nanoscale characterization of DNA conformation on a 3-D polymer surface. The system improved upon traditional single-color SSFM and white light reflectance spectrometry by using a substrate with a thicker SiO$_2$ layer (17.5 µm) and a LED illumination source. The newly designed substrate along with additional optical components enabled the use of multiple spectral bandwidths for precise axial localization of two different fluorophores and quantification of DNA surface density. We note that SSFM determines the axial location of fluorophores from the spectral oscillations, not from fluorescence intensity variations. Thus experimental conditions that can potentially cause fluorescence intensity variations do not affect the results.

Our results demonstrate that dual-color SSFM is advantageous over traditional single-color SSFM. For example, DNA surface density affects the orientation of surface-immobilized DNA. Previously, single-color SSFM determined orientations of surface-immobilized dsDNA by measuring the height differences between surface-distal and surface-proximal ends on separate DNA spots, which might have different immobilization densities and axial distributions. Also, dual-color SSFM increases the throughput of the height measurements under limited number of arrayed spots. Additionally, dual-color SSFM offers new assay designs. For instance, the location of
one fluorophore tagged to a DNA-binding protein can be compared to another labeled at a reference location on the DNA molecule to determine the protein binding site along the DNA sequence. Comparing to FRET\textsuperscript{51}, dual-color SSFM does not require a complex fluorescence normalization procedure in spectrum processing. Moreover, the axial heights measurable using dual-color SSFM ranges from subnanometer to more than 100 nanometers\textsuperscript{6} whereas the distance measurable using FRET is less than 10 nm\textsuperscript{51}. Comparing to acoustic wave sensors, the results of dual-color SSFM are directly interpreted as the geometric features of DNA conformation rather than mechanical properties of DNA films, such as modulus and viscosity\textsuperscript{139}. Moreover, LED-RS provides orthogonal information of the DNA surface density, which is often coupled with the signal generated by DNA conformational change for mechanical and optical resonance based sensors. Notably, the nanoscale DNA conformational change from flexible coils to rigid double helices was simultaneously monitored during the addition of complimentary sequence to ssDNA probes.

Our results of the orientation and surface density of dsDNA molecules indicated that dsDNA immobilized on a 3-D polymeric coating have more random rotation freedom under thermal fluctuations than those immobilized on a 2-D surface in hydrated environment. A precise model describing the relationship between the average orientation and surface density is unattainable at the moment due to lack of the exact axial distribution of dsDNA molecules using average ensemble measurements. Yet our results suggested that this larger degree of free random thermal rotation of the DNA molecules, and possibly of protein molecules, could cause the previously described high
performance of the polymeric coating, such as high surface density, bioactivity retaining, and easy accessibility of the immobilized molecular probes\textsuperscript{128,131}.

Although dual-color SSFM and LED-RS allowed for real-time simultaneous quantification of molecular binding and conformational change, a few limitations need to be overcome in future developments. First, other fluorescence labels with narrow spectrums, such as Quantum Dots (QDs) should be used to replace the fluorophores for real-time measurements. In this work, the green and red fluorescence spectrums were measured separately. The tail of the green fluorophore emission spectrum overlapped with that of the red fluorophore, which can affect the accuracy of red fluorophore localization if measured at the same time. Therefore, during DNA hybridization, we only measured the green fluorophore height after hybridization reached steady state in pure buffer without complementary sequences. Moreover, the fluorophores photobleach after limited number of measurements, limiting the time resolution for monitoring real-time binding processes. Also, FRET occurred between green fluorophores and the red fluorophores, and expedited the photobleaching of red fluorophores. Second, the flow cell employed for real-time hybridization measurements operated in a diffusion-limited regime\textsuperscript{140}. Thus, the observed effective rate constants of DNA binding and conformational change were indistinguishable, both of which were limited by the transportation rate of the complementary strands to the surface-immobilized ssDNA.

In summary, we have demonstrated the development of a dual-color SSFM combined with independent biomolecular sensing for the characterization of nanoscale conformation of DNA on a 3-D polymer surface. A newly designed substrate with a thick
SiO₂ layer (17.5 µm) allowed for the use of multiple spectral bandwidths for precise axial localization of two different fluorophores and quantification of biomolecule surface density. Using dual-color SSFM, we have shown that the conformation of surface-immobilized dsDNA on a 3-D polymeric surface was subject to surface density, buffer ionic strength, and dsDNA length. Furthermore, dsDNA immobilized on a polymeric surface possess more conformational freedom than those on a 2-D surface, providing additional evidence for the high density and easy accessibility of biomolecules on 3-D polymeric coating in molecular binding assays. Particularly, we have shown simultaneous and independent real-time monitoring of DNA conformational change and complementary strand binding during DNA surface hybridization. The overall optical setup is simple and compatible with regular fluorescence microscopes. All measurements were performed in a flow cell on DNA microarrays, compatible with multiplexed parallel molecular binding assays.
CHAPTER 3

DUAL-SPECTRAL IMAGING FOR THE STUDY OF PROTEIN-DNA INTERACTIONS

The precise characterization of the nanoscale conformation of surface-immobilized DNA established the foundation for detecting DNA conformational changes on the SSFM platform. This chapter introduces the approach of using quantified information of both DNA conformational change and molecular surface density for the study of protein-DNA interactions. First, to perform multiplexed and high-throughput assays, we demonstrate a one-dimensional spectral imaging configuration of SSFM and LED-RS that allows simultaneous measurement of multiple DNA spots. Then, we present the experiment design for the study of specific and nonspecific IHF-DNA binding. Finally, we show results of IHF-DNA binding at equilibrium and propose a model for further analysis of IHF-DNA interactions on the surface.

3.1 One-dimensional spectral imaging configuration

The microscope excitation and collection configuration of traditional SSFM and dual-color SSFM only permits measurement of one DNA spot for each acquired spectrum. Thus, the time for scanning of the entire DNA microarray linearly increases with the number of DNA spots. Also, this spot-by-spot scanning configuration limits the time resolution for real-time assays. To improve the throughput of SSFM and LED-RS, a one-dimensional (1-D) spectral imaging configuration was implemented, which enabled simultaneous acquisition of the interference spectrums of multiple DNA spots.

The 1-D spectral imaging configuration of SSFM was achieved by shaping
collimated circular laser beam into a line using a cylindrical lens before the objective (Figure 30). Thus, instead of focusing collimated circular laser beam to a diffraction-limited point, the objective focused a diffraction-limited laser line on the surface spanning multiple DNA spots. Previously, the use of a thick SiO$_2$ layer in the SSFM substrate allowed us to use two spectral bandwidths of fluorescence emission to determine the heights of two different fluorophore layers or to determine the height of one fluorophore layer and the bio-layer thickness. We continued to take advantage of the thick SiO$_2$ layer but only used the dual-spectral detection fashion that measures red fluorescence emission and LED interference spectra for the study of protein-DNA interactions. One rational is that the red fluorescence emission spectrum does not overlap with the LED emission spectrum so that it is feasible to acquire the two spectra from one measurement. Moreover, as discussed in the limitations of dual-color SSFM, FRET occurred between the green fluorophores and the red fluorophores, which consequently photobleached after a few measurements, faster than traditional SSFM. To characterize protein-DNA interaction, we need to make many measurements of fluorescence interference spectra at equilibrium and in real-time. Thus, green fluorophore and its interference spectrum were not used in the following work. In addition, in the 1-D spectral imaging configuration the red fluorophores do not photobleach as fast due to the reduced laser intensity resulted from linear expansion of a diffraction-limited laser spot on the surface. We can perform over a hundred of fluorescence spectral acquisitions before photobleaching affects the signal-to-noise ratio (SNR) of the interference spectra. However, to compensate for the reduced excitation intensity and its resulting reduced
fluorescence emission intensity, the exposure time for integrating fluorescence signal was increased.

A similar idea was implemented for LED-RS. Instead of imaging a circular pinhole placed in front of the LED to the center of one DNA spot, a vertical slit was imaged to span a line of DNA spots. The image of the slit and the laser line for fluorescence excitation were aligned and superimposed at the same location on the surface. Hence, using this 1-D spectral imaging configuration we can perform multiplexed detection of the conformational changes and molecular surface densities on multiple DNA spots of different sequences.

Figure 30 1-D spectral imaging configuration of SSFM. Collimated circular laser beam is focused by a cylindrical lens to a line at the back focal plane of the objective, which then performs Fourier Transform and focuses the laser line to a diffraction-limited lase line in the other dimension on the SSFM substrate. The SSFM substrate is placed in a customized flow cell.
3.2 Experimental approach

3.2.1 Optical setup

Fluorescence measurements (Figure 31). In the excitation path, helium-neon laser with peak intensity at 633 nm (Melles Griot, Carlsbad, CA) was collimated and expanded with two achromatic lenses with respective focal lengths of 30 mm and 150 mm (Thorlabs Inc., Newtown, NJ). The on-and-off of the laser was controlled by a mechanical shutter that opens and closes via a transit voltage signal (+5V or 0V). A dual-edge dichroic beam splitter (545/650 nm BrightLine, Semrock, Inc. NY) that specifically reflects the wavelength of the laser and transmits >90% of the fluorescence emission of Atto647N (red fluorophore) directed the collimated laser beam to the objective. Before the beam splitter reflected the collimated beam, a cylindrical lens focused the circular laser beam to a line at the back focal plane of the objective. The objective (Nikon 5x, numerical aperture (NA) = 0.13) then performed Fourier Transform of the laser line beam and focused the laser line beam to a diffraction-limited line in the other dimension on the sensor surface. In the collection path, the excited line of fluorescence emission on the surface was imaged onto the entrance slit of a spectrometer (SP-2150i, Princeton Instruments, Trenton, NJ) by using the objective, a beam splitter (92% reflection, 8% transmission), and a tube lens. A two-dimensional (2-D) CCD camera (PIXIS 256E, dimensions 26 mm x 6.7 mm, 1024 x 256, 26 µm x 26 µm pixels, Princeton Instruments, Trenton, NJ) was connected to the spectrometer to collect the line image of the fluorescence emission in the vertical dimension and the interference spectra
corresponding to each DNA spot position on the line image in the horizontal dimension. The 1-D spectral image acquired by the CCD camera of the fluorescence emission of multiple DNA spots along the line on the surface was then saved for spectrum analysis.

**LED-RS measurements** (Figure 31). In the illumination path, a yellow LED (M565L2, Thorlabs, Newton, NJ) was used as the illumination source. A rectangular slit 100 µm wide and 3 mm tall (S100R, Thorlabs, Newton, NJ) was placed in front of the LED on a rotation mount (CRM1, Thorlabs, Newton, NJ). An image of the slit was focused and superimposed on the focused laser line on the surface using a 50:50 beam splitter (Thorlabs, Newton, NJ) and the objective. A relay lens on a x, y translation mount (CXY1Q, Thorlabs, Newton, NJ) in front of the rectangular slit was placed at a position so that the LED slit image had the same focus with the laser line. Precise adjustment of the horizontal position of the slit image and its alignment with the direction of the laser line was made by tuning the x and y position of the 2-D translation mount and the rotation mount while examining their positions on a CMOS camera (ORCA-Flash2.8 C11440-10C Digital CMOS camera, Hamamatsu, Japan) at a height above the objective. Focusing of the laser line and the LED slit image on sensor surface was also achieved by observing their images on the CMOS camera and adjusting the stage longitudinally. Since the power of LED was far more sufficient, the reflection/transmission passband of beam splitters were selected to maximize the fluorescence collection efficiency. In the collection path, the LED slit image was then imaged in the same fashion as the excited fluorescence emission line onto the spectrometer entrance slit. The 1-D spectral image
acquired by the CCD camera of the LED vertical slit image on the surface was also saved for spectrum analysis.

![Diagram of optical setup](image)

Figure 31 Schematic illustration of the optical setup of the 1-D spectral imaging configuration of SSFM and LED-RS. Components are not to scale.

SSFM chip was placed in a custom flow cell fixed on a 4-axis stage that consists of an x, y translation micro-stage (Mad City Labs, Madison, WI), a manual z translation stage, and a 360° high-precision rotation mount (PR01, Thorlabs, Newton, NJ). The z translation was for fine focusing of the laser and LED slit image, the rotation mount was for SSFM chip alignment in the x, y direction, and the x, y translation was for DNA
microarray scanning. To perform automatic spectrum measurement of the DNA microarray, the in-house written MATLAB software application controlled and coordinated the spectrometer acquisition, the micro-stage movement, the laser shutter, and the LED.

The number of DNA spots that can be measured at the same time depends on the 1-D imaging field of view of the spectrometer and the size and pitch of the DNA spots. The dimensions of the CCD camera of the spectrometer and the magnification of the objective determine the field of view. For instance, in the current optical setup the dimension of the CCD camera of the spectrometer for the 1-D imaging is 6.7 mm. The magnification of the objective is 5x. Thus the field of view is 6.7/5=1.34 mm. The typical diameter of DNA spots on the polymer surface is about 120 µm and the pitch of the DNA microarray grid was selected to be 250 µm. Thus about 6 DNA spots were within the imaging field of view. A tube lens of a focal length of 180 mm instead of the focal length of the spectrometer, 150 mm, was placed in front of the spectrometer entrance slit. Thus the magnification was reduced to 5×150/180 = 4.17. Then imaging field of view was 6.7 mm/3.75=1.61 mm. Thus a total of 7 DNA spots along the imaging dimension were simultaneously measured within the field of view under current configuration of the optical setup. Using a larger CCD sensor array and an objective of smaller magnification, or performing 2-D spectral imaging can further increase the number of DNA spots that can be imaged simultaneously.
3.2.2 Spectrum analysis

Figure 32 1-D spectral imaging configuration of SSFM and LED-RS for parallel quantification of DNA conformation and molecular surface density. (a) The fluorescence image of an example microarrayed DNA spots labeled with red fluorophores on the polymer-functionalized SSFM substrate. The GenePix® 4000B Microarray Scanner (Molecular Devices, LLC, Sunnyvale, CA) acquired the image. The orange line indicates that both LED slit image and the laser line focus on the center of a line of DNA spots. (b) The 1-D spectral image of the LED slit image on the surface. (c) The 1-D spectral image of the line of excited fluorescence emission on the surface. (d) A sample LED interference spectrum integrated over 10 pixels at the center of one DNA spot, indicated by the yellow horizontal line in (b). (e) A sample fluorescence interference spectrum integrated over 10 pixels at the center of one DNA spot, indicated by the red horizontal line in (c).

To find the centers of DNA spots, as shown in Figure 32, a series of fluorescence spectral images were acquired by scanning from one edge of the DNA spots to the other. The middle position between the edges was determined to be the center of that line of
DNA spots. The center positions of the rest lines of DNA spot in the microarray were calculated from the determined center position and the microarray pitch distance.

The line of excited fluorescence emission and the LED slit image on the surface were both imaged onto the longitudinal dimension of the spectrometer CCD camera (Figure 32). The pixel number on the longitudinal axis matches to a unique position along the imaged line on the surface. The number of pixels spanned by each DNA spot can be approximated by $120 \mu m \times 4.17 \text{ (magnification)} / 26 \mu m \text{ (CCD pixel size)} = 17$ pixels. The fluorescence interference spectrum was integrated over 10 pixels at the center of each DNA spot to obtain sufficient SNR for fitting. Since the power of the LED is sufficient, the LED interference spectrum for each pixel along the line of the slit image was fit to obtain a thickness measurement. A moving average filter using a window length of 5 pixels was used to reduce the random noise and reveal thickness step changes. Background thickness measurements for the determination of DNA spot thicknesses were made separately at the middle between the lines of DNA spots. The DNA spot positions were visible along the longitudinal axis of the fluorescence spectral image, but were indistinguishable from nearby background in LED spectral image. However, we can assume the DNA spot position in the LED spectral image is the same as that in the fluorescence spectral image. This is because the LED slit was carefully aligned and superimposed with the laser line on the surface. Finally, the LED and integrated fluorescence interference spectra along the horizontal axis in the spectral image of multiple DNA spots were fit with custom algorithms previously described in Chapter 2.
3.2.3 Materials

Single-stranded 60-bp DNA was purchased from IBA GmbH (Goettingen, Germany), and single-stranded 20-bp and 40-bp DNA were purchased from IDT, Inc. (Coralville, IA). All DNA strands were purified with high-performance liquid chromatography (HPLC) after synthesis. For DNA surface immobilization, the first ssDNA were modified with amine-C6 linkers on the 5’ end while its complementary strand was modified with Atto647N on the 5’ or 3’ end of the complementary strand, corresponding to the surface-distal end or surface-proximal end on hybridized dsDNA immobilized on the polymer functionalized surface. We selected Atto647N fluorophore to label DNA strands because of its superior photostability than other organic dyes. The fluorescence emission spectrum of Atto647N is also distinctly separated from the yellow LED emission spectrum. The DNA sequences and their names used in the study of IHF-DNA binding are listed in Table 3. The dsDNA molecules were hybridized at 30 µM in 150 mM sodium phosphate buffer, made with DI water, sodium phosphate monobasic, and sodium phosphate bibasic powder (Sigma-Aldrich MO).
Table 3 DNA sequences and nomenclature used for the study of IHF-DNA interactions.

<table>
<thead>
<tr>
<th>Name</th>
<th>DNA sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>H(23)</td>
<td>5’-GAT AGG GCC AAA AAA GCA TTG CTT ATC AAT TTG TTG CAC CTG ACC GAT GAG CTG TTA GAA-3’</td>
</tr>
<tr>
<td>H(34)</td>
<td>5’-GCT GTT AGA AGA TAG GCC CAA AAA AGC ATT GCT TAT CAA TTT GTT GCA CCT GAC CGA TGA-3’</td>
</tr>
<tr>
<td>H(36)</td>
<td>5’-GAG CTG TTA GAA GAT AGG GCC AAA AAA GCA TTG CTT ATC AAT TTG TTG CAC CTG ACC GAT-3’</td>
</tr>
<tr>
<td>H(39)</td>
<td>5’-GCT CTG AGA AGA CAG TGA CCG GCC AAA AAA GCA TTG CTT ATC AAT TTG TTG CAC CTG ACC-3’</td>
</tr>
<tr>
<td>H(34)A</td>
<td>5’-GCT GTT AGA AGA TAG GCC CAA AAA AGC ATT GCT TAT CAA TTT GTA GCA CCT GAC CGA TGA-3’</td>
</tr>
<tr>
<td>Control</td>
<td>5’-CAA CAG CCC CGC TCT AGT TTG GGT TCA TAT ATC GGG ACA GGC CTC GGA ATC AAG TGC ATG-3’</td>
</tr>
<tr>
<td>20-bp DNA</td>
<td>5’-GCT GTT AGA AGA TAG GCC CA-3’</td>
</tr>
<tr>
<td>40-bp DNA</td>
<td>5’-ATC TGA ACC CAC CGC TAT TCC ATG CAC TTG ATT CCG AGG C-3’</td>
</tr>
</tbody>
</table>

Measurements of DNA conformation and IHF binding were made in 10 mM Tris, 50 mM or 150 mM NaCl, and 10 mg/mL bovine serum albumin (BSA) buffer prepared with Tris buffered saline tablets, NaCl powder, and BSA powder (Sigma-Aldrich, MO) unless specified. All buffer solutions were prepared with deionized water (DI water) filtered by Barnstead Nanopure Diamond (18.2 MΩ·cm\(^{-1}\) resistance, Thermo Scientific, Waltham, MA). SSFM substrates were prepared from silicon wafers with 17.5 µm thick thermally grown oxide layer (Silicon Valley Microelectronics, Santa Clara, CA), which were cut into 15 mm × 15 mm square chips. Before use, chips were rinsed with acetone, methanol, and subjected to sonication in acetone for 5 min, and oxygen plasma ashing (M4LTM, PVA TePla America, Corona, CA) for 10 min. DNA microarray spotting and surface functionalization of the SSFM substrate with a polymeric coating were prepared in the same fashion as described in Chapter 2.

Wild-type integration host factor (IHF) protein of *Escherichia coli* was a gift from Prof. Peter Dröge at Nanyang Technological University. IHF was stored in 10 mM Tris-HCl buffer at pH 7.6, containing 50 mM NaCl and 30% glycerol. IHF stock solution was
immediately stored in −30° freezer after shipping from Singapore to Boston University. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) assay was performed on 4-20% gradient Polyacrylamide Mini-PROTEAN® TGX™ Pre cast Gels (Bio-Rad, Hercules, CA) regularly to check the integrity and purity of the protein before each experiment (Figure 33). Two bands corresponding to the two subunits of IHF, each about 11 kDa, were visibly distinguishable. The concentration of the stock IHF solution was determined to be 83.2 µM by conducting a Bradford protein assay using NanoDrop 2000c UV-Vis Spectrophotometer (Thermal Scientific, Wilmington, DE) (Figure 34).

<table>
<thead>
<tr>
<th>wtIHF</th>
<th>scIHF</th>
<th>wtIHF</th>
<th>scIHF</th>
<th>Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>250 kD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>150 kD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75 kD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50 kD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25 kD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20 kD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15 kD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10 kD</td>
</tr>
</tbody>
</table>

Figure 33 Denaturing SDS-PAGE assay to check the integrity of IHF protein after shipment. Wild-type IHF (wtIHF) contains two subunits or amino acid chains and appears as two bands. Engineered single-chain IHF (scIHF) contains one amino acid chain and appears as one band.
Figure 34: Determine the concentration of IHF stock solution by Bradford assay. Two dilutions of standard IgG solution were prepared. One IgG dilution was used to generate the standard calibration curve, and the other was measured as a control sample in the same way as IHF stock sample dilutions.

3.2.4 Experiment design

We demonstrate the high-throughput multiplexed capability of DNA microarrays on the SSFM platform by performing measurements of IHF binding to hundreds of DNA spots in different conditions. The basic detection mechanism of using SSFM combined with LED-RS to detect DNA bending induced by IHF specific binding is shown in Figure 35. Specific binding of IHF to surface-immobilized dsDNA results in a height change of the fluorophores at the surface-distal ends, which causes a shift of the interference oscillations in the fluorescence emission spectrum. SSFM then matches this shift to an
average height change of the surface-distal fluorophores on the DNA. Similarly, IHF binding to surface-immobilized dsDNA, either specific or nonspecific, results in an accumulation of biomass on the surface and a change of the DNA spot thickness (Figure 35). The DNA spot thickness change in turn causes a shift of the interference fringes of the LED spectrum. LED-RS then matches this shift to a spot thickness change, which is converted to the molecular surface density of bound IHF as described previously.

![Graph showing fluorescence intensity vs. wavelength](image)

Figure 35 Schematic illustration of the detection mechanisms of DNA conformational change and IHF-DNA binding and using SSFM combined with LED-RS.
**DNA microarray design.** dsDNA molecules of different sequences and surface densities were spotted on the same chip with 10 replicates of each condition. First, we wanted to see whether we can distinguish IHF binding site locations on short dsDNA sequences. We designed 60-bp dsDNA sequences containing the 34-bp H’ site of bacteriophage λ, one of the best-characterized IHF binding sites. We designated the 34-bp H’ site to start at four different nucleotide positions in four of the 60-bp DNA sequences, which were named by adding the starting nucleotide position in parenthesis (Table 3). Given that 60-bp dsDNA is end-tethered at the first nucleotide to the surface and oriented, a smaller starting nucleotide position of the H’ site indicates a shorter distance of the H’ binding site to the surface. We speculated that larger fluorophore height change caused by IHF-induced DNA bending would be detected as the H’ site shifted closer to the surface (Figure 36).

![Diagram](image.png)

**Figure 36** The idea of distinguishing IHF binding site location on 60-bp dsDNA sequences from measuring fluorophore height changes by SSFM. When dsDNA bend at the same angle, lower axial binding site location results in a larger fluorophore height change measurement.
A 60-bp DNA Control sequence that does not have the H’ site and a 60-bp DNA sequence possessing the H’ site with a single nucleotide mutation (H’(34)A) were spotted together with the IHF binding sequences. The Control sequence was arbitrarily designed and optimized with Oligo Analyzer (Integrated DNA Technologies, Inc., Coralville, IA) to have 50% GC contents and minimized secondary structures. The mutated sequence named as H’(34)A is the same sequence as H’(34) except that a T was replaced by A of the consensus TTR element (Table 3). Designed in the same fashion as 60-bp Control sequence, 20-bp and 40-bp dsDNA sequences that do not include H’ site were also spotted to examine the nonspecific binding of IHF to DNA. The 20-bp and 40-bp dsDNA were not labeled with fluorophores and only LED-RS measurement was performed on these DNA spots.

Finally, we wanted to investigate the effect of DNA surface density on DNA conformational change and IHF binding by comparing results of the DNA spot of same sequence but different surface densities. Thus, all DNA sequences were spotted in three different concentrations, 5 µM, 7.5 µM, and 15 µM, to prepare dsDNA spots of each sequence at three surface density groups: low, medium and high.

**Quantification of IHF binding at different concentrations at equilibrium.** To demonstrate the detection of DNA conformational change and IHF binding, we performed equilibrium measurements of IHF binding to all DNA spots at a series of IHF concentrations ranging from 0 to 40 nM. The DNA microarray spots were consecutively immersed in binding buffer solutions containing IHF from low to high concentrations, which were constantly driven through the flow cell at a flow rate of 500 µL/min. SSFM
and LED-RS measurements were taken for a few times until equilibrium was reached, and longer incubation time was required for lower IHF concentrations. Equilibrium binding isotherms of both fluorophore height change and IHF binding were obtained by averaging ten DNA spot replicates of each condition. The equilibrium measurements were carried out in binding buffer solutions of two different NaCl concentrations, 50 mM and 150 mM. The accumulation of IHF binding was converted to a molecular binding ratio of IHF to DNA using molecular surface densities quantified by LED-RS. The equilibrium measurements of DNA spots of different sequences and surface densities in different NaCl concentrations were analyzed and compared.

**Dynamic IHF dissociation measurements.** Dissociation of bound IHF at equilibrium was monitored to study the salt-dependence of specific and nonspecific binding of IHF to DNA. Measurements of fluorophore height change and IHF dissociating from DNA spots were performed in buffer that contained unlabeled competitor dsDNA molecules of the sequence H’(34). The dissociation time constants of IHF binding sequence and Control sequence at different NaCl concentrations were analyzed and compared.

### 3.3 Quantification of IHF-DNA interaction

#### 3.3.1 Detection of IHF-DNA interaction at equilibrium

We measured the average height change of surface-distal end fluorophores and the IHF accumulation of 10 spot replicates of each DNA sequence at equilibrium for a range of IHF concentrations and three DNA surface density groups in two buffers
containing different salt concentrations. The accumulation of IHF was converted to the IHF to DNA molecular binding ratio using the added spot thickness and IHF molecular weight. To obtain the dissociation constant $K_d$ of IHF binding to DNA, we used nonlinear LSF to fit each equilibrium binding isotherm to an adapted Langmuir binding isotherm, which describes the relationship between the fraction of occupied ligand-binding sites to unbound ligand concentration. In the case of fluorophore height change measurement by SSFM, the equation for the adapted Langmuir binding isotherm is:

$$ \left\langle \Delta h \right\rangle = \left\langle \Delta h_{\text{max}} \right\rangle \frac{x}{K_d + x}, $$

where $\left\langle \Delta h \right\rangle$ is the measured average fluorophore height change of the DNA surface-distal end, $x$ is the IHF concentration, $K_d$ is the dissociation constant of IHF to the DNA sequence, and $\left\langle \Delta h_{\text{max}} \right\rangle$ is the maximum average fluorophore height change at equilibrium. In the case of IHF to DNA molecular binding ratio, the equation is:

$$ \left\langle R \right\rangle = \left\langle R_{\text{max}} \right\rangle \frac{x}{K_d + x}, $$

where $\left\langle R \right\rangle$ is the measured average IHF to DNA binding ratio, $x$ is the IHF concentration, $K_d$ is the dissociation constant of IHF to the DNA sequence, and $\left\langle R_{\text{max}} \right\rangle$ is the maximum IHF to DNA ratio at equilibrium. The binding isotherms were measured in Tris buffer solutions at pH 7.6 of two NaCl concentrations. An example of the equilibrium binding isotherms of DNA sequence H’(39) is shown in Figure 37.
Figure 37 Equilibrium binding isotherms of IHF to DNA sequence H'(39) measured by SSFM and LED-RS. (a) Equilibrium measurement of average fluorophore height change of 10 DNA spot replicates at different IHF concentrations in buffers of two NaCl concentrations. (b) Parallel equilibrium measurement of IHF to DNA molecular binding ratio of the same DNA spot replicates at different IHF concentrations in buffers of two NaCl concentrations.

Since the fluorophore height changes measured by SSFM and IHF to DNA binding ratios measured by LED-RS were both measurements of the same interaction in one binding assay, it was expected that the equilibrium isotherms should the same in principle. However, it is evident that the two equilibrium isotherms are different. By performing nonlinear LSF of the equilibrium isotherms, we obtained different $K_d$ values of the two measurements in both salt concentrations. The $K_d$ from fluorophore height change measurements was 0.7 nM in buffer containing 50 mM NaCl and 1.9 nM in buffer containing 150 mM NaCl. In contrast, the $K_d$ from IHF to DNA binding ratio measurements was 6.7 nM in buffer containing 50 mM NaCl and 17.6 nM in buffer containing 150 mM NaCl. The $K_d$ values from LED-RS measurements were about 10 times of those from SSFM measurements. This fitting discrepancy of the fluorophore
height change measurement and IHF to DNA binding ratio measurement was consistent across different DNA sequences containing the H’ IHF binding site at different densities.

We present the equilibrium isotherms of fluorophore height change and IHF to DNA binding ratio measured of all DNA sequences at different densities in the figures below and summarize the fitting results in the following tables (Figure 38, Figure 39, Figure 40, Figure 41, Figure 42, Figure 43, Table 4, Table 5, Table 6).
Figure 38 Equilibrium binding isotherms of fluorophore height changes of different DNA sequences immobilized at low surface density measured by SSFM.
Figure 39 Equilibrium binding isotherms of fluorophore height changes of different DNA sequences immobilized at medium surface density measured by SSFM.
Figure 40 Equilibrium binding isotherms of fluorophore height changes of different DNA sequences immobilized at high surface density measured by SSFM.
Figure 41 Equilibrium binding isotherms of IHF to DNA binding ratio of different DNA sequences immobilized at low surface density measured by LED-RS.
Figure 42 Equilibrium binding isotherms of IHF to DNA binding ratio of different DNA sequences immobilized at medium surface density measured by LED-RS.
Figure 43 Equilibrium binding isotherms of IHF to DNA binding ratio of different DNA sequences immobilized at high surface density measured by LED-RS.
Table 4 Nonlinear LSF fitting results of equilibrium binding isotherms of fluorophore height change and IHF to DNA binding ratio of different DNA sequences at low surface density in two buffers of different salt concentrations.

DNA surface density $\sim (1.73 \pm 0.24) \times 10^{12}/\text{cm}^2$

**LED-RS:**

<table>
<thead>
<tr>
<th>$K_d$ (nm)</th>
<th>Control</th>
<th>$H'(23)$</th>
<th>$H'(34)$</th>
<th>$H'(36)$</th>
<th>$H'(39)$</th>
<th>$H'(34)_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl 150mM</td>
<td>124.6±262</td>
<td>10.8±4.8</td>
<td>8.7±5.8</td>
<td>4.7±1.6</td>
<td>17.6±9.8</td>
<td>8.5±10</td>
</tr>
<tr>
<td>NaCl 50mM</td>
<td>22.2±6.4</td>
<td>10.6±2.4</td>
<td>7.1±2.0</td>
<td>5.4±1.3</td>
<td>6.7±1.9</td>
<td>4.5±1.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$R_{max}$</th>
<th>Control</th>
<th>$H'(23)$</th>
<th>$H'(34)$</th>
<th>$H'(36)$</th>
<th>$H'(39)$</th>
<th>$H'(34)_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl 150mM</td>
<td>1.5±2.5</td>
<td>0.9±0.1</td>
<td>0.6±0.1</td>
<td>0.7±0.1</td>
<td>1.1±0.3</td>
<td>0.5±0.2</td>
</tr>
<tr>
<td>NaCl 50mM</td>
<td>2.8±0.5</td>
<td>2.5±0.2</td>
<td>1.4±0.2</td>
<td>1.8±0.2</td>
<td>2.0±0.2</td>
<td>1.4±0.2</td>
</tr>
</tbody>
</table>

**SSFM:** (*end-time measurements@40nM*)

<table>
<thead>
<tr>
<th>$K_d$ (nm)</th>
<th>Control</th>
<th>$H'(23)$</th>
<th>$H'(34)$</th>
<th>$H'(36)$</th>
<th>$H'(39)$</th>
<th>$H'(34)_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl 150mM</td>
<td>Cannot fit</td>
<td>3.1±1.5</td>
<td>3.8±1.1</td>
<td>2.2±0.3</td>
<td>1.8±0.1</td>
<td>Cannot fit</td>
</tr>
<tr>
<td>NaCl 50mM</td>
<td>Cannot fit</td>
<td>1.1±0.1</td>
<td>0.7±0.1</td>
<td>0.6±0.1</td>
<td>0.7±0.1</td>
<td>1.8±0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\Delta h_{max}$ (nm)</th>
<th>Control</th>
<th>$H'(23)$</th>
<th>$H'(34)$</th>
<th>$H'(36)$</th>
<th>$H'(39)$</th>
<th>$H'(34)_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl 150mM</td>
<td>- 1.4nm</td>
<td>0.9±0.2</td>
<td>3.4±0.4</td>
<td>3.9±0.2</td>
<td>3.7±0.1</td>
<td>Cannot fit</td>
</tr>
<tr>
<td>NaCl 50mM</td>
<td>+ 0.9nm</td>
<td>3.2±0.1</td>
<td>6.3±0.1</td>
<td>5.5±0.1</td>
<td>4.9±0.1</td>
<td>4.4±0.1</td>
</tr>
</tbody>
</table>
Table 5 Nonlinear LSF fitting results of equilibrium binding isotherms of fluorophore height change and IHF to DNA binding ratio of different DNA sequences at medium surface density in two buffers of different salt concentrations.

**DNA surface density ~ (2.52 ± 0.34)x10^{12}/cm^2**

**LED-RS:**

<table>
<thead>
<tr>
<th>$K_d$ (nm)</th>
<th>Control</th>
<th>H'(23)</th>
<th>H'(34)</th>
<th>H'(36)</th>
<th>H'(39)</th>
<th>H'(34)A</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl 150mM</td>
<td>104.9±68.6</td>
<td>23.0±11.7</td>
<td>16.6±3.6</td>
<td>5.4±1.9</td>
<td>7.8±2.5</td>
<td>63.0±32.2</td>
</tr>
<tr>
<td>NaCl 50mM</td>
<td>11.9±3.2</td>
<td>7.8±2.1</td>
<td>9.0±1.5</td>
<td>6.5±1.7</td>
<td>6.2±1.6</td>
<td>8.6±2.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$R_{max}$</th>
<th>Control</th>
<th>H'(23)</th>
<th>H'(34)</th>
<th>H'(36)</th>
<th>H'(39)</th>
<th>H'(34)A</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl 150mM</td>
<td>1.7±0.9</td>
<td>1.1±0.3</td>
<td>1.1±0.1</td>
<td>0.6±0.1</td>
<td>0.7±0.1</td>
<td>1.8±0.6</td>
</tr>
<tr>
<td>NaCl 50mM</td>
<td>2.1±0.3</td>
<td>2.0±0.2</td>
<td>2.5±0.2</td>
<td>1.6±0.2</td>
<td>1.7±0.2</td>
<td>2.2±0.3</td>
</tr>
</tbody>
</table>

**SSFM: (end-time measurements@40nM)**

<table>
<thead>
<tr>
<th>$K_d$ (nm)</th>
<th>Control</th>
<th>H'(23)</th>
<th>H'(34)</th>
<th>H'(36)</th>
<th>H'(39)</th>
<th>H'(34)A</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl 150mM</td>
<td>Cannot fit</td>
<td>5.3±1.8</td>
<td>7.5±1.7</td>
<td>2.5±0.3</td>
<td>1.9±0.2</td>
<td>Cannot fit</td>
</tr>
<tr>
<td>NaCl 50mM</td>
<td>Cannot fit</td>
<td>1.2±0.1</td>
<td>0.9±0.2</td>
<td>0.8±0.2</td>
<td>0.8±0.1</td>
<td>2.0±0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\Delta h_{max}$ (nm)</th>
<th>Control*</th>
<th>H'(23)</th>
<th>H'(34)</th>
<th>H'(36)</th>
<th>H'(39)</th>
<th>H'(34)A</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl 150mM</td>
<td>− 1.0nm</td>
<td>1.1±0.2</td>
<td>3.9±0.4</td>
<td>4.0±0.2</td>
<td>3.5±0.2</td>
<td>Cannot fit</td>
</tr>
<tr>
<td>NaCl 50mM</td>
<td>+ 0.7nm</td>
<td>3.1±0.1</td>
<td>5.3±0.2</td>
<td>5.5±0.2</td>
<td>4.9±0.1</td>
<td>3.8±0.1</td>
</tr>
</tbody>
</table>
Table 6 Nonlinear LSF fitting results of equilibrium binding isotherms of fluorophore height change and IHF to DNA binding ratio of different DNA sequences at high surface density in two buffers of different salt concentrations.

DNA surface density \( \sim (3.38 \pm 0.44) \times 10^{12} \text{cm}^2 \)

**LED-RS:**

<table>
<thead>
<tr>
<th>( K_d ) (nm)</th>
<th>Control</th>
<th>( H'(23) )</th>
<th>( H'(34) )</th>
<th>( H'(36) )</th>
<th>( H'(39) )</th>
<th>( H'(34)A )</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl 150mM</td>
<td>Cannot fit</td>
<td>59.7( \pm )25.3</td>
<td>28.8( \pm )8.7</td>
<td>10.2( \pm )3.1</td>
<td>11.1( \pm )3.0</td>
<td>80.1( \pm )30.4</td>
</tr>
<tr>
<td>NaCl 50mM</td>
<td>16.7( \pm )3.3</td>
<td>14.4( \pm )2.9</td>
<td>11.4( \pm )2.0</td>
<td>9.1( \pm )1.4</td>
<td>8.7( \pm )1.8</td>
<td>9.9( \pm )1.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( R_{\text{max}} )</th>
<th>Control</th>
<th>( H'(23) )</th>
<th>( H'(34) )</th>
<th>( H'(36) )</th>
<th>( H'(39) )</th>
<th>( H'(34)A )</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl 150mM</td>
<td>Cannot fit</td>
<td>1.5( \pm )0.5</td>
<td>1.1( \pm )0.2</td>
<td>0.9( \pm )0.1</td>
<td>0.8( \pm )0.1</td>
<td>1.9( \pm )0.4</td>
</tr>
<tr>
<td>NaCl 50mM</td>
<td>2.6( \pm )0.3</td>
<td>2.2( \pm )0.2</td>
<td>2.1( \pm )0.2</td>
<td>2.0( \pm )0.1</td>
<td>1.8( \pm )0.2</td>
<td>2.1( \pm )0.2</td>
</tr>
</tbody>
</table>

**SSFM:** (*end-time measurements@40nM*)

<table>
<thead>
<tr>
<th>( K_d ) (nm)</th>
<th>Control</th>
<th>( H'(23) )</th>
<th>( H'(34) )</th>
<th>( H'(36) )</th>
<th>( H'(39) )</th>
<th>( H'(34)A )</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl 150mM</td>
<td>Cannot fit</td>
<td>12.9( \pm )5.1</td>
<td>7.0( \pm )0.7</td>
<td>3.2( \pm )0.2</td>
<td>2.3( \pm )0.1</td>
<td>Cannot fit</td>
</tr>
<tr>
<td>NaCl 50mM</td>
<td>Cannot fit</td>
<td>1.4( \pm )0.1</td>
<td>1.1( \pm )0.2</td>
<td>0.9( \pm )0.3</td>
<td>0.7( \pm )0.1</td>
<td>2.2( \pm )0.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \Delta h_{\text{max}} ) (nm)</th>
<th>Control*</th>
<th>( H'(23) )</th>
<th>( H'(34) )</th>
<th>( H'(36) )</th>
<th>( H'(39) )</th>
<th>( H'(34)A )</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl 150mM</td>
<td>(-0.8)nm</td>
<td>1.7( \pm )0.3</td>
<td>4.1( \pm )0.1</td>
<td>4.1( \pm )0.1</td>
<td>3.7( \pm )0.1</td>
<td>Cannot fit</td>
</tr>
<tr>
<td>NaCl 50mM</td>
<td>(+0.4)nm</td>
<td>3.1( \pm )0.1</td>
<td>5.4( \pm )0.2</td>
<td>5.2( \pm )0.2</td>
<td>4.9( \pm )0.1</td>
<td>3.6( \pm )0.2</td>
</tr>
</tbody>
</table>

To understand these results and the discrepancy between fluorophore height change and molecular binding ratio measurements, we need to trace the signal detected by the two types of measurements. The change of fluorophore height measured by SSFM was a result of DNA bending when IHF specifically binds to the DNA sequence. The change of IHF to DNA binding ratio measured by LED-RS was a result of IHF accumulation to the DNA spots, assuming DNA molecules did not come off from the covalent bonds to the polymer surface. Hence, it is probable that nonspecific binding of IHF to DNA also increases the IHF to DNA binding ratio.
3.3.2 Quantitative analysis

If both specific binding and nonspecific binding of IHF contributed to the IHF to DNA binding ratio, it is reasonable that the $K_d$ obtained from IHF to DNA binding ratio measurements was much larger than that obtained from fluorophore height change measurements. Hence, we assume that both specific and nonspecific binding of IHF were detected by LED-RS, and propose an additive Langmuir binding isotherm model to analyze the equilibrium binding isotherms of IHF to DNA binding ratio. The additive model simply adds two Langmuir binding isotherms, one describing the specific binding of IHF to DNA, and the other describing the nonspecific binding of IHF to DNA:

$$\langle R \rangle = \frac{\langle R_{specific} \rangle}{K_{d(specific)} + x} + \frac{\langle R_{nonspecific} \rangle}{K_{d(nonspecific)} + x}$$

where $\langle R \rangle$ is the measured IHF to DNA ratio, x is the IHF concentration, $K_{d(specific)}$ is the dissociation constant of specific IHF binding to DNA, $K_{d(nonspecific)}$ the dissociation constant of nonspecific IHF binding to DNA, $\langle R_{specific} \rangle$ is the maximum ratio of IHF specifically bound to DNA at equilibrium, and $\langle R_{nonspecific} \rangle$ is the maximum ratio of IHF nonspecifically bound to DNA at equilibrium.

We further assume that $K_d$ obtained from the equilibrium binding isotherm of fluorophore height changes resulted from DNA bending characterizes specific binding of IHF to DNA, and is equal to $K_{d(specific)}$. This assumption is based on measurements of the fluorophore height change of nonspecific Control sequence that did not show significant height changes associated with DNA conformational change (Figure 38,
We now resort to nonlinear LSF again and fit the equilibrium isotherm of IHF to DNA binding ratio to the additive model using $K_{d(specfic)}$ obtained from adapted Langmuir binding isotherm fitting of the equilibrium measurements of fluorophore height changes. The three fitting parameters obtained are $R_{specific}$, $R_{nonspecific}$, and $K_{d(nonspecific)}$ (Figure 44).

![Figure 44](attachment:figure44.png)

**Figure 44** Analysis of equilibrium isotherm of IHF to DNA binding ratio. (a) Fitting the equilibrium isotherm with a single Langmuir binding isotherm. (b) Fitting the equilibrium isotherm with an additive Langmuir binding isotherm model.

One might argue that adding parameters to a curve fit certainly results in a better fit, especially when there are only 7 data points used for fitting. This is a valid concern, and only increasing the data points can answer the question whether we are proposing a better model for fitting the equilibrium isotherm of IHF to DNA binding ratio. In statistics, different criteria are used to compare and select different statistical models, such as Akaike information criterion (AIC), AICc, Bayesian information criterion (BIC), and adjusted $R^2$. We will not go into deeper discussion for comparing the two models as we only have limited data points. Nonetheless, we calculated a simple criterion, the
adjusted $R^2$, which penalizes including extra variables in the model, to show that the additive Langmuir binding isotherm is a more practical model for our data (Figure 45). Moreover, we fit the equilibrium measurements of IHF to DNA binding ratio of the arbitrary Control sequence to the additive model and obtained a specific binding ratio of 0.05 and a nonspecific binding ratio of 2.6, which is the same IHF to DNA binding ratio obtained from the single Langmuir binding isotherm fit. Thus, we propose to use the additive Langmuir binding isotherm model to understand the experimental data measured by LED-RS and to further characterize specific and nonspecific IHF-DNA interactions.

![Adjusted $R^2$ of Least Squares Fitting](image)

**Figure 45** Compare adjusted $R^2$ of single Langmuir binding isotherm fit and additive Langmuir binding isotherm model fit for analyzing the equilibrium isotherms of IHF to DNA binding ratio.

Since the fluorophore height changes measured by SSFM is an ensemble average, it is critical to know the fraction of IHF specifically bound to the DNA molecules in order to calculate the DNA bending angle caused by IHF. Now that we can obtain both specific and nonspecific binding ratio from LED-RS measurements, we can normalize the measurement of ensemble fluorophore height changes. Hence, the measurements of DNA
bending induced by IHF using SSFM and overall IHF to DNA binding ratio using LED-RS complement each other by providing both conformational specific and nonspecific binding information.

We used the additive Langmuir binding isotherm model to fit again all the equilibrium isotherms of IHF to DNA binding ratio obtained by averaging measurements of 10 DNA spot replicates of each spotting condition in buffer containing 50 mM NaCl. As will be discussed later, nonspecific binding is inhibited in buffer containing high salt concentrations, we thus did not fit the equilibrium measurements of IHF to DNA binding ratio obtained in buffer containing 150 mM NaCl to the additive Langmuir binding isotherm model. Deeper interpretation of the measured and fitting results on specific and nonspecific IHF-DNA interactions will be discussed in the upcoming chapters.

3.4 Conclusions

In this chapter, we introduced the 1-D spectral imaging configuration of the combined SSFM and LED-RS, which allows us to simultaneously measure multiple DNA spots. We demonstrated the use of SSFM and LED-RS for parallel detection of DNA conformational change and protein binding. We proposed an additive Langmuir binding isotherm model to resolve conformational specific and nonspecific IHF-DNA binding using both measurements by SSFM and LED-RS. Equilibrium binding isotherms of fluorophore height changes and IHF-to-DNA binding ratios of DNA spots immobilized at different densities in buffers of two salt concentrations were analyzed. We will continue to analyze these results in more details in the following chapters.
CHAPTER 4

CHARACTERIZATION OF PROTEIN-DNA INTERACTIONS ON A SURFACE

Quantitative determination of the density and conformation of DNA molecules tethered to the surface can help understand and optimize DNA nanosensors and nanodevices, which use conformational or motional changes of surface-immobilized DNA molecules for detection or actuation. The dual-spectral detection modality of SSFM combined with LED-RS enables characterization of DNA conformation and density on the surface and constitutes a versatile approach for nanoscale solid-biochemical interface investigations. In this chapter we examine different factors that influence the detection and quantification of DNA conformational change and protein binding on a surface.

4.1 Motivation

The immobilization or synthesis of DNA molecules on a solid surface has stimulated the development of DNA sensors and nanodevices in wide-ranging biomedical applications. For instance, DNA microarrays, which allow for massive parallel multiplexed analysis, have become routine for gene-expression profiling in both research and clinical practice\(^{63,105,106}\). In recent years, utilizing surface-immobilized DNA, researchers have developed switch-based highly sensitive biological detectors and dynamic programmable DNA motors\(^{101,104,107-110}\). To understand and optimize the sensing and actuation performance of these DNA sensors and nanodevices, it is highly desirable to characterize two vital parameters, the DNA surface density and conformation, which dictate the behaviors of the surface-immobilized DNA molecules\(^{90,111-115}\).
Various optical, electrical and mechanical techniques have been developed to characterize the surface density, conformation or thickness of the DNA molecules immobilized on two-dimensional (2-D) functionalized surfaces such as gold, quartz crystal, silicon, silicon dioxide or diamond\textsuperscript{138,141-148}. In recent years, 3-D polymeric coatings, such as the copolymer surface used in this work, have been commonly used for surface functionalization given their high immobilization specificity and capacity, optimized biological activity, and simple immobilization procedure\textsuperscript{128,130,149-152}. The axial swelling of the 3-D polymeric coating in aqueous solutions increases the complexity for precise characterization of the biophysical properties of the solution-phase interfacial microenvironment. SSFM combined with LED-RS provides a convenient non-destructive method to directly determine both the nanoscale conformation and surface density of DNA immobilized on the 3-D polymeric surface.

We use the measured and fitting results from the IHF-DNA binding equilibrium titration experiments to investigate the biophysical factors affecting the detection and quantification of protein-DNA interactions on the SSFM platform. Investigating these factors not only demonstrates the versatility of the combined SSFM and LED-RS platform, but also helps us understand surface molecular interactions and optimize assay designs.
4.2 Factors affecting binding and conformational change

4.2.1 DNA orientation

Knowing the orientation of surface-immobilized DNA is critical to deduce DNA conformational changes from fluorophore height measurements by SSFM. SSFM detects height changes of fluorophores labeled on the surface-distal end of surface-tethered DNA molecules by assuming that the DNA molecules are rigid and oriented. When the DNA molecules bend or kink, the average height change of the fluorophores on the DNA surface-distal end drops, assuming the average DNA orientation remains the same. Given the same degree of DNA bending or kink, the fluorophore height change becomes larger at higher orientations. Thus, the orientation of DNA before protein binding affects the detection of DNA conformation on the surface. In fact, when the DNA orientation is low enough, SSFM cannot resolve the average fluorophore height change from random thermal fluctuations and measurement error. Low DNA orientations may also create steric hindrance to DNA conformational changes. Also, since the average DNA orientation has far less probability to be at high angles, when the DNA conformational change is small, such as a slight kink, SSFM cannot distinguish minor fluorophore height changes from noises from ensemble averaging and random thermal fluctuations.

In Chapter 2, we have shown the method to measure DNA orientation using fluorophore axial heights of surface-distal and surface-proximal ends of DNA measured by SSFM. Hence, we measure DNA orientation in each experiment before IHF binding. In an ideal scenario, increasing the DNA orientation increases the fluorophore height
change and reduces steric hindrance for DNA conformational change. Assuming there is no steric hindrance, we simulated the average fluorophore height change of 60-bp DNA molecules induced by IHF binding (DNA sequence H'(36), 50% of which bent at 160°) at different DNA orientations, represented by curves of different colors in Figure 46(a). It is also intuitively straightforward that at higher DNA orientations it is less effective to increase fluorophore height changes by increasing the DNA orientation from its initial value (Figure 46(b)).

![Figure 46](image)

Figure 46 Initial DNA orientation affects fluorophore height change measurements by SSFM. (a) Average fluorophore height change caused by DNA bending increases with DNA orientation to the surface. (b) It is less effective to increase fluorophore height changes by increasing the orientations of DNA that are already at high orientations.

To increase the SNR of measurable fluorophore height change over random thermal fluctuations and measurement error, we naturally want to increase the initial DNA orientation to the surface. In Chapter 2, we have shown that DNA orientation increases with DNA surface density and decreases with salt concentration. We did
measure larger fluorophore height changes in the buffer solutions containing lower salt concentration than that containing higher salt concentration (Figure 47). Yet even increasing the DNA surface density did increase the DNA orientation, it did not increase the measured average fluorophore height change after IHF bind and bend the DNA molecules (Figure 47). A probable explanation is that at higher surface densities, the space between DNA molecules are too small either for IHF to access its binding site or for the DNA molecules to undertake the sharp bending. We will discuss the effect of DNA surface density with more experimental results in the next section.

![Figure 47](image)

Figure 47 Increasing DNA orientation by increasing DNA surface density does not increase the measured fluorophore height change by SSFM. (a) The measured orientation of DNA sequence H'(36) increases with DNA surface density. (b) The measured average fluorophore height change of H'(36) decreases with DNA surface density.

Other methods to increase the orientation of surface-immobilized DNA include applying negative potentials to the surface, typically a gold electrode, such as the “Switching-DNA” platform\textsuperscript{124}. A properly applied negative potential on the surface repels and orient end-tethered DNA molecules while not affecting their biochemical
structures. Previously, Dr. Spuhler tried to deposit a thin electrode layer of indium tin oxide or gold on the SSFM substrate and apply negative potentials on the electrodes. However, adding extra layers to the substrate complicates the optical model for fluorophore height determination and the solid-biochemical microenvironment. The charged polymer surface (See 2.2.3) that orients DNA molecules by synthesizing negative charged groups in its polymeric scaffolds constitutes a potential approach to obtain higher DNA orientations while maintain the polymeric physiological-like surface environment. Characterization of protein-DNA interactions on the charged polymer surface could be an interesting piece of future work.

4.2.2 DNA surface density

It is well known that the surface density of ssDNA molecules affects their hybridization efficiency with complementary strands$^{115,153}$. The DNA surface hybridization regimes and mechanisms have been studied intensively to optimize DNA microarrays for detecting target DNA sequences. In recent years, more solid-based sensing methods have been used for large-scale investigation of protein-DNA interactions, such as Surface Plasmon Resonance, and DNA binding microarrays (See 1.2.2). However, the binding of proteins to surface-immobilized DNA has yet been well characterized. The results of IHF binding to surface-immobilized DNA at different surface densities may help understand the characteristics of surface protein-DNA binding assays.
We examined the effect of DNA surface density on the measured average fluorophore height change, average IHF to DNA binding ratio, and the $\langle \Delta h_{\text{max}} \rangle$, $\langle R_{\text{specific}} \rangle$, $\langle R_{\text{nonspecific}} \rangle$, $K_d(\text{specific})$, and $K_d(\text{nonspecific})$ obtained from fitting the equilibrium isotherms. First, we observed that while $\langle R_{\text{specific}} \rangle$ decreased with increased DNA surface density, $\langle R_{\text{nonspecific}} \rangle / \langle R_{\text{specific}} \rangle$ increased with increased DNA surface density (Figure 48). Similarly, we also observed that while $K_d(\text{specific})$ increased with DNA surface density, $K_d(\text{nonspecific})$ decreased with DNA surface density, although differences occurred on different DNA sequences (Figure 49). Also, the $\langle \Delta h_{\text{max}} \rangle$ and average fluorophore height change measured by SSFM decreased as DNA surface density increased for some DNA sequences (Figure 50). Although we showed results of difference binding sequences here, we will discuss the effect of binding site location in the next subsection. The results suggest that specific binding of IHF to the DNA layer is inhibited while nonspecific binding of IHF to the DNA layer is enhanced as DNA surface density increases.
Figure 48 The effect of DNA surface density on specific and nonspecific IHF-DNA binding ratio. (a) The IHF to DNA specific binding ratio decreases as DNA surface density increases. (b) The IHF to DNA nonspecific binding ratio decreases as DNA surface density increases.

Figure 49 The effect of DNA surface density on measured dissociation constants of IHF-DNA interaction. (a) The dissociation constant of IHF to DNA specific binding increases with increasing DNA surface density. (b) The dissociation constant of IHF to DNA nonspecific binding decreases with increasing DNA surface density.
Figure 50 The effect of surface density on fluorophore height changes measured by SSFM. (a) The maximum fluorophore height change at equilibrium from Hill equation fitting decreases with DNA surface density. (b) The measured average fluorophore height changes decreases with DNA surface density.

Based on these observations of the effect of DNA surface density on different binding properties, we propose two interpretations:

1. The decrease of $\langle R_{specific} \rangle$ as DNA surface densities increases means that IHF specific binding to DNA molecules is inhibited at high DNA surface densities. This indicates that the high-density DNA brushes on the surface obstructed the accessibility of IHF to the specific binding sites on the DNA molecules. This phenomenon resembles the macromolecular crowding effects in living cells, where DNA molecules are packaged tightly by various nucleoid associated proteins on the chromosome\textsuperscript{18}. The crowding of DNA molecules obstructs DNA-binding proteins and affects their binding dissociation constants, which was also observed in the effect of DNA surface density on $K_{d(specific)}$ in our results. On the other hand, the increase of $\langle R_{nonspecific} \rangle/\langle R_{specific} \rangle$ and the decrease of
\( K_{d(nonspecific)} \) provide evidence that nonspecific binding of IHF to DNA was enhanced. This is probably due to the reinforced negative surface potential of the DNA layer, enhancing electrostatic interactions between negatively charged DNA and positively charged proteins responsible for nonspecific interactions.

2. The decrease of maximum or measured average fluorophore height change for some DNA sequences as DNA density increases suggests that dense DNA layer might add additional steric hindrance for DNA bending. It is also plausible that the average fluorophore height change decreases with increased DNA surface density is a result of decreased \( \langle R_{specific} \rangle \) and ensemble averaging. However, the increase of DNA surface density also increases DNA orientation, which in turn could result in higher fluorophore height change. Thus, DNA surface density can have two opposite effects on fluorophore height changes caused by DNA bending. Ultimately, we need to calculate the specific DNA bending angle to examine whether a particular surface density causes steric hindrance to the DNA conformational change.

4.2.3 DNA length and binding site location

Besides DNA orientation, SSFM requires certain lengths of DNA to acquire fluorophore height changes large enough to distinguish from random noise and measurement errors and to further make quantifications of DNA conformational changes. Given a certain location along the DNA sequence where the bending or kink occurs, longer DNA sequence results in larger fluorophore height change. Nonetheless, when
DNA length comes close to the persistence length of DNA, the DNA molecules are no longer straight but curved. The ideal rigid rod model of surface-immobilized DNA does not apply, and more complicated models are required, adding difficulty for SSFM to quantify DNA conformational changes. Thus the DNA length should be long enough to enable detection of specific DNA conformational changes but short enough to be assumed as straight and rigid.

Additionally, a leader DNA sequence should be placed before the binding site to start the DNA sequence. On one hand, we have shown that DNA molecules can penetrate and are vertically distributed in the 3-D polymeric surface, the degree of which depends on the DNA length. Despite that longer DNA molecules penetrate less, the first few or ten base pairs can still reside in the polymeric matrix, which might hinder protein binding. Adding a certain length of leader DNA sequence shifts the binding site away from the surface and adds to the DNA length. Moreover, there exist lower and upper limits of the binding site location for SSFM measurements, depending on the length of the binding site and how much the DNA bends or kinks. Taking the lower limit of the 34-bp H’ binding site as an example. Assuming that we know the center of the H’ site on a 60-bp DNA sequence, but do not know how much and to which direction it would bend, we need to place the H’ site at a location that does not hinder any possible bending angle or direction (Figure 51). A simple geometric simulation shows that the lower limit of binding site location decreases with DNA orientation. For the typical range of DNA orientations on the polymeric surface in our IHF-DNA binding experiments, the lower limit of the center of the H’ binding site location is at about 32-bp to 35-bp (Figure 52).
Figure 51 Schematic illustration of the physical limits for the location of the 34-bp H' binding site along a 60-bp DNA sequence.

Figure 52 The lower limit of the binding site center location decreases as DNA orientation increases. The simulated relationship is for 34-bp H' binding site on a 60-bp DNA sequence. Inset: The lower limit of the H' binding site location in the typical range of DNA orientations on the polymeric surface in IHF-DNA binding experiments.
Previously, in our experiment design, we proposed to see whether we can distinguish shifted H’ binding site locations on 60-bp DNA (Figure 36). The idea is that given a certain DNA length and bending angle, the DNA sequence that has the binding site at a lower location would have a larger fluorophore height change caused by IHF-induced DNA bending. This has been detected by the “Switching-DNA” platform where the detected fluorescence intensity after IHF binding at equilibrium decreased as the H’ binding site on 80-bp DNA molecules shifted closer to the surface. But the “Switching-DNA” platform cannot distinguish less specific binding from lower fluorophore heights in its fluorescence intensity measurements, and thus was not able to resolve whether the decreased intensity was due to less IHF binding or a lower binding site location.

We thus compared measured average fluorophore height changes of DNA H’(34), H’(36), and H’(39), whose H’ binding site center locations are indicated in the parentheses of the DNA names. We did observe different fluorophore height changes of the three DNA sequences as expected on DNA spots immobilized at low surface densities (Figure 53). The average fluorophore height change of DNA H’(34) was larger than average fluorophore height change of H’(36), and both were larger than that of H’(39). As the DNA surface density increased and started to hinder IHF binding and DNA bending, the difference of measured fluorophore height changes between these DNA sequences decreased (Figure 53). We can further normalize the fluorophore height changes by the specific IHF to DNA binding ratio of each DNA sequence respectively and reveal even larger fluorophore height change differences between the DNA sequences (Figure 54). This normalization decouples the effect of less binding and demonstrates that the less
fluorophore height change observed was due to the lower binding site location. This result was also observed for \( \langle \Delta h_{\text{max}} \rangle \), the maximum average fluorophore height change at equilibrium shown in previous section (Figure 50(a)). Notice that the DNA sequence H’(23) did not have the largest fluorophore height change even its H’ binding site was closest to the surface. This is because that the center location of H’ binding site on the DNA sequence H’(23) was below the lower limit, when steric hindrance ruined the accurate representation of average fluorophore height changes for the DNA conformational changes.

If the binding site is short and at the end of the DNA sequence and thus resides at the top of the end-tethered DNA molecule, the fluorophore height change caused by DNA bending or kinks may be indistinguishable from measurement errors and random noises. Thus, short protein binding sites should be paid with extra attention.
Figure 53 SSFM can distinguish shifted H’ binding site location on 60-bp DNA by measuring IHF binding induced average fluorophore height changes of DNA. The difference of average fluorophore height changes between the different DNA sequences, H’(34), H’(36), and H’(39), reduces as surface density increases. The IHF binding site of DNA H’(23) is below the lower limit of binding site location thus cannot accurately represent DNA conformational changes. IHF to DNA binding ratio is presented to show that the measured different fluorophore heights are not due to less binding, but due to shifted binding site locations.
Figure 54 The difference of fluorophore height changes of DNA sequences containing shifted H’ binding site locations is more evident if normalized by the specific binding ratio. (a) Equilibrium binding isotherms of fluorophore height changes of the three DNA sequences. (b) Maximum fluorophore height changes Δℎₘₐₓ and maximum fluorophore height changes normalized by IHF to DNA specific binding ratio (𝑅_{specific}). This figure shows comparison of fluorophore height change measurements of DNA spots immobilized at low surface density.

4.2.4 Ionic strength and short spacer DNA

In Chapter 2, we have shown that under the same surface density, average DNA orientation decreases as buffer salt concentration increases on the polymer surface. The observation was in agreement with Manning’s counterion condensation theory\textsuperscript{119} that cations in the buffer can condense onto the negatively charged dsDNA backbones and shield the electrostatic repulsive forces between the dsDNA molecules. Therefore, increased concentration of cations in the buffer reduces the electrostatic repulsion between the dsDNA molecules, which resulted in more random rotational freedom of the dsDNA and therefore lowered average DNA orientation. This effect was also observed in the IHF-DNA binding experiments, in which the average orientation of the same DNA spot was smaller in buffer containing 150 mM NaCl than in buffer containing 50 mM NaCl. Therefore, assuming the DNA molecules bent at the same angle in the two buffers,
the average fluorophore height change was smaller in buffer containing 150 mM NaCl than in buffer containing 50 mM NaCl (Figure 46, Figure 47).

Increasing DNA surface density does increase DNA orientation, but decreases average fluorophore height change due to increased steric hindrance between DNA molecules in a denser DNA layer. Is there a way that we can increase the DNA orientation without increasing its surface density besides applying external forces? We mixed 60-bp DNA sequences that had the H' binding site with 20-bp spacer DNA that had no H' binding site at the same concentration and spotted them together on the polymer surface. The 60-bp DNA were labeled with fluorophores while the 20-bp DNA were not. Hence we measured the average orientation and fluorophore height changes of just the 60-bp DNA molecules with SSFM. We also spotted only 60-bp DNA sequences at the same concentration as control groups. We noticed that the DNA orientation of mixed 60-bp and 20-bp DNA spots was larger than the DNA orientation of lone 60-bp DNA spots in Tris buffer containing 150 mM NaCl (Figure 55(a)). We also measured a higher average fluorophore height change of 60-bp DNA in the mixed 60-bp and 20-bp DNA spots than the lone 60-bp DNA spots in Tris buffer containing 150 mM NaCl (Figure 55(b)). We did not observe the same effect of the 20-bp spacer DNA on the same DNA spots in Tris buffer containing 50 mM NaCl (Figure 56).
Figure 55 The effect 20-bp spacer DNA molecules on the DNA orientation and IHF-induced fluorophore height changes of 60-bp DNA molecules in buffer containing 150 mM NaCl. (a) Average DNA orientation is larger on spots containing mixed 60-bp and 20-bp DNA sequences than spots containing lone 60-bp DNA sequence. All the DNA sequences are spotted at 15 µM. (b) Average fluorophore height change is larger on spots containing mixed 60-bp and 20-bp DNA sequences than spots containing lone 60-bp DNA sequence.

Figure 56 The effect 20-bp spacer DNA molecules on the DNA orientation and IHF-induced fluorophore height changes of 60-bp DNA molecules in buffer containing 50 mM NaCl. (a) The average DNA orientation of spots containing mixed 60-bp and 20-bp DNA sequences and spots containing lone 60-bp DNA sequence are indistinguishable. (b) The average fluorophore height change of spots containing mixed 60-bp and 20-bp DNA sequences and spots containing lone 60-bp DNA sequence are indistinguishable.
Based on the above observations of the effect of 20-bp spacer DNA molecules on DNA orientation and fluorophore height change, we propose the following interpretation: in the buffer containing 150 mM NaCl, it is possible that 20-bp spacer DNA molecules increased the orientation of 60-bp DNA by increasing the surface density of the first 20-bp of the DNA layer. The increased surface density at the bottom 20-bp DNA layer did not add steric hindrance for DNA bending at the middle of the 60-bp DNA. Thus, the fluorophore height change increased as a result of increased DNA orientation without additional steric hindrance. On the other hand, in the buffer containing 50 mM NaCl, the effect of 20-bp spacer DNA on the orientation of 60-bp DNA was not observable. This could be that the added electrostatic repulsion between adjacent DNA molecules was too small to further increase the DNA orientation, which was already high and possessed high energy state in the low salt buffer.

Since 20-bp DNA and 60-bp DNA may have different immobilization dynamics, we cannot determine their individual surface density given that LED-RS measures ensemble average thicknesses. It is neither possible to calculate specific and nonspecific IHF to DNA binding ratio and use it to normalize average fluorophore height changes. Thus, although 20-bp spacer can increase measured fluorophore height changes in buffer containing higher salt concentrations, we did not continue to add them in the study of IHF-DNA interactions.
4.2.5 Conclusion

In summary, DNA orientation, surface density, and the binding site location affect the final results of IHF to DNA binding ratio and fluorophore height change measurements. In an ideal world, we would like to have high DNA orientation, low surface density, and optimized binding site location. DNA orientation and surface density change in opposite directions and optimizing binding site location requires prior knowledge of its length and sequence. Nevertheless, SSFM combined with LED-RS is a very handy tool to study and optimize these factors. Particularly, the microarray measurement fashion allows for the characterization of hundreds of DNA sequences of various lengths and different immobilization conditions on one chip. By characterizing the different factors affecting IHF-DNA interactions on the surface, we now can select an optimal interaction regime on the polymer surface for the study of specific and nonspecific IHF-DNA interactions and accurately analyze the results.
4.3 Effect of extra macromolecule layer on fluorophore height determination

The binding of protein molecules to the surface-immobilized DNA can have an effect on the precise determination of surface-distal end fluorophore height change. We examine this effect by reviewing the assumptions made to use SSFM to measure fluorophore heights tagged on DNA.

SSFM determines the height of fluorophores tagged on surface-immobilized DNA molecules on the polymer-functionalized two-layer SiO$_2$-Si substrate assuming that:

1. The refractive indices of the thermally grown SiO$_2$ layer at different wavelengths are predetermined;

2. The SiO$_2$ layer thickness can be obtained and used in the fitting algorithm to find the fluorophore height to the SiO$_2$ surface;

3. Extra organic layers adsorbed on the SiO$_2$ surface, such as the accumulated polymer scaffold, DNA, and protein, add to the SiO$_2$ layer thickness;

4. The refractive indices of the DNA and protein layers are the same as those of the SiO$_2$ layer.

The SiO$_2$ layer and DNA or protein layer thicknesses can be accurately measured by WLRS or LED-RS given the right refractive indices. The refractive indices of thermally grown SiO$_2$ layer have been well characterized as a dependent variable of wavelength in a deterministic equation. Many efforts have been spent to determine the refractive indices of DNA or protein layers on a surface, and only certain ranges were obtained, which change with layer thicknesses and buffer environments. By simply assuming the refractive indices of the DNA and protein layers to be the same with those
of the SiO$_2$ layer prohibits us to accurately measure the thickness of DNA and protein layers on top of the SiO$_2$ layer. However, WLRS and LED-RS can faithfully detect the optical thickness change of the SiO$_2$ layer upon adding DNA and polymer. Assuming the refractive indices of the DNA and protein layers to be the same with those of the SiO$_2$ layer, the optical thickness change can be converted to a thickness change, which is then added to the SiO$_2$ layer in SSFM fitting algorithm to determine the fluorophore height to the surface.

In SSFM dry measurements, all the organic molecules collapse to the surface and the fluorophores are on top of the organic molecular layer. We can obtain the absolute optical thickness between the fluorophores and the reflecting SiO$_2$-Si interface from fluorophore height measurement using SSFM. Assuming that fluorophores only constitutes a minimal percentage of the mass density of the organic layer, we can also obtain the absolute optical thickness using LED-RS. In the experimental setup, many factors, such as imperfect optical alignments, minor slope of chip surface, additional coated window between objective and chip surface, and imperfect spectrometer alignment, can all result in measurement errors. Thus the absolute optical thickness may not be precisely determined, but a relative change of optical thickness to a reference value can be accurately determined where systematic measurement errors can be cancelled in a subtraction. We thus use relative optical thickness change from an initial reference value in the same experiment as our signal. Therefore, we can obtain optical thickness increase of SiO$_2$ layer and fluorophore height change after depositing DNA by SSFM dry measurements or LED-RS.
In SSFM wet measurements, the fluorophores are not directly on top of the organic molecular layer, but are tagged to DNA molecules immobilized on the polymeric coating on the SiO$_2$ surface. The polymer swells and DNA molecules orient, lifting the fluorophores away from the surface. To determine the fluorophore heights to the surface, we need to know the absolute optical thickness of the SiO$_2$ layer added by DNA and polymer, and the refractive index of the aqueous solution in which the surface and the DNA are immersed. In fact, we cannot easily determine the refractive index of the aqueous solution that includes DNA molecules, polymeric scaffold, and re-distributed ions at the solid-biochemical environment. We approximated the refractive index of the aqueous environment with that of water or low salt buffer. Since the absolute optical thickness could be an inaccurate measurement and the solution refractive index was approximated, SSFM may not measure the exact absolute height of the fluorophore to the reflecting SiO$_2$-Si interface, but can accurately measure fluorophore height difference on of the same location on the surface where the absolute optical thickness remains the same.
We now examine the fluorophore height change measurement at the DNA surface-distal end, which was used to determine DNA bending by IHF. We illustrate all the heights and thicknesses to be used in Figure 57. Assuming that the SiO$_2$ and DNA layer total thickness is $D_1$, we use $D_1$ in SSFM fitting algorithm to determine the height of surface-distal fluorophore to the surface as $d_1$. After IHF binds to DNA, the total thickness of SiO$_2$ and DNA and IHF layer is $D_2$, which is then used in SSFM fitting algorithm to determine the height of surface-distal fluorophore to the surface as $d_2$. Before IHF binding, we add $D_1$ and $d_1$ to obtain $h_1$ as the absolute height of the surface-distal end fluorophore to the SiO$_2$-Si interface. After IHF binding, we add $D_2$ and $d_2$ to obtain $h_2$ as the absolute height of the surface-distal end fluorophore to the SiO$_2$-Si interface. We then determine the absolute height change of the surface-distal end fluorophores as:
\[ \Delta h' = h_1 - h_2 = (D_1 + d_1) - (D_2 + d_2) = (d_1 - d_2) - (D_2 - D_1), \]

where \((D_2 - D_1)\) is the optical thickness change of the SiO\(_2\) and DNA layer resulted from IHF binding. In the IHF experiments, \(\Delta h = (d_1 - d_2)\), in which \(d_1\) and \(d_1\) were obtained by using \(D_1\) in the SSFM fitting algorithm, was used as the measured average fluorophore height change rather than \(\Delta h'\), which could have resulted in an estimation error. We then simulated the estimation error of surface-distal fluorophore height change resulted from IHF-induced DNA bending as a function of IHF layer thickness, whose refractive index was assumed to be the same with SiO\(_2\). As IHF layer thickness increases, the estimation error increases to about 6% to 8% of the range of fluorophore height changes measured in experiments (Figure 58).

Figure 58 The estimation error of fluorophore height change resulted from IHF-induced DNA bending increases with accumulated IHF layer thickness.
However, we did not use this simulated estimation error to correct our fluorophore height change measurements. First, as discussed before, the absolute thicknesses $D_1$ and $D_2$ measured at two situations may not be precise, and can affect the fluorophore heights $d_1$ and $d_2$ obtained by using $D_1$ and $D_2$ in the SSFM fitting algorithm. Nevertheless, the fluorophore height difference of two axial positions is accurate when the optical thickness does not change. To avoid introducing extra measurement error, we thus only used $D_1$ in SSFM fitting algorithm to obtain $d_1$, $d_2$, and $\Delta h = (d_1 - d_2)$. Second this less than 10% estimation error is smaller than and cannot be distinguished from the SSFM measurement noise. Also, many plausible assumptions were made in this analysis.

For example, for the additional protein layer to increase the optical thickness of the SiO$_2$-DNA layer IHF must reside between the fluorophores and the SiO$_2$ and DNA layer after binding (Figure 59). However, in specific binding where IHF causes sharp bending of the DNA molecules, IHF resides above the fluorophores, which does not result in the phase difference between the direct fluorescence emission and reflected fluorescence emission, and thus does not result in the estimation error either. In our experimental results, we know that IHF binds both specifically and nonspecifically to the DNA molecules and the IHF to DNA binding ratio is more than one. But we cannot determine IHF distribution within the DNA layer and it impossible to account for the effect of IHF layer on fluorophore height changes without the distribution density of IHF on the DNA molecules in the nanoscale. Other factors, such as the polymer scaffold distribution and its density in aqueous solution, were neglected. Simply correcting experimental data using the simulated estimation error under particular assumptions is not convincing, and
thus we attribute the added protein layer effect into measurement error indicated by the standard deviation of our data.

Figure 59 The effect of IHF layer on the measured fluorophore height change depends on IHF locations on the DNA molecules. (a) In specific binding, IHF bends the DNA molecules and resides above the fluorophores and the SiO$_2$-DNA layer, which does not increase the optical path length difference between direct and reflected fluorescence emission. (b) In nonspecific binding, IHF binds to the DNA molecules and resides between the fluorophores and the SiO$_2$-DNA layer, which increases the optical path length difference between direct and reflected fluorescence emission.
CHAPTER 5

STUDY OF SPECIFIC AND NONSPECIFIC PROTEIN-DNA INTERACTIONS

5.1 Motivation

While most studies of protein-DNA interactions focus on specific binding, DNA-binding proteins generally have both specific and nonspecific binding modes. Depending on the cellular environment in vivo or the experimental conditions in vitro, the competition between specific and nonspecific binding varies, resulting in different contributions to the overall binding. In the cell, DNA-binding proteins function accurately by distinguishing specific and functional binding site from the vast genomic nonspecific potential sites. Moreover, it is suggested that in indirect readout, DNA-binding proteins achieve specific binding by first binding nonspecifically and diffusing along the 1-D DNA sequence until they recognize and capture certain DNA conformational fluctuations. Thus, it is imperative to resolve and examine both specific and nonspecific binding of protein-DNA complexes to understand their functions and binding mechanisms.

5.2 Parallel detection of specific and nonspecific binding of IHF

IHF functions in both specific and nonspecific binding complexes and its concentration and binding specificity vary with cell growth. In specific binding, IHF induces a sharp turn of a 34-bp long DNA, which facilitates many cellular processes by apposing distant DNA binding sites for other proteins. In nonspecific binding, IHF
binds to a 9 to 16-bp long DNA\textsuperscript{157} and acts as one of the major nucleoid-associated proteins to compact the 4.7 million-bp circular DNA into the bacterial chromatin\textsuperscript{47,48}. It is possible that the DNA might weakly bend in the nonspecific IHF-DNA complex or in an intermediate state from nonspecific to specific binding\textsuperscript{42,158}, but high-resolution structure of nonspecific IHF-DNA complex is yet available. The competition between IHF specific and nonspecific binding modes presumably results in its regulatory functions in prokaryote cells\textsuperscript{11}. Therefore, to characterize IHF-DNA interaction and understand its function, both specific and nonspecific binding should be measured and analyzed.

Biosensing techniques that measure protein-DNA interactions by detecting mass density changes, such as surface plasmon resonance (SPR), cannot resolve the individual contribution of specific binding and nonspecific binding to their signals and exploit various approaches to suppress nonspecific binding\textsuperscript{159-161}. On the other hand, biophysical techniques that measure DNA conformational changes, such as FRET, cannot provide quantitative information about nonspecific binding that does not contribute or adds noise to the signal\textsuperscript{22,50}. By combining SSFM and LED-RS and using the quantification model presented in Chapter 3, we demonstrate parallel measurement of specific and nonspecific IHF-DNA binding. The specific binding information is obtained from DNA bending induced by IHF measured by SSFM while the nonspecific binding is resolved from IHF to DNA binding ratio measured by LED-RS. We first examined IHF-DNA interaction specificity in buffers of different salt concentrations. Then, by differentiating specific binding and nonspecific binding, we calculated the DNA bending angle from ensemble conformational measurements. Finally we demonstrate differentiation between consensus
binding sequence and consensus sequence with a single nucleotide mutation under the
condition where nonspecific binding dominates.

5.2.1 Examining salt-dependence of IHF-DNA interactions

The affinity of protein-DNA interactions mostly exhibits a strong dependence of
salt concentration. This can be attributed to the polyelectrolyte nature of DNA and the ion
release upon protein-DNA complex formation, which results favorable entropic changes
to stabilize the complex. The X-ray crystal structure of the IHF-H' site complex reveals
that the DNA helix wraps around the sides of the protein, resulting in over 20 positively
charged amino acid side chains make ionic interactions with the negatively charged DNA
phosphate groups.\(^{43}\)

Given the many potential electrostatic interactions in the IHF-DNA complex, we
examined the salt-dependence of specific and nonspecific IHF-DNA binding. We
compared IHF-DNA binding in Tris buffer solutions of two salt concentrations, NaCl 50
mM (low salt buffer) and NaCl 150 mM (high salt buffer). First, we present evidence
showing that IHF binds nonspecifically to both arbitrary Control sequences and the 60-bp
IHF binding sequence including the 34-bp H' site (Figure 60, Figure 61). We quantified
IHF to DNA binding ratios of 20-bp, 40-bp, and 60-bp long arbitrary Control DNA
sequences (Figure 60). The DNA spots were immobilized at low surface densities and
measured in low salt or high salt buffers. Each data point is the average and standard
deviation of measurements of 10 DNA spot replicates. The IHF to DNA binding ratio at
equilibrium in 40 nM IHF solution in low salt buffer was about four times higher than
that in high salt buffer of all Control sequences. The result indicates that IHF nonspecifically binding to arbitrary DNA sequences of different lengths, which is inhibited in high salt buffer.

![Figure 60](image)

Figure 60 IHF binds nonspecifically to arbitrary Control sequences of different lengths. The IHF to DNA binding ratio at equilibrium in 40 nM IHF solution is about four times higher in the low salt buffer than in the high salt buffer.

Next, we compare IHF binding to 60-bp IHF binding sequence including the 34-bp H' site in low salt buffer and high salt buffer (Figure 61). IHF to DNA binding ratio reached equilibrium after 10 minutes of incubation in high salt buffer whereas it kept increasing in low salt buffer. However, in both low salt and high salt buffers, fluorophore height change reaches equilibrium at the same time scale with that of IHF to DNA binding ratio in high salt buffer. Since the kinetics of fluorophore height change
represents that of specific binding and is similar to that of the IHF to DNA binding ratio in high salt buffer, we speculate that in high salt buffers, IHF only binds specifically to the H' site in the DNA sequence whereas in low salt buffer IHF bindings both specifically and nonspecifically to the 60-bp DNA sequence.

Figure 61 Real-time measurement of IHF binding to DNA sequence H'(39) and Control sequence. Left: IHF to DNA binding ratio reaches equilibrium after 10 minutes of incubation in high salt buffer but keeps increasing in low salt buffer. Right: In both low salt and high salt buffers, fluorophore height change reaches equilibrium at the same time scale with that of IHF to DNA binding ratio in high salt buffer.

To further investigate IHF-DNA interaction in buffers of different salt concentrations, we measured dissociation kinetics of IHF from IHF-bound DNA spots at equilibrium. Both 60-bp IHF-binding sequence (H'(39)) and arbitrary Control sequence
were examined. Since our flow cell was not optimized to work in the reaction-limited regime, we did not examine the association kinetics. Also, to increase the dissociation speed, we incubated the IHF-bound DNA spots in buffer solutions containing unmodified DNA sequence H'(34) at a concentration of 400 nM to knock off the bound IHF faster by competing with the surface-immobilized DNA molecules. As IHF came off from the DNA spots, both the average fluorophore height change and IHF to DNA binding ratio decreased of the IHF binding sequence H'(39) (Figure 62). We thus obtained the dissociation rates from the kinetics of average fluorophore height change and IHF to DNA binding ratio by fitting the dissociation curve to a typical exponential decay equation (Table 7):

\[ y = y_0 + Ae^{-\frac{t}{t_0}} \]

where \( t_0 \) is the dissociation time constant.

The dissociation time constant of IHF to DNA binding ratio of the Control sequence is that of nonspecific binding, and the dissociation time constant of fluorophore height change is that of specific binding. We observed that at low salt buffer, the dissociation of nonspecifically bound IHF dominated the dissociation kinetics of DNA sequences containing the H’ binding site, whereas in high salt buffer the dissociation of specifically bound IHF dominated the dissociation kinetics.
Figure 62 Dissociation kinetics of IHF from DNA spots after binding at equilibrium. (a) The average fluorophore height change decreases as IHF comes off the DNA molecules, and its dissociation rate is larger than high salt buffer than in low salt buffer. (b) The average IHF to DNA binding ratio decreases as IHF comes off the DNA molecules, and its dissociation rate is similar in high salt buffer and in low salt buffer.

Table 7 Dissociation time constant of 60-bp IHF binding sequence and Control sequence measured by SSFM and LED-RS.

<table>
<thead>
<tr>
<th>NaCl (mM)</th>
<th>$t_{SSFM}$ (min)</th>
<th>$t_{LED-Consensus}$ (min)</th>
<th>$t_{LED-Control}$ (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>17.3</td>
<td>7.9</td>
<td>8.1</td>
</tr>
<tr>
<td>150</td>
<td>7.7</td>
<td>7.6</td>
<td>3.8</td>
</tr>
</tbody>
</table>

In agreement with other studies, our results show smaller binding affinity of both specific and nonspecific binding in high salt buffer than in low salt buffer (See Table 4, Table 5, Table 6). We measured the $\log(K_d)/\log([NaCl])$ slope of IHF to 60-bp DNA sequence containing the H' site between 6 to 7 for specific binding comparing to about 8 to 9 of IHF to 35-bp H' site measured by other techniques$^{162}$. We also observed that in high salt buffer nonspecific binding of IHF was much more suppressed than specific
binding. It is understood that specific binding has conformational, electrostatic, and thermodynamic contributions, whereas nonspecific protein-DNA interactions are primarily electrostatic\textsuperscript{11,29}. The increase of buffer ionic strength or salt concentration decreases the strength of electrostatic interactions, which explains the distinctive salt-dependence behaviors of specific and nonspecific binding.

5.2.2 Quantifying DNA bending angle induced by IHF specific binding

We continue to use the rigid rod model of 60-bp surface-immobilized dsDNA molecules for the quantification of DNA bending induced by IHF. The fluorophore height of the DNA surface-distal end reduces when the DNA bends upon specific binding to IHF. Since the fluorophore height of Control DNA sequence did not change, we assume that DNA molecules that bind nonspecifically to IHF do not form the specific bending while DNA molecules that bind specifically to IHF undertake a specific sharp bending. Since SSFM renders ensemble average, the measured average height change of the surface-distal end after IHF binding is an average of both straight and bent DNA molecules. Now that we can obtain IHF to DNA ratio of specific binding from LED-RS measurements, we can calculate the average fluorophore height change of just the bent DNA molecules. We thus propose a simple geometric model to use the average fluorophore heights measured by SSFM to calculate the bending angle of DNA induced by specific IHF binding.

We first define and use a few parameters for the calculation of the DNA bending angle as shown in Figure 63. We designate $\langle \Delta h_{\text{measured}} \rangle$ as the SSFM measured average
fluorophore height change, \( \langle \Delta h \rangle \) as the average fluorophore height change of DNA molecule after specific binding to IHF, \( h_0 \) as the initial height difference between DNA surface-distal and surface-proximal ends, \( \langle R_{\text{specific}} \rangle \) as the average of IHF to DNA molecular ratio of specific binding, \( \theta_{\text{DNA}} \) as the orientation of dsDNA molecule to the surface, and \( P \) as the nucleotide position of the center of the binding sequence from the DNA sequence first nucleotide. The distance between nucleotides of B-form DNA is 0.34 nm, so on a 60-bp long dsDNA, we have:

\[
L_1 = (P - 1) \times 0.34, \quad L_2 = (60 - P - 1) \times 0.34.
\]

The height of surface-distal end of bent DNA molecules is designated as \( h_1 \), a geometric average given that the DNA can have various bending directions. Thus we can give the relationship between the ensemble measured fluorophore height change and the fluorophore height change of DNA molecules specifically bound to IHF by:

\[
\langle \Delta h_{\text{measured}} \rangle = \langle h_0 \rangle - (1 - \langle R_{\text{specific}} \rangle) \cdot \langle h_0 \rangle + \langle R_{\text{specific}} \rangle \cdot \langle h_1 \rangle = \langle h_0 \rangle - \langle h_1 \rangle = \langle \Delta h \rangle
\]

Based on geometric calculation, on a single DNA molecule:

\[
\Delta h = (L_2 + L_2 \cdot \sin(\alpha_{\text{bend}} - 90^\circ)) \cdot \sin \theta_{\text{DNA}}.
\]

Thus the average DNA bending angle \( \langle \alpha_{\text{bend}} \rangle \) is:

\[
\langle \alpha_{\text{bend}} \rangle = \arcsin((\frac{\langle \Delta h_{\text{measured}} \rangle}{\langle R_{\text{specific}} \rangle \cdot \langle \theta_{\text{DNA}} \rangle} - L_2) / L_2) + 90^\circ.
\]

Here we note that in this geometric model, \( \langle \Delta h_{\text{measured}} \rangle, \langle h_0 \rangle, \langle h_1 \rangle, \langle \Delta h \rangle, \langle \theta_{\text{DNA}} \rangle \), and \( \langle R_{\text{specific}} \rangle \) are ensemble average values measured by SSFM and LER-RS, whereas \( h_0, h_1, \)
\( \Delta h, \theta_{\text{DNA}}, P, L_1, \) and \( L_2 \) are geometric parameters of each individual DNA molecule. As discussed in Chapter 2, we approximate the statistical average of \( \langle \theta_{\text{DNA}} \rangle \) by calculated average DNA orientation from \( \langle h_0 \rangle \):

\[
\langle \theta_{\text{DNA}} \rangle = \arcsin \left( \frac{\langle h_0 \rangle}{\sqrt{\langle r^2 \rangle}} \right)
\]

where \( \sqrt{\langle r^2 \rangle} \) is the root-mean-square (rms) end-to-end distance of the dsDNA based on the worm-like chain model\(^{118}\).

This simplified model of DNA bending induced by IHF we used does not reflect the real bending mechanism, in which two kinks of the DNA helix result in the overall bending. But the model is generalizable to other proteins whose conformational change mechanism remains to be discovered. In those cases, a fluorophore height change suggests a conformational change of the DNA and further accurate structural interpretations need to be examined by other methods, such as X-ray crystallography.
153

We thus calculated the DNA bending angle caused by IHF specific binding (Figure 64). Since we need to use the fitted parameters, the mean and standard errors of the parameters were used to calculate the estimated mean of the bending angle and its standard deviation. The estimated bending angle agrees with the DNA bending angles estimated by other methods, such as gel electrophoresis and X-ray crystallography\textsuperscript{43}. Our result shows that under low surface density, we can precisely determine the specific DNA bending angle caused by IHF binding.

However, the DNA bending angle, whose calculation already took into account of $\langle R_{specific} \rangle$, decreased as DNA surface density increased. This observation indicates that
for the DNA molecules that can still specifically binding to IHF at high surface densities, the DNA conformation change was compromised due to steric constraints (Figure 64).

![Figure 64 Estimation of the DNA bending angle induced by specific IHF binding.](image)

### 5.2.3 Discriminating consensus sequence with a single mutation

Studies have shown that a single T to A mutation in the center of the H' binding site TTR consensus region can destabilize the IHF-H' complex\(^\text{163}\) (Figure 65). As one of its indirect readout mechanisms, IHF distinguishes this mutation by recognizing different flexibility of the YpR step that disturbs stabilizing electrostatic interactions between IHF side chain and DNA\(^\text{163}\). By detecting both specific and nonspecific binding, we show
discrimination of IHF binding sequence with a single nucleotide mutation by quantifying different binding properties.

Figure 65 Crystal structure of the IHF–H’ DNA complex. (a) Ribbon view of the overall X-ray structure with the IHF α subunit in grey, the IHF β subunit in pink, the consensus sequence DNA bases in green and the less conserved bases in blue. (b) Stereo view of the contacts between IHF and the TTR element of the H’ site. (c) The sequence of the H’ binding site and the H’ binding site with a single nucleotide mutation. The numbering corresponds to bases 19–47 of bacteriophage λ, and arrows point to the position of artificial nicks needed introduced for crystallization (Source: Lynch, T. W., Read, E. K., Mattis, A. N., Gardner, J. F., & Rice, P. A. Integration Host Factor: Putting a Twist on Protein–DNA Recognition. Journal of Molecular Biology 2003, 330(3): 493–502).

First, we acquired the fluorophore height change and IHF to DNA binding ratio equilibrium isotherms of IHF binding sequence (sequence H'(34)), IHF binding sequence with a single nucleotide mutation (sequence H'(34)A), and an arbitrary Control sequence
The DNA spots were immobilized at low surface densities and measured in Tris buffer with 50mM NaCl. Each data point was the average and standard deviation of measurements of 10 DNA spot replicates. According to the quantitative analysis method described in Chapter 3, we obtained specific binding dissociation constant $k_{d \text{specific}}$ and maximum average fluorophore height change $\langle h_{\text{max}} \rangle$ by fitting the fluorophore height change equilibrium isotherm to the Hill equation:

$$\langle h \rangle = \langle h_{\text{max}} \rangle \frac{x}{x + k_{d \text{specific}}}$$

Also, we obtained the average IHF to DNA binding ratios $\langle R_{\text{specific}} \rangle$, $\langle R_{\text{nonspecific}} \rangle$, and nonspecific binding dissociation constant $K_{d \text{nonspecific}}$ by fitting the IHF to DNA ratio equilibrium isotherm to the dual-Hill equation model:

$$\langle R \rangle = \langle R_{\text{specific}} \rangle \frac{x}{K_{d \text{specific}} + x} + \langle R_{\text{nonspecific}} \rangle \frac{x}{K_{d \text{nonspecific}} + x}$$

SSFM can positively detect the mutated sequence from the difference of fluorophore height changes of the two DNA sequences at equilibrium in 40 nM IHF solution while mass density measurements cannot (Figure 66, Figure 67).
Figure 66 Detect single nucleotide mutation of IHF binding sequence. (a) Fluorophore height changes at equilibrium of IHF binding sequence H'(34), IHF binding sequence with a single mutation H'(34)A, and an arbitrary Control sequence. (b) IHF to DNA binding ratio at equilibrium of the three DNA sequences.

Figure 67 SSFM distinguishes IHF binding sequence containing a single nucleotide mutation H'(34)A from IHF consensus binding sequence H'(34) by measuring average fluorophore height changes.
To demonstrate the difference is because of less binding and ensemble averaging, and to investigate the effect of the single nucleotide mutation on the specific binding of IHF to DNA, we compare $k_d^{(specific)}$ and calculated DNA bending angle of the two DNA sequences (Figure 68). We observed that the dissociation constant of H'(34)A was about three times of that of H'(34) under experimental condition and the DNA bending angle induced by IHF binding of H'(34)A was slightly smaller than that of H'(34) (Figure 68).

![Bar chart](image)

Figure 68 Differentiation and comparison of IHF binding sequence H'(34) and IHF binding sequence containing a single nucleotide mutation H'(34)A. (a) The dissociation constant of H'(34)A is four times of that of H'(34) under experimental conditions. (b) The DNA bending angle induced by IHF binding of H'(34)A is smaller than that of H'(34) by 30°.

In agreement with previous studies, our results show that IHF discriminates against the mutant sequence and bind less tightly. Additionally, our results of the bending angle of the mutated DNA sequence is slightly smaller than the original H' binding site. This results agrees with previous results of the X-ray crystal structure of the IHF and single nucleotide mutated H' complex, which did not report significant DNA bending
angle difference from the IHF-H' complex\textsuperscript{163}. Recently, it was also suggested that the single T to A mutation does not affect the bendability of the DNA at the kinks\textsuperscript{42}. The observed smaller average bending angle of the mutated H' sequence in our experiments further indicates that the bending of the mutated DNA sequence may deform and fluctuate from tightly wrapped IHF-H' complex.

However, the specific binding affinity of IHF to the mutant sequence H'(34)A obtained from conformational detection was much smaller than previous EMSA results\textsuperscript{163}. We attribute this difference to the respective binding environment and specific detection methodology of each result. While EMSA measurements were made in non-equilibrium conditions and conformational variation could affect shifts of protein-DNA complexes, SSFM measurements were made at equilibrium specifically detecting conformational changes.

5.3 Specific and nonspecific protein-DNA binding and functional regulation

We have shown that parallel detection of specific and nonspecific binding facilitates our understanding of protein-DNA interactions. Most DNA-binding regulatory proteins have both specific and nonspecific binding modes, both of which need to be examined to elucidate the regulatory processes. The nucleus is highly compacted and enriched with DNA, RNA and proteins. This macromolecular crowding phenomenon has been suggested to affect protein diffusion and binding, and have regulatory effects on DNA replication and gene expression\textsuperscript{18,164-166}. In the crowded nucleus, nonspecific binding can be an important intermediate step for the specific recognition of regulatory
proteins. It was suggested that DNA-binding proteins locate their functional binding sites by nonspecifically binding to DNA and move in reduced dimension, such as 1-D sliding, short-range hopping, and interstrand transfer\textsuperscript{11}. Moreover, the competition between specific and nonspecific or functional or non-functional binding sites can possibly affect the activities of regulatory proteins, such as transcription factors.

Moreover, some nonspecific binding proteins also have sequence preferences that implicate regulatory utilities. For example, single-strand binding proteins (SSBP) and histone proteins mainly nonspecifically bind negatively charged DNA backbones through electrostatic interactions. SSBP binding often results in extended conformation of highly flexible ssDNA molecules, and therefore may prefer DNA sequences that are less coiled and deformed. Similarly, nucleosomes form preferentially at dsDNA sequences that have more flexibility and deformability. Thus, the sequence preference of nonspecific DNA-binding proteins may also have regulatory roles\textsuperscript{11}.

In short, the ultimate goal of characterizing both specific and nonspecific binding is to permit quantitative and mechanistic understanding of molecular regulatory functions.
CHAPTER 6

CONCLUSIONS AND FUTURE DIRECTIONS

6.1 Conclusions

Sixty years after the discovery of the structure of DNA double helix, numerous outstanding questions about how genomic DNA functions in the living cell remain to be answered. Protein-DNA interactions play critical roles in all aspects of genomic functions, but there is no simple coding rule to explain their specific recognition mechanisms. Notably, DNA-binding proteins distinguish functional binding sites through indirect readout of DNA conformational flexibility and deformations that facilitate their regulatory functions in the cell. Hence large-scale systematic investigations of protein-DNA interaction and their conformational specificity are essential to gain a complete insight to the complex and precise molecular regulatory mechanisms.

Traditionally, atomic scale structures of protein-DNA complexes are obtained to reveal important conformational and physical binding mechanisms, which tremendously advanced our understanding of protein-DNA interactions. Technological developments in the last decade stimulated the use of various nontraditional biophysical techniques to study the kinetics and conformational specificity of protein-DNA interactions. However, the wide use and throughput of these techniques are limited by their time-consuming and technical complexities. We thus presented a microarray-based high-throughput platform that enables effective and comprehensive parallel analysis of large-scale protein-DNA binding as well as conformational specificity.
We developed upon existing interferometric sensing techniques to quantify DNA conformational change and protein-DNA binding in parallel in a DNA microarray format. Spectral self-interference fluorescence microscopy (SSFM) allows for measurement of DNA conformational change by determining height changes of fluorophores tagged to specific nucleotides of surface-immobilized DNA molecules. White light reflectance spectroscopy (WLRS) permits quantification of surface molecular density by measuring thicknesses of accumulated biomolecules on the surface. We integrated SSFM and WLRS by designing a sensing substrate and substituting white light illumination source with a LED whose emission spectrum is distinctly separated from that of the fluorescence emission used by SSFM. By implementing a 1-D spectral imaging configuration, the combined system uses two spectral bandwidths in parallel, corresponding to two independent quantitative measurements.

Using the *E. coli* integration host factor protein as the molecular model system, we demonstrated parallel quantitative study of conformational specific protein-DNA interactions. We first presented the basic model for characterization and detection of conformational changes of surface-immobilized DNA. We also presented a quantitative model to resolve conformational specific binding and nonspecific binding. Based on the model, we evaluated factors affecting protein-DNA interactions on a surface, which should be carefully considered when for solid-phase molecular binding assay designs. We further demonstrated distinguished and parallel detection of conformational specific and nonspecific IHF-DNA binding, and discussed the importance of studying both specific
and nonspecific protein-DNA interactions for the understanding of their regulatory functions in the cell.

The capabilities of the presented platform for studying specific and nonspecific protein-DNA interactions can be vastly improved by future engineering efforts in improving system high-throughput and dynamics. However, any technique alone cannot equip us with all the information needed to elucidate protein-DNA interactions and their functions in the complex crowded cellular environment. Eventually, a workflow that includes iterative computational binding site prediction and high-throughput conformational specificity analysis, followed up by crystal structure characterization and \textit{in vivo} functional studies might serve as a viable and collaborative approach to discover the missing components of protein-DNA interactions. We discuss these suggested future technological development and scientific studies in detail in the following subsections.

### 6.2 Future platform developments

#### 6.2.1 Advantages and limitations

We have demonstrated parallel quantification of specific DNA conformational change and protein-DNA binding by integrating SSFM and LED-RS. The parallel detection approach enabled us to simultaneously study conformational specific binding and nonspecific binding, which facilitates understanding of protein-DNA interactions and their functions. The 1-D spectral imaging configuration permits high-throughput measurement of hundreds of DNA spots on one chip. Overall, the platform has been
demonstrated as a versatile tool for the study of conformational specific protein-DNA interaction in a fast and convenient manner.

On the other hand, the platform has limitations. First, SSFM has an inherent limit to its resolution of DNA conformational changes. The detection of DNA conformational change by SSFM is based on the rigid rod model of short surface-immobilized dsDNA molecules, assuming that DNA conformational change results in the height change of the surface-distal end fluorophore of the DNA. We have shown that various factors, such as DNA orientation and length, surface density, binding site location, and protein accumulation, can all potentially reduce the detected fluorophore height changes. To maximize single-to-noise ratio (SNR), these factors need to be optimized, which requires iterative experimentation and characterization. Also, even when LED-RS detects binding while SSFM does not detect fluorophore height changes, we can only conclude that no sharp bending or loops of the DNA occurs, but we cannot conclude whether the DNA undergoes small conformation deformation below SSFM detection limit. Thus, SSFM is only suitable to study protein-DNA complexes where protein binding induces large DNA conformation changes.

Second, the current SSFM platform has some technical limitations. The throughput of the platform is limited. Although the 1-D spectral imaging configuration allows for measurements of multiple DNA spots at the same time, greatly increased measurement efficiency, it still takes tens of measurements to scan the entire microarray, which in turn limits time resolution of dynamic experiments. Moreover, there is a trade-off between fluorescence intensity and number of measurements before fluorophores are
photobleached in SSFM experiments. To obtain sufficient SNR, the integration time for spectral imaging is typically from 1 to 5 seconds, and the total number of spectrum is less than 100 before photobleaching. This trade-off also limits the time-resolution of dynamic SSFM measurements. Finally, detailed kinetic study of protein-DNA interaction is unattainable due to the diffusion-limited flow-cell design and aforementioned real-time measurement limitations. Measurements of bending and unbending rates by SSFM and protein association and dissociation rates by LED-RS can provide extremely useful information for the understanding the biophysical mechanisms of protein-DNA interactions. In the next section, we propose solutions to solve these technical limitations.

### 6.2.2 Improve throughput and time resolution by implementing spectral filtering

To improve the throughput of each spectral acquisition, one potential approach is to implement spectral filtering techniques. In recent years, hyperspectral and spectral imaging techniques have been applied to fluorescence microscopy to improve measurement of FRET efficiency\(^{167}\) and live-cell imaging\(^{168}\). Hyperspectral imaging distinguishes from spectral imaging in terms of better spectral resolution and continuity. To achieve spectral data acquisition, different techniques are employed and can be categorized into four approaches: spatial scanning, spectral scanning, “snapshot” acquisition\(^{169}\), and spatiotemporal scanning\(^{170}\). Spatial scanning is no difference from our current 1-D spectral imaging configuration, snapshot hyperspectral imaging systems require high computational and manufacturing costs, and spatiotemporal scanning requires
moving of the whole system. We thus selected spectral scanning techniques to incorporate to our platform.

To achieve spectral scanning, wavelength tunable filters are commercially available, such as the acousto-optic tunable filter (AOTF). Several parameters of these spectral filters need to be carefully considered, including diffraction efficiency, wavelength tuning range, tuning speed, spectral resolution, and maximum beam dimension. The commercially available AOTFs can have wavelength tuning ranges from 200 nm to 1000 nm, diffraction efficiencies from 10% to 90%, tuning speeds from 4 to 35 microseconds. The spectral resolution varies from a few to tens of nanometers as a function of the square of the wavelength (Figure 69).

![Figure 69](Image)  
Figure 69. An example of the spectral resolution of acousto-optic tunable filter as function of the square of the wavelength. (Data from Panasonic acousto-optic tunable filter type EFLF specifications)
To implement spectral filtering to the combined SSFM and LED-RS system, detailed analysis of the effect of spectrum resolution and SNR are necessary for accurate determination of fluorophore axial heights and biomass density. The spectrum resolution also has a trade-off with the scanning speed over the fluorescence and LED spectrums for measuring protein-DNA binding kinetics. The throughput for each acquisition can be calculated by dividing the maximum beam size of the filter by the DNA spot center-to-center distance (pitch). For instance, a 5 x 5 mm beam size divided by 0.2 mm pitch renders 625 spots in each spectral image. Scanning two 60 nm (1000 cm⁻¹) spectral bandwidths with a 2 nm spectral resolution at a tuning speed of 10 microseconds requires 0.6 seconds. Suppose the integration time is 1 second, a coarse calculation using spectral filtering renders measurement of over 600 spots within 2 seconds for each SSFM and LED-RS spectral image acquisition. A tentative optical setup configuration is schematically illustrated in Figure 70.
6.2.3 Quantum Dots as substitutes for organic fluorophores

To further increase the time-resolution of SSFM, we need to reduce the integration time by either improving fluorescence collection efficiency or using brighter emitters. This is not an issue for LED-RS, for commercially available high-power LEDs are sufficient to provide high-intensity narrow bandwidth light for our purpose. For SSFM, non-blinking quantum dots\textsuperscript{171,172} possessing narrow non-overlapping spectrums...
can be a good substitute for the fluorophores in SSFM measurements. QDs have extraordinary brightness and photostability as compared to organic fluorophores, which have already been exploited in cell imaging and biological sensing\textsuperscript{173,174}. The emission of different QDs can be excited at the same laser wavelength, and the narrow distinguished bandwidths of QDs potentially permit simultaneous determination of multiple axial localizations. One potential concern of using QDs for SSFM measurements is the relatively large sizes of QDs to the DNA molecules might affect the resolution of conformational change detection. This might require using longer DNA sequences and applying external forces to orient surface-immobilized DNA molecules. Conjugation techniques of QDs and DNA and protein molecules are commercially available. In short, high-efficiency QDs combined with spectral filtering can be a viable approach to achieve sub-second resolution monitoring of protein-DNA binding kinetics.

6.2.4 Improving fluidic system design

The current flow cell employed for real-time hybridization measurements and IHF-DNA binding experiments operated in a diffusion-limited regime\textsuperscript{140}. The observed association kinetics of DNA binding and conformational change were slow and indistinguishable, constrained by the mass transport rate of the binding molecules to the surface-immobilized DNA probes. Moreover, in a high-throughput real-time measurement where a large-scale dense DNA microarray are detected simultaneously, the concentration of target protein molecule near the DNA surface may not be uniform along the direction of flow due to the depletion zones formation, which affect the measured
binding rates\textsuperscript{140}. Hence, for kinetic binding studies, the flow cell needs to be optimized to a regime where the rates of conformational change and binding are uniform across the chip to help determine the biophysical mechanisms of the reaction. A potential approach is to optimize the flow cell dimensions and introduce mixing or stirring strategies to overcome mass transport limit and bring uniform protein concentration to the DNA microarray surface to reveal true reaction kinetics.

The approaches discussed above only allow higher throughput of DNA probes. To simultaneously study of multiple proteins, our platform needs to be combined with microfluidic techniques to perform multi-channel assays. Such multiplexed detection of protein-DNA interactions not only increases throughput, but allows for simultaneous comparison of the binding affinities and specificities of various DNA-binding proteins.

6.3 Future studies in protein-DNA interactions

6.3.1 High-throughput systematic screening of regulatory binding sites

Perhaps the most intriguing regulatory proteins are transcription factors that bind specifically to functional binding sites and control transcription. One of the most important proteins in transcriptional regulation is TATA-binding protein (TBP), which is the first protein to binding to the transcription promoters to initiate transcription. Binding of TBP to the canonical TATA-box sequences induces a ~80° bending of the DNA helix mainly through minor groove interactions\textsuperscript{10,175} (Figure 2). This severe bending is stabilized by further interactions with flanking sequences and recruitment of other transcription factors\textsuperscript{175}. Recent studies have provided evidence that TBP can bind to
nontraditional binding sites other than the canonical TATA-box sequences, but with lower affinities\textsuperscript{176,177}. It was suggested that DNA conformational deformability or DNA bending might affect the functional distinction of TBP between TATA-box containing promoters and nontraditional TATA-less promoters\textsuperscript{178}. However, whether TBP induces bending to nontraditional TATA-less promoters remains undecided. Therefore, a systematic genome wide screening and analysis of TBP binding to potential TATA-less promoter sequences and their conformational changes could help advance our understanding of transcriptional regulation.

\textbf{6.3.2 Single-strand binding protein-DNA nonspecific binding}

Previously we have been using the rigid rod model of dsDNA and SSFM to measure DNA conformational changes that lower DNA surface-distal end fluorophore heights. On the other hand, we have shown detection of conformational change of ssDNA from random coils to rigid dsDNA during hybridization. Thus, besides dsDNA bending, we can also use SSFM to measure ssDNA stiffening, which elevates ssDNA surface-distal fluorophore heights. In particular, single-strand binding proteins (SSBP) nonspecifically binding to ssDNA segments resulted from transient duplex melting in the genome through electrostatic interactions. SSBP binding protects these ssDNA segments from nucleases digestion, premature annealing, or to remove secondary structures to facilitate other proteins during DNA replication, recombination, and repair (Figure 71). It was suggested that SSBP binding results in extended ssDNA conformation and also have binding preferences to ssDNA segments that require less conformational change to
achieve such extension. Thus conformational-based sequence preference may have regulatory meanings that need to be investigated. Therefore, a comprehensive screening and analysis of SSBP binding to ssDNA with various flexibility or secondary structures could help us elucidate SSBP functions and regulatory roles in cellular processes.

Figure 71 SSBP binding during DNA replication to avoid degradation and base-pairing of the exposed ssDNA.

6.3.3 A proposed workflow

A wide range of mechanistic and conformational behaviors of protein-DNA interactions concerning their biological functions remains to be fully understood. Our platform designed to achieve parallel quantification of DNA conformational specificity and protein binding provides researchers a convenient tool to screen and investigate conformational specific protein-DNA interactions. Nonetheless, the capability of the platform is still limited unless additionally combined with other \textit{in vivo}, \textit{in vitro}, and
computational techniques. We propose a workflow where the platform enhances effective and efficient study of the biophysical mechanisms and functions of conformational specific protein-DNA interactions. First, computational or bioinformatics techniques that predict sequence preferences of DNA-binding proteins are used to select candidate protein-binding sites in the genome. Then, high-throughput SSFM and LED-RS are used to examine the binding affinity and conformational specificity of targeted protein and candidate binding sites. The computational prediction and mechanistic modeling of protein-DNA interfaces can be performed iteratively with SSFM and LED-RS equilibrium and kinetics measurements to study in detail of the specific binding mechanisms. The first two steps in the workflow generate a small number of conformational specific protein-DNA complexes whose structures can be further analyzed by X-ray crystallography or NMR spectroscopy at the atomic scale. Alternatively, the functions of these protein-DNA complexes can be examined by site-specific mutagenesis studies in the cell to understand their regulatory roles in vivo.

Figure 72 A proposed workflow for effective and efficient study of conformational specific protein-DNA interactions and their functions.
BIBLIOGRAPHY


47. Ali, B. J.; Amit, R.; Braslavsky, I.; Oppenheim, A. B.; Gileadi, O.; Stavans, J. Compaction of Single DNA Molecules Induced by Binding of Integration Host


92. Akkahat, P.; Mekboonsoonglarp, W.; Kiatkamjornwong, S.; Hoven, V. P. Surface-Grafted Poly(Acrylic Acid) Brushes as a Precursor Layer for Biosensing


CURRICULUM VITAE

XIRUI ZHANG

Contact Information
Department of Biomedical Engineering
44 Cummington Mall,
Boston, MA 02215
Voice: (617) 816-7516
Email: xirui.zhang0@gmail.com

Education
2009-2014  Ph.D. and M.S. in Biomedical Engineering
College of Engineering, Boston University, Boston, MA
Advisor: Professor M. Selim Ünlü

2005-2009  B.S. in Biomedical Engineering and B.A. in Scientific English
Tianjin University, Tianjin, China

Research Experience
2010-2014  Graduate Research Assistant
Optical Characterization and Nanophotonics Laboratory
Boston University Photonics Center, Boston, MA
Advisor: Professor M. Selim Ünlü

August 2011  Visiting Scholar
Walter Schottky Institut
Technische Universität München, Garching, Germany
Collaborator: Professor Ulrich Rant

2009-2010  Biomedical Engineering Fellow
Biomedical Optics Laboratory, Boston University, Boston, MA
Advisor: Professor Irving Bigio

2008-2009  Undergraduate Research Assistant
Tianjin Key Laboratory of Biomedical Detecting Techniques and Instruments, Tianjin University, Tianjin, China
Advisor: Professor Feng Gao
Teaching Experience

Fall 2011  Teaching Assistant
Molecular Bioengineering, Boston University (Graduate class of 30)

Spring 2011  Teaching Assistant
Introduction to Probability, Boston University (Undergraduate class of 80)

Awards and Fellowship

January 2014  National Radio Science Meeting Travel Award
2009-2010  Distinguished Biomedical Engineering Fellowship, Boston University
2007-2009  National Scholarship, Tianjin University

Selected Publications

Book Chapters


Journal Publications


- S. Ahn, C. L. Huang, E. Ozkumur, X. Zhang, J. Chinnala, A. Yalcin, S. Bandyopadhyay, S. Russek, M. S. Ünlü, C. DeLisi. TATA binding proteins can


- For a complete list of publications see: 

**Selected Conference Proceedings**


