The fields of biological and medical physics and biomedical engineering are broad, multidisciplinary and dynamic. They lie at the crossroads of frontier research in physics, biology, chemistry, and medicine. The Biological and Medical Physics, Biomedical Engineering Series is intended to be comprehensive, covering a broad range of topics important to the study of the physical, chemical and biological sciences. Its goal is to provide scientists and engineers with textbooks, monographs, and reference works to address the growing need for information.

Books in the series emphasize established and emergent areas of science including molecular, membrane, and mathematical biophysics; photosynthetic energy harvesting and conversion; information processing; physical principles of genetics; sensory communications; automata networks, neural networks, and cellular automata. Equally important will be coverage of applied aspects of biological and medical physics and biomedical engineering such as molecular electronic components and devices, biosensors, medicine, imaging, physical principles of renewable energy production, advanced prostheses, and environmental control and engineering.

Editor-in-Chief:
Elias Greenbaum, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

Editorial Board:
Masuo Aizawa, Department of Bioengineering, Tokyo Institute of Technology, Yokohama, Japan
Olaf S. Andersen, Department of Physiology, Biophysics & Molecular Medicine, Cornell University, New York, USA
Robert H. Austin, Department of Physics, Princeton University, Princeton, New Jersey, USA
James Barber, Department of Biochemistry, Imperial College of Science, Technology and Medicine, London, England
Howard C. Berg, Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts, USA
Victor Bloomfield, Department of Biochemistry, University of Minnesota, St. Paul, Minnesota, USA
Robert Callender, Department of Biochemistry, Albert Einstein College of Medicine, Bronx, New York, USA
Britton Chance, Department of Biochemistry/ Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania, USA
Steven Chu, Lawrence Berkeley National Laboratory, Berkeley, California, USA
Louis J. DeFelice, Department of Pharmacology, Vanderbilt University, Nashville, Tennessee, USA
Johann Deisenhofer, Howard Hughes Medical Institute, The University of Texas, Dallas, Texas, USA
George Feher, Department of Physics, University of California, San Diego, La Jolla, California, USA
Hans Frauenfelder, Los Alamos National Laboratory, Los Alamos, New Mexico, USA
Ivar Giaever, Rensselaer Polytechnic Institute, Troy, New York, USA
Sol M. Gruner, Cornell University, Ithaca, New York, USA

For other titles published in this series, go to www.springer.com/series/3740
Contents

Part I  Binding DNA and Changing its Shape

1  Introduction to Biophysics of DNA–Protein Interactions: 
   From Single Molecules to Biological Systems ........................................... 3
   Mark C. Williams and L. James Maher, III

2  Single-Molecule Imaging of LacI Diffusing Along 
   Nonspecific DNA ....................................................................................... 9
   Y.M. Wang and R.H. Austin

3  How Proteins Slide on DNA ................................................................. 39
   Daniel Barsky, Ted A. Laurence, and Ėslovas Venclovas

4  Role of DNA Conformations in Gene Regulation ................................. 69
   Ralf Metzler, Bram van den Broek, Gijs J.L. Wuite, 
   and Michael A. Lomholt

5  Recognition of Nucleic Acids by Transcription Factor NF-κB .......... 85
   Gourisankar Ghosh, De-Bin Huang, and Tom Huxford

6  Dynamics and Mechanism of DNA-Bending Proteins 
   in Binding Site Recognition ................................................................... 107
   Anjum Ansari and Serguei V. Kuznetsov

7  Studies of Sequence-Nonspecific HMGB 
   DNA-Binding Proteins .......................................................................... 143
   L. James Maher, III

8  DNA Interactions with Single-Stranded DNA Binding 
   Proteins and Retroviral Nucleic Acid Chaperones by 
   Force Spectroscopy ................................................................................ 163
   Mark C. Williams and Ioulia Rouzina
## Part II Complex Protein-DNA Interactions

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>DNA Looping in Prophage Lambda: New Insight from Single-Molecule Microscopy</td>
<td>193</td>
</tr>
<tr>
<td></td>
<td>Laura Finzi, Carlo Manzo, Chiara Zurla, Haowei Wang, Dale Lewis, Sankar Adhya, and David Dunlap</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Dynamics of Protein–ssDNA Interactions in the Bacteriophage T4 Homologous Recombination System</td>
<td>213</td>
</tr>
<tr>
<td></td>
<td>Jie Liu and Scott W. Morrical</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Polymerase Switching in Response to DNA Damage</td>
<td>241</td>
</tr>
<tr>
<td></td>
<td>Jaylene N. Ollivierre, Michelle C. Silva, Jana Sefcikova, and Penny J. Beuning</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Topological Probes of a Cooperative, Nonspecific Protein–DNA Complex</td>
<td>293</td>
</tr>
<tr>
<td></td>
<td>Claire A. Adams, Manana Melikishvili, and Michael G. Fried</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>How to Think Like a Single Molecule: Obtaining Quantitative Measurements on Single DNA Molecules and Chromatin Fibers</td>
<td>307</td>
</tr>
<tr>
<td></td>
<td>Sanford H. Leuba and Richard A. Steinman</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Long-Range Chromatin Interactions in Cells</td>
<td>325</td>
</tr>
<tr>
<td></td>
<td>Guo Ling Zhou, Li Xin, and De Pei Liu</td>
<td></td>
</tr>
</tbody>
</table>

Index                                                                 | 341  |
Contributors

Claire A. Adams  
Department of Microbiology and Immunology,  
Center for Structural Biology, University of Kentucky, Lexington, KY 40536, USA

Sankar Adhya  
Laboratory of Molecular Biology, NCI, NIH, 37 Convent Drive, Bethesda, MD 20892-4264, USA

Anjum Ansari  
Physics Department, M/C 273, University of Illinois at Chicago,  
845 W. Taylor St., Chicago, IL, 60607 USA

R.H. Austin  
Department of Physics, Princeton University, Princeton, NJ, 08544

Daniel Barsky  
Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory,  
7000 East Avenue, Livermore, CA 94550, USA

Penny J. Beuning  
Department of Chemistry & Chemical Biology, Center for Interdisciplinary Research on Complex Systems, Northeastern University,  
360 Huntington Ave 102 Hurtig Hall, Boston, MA 02115, USA

Bram van den Broek  
Physics of Life Processes, Leiden Institute of Physics, Leiden University,  
Niels Bohrweg 2, 2333 CA, Leiden, The Netherlands

David Dunlap  
Cell Biology Department, Emory University, 615 Michael St.,  
Atlanta, GA 30322, USA

Laura Finzi  
Department of Physics, Emory University, 400 Dowman Dr.,  
615 Michael St., Atlanta, GA 30322, USA
Contributors

Michael G. Fried  
Center for Structural Biology, Department of Molecular and Cellular Biochemistry, University of Kentucky, 741 South Limestone, Lexington, KY 40536-0509, USA  
michael.fried@uky.edu

Gourisankar Ghosh  
Department of Chemistry and Biochemistry, University of California, 9500 Gilman Drive, La Jolla, San Diego, CA 92093-0375, USA

De-Bin Huang  
Department of Chemistry & Biochemistry, University of California, 9500 Gilman Drive La Jolla, San Diego, CA 92093-0375, USA

Tom Huxford  
Department of Chemistry & Biochemistry, San Diego State University, 5500 Campanile Drive, San Diego, CA, 92182-1030, USA

Serguei V. Kuznetsov  
Physics Department, M/C 273, University of Illinois at Chicago, 845 W. Taylor St., Chicago, IL 60607, USA

Ted A. Laurence  
Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA

Sanford H. Leuba  
Departments of Cell Biology and Physiology and Bioengineering, Hillman Cancer Center, Petersen Institute of NanoScience and Engineering and University of Pittsburgh Cancer Institute, University of Pittsburgh School of Medicine and Swanson School of Engineering, 5117 Centre Avenue, 2.26a, Pittsburgh, PA 15213, USA

Dale Lewis  
Laboratory of Molecular Biology, NCI, NIH, 37 Convent Drive, Bethesda, MD 20892-4264, USA

De Pei Liu  
National Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing100005, People's Republic of China

Jie Liu  
Section of Microbiology, University of California-Davis, Davis, CA 95616, USA

Michael A. Lomholt  
Department of Physics and Chemistry, MEMPHYS – Center for Biomembrane Physics, University of Southern Denmark, Campusvej 55, 5230, Odense M, Denmark
Contributors

L. James Maher, III
Department of Biochemistry and Molecular Biology,
Mayo Clinic College of Medicine, Rochester, MA, USA

Carlo Manzo
Institute for Bioengineering of Catalonia, Barcelona, Spain

Manana Melikishvili
Department of Molecular and Cellular Biochemistry, Center for Structural Biology, University of Kentucky, Lexington, KY 40536, USA

Ralf Metzler
Physics Department T30g, Technical University of Munich, 85747 Garching, Germany

Scott W. Morrical
Department of Biochemistry, University of Vermont College of Medicine,
Burlington, VT 05405, USA

Jaylene N. Ollivierre
Department of Chemistry & Chemical Biology, Northeastern University,
Boston, MA 02115, USA

Ioulia Rouzina
Department of Biochemistry, Molecular Biology, and Biophysics,
University of Minnesota, Minneapolis, MN, USA

Jana Sefcikova
Department of Chemistry & Chemical Biology, Northeastern University,
Boston, MA 02115, USA

Michelle C. Silva
Department of Chemistry & Chemical Biology, Northeastern University,
Boston, MA 02115, USA

Richard A. Steinman
Department of Medicine, Hillman Cancer Center, University of Pittsburgh School of Medicine, University of Pittsburgh Cancer Institute, 5117 Centre Avenue, Pittsburgh, PA 15213, USA

Česlovas Venclovas
Institute of Biotechnology, Graičiūno 8, LT-02241 Vilnius, Lithuania

Haowei Wang
Physics Department, Georgia Institute of Technology,
400 Dowman Dr, Atlanta, GA 30322, USA

Y.M. Wang
Department of Physics, Washington University in St. Louis,
Saint Louis, MO, 63130
Mark C. Williams  
Department of Physics, Northeastern University, Boston, MA, USA

Gijs J.L. Wuite  
Department of Physics and Astronomy, Faculty of Sciences, VU University,  
De Boelelaan 1081, 1081 HV, Amsterdam, The Netherlands

Li Xin  
251 Bayview Boulevard, Suite 100, Baltimore, MD 21224-2815, USA

Guo Ling Zhou  
Department of Genetics, Harvard Medical School, Center for Computational  
and Integrative Biology, Massachusetts General Hospital, 185 Cambridge Street,  
Boston, MA 02114-2790, USA

Chiara Zurla  
Department of Bioengineering, Georgia Institute of Technology, Atlanta, GA, USA
Part I

Binding DNA and Changing its Shape
Chapter 1
Introduction to Biophysics of DNA–Protein Interactions: From Single Molecules to Biological Systems

Mark C. Williams and L. James Maher, III

The field of biophysics grows as new methods are applied to increasingly complex biological problems. For example, the field of single molecule biophysics has made new contributions to many areas of biology, from molecular to cellular levels. These insights enhance our quantitative understanding of biology. Recent studies that combine in vitro bulk biochemical methods, single molecule experiments, and in vivo approaches to molecular interactions have advanced many areas of biology. Particularly relevant illustrative examples are provided by the field of DNA–protein interactions. This book focus on some of these examples, involving application of in vitro bulk solution methods to probe interactions at the molecular level, single molecule methods, and in vivo measurements in the same systems. The book presents an overview of modern quantitative methods used for studying DNA–protein interactions, and displays the various levels of complexity involved in studying these interactions in biological systems. These chapters should therefore be of interest to all workers in the field of DNA–protein interactions.

Section 1 covers DNA–protein interactions that involve only one or two proteins, illustrating approaches to deduce mechanistic insights into the biophysical principles at play in these “simple” interactions. One important question concerns the mechanism by which binding proteins search for and identify specific binding sites. This mechanism is important for many proteins that recognize specific sequences and bind to them to regulate cellular processes. The most commonly studied example is that of a transcription factor, which recognizes specific sequences that regulate transcription of a gene. It is now appreciated that many DNA-binding proteins reach their specific sites on DNA using mechanisms more complex than
three-dimensional random diffusion from solution. Alternative mechanisms involve a reduction of the volume to be searched. This volume reduction likely comes from a sequence-nonspecific DNA-binding mode allowing diffusion in one dimension along the DNA. Thus, Wang and Austin in Chap. 2 describe experiments to measure how the model Lac repressor protein searches in one dimension to find a specific binding site. In their study, they directly image a fluorescent Lac repressor molecule as it searches a single DNA molecule, remaining close to the DNA for long periods of time. They also address the question of how the protein molecule slides along the DNA – Does the protein remain in DNA contact and truly slide, or does it exhibit small micro-hops in order to sample DNA-binding states? While this question has not yet been conclusively answered, it may be possible to test predictions of different hypothetical mechanisms using further developments in fluorescence imaging methodology or other single molecule methods.

The mechanism of DNA sliding is further addressed in Chap. 3 by Barsky et al., which discusses how it is physically possible to slide along DNA and what biophysical aspects of DNA–protein interactions may affect sliding rates. This chapter also provides a concise overview of different systems in which DNA sliding may be important. These systems extend well beyond the canonical transcription factor search problem. For example, the processivity clamp is a critical component of the DNA replication fork, and this protein encircles DNA and must slide at high rates within the replication complex.

Beyond DNA sliding mechanisms, Chap. 4 by Metzler et al. discusses the advantages of sliding compared to searching the fluid volume. While it is clear that some enhancement of the DNA search rate is obtained by searching only the DNA itself, the one-dimensional search process proceeds slowly, as diffusion along DNA is slower than diffusion in solution. Measuring one-dimensional diffusion in single molecule experiments typically involves stretched DNA. However, it is clear that DNA in solution and in vivo is always compacted by naturally entropic elasticity and packaging proteins. How does DNA compaction affect the DNA search time? It has been shown that the theoretical optimum search time must involve both one-dimensional diffusion along DNA as well as some diffusion in solution. This chapter further describes single molecule experiments that directly demonstrate compaction effects by measuring protein search times while the DNA tension is varied in an optical tweezers instrument. When the DNA is under less tension, proteins hop quickly from one part of the molecule to another, either through solution or by direct transfer from one DNA segment to another part. To measure diffusion, the authors determine the time required for restriction endonuclease cleavage of the target DNA molecule. The authors demonstrate that the rate of restriction endonuclease digestion is increased on more compact DNA molecules under little tension than when DNA is stretched. This suggests that the localization of a specific sequence by a restriction endonuclease occurs through a combination of sliding along DNA and other diffusion mechanisms.

Once a protein arrives at a specific DNA site, recognition of that site is a complex problem that is not yet well understood from a biophysical standpoint. In Chap. 5, Ghosh et al. describe exemplary X-ray crystallography experiments that probe in detail how members of the dimeric NF-κB transcription factor family recognize
related DNA-binding site sequences. These authors have compared structures in atomic detail for an extensive array of different complexes. This provides a nearly unique depth of understanding for the NF-κB family of proteins. Emphasized by this work is the fact that the binding site sequence, in itself, may be insufficient to guide homing to unique sites. Additional specification is likely provided by cooperative contacts with additional bound proteins. The authors touch on the interesting observation that transcription factors such as NF-κB can recognize folded RNA molecules with sequence and structure specificity.

It is increasingly understood that DNA site recognition may involve mutual structural reorganization by the protein and DNA partner. This reorganization can dramatically alter the dynamic structure and trajectory of DNA. Two chapters are devoted to these issues. In Chap. 6, Ansari and Kuznetsov describe extremely fast kinetic experiments measuring the dynamics of sequence-specific DNA-bending proteins during their interaction with DNA. At issue is the mechanistic question of whether site binding and DNA bending are concerted or separable events. Laser-induced temperature jump relaxation kinetics with fluorescence resonance energy transfer detection has been applied to this fascinating problem.

While the mechanism by which sequence-specific binding proteins locate and recognize their DNA sites is well studied and of great interest, many proteins that bind DNA do not recognize specific double-stranded DNA (dsDNA) sequences. For example, a class of DNA-bending proteins known as high mobility group type B (HMGB) proteins bind dsDNA at any sequence and bend the DNA at that location. The biophysics of the bending process is of particular interest because protein binding changes the character of DNA, endowing DNA with a collection of rapidly interconverting new shapes. In effect, the DNA appears to be more flexible in the presence of HMGB proteins. In Chap. 7, Maher reviews the biophysics of HMGB proteins, including in vivo data from bacterial experiments. This work indicates that so-called architectural DNA-binding proteins (e.g., the bacterial HU protein, or HMGB proteins introduced experimentally into bacteria from higher cells) play important roles in regulating genetic switches that require strong DNA bending and looping.

Rather than recognizing a specific sequence of dsDNA, single-stranded DNA (ssDNA)-binding proteins (SSBs) recognize DNA in its denatured form. Because ssDNA is so different structurally from dsDNA, the ability of these proteins to distinguish ssDNA is not difficult to imagine. These proteins typically contain oligonucleotide/oligosaccharide binding folds (OB-folds) that facilitate simultaneous stacking of aromatic residues with ssDNA bases as well as cationic binding to the ssDNA backbone. However, the proteins must locate and bind ssDNA that occurs transiently during DNA replication or repair. For example, ssDNA is transiently exposed by helicases and other motor proteins at the replication fork. In the case of bacteriophage replication, the fork moves extremely rapidly, so bacteriophage SSBs must search DNA optimally in order to facilitate rapid DNA replication. In Chap. 8 Williams and Rouzina describe the biophysical mechanisms that govern the interactions of SSBs from bacteriophage T4 (T4 gene 32 protein) and T7 (T7 gene 2.5 protein), as determined through single DNA molecule stretching experiments, combined with the results of other biochemical measurements. They describe how these
SSBs search dsDNA primarily via one-dimensional diffusion along the dsDNA molecule. Because high concentrations of these proteins are present in the infected cell, many proteins simultaneously search dsDNA, resulting in a rapid process that requires little contribution from solution diffusion. Thus, SSBs bind initially to dsDNA, sliding to locate transient sites of ssDNA as DNA replication forks propagate along the molecule. These authors also discuss the mechanism by which both T4 gene 32 protein and T7 gene 2.5 protein binding to DNA is regulated by salt-dependent conformational changes involving their C-termini, a common protein-binding regulatory mechanism. The authors go on to compare traditional replication SSBs from bacteriophages to retroviral nucleocapsid (NC) proteins, which also bind preferentially to ssDNA. However, retroviral NC proteins differ significantly from SSBs in that their overall preferential binding to ssDNA relative to dsDNA is weak, and they therefore only weakly destabilize dsDNA. Termed “nucleic acid chaperones,” these proteins facilitate single-stranded nucleic acid folding and secondary structure rearrangement during the retroviral replication process. Single molecule studies are ideal for examining the mechanism by which these proteins facilitate nucleic acid melting and reannealing. The described experiments illustrate the capability of single molecule force spectroscopy to measure DNA binding thermodynamics and kinetics, which quantitatively elucidate the biophysics of DNA–protein interactions.

Thus, the first 8 chapters describe biophysical mechanisms by which individual proteins interact with DNA in simplified experimental system. Increasingly relevant (but less tractable) systems attempt to mimic multi-component complexes as found in living cells. Studies of single proteins interacting with DNA reveal the mechanisms of individual proteins, but it is desirable to understand these results in the context of more complex biological systems. The second half of the book describes examples of these complex protein–DNA interactions in the context of important biophysical processes.

Chapter 9 by Finzi et al. reports a detailed biophysical analysis of the molecular mechanism of the bacteriophage lambda epigenetic switch, which regulates the choice between the quiescent lysogenic state and the violently reproductive lytic state in the host cell. This switch mechanism is a model for other more complex regulatory processes. It involves the interaction of multiple protein clusters positioned at a distance along DNA. Using single molecule tethered particle methods, Finzi et al. characterize the functional interactions between a single DNA molecule and the lambda repressor protein (CI). The authors characterize the evidence for protein–protein interactions tethered by a long segment of looped viral DNA. Estimation of kinetic parameters permits characterization of some of the biophysical features of this complex system.

In Chap. 10, Liu and Morrical describe some of the complex DNA–protein interactions involved in bacteriophage T4 recombination processes. During late stages of bacteriophage infection, replication occurs rapidly through recombination, allowing replication of several phage genomes to occur in parallel. This recombination-dependent replication involves multiple protein–DNA and protein–protein interactions, including both DNA-binding proteins, such as T4 UvsY and T4 gene 32 protein, as well as DNA motors such as the helicase UvsX.
These proteins work together to convert SSB-coated homologous ssDNA strands into a presynaptic filament competent for the recombination events that are required for additional replication forks to be created. The chapter describes how characteristics of the DNA–protein interactions, such as preferentially binding to wrapped or stretched ssDNA or dsDNA, facilitate transitions during the recombination-dependent replication process.

While recombination-dependent replication in bacteriophage T4 involves the reorganization of several DNA–protein interactions, the *Escherichia coli* replication fork presents a dauntingly complex system of interactions between proteins and DNA. Bacterial DNA replication is much more complex and regulated than bacteriophage replication, and the required proteins are in constant exchange as replication proceeds. In Chap. 11, Ollivierre et al. describe interactions between several of the *E. coli* replication proteins that form the holoenzyme, the core protein complex that replicates the bacterial genome. This chapter describes examples of the molecular details of protein–protein and protein–DNA interactions involved in replication. The focus is placed on the replication of damaged DNA. Here regulated DNA polymerase switching is observed. In this process, specialized DNA polymerases are substituted for the replicative polymerase in order to copy across the damaged DNA template and allow replication to continue past a lesion. The authors briefly describe much more complex DNA replication processes in eukaryotic cells. The results illustrate how a wide array of biochemical and biophysical methods, as well as complementary in vivo measurements, are needed to shed light on the extremely complex biological mechanisms required for DNA replication in living systems.

The concept of DNA damage recognition and repair in eukaryotes is again emphasized by Adams et al. in Chap. 12. The authors discuss their pioneering work on the entirely sequence-nonspecific repair protein O⁶-alkylguanine-DNA-alkyltransferase (AGT). This protein reverses DNA damage through an irreversible transfer of the DNA alkyl lesion to a protein side chain. Using clever biophysical methods including protein–protein cross-linking and analysis of changes in DNA topology, the authors build a compelling model for a highly cooperative protein–DNA filament. As with all good models, the work raises many new questions about the mechanism by which this sequence-nonspecific protein functions.

In eukaryotic cells, processes such as replication and transcription, replete with protein–protein and protein–DNA interactions as discussed above, must also occur in the context of highly packaged genomic DNA. Eukaryotic DNA is packaged into chromatin, a highly organized structure with multiple levels of compaction. The exact structure of biological chromatin is not known in detail, but many aspects are beginning to be understood. It is known, for example, that the first level of DNA compaction involves histone octamers around which DNA is tightly wrapped. The forces required to disrupt nucleosome structures have been quantified, and monomeric nucleosome core particles are thermodynamically stable in vitro. Higher order chromatin structure is not very well understood, and in vivo chromatin dynamics remain mysterious. In Chap. 13, Leuba and Steinman introduce single molecule experiments on artificial chromatin structures, which may shed light on their dynamics.
Finally, Zhou et al. describe in Chap. 14 the application of low-resolution but extremely powerful genome-wide methods to explore the overall organization of eukaryotic DNA within the nuclei of cells. Practitioners of the Chromatin Conformation Capture (3C) method, these authors summarize how this approach, and its new relatives, can be applied to determine which sequences are held close together in three-dimensions within the nuclear compartment. These methods provide semi-quantitative measurements, ultimately promising a series of distance constraints that might do for cell biology what the Nuclear Overhauser Effect has done for nuclear magnetic resonance spectroscopy. The present chapter surveys the concept of long-range DNA regulatory loops, reviews classic examples, identifies key implicated proteins, and offers a look at where the future may lead in solving this ultimate problem in protein–DNA biophysics.
2.1 Introduction

Transcription factors, restriction enzymes, and RNA polymerases are proteins that function by binding to their specific target sites on DNA [1, 2]. The DNA targets for these proteins are typically a few tens of base pairs long, while the chromosomes contain over a million base pairs of DNA (E. coli, for example, has 4.6 million base pairs); therefore, before reaching their targets, it is inevitable that DNA-binding proteins encounter nonspecific DNA first. In this process, protein–nonspecific-DNA binding does occur (although with weaker affinity than DNA target binding [3]) and this interaction affects the specific-DNA targeting rate of the protein. In order to regulate the targeting rate of DNA-binding proteins, which is an important step for gene expression regulations, the mechanisms of protein interaction with nonspecific DNA must be elucidated.

The notion that nonspecific DNA influences the targeting rate of DNA-binding proteins gained renewed credence when faster-than-diffusion target binding of LacI (Lactose repressor protein) was observed in 1970. LacI binds to its lactose operator target (lacO) 100 times faster than allowed by the 3D diffusion limit [4]. This faster-than-diffusion binding was explained by the facilitated-diffusion model, in which a DNA-binding protein interacts with nonspecific DNA before reaching its target [5–18].

A facilitated-diffusion process is composed of three main types of protein’s motion around DNA: sliding, where the protein translocates (or slides) along nonspecific DNA base pairs without losing contact (Fig. 2.1, green arrows); hopping, where the protein dissociates from DNA briefly, performing free 3D diffusion, and lands back on DNA at a location that is shorter than the DNA’s persistence length away from the dissociation site (Fig. 2.1, black arrows); and jumping, where the protein’s DNA landing location is not correlated to the dissociation site [15].
Figure 2.1 shows that for the typical elongated DNA geometry used in single-molecule fluorescence imaging experiments, a protein performs a series of alternating sliding and hopping motions along DNA. In a cell, the targeting process can be further complicated by road blocks of other bound proteins on DNA and different configurations of the chromosomal DNA due to confinement [19–21]. In this chapter, we focus on sliding and hopping diffusions of proteins on DNA since most single-molecule fluorescence imaging experiments use low concentrations of elongated DNA molecules for which jumping and in vivo effects are absent.

In order to calculate the effects of facilitated diffusion on the targeting rate of proteins, all three components of the facilitated-diffusion process as well as the switching kinetics among the different components should be quantified. Hopping and jumping motions are 3D Brownian motions of proteins in water, and the dynamics of 3D Brownian motions are relatively well understood, in contrast to sliding. For the sliding motion, two major issues await investigation due to limitations of current single-particle-tracking techniques: (1) The translocation (or sliding) mechanism of a protein on DNA sequences is not fully understood. There are two main translocation models: one is the DNA-sequence-independent model, where proteins are insensitive to DNA sequences and perform Brownian diffusion on DNA; the other is the DNA-sequence-dependent model, where the sliding diffusion characteristics are time dependent – anomalous subdiffusion in short millisecond timescales and Brownian diffusion in long second timescales [22]. (2) The mean sliding time and hopping time, or the protein-nonspecific-DNA dissociation rate constant (or nonspecific-DNA dissociation time, or sliding time) \( t_d \) and 3D hopping time \( t_{hop} \), respectively, are not precisely known. Uncertainties in these parameters will decrease the accuracy of protein targeting rate calculations.

This chapter emphasizes studies of the sliding component of facilitated diffusion (and along with it the sliding–hopping alternation kinetics) using single-molecule fluorescence imaging methods. LacI is used as a model protein, and DNA molecules are elongated along fused-silica surfaces. The elongated DNA configuration offers a simple platform for revealing the dynamics of protein translocation on DNA base
pairs during sliding, which is essential for understanding facilitated diffusion for any DNA configurations in cells. Why use the single molecule method? Conventional bulk measurements have been very powerful in validating and characterizing facilitated diffusion [4, 8, 10, 13, 16, 18, 23–25]; however, in order to characterize the sliding mechanisms of proteins on DNA, the motion of proteins on DNA has to be tracked. Thus, single-molecule imaging is an ideal candidate for these studies.

This chapter is divided into three parts: First, we present our initial study of LacI diffusion on nonspecific DNA in the timescale of seconds performed in 2006. Prior to this study, two articles had reported on the direct imaging of proteins diffusion on DNA. One was by Nobuo Shimamoto [12] and the other was by Yoshie Harada et al. [26]. In the first article, it is not clear whether the proteins interact with a single DNA molecule, or an array of DNA molecules; and in the second article the diffusion characteristics are not analyzed. Our study addresses both issues. Next, we discuss the limitations of current instrumentation and the single-molecule point spread function (PSF) centroid tracking method for technically demanding studies of molecule sliding, which require millisecond temporal resolution and nanometer spatial resolution. We introduce our new single-molecule image deconvolution (SMID) method, which meets these technical demands and we present the results of our preliminary application of SMID to LacI sliding studies. Using the SMID method, sliding and sliding–hopping alternation characteristics can be extracted from single protein-diffusing-on-DNA images in millisecond and sub-millisecond exposure times; and thus the SMID method increases the temporal resolution of single-particle tracking by at least 100-fold from $\approx 300$ ms to milliseconds and sub-milliseconds (Sect. 2.4.1). Lastly, we outline additional studies necessary for single-molecule investigations of protein sliding on DNA using SMID.

### 2.2 Early LacI Diffusion Experiments in the Timescale of Seconds

#### 2.2.1 Facilitated Target Association Rate Calculation Assuming Brownian Sliding

In this section, we discuss our single-molecule imaging studies of one-dimensional diffusion of LacI repressor proteins along elongated DNA in the timescale of seconds. Our analysis of the LacI transcription factors’ diffusion yielded four main results: (1) LacI diffuses along nonspecific DNA in the form of 1D Brownian motion, (2) the observed 1D diffusion coefficients $D_1$ vary over an unexpectedly large range, from $2.3 \times 10^{-12}$ cm$^2$/s to $1.3 \times 10^{-9}$ cm$^2$/s, (3) the lengths of DNA covered by these 1D diffusions vary from 120 nm to 2920 nm, and (4) assuming that the LacI sliding is Brownian for all timescales, the mean values of $D_1$ and the diffusional lengths indeed predict a LacI target binding rate 90 times faster than the 3D diffusion limit.
The expected association rate $k_{a(3D)}$ by which DNA-binding proteins find their specific target sequences on double-stranded DNA in a random 3D search is $4\pi D_3 l_{seq}$ per unit protein concentration, where $l_{seq}$ is the effective DNA target length, and $D_3 = k_B T / 3\pi \eta a = 9 \times 10^{-7}$ cm$^2$/s is the 3D diffusion coefficient of the protein in solution [4, 15, 27], where $k_B$ is the Boltzmann constant, $T$ is the temperature, $\eta$ is the viscosity of the solvent, and $a \approx 5$ nm is the typical diameter of the protein. With $l_{seq} \approx 3$ bp (or 1 nm), the protein–DNA association rate $k_{a(3D)}$ should be $10^9$/M/s. The original in vitro study on LacI-lacO binding by Riggs et al. was with 45.5 kbp DNA of 15.5 μm in length, and the lacO association rate $k_{a(Exp)}$ was measured to be $10^{10}$/M/s, 100 times higher than the diffusion limit of $k_{a(3D)} \approx 10^9$/M/s [4] (the $10^{10}$/M/s binding rate was also reported in [13, 15, 28]).

It has been proposed that such high rates can be achieved if the protein undergoes a facilitated-diffusion process in which the protein performs a combination of 1D diffusion along the DNA and 3D diffusion in solution. In this model, the key to faster targeting lies with the nonspecific DNA sequences that flank the target site. By 3D diffusion, a protein most likely will run into a segment of nonspecific DNA first. After nonspecific binding, the protein will diffuse along the DNA for a certain time and eventually dissociate. By doing so, the effective concentration of protein near the DNA increases, and thus the targeting rate should change. This facilitated-diffusion modified protein-target association rate $k_a$ per protein concentration has been derived by Halford and Marko [15]:

$$k_a = \left(\frac{1}{D_3 l_d} + \frac{L c}{D_1}\right)^{-1} = k_{a,3D} \left(\frac{l_{seq}}{l_d} + \frac{D_3}{D_1} l_{seq} L d c\right)^{-1}, \quad (2.1)$$

where $D_1$ is the 1D diffusion coefficient of the nonspecifically bound protein along the DNA, $L$ is the total length of the DNA molecule, $l_d$ is the maximum DNA contour distance $x_{max} - x_{min}$ covered by the protein before dissociation, $c$ is the concentration of the target, and $\left(\frac{l_{seq}}{l_d} + \frac{D_3}{D_1} l_{seq} L d c\right)^{-1}$ is the acceleration factor to $k_{a,3D} = \frac{D_3 l_{seq}}{l_d}$. In order to evaluate the facilitated diffusion model directly, it is necessary to know $D_1$ and $l_d$, which can only be obtained by imaging protein-DNA binding dynamics using single-molecule measurements. In fact, if these values do not fall within a certain range, “facilitated” diffusion can actually slow the search times.

Note that in this facilitated-targeting rate calculation (and other numerous calculations), sliding was assumed to be the DNA sequence-insensitive Brownian diffusion for all timescales [15, 21, 29, 30]. In the Brownian sliding model, the 1D mean square displacement of proteins on DNA is $\langle \Delta x^2 \rangle = 2D_1 t$, where $D_1$ is the 1D sliding diffusion coefficient and $t$ is the observation time. Should protein translocation follow a different model from Brownian motion (i.e., if $D_1$ is time dependent as $D_1(t)$, or if the time dependence $\langle \Delta x^2 \rangle$ of differ from the power of one), the protein search speed through DNA sequences will change. As a consequence, the effect of facilitated diffusion on the protein’s target association rate will vary.
2.2.2 Experiments

We used a LacI fusion-protein consisting of a green fluorescent protein (GFP)–GFP13 (S65T):lacI-I12 fusion (GFP-LacI), and stained the DNA with the dimeric cyanine dye BOBO-3. DNA constructs of Lambda Zap vector with 256 tandem copies of lacO (lacO\textsubscript{256}) were used. LacO\textsubscript{256}-DNA was 42.06 kbp long with a contour length of 14.3 µm, and the 9.22 kbp lacO\textsubscript{256} insertion started at 24.02 kbp. The nonspecific sequences of the DNA construct are identical to that of λ DNA. The synthesis methods for the fusion protein and the lacO\textsubscript{256}-DNA, and the sample preparation method are described in [31]. There were lacO\textsubscript{256}-DNA dimers as well as monomers in the solution; the dimers were formed by the sticky-end-hybridization of two lacO\textsubscript{256}-DNA monomers (Fig. 2.2a). After the LacI-DNA and BOBO-3 incubation, the GFP-LacI concentration was 50 nM and the lacO\textsubscript{256}-DNA concentration was 11 pM (0.3 µg/ml). The DNA intercalating cyanine dyes are known to stretch DNA by 30% in length at 1 dye/5 bp [32], so at our concentration of 1 dye/10 bp, the DNA molecules were stretched by 15% to 16 µm. Since BOBO-3 produced no obvious effect on the DNA-configuration-dependent LacI-DNA

![Fig. 2.2](image-url) (a) Schematics of a GFP-LacI (green) bound lacO\textsubscript{256}-DNA monomer and dimer (red). (b) Elongation of the DNA. (c) Frame-averaged superposed image of GFP-LacI bound to an elongated lacO\textsubscript{256}-DNA dimer. The scale bar is 1 µm. (d) A GFP-LacI monomer of frequent blinking and unitary bleaching. (e) A GFP-LacI monomer that blinked, recovered the first bleaching in 3 s, and finally irreversibly bleached. (f) The GFP-LacI dots for the first 12 frames of (e), showing blinking at frames 2 and 7, and bleaching at frame 10. (g) A GFP-LacI dimer with two bleaching events.
specific binding [31], we expect that its effect on the LacI DNA-specific binding (which is less DNA-configuration dependent) will be negligible [33]. A catalytic oxygen scavenging solution was used to maximize dye lifetimes [31]. 1 μl of the DNA+LacI solution and 4 μl of the oxygen scavenging solution were deposited onto a fused-silica chip.

A glass cover slip was used to flatten the solvent, and the edges of the cover slip were then sealed with nail polish. As the cover slip flattened the droplet, hydrodynamic flow elongated the DNA dimers, and the two LacI-lacO256 sites stuck to the surface, creating an anchored elongated DNA molecule (Figs. 2.2b, c, and 2.3c) stretched up to 90% of its native contour length. The tension on DNA was a few pico-Newton [34]. DNA was not observed to stick to fused-silica surfaces at our pH of 8.0 and BSA concentrations, and the elongated DNA molecules were effectively suspended from the surface, as evidenced by the DNA’s transverse motion of ± 50 nm (data not shown). Thus, unbound GFP-LacI molecules interacted only with free unattached and nonspecific DNA. Note that the sticking of GFP-LacI to fused-silica surfaces occurred only at the deposition step as the air–water interface moved over the chip surface. After the cover slip was sealed, the free GFP-LacI molecules (≈2 nM) diffused in the solution freely and did not stick to the surface, as evidenced by observation of the freely diffusing GFP-LacI near the surface (data not shown).

The single-molecule experiments were performed using a prism-type Total Internal Reflection Fluorescence Microscopy (TIRFM) method (Fig. 2.3a and b). The laser excitation was synchronized to the 3.4 Hz data acquisition rate of the I-CCD camera. The emitted photons from BOBO-3 and GFP were collected using a 100X TIRF oil-immersion objective (N.A. = 1.45), went through a custom-designed dichroic mirror and emission filter set (Chroma Technology

![Fig. 2.3](image-url)

**Fig. 2.3** Schematics of our experimental setup. (a) Light goes from air into a prism. At the prism–water interface, TIRF occurs at an inclination above the critical angle. (b) The decay of the evanescent light intensity \( I(z) \) in \( z \) direction. The penetration depth \( d \) is \( \approx 250 \) nm. (c) Schematic of our protein–DNA interaction system. A single DNA molecule (orange) is elongated along the fused-silica surface (grey) for TIRF imaging (blue light), and the 3D diffusing LacI molecules (green) run into the DNA molecule and bind to it. (d) Schematics of our imaging system. DNA and proteins are excited by yellow and blue laser light, respectively, and the emission photons go through an 100X objective and are collected by a single-photon sensitive CCD camera.
Single-Molecule Imaging of LacI Diffusing Along Nonspecific DNA

Corp, Rockingham, VT), and were recorded by an I-CCD camera (I-PentaMAX:HQ Gen III, Princeton Instruments, Trenton, NJ; Fig. 2.3d). The PSF width of the diffraction-limited optical system was measured to be 280 nm, and the imaging pixel size was 117 nm. The pixel count of the camera was converted to a photon count using known conversion factors [31]. The mean 488 nm illumination intensity over the illumination areas of 30 \( \mu \text{m} \times 50 \mu \text{m} \) was 1000 W/cm\(^2\). The centroid location of a GFP-LacI dot was determined by fitting its 1D fluorescence intensity profile to a Gaussian. The number of detected photons per PSF per frame (between 50 and 300 photons) limited the position measurement accuracy to be between 10 and 50 nm [35].

Knowledge of the fluorescence characteristics of single free GFP-LacI monomers and dimers attached to fused-silica surfaces is essential in determining the single-molecule nature of a bound protein. GFP-LacI monomers blink frequently (short fluorescence dips to near instrumental noise level), and bleach with no recovery (Fig. 2.2d). At our excitation intensity of 1000 W/cm\(^2\), mean exposure time of 10 ms, and synchronized imaging frequency of 3.4 Hz, the mean net observation time of each GFP-LacI molecule was 5 s before it bleached (giving a total laser exposure time of 0.15 s). The mean number of photons emitted by the bound GFP-LacI molecules before bleaching was \( \approx 4 \times 10^4 \) photons. This 5 s observation time gave the instrumental limit to the maximum mean distance we observed GFP-LacI motion on DNA in this experiment.

2.2.3 Results and Analysis

An image sequence of a single GFP-LacI molecule diffusing along DNA is shown in Fig. 2.4b. This is 1 out of 70 walks that were observed and chosen for its large net displacement. Figure 2.4a shows the frame-averaged superposed image of the anchored DNA and the diffusing GFP-LacI on DNA. Time-lapse images of the diffusing protein show clear relative displacements (Fig. 2.4b), with one immobile anchoring site used as a reference point. We know that we were observing a single GFP-LacI dimer from the fluorescence time trace in Fig. 2.4d, which clearly shows two bleaching steps. Both GFP-LacI monomers (80%) and dimers (20%) have been observed to diffuse on DNA. As is evident in Fig. 2.4d, fluorescence time traces of bound GFP-LacI molecules were identical to that of single immobile GFP-LacI (Fig. 2.2d-g), with the same blinking rate and characteristic bleaching time of \( \approx 0.15 \text{ s} \) (5 s net observation time). The DNA locations of the diffusing protein at different frames are correlated and localized, thus at \( D_3 = 10^8 \text{ nm}^2/\text{s} \) and our protein concentration of a few proteins/\( \mu \text{m}^3 \), the chance for two different proteins landing consecutively on the same location of DNA is 1 in 1000. Figure 2.4d plots the distribution of all relative displacements \( \langle D_x \rangle \) of the walk. This is a Gaussian of SD = 130 nm centered near zero, which is typical for Brownian motion with limited data points.

Now, we discuss our analysis showing that individual protein diffusion trajectories, which consist of multiple measurements \( x_i \) until the protein disassociates, are
Brownian in nature, and we obtain 1D diffusion constant $D_1$ for these trajectories. Qian et al. have derived an expression to tell (1) whether a single diffusion trajectory is Brownian and if so (2) obtain the diffusion constant of the trajectory [36]. This method calculates the mean square displacement MSD($n, N$) for all available time intervals of a single diffusion trajectory

$$MSD(n, N) = \frac{1}{N-n} \sum_{i=1}^{N-n} (x_{i+n} - x_i)^2 = 2D_t n \Delta t + 2\sigma_s^2,$$  \hspace{1cm} (2.2)

where, $N$ is the total number of positions measured, $n$ is the measurements index going from 1 to $N$, $\Delta t$ is the time interval between two consecutive position measurements, and $\sigma_s$ is the measurement accuracy associated with each $x_i$. We can obtain $D_1$ of a single diffusion trajectory from its MSD($n, N$) to high precision by weighted linear-fitting MSD($n, N$) to $n$, taking MSD($n, N$)'s variances at different $n$ into consideration. As $n$ increases, the number of available measurement points for MSD($n, N$) averaging decreases, and the variance in MSD($n, N$) increases as

$$\sigma_{n,N}^2 = (2D_t n \Delta t)^2 (2n^2 + 1) / [3n(N-n+1)].$$  \hspace{1cm} (2.3)

If a single trajectory is Brownian, then its MSD($n, N$) at $n$ below a cutoff $n_c$ will be a linear function of $n$, with $n_c$ determined by a set fractional MSD($n, N$) uncertainty in Eq. 2.3. We chose $n_c$ to be where $\sigma_{n,N}^2(2D_t n \Delta t)$ is 50%. We plot MSD($n, N$) versus $n$
for trajectories with N > 10, where there are at least three MSD\(_{(n, N)}\) values whose fractional variances are <50% Eq. 2.3. We also used only trajectories with less than five contiguous GFP blinks. Since \(\sigma_s < 50\) nm, MSD\(_{(n, N)}\), which is the square of the difference of two position measurements, has an offset of 2500 nm\(^2\) < 2\(\sigma_s^2\) < 5000 nm\(^2\). These photon noise offsets were subtracted in the MSD\(_{(n, N)}\) versus \(n\) curves.

Figure 2.5a plots displacement \(x\) versus time for 70 trajectories. The 15 trajectories in color are the walks for which we have obtained \(\text{D}_1\), and the center black line is a stationary GFP-LacI stuck to the fused-silica surface (not DNA). Figure 2.5c plots MSD\(_{(n, N)}\) versus \(n\) for these 15 trajectories in linear scale and Fig. 2.5d in log–log scale at low \(n\) values, respectively. The log–log plots are all straight lines with the slope of 1 at low \(n\), clearly indicating that the 1D trajectories are Brownian motions. The dashed line in Fig. 2.5d is a fit of 2 with weighted error 3 to all \(n\) points below \(n_c\) of the topmost trajectory. The intercepts at \(n = 1\) are 2\(\text{D}_1\)\(\Delta t\) for each particular walk, as can be seen by inspection of 2 (Fig. 2.5d). Thus, while all the walks are Brownian in nature, the different intercepts at \(n = 1\) indicate that there is a large distribution in diffusion coefficients and there is not a unique, single value for \(\text{D}_1\). We also plotted the distributions of nondegenerate relative displacements \(x_i - x_{i-n}\) for the first 15 positions of all 70 trajectories for \(n = 1, 2,\) and 3 in Fig. 2.5b; the displacements are all Gaussians centered at zero with SD increasing with \(n\). This result further demonstrates that LacI’s diffusion trajectories are truly Brownian in nature, regardless of the variations in individual diffusion coefficients. Two other papers on single-molecule imaging studies of

---

Fig. 2.5 (a) \(x\) versus \(t\) for 70 trajectories. (b) Nondegenerate \(x_i - x_{i-n}\) distributions. (c) MSD\(_{(n, N)}\) versus \(n\) for the 15 colored walks in linear scale and (d) in log–log scale. The arrows in (a), (c), and (d) denote the walk in Fig. 4. The dashed lines in (c) and (d) are the fit of 2 to the top trajectory.
protein diffusion on DNA were published in the same year (2006) [37, 38], with
[38] reporting a similar large distribution in 1D diffusion coefficients for Rad51
on aligned DNA molecules, so our result here may be of some generality.

Figure 2.6a shows the distribution of the 15 $D_1$ values (corrected to DNA contour
length), which span a large range from $2.3 \times 10^2$ nm$^2$/s to $1.3 \times 10^5$ nm$^2$/s. Figure 2.6b
shows that the different $D_1$ values are distributed randomly along the DNA, showing
a lack of correlation between the $D_1$ and the regions on the lambda DNA on which the
protein has diffused. Figure 2.6c shows the distribution of the $l_d$ in DNA contour
length. Because lambda DNA has large sequence variance in the nonspecific region with
$\pm 30\%$ difference in local AT and CG concentrations, it is possible that the diffusion
constants are a function of local sequence. It is also possible that the large distribu-
tion in $D_1$ is caused by conformational distributions in the protein [39]. Further
experiments are needed to answer these questions.

Finally, we use our data to examine the question of the extent to which facilitated
diffusion can enhance the LacI target binding rate. Just as there is a distribution in
the 1D diffusion coefficient $D_1$, there is also a distribution in the diffusion lengths $l_d$, 
whose value is further compromised by the mean observation time to bleaching of
the GFP of 5 s. Since the final target binding is the result of many diffusion events

![Figure 2.6](image-url)

**Fig. 2.6** (a) $D_1$ distribution of the 15 trajectories. (b) $D_1$ versus fractional bound location on the
nonspecific segment of the LacO$_{256}$-DNA dimer. The error bars were obtained from the fit of
MSD$_{n,N}$ to $n$ with weighted errors at each $n$ given by 3. (c) Histogram of $x_{\text{max}} - x_{\text{min}} = l_d$ for
the 70 trajectories. The solid line is a Gaussian fit with a mean of $500 \pm 220$ nm (mean $\pm$ SD).
Values in (a) and (c) have been adjusted to DNA contour length.
on nonspecific DNA, we use the mean $< l_d >$ (probably a lower bound due to bleaching) of 500 nm, and the mean diffusion coefficient of $< D_1 > = 2.1 \times 10^{-10}$ cm$^2$/s in 1. Also using Riggs’ concentration of 1 lacO/1670 μm$^3$, $D_s = 4 \times 10^{-7}$ cm$^2$/s for LacI tetramers (a ≈ 10 nm), and $L = 15.5$ μm, the accelerating factor in 1 is 93 ± 20, which thus resolves the 100-fold discrepancy between the theory and the experimental data. We conclude from these measurements that facilitated diffusion increases the LacI-lacO binding rate well over the apparent diffusion limit. This result demonstrates that facilitated diffusion in the form of 1D Brownian motion is the mechanism responsible for the faster-than-diffusion binding of LacI to lacO, and quite possibly, the reason also for the observed faster-than-diffusion binding in other protein–DNA interactions.

### 2.2.4 Concern for the Interpretation of the “Sliding Length” $< l_d >$ and $D_1$

There are two concerns for the above interpretation of experimental results: (1) The observed mean “sliding length” $< l_d >$, which is the maximum DNA contour distance $x_{max} - x_{min}$ covered by the protein before dissociation, quite likely is not the pure sliding length defined in Eq. 2.1; rather it is the combined distance of many sliding and hopping cycles before the permanent dissociation of proteins from DNA. This statement stems from the estimation that LacI’s nonspecific DNA dissociation rate constant is on the order of milliseconds (≈ 0.6 ms to 5 ms, or milliseconds to seconds) [3, 19, 21, 30]. (2) For the same reason, $D_1$ is the “effective” 1D diffusion coefficient of the combined sliding and hopping trajectory in the timescale of seconds, rather than the pure sliding diffusion coefficient.

With the three questionable parameters in Eq. 2.1 – (1) Brownian sliding (Sect. 2.2.1), (2) the sliding length (Sect. 2.2.3), and (3) the 1D diffusion coefficient $D_1$ (Sect. 2.2.3) – the LacI target association rate calculation due to facilitated diffusion should be reevaluated. In order to correctly calculate facilitated target association rate using Eq. 2.1, the true sliding displacement versus time relation, sliding length, and sliding dissociation rate constant should be obtained.

### 2.3 New Challenge: Millisecond Timescale Single-Molecular Tracking

To reiterate key limitations of the above LacI diffusion experiments in the timescale of seconds for sliding mechanism studies of LacI on DNA: (1) We know that the LacI proteins stay around DNA for the observation time of seconds; what we do not know is whether the proteins slide for the whole time or many hopping cycles are convolved. (2) If the protein diffusion pathway on DNA is a combination of sliding and hopping, the sliding–hopping-alternation kinetics is not known
(i.e., the mean sliding time \( t_d \) and hopping time \( t_{hop} \) are not known). (3) In the sliding motion, how the protein displacement changes with time is not known. These three factors must be addressed for correct calculation of the target association rate of proteins.

Faster single-molecule tracking than the current centroid tracking method in seconds timescales appears to be the answer to the above three questions. Then, how much faster is fast enough? For (1) and (2), millisecond tracking resolution may be sufficient according to Sect. 2.2.4. For (3), the sliding motion characteristics studies, is milliseconds tracking also sufficient? Below, we list predicted displacement versus time characteristics of different sliding models in \( 10^{-7} \) s to 1 s timescales.

Other than the DNA-sequence-independent Brownian sliding model, the alternative sliding models are the DNA-sequence-dependent sliding models. There are four simple DNA-sequence-dependent models describing four different protein translocation energy landscapes along a stretch of DNA sequences (Fig. 2.7a I–IV): (1) The energy at each binding site \( n \) is independent of others. The translocation energy barrier from site \( n \) to site \( n' = n \pm 1 \) is the difference between the protein binding energies of the two sites, if positive and zero, if negative [22, 40]. (2) In order to move from site \( n \) to site \( n' \), the protein needs to completely dissociate from the DNA first over a threshold level \( EM = \text{Max}[E(n)] \). The translocation energy barrier is \( EM - E(n) \) [22]. (3) The threshold energy \( Et \) is lower than \( EM \), and the translocation energy barrier is the maximum of the energy differences and zero – \( \text{Max}[Et - E[n], E[n'] - E[n], 0] \) [22, 41, 42]. (4) Two-state model in which \( Et \) separates the reading regime (\( E[n] < Et \)), where the translocation is the same as in model III from a sliding regime (\( E[n] > Et \)), where the sliding motion is on a flat energy landscape [22, 43, 44].

In contrast to the DNA-sequence-independent sliding model, where the translocation mechanism yields Brownian diffusion for all timescales as \( \langle n^2 \rangle = 2D_t t \) the four DNA-sequence-dependent protein translocation mechanisms predict distinct diffusion patterns at different timescales. Figure 2.7b shows the mean square displacement versus time (\( \langle n^2 \rangle \) versus \( t \)) log–log plots of the four sequence-dependent diffusion models from \( t = 5 \times 10^{-7} \) s to 1 s. All diffusions are Brownian with the \( \langle n^2 \rangle \) versus \( t \) slope of one in the log–log plot at \( t > 30 \) ms (left vertical dashed line). At \( t < 30 \) ms, all models exhibit subdiffusion behavior, in which \( \langle n^2 \rangle \) is proportional to \( A(t) t^b \), with \( b < 1 \) (\( b(t) \) being the slope of the lines at time \( t \) and \( \log(A(t)) \) being the vertical offset of the line, according to the expression \( \log \langle n^2 \rangle = \log A(t) + b(t) \log t \)).

It is clear that diffusion studies of sliding below 30 ms are necessary to differentiate Brownian sliding from subdiffusive sliding, provided that sliding lasts longer than 30 ms. This 30-ms tracking resolution cannot be afforded by the current centroid method, which is limited to temporal resolution of 300 ms (see Sect. 2.4.1), let alone 1 ms sliding time. While the diffusion of proteins on DNA in the timescales of seconds has been reported in many recent single-molecule protein–DNA interaction studies [20, 37, 38, 45–48, 48–56], no millisecond timescales studies have been reported. Alternative higher temporal resolution single-molecule