Exploring Microscopic Mechanisms of Chemical and Biological Processes using Discrete-State Stochastic Models

विद्या वाचस्पति की

उपाधि की अपेक्षाओं की आंशिक पूर्ति में प्रस्तुत शोध प्रबंध

A Thesis Submitted in Partial Fulfillment

of the Requirements of the Degree of

Doctor of Philosophy

द्वारा / *by*

भावाक्षी पुनिया / Bhawakshi Punia

पंजीकरण सं. / Registration No.: 20193674

शोध प्रबंध पर्यवेक्षक / Thesis Supervisor: Dr. Srabanti Chaudhury



भारतीय विज्ञान शिक्षा एवं अनुसंधान संस्थान पुणे

INDIAN INSTITUTE OF SCIENCE EDUCATION AND

RESEARCH PUNE

2024

I dedicate this thesis to
my grandparents.

CERTIFICATE

Certified that the work incorporated in the thesis entitled **"Exploring Microscopic Mechanisms of Chemical and Biological Processes using Discrete-State Stochastic Models"** submitted by **Bhawakshi Punia** (**Roll No: 20193674**) was carried out by the candidate, under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or institution.

S. Chaudhury

Dr. Srabanti Chaudhury

Date: 16.07.24

DECLARATION

Name of the student: Bhawakshi Punia

Registration Number: 20193674

Thesis supervisor: Dr. Srabanti Chaudhury

Department: Chemistry

Date of joining program: August 01, 2019

Date of Pre-Synopsis Seminar: April 25, 2024

Title of Thesis: Exploring Microscopic Mechanisms of Chemical and Biological Processes using Discrete-State Stochastic Models

I declare that this written submission represents my idea in my own words and where others' ideas have been included; I have adequately cited and referenced the original sources. I declare that I have acknowledged collaborative work and discussions wherever such work has been included. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

The work reported in this thesis is the original work done by me under the guidance of Dr./Prof. Srabanti Chaudhury

Date: 16.07.2024

Signature of the student Bhawakshi

ACKNOWLEDGEMENTS

As I reach the end of this journey, I am reminded of my uncle's words: "It is a unique experience. You'll remember it forever." The uniqueness of this experience is truly shaped by the wonderful people I met and the memories I made, for which I am deeply grateful.

To begin, I want to express my sincere thanks to my advisor, Dr. Srabanti Chaudhury, whose guidance, support, mentorship, and patience have been instrumental in navigating this journey. I still remember your words on our orientation day: "PhD is a lot more about hard work and perseverance." Thank you for providing me with life-changing opportunities and allowing me to grow through them all.

I am thankful to my Research Advisory Committee (RAC) members, Prof. Sarika M. Bhattacharyya (CSIR-NCL Pune) and Dr. Bijay Agarwalla (IISER Pune), for their time, comments, feedback, and continuous evaluation. Additionally, I extend my gratitude towards Prof. Anatoly B. Kolomeisky (Rice University, U.S.A.) for a wonderful collaboration. Your expertise has enhanced my conceptual understanding and pushed me to think more deeply and critically.

To my beloved parents, Neeraj and Manoj Punia, my constant support system throughout life, whose love and belief provided me with confidence. My dearest brother Yash, who has been my anchor through the highs and lows of this journey. Thank you for being available for all kinds of discussions, at any place or time. Special thanks to my grandmother Premlata Punia and my uncles Dr. Sukhwinder Dabar and Dr. Mukesh Punia, who nudged me to take up this valuable and unique journey.

Having a cheerful and supportive lab environment is a blessing. I was welcomed by my seniors, Dr. Bappa Ghosh and Dr. Divya Singh, along with Rugwed, our shared lab space member. I am thankful to Bappa for sharing his wisdom whenever I was in need. Extending my heartfelt gratitude to Divya for her help while I was learning the tools and techniques of the lab, for the dedicated jokes, and tea-time discussions. You were the warmest seniors I could ever imagine! I thank Kinjal for clearing technical doubts in my first project. The present members of the lab − Pankaj and Akshay, for a fun environment. I will cherish our laugh sessions, watching cringe content together, and sharing weird and funny stories. I wish you all the success in life! The current and former members of the neighboring lab − Ardhra, Dr. Prabhat, and Dr. Uttama, for scientific, non-scientific and silly discussions that offered a breather.

Special thanks to my colleagues and friends at IISER Pune − Deepak, Rashi, Ruksana, Sarika, Umashis, Nahid, Alpana, Rishu, Cavya, and Sandeep. The camaraderie, occasional dinner parties, and shared experiences have not only contributed to my academic journey but also made it an enjoyable and memorable one. I am especially grateful to Deepak for his unfailing companionship. Be it the movies, temple visits, stand-up comedy shows, watching web-series and sports, or lending an ear during both happy and difficult times, your presence was constant throughout my PhD journey. Also, I thank Rashi for being the fun, charming, ever-smiling hostel room neighbor and my confidante. Our late-night chit-chats and warm hugs will never be forgotten. I thank Sayali and Sujit for their occasional visits. This list is incomplete without my friends from back home. Jatin, Simran, and Arushi, I am genuinely blessed to have you guys for keeping me *insanely* sane and checking up on me over time. To Sandeep and Aayush, I acknowledge our decade-long friendship, support, laughter, and shared memories. Arjun, thank you for the never-ending conversations from zero to infinity (quite literally and poetically). I wish my friends good health and prosperity in life.

I extend my heartfelt appreciation to all members of IT, Accounts, and the Academic office at IISER Pune for their assistance in ensuring that the logistics of my research and national/international conference/workshop visits proceeded smoothly. I genuinely thank Sayalee Ma'am, Mayuresh Sir, Swapnil Sir, Tushar Sir, and Sanjay Sir for their availability and efficiency. A special mention to the housekeeping staff at IISER Pune for keeping our campus, especially the reasearch labs well as our hostels, clean and tidy.

Additionally, I would like to extend my heartfelt gratitude to the esteemed Prime Minister Research Fellowship (PMRF) for providing financial support for my research during my doctoral studies. This distinguished fellowship also enabled me to participate in a diverse array of national and international conferences, which expanded my academic horizons and provided me networking opportunities. The experiences offered by PMRF have significantly contributed to the progress of my academic studies as well as a successful completion of my thesis.

With the submission of this thesis, I would like to express my sincere gratitude to all the people who have been my comrades, companions, colleagues, acquaintances, and well-wishers throughout this journey. Thank you for your presence.

SYNOPSIS OF THE THESIS

Chemical and biological processes are inherently stochastic and heterogeneous at microscopic and mesoscopic levels. This thesis theoretically investigates the microscopic dynamics of selected processes through discrete-state stochastic modelling frameworks.

Biological cells are extremely complex where multiple stochastic processes occur simultaneously. Given that biomolecules occupy 40% of intra-cellular volume, they may act as crowding agents and influence biological events. The search by DNA-binding proteins (DBPs) for specific target DNA sequences initiates key biological processes. While extensively studied in vitro, bulk macromolecules tend influence in vivo search by altering DBPs' diffusion. This thesis examines these effects using a discrete-state stochastic framework considering factors such as the spatial positioning of crowders relative to the target site and their mobility. Strong non-specific crwoder-protein interactions can also accelerate target search. Theoretical predictions align with existing studies and are supported by Monte Carlo simulations, providing a deeper understanding of the process. In addition, proximity effects between a sub-diffusing protein and its target are also investigated through a continuum approach, yielding distance-dependent first passage time distributions. Another process probed in this thesis is the transport of ssDNA through biological nanopore sensors which are often used for genetic sequencing. Studies have shown that poly(ethylene glycol) (PEG) crowders increase the capture rates and translocation times of ssDNA through α-hemolysin nanopores, enhancing throughput and sensing accuracy. A discrete-state stochastic model based on cooperative partitioning of PEGs within the nanopore cavity is developed. Apparent electrostatic interactions between DNA and PEGs are hypothesized to control the dynamics of this processes.

The thesis also delves into stochastic chemical processes, particularly focusing on single-molecule (SM) experiments that have revealed detailed catalytic properties of individual nanoparticles (NPs). These experiments provide insights that are obscured in ensemble-averaged bulk measurements. Extending a minimal theoretical framework, a comprehensive description of reaction dynamics on catalysts with multiple active sites is discussed using a discrete-state stochastic approach. It examines how the influence of active sites, intermediate species, and reaction topologies influence catalytic dynamics, providing quantitative bounds for realistic catalytic processes that can be applied to experimental data. In a notable SM experiment, redox reactions catalyzed by individual metal NPs exhibit cooperative effects with finite lifetimes and communication distances. This

thesis proposes that surface diffusion of positively charged holes formed during redox reactions stimulates subsequent reactions. An exactly solvable discrete-state model is used to capture their dynamics, with analytical results matching experimental observations, supporting the theoretical approach. This work is extended to account for spatial complexity of NPs, predicting an enhanced cooperative communication which is restricted to smaller regions. Monte Carlo simulations corroborate these findings, clarifying molecular mechanisms of catalytic processes.

The interdisciplinary investigation presented in this thesis aims to highlight the strength of discrete-state stochastic modelling techniques in understanding complex chemical and biological systems, offering crucial insights for advancing both theoretical and experimental efforts.

Contents

Certificate i

Declaration ii

Acknowledgements iii

Synopsis of the Thesis v

1 Introduction 1

1.1 Stochastic Processes in Chemistry and Biology 1

1.2 Single Molecule (SM) Measurements 4

1.3 Stochastic Simulation Algorithm 7

1.4 Chemical Master Equations and First Passage Time Distributions 9

1.5 Quantifying Stochastic Fluctuations from Probability Distributions 13

1.6 Proteins' Search for Target DNA Sequences 16

1.7 Transport of ssDNA through Biological Nanopores 18

1.8 SM Nanoparticle Catalysis 21

1.9 Thesis Chapters in Brief 22

Bibliography 24

2 Theoretical Study of the Role of Dynamic Bulk Crowders in the Protein Search for Targets on DNA 41

2.1 Introduction . . . 42

2.2 Theoretical Model . . . 44

2.3 Results and Discussion . . . 49

2.4 Conclusion . . . 54

Bibliography . . . 56

3 Influence of Nonspecific Interactions between Proteins and In Vivo Cytoplasmic Crowders in Facilitated Diffusion of Proteins: Theoretical Insights 60

3.1 Introduction . . . 61

3.2 Theoretical Model . . . 64

3.3 Results and Discussion . . . 70

3.3.1 Effect of the Strength of the Protein-Crowder Associations . . . 71

3.3.2 Effect of the Position of the Crowded Region around the Target . . . 74

3.3.3 Effect of the Size of the Crowded Region . . . 75

3.4 Comparisons with Experiments and Computational Studies . . . 76

3.5 Conclusions . . . 77

Bibliography . . . 78

4 Theoretical Insights into the Full Description of DNA Target Search by Subdiffusing Proteins 85

4.1 Introduction . . . 87

4.2 Theoretical Model . . . 89

4.3 Results and Discussion . . . 93

4.4 Application to Real Systems 98

4.5 Conclusions 100

Bibliography 101

5 Microscopic Mechanism of Macromolecular Crowder-Assisted DNA Capture and Translocation through Biological Nanopores 107

5.1 Introduction 108

5.2 Theoretical Model 110

5.3 Results and Discussion 117

5.4 Conclusions 120

Bibliography 122

6 Macromolecular Crowding Facilitates ssDNA Capture within Biological Nanopores: Role of Size Variation and Solution Heterogeneity 126

6.1 Introduction 127

6.2 Theoretical Model 129

6.2.1 Effect of Macromolecular Crowder Size on ssDNA Capture 129

6.2.2 Effect of Heterogeneity in Crowded Solutions on ssDNA Capture . . 131

6.3 Results and Discussion 134

6.4 Conclusions 138

Bibliography 139

7 Understanding the Reaction Dynamics on Heterogeneous Catalysts Using a Simple Stochastic Approach 143

7.1 Introduction 144

7.2 Theoretical Model 146

7.3 Results and Discussion . . . 149

7.4 Conclusions . . . 160

Bibliography . . . 161

8 Microscopic Mechanisms of Cooperative Communications within Single Nanocatalysts 164

8.1 Introduction . . . 165

8.2 Theoretical Model . . . 167

8.3 Results and Discussion . . . 170

8.3.1 Catalytic Communications in Metallic Nanorods . . . 170

8.3.2 The Effect of External Electric Field . . . 173

8.4 Conclusions . . . 176

Bibliography . . . 179

9 How Heterogeneity Affects Cooperative Communications within Single Nanocatalysts 182

9.1 Introduction . . . 183

9.2 Theoretical Model . . . 184

9.3 Results and Discussion . . . 188

9.4 Conclusions . . . 196

Bibliography . . . 197

10 Conclusions and Future Directions 201

Appendix A 209

Appendix B 214

Appendix C **224**

Appendix D **227**

Appendix E **231**

Appendix F **235**

Appendix G **238**

List of Publications **245**

Rights and Permissions **246**

Chapter 1

Introduction

"Almost all life depends on probabilities."

Voltaire

1.1 Stochastic Processes in Chemistry and Biology

Whether it's choosing what to eat, waiting for a bus, checking stock market prices, or a reader's decision to browse through this thesis, the outcomes of these events are influenced by probability, which guides us through this world of uncertainty and chance. This is the essence of a stochastic process. The term 'stochastic' is a jargon for 'random' and thus, a stochastic process describes any event that evolves randomly over time.[1] Based on how a system under study is being monitored, these events are classified into two main categories: discrete state and continuous state stochastic processes.[2] In discrete-state stochastic events, the possible states of a system are distinct, finite and countable, whereas the system may observe infinitely uncountable states as it evolves temporally during continuous-state stochastic processes.[1, 2] Interestingly, these events may be monitored in both ways. For instance, let us consider a simple example of a bus ride. The bus's journey may be represented by its transitions between a finite set of stops, which are probabilistic and can be influenced by factors such as schedule adherence, traffic conditions, etc. This monitoring approach captures the bus's movement in discrete states, i.e., bus stops. On the other hand, the exact position of the bus may be represented as a continuous function of time, capturing its ongoing movement along the route. Changes in its

speed and traffic conditions introduce stochasticity, which can be modeled as a continuous state stochastic process. Irrespective of the adopted approach, the system's variability results in a collection of random variables, X, that describe the fate or state of a system as it evolves in time t and thus stochastic processes are expressed as $X(t)$.[1, 3] Since each random variable is associated with a probability, one can probe the respective probability distributions to quantify the likelihood of stochastic outcomes using several experimental, simulation and mathematical frameworks.

Our comprehension of natural events in our surroundings generally begins with macroscopic observations, where systems appear to operate deterministically within the constraints of classical physics. However, a closer examination at the microscopic level reveals that these seemingly deterministic processes are inherently stochastic due to the random interactions of individual particles. Chemical reactions and biological processes are guided by the principles of statistical mechanics at the microscopic resolution, where the random collisions and interactions among molecular constituents drive variable outcomes.[4]

Since we encounter chemical reactions macroscopically, which involves a large collection of molecules, the time evolution of each chemical species is generally expressed with respect to their concentrations, based on rate laws. Often certain approximations are introduced to simplify the set of differential equations, for example the steady state approximation. This study is commonly referred to macroscopic chemical kinetics.[5, 6] However, in all chemical reactions, individual reactant species collide and interact with each other to produce desired products. Such random collisions lead to a variability in the timescales of the microscopic reaction events contingent upon a sequence of successive probabilistic occurrences with multiple possible time evolutions for an initial number of molecules.[4, 7–9] The respective probabilities can be determined by observing a significant number of events, resulting in a probabilistic rate law. Thus, deterministic and stochastic modeling are closely related, with deterministic modeling's reaction rate constants being interpreted in terms of probabilities in the stochastic approach.[8, 9] It is important to note that the stochastic approach is crucial for describing reaction kinetics of systems with small concentrations because traditional deterministic models usually fail to capture the random fluctuations that can significantly impact reaction outcomes in such systems.[7] Delbrück initially modelled an autocatalytic reaction by developing a corresponding stochastic model, paving the way for a stochastic treatment of several different types of chemical reactions.[6, 10–12] Probing the classic Michaelis-Menten (MM) reaction mechanism at the molecular level has revolutionized the field of single molecule

enzymology.[13–17] Theoretical studies show that while individual enzyme molecules obey classic MM kinetics, it has a completely different interpretation at the microscopic resolution.[18, 19] The reaction rates of stochastic events can also be interpreted as a barrier crossing problem of Brownian particles, popularly known as the Kramers model.[20–22] Electrochemical processes like electron transfer, ion transport and molecular interactions at electrode surfaces naturally possess random fluctuations. Their study provides significant insights to electrodeposit formation and interface dynamics, ultimately improving the design and optimization of electrochemical systems.[23–27] Another example is fluorescence intermittency, also known as "blinking", observed in individual fluorescent molecules or quantum dots where the signal randomly switches between "on" and "off" states. Understanding this phenomenon is crucial for applications in single-molecule spectroscopy and the development of advanced fluorescent probes.[28–30]

Stochastic processes are particularly important in small-volume systems like biological cells which are inherently noisy and thus, are necessary for understanding how organisms function at the molecular and cellular levels.[4] Multiple biological events simultaneously occur in each cell.[31, 32] For instance, gene expression is stochastic due to protein-coding genes being present in few copies, transcription initiation being a multi-step process, and the possible arrival of multiple transcription factors.[33] This results in fluctuations in mRNA and protein abundances, leading to cell-to-cell variability in expression levels.[33, 34] The random diffusion of bio-macromolecules within a cell influences reaction rates and efficiency of cellular processes like signal transduction and metabolic pathways. These are complex systems of biochemical reactions controlled by interacting species and they regulate various cellular functions in response to stimuli.[35] Examples include protein-DNA interactions,[36, 37] protein-protein interactions,[38, 39] G-protein coupled receptor (GPCR) signalling,[40] etc. These stochastic events serve as the cornerstone of molecular biology, shaping the dynamic landscape of molecular transformations and cellular behaviours.

To probe these chemical and biological stochastic processes effectively, researchers seek to obtain probability time distributions that characterize the likelihood of different outcomes occurring over time.[17] These distributions can be derived both experimentally and theoretically. In the following sections, these methods are discussed in detail, followed by an examination of how to analyze the resulting probability distributions to gain deeper insights into the underlying stochastic processes.

1.2 Single Molecule (SM) Measurements

Even though large-scale experiments typically involve observing ensembles of molecules subjected to identical physiochemical constraints over extended periods, their interpretation involves discussions of molecular events where single molecules often take center stage. Bulk measurement techniques average the properties of numerous molecules under similar conditions, obscuring individual molecule behaviors which may not always act the same way.[17, 41] Averages, therefore, do not accurately represent the contribution of each molecule. The emergence of SM studies addresses this limitation by eliminating ensemble averaging, providing direct access to statistical distributions resulting from individual molecules. In the early 1980s, the scanning tunnelling microscope revolutionized the field by resolving the atomic structure of a conducting silicon surface in real space, revealing its atomic-scale features.[42] The 2014 Nobel Prize in Chemistry was awarded to William E. Moerner, Eric Betzig and Stefan W. Hell for their ground-breaking work in single molecule detection and the development of high-resolution microscopy techniques, underscoring the significant impact of SM experiments on advancing scientific understanding.[43–48] In SM fluorescence techniques, as shown in 1.1(a), individual molecules are detected by observing the light emitted by attached fluorophores which are first excited by absorbing light from an external source followed by internal conversion and vibrational relaxation, resulting in the emission of fluorescent light.[49–53] Single-molecule Förster Resonance Energy Transfer (smFRET) is a powerful technique that monitors conformational changes and molecular interactions within biomolecules, in real-time, occurring over the distances of 1-10 nm by measuring energy transfer between two fluorescent dyes attached to a molecule or molecular complex.[54, 55] Apart from these fluorescence SM studies, mechanical force-based experiments are also performed. Atomic force microscopy (AFM) provides detailed surface topography at the nanometer scale by scanning a sharp tip over a sample's surface to sense near-field physical interactions which generates three-dimensional images. As presented in Fig. 1.1(b), AFM uses a force-sensing tip mounted on a cantilever to scan the sample, causing deflections that are measured using a laser beam.[56] Optical tweezers, a pioneering discovery by Ashkin et. al. in 1986, that was eventually recognized as the 2018 Nobel Prize in Physics, is another force-based technique which employs tightly focused laser beams to hold (or trap) and manipulate microscopic objects.[57, 58] Unlike typical tweezers, they are contactless, causing minimal damage to the delicate biological or chemical systems under observation allowing researchers to conduct precise experiments.[59–61]

This technique is capable of trapping microscopic objects as small as 5 nm while exerting forces over 100 pN with resolutions as precise as 100 aN.[62] These breakthroughs laid the foundation for advancements in SM techniques that have revolutionized fields like molecular biology, biophysics, nanotechnology, and medicine, providing new insights into individual molecules' dynamics and functions.[17, 63]

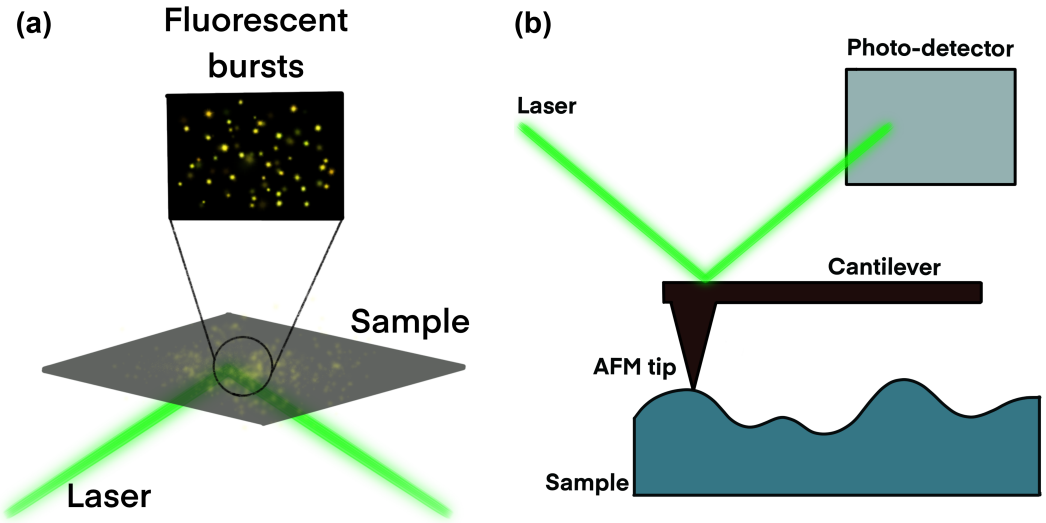


Figure 1.1: A simple schematic to show the working principle of (a) single molecule fluorescence imaging and (b) atomic force microscopy to scan the surface of a given sample.

The intrinsic dynamics within single proteins and nucleic acid molecules between their compact globular structure (folded state) and the denatured (unfolded state) are efficiently captured through these powerful imaging techniques. This allows researchers to identify and analyze transient intermediate states, different folding pathways, rare folding events and the sequence of structural changes.[64–70] Optical tweezers monitor folding-unfolding dynamics of such macromolecules via their extension lengths, allowing direct observation of any possible intermediate stable states, transition states as well as the microscopic transition times that eventually aids in understanding the kinetic mechanism.[71–74] Such conformational changes are also shown to influence ligand binding affinities and alternatively, sometimes these protein-ligand interactions may also induce conformational fluctuations through allosteric modulations.[75–78] Importantly, the complex process of G-protein coupled receptor activation via ligand/drug binding and consequent signal transduction pathways have also been critically explored through smFRET, proving their usefulness in pharmacology.[79–81] Ever since Ashkin et. al. demonstrated the use of optical tweezers in probing the behaviour and mechanical properties of single red blood cells,[58] it has revolutionized blood research by providing significant insights into the complex underlying processes.[82, 83] This allows for detailed analysis of the changes in the mechanical

properties of the red blood cells under various conditions while they pass through narrow capillaries, interact with other cells as well as in diagnosing diseases like malaria or sickle cell anaemia.[84–88] SM enzymology performed on individual enzymes such as motor proteins, nucleases, flavoenzymes, etc. uncovered significant insights about enzyme kinetics at the microscopic scale.[13, 16, 18, 19] Another example is the imaging of protein-DNA interactions which eventually serves as the initial step of several biochemical and biophysical processes in cells such as duplex DNA unwinding, damage repair, DNA degradation, etc.[89–93] Chen and coworkers have used SM fluorescence microscopy to observe individual product formation events catalysed by single nanocatalysts, providing insights into their catalytic activity and underlying molecular mechanisms.[94, 95] By measuring the subtle changes in the AFM tip's vibration frequency caused by the formation of chemical bonds between the tip and the surface atoms, Sugimoto et al. were able to chemically identify individual atoms of silicon, tin, and lead (that have similar chemical properties) on an alloy surface.[96] This technique is also powerful in resolving individual atoms within organic molecules, for instance, pentacene molecule adsorbed on both conducting and non-conducting surfaces.[97] Polymer growth and crystallisation can be viewed in real-time using AFM, shedding light on their morphologies and timescales of the involved phase transitions.[98, 99] Conceptually similar to optical tweezers, magnetic tweezers have been used to monitor growth of a single polymer in real-time which, interestingly, displays stochastic wait-and-jump events instead of linear and continuous growth.[100–102]

In conclusion, SM experiments have allowed researchers to isolate and manipulate individual molecules or particles, unveiling entirely new discoveries and revolutionizing the scientific landscape. Irrespective of the experimental procedure adopted, these measurements yield extensive quantitative data, for example the microscopic timescales of individual events, that reflect the minute details of stochastic processes under observation. These data are sufficient to produce probability distributions and their close exploration can offer deep insights into the underlying microscopic mechanisms and complexities of these processes, uncovering the fundamental intricacies and dynamic behaviors in both chemistry and biology. Additionally, these probability distributions are not limited to experimental data alone and they can also be derived using computer simulations and theoretical tools of statistical mechanics, which will be discussed in the following sections.

1.3 Stochastic Simulation Algorithm

Stochastic modelling of chemical and biological processes becomes relevant for small-volume systems since classical rate laws in terms of the species' concentrations fail to describe their kinetics and dynamic behaviour accurately. Apart from experimental techniques, simulation methods are also utilized to model stochastic systems which efficiently capture their discrete population levels while the system evolves in time.[103] One of the popular approach is the molecular dynamics (MD) simulation which tracks the position and velocity of all the interacting species in a fixed volume, during a chemical reaction, where individual molecules collide with each other.[104, 105] MD simulations tend to be computationally expensive since non-reactive molecular collisions are also tracked which can be easily ignored in the Stochastic Simulation Algorithm (SSA) technique developed by Gillespie.[7, 103, 106, 107] This numerical simulation procedure works under the assumption that the system is "well stirred" such that the positions of the interacting molecules is completely randomised. Thus, their movements need not be tracked as the system evolves and one can monitor the specific events that alter the reactant populations.[7] The algorithm is described below.[7, 9, 103, 106]

Within a fixed volume, consider a well-stirred system of N molecular species (in thermal equilibrium) described as $S_1, S_2, ..., S_N$ such that their exact population at any time t are represented as $X(t) \equiv \{X_1(t), X_2(t), ..., X_N(t)\}$. Thus, $X(t)$ describes the state of the system at any instant in time and $X_{i=1,N}(t)$ is a positive integer. During the course of reaction, all the species are allowed to interact through M chemical reaction events expressed via $R_1, R_2, ..., R_M$. In SSA, each reaction R_i is associated with two quantities. First, the state change vector $\nu_i \equiv \{\nu_{1,i}, \nu_{2,i}, ..., \nu_{N,i}\}$ which accounts for the modifications in the population of each species as a consequence of the reaction event R_i. Second, the propensity function a_i which describes the probability $a_i(\mathbf{x})dt$ that the system will undergo a reaction event R_i within an infinitesimal time step dt. Starting from a known current state, $X(t) = \mathbf{x}$, the numerical simulation procedure of the SSA determines two unknowns as the system evolves: the time interval for the next reaction to occur which is denoted by τ and the exact reaction event R_j that will proceed during this time. Note that the chosen reaction is denoted by the index j. Considering $a_0(\mathbf{x}) \equiv \sum_{i=1}^{M} a_i(\mathbf{x})$, these two unknowns are determined by generating two random numbers, r_1 and r_2 in the SSA. From the state $\mathbf{x}$ at time t, the time interval of the next reaction is

$$\tau = \frac{\ln(1/r_1)}{a_0(\mathbf{x})} \tag{1.1}$$

and the index j of the reaction is the smallest integer which satisfies the condition

$$\sum_{i=1}^{j} a_i(\mathbf{x}) > r_2 a_0(\mathbf{x}). \tag{1.2}$$

The system's time and population is then updated accordingly to $t + \tau$ and $\mathbf{x} + \nu_j$, respectively. The procedure is repeated until the desired output is achieved or the final time is reached. Detailed steps of the SSA are described below.

1. **System initialization:** Define all the N chemical species, $S_1, S_2, ..., S_N$ and the M reactions in the system, $R_1, R_2, ..., R_M$. Assign reaction rate constants (or the propensities) to each reaction. At the initial time t_0, specify the initial number of molecules for each species, $X(t_0) \equiv \{X_1(t_0), X_2(t_0), ..., X_N(t_0)\} \equiv \mathbf{x}_0$.

2. **Calculate propensity functions:** For each reaction, calculate the propensity functions $a_1(\mathbf{x}), a_2(\mathbf{x}), ..., a_M(\mathbf{x})$ from the product of the reaction rate constants and the population of the involved species at that instant. This quantifies how likely each reaction is to occur in a given time interval. The sum of all propensity functions is also determined, $a_0(\mathbf{x})$.

3. **The next reaction and its time interval:** Generate two uniformly distributed random numbers, r_1 and r_2. The time interval for the next reaction is determined from r_1 using Eq. (1.1) and the reaction that will be occur during this time is obtained by using r_2 in Eq. (1.2).

4. **Update system state:** Advance the simulation time by τ and update the number of molecules of each species according to the state change vector of the chosen reaction, ν_j.

6. **Repeat:** The process is repeated until the desired simulation time is reached or a stopping criterion is met.

The simulation steps 1-6 are performed several times to generate a large amount of quantitative data for the timescales of a stochastic process, which can be explored in detail further. This computation modelling approach is also known as the kinetic Monte Carlo

simulation technique.[7] SSA is widely applied to explore complex processes in chemistry and biology where randomness plays a crucial role. These numerical simulations have revealed diversity in identical cells due to underlying stochastic fluctuations in gene regulatory networks, allowing significant insights into how cells differentiate and adapt to environmental changes.[108–110] Importantly, the role of feedback loops in these networks is accurately captured via SSA and its modifications.[111, 112] Detailed understanding of dynamic fluctuations in enzymes, the consequent temporal variations as well as stochastic bursts of product formations is possible through this technique.[113–116] Lattice-based kinetic Monte Carlo simulations are popularly adopted to model stochastic chemical reaction networks occurring at surfaces such as complex heterogeneous catalytic processes, where mean-field approaches usually fail.[117–119] Such simulations are especially powerful in probing the intricacies of crystal growths on surfaces, also called epitaxy growth, which can be directly compared to existing experimental observations.[120, 121] All these examples illustrate the versatility of the SSA in offering crucial insights to complex stochastic processes, enabling ways to optimize materials and probing noise-induced fluctuations in biochemical reactions at the molecular scale.

1.4 Chemical Master Equations and First Passage Time Distributions

Although numerous analytical approaches have been developed in the literature to model stochastic processes and determine the distributions of relevant properties, such as microscopic timescales, this thesis will focus solely on those methods employed in the subsequent chapters. It is also stressed that this section is focused on discrete-state stochastic processes only, where all the possible states of a system are finite and distinct. The inherent stochasticity of any system causes all its states to be associated with some probability and thus, the evolution of these distinct states are described by a set of linear differential equations of the respective conditional probabilities. These set of equations are called the chemical master equations (CMEs). The simplest way to understand their mathematical formulation is by considering a birth-death process.[3] Consider a system consisting of a species X whose number population is expressed by x at any time t. The conditional probability of this state can be written as $P[x,t|x',t']$ which means that the species X constitutes x numbers at time t, given that it had x' population at the initial time t'. In a birth-death process, a finite number of X species are either generated or removed from

the system. Thus, from this current population x, it is assumed that either there is an increment by one unit with a transition probability $r^+(x)dt$ or the population decreases by one unit with a transition probability $r^-(x)dt$ during a small time-step dt.[3, 122] These situations can be illustrated as

$$\begin{aligned} x &\xrightarrow{r^+(x)dt} x+1 \\ x &\xrightarrow{r^-(x)dt} x-1. \end{aligned} \tag{1.3}$$

Under these considerations, one can monitor the evolution of the conditional probability $P[x,t|x',t']$ to arrive at either the forward master equation or the backward master equation. The mathematical framework to derive both these equations is discussed below.

A forward master equation (FME) is derived by focusing on all possible transitions in the population of the species X at the time t such that the system achieves a number population of x after a small time step dt in the future, resulting in the time $t+dt$. As illustrated in Fig. 1.2, the population x at time $t+dt$ can result in three possible ways during the time dt. First, the population $x+1$ at time t decreases by one unit with the probability $r^-(x+1)dt$. Second, the number $x-1$ at time t increases by one unit with the probability $r^+(x-1)dt$ and finally, the system could already possess x number of species at time t such that it remains unchanged even after a small time step in the future at time $t+dt$. Note that these transition probabilities are represented as r^-, r^+ and $1-(r^++r^-)$ in Fig. 1.2 for convenience. Now, the probability of finding x number of X species at some time $t+dt$, provided there were x' numbers of X at the initial time t' already can now be described as[3, 122]

$$\begin{aligned} P[x,t+dt|x',t'] = {} & r^+(x-1)dtP[x-1,t|x',t'] \\ & + r^-(x+1)dtP[x+1,t|x',t'] \\ & + \big\{1-\big(r^+(x)+r^-(x)\big)dt\big\}P[x,t|x',t']. \end{aligned} \tag{1.4}$$

In the above equation, the first term on the right side corresponds to an increase in the number population (during the time dt) by one unit along with the associated transition probability $r^+(x-1)dt$ whereas the second term represents the probability of a decrease in the number population by one unit associated with transition probability $r^-(x+1)dt$ during the small time dt. The third term represents the probability that the system population remains unchanged. Further simplification of Eq. (1.4) using the First Principle results in

$$\frac{dP[x,t|x',t']}{dt} = r^+(x-1)P[x-1,t|x',t'] + r^-(x+1)P[x+1,t|x',t'] - \{r^+(x)+r^-(x)\}P[x,t|x',t']. \tag{1.5}$$

This is the desired FME which represents the changes in a state observed 'forward' in time. For simplicity, the conditional probabilities in Eq. (1.5) may also be represented as $P(x,t)$ instead.

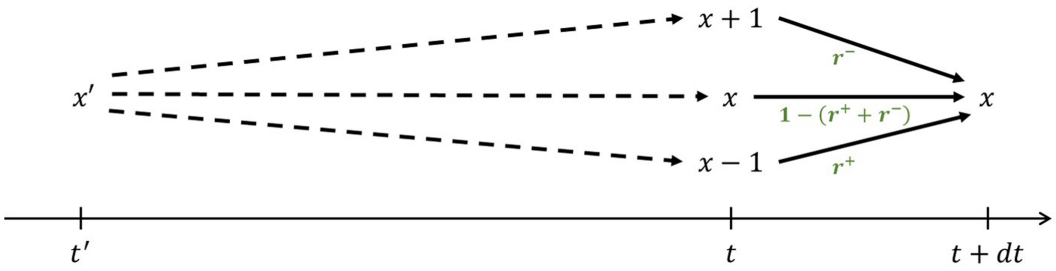


Figure 1.2: Possible transitions in the population of X species at the time t during a small time step of dt that eventually lead to a state with x numbers at the time $t+dt$. This consideration helps to derive the forward master equation of any stochastic process.

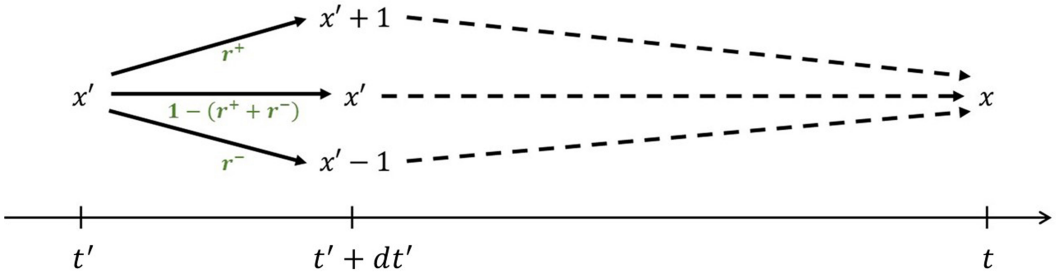


Figure 1.3: Possible transitions in the initial population of x' at the initial time t' after a small time step of dt' that eventually led to the current population of x at the time t. This consideration is useful in deriving the backward master equation of any stochastic process.

To derive a backward master equation (BME), all the possible transitions in the starting population x' at the initial time t' are taken into consideration.[3, 122] As shown in Fig. 1.3, during the small time step dt', the population x' at time t' can either increase by one unit to $x'+1$ or decrease by one unit resulting in $x'-1$ at time $t'+dt'$. Another possibility is that the population remains unchanged, x', even after a small time step dt'. These transitions are associated with respective transition probabilities $r^+(x')dt'$, $r^-(x')dt'$ and $1-[r^+(x')+r^-(x')]dt'$. It is to be noted that these transition probabilities are represented as r^+, r^- and $1-(r^++r^-)$ in Fig. 1.3 for convenience. Now, the probability of finding x number of X species at the time t, provided there were x' numbers of X at the initial time t' already can now be described as

$$\begin{aligned} P[x,t|x',t'] &= r^{+}(x')dt'P[x,t|x'+1,t'+dt'] + r^{-}(x')dt'P[x,t|x'-1,t'+dt'] \\ &+ \left\{1-\left(r^{+}(x')+r^{-}(x')\right)dt'\right\}P[x,t|x',t'+dt']. \end{aligned} \tag{1.6}$$

Further simplification of the above equation under the stationary state condition suggests that the conditional probabilities $P[x,t|x',t'+dt'] = P[x,t-dt|x',t']$ and thus, using the First Principle approach afterwards results in[3, 122]

$$\begin{aligned} \frac{dP[x,t|x',t']}{dt} &= r^{+}(x')P[x,t|x'+1,t'] + r^{-}(x')P[x,t|x'-1,t'] \\ &- \left\{r^{+}(x')+r^{-}(x')\right\}P[x,t|x',t']. \end{aligned} \tag{1.7}$$

The above equation is the desired BME which is very useful in determining the first passage time probability distributions (FPTDs) of a stochastic process. The time required by a stochastic process to reach a specified state for the first time, provided the initial state of the system was known, is termed as a first passage time.[122, 123] The inherent variability of these processes results in a collection of such times arising from multiple realizations of the same event, using which one develops a first passage time distribution to explore the system further. Importantly, since the course of any stochastic event is generally guided by probability, first passage time probability distributions are also constructed. This a powerful tool for understanding and predicting the timescales of stochastic events, offering significant insights to the underlying mechanisms of chemical and biological processes as well as the factors that influence these events.[122, 124–126] For instance, the search process of cellular proteins for their respective target sequences on the DNA is conveniently probed as a first passage event which allows researchers to efficiently explore the complex search dynamics at the microscopic scale.[127–129] Molecular motors transport cellular cargo along cytoskeletal filaments by converting chemical energy to mechanical work and a first passage time analysis allows for a detailed understanding of the stochastic timescales and the underlying kinetics of this intracellular transport mechanism.[130–132] Since the determination of first passage times require the knowledge of the system's initial state, it is sometimes challenging to observe all possible states in single molecule experiments and the microscopic time duration between successive desired events (or waiting times) is monitored instead, resulting in waiting time distributions.[13, 14, 133] The stochastic fluctuations in the microscopic reaction steps of catalytic processes, on either enzymes or nanoparticles, manifest in the waiting time as well as first passage time distributions and they provide significant insights into the reaction dynamics of the catalytic events at the

single molecule level.[17, 134–137] Exploring cell size regulation via stochastic growth and division of individual cells is crucial in identifying the factors that cause cells to maintain homeostasis despite molecular variations.[138, 139] While SM experiments capture the microscopic residence times of proteins in their folded and unfolded states, a first passage time analysis helps identify and predict the detailed dynamics of any possible intermediate states along with their associated pathways.[140–143] A recent theoretical study explored the influence of such folding/unfolding events on protein-ligand binding events through a detailed analytical framework.[144]

To summarise, the theoretical analysis of first passage times and their distributions offer a powerful tool for elucidating the complex temporal behaviours of stochastic processes in chemistry and biology, where the time evolution of specific events is critical. The investigation of such FPTDs allows researchers to predict the time scales of stochastic processes while identifying the factors that may influence their variability, thereby providing important insights into their underlying mechanisms. Such insights are crucial for optimizing experiments. Importantly, a comparison of theoretical FPTDs with corresponding experimental data is useful in validating analytical models and inferring associated parameter values which eventually assists in uncovering necessary aspects of the system's dynamics.

1.5 Quantifying Stochastic Fluctuations from Probability Distributions

Whether we perform single molecule experiments or analytical calculations, the existence of probability distributions arising from multiple timescales of the same stochastic event is clear evidence of inherent variability of these processes. The heterogeneity that persists in the reactivity of individual molecules in a population is referred to static disorder. On the other hand, dynamic disorder corresponds to the temporal fluctuations in the microscopic rate constants that have comparable or longer rates than the timescales of the entire stochastic event.[17, 133] For example, the dynamic disorder in single molecule enzyme kinetic reactions arise due to slow conformational changes ($\sim 10^{-3}$ to 10 seconds) of the single enzyme.[19] Individual metal nanoparticle-catalysed reactions possess temporal fluctuations due to dynamic surface restructuring that happens at rates comparable to the turnover times.[145]

Since the obtained probability distributions (say, $f(t)$) consist of the signatures of a

stochastic process, we may quantify them using well-known statistical quantitative measures called the moments of distribution. One can obtain these through the moment generation formula

$$\langle t^n \rangle = \int_0^\infty t^n f(t) dt. \tag{1.8}$$

Note that for single molecule experiments, t corresponds to the microscopic times between successive stochastic events (also called the waiting time) whereas it could also be the first passage time if the distribution is derived analytically. The first moment ($n = 1$) of the probability distribution provides us with its expected value, $\langle t \rangle$, which is also called the mean time of the distribution. It is now known that this average value is quantitatively similar to the reaction rates determined in ensemble measurements of the same process.[19, 145–147] In other words, the signatures of stochastic events are masked in ensemble measurements and the average time obtained from the distributions is not an appropriate quantity to probe intrinsic fluctuations of the system. The determination of the second moment ($n = 2$) of the probability distribution leads us to an important statistical quantity known as the randomness parameter.[17, 146] It is generally denoted as R and is a dimensionless ratio of the variance of the distribution to the square of the mean value,[146, 148]

$$R = \frac{\langle t^2 \rangle - \langle t \rangle^2}{\langle t \rangle^2}. \tag{1.9}$$

Randomness parameter is unity in the absence of stochastic fluctuations in the system and a single rate-determining step dominates the reaction events of the system. On the contrary, R deviates from unity in the presence of any stochastic fluctuations that arise due to multiple rate-determining steps (or dynamic disorder) in the process.[17, 146] Interestingly, earlier studies have shown that R's numerical value may also indicate the number of intermediate steps in a stochastic event.[146, 148, 149] The inverse of R provides an estimate of the number of rate-limiting discrete states. Importantly, the randomness parameter and its inverse serve as quantitative indicators of the shape of the probability time distributions.[146, 148] A single rate-determining step results in a distribution that follows a single exponential decay and R is equal to one in this situation. If the distribution exhibits a peak, it often suggests that the variance is less than the square of the mean value and the randomness parameter is less than unity. The value of $1/R$ will be

more than one in this case. Conversely, long and multi-exponential tailed distributions have a value of R greater than unity (or $1/R$ less than one).[148] Thus, $R < 1$ suggests a peaked distribution, $R > 1$ indicates the presence of multi-exponential tails and $R = 1$ corresponds to a single-exponential distribution. In this manner, the randomness parameter allows for the characterization of time distributions. A crucial feature of R is that it can be determined conveniently from SM experiments as well, allowing detailed interpretation of microscopic reaction dynamics obscured by ensemble measurements. English et. al. have shown that a single enzyme's catalytic reaction mechanism undergoes stochastic fluctuations at intermediate substrate concentrations where R is less than one. At low and high substrate concentrations, the randomness parameter is close to unity and the system has no dynamic disorder.[19]

There are other useful statistical parameters, as well, that are occasionally utilised to probe stochastic fluctuations in the system. Fano factor is widely used to quantify the flux-associated fluctuations in any stochastic process.[150] For example, the microscopic mechanisms of transcriptional bursting in cells can be explored through fano factor which quantifies the noise with respect to the number of RNA transcripts (or the flux) produced.[151, 152] It is mathematically represented as

$$F = \frac{\langle n^2 \rangle - \langle n \rangle^2}{\langle n \rangle} \tag{1.10}$$

where n corresponds to the number of molecules produced or the net flux in a single stochastic event. Another common parameter is the Poisson indicator which can determine whether the statistical behavior of the available data is Poissonian or non-Poissonian and it is expressed as

$$P = \frac{\langle t^2 \rangle - 2\langle t \rangle^2}{\langle t \rangle^2}. \tag{1.11}$$

P equals zero if the mean of the distribution is equal to its variance, i.e., the distribution reflects a Poissonian behavior. Any deviation from this zero value suggests the presence of multiple competing reaction timescales in the system.[153]

The usual steps of theoretical analysis employed in this thesis are as follows. A function $F_{x'}(t)$ is generally defined to mathematically denote the first passage time probability distribution of a system to reach a desired state x at the time t for the first time, given that

it was initially in the state x'. The time evolution of these probabilities, for all states of the system, can then be described using the CME approach. By invoking necessary boundary conditions, the resulting set of linear differential equations are conveniently solved in the Laplace space by performing a Laplace transform, $\tilde{F}_{x'}(s) = \int_0^\infty F_{x'}(t)e^{-st}dt$. It must be noted that irrespective of the approach adopted to write these master equations, FME or BME, the same first passage time distribution will be obtained. The underlying properties of the stochastic process being observed can be successfully quantified by exploring the characteristics of the resulting distributions. In the next few sections, selected stochastic processes that are explored in this thesis will be briefly discussed.

1.6 Proteins' Search for Target DNA Sequences

Genetic blueprints for performing various biological processes are constituted in large polynucleotide sequences that form deoxyribonucleic acid or DNA. These important processes are usually initiated by a successful binding of proteins to their specific DNA-binding target sequences.[31, 32] For example, DNA-repair enzymes must find damaged sequences buried amid many undamaged nucleotide residues of the genome in order to preserve genomic integrity.[154] It is required that transcription factor proteins bind to their corresponding gene promoter regions in order to activate or repress a gene.[31, 155] Hence, comprehending these complex protein-DNA interactions will enhance our understanding of diverse proteins with their distinct search strategies.

In 1970, Riggs et al. made a ground-breaking discovery that the lac repressor protein's rate of association with its target is faster than its 3D diffusion rate by about two orders in magnitude.[156] Clearly, this suggests that the search process is much more complex and proteins might not perform simple diffusion within the cell. These striking findings paved way for a thorough investigation of the underlying dynamics of this complex target search process of the proteins, both experimentally and theoretically.[128, 157–162] The articles by Berg, Winter and von Hippel are widely considered to be the most influential work in this field where they postulated that proteins locate their targets on DNA by several search mechanisms, including sliding and hopping.[163–165] Single molecule investigations have significantly enhanced our understanding of protein-DNA interactions and the process of target search on DNA in the recent years.[158, 161, 166–169] It is now well-established through a large amount of SM experimental studies, supported by theoretical efforts, that proteins find their respective targets via facilitated diffusion mechanism which constitutes

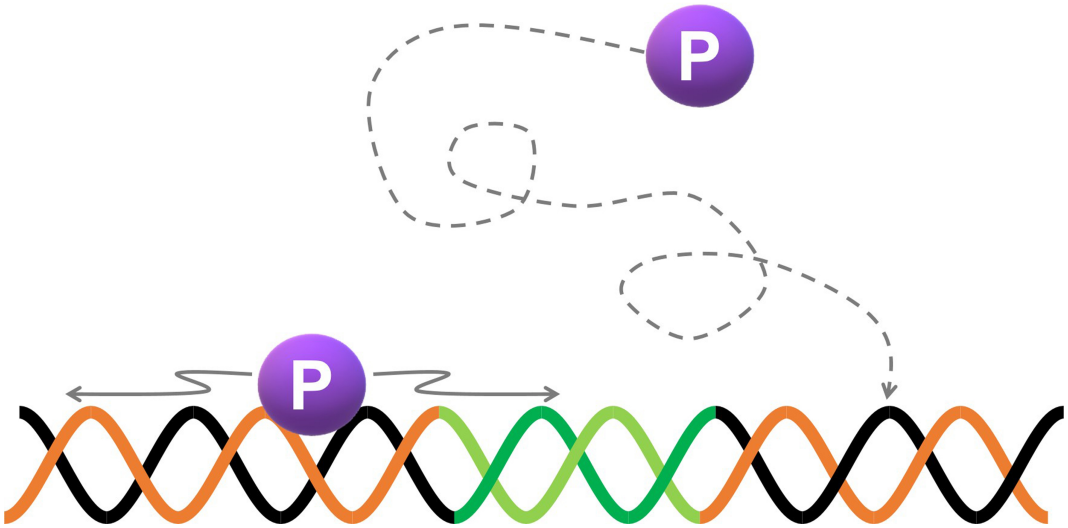


Figure 1.4: A simple schematic depicting the target search of a DNA-binding protein (in purple) via facilitated diffusion mechanism. The target sequence on the DNA is in green. Solid and dashed grey arrows represent the 1D and 3D diffusion based search modes of the protein, respectively.

3D and 1D diffusion based search modes in biological cells.[157, 158, 161, 166, 170, 171] In the 3D search pathway, as shown in Fig. 1.4, the protein performs 3D diffusion within the cell cytoplasm while constantly binding and unbinding from the DNA until the target site is recognised. Once bound to the DNA, the protein may also perform a relatively slower 1D random walk until it finds the target. It is believed that a protein adopts a combination of these search pathways in order to accomplish its search.

Nevertheless, there lacks a complete elucidation for this complex process's underlying dynamics and various theoretical concepts about facilitated diffusion have been examined and addressed. These include electrostatic effects, exploring 3D and 1D pathways, conformational transitions, etc.[170, 172] Continuum models indicate that the mean search time to discover the target, without sliding, is infinite. However, this prediction is nonphysical since the protein may still explore the DNA in the three-dimensional regime.[173] Among several theoretical and simulation efforts to explore the intricacies of this complex process, Kolomeisky and coworkers adopted a discrete-state stochastic approach to model the system.[173] Assuming only a single protein and a DNA being present in the system, in this study, the entire DNA was divided into distinct states consisting equal-sized sequences and the cytoplasm constituted a separate state. The protein molecule, as a consequence, can occupy any of these discrete states at any instant as the system evolves. Considering the search of a protein begins from the cytosol, a first passage time analysis was performed to evaluate the average target search time. This simple theoretical approach was successful

in exploring the dynamics through 1D and 3D search modes as proposed in the facilitated diffusion mechanism. However, the analytical picture was far from in vivo conditions. This model has thus been extended in several directions to probe the influence of more realistic cellular conditions to the search, for example, in the presence of molecular traps,[174] DNA looping,[175, 176] multiple conformers,[177] sequence heterogeneity of DNA,[178] etc.

Since macromolecules comprise around 10-40% of the total intracellular volume,[179–181] they can influence ongoing biological processes in the cell. In the context of the target recognition problem, these macromolecules can either act as obstacles while being attached to the DNA or they can exist in the cellular environment and impact the protein's diffusion. SM studies have shown that crowding introduced by poly(ethylene glycol) (PEG) polymers enhance DNA hydrolysis via endonucleases, while it remains unaffected for exonucleases.[182] DNA translocation by uracil glycosylase is significantly enhanced in human cells compared to corresponding in vitro studies, highlighting the importance of nonspecific interactions in the nuclear environment.[183] Extensive simulations show protein-DNA affinity changes with crowder molecules, with their size and mass affecting the search for different salt concentrations.[184–186] Through extensive computer simulations, Bhattacherjee and coworkers have thoroughly examined the impact of molecular crowders and their properties in this process.[187–190] Although the influence of obstacles has been investigated theoretically using discrete-state models,[175, 191, 192] the role of bulk crowders is still not well-understood and needs a close attention which is addressed in the present thesis.

1.7 Transport of ssDNA through Biological Nanopores

Selective transport of various biomolecules across cellular membranes forms an essential life process. Protein nanopores allow the translocation of charged polymers, including RNA and single-stranded DNA (ssDNA), through cell membranes.[193] These polymers are first captured at the mouth of individual nanopores, which is the rate-limiting stage of the transport process, before they can be translocated across membranes as shown in Fig. 1.5. Although a difficult process to monitor in vivo, Kasianowicz et al. detected individual ssDNA or RNA molecules, driven by an applied electric field, as they traverse an α-hemolysin (αHL) protein nanopore.[194] This heptameric pore has a length of around 10nm and a width of 2 nm.[195] As shown in Fig. 1.5, it spans a biological membrane, dividing it into two chambers known as the *cis* side and the *trans* side. Distinctive charge distribution

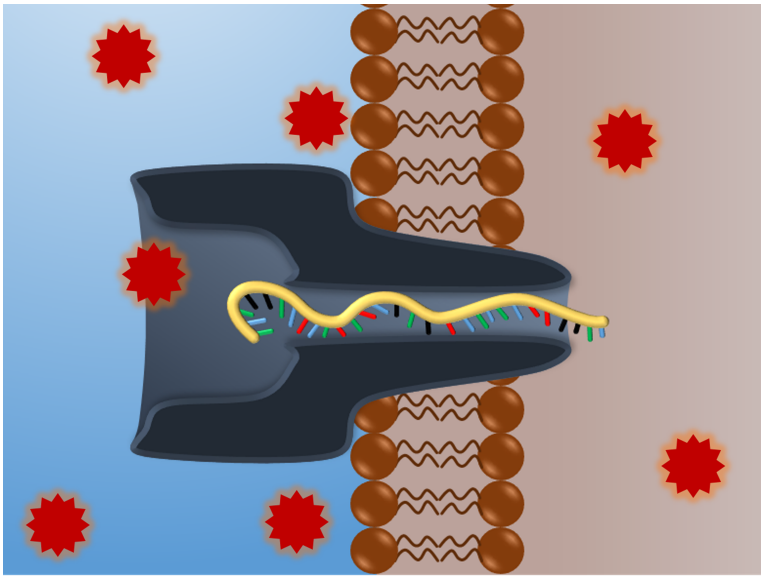


Figure 1.5: A cartoon description of a ssDNA transport across a cell membrane through an α-hemolysin protein nanopore, from the *cis* (left) to the *trans* (right) side. Red colored shapes represent other biomolecules, present in the cytoplasm, that act as bulk crowders during this process.

on each nucleotide allowed for the tracking of current variations as individual base pairs pass through the nanopores. Additionally, the duration of time for a single polymer chain to translocate the nanopore can be determined from the current blockades, eventually providing the translocation time. This technique eventually led to the utilisation and development of biological and synthetic nanopore sensors during genome sequencing.[196, 197]

To efficiently characterize the produced current profiles, it is critical to enhance the precision of nanopore sensors by controlling the system's charge distribution.[198–200] While rapid polymer capture is essential for high sensor throughput, it should stay in the nanopores long enough to be individually detected for each charged nucleotide.[201] Several possible solutions have been proposed to meet these ideal sensor requirements. It seems reasonable that a stronger applied electric field throughout the system would lead to a faster capture but it also makes the translocation faster, making sequencing more challenging. Electric field manipulation,[202] optical tweezers to slow down translocation,[203] salt gradients throughout the nanopores,[204] and higher solution flow rates to promote rapid DNA capture[205] are some of the solutions put forward for this problem. Additionally, polymer translocation has been extensively explored using large scale simulation and analytical techniques with a specific attention on the determination of translocation time (τ). Sung and Park performed the initial theoretical investigation in this field and

then Muthukumar expanded the same, specifically focusing on the scenario of unbiased translocation.[206, 207] The translocation process was conceptualized as a 1D diffusion problem, with the monomers on the *trans* side of the pore serving as the reaction coordinate. For an unimpeded motion of a polymer having length N, 2D lattice simulations show that $\tau \sim N^{2.5}$ whereas if there exists a chemical potential difference ($\Delta\mu$) across the membrane, then $\tau \sim N^{1.53}/\Delta\mu$. While Chaudhury et al. adopted a generalized Langevin equation (GLE) based formalism to model the anomalous dynamics of polymer translocation and eventually obtain the scaling behaviors of τ,[208] Dubbeldam et al. have used a fractional Fokker Planck equation (FPPE) based approach of probing the anomalous transport.[209, 210] All these studies provide useful insights about the intricate nature of the process, especially translocation.

Owing to the ubiquitous nature of bio-macromolecules in real cellular conditions, they may also influence this transport process. SM studies performed in vitro by using poly(ethylene glycol) (PEG) polymers in solution show that other biomolecules may also partition within the nanopores, apart from the desired polymer's transport.[211–214] PEG crowders in solutions are shown to naturally assist the transport of charged nucleotides across membranes. As demonstrated by Freedman et al., the apparent crowding of proteins near the nanopore increases the capture rate.[215] Rapid capture of the polypeptide Syn B2 within the β-barrel of the αHL nanopore is facilitated by highly concentrated large-size PEGs in the solutions.[216] Yao et al. have investigated the capture and translocation of ssDNA by introducing PEG crowders on the *cis* side of the αHL nanopore in different sizes and concentrations.[217] It was observed that ssDNA capture is facilitated by high concentrations of optimum-size crowders, whereas translocation via the nanopore is slowed down. Not limited to biological nanopores only, nucleic acid and protein detection is enhanced in solid-state nanopores in the presence of macromolecular crowding as well.[218, 219] Chen and Luo have theoretically demonstrated that the direction of polymer translocation is guided by the size of the crowding agents if their area fractions are equal on both sides of the nanopore, mostly because of their conformational entropy.[220] Despite these available results from SM experiments, a microscopic understanding of the crowders' role in assisting ssDNA transport is missing. By adopting the concepts of statistical mechanics and discrete-state modelling strategies, in this thesis, the capture and translocation of ssDNA through an αHL nanopore is probed in crowded conditions.

1.8 SM Nanoparticle Catalysis

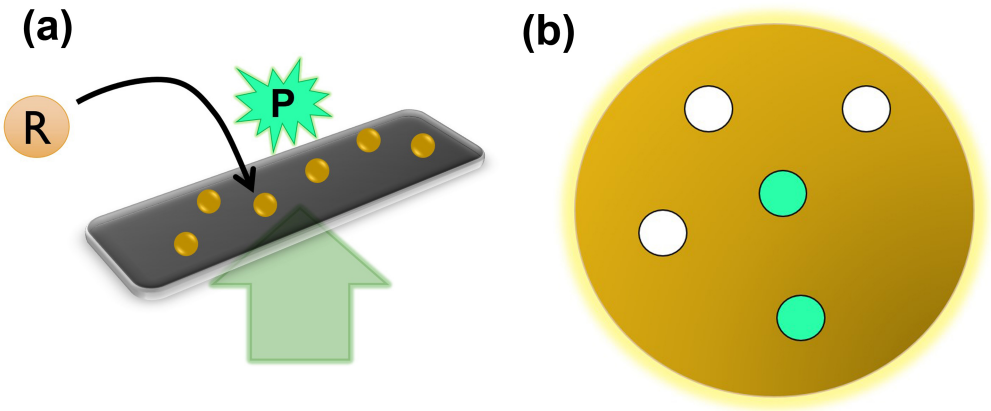


Figure 1.6: (a) Schematic of a chemical reaction catalyzed by metal nanoparticles, illuminated by laser, using SM fluorescence microscopy technique. Metal nanoparticles (in yellow) are immobilized on a surface (in black). A non-fluorescent reactant (R) molecule transforms into a fluorescent product (P) molecule. (b) Magnified view of a single nanoparticle with many active sites. Green filled circles represent newly formed products at an active site while white filled circles correspond to the situation where no products are formed yet.

Catalysts are vital in current industrial operations and scientific research activities since they expedite necessary chemical reactions.[221, 222] Although catalysis is of great relevance and has been extensively studied for many years, the molecular principles behind catalytic events are still not well-understood.[223, 224] Recent advancements in experimental techniques for investigating catalytic reactions, particularly SM methods, have significantly contributed to the discovery of molecular characteristics of catalytic systems. They have revealed remarkable molecular details of chemical reactions on single metal nanoparticle catalysts that possess multiple active sites.[225, 226] The powerful technique of SM fluorescence microscopy, see Fig. 1.6(a), captures microscopic details of catalytic reactions on single nanoparticles.[145] As shown in Fig. 1.6(b), individual nanocatalysts possess a large amount of active sites on their surface where chemical reactions occur simultaneously and each product formation event is successfully detected at the molecular resolution using SM fluorescence microscopy. These nanocatalysts undergo dynamic restructuring,[145] show size-dependent,[227] facet-dependent catalytic reactivity[228] at SM resolution leading to spatio-temporal fluctuations which is also known as dynamic disorder. SM studies have also proven useful in revealing kinetic properties and previously unknown intermediate steps during chemical reaction catalyzed by gold nanoparticles that are usually obscured in ensemble measurements.[229]

Such large quantitative data from these experiments has allowed for several theoretical descriptions for the underlying complex dynamics of these nanocatalysts. Xu et al. proposed the first study in this direction based on Langmuir-Hinshelwood theory, which postulates an equilibrium between the binding and unbinding of the substrates to the catalyst surface. This approach considers all active sites as a unified catalytic site.[94] Nevertheless, this methodology was unable to provide any explanation for the experimentally observed variations in catalytic activity and underlying stochasticity. Later, a more sophisticated theoretical approach was presented by Chaudhury and coworkers, utilizing first-passage analysis which considers the number of active sites and the associated stochastic effects to some extent.[134, 230] A novel discrete-state theoretical framework was later established to study the kinetics of chemical reactions including intermediate states on individual catalytic particles.[231] The proposed method considered the stochastic effects of individual chemical reactions at each catalytic site, allowing a direct interpretation of SM experimental observations. This thesis investigates an effective system of discrete states to probe catalytic reactions on a single metal nanoparticle.

In one remarkable study, Zou et al. observed that redox chemical reactions catalysed by individual nanorods and nanoplates display cooperative effects.[232] It was demonstrated that the chemical processes occurring at a particular active site can efficiently trigger similar reactions at nearby locations. After dividing each nanoparticle into equal-sized segments, statistical analysis of Pearson's cross-correlation coefficients between subsequent time-dependent reactions were determined. The results revealed the existence of apparent intra-particle communications between spatially close catalytic sites resulting in stimulation of chemical reactions.[232] These cooperative effects are universally observed for different redox chemical reactions studied in experiments and are characterised by lifetimes as well as communication distances. Through electric field variations, it was shown that these intra-particle communications might originate via some positively charged species but the molecular mechanism of these surprising observations remain not well-understood. The present thesis addresses the microscopic origins of these cooperative effects through the theoretical tools introduced in previous sections.

1.9 Thesis Chapters in Brief

This thesis investigates the microscopic mechanisms of stochastic processes relevant to chemistry and biology using discrete-state stochastic modeling frameworks. Broadly, three

different processes are studied. In **chapters 2-4**, the DNA target search process of proteins in cellular conditions is explored. The impact of macromolecular crowders on a protein molecule's search for target sequences on DNA is probed, revealing how the nature of crowders, the length and position of crowding and non-specific crowder-protein interactions may control the search process. The ubiquitous nature of these crowders may influence other biological processes as well and in **chapters 5 and 6**, the capture and translocation of ssDNA through biological nanopores in crowded conditions is probed. This investigation provides precise analytical solutions applicable to single-molecule experiments. In the context of chemical processes, the final section of thesis, including **chapters 7-9**, focuses on understanding the molecular mechanisms of heterogeneous catalytic reactions mediated by individual nanoparticles. Our theoretical investigation provides a quantitative microscopic understanding of catalytic cooperativity between different active sites on a single nanocatalyst.

Chapters 2-4 explore the search of a single DNA-binding protein molecule for its respective target sequence on the DNA in real cellular conditions. By dividing the entire system into discrete states in which a protein molecule can be found, **chapters 2 and 3** discuss the effect of bulk macromolecular crowders on the search process with respect to their sizes, mobility rates and non-specific interactions with protein are probed. With supporting Monte Carlo simulations, the analytical results of the average target search time provide a microscopic understanding of the favourable situations that may accelerate or retard the process, thereby, offering physical explanations of existing single-molecule experimental observations and simulation studies. **Chapter 4** studies the proximity effects between a single sub-diffusing protein and its target, at the beginning of the search, by adopting a continuum description instead of discrete-state approach. This work analytically calculated the distance-dependent first passage time distributions and eventually quantified the most-probable first passage times in addition to the mean values. A direct application of our results to existing single-molecule measurements is also illustrated.

Given the ubiqutous nature of bulk cellular crowders, they may also impact other biological processes. Accordingly, in **chapters 5 and 6**, a theoretical investigation on the transport of a single-stranded DNA (ssDNA) through biological nanopores under crowded conditions is performed. A novel discrete-state stochastic model is developed to probe the capture as well translocation of a ssDNA as is passes through an α-hemolysin nanopore (αHL), while the solution is crowded with polyethylene glycol (PEG) molecules. Exact analytical quantities are evaluated that are directly applicable to single-molecule exper-

iments. While **chapter 5** specifically explores the role of crowder concentration on the transport process, **chapter 6** addresses the effect of PEG size as well as the influence of heterogeneously crowded solutions on the capture process specifically.

The final **chapters 7-9** address the molecular mechanisms of heterogeneous catalytic reactions facilitated by individual metal nanoparticles. In **chapter 7**, a discrete-state stochastic model is presented that probes the reaction dynamics of catalytic events occurring on a single nanoparticle that possesses multiple active sites. The influence of reaction intermediates, number of active sites and reaction scheme (topology) on the mean turnover times, first passage time distributions and the randomness parameter are explored analytically as well as through supporting Monte Carlo computer simulations. The final chapters of the thesis, **chapters 8 and 9**, discuss the probable microscopic mechanisms for the origins of intra-particle cooperative communications, observed through subsequent product formation events arising from redox reactions, between spatially close active sites of nanocatalysts.[232] **Chapter 8** describes the proposed theoretical approach where the probability for a redox reaction to fire up depends on the local concentration of charged holes originating from each reaction event. Since most of the experiments were performed on nanorods, a 1D description was adopted to probe the dynamics of these holes through an exactly solvable discrete-state stochastic framework by monitoring the evolution of a single charged hole until it disappears. The analytical results quantitatively agree with experimentally observed values of communication times (t_0) and distances (x_0). In **chapter 9**, this simple 1D discrete-state stochastic approach is extended to probe the role of surface complexity of nanoparticles, due to their intrinsic heterogeneity, on the catalytic communications. Obtained results predicted a decrease in x_0 and t_0 due to surface heterogeneity while a simultaneous increase in the communication strengths, clarifying several aspects of intra-particle cooperative communications on heterogeneous nanoparticles.

In summary, the interdisciplinary research in this thesis highlights the effectiveness of discrete-state stochastic modeling techniques in revealing the intricacies of complex chemical and biological systems, which cannot be achieved through ensemble measurements, offering valuable insights that enhance both theoretical and experimental research.

Bibliography

(1) Van Kampen, N. G., *Stochastic Processes in Physics and Chemistry*; Elsevier: 1992; Vol. 1.

(2) Bressloff, P. C., *Stochastic Processes in Cell Biology*; Springer: 2014; Vol. 41.

(3) Gardiner, C. W., *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*; Springer: 2004.

(4) Teimouri, H.; Kolomeisky, A. B. Power of Stochastic Kinetic Models: From Biological Signaling and Antibiotic Activities to T Cell Activation and Cancer Initiation Dynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2022**, *12*, e1612.

(5) Laidler, K. J., *Chemical Kinetics*; Pearson Education: 1987.

(6) Érdi, P.; Lente, G., *Stochastic Chemical Kinetics: Theory and (Mostly) Systems Biological Applications*; Springer: 2014.

(7) Gillespie, D. T. Stochastic Simulation of Chemical Kinetics. *Annu. Rev. Phys. Chem.* **2007**, *58*, 35–55.

(8) Mira, J. et al. Two Examples of Deterministic Versus Stochastic Modeling of Chemical Reactions. *J. Chem. Educ.* **2003**, *80*, 1488.

(9) Kumar, A. B.; Ramaswamy, R. Chemistry at the Nanoscale: When Every Reaction is a Discrete Event. *Resonance* **2018**, *23*, 23–40.

(10) Delbrück, M. Statistical Fluctuations in Autocatalytic Reactions. *J. Chem. Phys.* **1940**, *8*, 120–124.

(11) McQuarrie, D. A. Stochastic Approach to Chemical Kinetics. *J. Appl. Probab.* **1967**, *4*, 413–478.

(12) Bartholomay, A. F. A Stochastic Approach to Statistical Kinetics with Application to Enzyme Kinetics. *Biochemistry* **1962**, *1*, 223–230.

(13) Xie, X. S.; Lu, H. P. Single-Molecule Enzymology. *J. Biol. Chem.* **1999**, *274*, 15967–15970.

(14) Min, W. et al. Fluctuating Enzymes: Lessons from Single-Molecule Studies. *Acc. Chem. Res.* **2005**, *38*, 923–931.

(15) Schwabe, A. et al. Exploration of the Spontaneous Fluctuating Activity of Single Enzyme Molecules. *FEBS Lett.* **2013**, *587*, 2744–2752.

(16) Smiley, R. D.; Hammes, G. G. Single Molecule Studies of Enzyme Mechanisms. *Chem. Rev.* **2006**, *106*, 3080–3094.

(17) Singh, D. et al. Theoretical Tools to Quantify Stochastic Fluctuations in Single-Molecule Catalysis by Enzymes and Nanoparticles. *ACS Omega* **2022**, *7*, 47587–47600.

(18) Lu, H. et al. Single-Molecule Enzymatic Dynamics. *Science* **1998**, *282*, 1877–1882.

(19) English, B. P. et al. Ever-Fluctuating Single Enzyme Molecules: Michaelis Menten Equation Revisited. *Nat. Chem. Biol.* **2006**, *2*, 87–92.

(20) Kramers, H. A. Brownian Motion in a Field of Force and the Diffusion Model of Chemical Reactions. *Physica* **1940**, *7*, 284–304.

(21) Pollak, E.; Miret-Artés, S. Recent Developments in Kramers' Theory of Reaction Rates. *ChemPhysChem* **2023**, *24*, e202300272.

(22) Hänggi, P. et al. Reaction-Rate Theory: Fifty Years after Kramers. *Rev. Mod. Phys.* **1990**, *62*, 251.

(23) Singh, P. S.; Lemay, S. G. Stochastic Processes in Electrochemistry. *Anal. Chem.* **2016**, *88*, 5017–5027.

(24) Gabrielli, C. et al. Fluctuations in Electrochemical Systems. I. General Theory on Diffusion Limited Electrochemical Reactions. *J. Chem. Phys.* **1993**, *99*, 7232–7239.

(25) Lu, S.-M. et al. Understanding the Dynamic Potential Distribution at the Electrode Interface by Stochastic Collision Electrochemistry. *J. Am. Chem. Soc.* **2021**, *143*, 12428–12432.

(26) Xiong, J. et al. Ion Transport within High Electric Fields in Nanogap Electrochemical Cells. *ACS Nano* **2015**, *9*, 8520–8529.

(27) Siddique, N. et al. Stochastic Reconstruction and Electrical Transport Studies of Porous Cathode of Li-ion Batteries. *J. Power Sources* **2012**, *217*, 437–443.

(28) Kuno, M. et al. "On"/"Off" Fluorescence Intermittency of Single Semiconductor Quantum Dots. *J. Chem. Phys.* **2001**, *115*, 1028–1040.

(29) Haase, M. et al. Exponential and Power-Law Kinetics in Single-Molecule Fluorescence Intermittency. *J. Phys. Chem. B* **2004**, *108*, 10445–10450.

(30) Heuff, R. F. et al. Fluorescence Correlation Spectroscopy using Quantum Dots: Advances, Challenges and Opportunities. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1870–1880.

(31) Alberts, B. et al., *Molecular Biology of the Cell*; Garland Science, Taylor and Francis Group: 2014.

(32) Lodish, H. F. et al., *Molecular Cell Biology*; WH Freeman New York: 2006.

(33) Zhou, T.; Liu, T. Quantitative Analysis of Gene Expression Systems. *Quant. Biol.* **2015**, *3*, 168–181.

(34) Raj, A.; van Oudenaarden, A. Single-Molecule Approaches to Stochastic Gene Expression. *Annu. Rev. Biophys.* **2009**, *38*, 255–270.

(35) Rangamani, P.; Iyengar, R. Modelling Cellular Signalling Systems. *Essays Biochem.* **2008**, *45*, 83–94.

(36) Bigman, L. S.; Levy, Y. Protein Diffusion Along Protein and DNA Lattices: Role of Electrostatics and Disordered Regions. *Annu. Rev. Biophys.* **2023**, *52*, 463–486.

(37) Chua, G. N.; Liu, S. When Force Met Fluorescence: Single-Molecule Manipulation and Visualization of Protein–DNA Interactions. *Annu. Rev. Biophys.* **2024**, *53*.

(38) Tan, X. et al. Single-Molecule Study of Protein-Protein Interaction Dynamics in a Cell Signaling System. *J. Phys. Chem. B* **2004**, *108*, 737–744.

(39) Cukuroglu, E. et al. Hot Spots in Protein–Protein Interfaces: Towards Drug Discovery. *Prog. Biophys. Mol. Biol.* **2014**, *116*, 165–173.

(40) Calebiro, D.; Sungkaworn, T. Single-Molecule Imaging of GPCR Interactions. *Trends Pharmacol. Sci.* **2018**, *39*, 109–122.

(41) Tamarat, P. et al. Ten Years of Single-Molecule Spectroscopy. *J. Phys. Chem. A* **1999**, *104*, 1–16.

(42) Binnig, G. et al. 7× 7 Reconstruction on Si (111) Resolved in Real Space. *Phys. Rev. Lett.* **1983**, *50*, 120.

(43) Moerner, W. E.; Kador, L. Optical Detection and Spectroscopy of Single Molecules in a Solid. *Phys. Rev. Lett.* **1989**, *62*, 2535.

(44) Dickson, R. M. et al. On/Off Blinking and Switching Behaviour of Single Molecules of Green Fluorescent Protein. *Nature* **1997**, *388*, 355–358.

(45) Hell, S. W.; Wichmann, J. Breaking the Diffraction Resolution Limit by Stimulated Emission: Stimulated-Emission-Depletion Fluorescence Microscopy. *Opt. Lett.* **1994**, *19*, 780–782.

(46) Betzig, E. et al. Near Field Scanning Optical Microscopy (NSOM): Development and Biophysical Applications. *Biophys. J.* **1986**, *49*, 269–279.

(47) Betzig, E. et al. Breaking the Diffraction Barrier: Optical Microscopy on a Nanometric Scale. *Science* **1991**, *251*, 1468–1470.

(48) Betzig, E.; Chichester, R. J. Single Molecules Observed by Near-Field Scanning Optical Microscopy. *Science* **1993**, *262*, 1422–1425.

(49) Shashkova, S.; Leake, M. C. Single-Molecule Fluorescence Microscopy Review: Shedding New Light on Old Problems. *Biosci. Rep.* **2017**, *37*, BSR20170031.

(50) Weiss, S. Fluorescence Spectroscopy of Single Biomolecules. *Science* **1999**, *283*, 1676–1683.

(51) Ha, T. Single-Molecule Fluorescence Methods for the Study of Nucleic Acids. *Curr. Opin. Struct. Biol.* **2001**, *11*, 287–292.

(52) Cordes, T.; Blum, S. A. Opportunities and Challenges in Single-Molecule and Single-Particle Fluorescence Microscopy for Mechanistic Studies of Chemical Reactions. *Nat. Chem.* **2013**, *5*, 993–999.

(53) Kondo, T. et al. Single-Molecule Fluorescence Spectroscopy of Photosynthetic Systems. *Chem. Rev.* **2017**, *117*, 860–898.

(54) Sasmal, D. K. et al. Single-Molecule Fluorescence Resonance Energy Transfer in Molecular Biology. *Nanoscale* **2016**, *8*, 19928–19944.

(55) Bacic, L. et al. Recent Advances in Single-Molecule Fluorescence Microscopy Render Structural Biology Dynamic. *Curr. Opin. Struct. Biol.* **2020**, *65*, 61–68.

(56) Dufrêne, Y. F. Towards Nanomicrobiology using Atomic Force Microscopy. *Nat. Rev. Microbiol.* **2008**, *6*, 674–680.

(57) Ashkin, A. et al. Observation of a Single-Beam Gradient Force Optical Trap for Dielectric Particles. *Opt. Lett.* **1986**, *11*, 288–290.

(58) Ashkin, A. et al. Optical Trapping and Manipulation of Single Cells using Infrared Laser Beams. *Nature* **1987**, *330*, 769–771.

(59) Moffitt, J. R. et al. Recent Advances in Optical Tweezers. *Annu. Rev. Biochem.* **2008**, *77*, 205–228.

(60) Gennerich, A., *Optical Tweezers*; Springer: 2017; Vol. 1486.

(61) Bustamante, C. J. et al. Optical Tweezers in Single-Molecule Biophysics. *Nat. Rev. Methods Primers* **2021**, *1*, 25.

(62) Grier, D. G. A Revolution in Optical Manipulation. *Nature* **2003**, *424*, 810–816.

(63) Ritort, F. Single-Molecule Experiments in Biological Physics: Methods and Applications. *J. Phys. Condens. Matter* **2006**, *18*, R531.

(64) Orte, A. et al. Evidence of an Intermediate and Parallel Pathways in Protein Unfolding from Single-Molecule Fluorescence. *J. Am. Chem. Soc.* **2008**, *130*, 7898–7907.

(65) Kamagata, K. et al. Long-Term Observation of Fluorescence of Free Single Molecules to Explore Protein-Folding Energy Landscapes. *J. Am. Chem. Soc.* **2012**, *134*, 11525–11532.

(66) Michalet, X. et al. Single-Molecule Fluorescence Studies of Protein Folding and Conformational Dynamics. *Chem. Rev.* **2006**, *106*, 1785–1813.

(67) Banerjee, P. R.; Deniz, A. A. Shedding Light on Protein Folding Landscapes by Single-Molecule Fluorescence. *Chem. Soc. Rev.* **2014**, *43*, 1172–1188.

(68) Zhuang, X. et al. A Single-Molecule Study of RNA Catalysis and Folding. *Science* **2000**, *288*, 2048–2051.

(69) Pirchi, M. et al. Single-Molecule Fluorescence Spectroscopy Maps the Folding Landscape of a Large Protein. *Nat. Commun.* **2011**, *2*, 493.

(70) Dufrêne, Y. F. et al. Imaging Modes of Atomic Force Microscopy for Application in Molecular and Cell Biology. *Nat. Nanotechnol.* **2017**, *12*, 295–307.

(71) Neupane, K. et al. Direct Observation of Transition Paths During the Folding of Proteins and Nucleic Acids. *Biophys. J.* **2016**, *110*, 517a.

(72) Rebane, A. A. et al. Structure-Based Derivation of Protein Folding Intermediates and Energies from Optical Tweezers. *Biophys. J.* **2016**, *110*, 441–454.

(73) Wang, H. et al. Mechanical Unfolding and Folding of a Complex Slipknot Protein Probed by using Optical Tweezers. *Biochem.* **2019**, *58*, 4751–4760.

(74) Bustamante, C. et al. Single-Molecule Studies of Protein Folding with Optical Tweezers. *Annu. Rev. Biochem.* **2020**, *89*, 443–470.

(75) Kim, E. et al. A Single-Molecule Dissection of Ligand Binding to a Protein with Intrinsic Dynamics. *Nat. Chem. Biol.* **2013**, *9*, 313–318.

(76) Seo, M.-H. et al. Protein Conformational Dynamics Dictate the Binding Affinity for a Ligand. *Nat. Commun.* **2014**, *5*, 3724.

(77) Maus, H. et al. A Competition smFRET Assay to Study Ligand-Induced Conformational Changes of the Dengue Virus Protease. *Prot. Sci.* **2023**, *32*, e4526.

(78) Wu, S. et al. Dissecting the Protein Dynamics Coupled Ligand Binding with Kinetic Models and Single-Molecule FRET. *Biochem.* **2022**, *61*, 433–445.

(79) Gregorio, G. G. et al. Single-Molecule Analysis of Ligand Efficacy in β2AR–G-Protein Activation. *Nature* **2017**, *547*, 68–73.

(80) Maslov, I. et al. Sub-Millisecond Conformational Dynamics of the A2A Adenosine Receptor Revealed by Single-Molecule FRET. *Commun. Biol.* **2023**, *6*, 362.

(81) Chavez-Abiega, S. et al. Physical Biology of GPCR Signalling Dynamics Inferred from Fluorescence Spectroscopy and Imaging. *Curr. Opin. Struct. Biol.* **2019**, *55*, 204–211.

(82) Zhu, R. et al. Optical Tweezers in Studies of Red Blood Cells. *Cells* **2020**, *9*, 545.

(83) Avsievich, T. et al. The Advancement of Blood Cell Research by Optical Tweezers. *Rev. Phys.* **2020**, *5*, 100043.

(84) Zhong, M.-C. et al. Trapping Red Blood Cells in Living Animals using Optical Tweezers. *Nat. Commun.* **2013**, *4*, 1768.

(85) Dao, M. et al. Mechanics of the Human Red Blood Cell Deformed by Optical Tweezers. *J. Mech. Phys. Solids* **2003**, *51*, 2259–2280.

(86) Kapkov, A. A. et al. Forces of RBC Interaction with Single Endothelial Cells in Stationary Conditions: Measurements with Laser Tweezers. *J. Innov. Opt. Health Sci.* **2021**, *14*, 2142005.

(87) Paul, A. et al. Diagnosis of Malarial Infection using Change in Properties of Optically Trapped Red Blood Cells. *Biomed. J.* **2017**, *40*, 101–105.

(88) Brandao, M. et al. Optical Tweezers for Measuring Red Blood Cell Elasticity: Application to the Study of Drug Response in Sickle Cell Disease. *Eur. J. Haematol.* **2003**, *70*, 207–211.

(89) Syed, S. et al. Single-Molecule Fluorescence Reveals the Unwinding Stepping Mechanism of Replicative Helicase. *Cell Rep.* **2014**, *6*, 1037–1045.

(90) Lu, H. P. et al. Single-Molecule Conformational Dynamics of Fluctuating Noncovalent DNA-Protein Interactions in DNA Damage Recognition. *J. Am. Chem. Soc.* **2001**, *123*, 9184–9185.

(91) Lee, G. et al. Single-Molecule Analysis Reveals Three Phases of DNA Degradation by an Exonuclease. *Nat. Chem. Biol.* **2011**, *7*, 367–374.

(92) Hilario, J.; Kowalczykowski, S. C. Visualizing Protein–DNA Interactions at the Single-Molecule Level. *Curr. Opin. Chem. Biol.* **2010**, *14*, 15–22.

(93) Ivanovaitė, Š. et al. smFRET Detection of Cis and Trans DNA Interactions by the BfiI Restriction Endonuclease. *J. Phys. Chem. B* **2023**, *127*, 6470–6478.

(94) Xu, W. et al. Single-Molecule Kinetics of Nanoparticle Catalysis. *Nano Res.* **2009**, *2*, 911–922.

(95) Sambur, J. B.; Chen, P. Approaches to Single-Nanoparticle Catalysis. *Annu. Rev. Phys. Chem.* **2014**, *65*, 395–422.

(96) Sugimoto, Y. et al. Chemical Identification of Individual Surface Atoms by Atomic Force Microscopy. *Nature* **2007**, *446*, 64–67.

(97) Gross, L. et al. The Chemical Structure of a Molecule Resolved by Atomic Force Microscopy. *Science* **2009**, *325*, 1110–1114.

(98) Hobbs, J. K. et al. How Atomic Force Microscopy has Contributed to Our Understanding of Polymer Crystallization. *Polymer* **2009**, *50*, 4281–4292.

(99) Nguyen-Tri, P. et al. Recent Applications of Advanced Atomic Force Microscopy in Polymer Science: A Review. *Polymers* **2020**, *12*, 1142.

(100) Liu, C. et al. Real-Time Single-Polymer Growth Towards Single-Monomer Resolution. *Trends Chem.* **2021**, *3*, 318–331.

(101) Liu, C. et al. Single Polymer Growth Dynamics. *Science* **2017**, *358*, 352–355.

(102) Xie, J. et al. Characterization of Single Synthetic Polymers via Magnetic Tweezers. *J. Polym. Sci.* **2024**, *62*, 1260–1276.

(103) Cao, Y.; Samuels, D. C. Discrete Stochastic Simulation Methods for Chemically Reacting Systems. *Methods Enzymol.* **2009**, *454*, 115–140.

(104) Binder, K. et al. Molecular Dynamics Simulations. *J. Condens. Matter Phys.* **2004**, *16*, S429.

(105) Karplus, M.; Petsko, G. A. Molecular Dynamics Simulations in Biology. *Nature* **1990**, *347*, 631–639.

(106) Gillespie, D. T. A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. *J. Comput. Phys.* **1976**, *22*, 403–434.

(107) Gillespie, D. T. Exact Stochastic Simulation of Coupled Chemical Reactions. *J. Phys. Chem.* **1977**, *81*, 2340–2361.

(108) El Samad, H. et al. Stochastic Modelling of Gene Regulatory Networks. *Int. J. Robust Nonlinear Control.* **2005**, *15*, 691–711.

(109) Murrugarra, D. et al. Modeling Stochasticity and Variability in Gene Regulatory Networks. *EURASIP J. Bioinform. Syst. Biol.* **2012**, *2012*, 1–11.

(110) Inoue, M.; Kaneko, K. Entangled Gene Regulatory Networks with Cooperative Expression Endow Robust Adaptive Responses to Unforeseen Environmental Changes. *Phys. Rev. Res.* **2021**, *3*, 033183.

(111) Zhang, L. et al. Kinetics and Mechanisms of Charge Transfer Processes in Photocatalytic Systems: A Review. *J. Photochem. Photobiol. C: Photochem. Rev.* **2012**, *13*, 263–276.

(112) Ribeiro, A. et al. A General Modeling Strategy for Gene Regulatory Networks with Stochastic Dynamics. *J. Comput. Biol.* **2006**, *13*, 1630–1639.

(113) Kierzek, A. M. STOCKS: STOChastic Kinetic Simulations of Biochemical Systems with Gillespie Algorithm. *Bioinformatics* **2002**, *18*, 470–481.

(114) Kumar, A. et al. Emergence of Dynamic Cooperativity in the Stochastic Kinetics of Fluctuating Enzymes. *J. Chem. Phys.* **2016**, *145*.

(115) Kang, H.-W. et al. Quasi-Steady-State Approximations Derived from the Stochastic Model of Enzyme Kinetics. *Bull. Math. Biol.* **2019**, *81*, 1303–1336.

(116) Tonn, M. K. et al. Stochastic Modelling Reveals Mechanisms of Metabolic Heterogeneity. *Commun. Biol.* **2019**, *2*, 108.

(117) Andersen, M. et al. A Practical Guide to Surface Kinetic Monte Carlo Simulations. *Front. Chem.* **2019**, *7*, 202.

(118) Dybeck, E. C. et al. Generalized Temporal Acceleration Scheme for Kinetic Monte Carlo Simulations of Surface Catalytic Processes by Scaling the Rates of Fast Reactions. *J. Chem. Theory Comput.* **2017**, *13*, 1525–1538.

(119) Jørgensen, M.; Gronbeck, H. Scaling Relations and Kinetic Monte Carlo Simulations to Bridge the Materials Gap in Heterogeneous Catalysis. *ACS Catal.* **2017**, *7*, 5054–5061.

(120) Xu, D. et al. Kinetic Monte Carlo Simulations of GaN Homoepitaxy on c-and m-Plane Surfaces. *J. Chem. Phys.* **2017**, *146*.

(121) Xu, L. et al. Kinetic Monte Carlo Simulations of Pd Deposition and Island Growth on MgO (1 0 0). *Surf. Sci.* **2007**, *601*, 3133–3142.

(122) Iyer-Biswas, S.; Zilman, A. First-Passage Processes in Cellular Biology. *Adv. Chem. Phys.* **2016**, *160*, 261–306.

(123) Redner, S., *A Guide to First-Passage Processes*; Cambridge university press: 2001.

(124) Chou, T.; D'Orsogna, M. R. In *First-Passage Phenomena and their Applications*; World Scientific: 2014, pp 306–345.

(125) Szabo, A. et al. First Passage Time Approach to Diffusion Controlled Reactions. *J. Chem. Phys.* **1980**, *72*, 4350–4357.

(126) Li, X.; Kolomeisky, A. B. Mechanisms and Topology Determination of Complex Chemical and Biological Network Systems from First-Passage Theoretical Approach. *J. Chem. Phys.* **2013**, *139*.

(127) Iwahara, J.; Kolomeisky, A. B. Discrete-State Stochastic Kinetic Models for Target DNA Search by Proteins: Theory and Experimental Applications. *Biophys. Chem.* **2021**, *269*, 106521.

(128) Coppey, M. et al. Kinetics of Target Site Localization of a Protein on DNA: A Stochastic Approach. *Biophys. J.* **2004**, *87*, 1640–1649.

(129) Zhang, Y.; Dudko, O. K. First-Passage Processes in the Genome. *Annu. Rev. Biophys.* **2016**, *45*, 117–134.

(130) Kolomeisky, A. B. Motor Proteins and Molecular Motors: How to Operate Machines at the Nanoscale. *J. Condens. Matter Phys.* **2013**, *25*, 463101.

(131) Mogre, S. S. et al. Getting Around the Cell: Physical Transport in the Intracellular World. *Phys. Biol.* **2020**, *17*, 061003.

(132) Mlynarczyk, P. J.; Abel, S. M. First Passage of Molecular Motors on Networks of Cytoskeletal Filaments. *Phys. Rev. E* **2019**, *99*, 022406.

(133) Ye, R. et al. Analogy between Enzyme and Nanoparticle Catalysis: A Single-Molecule Perspective. *ACS Catal.* **2019**, *9*, 1985–1992.

(134) Das, A.; Chaudhury, S. Modeling the Heterogeneous Catalytic Activity of a Single Nanoparticle Using a First Passage Time Distribution Formalism. *Chem. Phys. Lett.* **2015**, *641*, 193–198.

(135) Jangid, P. et al. Theoretical Understanding of Dynamic Catalysis. *J. Phys. Chem. C* **2024**.

(136) Chaudhury, S. et al. Dynamics of Chemical Reactions on Single Nanocatalysts with Heterogeneous Active Sites. *J. Chem. Phys.* **2023**, *158*, 074101.

(137) Sakref, Y. et al. On Kinetic Constraints that Catalysis Imposes on Elementary Processes. *J. Phys. Chem. B* **2023**, *127*, 10950–10959.

(138) Ghusinga, K. R. et al. A Mechanistic Stochastic Framework for Regulating Bacterial Cell Division. *Sci. Rep.* **2016**, *6*, 30229.

(139) Teimouri, H. et al. Stochastic Mechanisms of Cell-Size Regulation in Bacteria. *J. Phys. Chem. Lett.* **2020**, *11*, 8777–8782.

(140) Chung, H. S.; Gopich, I. V. Fast Single-Molecule FRET Spectroscopy: Theory and Experiment. *Phys. Chem. Chem. Phys.* **2014**, *16*, 18644–18657.

(141) Chung, H. S. et al. Experimental Determination of Upper Bound for Transition Path Times in Protein Folding from Single-Molecule Photon-by-Photon Trajectories. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 11837–11844.

(142) De Sancho, D. et al. Folding Kinetics and Unfolded State Dynamics of the GB1 Hairpin from Molecular Simulation. *J. Chem. Theory Comput.* **2013**, *9*, 1743–1753.

(143) Best, R. B.; Mittal, J. Microscopic Events in β-Hairpin Folding from Alternative Unfolded Ensembles. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 11087–11092.

(144) Singh, D.; Chaudhury, S. A Single-Molecule Stochastic Theory of Protein-Ligand Binding in the Presence of Multiple Unfolding/Folding and Ligand Binding Pathways. *Biophys. Chem.* **2022**, *285*, 106803.

(145) Xu, W. et al. Single-Molecule Nanocatalysis Reveals Heterogeneous Reaction Pathways and Catalytic Dynamics. *Nat. Mater.* **2008**, *7*, 992–996.

(146) Kou, S. et al. Single-Molecule Michaelis–Menten Equations. *J. Phys. Chem. B* **2005**, *109*, 19068–19081.

(147) Xu, W. et al. Single-Molecule Kinetic Theory of Heterogeneous and Enzyme Catalysis. *J. Phys. Chem. C* **2009**, *113*, 2393–2404.

(148) Moffitt, J. R.; Bustamante, C. Extracting Signal from Noise: Kinetic Mechanisms from a Michaelis-Menten-Like Expression for Enzymatic Fluctuations. *FEBS J.* **2013**, *281*, 498–517.

(149) Floyd, D. L. et al. Analysis of Kinetic Intermediates in Single-Particle Dwell-Time Distributions. *Biophys. J.* **2010**, *99*, 360–366.

(150) Barato, A. C.; Seifert, U. Universal Bound on the Fano Factor in Enzyme Kinetics. *J. Phys. Chem. B* **2015**, *119*, 6555–6561.

(151) Klindziuk, A.; Kolomeisky, A. B. Theoretical Investigation of Transcriptional Bursting: a Multistate Approach. *J. Phys. Chem. B* **2018**, *122*, 11969–11977.

(152) Klindziuk, A.; Kolomeisky, A. B. Understanding the Molecular Mechanisms of Transcriptional Bursting. *Phys. Chem. Chem. Phys.* **2021**, *23*, 21399–21406.

(153) Chaudhury, S. Poisson Indicator and Fano Factor for Probing Dynamic Disorder in Single-Molecule Enzyme Inhibition Kinetics. *J. Phys. Chem. B* **2014**, *118*, 10405–10412.

(154) Hakem, R. DNA-Damage Repair; the Good, the Bad, and the Ugly. *EMBO J.* **2008**, *27*, 589–605.

(155) Serebreni, L.; Stark, A. Insights into Gene Regulation: From Regulatory Genomic Elements to DNA-Protein and Protein-Protein Interactions. *Curr. Opin. Cell Biol.* **2021**, *70*, 58–66.

(156) Riggs, A. D. et al. The Lac Represser-Operator Interaction: III. Kinetic Studies. *J. Mol. Biol.* **1970**, *53*, 401–417.

(157) Von Hippel, P. H.; Berg, O. G. Facilitated Target Location in Biological Systems. *J. Biol. Chem.* **1989**, *264*, 675–678.

(158) Elf, J. et al. Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell. *Science* **2007**, *316*, 1191–1194.

(159) Cherstvy, A. et al. Protein-DNA Interactions: Reaching and Recognizing the Targets. *J. Phys. Chem. B* **2008**, *112*, 4741–4750.

(160) De la Rosa, M. A. D. et al. Dynamic Strategies for Target-Site Search by DNA-Binding Proteins. *Biophys. J.* **2010**, *98*, 2943–2953.

(161) Hammar, P. et al. The Lac Repressor Displays Facilitated Diffusion in Living Cells. *Science* **2012**, *336*, 1595–1598.

(162) Izeddin, I. et al. Single-Molecule Tracking in Live Cells Reveals Distinct Target-Search Strategies of Transcription Factors in the Nucleus. *Elife* **2014**, *3*, e02230.

(163) Berg, O. G. et al. Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 1. Models and Theory. *Biochem.* **1981**, *20*, 6929–6948.

(164) Winter, R. B.; Von Hippel, P. H. Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 2. The Escherichia Coli Lac Repressor-Operator Interaction: Equilibrium Measurements. *Biochem.* **1981**, *20*, 6948–6960.

(165) Winter, R. B. et al. Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 3. The Escherichia Coli Lac Repressor-Operator Interaction: Kinetic Measurements and Conclusions. *Biochem.* **1981**, *20*, 6961–6977.

(166) Gowers, D. M. et al. Measurement of the Contributions of 1D and 3D Pathways to the Translocation of a Protein Along DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15883–15888.

(167) Wang, Y. et al. Single Molecule Measurements of Repressor Protein 1D Diffusion on DNA. *Phys. Rev. Lett.* **2006**, *97*, 048302.

(168) Tafvizi, A. et al. Dancing on DNA: Kinetic Aspects of Search Processes on DNA. *ChemPhysChem* **2011**, *12*, 1481–1489.

(169) Normanno, D. et al. Probing the Target Search of DNA-Binding Proteins in Mammalian Cells using TetR as Model Searcher. *Nat. Commun.* **2015**, *6*, 7357.

(170) Mirny, L. et al. How a Protein Searches for its Site on DNA: The Mechanism of Facilitated Diffusion. *J. Phys. A: Math. Theor.* **2009**, *42*, 434013.

(171) Kolomeisky, A. B. Physics of Protein–DNA Interactions: Mechanisms of Facilitated Target Search. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2088–2095.

(172) Halford, S. E.; Marko, J. F. How do Site-Specific DNA-Binding Proteins Find their Targets? *Nucleic Acids Res.* **2004**, *32*, 3040–3052.

(173) Veksler, A.; Kolomeisky, A. B. Speed-Selectivity Paradox in the Protein Search for Targets on DNA: Is it Real or Not? *J. Phys. Chem. B* **2013**, *117*, 12695–12701.

(174) Lange, M. et al. Dynamics of the Protein Search for Targets on DNA in the Presence of Traps. *J. Phys. Chem. B* **2015**, *119*, 12410–12416.

(175) Shvets, A. et al. Role of Static and Dynamic Obstacles in the Protein Search for Targets on DNA. *J. Phys. Chem. B* **2016**, *120*, 5802–5809.

(176) Felipe, C. et al. DNA Looping and DNA Conformational Fluctuations can Accelerate Protein Target Search. *J. Phys. Chem. B* **2021**, *125*, 1727–1734.

(177) Mondal, K.; Chaudhury, S. Effect of DNA Conformation on the Protein Search for Targets on DNA: A Theoretical Perspective. *J. Phys. Chem. B* **2020**, *124*, 3518–3526.

(178) Shvets, A. A.; Kolomeisky, A. B. Sequence Heterogeneity Accelerates Protein Search for Targets on DNA. *J. Chem. Phys.* **2015**, *143*.

(179) Zimmerman, S. B.; Trach, S. O. Macromolecular Crowding Extends the Range of Conditions Under Which DNA Polymerase is Functional. *Biochim. Biophys. Acta, Gene Struct. Expression* **1988**, *949*, 297–304.

(180) Zimmerman, S. B.; Trach, S. O. Estimation of Macromolecule Concentrations and Excluded Volume Effects for the Cytoplasm of Escherichia Coli. *J. Mol. Biol.* **1991**, *222*, 599–620.

(181) Ellis, R. J. Macromolecular Crowding: Obvious but Underappreciated. *Trends Biochem. Sci.* **2001**, *26*, 597–604.

(182) Sasaki, Y. et al. Regulation of DNA Nucleases by Molecular Crowding. *Nucleic Acids Res.* **2007**, *35*, 4086–4093.

(183) Esadze, A. et al. Measurement of Nanoscale DNA Translocation by Uracil DNA Glycosylase in Human Cells. *Nucleic Acids Res.* **2017**, *45*, 12413–12424.

(184) Brackley, C. et al. Intracellular Facilitated Diffusion: Searchers, Crowders, and Blockers. *Phys. Rev. Lett.* **2013**, *111*, 108101.

(185) Krepel, D. et al. Mechanism of Facilitated Diffusion During a DNA Search in Crowded Environments. *J. Phys. Chem. B* **2016**, *120*, 11113–11122.

(186) Krepel, D.; Levy, Y. Protein Diffusion Along DNA: On the Effect of Roadblocks and Crowders. *J. Phys. A: Math. Theor.* **2016**, *49*, 494003.

(187) Dey, P.; Bhattacherjee, A. Role of Macromolecular Crowding on the Intracellular Diffusion of DNA Binding Proteins. *Sci. Rep.* **2018**, *8*, 844.

(188) Dey, P.; Bhattacherjee, A. Disparity in Anomalous Diffusion of Proteins Searching for Their Target DNA Sites in a Crowded Medium is Controlled by the Size, Shape and Mobility of Macromolecular Crowders. *Soft Matter* **2019**, *15*, 1960–1969.

(189) Dey, P.; Bhattacherjee, A. Mechanism of Facilitated Diffusion of DNA Repair Proteins in Crowded Environment: Case Study with Human Uracil DNA Glycosylase. *J. Phys. Chem. B* **2019**, *123*, 10354–10364.

(190) Dey, P.; Bhattacherjee, A. Structural Basis of Enhanced Facilitated Diffusion of DNA-Binding Protein in Crowded Cellular Milieu. *Biophys. J.* **2020**, *118*, 505–517.

(191) Shvets, A. A.; Kolomeisky, A. B. The Role of DNA Looping in the Search for Specific Targets on DNA by Multisite Proteins. *J. Phys. Chem. Lett.* **2016**, *7*, 5022–5027.

(192) Felipe, C. et al. The Effect of Obstacles in Multi-Site Protein Target Search with DNA Looping. *J. Chem. Phys.* **2020**, *152*.

(193) Muthukumar, M. Mechanism of DNA Transport Through Pores. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 435–450.

(194) Kasianowicz, J. J. et al. Characterization of Individual Polynucleotide Molecules using a Membrane Channel. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13770–13773.

(195) Deamer, D. W.; Branton, D. Characterization of Nucleic Acids by Nanopore Analysis. *Acc. Chem. Res.* **2002**, *35*, 817–825.

(196) Branton, D. et al. The Potential and Challenges of Nanopore Sequencing. *Nat. Biotechnol.* **2008**, *26*, 1146–1153.

(197) Mayer, S. F. et al. Biological Nanopores for Single-Molecule Sensing. *iScience* **2022**, *25*.

(198) Smeets, R. M. et al. Salt Dependence of Ion Transport and DNA Translocation through Solid-State Nanopores. *Nano Lett.* **2006**, *6*, 89–95.

(199) Ayub, M.; Bayley, H. Engineered Transmembrane Pores. *Curr. Opin. Chem. Biol.* **2016**, *34*, 117–126.

(200) Shi, W. et al. Nanopore Sensing. *Anal. Chem.* **2017**, *89*, 157–188.

(201) He, Y. et al. DNA Capture in Nanopores for Genome Sequencing: Challenges and Opportunities. *J. Mater. Chem.* **2012**, *22*, 13423–13427.

(202) He, Y. et al. Gate Manipulation of DNA Capture into Nanopores. *ACS Nano* **2011**, *5*, 8391–8397.

(203) Trepagnier, E. H. et al. Controlling DNA Capture and Propagation through Artificial Nanopores. *Nano Lett.* **2007**, *7*, 2824–2830.

(204) Leong, I. W. et al. Salt Gradient Control of Translocation Dynamics in a Solid-State Nanopore. *Anal. Chem.* **2021**, *93*, 16700–16708.

(205) Sohi, A. N. et al. DNA Capture by Nanopore Sensors under Flow. *Anal. Chem.* **2020**, *92*, 8108–8116.

(206) Sung, W.; Park, P. Polymer Translocation through a Pore in a Membrane. *Phys. Rev. Lett.* **1996**, *77*, 783.

(207) Muthukumar, M. Polymer Translocation through a Hole. *J. Chem. Phys.* **1999**, *111*, 10371–10374.

(208) Chaudhury, S.; Cherayil, B. J. A Model of Anomalous Chain Translocation Dynamics. *J. Phys. Chem. B* **2008**, *112*, 15973–15979.

(209) Dubbeldam, J. et al. Driven Polymer Translocation through a Nanopore: A Manifestation of Anomalous Diffusion. *Europhys. Lett.* **2007**, *79*, 18002.

(210) Dubbeldam, J. et al. Driven Translocation of a Polymer: Fluctuations at Work. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **2013**, *87*, 032147.

(211) Bezrukov, S. M. et al. Dynamics and Free Energy of Polymers Partitioning into a Nanoscale Pore. *Macromolecules* **1996**, *29*, 8517–8522.

(212) Movileanu, L. et al. Partitioning of Individual Flexible Polymers into a Nanoscopic Protein Pore. *Biophys. J.* **2003**, *85*, 897–910.

(213) Aksoyoglu, M. A. et al. Size-Dependent Forced PEG Partitioning into Channels: VDAC, OmpC, and α-Hemolysin. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 9003–9008.

(214) Gurnev, P. A. et al. Poly(ethylene glycol)s in Semidilute Regime: Radius of Gyration in the Bulk and Partitioning into a Nanopore. *Macromolecules* **2017**, *50*, 2477–2483.

(215) Freedman, K. J. et al. Nonequilibrium Capture Rates Induce Protein Accumulation and Enhanced Adsorption to Solid-State Nanopores. *ACS Nano* **2014**, *8*, 12238–12249.

(216) Larimi, M. G. et al. Interactions of a Polypeptide with a Protein Nanopore under Crowding Conditions. *ACS Nano* **2019**, *13*, 4469–4477.

(217) Yao, F. et al. Crowding-Induced DNA Translocation through a Protein Nanopore. *Anal. Chem.* **2020**, *92*, 3827–3833.

(218) Chau, C. C. et al. Macromolecular Crowding Enhances the Detection of DNA and Proteins by a Solid-State Nanopore. *Nano Lett.* **2020**, *20*, 5553–5561.

(219) Marcuccio, F. et al. Mechanistic Study of the Conductance and Enhanced Single-Molecule Detection in a Polymer–Electrolyte Nanopore. *ACS Nanosci. Au* **2023**, *3*, 172–181.

(220) Chen, Y.; Luo, K. Dynamics of Polymer Translocation through a Nanopore Induced by Different Sizes of Crowding Agents. *J. Chem. Phys.* **2013**, *138*.

(221) Somorjai, G. A.; Li, Y., *Introduction to Surface Chemistry and Catalysis*; John Wiley & Sons: New York, 2010.

(222) Ross, J. R., *Heterogeneous Catalysis: Fundamentals and Applications*; Elsevier: Amsterdam: The Netherlands, 2011.

(223) Friend, C. M.; Xu, B. Heterogeneous Catalysis: A Central Science for a Sustainable Future. *Acc. Chem. Res.* **2017**, *50*, 517–521.

(224) Zhang, Z. et al. Ensembles of Metastable States Govern Heterogeneous Catalysis on Dynamic Interfaces. *Acc. Chem. Res.* **2020**, *53*, 447–458.

(225) Chen, P. et al. Spatiotemporal Catalytic Dynamics within Single Nanocatalysts Revealed by Single-Molecule Microscopy. *Chem. Soc. Rev.* **2014**, *43*, 1107–1117.

(226) Makarov, D. E., *Single Molecule Science: Physical Principles and Models*; CRC Press: 2015.

(227) Zhou, X. et al. Size-Dependent Catalytic Activity and Dynamics of Gold Nanoparticles at the Single-Molecule Level. *J. Am. Chem. Soc.* **2010**, *132*, 138–146.

(228) Chen, T. et al. Single-Molecule Nanocatalysis Reveals Facet-Dependent Catalytic Kinetics and Dynamics of Pallidium Nanoparticles. *ACS Catal.* **2017**, *7*, 2967–2972.

(229) Shen, H. et al. Single-Molecule Kinetics Reveals a Hidden Surface Reaction Intermediate in Single-Nanoparticle Catalysis. *J. Phys. Chem. C* **2014**, *118*, 26902–26911.

(230) Singh, D.; Chaudhury, S. A Stochastic Theoretical Approach to Study the Size-Dependent Catalytic Activity of a Metal Nanoparticle at the Single Molecule Level. *Phys. Chem. Chem. Phys.* **2017**, *19*, 8889–8895.

(231) Chaudhury, S. et al. Theoretical Investigations of the Dynamics of Chemical Reactions on Nanocatalysts with Multiple Active Sites. *J. Phys. Chem. Lett.* **2020**, *11*, 2330–2335.

(232) Zou, N. et al. Cooperative Communication within and between Single Nanocatalysts. *Nat. Chem.* **2018**, *10*, 607–614.

Chapter 2

Theoretical Study of the Role of Dynamic Bulk Crowders in the Protein Search for Targets on DNA

The content of this chapter is reprinted (adapted) with permission from (*J. Stat. Mech.*, 2021, 073502). Copyright © 2021 IOP Publishing. All rights reserved.

2.1 Introduction

A protein-DNA interaction is very important and plays a crucial role in many biological processes, such as DNA transcription, DNA unwinding, gene expression, and regulation[1–3]. In vitro experimental studies by Riggs et al have shown that the association rate between the *lac* repressor and its target is about two orders of magnitude faster than the rate for the three-dimensional (3D) diffusion limit[4]. Based on this interesting finding, it was proposed that when the protein molecule is bound to a non-specific site on the DNA, it slides to the target through the DNA molecule by performing a one-dimensional (1D) random walk. Thus the protein undergoes 3D diffusion along with 1D random walk during this search process and this combined mechanism is called facilitated diffusion[5, 6]. This phenomenon has been supported experimentally by bulk measurements[7, 8] and recent single-molecule experiments[9–11] both in vitro and in vivo. In most of the theoretical models on facilitated diffusion of target search dynamics, one usually considers an oversimplified picture where there is unobstructed access to all sites on DNA and the protein finds its target site on the DNA through a combination of 1D and 3D diffusion.

In living cells, DNA is heavily crowded and macromolecules occupy 10%-40% of the total cellular volume. Hence, they are expected to have a significant effect on the search dynamics in various ways. For example, macromolecular crowders can act as obstacles and they can cover the DNA chain and can block the 1D dynamics of the searching protein along the DNA (in vivo crowding)[12–16]. This situation has been studied theoretically using a discrete state theoretical model by Kolomeisky and co-workers where it has been shown that the crowding on DNA can accelerate or slow down the search dynamics depending on the size and spatial position of the obstacles, the spatial positions of the target, and whether the obstacles are static or dynamic[17, 18]. In the case of dynamic crowding, the protein search dynamics depend on the mobility of crowding agents whereas computational studies indicated that under certain conditions such blockers can lead to faster search dynamics[15, 16]. There is also another type of crowding where the crowders are found in the solution and this is known as bulk crowding. In living cells, the DNA is densely covered by proteins, nucleic acids, and other cellular organelles that can make some regions of the DNA partially inaccessible to other DNA-binding proteins. This crowded intracellular environment is very complex and is expected to have a significant effect on the target search process[19–22]. It is anticipated that such bulk crowders will strongly bias the search toward 1D diffusion. In some studies, it has been shown that the

target search time remains unchanged in the presence of such crowders[13] whereas some recent experimental studies reported that such crowding agents could promote 'facilitated diffusion' in human DNA glycolysases[23]. Extensive 2D/3D Langevin dynamics simulations have successfully captured important insights into the protein target search in a crowded cellular medium[24]. Using coarse-grained molecular dynamics and Monte Carlo simulations it was shown that crowders can promote facilitated diffusion and accelerate search kinetics and the target search dynamics depending on the crowding occupancy and the properties of the crowding particles and the salt concentration[25]. Extensive MD simulation studies have shown that the target search dynamics of a protein depend on the properties of the crowding particles such as the size, shape, and mobility of the crowding molecules adjacent to DNA and far from DNA[26]. The lack of consistency and universality in all these studies reflects that the underlying molecular mechanism governing the effect of bulk crowding and its mobility on protein-DNA interaction is complex and unclear[22]. To provide a deeper understanding of the role of facilitated diffusion in target search dynamics, Mondal and Chaudhury have developed a theoretical method that explores the effect of bulk crowding during the protein search[27]. It was found that when the crowders block the target site, the protein dynamics is significantly modified and the search time increases. But this theory was restricted only to static crowders in solution and one needs to concentrate on a more realistic situation in the cellular environment where the crowders are dynamic. Also, in simulations for recognition of a target site by the protein under crowded conditions, the crowding particles are considered to be mobile[13, 25, 26].

In this chapter, a theoretical method is devised to analyze the effect of mobile bulk crowders on the 3D and 1D target search dynamics of DNA binding proteins. The role of crowding molecules in the bulk on the protein dynamics adjacent to DNA is explored through an analytical description of the search process in the presence of dynamic crowders by using a discrete-state stochastic framework. This particular discrete state model has been used successfully in the context of protein search dynamics and can account for important chemical and physical processes[17, 18, 27–31]. The analysis begins with the case of dynamic bulk crowding that can block a single binding site and is then extended to crowders that block the protein to bind to multiple sites on the DNA chain.

2.2 Theoretical Model

A single DNA molecule is represented as a chain consisting of L binding sites where one of the sites located at the position m serves as a target site. The protein always starts from the solution labeled as state 0. It is assumed that the DNA chain is coiled in the solution, and the protein diffuses very fast in the volume around the DNA. The searching protein can bind to any vacant site on DNA with equal probability with a binding rate k_{on} per each site. The protein bound to any site can then slide along the DNA chain with a diffusion rate u and can also dissociate from DNA into the solution with a rate k_{off}. For realistic values of DNA lengths ($L = 10^6 - 10^9$), the search process is independent of the target position as long as the target is not at the end sites of the DNA chain. In this theoretical scheme, the target is fixed at the middle of the DNA chain as from symmetry arguments this leads to the shortest search time[28]. Bulk crowding in the cellular medium is quite large (up to 30% - 40% by volume). To analyze the role of bulk crowders, it is assumed that the crowder can block one or more sites on the DNA and will prevent the protein to bind to that site/sites or dissociate from that site/s into the solution. Thus, crowders will significantly slow down the coming of protein from the bulk. The protein will have unhindered access to only those binding sites that are not blocked by the bulk crowders. Once the protein is already bound to DNA, it can also slide to the region that is blocked from outside by the crowder. Thus, the bulk crowding prevents the association and dissociation of the protein from the bulk but once bound to any site on the DNA, the protein can slide through the crowded region as well via 1D motion (figure 2.1).

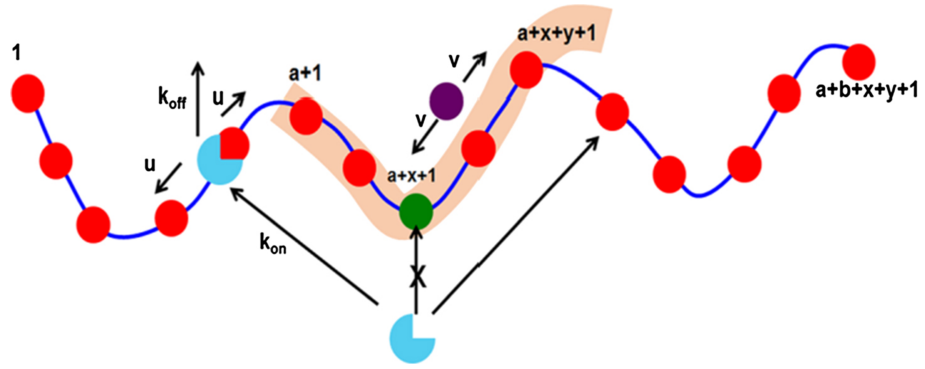


Figure 2.1: Schematic to depict protein target search on DNA with a mobile crowder in the bulk. The DNA is of length $L = a + x + y + b + 1$, with the target in the middle at $m = a + x + 1$. The number of sites blocked on the left and right side of the target are x and y, respectively. Similarly, there are a and b unblocked sites on the left and right side of the DNA respectively. Effectively, the crowder blocks $\Delta = x + y + 1$ sites of the DNA and it can diffuse in the bulk adjacent to the DNA with rate v.

Since in in vivo conditions, the cellular matrix contains a large number of species that can act as crowding molecules, the present analysis is extended to multiple dynamic crowders in the bulk. The crowder is mobile and assumed to diffuse in the bulk adjacent to the DNA chain with a rate v. When the size of the crowder is such that it blocks only a single site on the DNA, the probability of blocking the protein to bind to the target site is very less. The total search process can be viewed as a combination of the search happening when the protein experiences no effect from the bulk crowding and the search being affected as a result of the crowding. So, the total search time is a combination of the average search time for the system without a crowder and the blocking time in the presence of a crowder. This can be approximated as

$$\langle T \rangle \cong \langle T_0 \rangle + \langle T_{bl} \rangle. \tag{2.1}$$

Here, $\langle T_0 \rangle$ is the search time when the DNA chain is free from crowding. The analytical expression of the search time when the DNA chain is free from crowding has been obtained earlier[28] and is given as,

$$\langle T_0 \rangle = \frac{1}{k_{on}} \frac{1}{S_0(s=0)} + \frac{1}{k_{off}} \left(\frac{L}{S_0(s=0)} - 1 \right) \tag{2.2}$$

where the function $S_0(s=0)$ relates to the different sites that a protein molecule scans on an average while searching for a specific site and has the form

$$S_0(s) = \frac{z(1+z)(z^{-L} - z^{L})}{(1-z)(z^{1-m} + z^{m})(z^{m-L} + z^{L+1-m})} \tag{2.3}$$

with $z = \frac{s+2u+k_{off} - \sqrt{(s+2u+k_{off})^2 - 4u^2}}{2u}$.

The above expression for $\langle T_0 \rangle$ is valid as long as the target is not situated at the end of the DNA chain. The second term on the right-hand side of Eq. (2.1), $\langle T_{bl} \rangle$ is the delay time which the protein experiences to recognize and find the target due to the bulk crowding and it has the form $\langle T_{bl} \rangle = P_m T_m$ where P_m is the probability of blocking a binding site and T_m is the time for which the site is blocked. The probability that any site is blocked during a search process is Δ/L where Δ is the length of the blocked region. For the case when the crowder size is such that only one site on the chain is blocked, $\Delta = 1$. The crowder diffuses from one site to the other in a time $1/2v$ where the factor of 2 signifies

that the crowder can diffuse to either of the sides available. Since the probability to block any site during a search process is $1/L$, the expression for the average search time when a single site is blocked by a dynamic crowder is

$$\langle T \rangle \cong \langle T_0 \rangle + \langle T_{bl} \rangle \cong \langle T_0 \rangle + \frac{1}{2vL} \approx \langle T_0 \rangle. \tag{2.4}$$

For long chain lengths, the contribution of the second term on the right-hand side of equation (2.4) is negligible and the search time in the presence of a single blocking site (which may be the target site) remains the same as that of the case when there are no crowders. As shown in figure 2.3, for this case, the search process is also independent of the velocity of the crowder particle and the search time remains unaltered and is equal to the case when no crowders are present. This is expected because for a long DNA chain with many binding sites, the crowding particle cannot always block the protein from finding the target site. This is in contrast to static crowding where the search time is affected even when one site on the DNA is blocked provided that site is the target site.

Next, let us extend the results to dynamic bulk crowders of any size that can take into account more realistic situations. In this case $\langle T_{bl} \rangle$ in equation (2.1) will have the contributions from all the times where the mobile crowders in the bulk will block different sites on the DNA chain at a time—the non-specific sites or/and the target site. At any instant of time, when some binding site is blocked by these bulk crowders, the search process continues while the crowder is diffusing away from that particular binding site.

In this situation, two independent search processes are going on: when the crowder is diffusing away from any binding site, t_1, and when the crowder is blocking the binding site, t_2 (equivalent to static crowding). However, as discussed before in the previous study on static bulk crowding, a significant delay in the search process will occur when the crowders block the target site[27]. Thus $\langle T_{bl} \rangle$ approximated as $\langle T_{bl} \rangle \approx P_m T_m$ is calculated by considering the case that one of the blocked sites is the target site. As a consequence, P_m will determine the probability of blocking the target site, and the time T_m is the additional time taken by the protein to find the target site due to the crowding. For a given length of the blocked region Δ, there are several ways of blocking the target site which are presented in Table 2.1.

Due to the symmetry, some of the configurations in Table 2.1 are equally probable. $P_1 = P_\Delta, P_2 = P_{\Delta-1}, ..., P_{\Delta/2} = P_{(\Delta/2)+1}$. Some of these configurations are shown in

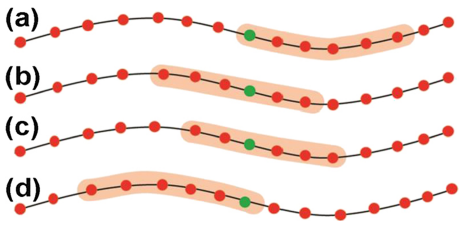


Figure 2.2: A schematic view of different configurations of bulk crowders of size λ blocking the target site while diffusing over the DNA chain of length L with the target site (in green) in the middle of the DNA chain. The parameters x and y for each configuration are (a) $x = 0$ and $y = \lambda - 1$ (b) $x = \lambda/2$ and $y = (\lambda/2) - 1$ (c) $x = (\lambda/2) - 1$ and $y = \lambda/2$ and (d) $x = \lambda - 1$ and $y = 0$.

figure 2.2.

S. No.	**Number of sites blocked on the left of the target site (x)**	**Number of sites blocked on the right of the target site (y)**
1 (Fig. 2.2a)	0	$\Delta - 1$
2	1	$\Delta - 2$
3	2	$\Delta - 3$
–	–	–
$\Delta/2$ (Fig. 2.2c)	$\Delta/2 - 1$	$\Delta/2$
$\Delta/2 + 1$ (Fig. 2.2b)	$\Delta/2$	$\Delta/2 - 1$
–	–	–
$\Delta - 1$	$\Delta - 2$	1
Δ (Fig. 2.2d)	$\Delta - 1$	0

Table 2.1: A tabulated view of different configurations of the chain in the presence of dynamic crowders restricting the protein to bind/unbind to different sites on DNA including the target site. The length of the crowder $\Delta = x + y + 1$. For a given value of x, the parameters $a = m - x - 1$, $y = \Delta - (x + 1)$ and $b = L - m - y$.

The average time required by the crowder to diffuse away from any binding site in a particular configuration, t_1 would be the product of the number of sites it has to cover and the time taken to diffuse over those sites to unblock a required binding site. The time taken to diffuse over a single binding site is $1/2v$. So, the average time taken by the crowders to make any site free from blocking such that the protein molecule can bind to it directly would be the average of diffusing time from all the configurations.

$$\begin{aligned} t_1 &= \frac{1}{2v\Delta}\left[\Delta + 2(\Delta - 1) + 3(\Delta - 2) + ... + 2(\Delta - 1) + \Delta\right] \\ &= \frac{1}{v\Delta}\sum_{n=0}^{n=\frac{\Delta}{2}-1}(n+1)(\Delta - n) = \frac{(1+\Delta)(2+\Delta)}{12v}. \end{aligned} \tag{2.5}$$

The expression for t_2 will be the average of the search times for all the configurations considering the blocking is static. In a previous study on static bulk crowding, it was

shown that symmetrical crowding about the target side has a stronger effect on the target search process than asymmetric crowding[27]. So, among all the possible configurations given in table 2.1, the greatest contribution to t_2 will be for the case when the crowders are diffusing over the whole DNA chain with the target site in the center of the blocked region. As a result, one can approximate t_2 as

$$t_2 = \frac{1}{\Delta}\sum_{x=0}^{x=\Delta-1}\langle T_x\rangle \approx \langle T_x\rangle \begin{cases} x = \frac{\Delta}{2} & \text{when } \Delta = \text{even} \\ x = \frac{\Delta-1}{2} & \text{when } \Delta = \text{odd.} \end{cases} \tag{2.6}$$

The analytical expression for $\langle T_x\rangle$ has been obtained earlier (Eq. (4) in[27]) where x is the number of non-specific sites blocked from one side of the target

$$\begin{aligned}\langle T_x\rangle = {} & \frac{1}{k_{on}}\left[\frac{1}{K(a,x,0)+K(b,y,0)}\right] + \frac{1}{k_{off}}\left[\frac{L-\Delta}{K(a,x,0)+K(b,y,0)} - 1\right] \\ & + \frac{x(1+x)}{2u}\left[\frac{K(a,x,0)}{K(a,x,0)+K(b,y,0)}\right] + \frac{y(1+y)}{2u}\left[\frac{K(b,y,0)}{K(a,x,0)+K(b,y,0)}\right].\end{aligned} \tag{2.7}$$

In the above expression,

$$\begin{aligned} K(a,x,s) &= \frac{z^{1-a}(z^{2a}-1)}{z-1} \times \frac{1}{(z^{1+a}+z^{-a})(1+x) - x(z^a+z^{1-a})} \\ K(b,y,s) &= \frac{z^{1-b}(z^{2b}-1)}{z-1} \times \frac{1}{(z^{1+b}+z^{-b})(1+y) - y(z^b+z^{1-b})} \end{aligned}$$

where $z = \frac{s+2u+k_{off}-\sqrt{(s+2u+k_{off})^2-4u^2}}{2u}$. The physical meaning of each term in the above expression for $\langle T_x\rangle$ has also been discussed in detail in[27]. Thus, the delay in the search process due to two independent events happening simultaneously due to the presence of dynamic crowders in the bulk is determined approximately by the following equation:

$$\frac{1}{T_m} = \frac{1}{t_1} + \frac{1}{t_2} \tag{2.8}$$

and the total search time in the presence of mobile bulk crowders can be approximated as

$$\langle T\rangle \cong \langle T_0\rangle + P_m T_m = \langle T_0\rangle + \frac{\Delta}{L}\left(\frac{t_1 t_2}{t_1+t_2}\right). \tag{2.9}$$

The analytical expression of $\langle T_0 \rangle$ is given by Eq. 2.2 and it is the search time in the absence of bulk crowders. Equation 2.9 is the general approximate expression of the average search time in the presence of dynamic crowders adjacent to the DNA in bulk. In the presence of a single dynamic crowder around the DNA chain ($\Delta = 1$), Eq. 2.9 reduces to Eq. 2.4 as shown in the next section. Equation 2.9 is valid for all binding sites occupied by the target except for the one near the ends of the DNA chain such that the position of the target is less than the magnitude of $\Delta/2$ (i.e. $m < \Delta/2$ or $(L - m) < \Delta/2$). For this case, symmetric crowding can never be achieved for the target site which has a major contribution to the time t_2 in this analysis of T_m. Also if the crowders are present at the ends of the DNA ($m = 1$, $m = L$), the crowders diffuse linearly only in one direction along the DNA to make the target site free from blocking. This will lead to a reduction in the blocking time causing the search time to be almost independent of the presence of crowders.

2.3 Results and Discussion

To investigate the role of dynamic cellular crowders in the protein search process, it is assumed that the crowders are in the vicinity of the DNA and are sliding along the DNA chain in bulk both in forward and backward direction with a rate v. Three dynamic search regimes are identified depending on the relative values of the scanning length $\lambda = \sqrt{u/k_{off}}$ that corresponds to the distance the protein moves on the DNA chain before it dissociates into the solution, the length of DNA, the target size which is considered to be equal to one. For simplicity, let us first consider that the crowder is of size $\Delta = 1$, and then the results are extended to dynamic crowders of any size. As shown in figures 2.3 and 2.4, there exist three different search regimes depending on the range of parameters for the case of a single crowder ($\Delta = 1$) and also for the more realistic situation when the size of the bulk crowders is such that they block more than one site of the DNA chain. From experimental measurements of protein search dynamics, for example, for *lac* repressor proteins, the transition rates u, k_{on}, and k_{off} were estimated as $u \approx 10^3 - 10^6$ s^{-1} , $k_{off} \approx 200 - 3000$ s^{-1}, $k_{on} \approx 10^4 - 10^6$ s^{-1}[10, 11].

When the crowders are present in the bulk in the close vicinity of DNA, they obstruct the protein molecule to bind to the sites that are blocked and the protein can access (bind/unbind) only those sites that are free from crowding. When the scanning length is less than the size of the target ($\lambda < 1$), the protein molecule mostly finds the target

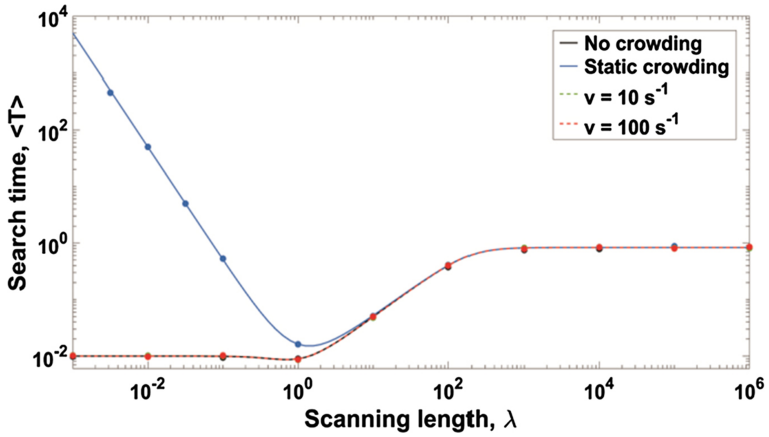


Figure 2.3: Average search times without any crowding, with static crowding ($v = 0$), and in the presence of a single mobile crowder as a function of scanning length λ for different rates of diffusion of the crowder. Other chosen parameters are $L = 10^3$, $m = L/2 = 500$, $k_{on} = 10^2$ s^{-1} and $u = 10^5$ s^{-1}. For static crowding, the blocked site is the target site. The curves are the theoretical results and the symbols correspond to Monte Carlo simulations.

site by the 3D pathway with continuous binding and unbinding events at the unblocked sites. For a bulk crowder of size Δ, there are $L - \Delta$ sites that are available for the protein molecule to bind. When the mobile bulk crowders block just one site of the chain ($\Delta = 1$), then $L - 1(\approx L)$ sites are free from crowding and the protein locates the target site irrespective of the position of the crowders as it can access almost all the sites except one via 3D binding/unbinding. Though blocking the target site has a maximum effect on the search dynamics, it is very unlikely that always the target site will be blocked by the single crowder in all the search processes to make a significant contribution to the search dynamics on average.

In the limit of $\lambda < 1$ where the unbinding rate (k_{off}) is very high, the function $S_0(s)$ in Eq. 2.3 at $s \to 0$ is $S_0(0) \simeq 1$ and this results in the first term of Eq. 2.9 to be approximately $\langle T \rangle \approx 1/k_{on}$[28]. In the second term on the right-hand side of Eq. 2.9, only t_2 is a function of the scanning length as can be observed from Eq. 2.7 since

$$z(s=0) = 1 + \frac{1}{2\lambda^2} - \sqrt{\frac{1}{4\lambda^4} + \frac{1}{\lambda^2}} \tag{2.10}$$

and t_2 in this limit can be written as $t_2 = k_{on}^{-1}\left[\frac{1}{1+x(1-z(0))} + \frac{1}{1+y(1-z(0))}\right]^{-1}\left(\frac{1}{z(0)} - 1\right)$ as shown in[27]. For lower scanning lengths, $z(0) \to 0$, $1 - z(0) \to 1$, and $\left(\frac{1}{z(0)} - 1\right) \approx \lambda^{-2}$. Using the expression of t_1 defined in Eq. 2.5 and combining all these approximated expressions, the average search time in presence of dynamic crowders in the limit of $\lambda < 1$ is given by

$$\langle T\rangle \approx \frac{1}{k_{on}} + \frac{\Delta}{L}\left[\frac{12v}{(\Delta+1)(\Delta+2)} + k_{on}\lambda^2\Big(\frac{1}{1+x}+\frac{1}{1+y}\Big)\right]^{-1}. \tag{2.11}$$

When the crowder size is $\Delta = 1$, the constants $x = y = 0$, which results in the Eq. 2.11 to be modified as,

$$\langle T\rangle \simeq \frac{1}{k_{on}} + \frac{1}{2L}\left(\frac{1}{v + k_{on}\lambda^2}\right) \approx \frac{1}{k_{on}} + \frac{1}{2L} \tag{2.12a}$$

For long chain lengths,

$$\langle T\rangle \simeq \frac{1}{k_{on}}. \tag{2.12b}$$

So, for a static crowder that blocks only the target site, the search time is largely affected in this 3D regime whereas the overall search time remains constant when such bulk crowders are mobile. This is also evident in figure 2.3 where the search time is plotted as a function of scanning length for static (the blue line) as well as mobile crowders for $\Delta = 1$ with varying mobility rate constants.

However, as the size of the blocked region increases, the probability of the target site being blocked increases which will affect the search process since in this 3D diffusion regime the protein finds the target mostly by independent association/dissociation events. If the target site is blocked, the search continues until and unless the crowders move away from the target site, and hence the search time increases. So, it is important to understand the role of the size and the mobility of the crowders in this regime. When the unbinding rate of the protein is very high such that the protein immediately dissociates after binding and never takes up a 1D pathway ($\lambda << 1$), the 3rd term on the right-hand side of Eq. 2.11 has no contribution to the total search time. In this case, the protein will find the target site only when the crowder diffuses away from the target making it available for the protein to recognize and bind. So Eq. 2.11 then reduces to $\langle T\rangle \approx \frac{1}{k_{on}} + \frac{\Delta(\Delta+1)(\Delta+2)}{12vL}$. This is an important result. The search time becomes independent of the scanning length when $\lambda << 1$ and the search time then depends only on the properties of the crowders such as their size and mobility. The search process becomes slower with an increase in the crowder size and with the decrease in the mobility of the crowder as can be observed from figures 2.4(a) and (b) respectively. Within this 3D regime, as the scanning length increases (keeping $\lambda < 1$), the protein molecule can undergo some sliding motion on the DNA chain, and then t_2 (the third term on the RHS of Eq. 2.11) also contributes to the total search

time along with t_1. This additional contribution though small in magnitude aids in finding the target site faster via an assisted 1D sliding pathway. This leads to a slight decrease in the total search time in this jumping regime and an inverse dependence of the search time with the scanning length is observed. From Eq. 2.11, it is evident that for fixed values of x and y, the t_2 term will make some contribution to the average search time only when the quantity $k_{on}\lambda^2 > 1$. Thus, the search time remains constant till $\lambda < 1/\sqrt{k_{on}}$, and then it decreases with an increase in λ until it achieves a minimum near $\lambda \sim 1$ because around this value both the 1D and 3D pathway contribute equally to the total search time. From figures 2.4(a) and (b), with the binding rate per each site considered as 100 s^{-1}, the search time remains constant till $\lambda \approx 0.1$ after which it decreases and then attains a minimum value at around $\lambda \approx 1$. This observation is quite important because, within the jumping regime, two different behaviors of the search time are observed. Initially, the search time remains constant as a function of the scanning length and later decreases after attaining some particular value of the scanning length. This is in contrast to the search dynamics of the protein in the presence of static crowders where the search time is inversely dependent on the scanning length within this entire 3D search regime.

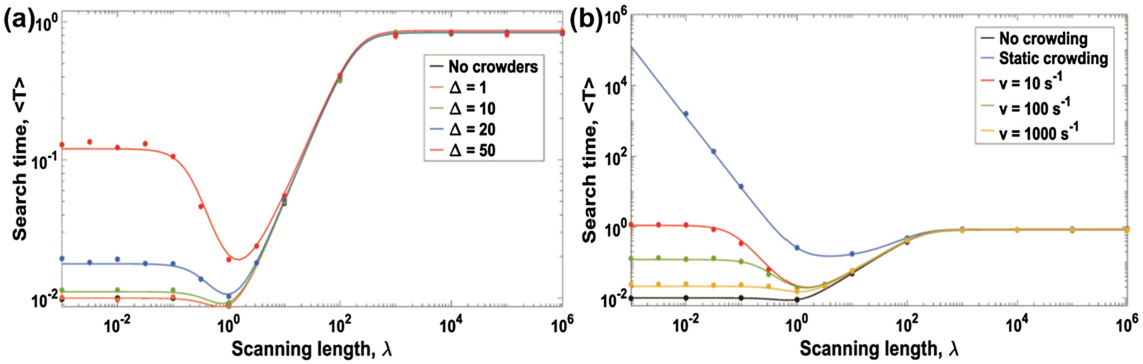


Figure 2.4: Average target search times as a function of scanning length λ without any crowders in the bulk solution and in the presence of mobile crowders of varying (a) lengths with their mobility rate constant at $v = 10^2$ s^{-1} and (b) mobility rate while keeping the length of the blocking region $\Delta = 50$. The curves correspond to the theoretical results and the symbols correspond to Monte Carlo computer simulations. Other chosen parameters are $L = 10^3$, $m = 500$, $k_{on} = 10^2$ s^{-1} and $u = 10^5$ s^{-1}.

Next, the search dynamics in the sliding regime ($1 \leq \lambda \leq (L - \Delta)$) is discussed, where the search process is a combination of 3D and 1D motion. The protein binds to a non-specific site, slides along the DNA, dissociates and this process continues until the target site is reached. In the presence of molecular crowding, for $\Delta = 1$, there are $L - 1(\approx L)$ unhindered sites on which the protein molecule can independently bind and can slide along the chain to reach the target. Thus, there is no effect of the crowder

and the search process remains unaffected by the presence of the crowder irrespective of its mobility rate. As discussed before, the search process is delayed with an increase in the length of the blocked region and with a decrease in the mobility of the crowders in the 3D pathway where the protein molecule has to bind to the different sites of the DNA from the bulk solution. However, due to added contribution from the 1D sliding event in this sliding regime, the bulk crowders affect the search process to a less extent than in the jumping regime. The effect of the crowders of length $\Delta > 1$ is primarily observed at low scanning lengths where the 3D search pathway dominates and the search process then is influenced due to a relatively high number of dissociation events taking place before the target site is finally reached either by 3D pathway or a 1D pathway. As the scanning length gradually increases, the 1D search pathway takes over, and then the protein molecule can slide along the DNA chain to arrive at the target site without unbinding frequently from any non-specific sites. Since bulk crowders do not affect the 1D sliding pathway, their effect decreases as λ increases. Mathematically, the first term of Eq. 2.9 is approximately $\langle T_0 \rangle \approx \frac{L\lambda}{2u}$ since the function $S_0(0) \approx \sqrt{1+4\lambda^2}$ when the scanning length varies as $1 < \lambda < L - \Delta$[28]. The search time when the crowder is static is approximately, $t_2 \approx \frac{L-\Delta}{u}\left(\frac{1}{x+\lambda} + \frac{1}{y+\lambda}\right)^{-1}$ [27]. So, the final expression of search time (Eq. 2.9) can be approximated in this limit to be

$$\langle T \rangle \simeq \frac{L\lambda}{2u} + \frac{\Delta}{L}\left[\frac{12v}{(\Delta+1)(\Delta+2)} + \frac{u\left(\frac{1}{x+\lambda} + \frac{1}{y+\lambda}\right)}{L-\Delta}\right]^{-1}. \tag{2.13}$$

For the set of chosen parameter values and for a given chain length of the DNA which is considered to be long, the contribution of the last two terms in Eq. 2.13 is insignificant and $\langle T \rangle \approx \frac{L\lambda}{2u} \approx \langle T_0 \rangle$. Thus mathematically one can show that neither the size nor the mobility of the crowders affects the search dynamics in this 1D + 3D sliding regime.

Lastly, when the scanning length is high enough, $\lambda > L - \Delta$, once the protein molecule binds to any site on the DNA chain, it finds the target site predominately by a 1D diffusion mechanism along the chain backbone. Thus, the bulk crowders do not affect the target search process in this random walk regime regardless of their size or mobility which is also observed in figures 2.4(a) and (b). From Eq. 2.9, $\langle T_0 \rangle \approx \frac{L^2}{12u}$ when $\lambda > L - \Delta$[28]. The expressions of $K(a, x, s)$ and $K(b, y, s)$ can be expanded linearly in this limit at $s \to 0$ as

$$K(a,x,0) = a - \frac{a\lambda^{-2}}{6}(1+3a+2a^2+6ax) + O[\lambda^{-4}] \tag{2.14a}$$

$$K(b,y,0) = b - \frac{b\lambda^{-2}}{6}(1+3b+2b^2+6by) + O[\lambda^{-4}] \tag{2.14b}$$

where the constants a, b and y have been defined previously for a known value of x. This will result in $t_2 \approx \frac{L^3}{12u(L-\Delta)}$[27]. For large values of the L, the total search time in this regime becomes

$$\langle T \rangle \simeq \frac{L^2}{12u} + \frac{\Delta}{L}\left[\frac{12v}{(\Delta+1)(\Delta+2)} + \frac{12u(L-\Delta)}{L^3}\right]^{-1}. \tag{2.15}$$

For the chosen parameters and long DNA chain lengths $\langle T \rangle \approx \frac{L^2}{12u} \approx \langle T_0 \rangle$. Thus the search time in this regime is independent of the scanning length and is also not affected by the change in the crowder size or its mobility (figures 2.3 and 2.4). Presented theoretical results are complemented with extensive Monte Carlo simulations.

2.4 Conclusion

This chapter discussed a theoretical approach to explicitly investigate the role of bulk crowding on protein search dynamics when dynamic crowders are located adjacent to the DNA. Three different dynamic search phases were explored based on the scanning length. The effect of the dynamic crowders was different in these search regimes depending on the size and the mobility of the crowder. From theoretical/simulation results it was confirmed that for single-site crowding, the crowder does not modify the search dynamics and the search time is invariant of the crowder properties such as its mobility. This is in contrast to static crowding, where it was found that static crowding agents affect the target search dynamics of the protein[27], provided this single blocking site is the target site. For mobile bulk crowders that block more than one binding site, three search regimes with different dynamics were identified. In the random walk regime where the search is taking place only via 1D sliding events, the crowder does not modify the search dynamics. In the sliding regime, where the search is a combination of 3D and 1D modes, initially, the search process is slower in the presence of crowders but with the increase in the scanning length, this effect diminishes. However, these crowders significantly decelerate the protein search in the jumping regime where the search is taking place only via 3D binding/unbinding events.

Here two important factors can modulate the search process: the mobility of the crowding molecules and the size of the crowder. The lower the sliding velocity of the crowder, the longer is the search time. An increase in the sliding velocity of the crowder increases the probability to diffuse away from the target site and facilitates the protein to bind to the target site. Thus, highly mobile crowders will not affect the search dynamics even if there is significant crowding. Thus for dynamic crowders, if there is a single site crowding or if the sliding velocity of a crowder that blocks multiple binding sites is very high, the search dynamics remain unaffected and this can provide a molecular-level understanding of the Brownian dynamics simulation predictions as reported in[13].

At interim mobility rates, increasing the crowder size delays the search process because the target site will remain blocked for a longer time. For a given length of the crowder, the search is faster in systems surrounded by mobile crowders in comparison with the static crowders. But this effect can be observed only in the 3D jumping regime, where the scanning length is shorter than the size of the target. The approximate theoretical picture shows that in this regime, the search dynamics is different for the systems with dynamic crowders in comparison with the systems with static crowders. For mobile crowders, the search time initially remains constant with respect to the scanning length in this 3D regime and this later shows an inverse dependence with the scanning length along with being influenced by the crowder size and their mobility. For static crowders, the search time is a rapidly decreasing function of the scanning length by few orders of magnitude.

Coarse-grained MD and Monte Carlo simulations in[25] reflected that small sized and lighter crowder particles (having high mobility) enhance the search process which agrees with the presented theoretical predictions. This simulation study also showed that facilitated diffusion of the proteins in a crowded environment is promoted due to the contribution of high 1D diffusion in the sliding and the hopping regime whereas this theory suggests that with dynamic crowders in the bulk, the 3D search pathway is hindered while the linear 1D search pathway remains unaffected. The latter agrees with extensive Langevin dynamics simulation results reported in[26], where it is observed that during 1D diffusion, the search of the protein for its target site on the DNA molecule is independent of the size, shape, and mobility of the crowder. In analyzing the role of crowding molecules on the 3D diffusion of a protein, it was shown that the 3D protein diffusion depends on the physical properties of the bulk crowder molecules which is also confirmed from this work. The theoretical predictions are fully validated by extensive Monte Carlo computer simulations. Corresponding microscopic explanation of the role of bulk crowders in the

protein search are also discussed. The bulk crowders can accelerate or decelerate or even lead to no changes at all depending on the nature of crowders (static or mobile), the length of the crowding, or their mobility rate. In the 3D diffusion regime, the dynamics can be similar to the case of the search on a DNA with no bulk crowding under three scenarios: static crowders that block non-specific sites, dynamic crowders that block a single site on the DNA, or dynamic crowders having high mobility. It is believed that this approximate analytical approach is the first of its kind that explicitly takes into account the role of dynamic crowders in the vicinity of the DNA and analyzes the effect of such crowders in the bulk solution in the protein search dynamics.

To understand the importance of the theoretical results for realistic situations, it is crucial to look into in vitro experiments carried out via repair enzymes on human DNA which indicate that the search process involves a combination of 3D and 1D motion[23]. Experimental study on the translocation probability of repair enzymes between two damaged sites on the DNA shows that the crowding effect increases both the efficiency and average distance of DNA chain translocation by hindering escape of the enzymes to the bulk solution. It was shown that crowding biased the search process toward the sliding and hopping mode over the 3D diffusion mode and this has been validated through theoretical and computational studies. The current study also shows that with the increase in crowding, it is difficult to adopt the 3D search pathway as compared to the sliding and hopping pathway. Moreover, the experimental study also monitored the association times of the enzymes on the DNA with the change in the viscosity of different crowders and the experimental results indicate that slow moving crowders (high viscosity) tend to delay the association process of the enzymes to the DNA which can directly be explained from the analytical results because the slow mobility of the crowders increases the target search time in the 3D search regime which is governed by the association rate of the protein on the DNA.

Thus, this proposed simple theoretical method provides a physical-chemical understanding and a complete picture of the role of bulk crowding in a cellular environment on the target search dynamics, static as well as dynamic crowding. But the intracellular environment around DNA is quite complex and many important phenomena during the protein-DNA search process such as the interactions between the crowder and the DNA, interactions between the searching proteins and crowders in the solution[32] are not taken into account and it is expected to affect the facilitated diffusion of DNA binding proteins.

Bibliography

(1) Alberts, B. et al., *Molecular Biology of the Cell*; Garland Science, Taylor and Francis Group: 2014.

(2) Lodish, H. F. et al., *Molecular Cell Biology*; WH Freeman New York: 2006.

(3) Phillips, R. et al., *Physical Biology of the Cell*; Garland Science: 2012.

(4) Riggs, A. D. et al. The Lac Represser-Operator Interaction: III. Kinetic Studies. *J. Mol. Biol.* **1970**, *53*, 401–417.

(5) Mirny, L. et al. How a Protein Searches for its Site on DNA: The Mechanism of Facilitated Diffusion. *J. Phys. A: Math. Theor.* **2009**, *42*, 434013.

(6) Kolomeisky, A. B. Physics of Protein–DNA Interactions: Mechanisms of Facilitated Target Search. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2088–2095.

(7) Marius Clore, G. Exploring Translocation of Proteins on DNA by NMR. *J. Biomol. NMR* **2011**, *51*, 209–219.

(8) Pollak, A. J. et al. Distinct Facilitated Diffusion Mechanisms by E. Coli Type II Restriction Endonucleases. *Biochem.* **2014**, *53*, 7028–7037.

(9) Gowers, D. M. et al. Measurement of the Contributions of 1D and 3D Pathways to the Translocation of a Protein Along DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15883–15888.

(10) Wang, Y. et al. Single Molecule Measurements of Repressor Protein 1D Diffusion on DNA. *Phys. Rev. Lett.* **2006**, *97*, 048302.

(11) Elf, J. et al. Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell. *Science* **2007**, *316*, 1191–1194.

(12) Mahmutovic, A. et al. What Matters for Lac Repressor Search In Vivo — Sliding, Hopping, Intersegment Transfer, Crowding on DNA or Recognition? *Nucleic Acids Res.* **2015**, *43*, 3454–3464.

(13) Brackley, C. et al. Intracellular Facilitated Diffusion: Searchers, Crowders, and Blockers. *Phys. Rev. Lett.* **2013**, *111*, 108101.

(14) Gomez, D.; Klumpp, S. Facilitated Diffusion in the Presence of Obstacles on the DNA. *Phys. Chem. Chem. Phys.* **2016**, *18*, 11184–11192.

(15) Marcovitz, A.; Levy, Y. Obstacles May Facilitate and Direct DNA Search by Proteins. *Biophys. J.* **2013**, *104*, 2042–2050.

(16) Liu, L.; Luo, K. DNA-Binding Protein Searches for its Target: Non-Monotonic Dependence of the Search Time on the Density of Roadblocks Bound on the DNA Chain. *J. Chem. Phys.* **2015**, *142*.

(17) Shvets, A. A.; Kolomeisky, A. B. Crowding on DNA in Protein Search for Targets. *J. Phys. Chem. Lett.* **2016**, *7*, 2502–2506.

(18) Shvets, A. et al. Role of Static and Dynamic Obstacles in the Protein Search for Targets on DNA. *J. Phys. Chem. B* **2016**, *120*, 5802–5809.

(19) Gorman, J. et al. Visualizing One-Dimensional Diffusion of Eukaryotic DNA Repair Factors Along a Chromatin Lattice. *Nat. Struct. Mol. Biol.* **2010**, *17*, 932–938.

(20) Li, G.-W. et al. Effects of Macromolecular Crowding and DNA Looping on Gene Regulation Kinetics. *Nat. Phys.* **2009**, *5*, 294–297.

(21) Cheung, M. S. et al. Molecular Crowding Enhances Native State Stability and Refolding Rates of Globular Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 4753–4758.

(22) Elcock, A. H. Models of Macromolecular Crowding Effects and the Need for Quantitative Comparisons with Experiment. *Curr. Opin. Struct. Biol.* **2010**, *20*, 196–206.

(23) Cravens, S. L. et al. Molecular Crowding Enhances Facilitated Diffusion of Two Human DNA Glycosylases. *Biophys. J.* **2015**, *108*, 76a.

(24) Dey, P.; Bhattacherjee, A. Role of Macromolecular Crowding on the Intracellular Diffusion of DNA Binding Proteins. *Sci. Rep.* **2018**, *8*, 844.

(25) Krepel, D. et al. Mechanism of Facilitated Diffusion During a DNA Search in Crowded Environments. *J. Phys. Chem. B* **2016**, *120*, 11113–11122.

(26) Dey, P.; Bhattacherjee, A. Disparity in Anomalous Diffusion of Proteins Searching for Their Target DNA Sites in a Crowded Medium is Controlled by the Size, Shape and Mobility of Macromolecular Crowders. *Soft Matter* **2019**, *15*, 1960–1969.

(27) Mondal, K.; Chaudhury, S. A Theoretical Study of the Role of Bulk Crowders on Target Search Dynamics of DNA Binding Proteins. *J. Stat. Mech.: Theory Exp.* **2020**, *2020*, 093204.

(28) Veksler, A.; Kolomeisky, A. B. Speed-Selectivity Paradox in the Protein Search for Targets on DNA: Is it Real or Not? *J. Phys. Chem. B* **2013**, *117*, 12695–12701.

(29) Lange, M. et al. Dynamics of the Protein Search for Targets on DNA in the Presence of Traps. *J. Phys. Chem. B* **2015**, *119*, 12410–12416.

(30) Shvets, A. A.; Kolomeisky, A. B. Sequence Heterogeneity Accelerates Protein Search for Targets on DNA. *J. Chem. Phys.* **2015**, *143*.

(31) Mondal, K.; Chaudhury, S. Effect of DNA Conformation on the Protein Search for Targets on DNA: A Theoretical Perspective. *J. Phys. Chem. B* **2020**, *124*, 3518–3526.

(32) Krepel, D.; Levy, Y. Protein Diffusion Along DNA: On the Effect of Roadblocks and Crowders. *J. Phys. A: Math. Theor.* **2016**, *49*, 494003.

Chapter 3

Influence of Nonspecific Interactions between Proteins and In Vivo Cytoplasmic Crowders in Facilitated Diffusion of Proteins: Theoretical Insights

The content of this chapter and associated content in Appendix A are reprinted (adapted) with permission from (*J. Phys. Chem. B* 2022, 126, 16, 3037-3047). Copyright © 2022 American Chemical Society.

3.1 Introduction

The DNA is composed of polynucleotide sequences that hold genetic blueprints for carrying out multiple biological processes such as DNA damage repair, gene regulation, transcription, etc., which are initiated by the association of DNA-binding proteins to specific target sequences on the DNA.[1–3] Ever since the in vitro experiments carried out by Riggs et al.[4] showed that the lac repressor protein's target search is about 2 orders of magnitude faster than its three-dimensional (3D) diffusion rate, there have been numerous experimental and theoretical studies to understand the mechanism of this search process. It was proposed that the protein molecule recognizes its target via facilitated diffusion, where the protein performs one-dimensional (1D) random walk diffusion on the DNA once it nonspecifically binds to a DNA-binding site as a result of 3D diffusion.[5–7] This mechanism of facilitated diffusion has been successfully captured by bulk measurements[8–13] and single-molecule experiments[14–18] both in vitro and in vivo.

Usual theoretical models of the facilitated diffusion consider an oversimplified picture of the target search dynamics by considering only the DNA and a protein molecule.[19–21] However, in living cells, the cytosol is densely packed, and macromolecules alone occupy about 10-40% of the entire intracellular volume that can affect the search dynamics of the protein.[17, 22–26] These macromolecules may either be present on the DNA and pose as "roadblocks" for the search process, or they may lie in the cellular medium and affect the diffusion of the protein. The role of roadblocking obstacles in the target search process has been experimentally and theoretically investigated before.[18, 23, 27–36] It has also been shown that the effect of such roadblocks on the DNA can be circumvented by DNA looping.[33, 37] In this work, the main focus is on the search process which is affected by the macromolecular crowders that lie in the cytoplasm of the cell. To capture the effect of such crowded environments on the target search of the protein, various experiments have been performed by tuning the cellular medium with different kinds of biomacromolecules or ionic environments. It was shown that the translocation of the DNA repair enzymes was enhanced under crowded conditions by modifying the in vitro concentrations of different crowding agents within the medium.[38] The hydrolysis of DNA via endonucleases was enhanced in the presence of poly(ethylene glycol) (PEG) crowders, while it remained unaffected for exonucleases.[39] The molecular reason behind these distinct observations is still not well-understood. The effect of macromolecular bulk crowding was also studied using computational and analytical techniques. By performing extensive molecular dynamics

simulations, important insights into the target search of the repair proteins in crowded environments were provided.[40] Large-scale Brownian dynamics simulations demonstrated an increase or decrease in the search time concerning the protein-DNA affinity in the presence of crowder molecules diffusing within the cytosol.[30] The effect of size and mass of the bulk crowders on the DNA target search for different salt concentrations was closely studied using CGMD and Monte Carlo simulations.[41] Through analytical techniques and Langevin dynamics simulations, it was determined that the search time of the protein increased due to molecular crowding.[42] With the help of a model that strongly drives from the cytoplasm length scale-dependent viscosity, the effect of molecular bulk crowders on the diffusion of the transcription factors along the DNA was probed.[43] Bhattacherjee and co-workers have rigorously studied the effect of molecular crowders and their properties on the DNA target search process through extensive computer simulations.[44–46] The effect on the search dynamics due to bulk crowders lying in the close vicinity of the DNA was investigated theoretically using discrete-state stochastic models by considering the bulk crowders to be either static or dynamic in nature.[47–49]

In all the experimental and theoretical studies discussed above, the bulk crowders are considered to be inert, i.e., these macromolecules do not interact with the searching protein or the DNA in any manner, and their contribution to the search dynamics is entropically driven through excluded volume interactions. However, several pieces of evidence show that cellular crowding is not really "inert", and these macromolecules may nonspecifically interact with the protein, affecting its native stability.[24, 50–52] Senske et al. have proposed a unified framework to describe the stabilizing effects of the macromolecular crowding on the proteins, which is enthalpy-driven rather than entropy-based volume exclusion effects alone.[53] Their work strongly supports the existence of interactions between the protein and intracellular crowders. Importantly, it is again stressed that supposedly "inert" macromolecular crowders such as PEG which are used in vitro tend to interact with the proteins. Several other studies are also focusing on the effect of macromolecular crowding on different biomolecules, their stability, and related biological and biochemical processes.[52, 54, 55]

Such protein-crowder interactions and their subsequent effect on the search dynamics can be observed commonly through in vivo studies. The facilitated diffusion of the lac repressor protein in the presence of the inducer molecules that act as bulk crowders within the medium of Escherichia coli bacteria was monitored using single-molecule techniques.[16] DNA translocation by uracil glycosylase was greatly enhanced in human cells

when compared with in vitro studies that incorporated homogeneous "inert" crowding agents such as PEG, which suggests the importance of nonspecific interactions of crowders in the nuclear environment.[56] The influence of electrostatic interactions due to ionic environments on the DNA repair enzyme's search pathway was examined through in vitro experiments.[57] It was observed that the enzyme-undamaged site interaction was electrostatically driven at an optimum cation concentration, while the enzyme-damaged site interaction was not. Later on, by varying the concentration of salt ions and molecular crowding simultaneously, a more realistic scenario was closely studied via in vitro DNA damage search experiments.[58] The molecular origin of different activities of endonucleases and exonucleases in the presence of PEG crowders on the DNA hydrolysis might be attributed to the apparent PEG interactions with the biomolecules present in the system that might eventually affect the relative activity of the DNA nucleases.[39] The review on the human 8-oxoguanine DNA glycosylase 1 (OGG1) repair enzyme's search for its substrate on the DNA in crowded environments suggests that the nuclear milieu may impact the search process either through crowder-interacting proteins or by confining the protein near the DNA preventing its escape to the bulk.[59] Some computational studies have also been performed recently to incorporate the effects of such nonspecific interactions on the target search process. Extensive Langevin dynamics simulations show that the protein-crowder interactions in the bulk play an important role in enhancing the facilitated diffusion of the protein within a crowded environment. Such all-atom simulations show that the protein can float along the crowders over the DNA under the influence of nonspecific interactions with bulk crowder molecules.[60] This work also emphasizes that attractive interactions are present in PEG-crowded medium, which were otherwise considered as an inert crowding agent. By considering the contribution of electrostatic interactions, it was shown using coarse-grained simulations that the presence of nucleosome crowding in the chromatin enhances the target search but slows down the protein diffusion in the bulk.[61] In a recent computational study, realistic geometry of cells has been reconstructed to understand the cell signaling properties at a much closer level using soft X-ray tomographic images.[62] The effect of different cell organelles that modify the diffusion properties of the biomolecules, which also affects cell signaling and search processes, can be studied using such techniques.

The effect of protein-crowder interactions on protein stability has been extensively studied theoretically.[63–66] However, there are no theoretical studies that explicitly take into account these interactions between the protein and the bulk crowders to probe their

influence on the target search process. Thus, this chapter presents a comprehensive theoretical method of analyzing the role of protein-crowder associations in the DNA target search process. The presented approach is based on a discrete-state stochastic framework that allows us to obtain fully analytical solutions for all the dynamic properties. This theoretical approach has been successfully applied to study the role of static and dynamic inert obstacles,[27, 28] inert bulk crowders,[47–49] and other factors[33, 67–74] in the target search process. This stochastic framework takes into account the most relevant chemical and physical processes of the search process in the presence of interacting bulk crowders lying in the close vicinity of the DNA. The obtained results are supported by Monte Carlo computer simulations. By mimicking in vivo cellular conditions closely, this analysis attempts to offer a deeper understanding of the search process and provide the conditions where such associations with the crowders may or may not affect the protein search on DNA.

3.2 Theoretical Model

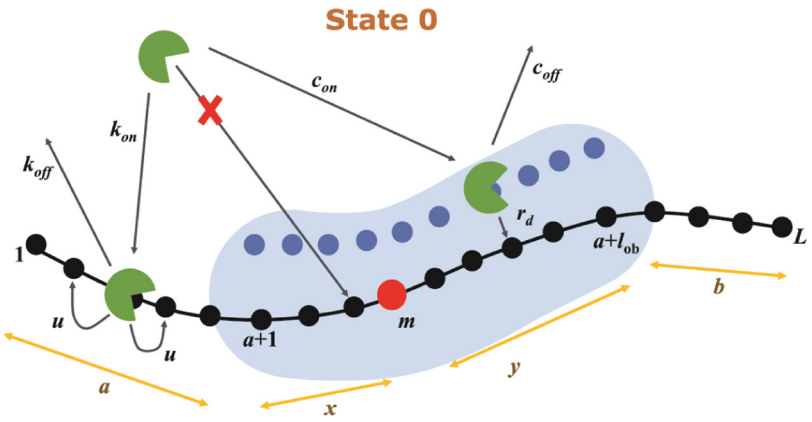


Figure 3.1: Schematic of the target search process of the protein (in green) in the presence of interacting bulk crowders. There are a total of L binding sites (solid black circles) on the DNA chain (black line), and the yellow arrows indicate the number of binding sites within that particular region. The blue-shaded region depicts the crowded region of the DNA, and the solid blue circles represent the bulk crowders that can potentially bind to the protein molecule. The total number of blocked sites is l_{ob}, which includes the target site situated at position m. The unblocked sites on the DNA are $L - l_{ob}$ in total. The protein begins its search from the bulk solution labeled as "state 0." The black arrows represent all the transitions that the protein molecule can undergo. The transitions along with the associated rate constants have been discussed in the main text.

This study presents a simple discrete-state stochastic model to probe the target search dynamics of the protein in the presence of protein-interacting bulk crowders lying close to the DNA, which is schematically shown in Figure 3.1. Let us consider a single protein

molecule that is searching for its specific target site on a DNA chain consisting of L binding sites (L discrete states), out of which the binding site m serves as the target site. The protein always starts from the cellular matrix, also called the bulk solution that is labeled as state 0. Although the DNA-binding protein is capable of multimerization within the biological systems while searching for its respective target site on the DNA,[16, 75–77] this analysis considers a single biological entity (the protein) searching for its target site. This DNA-binding entity can exist as a protein molecule, a dimer, an oligomer, etc., within the biological system. Therefore, the role of multimerization of the proteins is not included in this study. The target-searching protein can bind to any site on the DNA with equal probability, owing to the assumption that the DNA chain is coiled in the solution and the protein diffuses very fast in the volume around the DNA. This protein associates with the DNA directly with a binding rate k_{on} per DNA-binding site. While being on the DNA, the protein can slide along the chain with equal probability in both directions with a diffusion rate u or it can dissociate from the DNA to the bulk solution with a rate k_{off}. The protein molecule can undergo hopping events where after the protein-DNA complex dissociates into the bulk, the protein rebinds to the DNA again at a nearby site. This is referred to as a "microhop". If the rebinding happens at some other distant sites, it is termed as a "macrohop".[12, 78] In the current model, "microhops" and "macrohops" cannot be distinguished as the protein may bind to any DNA site that is free from crowding. This is due to the assumption that the DNA chain is coiled in the solution, which assures that the next association event is independent.

In this work's theoretical model, the target is situated at the middle of the DNA chain, as from symmetry arguments, this leads to the shortest search time.[19] The bulk crowders are lying in vicinity of the DNA molecule such that they block l_{ob} number of binding sites on the DNA and the protein cannot bind to these sites directly (see Figure 3.1). A previous theoretical study has confirmed that there is a significant change in the target search dynamics when bulk crowders block the target site compared to the case when the target site is free from any crowding.[48] Based on this important result, the current study assumes that l_{ob} includes the target site as well. To investigate the role of nonspecific protein-crowder associations in the target search, the protein lying in the bulk solution may associate with the crowder via rate c_{on}. A larger value of c_{on} will lead to a faster association of the protein with the crowder. This protein can then either dissociate back to the bulk solution with the rate c_{off} or access the DNA binding sites lying in the blocked region. The protein can bind to the crowded DNA from the crowder-protein

complex in multiple ways such as direct transfer, multiple reversible bindings, through a transition state, etc.[16, 18, 38, 56, 60, 79] Since a clear understanding of the protein's transfer to the crowded DNA is unknown, a simplified situation is considered where the rate of binding to the DNA region that is blocked by the crowder is represented by a single parameter r_d that encompasses all these types of binding modes. Thus, the rates c_{on}, c_{off}, and r_d take care of all the processes that the protein may undergo before it finally binds to the DNA in the crowded region as shown in Figure 3.1. The search process ends when the protein reaches the target for the first time.

To proceed with the analysis, the powerful method of first-passage probability densities is utilised, which has been successfully employed before on previous protein-DNA target search studies.[19, 27, 28, 48, 49, 67, 72] Figure 3.1 depicts this work's proposed discrete-state stochastic model for the target search dynamics of the protein in the presence of interacting bulk crowders. One can introduce a probability density function $F_n(t)$ that describes the probability of reaching the target site m at time t for the first time provided initially, the protein was either at the DNA binding site n or in the bulk (state $n = 0$) at $t = 0$. Similarly, the function $F_n^c(t)$ describes the probability distribution to find the target for the first time provided the protein-crowder complex blocks the DNA binding site n at $t = 0$. The temporal evolution of all these functions is described by a set of backward master equations.

For the unblocked DNA binding sites when $n \in [2, a] \cup [a + l_{ob} + 1, L - 1]$

$$\frac{dF_n(t)}{dt} = uF_{n+1}(t) + uF_{n-1}(t) + k_{off}F_0(t) - (2u + k_{off})F_n(t) \tag{3.1a}$$

while at DNA ends ($n = 1$ and $n = L$),

$$\frac{dF_1(t)}{dt} = uF_2(t) + k_{off}F_0(t) - (u + k_{off})F_1(t) \tag{3.1b}$$

$$\frac{dF_L(t)}{dt} = uF_{L-1}(t) + k_{off}F_0(t) - (u + k_{off})F_L(t) \tag{3.1c}$$

In addition, for the crowded region of the DNA when $n \in [a+1, a+l_{ob}]$ except for $n = m$, the backward master equations are

$$\frac{dF_n(t)}{dt} = uF_{n+1}(t) + uF_{n-1}(t) - (2u)F_n(t) \tag{3.1d}$$

$$\frac{dF_n^c(t)}{dt} = r_dF_n(t) + c_{off}F_0(t) - (r_d + c_{off})F_n(t) \tag{3.1e}$$

for protein-DNA bound states and protein-crowder bound states, respectively. When the protein lies in the bulk solution, $n = 0$, one writes

$$\frac{dF_0(t)}{dt} = k_{on}\left[\sum_{n=1}^{n=a} F_n(t) + \sum_{n=a+l_{ob}+1}^{n=L} F_n(t)\right] + c_{on}\sum_{n=a+1}^{n=a+l_{ob}} F_n^c(t) - \Big[(L-l_{ob})k_{on} + l_{ob}c_{on}\Big]F_0(t). \tag{3.1f}$$

Furthermore, the final condition corresponds to the situation when the target search is completed.

$$F_m(t) = \delta(t) \tag{3.1g}$$

These backward master equations are solved by employing the Laplace transformations of the first-passage probability density functions. Thus, for the crowding-free DNA binding sites $n \in [2, a] \cup [a + l_{ob} + 1, L - 1]$,

$$(s + 2u + k_{off})\tilde{F}_n(s) = u\tilde{F}_{n+1}(s) + u\tilde{F}_{n-1}(s) + k_{off}\tilde{F}_0(s) \tag{3.2a}$$

while at the boundary sites of the DNA which are $n = 1$ and $n = L$, respectively,

$$(s + u + k_{off})\tilde{F}_1(s) = u\tilde{F}_2(s) + k_{off}\tilde{F}_0(s) \tag{3.2b}$$

$$(s + u + k_{off})\tilde{F}_L(s) = u\tilde{F}_{L-1}(s) + k_{off}\tilde{F}_0(s) \tag{3.2c}$$

Within the crowded region $n \in [a + 1, a + l_{ob}]$ except $n = m$, for the protein-DNA bound states and the protein-crowder bound states, respectively,

$$(s + 2u)\tilde{F}_n(s) = u\tilde{F}_{n+1}(s) + u\tilde{F}_{n-1}(s) \tag{3.2d}$$

$$(r_d + c_{off})\tilde{F}_n(s) = r_d\tilde{F}_n(s) + c_{off}\tilde{F}_0(s) \tag{3.2e}$$

When the protein lies in the bulk solution, $n = 0$, or is bound to the target site, $n = m$, the Laplace transforms are, respectively

(3.2f)

$$\Big[s + (L-l_{ob})k_{on} + l_{ob}c_{on}\Big]\tilde{F}_0(s) = k_{on}\left[\sum_{n=1}^{n=a}\tilde{F}_n(s) + \sum_{n=a+l_{ob}+1}^{n=L}\tilde{F}_n(s)\right] + c_{on}\sum_{n=a+1}^{n=a+l_{ob}}\tilde{F}_n^c(s). \tag{3.2g}$$

$$\tilde{F}_m(s) = 1. \tag{3.2h}$$

This set of linear equations can now be solved to obtain analytical expressions for the Laplace transforms of the first-passage probability functions as shown in the Appendix A. The explicit expressions for the first-passage probability distribution functions in the Laplace space provide us with a full dynamic description of the protein's search for its target. The aim of this study is to calculate the average target search time of the protein if its original position was in bulk solution, i.e., state 0. This can be obtained using

$$\begin{aligned} T_0 &= -\frac{\partial \tilde{F}_0(s)}{\partial s}\bigg|_{s=0} \\ &= \frac{1+k_{on}^c l_{ob}/r_d}{k_{on}S_3(0)+k_{on}^c\big[l_{ob}-S_4(0)\big]} + \frac{1}{k_{off}}\left\{\frac{k_{on}(L-l_{ob})+k_{on}^c l_{ob}}{k_{on}S_3(0)+k_{on}^c\big[l_{ob}-S_4(0)\big]}\right\} \\ &+ \frac{1}{k_{on}S_3(0)+k_{on}^c\big[l_{ob}-S_4(0)\big]}\left[\frac{x(x+1)}{2u}\left\{k_{on}S_{3,ax}(0)+k_{on}^c\Big(\frac{1+2x}{3}-S_{4,ax}(0)\Big)\right\}\right. \\ &\left. + \frac{y(y+1)}{2u}\left\{k_{on}S_{3,by}(0)+k_{on}^c\Big(\frac{1+2y}{3}-S_{4,by}(0)\Big)\right\}\right] \end{aligned} \tag{3.3}$$

The analytical expressions of the auxiliary functions $S_3(s)$ and $S_4(s)$ have been derived and discussed in the Appendix A.

$$S_{3,ax}(s=0) = \frac{p(p^{2a}-1)}{(p-1)\Big[(1+p^{1+2a})(1+x)-(p+p^{2a})x\Big]}\bigg|_{p=p(s=0)} \tag{3.4a}$$

$$S_{4,ax}(s=0) = \frac{x(x+1)}{2}\left[\frac{(p-1)(p^{2a}-1)}{(1+p^{1+2a})(1+x)-(p+p^{2a})x}\right]\bigg|_{p=p(s=0)} \tag{3.4b}$$

where

$$p(s=0) = \frac{2u+k_{off}-\sqrt{(2u+k_{off})^2-4u^2}}{2u}. \tag{3.4c}$$

Additionally,

$$k_{on}^c = c_{on}\frac{r_d}{r_d+c_{off}} \tag{3.4d}$$

is defined as an effective rate constant for the protein molecule lying in the bulk to bind to the blocked DNA sites, Figure 3.1, via the protein-crowder complex.

The physical meaning of each term of the average search time expression in Eq. 3.3 can be explained as follows. The parameter $S_3(s=0)$ describes the average number of free DNA binding sites scanned by the protein during a single search event. The parameter

$S_4(s=0)$ defines the average number of protein-crowder binding events that happen in one search. Consequently, the quantity $(l_{ob} - S_4(0))$ describes the average sites that lie in the crowded region and are scanned by the protein throughout a single search. The first term in Eq. 3.3 represents the contribution of the average number of association events that occurred during the search process to the total search time. The second term corresponds to the average time of the dissociation events of the protein back into the bulk solution. This quantity is lesser by one than the number of search cycles because the last association to DNA will lead to the recognition of the target site and no further dissociations from DNA will take place. The last two terms of Eq. 3.3 represent the protein's search on the DNA within the crowded region from either side of the target site.

When there are no associations between the protein and the bulk crowder molecules, i.e., $c_{on} = 0$, the average search time exactly agrees with the case of inert static bulk crowders that block the target site of the DNA.[48]

$$T_0\Big|_{c_{on}=0} = \frac{1}{k_{on}S_3(0)} + \frac{1}{k_{off}}\left[\frac{L-l_{ob}}{S_3(0)} - 1\right] + \frac{x(x+1)S_{3,ax}(0) + y(y+1)S_{3,by}(0)}{2uS_3(0)} \tag{3.5}$$

In the absence of bulk crowders,[19] $l_{ob} = x = 0$ in the Eq. 3.3 and one obtains

$$T_0\Big|_{l_{ob}=x=0} = \frac{1}{k_{on}S'(0)} + \frac{1}{k_{off}}\left[\frac{L}{S'(0)} - 1\right] \tag{3.6a}$$

with

$$S'(0) = \frac{p(1+p)(1-p^{2L})}{(1-p)(p+p^{2m})[1+p^{1+2(L-m)}]}. \tag{3.6b}$$

Thus, the analytical result in Eq. 3.3 successfully offers an agreement with previous theoretical studies under certain limits.

From this theory, one can also analyze the first-passage time distribution (FPTD) of the search process provided the protein starts from the bulk solution, $F_0(t)$, by carrying out a numerical inverse Laplace transform of $\tilde{F}_0(s)$ and plotting it against time (see Fig. A.1). The distributions span over decades of time scales. However, this is not the entire distribution as discussed in some existing theoretical studies.[80, 81] The probable reasons for not observing three distinct times as shown in these studies can be due to several factors. The geometry of the system is not relevant in this study and as a result, the

external boundary of the system (which represents the boundary of the cell) cannot be defined. The current theory is based on a discrete-state stochastic approach where the protein is constrained to exist in a limited number of distinct states. Thus, the resulting FPTD would not be the same as the one described using a continuous modeling approach as discussed in the above-mentioned references. In this work's model, the entire cellular matrix is considered to be a discrete state (0^{th} state). Therefore, one cannot determine the initial distance between the protein and the target/DNA at the beginning of the search process.

3.3 Results and Discussion

Since this study assumes the protein to have an affinity for the crowder molecules present in the bulk, the existence of a new crowder-assisted search pathway for the protein to find its target is proposed. This crowder-assisted pathway involves two steps - the association of the protein to the crowders with rate c_{on} followed by the formation of a complex between the protein and the blocked DNA sites with rate r_d. It is important to note that the protein-crowder complex can also dissociate, allowing the protein to return to the bulk solution with rate c_{off} without forming a protein-DNA complex as shown in Figure 3.1. A sufficient protein-crowder binding rate c_{on} will ensure frequent associations of the protein molecule with the crowder in the bulk solution. After this first step of the formation of a protein-crowder complex, it is necessary to transform it into a DNA-protein complex readily by binding of the protein to the blocked DNA sites without dissociating to the bulk solution. To accomplish this, $r_d > c_{off}$. The protein molecule can then scan the blocked region of the DNA via 1D sliding to complete its search.

Next, let us consider some more realistic situations of the existence of protein-interacting crowders within the proposed discrete-state model. How does the strength of the protein-crowder associations affect the target search process? Will the size and positioning of the crowders about the DNA target site modify the search of the protein in any manner? Is there a possibility that such bulk crowders will not modify the search process at all? To provide an answer to these questions, the dynamic properties of the system are explored using a theoretical approach that allows us to obtain a comprehensive description of a protein's target search process. Let us introduce four length scales in the system: the length of the DNA, the size of the target site which is assumed to be unity, the scanning length parameter $\lambda = \sqrt{u/k_{off}}$ that corresponds to the average distance

that the protein hops along the DNA chain before it dissociates back into the bulk, and the size of the bulk crowders l_{ob} ($1 < l_{ob} < L$) that block a certain region of the DNA including the target site.

As described previously in other simple protein-DNA interaction models, the search time is studied as a function of the scanning length. It is convenient to use this parameter for analysis because it provides an idea of the strength of the interactions between the protein and the DNA during the search process. The scanning length covers a wide range as it can be less than one (the size of the target) and greater than L (the length of the DNA). Depending on its values, three dynamic search regimes for the proteina are determined. In the limit $\lambda < 1$, when the scanning length is less than the target size, the protein adopts a 3D search pathway where it constantly binds to/unbinds from the DNA to locate its target and little to no sliding on the DNA can take place. This dynamic phase is called the jumping regime. When $\lambda > L$, the protein molecule has a strong affinity to nonspecifically bind to the DNA and it does not dissociate until the target is found. The DNA-bound protein performs a simple random walk and finds the target in only one search cycle. Due to the redundancy of the random walk, many sites are repeatedly visited and the search in this regime is slow. This dynamic phase is called the random walk regime. For intermediate scanning lengths, $1 < \lambda < L$, the protein molecule performs a combined 1D + 3D search. The protein can scan some length on DNA in every search cycle. In this case, the protein molecule visits the DNA several times before finding the target site, and the time for each cycle increases with the scanning length. This search phase is called the sliding regime. Overall, such a phase diagram provides a full description of the protein search dynamics and is employed it to study the search process in the presence of nonspecific associations between bulk crowders and the protein molecule.

3.3.1 Effect of the Strength of the Protein-Crowder Associations

Figure 3.2 shows the three dynamic search regimes for varying strengths of nonspecific associations between the protein and bulk macromolecular crowders.

When $\lambda < 1$ (jumping regime), the protein adopts a 3D search pathway where it will occasionally bind to the DNA, but it cannot slide and will soon dissociate back into cellular medium. However, the search process can continue via the proposed crowder-assisted pathway. When the protein-crowder associations are strong (high c_{on}), the search becomes relatively faster as the protein can indirectly access the blocked DNA sites through

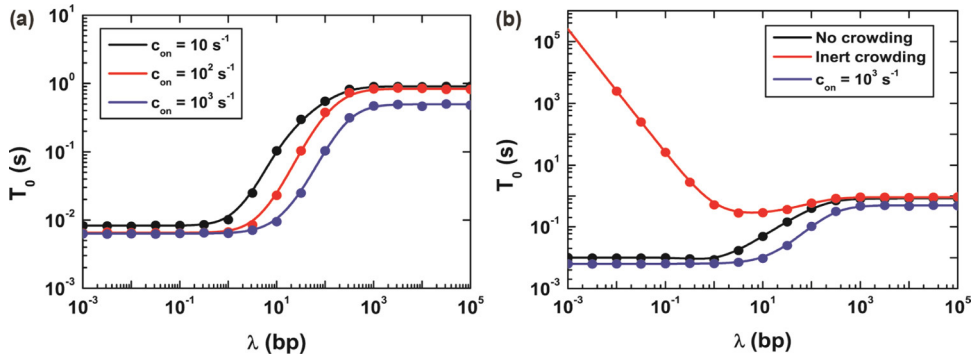


Figure 3.2: Target search time as a function of the scanning length for (a) varying protein-crowder association rate c_{on} and (b) high affinity of the protein and bulk crowders ($c_{on} = 10^3$ s^{-1}) while maintaining $r_d > c_{off}$ compared to the case of inert bulk crowders and in the absence of crowders. Other parameters are $L = 1001$, $m = 501$, $l_{ob} = 101$, $u = 10^5$ s^{-1}, $k_{on} = 10^2$ s^{-1} per each site, $c_{off} = 1$ s^{-1}, $r_d = 10^3$ s^{-1}, and $x = 50$. The solid curves are from the analytical expression in Eq. 3.3, while the symbols correspond to Monte Carlo computer simulations. The details of the simulations are given in the Appendix A.

this new crowder-mediated pathway along with the protein's usual 3D search mode. Figure 3.2a shows the search time for different values of c_{on} by maintaining $r_d > c_{off}$. As shown in Figure 3.2b, the search becomes relatively faster in comparison to inert crowders for strong protein-crowder associations in the jumping regime due to major contributions from the new crowder-assisted search pathway along with the protein's usual relatively slow 3D search mode.

From the definition of scanning length and Eq. 3.4c, one can show that $p(s = 0) \to \lambda^2$ in the limit of $\lambda < 1$. The auxiliary functions $S_i(s = 0)$ in Eq. 3.3 in this limit can be approximated to $S_3(0) \approx \frac{\lambda^2}{1+x} + \frac{\lambda^2}{1+y}$ and $S_4(0) \approx (x + y)/2$. Using these simplifications, the average search time reduces to

$$T_0 \approx \frac{1 + k^c_{on} l_{ob}/r_d + \frac{k^c_{on}}{12u}\left[x(x+1)(x+2) + y(y+1)(y+2)\right]}{(1 + l_{ob})\left[\frac{k_{on}\lambda^2}{(1+x)(1+y)} + \frac{k^c_{on}}{2}\right]}. \tag{3.7}$$

Equation 3.7 shows that the protein can participate in the search process via the new crowder-assisted search pathway, in addition to the 3D search. For such short scanning lengths in this regime, the search times are effectively independent of λ. As c_{on} increases such that $c_{on} > k_{on}$, the effective rate k^c_{on} will be high, and the protein will eventually adopt the crowder-mediated pathway exclusively. The average search time in the jumping regime further reduces to

$$T_0 \cong \frac{2}{1+l_{ob}}\left[\frac{l_{ob}}{r_d}+\frac{x(x+1)(x+2)+y(y+1)(y+2)}{12u}\right] \tag{3.8}$$

which is independent of the scanning length and the effective rate k_{on}^c.

In the second search phase, when $1 < \lambda < L - l_{ob}$, in addition to 3D associations/dissociations, the protein can also slide along the DNA chain in a 1D random walk motion to find the target. As the scanning length increases, the search process will be delayed due to redundant visits to the same binding sites by the protein during its 1D random walk. As shown in Figure 3.2, the search dynamics is accelerated in the presence of intermediate and strong associations between the protein and the crowders in comparison with the search in the absence of crowders or in the presence of inert crowders. This is because the crowder-assisted search pathway provides an additional channel to the protein to access the DNA sites that are otherwise blocked by the crowders, thereby allowing it to scan the DNA chain more efficiently and recognize the target site. In the sliding regime, $p(s=0) \approx (1-\lambda^{-1})$ and the functions $S_i(s=0)$ approximately become $S_3(0) \approx \frac{\lambda^2}{\lambda+x}+\frac{\lambda^2}{\lambda+y}$ and $S_4(0) \approx \frac{1}{2}\left[\frac{x(x+1)}{\lambda+x}+\frac{y(y+1)}{\lambda+y}\right]$ provided that the target is not at the chain ends, which is true for the current case. The second term in Eq. 3.3 will have a major contribution, and the average target search time scales as

$$T_0 \sim \frac{1}{u} \times \frac{\lambda^2 k_{on}(L-l_{ob})+k_{on}^c l_{ob}}{2k_{on}\lambda + k_{on}^c\left\{l_{ob}-\frac{1}{2}\left[\frac{x(x+1)}{\lambda+x}+\frac{y(y+1)}{\lambda+y}\right]\right\}} \tag{3.9a}$$

Equation 3.9a shows that the search time increases with the scanning length due to the sliding of the protein on the DNA in this dynamic regime. Since this study assumes $r_d > c_{off}$ and for strong protein-crowder associations $k_{on}^c > k_{on}$, the average search time can be further approximated to

$$T_0 \sim \frac{1}{u}\frac{\lambda^2 k_{on}(L-l_{ob})/k_{on}^c + l_{ob}}{l_{ob}-\frac{1}{2}\left[\frac{x(x+1)}{\lambda+x}+\frac{y(y+1)}{\lambda+y}\right]}. \tag{3.9b}$$

For a given value of λ in the sliding regime, the search time decreases with an increase in the protein-crowder association rates (see Figure 3.2a). The crowder-assisted pathway accelerates the search process due to frequent protein-crowder associations, which decreases the search time in this sliding regime.

Finally, when $\lambda > L - l_{ob}$, the protein will find its target mainly by performing a simple 1D random walk on the DNA. The protein molecule will redundantly visit the same

sites in a single search cycle, eventually making the search time high and independent of the scanning length λ. Along with this simple 1D random walk, the search can also be accomplished via the crowder-mediated pathway. As seen in Figure 3.2b, this alternate search pathway may lead to a faster search compared to the case of no crowders in the bulk solution, provided both the conditions $c_{on} > k_{on}$ and $r_d > c_{off}$ hold. In this random walk regime, the functions $S_i(s = 0)$ in Eq. 3.3 can be expressed as $S_3(0) \approx a - \frac{a(1+3a+2a^2+6ax)}{6\lambda^2} + b - \frac{b(1+3b+2b^2+6by)}{6\lambda^2}$ and $S_4(0) \approx \frac{ax(1+x)}{2\lambda^2} + \frac{by(1+y)}{2\lambda^2}$. Consequently, the average search time in this 1D regime can be expressed approximately as

$$T_0 \cong \frac{k_{on}\big[a+b+3(a^2+b^2)+2(a^3+b^3)+6(a^2x+b^2y)\big]}{6u\big[k_{on}(L-l_{ob})+k^c_{on}l_ob\big]} + \frac{3k^c_{on}\big[ax(1+x)+by(1+y)\big]}{6u\big[k_{on}(L-l_{ob})+k^c_{on}l_ob\big]}. \tag{3.10}$$

This time is independent of the scanning length. Since the target site is assumed to be situated in the middle of the DNA chain, Eq. 3.10 can further be simplified to

$$T_0 \sim \frac{L}{12u} \times \frac{k_{on}L^2 + 3k^c_{on}\big[(l_{ob}-x)^2+x^2\big]}{k_{on}(L-l_{ob})+k^c_{on}l_ob}. \tag{3.11}$$

Thus, one can conclude that the protein-crowder associations modify the target search depending on how easily the crowder-assisted pathway allows the protein to access the blocked DNA binding sites. Stronger protein-crowder binding rates may accelerate the search provided the protein makes a smooth transition from the bulk crowders to the blocked DNA binding sites.

3.3.2 Effect of the Position of the Crowded Region around the Target

The previous results of the target search in the presence of inert bulk crowders that block the target site have shown that the average search time is higher in the 3D search regime compared to the case when there are no crowders.[48] This is because the protein only performs a 1D random walk to reach the target after it enters into the crowded region from either side of the DNA chain. The search time was found to be higher for the case of symmetrical crowders compared to unsymmetrical crowding around the target site. In present study, the protein-interacting crowders allow the protein to access the blocked DNA sites through an additionally new crowder-mediated search pathway, thereby

reducing the search times (see Figure 3.2). This acceleration of the search is observed mostly at lower scanning lengths. As shown in Figure 3.3a, the fastest dynamics is found when the bulk crowding is symmetric to the location of the target. For high values of c_{on}, the crowder-assisted search pathway predominates, and one can approximate it as the only search mode adopted by the protein. Effectively, one can then consider that the region on the DNA chain that is scanned by the protein is not the entire length L but the crowded region l_{ob}. Based on symmetry arguments proposed by Kolomeisky and co-workers,[19] the search time will be least when the target position is symmetric around the crowding.

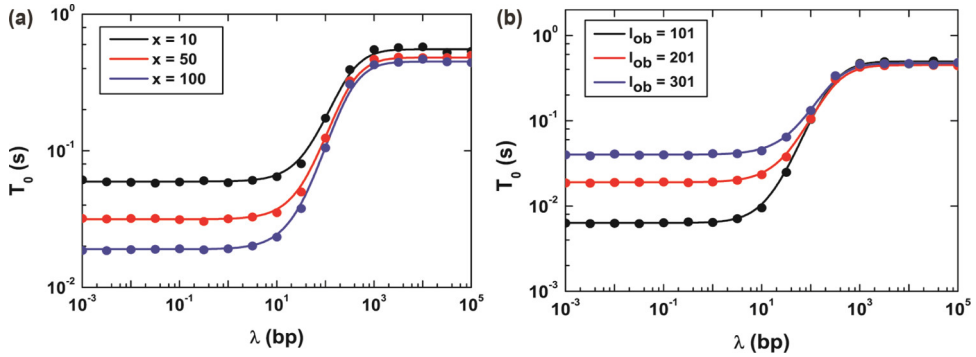


Figure 3.3: Effect of (a) position and (b) size of the crowded region on the target search time as a function of the scanning length . For (a), the position of crowding is changed by varying the parameter x (see Figure 3.1) for fixed $l_{ob} = 201$, and for (b), the crowding is kept symmetric. Other parameters chosen are as follows: $L = 1001$, $m = 501$, $u = 10^5$ s^{-1} and $k_{on} = 10^2$ s^{-1} per each site, $c_{on} = 10^3$ s^{-1}, $c_{off} = 1$ s^{-1} and $r_d = 10^3$ s^{-1}. The solid curves are from the analytical expression in eq 3, and the symbols correspond to Monte Carlo computer simulations. The details of the simulations are presented in the Appendix A.

3.3.3 Effect of the Size of the Crowded Region

Although strong protein-crowder associations introduce an additional crowder-assisted pathway that accelerates the target search, the process is delayed with an increase in the size of the crowded region, especially in the jumping regime. This is because the search via this crowder-mediated pathway involves extra steps before the protein associates with the DNA in the crowded region compared to the case where the protein can simply access the free DNA sites via 3D and/or 1D motion. As shown in Figure 3.3b, such direct protein-DNA associations are hindered, especially in the jumping search regime that delays the search with an increase in the size of the crowded region.

3.4 Comparisons with Experiments and Computational Studies

It has been shown experimentally that DNA translocation by the glycosylase enzyme, which is considered to be a DNA repair enzyme, was accomplished much more easily in human cells (in vivo) as compared to studies where supposedly "inert" PEG molecules were used as macromolecular crowders (in vitro).[38, 56] This observation is supported by this theory, which shows that the search is relatively faster under in vivo conditions because stronger enzyme-crowder associations may exist that cause the enzyme to recognize the damaged sites relatively faster via the predicted crowder-assisted search pathway. Alternatively, the slow search process in the presence of PEG may be accounted for by assuming weak associations between the searcher enzyme and the macromolecular crowders. This also ascertains the fact that PEG, which was mostly considered to be "inert", actually exhibits weak nonspecific interactions with the protein. This has been established via other experimental studies as well.[50, 51, 53] A separate experimental study showed that the hydrolysis of double-stranded DNA by endonucleases was enhanced in PEG macromolecular crowded environment, while the activity of the exonucleases remained unaffected.[39] It was suspected that a difference in the reaction mechanism may account for the different effects of molecular crowding on endonucleases and exonucleases. This theory predicts that the difference in activity may be due to varying strengths of associations of the PEG crowders with the endonucleases and the exonucleases that ultimately affect the ease of hydrolysis.

There exist a few computational studies that incorporate nonspecific interactions between the protein and the bulk crowders while studying the target search process. Nucleosomal crowding via CGMD simulations was shown to facilitate the protein-DNA binding,[61] which can be easily explained through crowder-mediated search pathway proposed in this work. The protein could easily bind to the crowded chromatin due to their strong associations and accelerate the search process, which is not possible in an inert crowded environment. By introducing attractive protein-crowder interactions in the protein's target search system, extensive Langevin dynamics simulations predicted the existence of apparent one-dimensional floating of the protein on the crowding above the DNA, which enhances the search at optimum attractive strengths.[60] In this chapter, a new crowder-assisted search pathway is also proposed that facilitates the target search process depending on the protein's association with the bulk crowders. High protein-crowder association

rates predominantly enhance the 3D search pathway, leading to an overall faster search that was also observed in the simulation study.

In living E. coli, the lac repressor binds to the chromosomal lac operator. In a single-molecule study, the transcription factor dynamics of the lac repressor were studied under cellular conditions under the effect of varying inducer IPTG (isopropyl β-D-1-thiogalactopyranoside) concentrations.[16] With the measurements of diffusion constants and other rate parameters such as the association rate, the dissociation rate, the association rate of the repressor to the macromolecular crowder IPTG, and the crowder dissociation rate, the lower and upper bounds of the search time were determined experimentally. Using the presented theoretical model, one can use Eq. 3.3 and determine the rate of formation of the LacI-operator complex from the IPTG-bound LacI that may satisfy these upper and lower limits of the search time for a given length of the crowded region over the target site. The values of r_d predicted from this theory may be checked experimentally to further validate this work.

3.5 Conclusions

To summarize, this chapter presentEd a theoretical investigation on the role of attractive interactions between the protein and the macromolecular bulk crowders that lie in the close vicinity of the DNA in protein's search for its target on DNA. By constructing a discrete-state stochastic model for the underlying processes, the dynamics of the system is explicitly analyzed using a method of first-passage probability densities. An exact analytical expression for the average target search time is obtained and theoretical calculations are also supported by Monte Carlo computer simulations. This theory suggests that protein-crowder associations introduce a new crowder-mediated search pathway for the protein molecule where it is capable of indirectly accessing the blocked DNA binding sites via the formation of protein-crowder complexes. A smooth transition of the protein from the crowder-bound complex to a DNA-bound one, without any dissociation into the bulk, ensures the ease of the adoption of this newly proposed search pathway assisted by the bulk crowders. Obtained results show that these associations play an important role in modifying the protein's search dynamics. The search is much faster in the presence of such attractive associations especially in the 3D search regime compared to an inert cellular crowded environment. In all the dynamic search regimes, the effect is stronger depending on the strength of the protein-crowder attractive associations (c_{on}). This work's

main idea is that the protein target search becomes faster because of the involvement of an additional channel for finding the target in the system. This study could explain the role of protein-crowder associations in enhancing the facilitated diffusion of protein in the presence of molecular crowding agents inside a crowded environment. This theoretical approach can capture some realistic features of protein-DNA dynamics in living cells where attractive interactions are found to exist between the protein and molecular bulk crowders. By incorporating such models into experimental investigations, kinetic rate constants and other parameters can be determined.

While this theory provides a clear physical picture of the effect of protein-crowder associations in the search of the protein for targets on DNA, one should note that the model is rather simplified and many realistic processes are not taken into account. This work did not consider the possibility of the sliding motion of the protein molecule on the bulk crowders or any reversible transitions between protein-crowder and protein-DNA complexes within the crowded region. It will be critically important to address these issues in more advanced theoretical studies. Additionally, this study concentrated only on those crowders that lie in the bulk solution and any type of DNA-bound crowders was not considered. Thus, the role of "microhops" in the search process is not explored in this work.[12, 78] The bulk solution is extremely complex and the presented model adopts spatial resolution only to the 1D lattice of the DNA. Studies have shown that proteins are inhomogenously distributed within the bulk depending on the local state of the DNA and the target region.[82] Furthermore, due to the spatial inhomogeneity, the initial position of the protein with respect to their targets tend to affect the search process.[81, 83, 84] Despite these experimental and theoretical pieces of evidence, this work assumed the cellular matrix to be uniform representing a single discrete state for the system. While this assumption may not be accurate physically, it helps us to simplify this theoretical model and obtain analytical results, which are in good qualitative agreement with previously obtained experimental and computational observations. This theory successfully captures the essence of in vivo search dynamics of the protein by considering attractive interactions between the protein and the bulk crowders.

Bibliography

(1) Alberts, B. et al., *Molecular Biology of the Cell*; Garland Science, Taylor and Francis Group: 2014.

(2) Lodish, H. F. et al., *Molecular Cell Biology*; WH Freeman New York: 2006.

(3) Phillips, R. et al., *Physical Biology of the Cell*; Garland Science: 2012.

(4) Riggs, A. D. et al. The Lac Represser-Operator Interaction: III. Kinetic Studies. *J. Mol. Biol.* **1970**, *53*, 401–417.

(5) Mirny, L. et al. How a Protein Searches for its Site on DNA: The Mechanism of Facilitated Diffusion. *J. Phys. A: Math. Theor.* **2009**, *42*, 434013.

(6) Kolomeisky, A. B. Physics of Protein–DNA Interactions: Mechanisms of Facilitated Target Search. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2088–2095.

(7) Von Hippel, P. H.; Berg, O. G. Facilitated Target Location in Biological Systems. *J. Biol. Chem.* **1989**, *264*, 675–678.

(8) Marius Clore, G. Exploring Translocation of Proteins on DNA by NMR. *J. Biomol. NMR* **2011**, *51*, 209–219.

(9) Pollak, A. J. et al. Distinct Facilitated Diffusion Mechanisms by E. Coli Type II Restriction Endonucleases. *Biochem.* **2014**, *53*, 7028–7037.

(10) Iwahara, J. et al. NMR Structural and Kinetic Characterization of a Homeodomain Diffusing and Hopping on Nonspecific DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15062–15067.

(11) Winter, R. B.; Von Hippel, P. H. Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 2. The Escherichia Coli Lac Repressor-Operator Interaction: Equilibrium Measurements. *Biochem.* **1981**, *20*, 6948–6960.

(12) Berg, O. G. et al. Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 1. Models and Theory. *Biochem.* **1981**, *20*, 6929–6948.

(13) Winter, R. B. et al. Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 3. The Escherichia Coli Lac Repressor-Operator Interaction: Kinetic Measurements and Conclusions. *Biochem.* **1981**, *20*, 6961–6977.

(14) Gowers, D. M. et al. Measurement of the Contributions of 1D and 3D Pathways to the Translocation of a Protein Along DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15883–15888.

(15) Wang, Y. et al. Single Molecule Measurements of Repressor Protein 1D Diffusion on DNA. *Phys. Rev. Lett.* **2006**, *97*, 048302.

(16) Elf, J. et al. Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell. *Science* **2007**, *316*, 1191–1194.

(17) Gorman, J. et al. Visualizing One-Dimensional Diffusion of Eukaryotic DNA Repair Factors Along a Chromatin Lattice. *Nat. Struct. Mol. Biol.* **2010**, *17*, 932–938.

(18) Hammar, P. et al. The Lac Repressor Displays Facilitated Diffusion in Living Cells. *Science* **2012**, *336*, 1595–1598.

(19) Veksler, A.; Kolomeisky, A. B. Speed-Selectivity Paradox in the Protein Search for Targets on DNA: Is it Real or Not? *J. Phys. Chem. B* **2013**, *117*, 12695–12701.

(20) Coppey, M. et al. Kinetics of Target Site Localization of a Protein on DNA: A Stochastic Approach. *Biophys. J.* **2004**, *87*, 1640–1649.

(21) Lomholt, M. A. et al. Facilitated Diffusion with DNA Coiling. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 8204–8208.

(22) Ellis, R. J. Macromolecular Crowding: Obvious but Underappreciated. *Trends Biochem. Sci.* **2001**, *26*, 597–604.

(23) Li, G.-W. et al. Effects of Macromolecular Crowding and DNA Looping on Gene Regulation Kinetics. *Nat. Phys.* **2009**, *5*, 294–297.

(24) Cheung, M. S. et al. Molecular Crowding Enhances Native State Stability and Refolding Rates of Globular Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 4753–4758.

(25) Zimmerman, S. B.; Trach, S. O. Estimation of Macromolecule Concentrations and Excluded Volume Effects for the Cytoplasm of Escherichia Coli. *J. Mol. Biol.* **1991**, *222*, 599–620.

(26) Zimmerman, S. B.; Trach, S. O. Macromolecular Crowding Extends the Range of Conditions Under Which DNA Polymerase is Functional. *Biochim. Biophys. Acta, Gene Struct. Expression* **1988**, *949*, 297–304.

(27) Shvets, A. A.; Kolomeisky, A. B. Crowding on DNA in Protein Search for Targets. *J. Phys. Chem. Lett.* **2016**, *7*, 2502–2506.

(28) Shvets, A. et al. Role of Static and Dynamic Obstacles in the Protein Search for Targets on DNA. *J. Phys. Chem. B* **2016**, *120*, 5802–5809.

(29) Krepel, D.; Levy, Y. Protein Diffusion Along DNA: On the Effect of Roadblocks and Crowders. *J. Phys. A: Math. Theor.* **2016**, *49*, 494003.

(30) Brackley, C. et al. Intracellular Facilitated Diffusion: Searchers, Crowders, and Blockers. *Phys. Rev. Lett.* **2013**, *111*, 108101.

(31) Liu, L.; Luo, K. DNA-Binding Protein Searches for its Target: Non-Monotonic Dependence of the Search Time on the Density of Roadblocks Bound on the DNA Chain. *J. Chem. Phys.* **2015**, *142*.

(32) Murugan, R. Theory of Site-Specific DNA-Protein Interactions in the Presence of Nucleosome Roadblocks. *Biophys. J.* **2018**, *114*, 2516–2529.

(33) Felipe, C. et al. The Effect of Obstacles in Multi-Site Protein Target Search with DNA Looping. *J. Chem. Phys.* **2020**, *152*.

(34) Mahmutovic, A. et al. What Matters for Lac Repressor Search In Vivo — Sliding, Hopping, Intersegment Transfer, Crowding on DNA or Recognition? *Nucleic Acids Res.* **2015**, *43*, 3454–3464.

(35) Koslover, E. F. et al. Crowding and Hopping in a Protein's Diffusive Transport on DNA. *J. Phys. A: Math. Theor.* **2017**, *50*, 074005.

(36) Meroz, Y. et al. Facilitated Diffusion in a Crowded Environment: From Kinetics to Stochastics. *J. Phys. A: Math. Theor.* **2009**, *42*, 434012.

(37) Bauer, M. et al. Real Sequence Effects on the Search Dynamics of Transcription Factors on DNA. *Sci. Rep.* **2015**, *5*, 10072.

(38) Cravens, S. L. et al. Molecular Crowding Enhances Facilitated Diffusion of Two Human DNA Glycosylases. *Biophys. J.* **2015**, *108*, 76a.

(39) Sasaki, Y. et al. Regulation of DNA Nucleases by Molecular Crowding. *Nucleic Acids Res.* **2007**, *35*, 4086–4093.

(40) Dey, P.; Bhattacherjee, A. Mechanism of Facilitated Diffusion of DNA Repair Proteins in Crowded Environment: Case Study with Human Uracil DNA Glycosylase. *J. Phys. Chem. B* **2019**, *123*, 10354–10364.

(41) Krepel, D. et al. Mechanism of Facilitated Diffusion During a DNA Search in Crowded Environments. *J. Phys. Chem. B* **2016**, *120*, 11113–11122.

(42) Liu, L.; Luo, K. Molecular Crowding Effect on Dynamics of DNA-Binding Proteins Search for Their Targets. *J. Chem. Phys.* **2014**, *141*.

(43) Tabaka, M. et al. Quantitative Influence of Macromolecular Crowding on Gene Regulation Kinetics. *Nucleic Acids Res.* **2013**, *42*, 727–738.

(44) Mondal, A.; Bhattacherjee, A. Searching Target Sites on DNA by Proteins: Role of DNA Dynamics Under Confinement. *Nucleic Acids Res.* **2015**, *43*, 9176–9186.

(45) Dey, P.; Bhattacherjee, A. Role of Macromolecular Crowding on the Intracellular Diffusion of DNA Binding Proteins. *Sci. Rep.* **2018**, *8*, 844.

(46) Dey, P.; Bhattacherjee, A. Disparity in Anomalous Diffusion of Proteins Searching for Their Target DNA Sites in a Crowded Medium is Controlled by the Size, Shape and Mobility of Macromolecular Crowders. *Soft Matter* **2019**, *15*, 1960–1969.

(47) Ghosh, S. et al. First-Passage Processes on a Filamentous Track in a Dense Traffic: Optimizing Diffusive Search for a Target in Crowding Conditions. *J. Stat. Mech.: Theory Exp.* **2018**, *2018*, 123209.

(48) Mondal, K.; Chaudhury, S. A Theoretical Study of the Role of Bulk Crowders on Target Search Dynamics of DNA Binding Proteins. *J. Stat. Mech.: Theory Exp.* **2020**, *2020*, 093204.

(49) Punia, B.; Chaudhury, S. Theoretical Study of the Role of Dynamic Bulk Crowders in the Protein Search for Targets on DNA. *J. Stat. Mech.: Theory Exp.* **2021**, *2021*, 073502.

(50) Miklos, A. C. et al. Protein Crowding Tunes Protein Stability. *J. Am. Chem. Soc.* **2011**, *133*, 7116–7120.

(51) Miklos, A. C. et al. Volume Exclusion and Soft Interaction Effects on Protein Stability Under Crowded Conditions. *Biochem.* **2010**, *49*, 6984–6991.

(52) Feig, M. Virtual Issue on Protein Crowding and Stability. *J. Phys. Chem. B* **2021**, *125*, 10649–10651.

(53) Senske, M. et al. Protein Stabilization by Macromolecular Crowding Through Enthalpy Rather than Entropy. *J. Am. Chem. Soc.* **2014**, *136*, 9036–9041.

(54) Nakano, S.-i. et al. Effects of Molecular Crowding on the Structures, Interactions, and Functions of Nucleic Acids. *Chem. Rev.* **2014**, *114*, 2733–2758.

(55) Esadze, A.; Stivers, J. T. Facilitated Diffusion Mechanisms in DNA Base Excision Repair and Transcriptional Activation. *Chem. Rev.* **2018**, *118*, 11298–11323.

(56) Esadze, A. et al. Measurement of Nanoscale DNA Translocation by Uracil DNA Glycosylase in Human Cells. *Nucleic Acids Res.* **2017**, *45*, 12413–12424.

(57) Cravens, S. L. et al. Electrostatic Properties of Complexes Along a DNA Glycosylase Damage Search Pathway. *Biochem.* **2014**, *53*, 7680–7692.

(58) Cravens, S. L.; Stivers, J. T. Comparative Effects of Ions, Molecular Crowding, and Bulk DNA on the Damage Search Mechanisms of hOGG1 and hUNG. *Biochem.* **2016**, *55*, 5230–5242.

(59) D'augustin, O. et al. Lost in the Crowd: How Does Human 8-Oxoguanine DNA Glycosylase 1 (OGG1) Find 8-Oxoguanine in the Genome? *Int. J. Mol. Sci.* **2020**, *21*, 8360.

(60) Dey, P.; Bhattacherjee, A. Structural Basis of Enhanced Facilitated Diffusion of DNA-Binding Protein in Crowded Cellular Milieu. *Biophys. J.* **2020**, *118*, 505–517.

(61) Kanada, R. et al. Nucleosome Crowding in Chromatin Slows the Diffusion but can Promote Target Search of Proteins. *Biophys. J.* **2019**, *116*, 2285–2295.

(62) Ma, J. et al. Strong Intracellular Signal Inactivation Produces Sharper and More Robust Signaling from Cell Membrane to Nucleus. *PLoS Comput. Biol.* **2020**, *16*, e1008356.

(63) Rosen, J. et al. Modest Protein-Crowder Attractive Interactions can Counteract Enhancement of Protein Association by Intermolecular Excluded Volume Interactions. *J. Phys. Chem. B* **2011**, *115*, 2683–2689.

(64) Bhattacharya, A. et al. Protein–Protein Interactions in a Crowded Environment. *Biophys. Rev.* **2013**, *5*, 99–108.

(65) Kim, Y. C.; Mittal, J. Crowding Induced Entropy-Enthalpy Compensation in Protein Association Equilibria. *Phys. Rev. Lett.* **2013**, *110*, 208102.

(66) Harada, R. et al. Reduced Native State Stability in Crowded Cellular Environment Due to Protein–Protein Interactions. *J. Am. Chem. Soc.* **2013**, *135*, 3696–3701.

(67) Shvets, A. A.; Kolomeisky, A. B. Sequence Heterogeneity Accelerates Protein Search for Targets on DNA. *J. Chem. Phys.* **2015**, *143*.

(68) Lange, M. et al. Dynamics of the Protein Search for Targets on DNA in the Presence of Traps. *J. Phys. Chem. B* **2015**, *119*, 12410–12416.

(69) Mondal, A.; Bhattacherjee, A. Understanding the Role of DNA Topology in Target Search Dynamics of Proteins. *J. Phys. Chem. B* **2017**, *121*, 9372–9381.

(70) Kochugaeva, M. P. et al. Optimal Length of Conformational Transition Region in Protein Search for Targets on DNA. *J. Phys. Chem. Lett.* **2017**, *8*, 4049–4054.

(71) Shin, J.; Kolomeisky, A. B. Surface-Assisted Dynamic Search Processes. *J. Phys. Chem. B* **2018**, *122*, 2243–2250.

(72) Mondal, K.; Chaudhury, S. Effect of DNA Conformation on the Protein Search for Targets on DNA: A Theoretical Perspective. *J. Phys. Chem. B* **2020**, *124*, 3518–3526.

(73) Felipe, C. et al. DNA Looping and DNA Conformational Fluctuations can Accelerate Protein Target Search. *J. Phys. Chem. B* **2021**, *125*, 1727–1734.

(74) Shvets, A. A.; Kolomeisky, A. B. The Role of DNA Looping in the Search for Specific Targets on DNA by Multisite Proteins. *J. Phys. Chem. Lett.* **2016**, *7*, 5022–5027.

(75) Thompson, M. A. et al. Three-Dimensional Tracking of Single mRNA Particles in Saccharomyces Cerevisiae using a Double-Helix Point Spread Function. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 17864–17871.

(76) Siggers, T.; Gordân, R. Protein–DNA Binding: Complexities and Multi-Protein Codes. *Nucleic Acids Res.* **2014**, *42*, 2099–2111.

(77) Tarunina, M.; Jenkins, J. Human p53 Binds DNA as a Protein Homodimer but Monomeric Variants Retain Full Transcription Transactivation Activity. *Oncogene* **1993**, *8*, 3165–3173.

(78) De la Rosa, M. A. D. et al. Dynamic Strategies for Target-Site Search by DNA-Binding Proteins. *Biophys. J.* **2010**, *98*, 2943–2953.

(79) Morelli, M. J. et al. Effects of Macromolecular Crowding on Genetic Networks. *Biophys. J.* **2011**, *101*, 2882–2891.

(80) Grebenkov, D. S. et al. Strong Defocusing of Molecular Reaction Times Results from an Interplay of Geometry and Reaction Control. *Commun. Chem.* **2018**, *1*, 96.

(81) Pulkkinen, O.; Metzler, R. Distance Matters: The Impact of Gene Proximity in Bacterial Gene Regulation. *Phys. Rev. Lett.* **2013**, *110*, 198101.

(82) Kuhlman, T. E.; Cox, E. C. Gene Location and DNA Density Determine Transcription Factor Distributions in Escherichia Coli. *Mol. Syst. Biol.* **2012**, *8*, 610.

(83) Képes, F. Periodic Transcriptional Organization of the E. Coli Genome. *J. Mol. Biol.* **2004**, *340*, 957–964.

(84) Kolesov, G. et al. How Gene Order is Influenced by the Biophysics of Transcription Regulation. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 13948–13953.

Chapter 4

Theoretical Insights into the Full Description of DNA Target Search by Subdiffusing Proteins

The content of this chapter and associated content in Appendix B are reproduced (adapted) from Ref. *Phys. Chem. Chem. Phys.*, 2022, 24, 29074 with permission from the Royal Society of Chemistry.

4.1 Introduction

An encounter of DNA-binding proteins (DBPs) and their respective target binding sites on the DNA followed by a successful reaction initiates several important biological processes such as gene regulation, transcription, damage repair, etc.[1, 2] The findings of in vitro experiments on lac repressor protein's target search by Riggs et al.[3] paved the way for numerous efforts to comprehend the underlying mechanism of the search process. The target site recognition by DBPs happens via the widely accepted facilitated diffusion mechanism which comprises a balance of 1D and 3D diffusion-based search modes of proteins.[4–6] The facilitated diffusion of DBPs is successfully captured in vitro through bulk measurements[7–10] and in vivo via single-molecule experiments.[11–15]

Theoretically, such target search studies at the single-molecule level are preferentially analyzed through the first passage time distributions (FPTDs) of a DBP to recognize its target site for the first time, provided its search started from some fixed initial position.[16, 17] An important quantity of interest has been the average of FPTD called the mean first passage time (MFPT) which is often represented as a function of the mean reaction rate constant owing to the assumption that the possible initial positions of a DBP are uniformly distributed within the cytoplasm.[18–20] However, several examples demonstrate a high local concentration of proteins near their promoter sites which facilitates the search.[21–24] The geometry of a nucleus, its components, and underlying properties may collectively affect the search dynamics of transcription factors (TFs).[25, 26] Inspired by these evidences, important insights provided through theoretical techniques, especially the MFPT, show that the knowledge of the initial position of DBPs is vital for a deeper understanding of the search process.[27–29] The FPTD and its mean depend on the geometry of the system as well as the size of the target region.[30–33]

Interestingly, the mean of the FPTD is a manifestation of the most-probable events only if the DBP starts its search from a large distance away from the target.[34] As this initial distance grows smaller, the standard deviation about the mean value of the FPTD becomes higher. Based on this effect of distance, the searching DBP may adopt two different kinds of trajectories.[35] Firstly, all the paths that lead the protein straight to its target site fall into the category of "direct trajectories". During this search process, the DBP does not interact with the confining boundary of the system and thus, such trajectories are important when the DBPs lie close to the target site. For the second kind of search path called "indirect trajectories", the DBPs interact with the outer boundary of

the system during their search for the target. These trajectories tend to explore the entire system before the search is complete. A proper inspection of the complete FPTDs revealed that there lies a vast time-scale separation between these two different trajectories, thereby highlighting the importance of proximity in the search process.[35] Full FPTDs reveal that the MFPT is controlled by the distribution at long times and are a result of indirect trajectories. The MFPT masks the typical (most probable) FPTs that are observed in the distribution at much shorter times, being prominently governed by the direct trajectories.[36] By considering simple spherical and cylindrical geometries, qualitatively similar FPTDs were observed from which the most-probable time, as well as the MFPT, were analytically determined for appropriate quantification of search processes.[37–39]

In biological systems, the "by default" assumed normal Brownian diffusion motion of DBPs within the cytosol can deviate and lead to anomalous subdiffusion.[40–42] Anomalous subdiffusion of biomolecules is generally either a consequence of viscoelastic properties of the cytoplasm or due to a heavily crowded cellular environment, but it can have other microscopic origins as well.[43–45] Therefore, incorporating the subdiffusive nature of DBPs within the cytoplasm, in theoretical and simulation studies, would allow for a general description of their search dynamics. The mean-squared displacement (MSD) for a particle generally scales with time as $\langle x \rangle \sim t^{\alpha}$, where α is an anomaly exponent and lies within the range $0 < \alpha < 1$ for subdiffusion and $\alpha = 1$ corresponds to normal diffusion. In the context of the initial position of the DBPs from their targets, there are only a few works that focus on the target search process of subdiffusing DBPs. Through single particle tracking experiments, it was observed that diffusion properties of TFs, namely c-Myc and P-TEFb, cause them to adopt separate searching strategies within the cell nucleus, which is also governed by the initial distance.[46] Single-molecule studies on the DNA repair heterodimer Rad4-Rad23 showed that the subdiffusing protein displays a constrained "recognition-at-a-distance" mechanism for an efficient repair.[47] Theoretical models supported by simulations show that search dynamics of TFs near the nucleosome roadblocks on the DNA lattice displays subdiffusion which enhances the search process.[48] Thus, the subdiffusive motion of TFs allows them to stay near the DNA which may be helpful for in vivo gene regulation.[49]

In all these experimental and simulation studies, the focus has been on determining the MFPT but obtaining an analytical expression of the same is quite challenging. The FPTDs of subdiffusing particles show fat tails at large times that ultimately cause the MFPTs to diverge and thus it becomes difficult to obtain analytical solutions.[50] However,

it is expected that an entire description of FPTDs might provide analytical expressions of timescales that are biologically relevant. Such a theoretical picture will also provide a general description of FPTDs and the underlying information about the contributions from the direct and indirect trajectories. In this direction, this chapter aims to find a general FPTD of the target search process by DBPs and arrive at results that are universally applicable.

4.2 Theoretical Model

This work describes the search process in cylindrical domains because it is a preferred geometry in computer simulations to accommodate elongated structures of DNAs.[51–53] In in vitro experimental setups also such geometry has been considered due to long strands of DNAs.[54, 55] Based on these observations, a cylindrical geometry is opted, in this study, to probe the diffusion of DBPs for realistic situations. Let us consider a single subdiffusing DBP within the domain $\Omega = \left\{(x, y, z) \in R^3 \colon \rho < r = \sqrt{x^2 + y^2} < R, 0 < z < L\right\}$ formed between two co-axial right circular cylinders of radii ρ and R with height L (Fig. 4.1). The inner cylinder represents an elongated DNA molecule while the outer cylinder is the boundary of cell/nuclei. The annular region $\Gamma = \left\{(x, y, z) \in R^3 \colon \sqrt{x^2 + y^2} = r = \rho, 0 < z < \epsilon\right\}$ serves as a target on the DNA where the DBP may arrive and either recognize the target by reacting with it or get reflected into the cytoplasm without a successful reaction. All boundaries of Ω are reflecting in nature except for the target region Γ which is partially absorbing. The search process ends when the DBP arrives and successfully reacts with the target region on the DNA.

For a target situated in the middle of the DNA, one can half L and ϵ as a result of reflecting planes at the top and bottom of the cylinders, to represent one-half of the entire DNA. It should be noted that the current model is similar to the one studied previously[33, 37] with a major difference being the DBP is allowed to exhibit subdiffusion to mimic in vivo conditions. In this work, the cause of subdiffusion of DBPs is attributed to the viscoelastic cellular medium and the macromolecular crowders are not considered explicitly.[45, 56] Additionally, since the DNA and the DBP are charged species, there are electrostatic interactions that may facilitate the target search. Several studies have demonstrated that these ionic interactions allow for a smooth 1D sliding of the DBPs along the DNA through non-specific contacts.[57–59] They cause the DBPs to stay close to the DNA as observed through the electrostatic potential energy distributions.[60, 61]

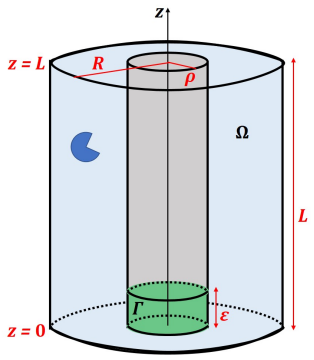


Figure 4.1: Schematic representation of a single DBP (blue shape) subdiffusing within the cellular medium in the domain Ω lying between two concentric cylinders of radii ρ (representing an elongated DNA) and R (outer boundary of the cell/nuclei). The cylinders are capped by reflecting planes at $z = 0$ and $z = L$. The target region, Γ, is marked in green which lies on the DNA having radius ρ and height ϵ.

However, it is argued that in cellular conditions, these ionic interactions do not affect the binding of DBP and the DNA owing to significant screening of charges due to the presence of the counter ions. The Debye length is roughly 1-2 nm in the cell which is less than the size of the target region of the DNA (at least 15-20 base pairs).[28, 62, 63] At such small Debye lengths, electrostatic effects will play a minimal role in binding the DBP to the DNA and will be relevant only after the collision between the two. Thus, even though the electrostatic interactions cause the DBP to slide along the DNA easily, they do not affect the binding of DBP and the DNA. It is also believed that crowded cellular conditions may cause such short-range screening to be significant due to reduced particle distances.[63, 64] In the interest of keeping the theoretical analysis simple while capturing the role of subdiffusing DBPs, this work does not consider the presence of electrostatic interactions in the system.

The earliest theories that were developed to model subdiffusion for any system promoted the use of fractional diffusion equations.[50, 65, 66] Thus, in cylindrical coordinates, the survival probability $S(r, z, \phi; t)$ of a subdiffusing DBP within the domain Ω with anomaly exponent α will satisfy the Fractional Fokker-Planck Equation (FFPE)

$$\frac{\partial S(r, z, \phi; t)}{\partial t} = D_{\alpha 0} D_t^{\alpha} \Delta S(r, z, \phi; t). \tag{4.1}$$

Here, ${}_0D_t^{\alpha}$ is the Riemann-Liouville fractional differential operator defined as

$${}_0D_t^{\alpha} g(t) = \frac{1}{\Gamma(\alpha)} \frac{\partial}{\partial t} \int_0^t dt' \frac{g(t')}{(t - t')^{1-\alpha}} \tag{4.2}$$

and the Laplacian operator is $\Delta = \partial_r^2 + r^{-1}\partial_r + \partial_z^2 + r^{-2}\partial_\phi^2$. It is worth mentioning that single molecule fluorescence measurements have shown that non-specifically bound DBPs tend to rotate around the helical path of the DNA while they slide on it, thereby, displaying a rotation-coupled 1D diffusion.[67] This phenomenon has been captured through other experiments and computational studies as well.[68, 69] Through coarse-grained and extensive atomistic MD simulations, it was shown that these rotational and translational motions may even be uncoupled.[70] Such observations in the sliding of DBPs and their consequent effects cannot be captured by this theory because of the axial symmetry (see Fig. 4.1) in the model. There is no dependency on the polar angle ϕ in the present model. Additionally, the axial coordinate z is replaced by $\theta = \pi z/L$ for further simplification of the notations.[33, 37] Due to subdiffusion of DBPs, the generalized diffusion coefficient D_α (units $\mathrm{m}^2\ \mathrm{s}^{-\alpha}$) is considered as

$$D_\alpha = \frac{k_B T}{m\gamma\Gamma(3-\alpha)} = \frac{D_1}{\Gamma(3-\alpha)} \tag{4.3}$$

with m as the mass of the DBP, γ as the friction coefficient, D_1 as the diffusion coefficient for normal diffusion ($\alpha = 1$), the temperature T and k_B being the Boltzmann constant. To proceed, it is convenient to perform a Laplace transform of Eq. 4.1 while applying the convolution property on Eq. 4.2 such that $\mathcal{L}\left[{}_0D_t^\alpha\ g(t)\right] = s^{1-\alpha}\tilde{g}(s)$ and one gets

$$\left(s - D_\alpha\ s^{1-\alpha}\right)\tilde{S}(r,\theta;s) = 1. \tag{4.4}$$

The main aim of this analysis is to solve Eq. 4.4 to obtain the survival probability, $\tilde{S}(r,\theta;s)$, using which the FPTD, $\tilde{F}(r,\theta;s)$, for a subdiffusing DBP to arrive at and react with the target region on the DNA can be determined. Due to reflecting outer boundaries of the system, $\left(\partial_r\tilde{S}(r,\theta;s)\right)_{r=R} = 0$ and $\left(\partial_\theta\tilde{S}(r,\theta;s)\right)_{\theta=0,\pi} = 0$. The mixed boundary condition on the inner cylinder, owing to the partially absorbing target region on the DNA is written as

$$D_\alpha\left(\frac{\partial\tilde{S}(r,\theta;s)}{\partial r}\right)_{r=\rho} = \begin{cases} \kappa\tilde{S}(r=\rho,\theta;s) & (0<\theta<\epsilon) \\ 0 & (\epsilon<\theta<\pi) \end{cases} \tag{4.5}$$

where κ (units $\mathrm{m\ s}^{-1}$) is the reactivity coefficient under the "radiation" boundary condition. This quantity represents the intrinsic reactivity of the target recognition (after the

DBP locates the target) governed by different processes that may hamper the possibility of perfect reactions. Some examples include conformation changes within the DNA or the DBP, structural incompatibility, presence of stochastically gated channels, etc.[33, 71, 72] For a perfectly reactive target, $\kappa = \infty$ while a non-reactive target will have $\kappa = 0$. Note that here the sub-diffusive behaviour of the DBP is not affecting its reactivity with the target site. Due to the mixed boundary condition in Eq. 4.5, it is difficult to obtain an exact solution for $\tilde{S}(r, \theta; s)$. However, one can adopt the procedure followed previously by introducing self-consistent approximation (SCA)[37, 73] which will allow for an approximate solution of $\tilde{S}(r, \theta; s)$ (see the details in Appendix B). Since FPTD is a negative time derivative of the survival probability,

$$\tilde{F}(r, \theta; s) = 1 - s\tilde{S}(r, \theta; s) \tag{4.6a}$$

$$= \frac{\tilde{\eta}(\rho; s)}{\tilde{b}_0(\rho; s)} \left[\tilde{b}_0(r; s) + 2\tilde{b}_0'(\rho; s) \sum_{n=1}^{n=\infty} \frac{\tilde{b}_n(r; s)}{\tilde{b}_n'(\rho; s)} \frac{\sin(n\epsilon)}{n\epsilon} \cos(n\theta) \right] \tag{4.6b}$$

where $\tilde{b}_n'(r; s)$ represents the first derivative with respect to r. The radial function $\tilde{b}_n(r; s)$ is a linear combination of modified Bessel functions of the first, $I_\nu(x)$, and the second kind, $K_\nu(x)$, such that

$$\tilde{b}_n(r; s) = T_0(rk_n)K_1(Rk_n) + K_0(rk_n)I_1(Rk_n) \tag{4.7a}$$

where

$$k_n = \sqrt{\frac{n^2\pi^2}{L^2} + \frac{s^\alpha}{D_\alpha}}. \tag{4.7b}$$

Additionally, the function $\tilde{\eta}(\rho; s)$ is defined by

$$\tilde{\eta}(\rho; s) = \left[1 - \left\{ \frac{D_\alpha}{\kappa} \frac{\pi}{\epsilon} + \tilde{\xi}(r = \rho; s) \right\} \frac{\tilde{b}_0'(\rho; s)}{\tilde{b}_0(\rho; s)} \right]^{-1} \tag{4.8a}$$

with

$$\tilde{\xi}(r; s) = -2 \sum_{n=1}^{n=\infty} \frac{\tilde{b}_n(r; s)}{\tilde{b}_n'(\rho; s)} \left(\frac{\sin(n\epsilon)}{n\epsilon} \right)^2. \tag{4.8b}$$

The FPTD in Eq. 4.6 is one of the main results of this work for understanding the effect of a single subdiffusing DBP on the target search process provided the search started from a fixed initial position. On comparing the obtained result in Eq. 4.6 for the FPTD,

$\tilde{F}(r,\theta;s)$, of a subdiffusing DBP with the previously determined result for a normally diffusing DBP (Eq. (7) in ref.[37]), it is observed that the current analytical result varies only by changes in the anomaly parameter α (existing as s^{α} in the resulting expressions in the Laplace space) such that $\alpha = 1$ reflects normally diffusing DBPs.[37] Thus, the subdiffusion behaviour offers a change in "time clocks"[74] with the variation of α and the qualitative behaviour usually remains unaffected. This model offers a general description for any type of diffusing particle searching for its target in cylindrical domains, whether in vivo or in vitro.

4.3 Results and Discussion

To determine the characteristic timescales of the direct and indirect trajectories, one can look into the distance-dependent FPTDs. In single-particle tracking experiments, the search dynamics of any DBP situated at a fixed distance from the target is probed.[46, 75] Various theoretical studies, as mentioned earlier, have also preferred to choose the initial distance between the DBP and the target site to study the search process.[31, 32, 35] In a similar spirit, this work monitors the search process when a subdiffusing DBP starts its search from a fixed radial distance, r, away from the target site (not the entire DNA). The resulting FPTD, $\tilde{F}(r;s)_{\epsilon}$, i.e., the distance-dependent FPTD is determined from $\tilde{F}(r,\theta;s)$ by

$$\tilde{F}(r;s)_{\epsilon} = \frac{1}{\epsilon}\int_0^{\epsilon} \tilde{F}(r,\theta;s)d\theta = \tilde{\eta}(\rho;s)\left(\frac{\tilde{b}_0(r;s)}{\tilde{b}_0(\rho;s)} - \tilde{\xi}(r;s)\frac{\tilde{b}_0'(\rho;s)}{\tilde{b}_0(\rho;s)}\right) \tag{4.9}$$

The FPTD determined in Eq. 4.9 considers the radial distance between the target region and subdiffusing DBP only. All other possible orientations that might result in the same initial distance r are not considered. Further, the FPTD in Eq. 4.9 is not the surface-averaged FPTD determined previously for a normally diffusing DBP as this study is not focusing on the distance between the DBP and the entire DNA.[37] The details for determining the surface-averaged FPTD are present in the Appendix B.

If the DBP starts its search from the surface of the target site, i.e., $r = \rho$, then Eq. 4.9 reduces to

$$\tilde{F}(r;s)_{\epsilon} = 1 + \frac{D_{\alpha}\pi}{\kappa\epsilon} \times \frac{\tilde{\eta}(\rho;s)\tilde{b}_0'(\rho;s)}{\tilde{b}_0(\rho;s)} \tag{4.10}$$

such that for perfectly reactive targets where $\kappa = \infty$, $\tilde{F}(r;s)_\epsilon = 1$ and the search is instantly completed. To validate the obtained analytical expression of $\tilde{F}(r;s)_\epsilon$ within SCA, numerical simulations have also been performed. Excellent agreements are obtained between Eq. 4.9 and corresponding numerically obtained values through simulations (see Fig. B.1 in Appendix B). Next, by performing the numerical inverse Laplace transform through Gaver-Stehfest method,[76] it is found that the distribution $F(r;t)_\epsilon$ spans over decades of orders in time (Fig. 4.2) irrespective of the anomaly exponent α and the qualitative behaviour is the same. This chapter now proceeds to determine the most-probable time and the MFPT of a single subdiffusing DBP from the resulting FPTD.

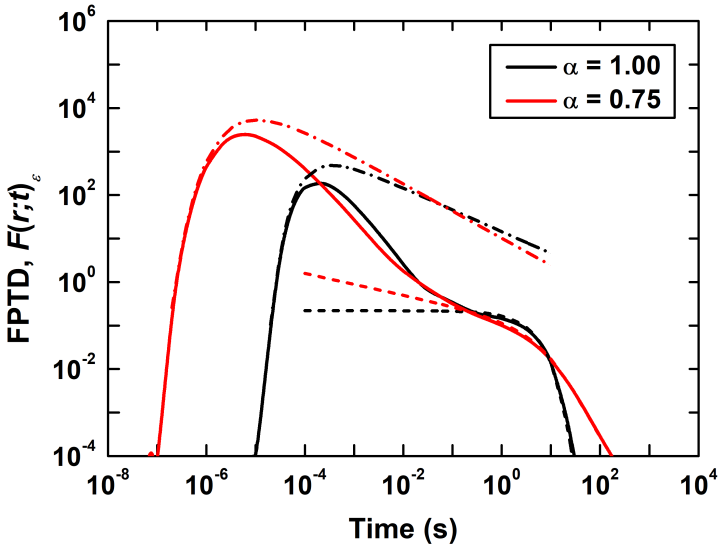


Figure 4.2: Distance dependent FPTD function, $F(r;t)_\epsilon$, (solid lines) with respect to time t determined for the anomaly exponents $\alpha = 0.75$ (red) and 1 (black). Dash-dotted lines correspond to the short-time asymptotic expression, $F(r;t \to 0)_\epsilon$ in Eq. 4.11a, and the dashed lines represent the long-time asymptotic expression for the FPTD, $F(r;t \to 0)_\epsilon$ in Eq. 4.14, for $\alpha = 0.75$ and 1. Parameters chosen are: $R = 1$ μm, $L = \pi$ μm, $D_1 = 10$ μm^2 s^{-1}, $\kappa = \infty$, $\rho = 2$ nm, $r = 0.1$ μm and $\epsilon = 0.2$ μm $\equiv 0.2$.

At shorter times, there lies a maximum that reflects the most-probable time, t_{mp}, of the distribution that results from direct trajectories of the subdiffusing DBP to the target. The time t_{mp} may be obtained by solving for the time when $\partial_t F(r;t)_\epsilon = 0$ but an exact analytical form of $F(r;t)_\epsilon$ is not known. Instead, as shown in Fig. 4.2, it is observed that there exists a good agreement of $F(r;t)_\epsilon$ with the short-time asymptotic expression when $t \to 0$ which includes the peak of the distribution as well. For perfect reactions ($\kappa = \infty$), the short-time expression reads as

$$F(r;t\to 0)_{\epsilon} \simeq \sqrt{\frac{\rho}{rt^2}}\Bigg[W_{-\frac{\alpha}{2},0}\Big(-\frac{r-\rho}{\sqrt{D_\alpha t^\alpha}}\Big)+\frac{(\pi-\epsilon)(r-\rho)\sqrt{D_\alpha t^\alpha}}{8\pi r\rho} \times W_{-\frac{\alpha}{2},\frac{\alpha}{2}}\Big(-\frac{r-\rho}{\sqrt{D_\alpha t^\alpha}}\Big)\Bigg] \tag{4.11a}$$

and for imperfect reactions ($\kappa<\infty$),

$$F(r;t\to 0)_{\epsilon} \simeq \sqrt{\frac{\rho\kappa^2 t^{\alpha-2}}{rD_\alpha}}W_{-\frac{\alpha}{2},\frac{\alpha}{2}}\Big(-\frac{r-\rho}{\sqrt{D_\alpha t^\alpha}}\Big) \tag{4.11b}$$

where $W_{\lambda,\mu}(x)=\sum_{n=0}^{n=\infty}x^n/\Big\{n!\Gamma(\mu+\lambda n)\Big\}$ describes the Wright function[77]: see Appendix B for more details. Thus, t_{mp} of the system can be obtained through the method of derivatives if one solves for the time when $\partial_t F(r;t\to 0)_{\epsilon}=0$. Finally, this time t_{mp} is given by

$$t_{mp}=\left[\frac{(r-\rho)\mu}{\sqrt{D_\alpha(1-\mu)^{1-\mu}}}\right]^{1/\mu} \tag{4.12}$$

where $\mu=\alpha/2$. Using the parameters defined in Fig. 4.2, one finds $t_{mp}(\alpha=0.75)=1.79\times10^{-5}$ s and $t_{mp}(\alpha=1)=4.80\times10^{-4}$ s. The underlying memory effects in the subdiffusive nature of a DBP allow it to scan the local volume properly due to a redundant motion causing the direct trajectories to be more frequent if the DBPs are subdiffusing. As a result, slower diffusion of DBPs due to lower α value decreases the t_{mp} (see Fig. B.2(a) in Appendix B). The size of the target region, ϵ, does not affect the t_{mp} which is also verified from the observation that the maximum of the FPTD in Fig. 4.3a shows no variation with changes in the target size. For a fixed α, the t_{mp} gradually increases with the initial distance between the DBP and the target, r, due to a delay in completing direct trajectories which is evident from the gradually vanishing maximum of the curve while it shifts towards higher times (Fig. 4.3(b) and Fig. B.2(b) in Appendix B). Additionally, from Eq. 4.12, the time t_{mp} is found to be independent of the dimensions of the boundary of the cell/nuclei, R and L, as well as the reactivity of the system, κ.

Now that the most-probable time of the search process is determined, let us proceed to find the MFPT. This quantity is analytically obtained from the FPTD, in Eq. 4.9, by evaluating $\partial_s\tilde{F}(r;s)_{s=0}=\tilde{S}(r;s=0)$. By performing Taylor expansion, using Eq. 4.6a and 4.9, one can analytically show that the MFPT is infinite: see Appendix B. The current model consists of a single subdiffusing DBP searching for its target on the DNA

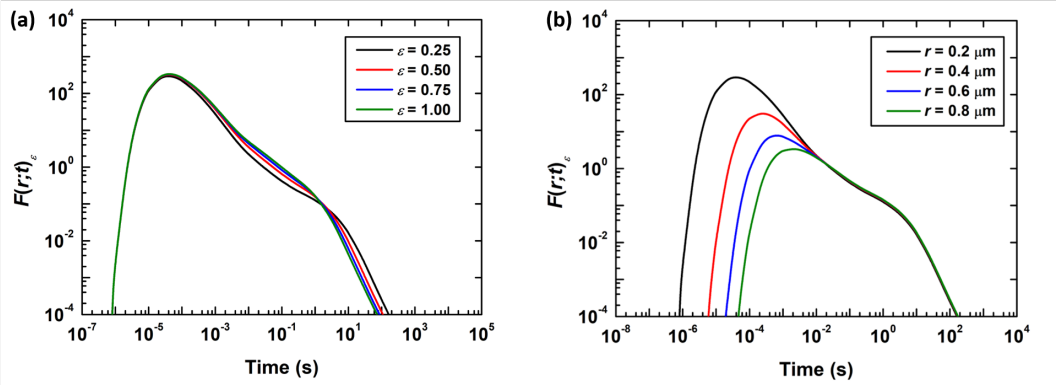


Figure 4.3: The function $F(r;t)_\epsilon$ determined by performing numerical inverse Laplace transform of Eq. 4.9 with respect to time for different (a) target sizes, ϵ and (b) initial position of the DBP, r. Parameters chosen are: $R = 1$ μm, $L = \pi$ μm, $D_1 = 10$ μm^2 s^{-1}, $\rho = 2$ nm and $\kappa = \infty$. For (a) $r = 0.1$ μm and (b) $\epsilon = 0.25$ μm $\equiv 0.25$.

and in such a scenario, the theoretical mean time always diverges due to the presence of long tails in the probability distribution.[74] These long tails are a result of certain rare trajectories that were delayed due to the subdiffusion of DBP. In the presence of multiple independently subdiffusing DBPs searching for the target, it is expected that the mean time would be a finite quantity.[74] Further, if one observes the search process from the experimental point where the observations are reported until some fixed time,[46, 75] then it is possible to obtain a finite mean time of a single subdiffusing DBP. In computational studies as well, a maximum time is chosen to ensure that enough trajectories are recorded that will properly capture the entire system dynamics due to a subdiffusing DBP.[43, 78, 79] Similarly, one can consider a fixed observation time, t_{max}, which will correspond to the maximum time for which the system is kept under observation in experiments. Analytically, this argument then gives rise to a conditional FPTD[43, 79] written as

$$F_c(r;t)_\epsilon = \frac{F(r;t)_\epsilon}{1 - \int_{t_{max}}^{\infty} F(r;t')_\epsilon} dt' \tag{4.13}$$

such that the denominator describes all those trajectories that are successful within the time t_{max}. Since an analytical expression of $F(r;t)_\epsilon$ is not available, the asymptotic expression of the same is utilised. In the limit $t \to \infty$,

$$F(r;t \to \infty)_\epsilon \cong \frac{A + B - C}{A^2} t^{\alpha-1} E_{\alpha,\alpha}(-t^\alpha/A) \tag{4.14}$$

such that the constants A, B, and C are

$$A = \left[\frac{D_\alpha \pi}{\kappa\epsilon} + \tilde{\xi}(\rho; s=0)\right] \frac{\left(R^2 - \rho^2\right)}{2\rho D_\alpha} \tag{4.15a}$$

$$B = \frac{R^2 \ln r/\rho}{2D_\alpha} \tag{4.15b}$$

$$C = \tilde{\xi}(r; s=0) \times \frac{\left(R^2 - \rho^2\right)}{2\rho D_\alpha} \tag{4.15c}$$

The details of obtaining these constants are presented in the Appendix B. $E_{p,q}(y)$ is the Mittag-Leffler function defined by $E_{p,q}(y) = \sum_{n=0}^{n=\infty} y^n/\Gamma(q+pn)$.[80] Eq. 4.14 shows an excellent agreement with numerically obtained $F(r;t)_\epsilon$ for times $t > 0.1$ s (see Fig. 4.2). This explains how the most-probable trajectories are masked in the evaluation of the mean times as they occur at much shorter times. The conditional MFPT, T_c, is then evaluated from Eqs. 4.13 and 4.14 such that one obtains

$$T_c = \int_0^{t_{max}} t' F_c(r; t' \to \infty)_\epsilon dt' = \frac{\int_0^{t_{max}} t' F(r; t' \to \infty)_\epsilon dt'}{1 - \int_{t_{max}}^{\infty} F(r; t')_\epsilon dt'}. \tag{4.16}$$

From the parameters defined in Fig. 4.2, it is found that $T_c(\alpha = 0.75) = 20.88$ s and $T_c(\alpha = 1) = 2.90$ s for an observation time of $t_{max} = 3600$ s (1 hour). The conditional time T_c decreases as the size of the target region increases which is in accordance with a previous study.[37] As the initial distance of the DBP from DNA grows, the search is slightly delayed because the DBP will require more time to come in close vicinity of the DNA to eventually recognize its target (Fig. B.3(a) in Appendix B). For any value of α, lying in the biologically relevant range of $0.6 < \alpha < 0.8$,[44] T_c increases with t_{max} because the proportion of successful trajectories that were delayed due to the redundant search gets included within a higher t_{max} value (Fig. B.3(b) in Appendix B). For the parameters chosen, the target search eventually becomes faster as $\alpha \to 1$ for a normally diffusing protein which, unlike subdiffusing DBPs, has no redundant motions and thus there are no characteristic fat tails observed in the FPTD. For an experimentally sound analysis of the target search of a subdiffusing DBP, it is important that the observation time is decided carefully such that the inclusion of rare trajectories won't lead to erroneous results.

From the FPTD determined in Eq. 4.9, the most-probable time, in Eq. 4.12, as well as the conditional MFPT, Eq. 4.16, are obtained for a single subdiffusing DBP searching for its target region from a known initial distance. Apart from determining these characteristic times, it is also observed in Fig. 4.2 that the FPTD becomes broader as the subdiffusive nature of the DBP increases (i.e., as α decreases) which is in accordance with

previous studies that explored the subdiffusive enzymatic reaction events[81] and exit-time distributions of subdiffusing particles.[71] In a subdiffusion process, as seen from Eq. 4.12, most of the target search events are complete within a very small time due to the underlying memory effects that lead to a successful local exploration without missing any area. This is evident from the obtained results in Fig. 4.2 as well where a relatively broader maximum is towards very small-time scales (which corresponds to the most-probable events) for the subdiffusing DBPs compared to the normally diffusing ones. Despite a relatively small most-probable time, the conditional MFPT of the subdiffusion process may be higher than that for the normally diffusing DBPs. This is due to some trajectories that are delayed as a result of redundant motions of the DBPs even though the probability of occurrence of these events is negligible when compared to most-probable events. These trajectories manifest as fat tails in the probability distributions for the subdiffusing particles which converge only at very large times.[50, 65] From Eq. 4.16, one can find that the conditional MFPT is dependent on the observation time (t_{max}) of the system in addition to other parameters such as the dimensions of the cell, target size, etc. On increasing the value of t_{max}, the contribution due to the trajectories present in the long tails of the distribution increases the average time (Fig. B.3(b) in Appendix B). Thus, the underlying memory effects of a subdiffusing DBP cause the most-probable time to decrease and the conditional MFPT to increase if the observation time is high enough.

4.4 Application to Real Systems

To illustrate the potential application of this theoretical approach, let us consider an experimental study where it was observed through single particle tracking experiments that distinct diffusion properties of TFs, namely c-Myc and P-TEFb, within the nucleus of U2OS cells cause them to adopt separate target searching strategies.[46] The transcription factor c-Myc searches for the target in a 'non-compact' way such that it explores the entire volume of the nucleus. On the other hand, P-TEFb recognises its target in a 'compact' manner by exploring a local volume such that the TF is bound to follow paths that guides it to nearby targets. It was shown that the elongation transcription factor P-TEFb displays subdiffusion within the nucleus. The MFPT was evaluated analytically after studying the diffusion properties of the TF experimentally.[46] As shown in Fig. 4.4(a), the mean squared displacement (MSD) data is fitted with the well-known equation for a particle diffusing in 3D,[50]

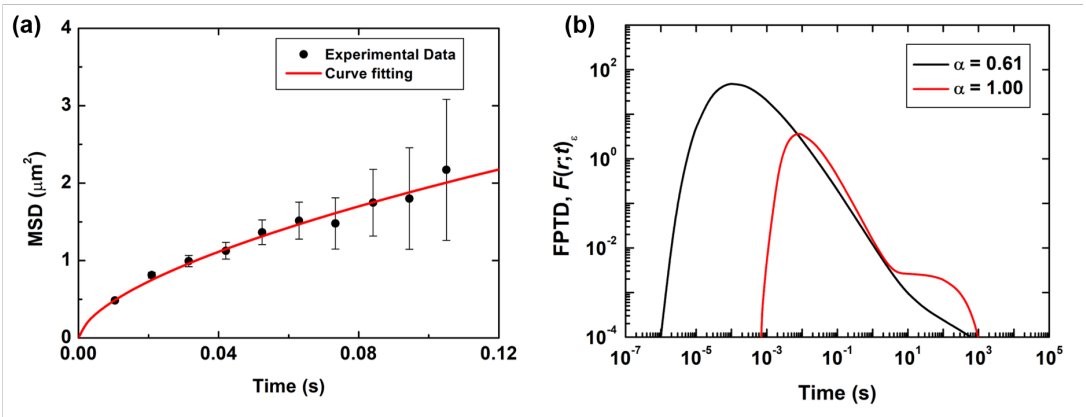


Figure 4.4: Theoretical analysis of the target search performed by the TF named P-TEFb within the nucleus of U2OS cells. (a) MSD data plotted against the time. The black symbols correspond to the experimental data (Fig. 3(c) of Ref.[46]) and the red line is the fitting curve obtained by using Eq. 4.17 to determine the parameters $\alpha = 0.61$ and $D_1 = 1.46\mu\text{m}^2\ \text{s}^{-1}$. (b) The distance-dependent FPTD, $F(r;t)_\epsilon$, determined by taking the numerical inverse Laplace transform of Eq. 4.9 for $\alpha = 0.61$ (black curve) compared with the hypothetical case of $\alpha = 1$ (red curve). As discussed in the Appendix B, the parameters considered are $R = 5\mu$m, $L = 3\mu$m, $D_1 = 1.46\mu\text{m}^2\ \text{s}^{-1}$, $r = 0.25\mu$m, $\rho = 2$ nm, $\epsilon = 0.5\mu\text{m} \equiv 0.52$ and $\kappa = \infty$.

$$\langle x^2 \rangle = \frac{6D_\alpha t^\alpha}{\Gamma(1+\alpha)} = \frac{6D_1 t^\alpha}{\Gamma(3-\alpha)\Gamma(1+\alpha)} \tag{4.17}$$

to determine the parameters $\alpha = 0.61$ and $D_1 = 1.46\mu\text{m}^2\ \text{s}^{-1}$. By using Eq. 4.9, one can determine the distance-dependent FPTD, $F(r;t)_\epsilon$, by performing a numerical inverse Laplace transform using Gaver-Stehfest method.[76] For a comparative analysis with a normally diffusing DBP, two values of the anomaly exponent are considered, $\alpha = 0.61$ and a hypothetical $\alpha = 1$. All the other parameters that are used to determine the FPTDs are detailed in the Appendix B.

The resulting FPTDs for these two situations are shown in Fig. 4.4(b) and it is observed (from the broad peak) that a majority of the events are completed under 1s for the subdiffusion process. For the chosen set of parameters, the most-probable times and the conditional MFPTs for both α values are determined in the Appendix B and it is observed that the time $t_{mp}(\alpha = 1) > t_{mp}(\alpha = 0.61)$ whereas the time $T_c(\alpha = 1) < T_c(\alpha = 0.61)$ as predicted from this theory.

4.5 Conclusions

The target search process of DBPs and their average search time has been studied extensively over the years. In this work, the target search process of a single DBP subjected to in vivo conditions is probed, where biomolecules display anomalous diffusion. The model applies to systems with cylindrical geometry which are common to protein-DNA target search studies. Since the heterogeneity of the cellular medium affects the search by modulating the initial position of the DBPs, the focus of this work was to understand the effect of subdiffusing DBPs searching for their targets situated at a known initial distance from the target site on the DNA.

From the full distribution of the FPTs obtained under the SCA, it is found that the characteristic timescales of the direct (most-probable time) and indirect (conditional mean time) trajectories are decades of order apart in time displaying the significance of the initial position of the DBP. The most-probable time lies at much shorter times compared to the MFPT if the initial position of the DBPs is close to the target region. The underlying memory effects in the subdiffusion process allow for a greater degree of redundancy in the motion of DBPs which leads to successful local exploration. This causes a majority of the events to be accomplished within a smaller time frame when compared with the search of normally diffusing proteins. Consequently, a broader maximum (which corresponds to the most-probable events) is observed in the FPTD of the subdiffusing proteins compared to the normally diffusing ones. However, the inherent redundancy causes certain rare events to complete in much larger times which results in long-tails in the FPTD leading to an increase in the conditional MFPT as the observation time increases.

This work also presented a straightforward application of obtained analytical results to single particle tracking experiments performed on the target search process of subdiffusing TFs. Despite the easy application of the results to experiments, there are certain assumptions associated with this work that restrict its direct quantitative application. To keep the calculations and results fairly simple, the DBP was assumed to be situated initially at a fixed radial distance from the target whereas other initial orientations with the same distance are also possible. The results are obtained under the assumption of axial symmetry and thus, the effect of rotation-coupled translational diffusion of the DBPs along the DNA is not considered. Electrostatic interactions, although present in cellular conditions, are ignored in this study and any consequent effects on the anomalous diffusion of DBPs are thus not observed. In this work, only a single DBP is searching for the target site

on the DNA which is not always true in biological systems as multiple copies of DBPs search for the target. Due to the consideration of a crowder-free cellular medium, the anomaly exponent remains at a constant value independent of the distance of the DBP from the target site. The presence of explicit crowders may help in proper quantification of the anomaly exponent depending on the varying density of the crowded regions in the cytoplasm.[22, 23, 44]

Irrespective of the assumptions involved, the obtained results are useful in understanding of the search for a single subdiffusing DBP from a fixed initial distance. The quantitative results can prove to be helpful in providing estimates related to the target search timescales of the process under observation. When $\alpha = 1$, the results for a normally diffusing particle are successfully reproduced.[37] The obtained analytical expressions are universally applicable to any kind of DBP with respect to their diffusion behaviour.

Bibliography

(1) Alberts, B. et al., *Molecular Biology of the Cell*; Garland Science, Taylor and Francis Group: 2014.

(2) Phillips, R. et al., *Physical Biology of the Cell*; Garland Science: 2012.

(3) Riggs, A. D. et al. The Lac Represser-Operator Interaction: III. Kinetic Studies. *J. Mol. Biol.* **1970**, *53*, 401–417.

(4) Mirny, L. et al. How a Protein Searches for its Site on DNA: The Mechanism of Facilitated Diffusion. *J. Phys. A: Math. Theor.* **2009**, *42*, 434013.

(5) Kolomeisky, A. B. Physics of Protein–DNA Interactions: Mechanisms of Facilitated Target Search. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2088–2095.

(6) Von Hippel, P. H.; Berg, O. G. Facilitated Target Location in Biological Systems. *J. Biol. Chem.* **1989**, *264*, 675–678.

(7) Winter, R. B.; Von Hippel, P. H. Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 2. The Escherichia Coli Lac Repressor-Operator Interaction: Equilibrium Measurements. *Biochem.* **1981**, *20*, 6948–6960.

(8) Marius Clore, G. Exploring Translocation of Proteins on DNA by NMR. *J. Biomol. NMR* **2011**, *51*, 209–219.

(9) Iwahara, J. et al. NMR Structural and Kinetic Characterization of a Homeodomain Diffusing and Hopping on Nonspecific DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15062–15067.

(10) Pollak, A. J. et al. Distinct Facilitated Diffusion Mechanisms by E. Coli Type II Restriction Endonucleases. *Biochem.* **2014**, *53*, 7028–7037.

(11) Yu, J. et al. Probing Gene Expression in Live Cells, One Protein Molecule at a Time. *Science* **2006**, *311*, 1600–1603.

(12) Elf, J. et al. Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell. *Science* **2007**, *316*, 1191–1194.

(13) Hammar, P. et al. The Lac Repressor Displays Facilitated Diffusion in Living Cells. *Science* **2012**, *336*, 1595–1598.

(14) Wang, Y. et al. Single Molecule Measurements of Repressor Protein 1D Diffusion on DNA. *Phys. Rev. Lett.* **2006**, *97*, 048302.

(15) Gorman, J. et al. Visualizing One-Dimensional Diffusion of Eukaryotic DNA Repair Factors Along a Chromatin Lattice. *Nat. Struct. Mol. Biol.* **2010**, *17*, 932–938.

(16) Redner, S., *A Guide to First-Passage Processes*; Cambridge university press: 2001.

(17) Metzler, R. et al., *First-Passage Phenomena and their Applications*; World Scientific: 2014; Vol. 35.

(18) Iwahara, J.; Kolomeisky, A. B. Discrete-State Stochastic Kinetic Models for Target DNA Search by Proteins: Theory and Experimental Applications. *Biophys. Chem.* **2021**, *269*, 106521.

(19) Tafvizi, A. et al. Dancing on DNA: Kinetic Aspects of Search Processes on DNA. *ChemPhysChem* **2011**, *12*, 1481–1489.

(20) Grigoriev, I. V. et al. Kinetics of Escape Through a Small Hole. *J. Chem. Phys.* **2002**, *116*, 9574–9577.

(21) Dröge, P.; Müller-Hill, B. High Local Protein Concentrations at Promoters: Strategies in Prokaryotic and Eukaryotic Cells. *Bioessays* **2001**, *23*, 179–183.

(22) Kolesov, G. et al. How Gene Order is Influenced by the Biophysics of Transcription Regulation. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 13948–13953.

(23) Kuhlman, T. E.; Cox, E. C. Gene Location and DNA Density Determine Transcription Factor Distributions in Escherichia Coli. *Mol. Syst. Biol.* **2012**, *8*, 610.

(24) Hansen, A. S. et al. Guided Nuclear Exploration Increases CTCF Target Search Efficiency. *Nat. Chem. Biol.* **2020**, *16*, 257–266.

(25) Woringer, M. et al. Geometry of the Nucleus: A Perspective on Gene Expression Regulation. *Curr. Opin. Chem. Biol.* **2014**, *20*, 112–119.

(26) Wunderlich, Z.; Mirny, L. A. Spatial Effects on the Speed and Reliability of Protein–DNA Search. *Nucleic Acids Res.* **2008**, *36*, 3570–3578.

(27) Pulkkinen, O.; Metzler, R. Distance Matters: The Impact of Gene Proximity in Bacterial Gene Regulation. *Phys. Rev. Lett.* **2013**, *110*, 198101.

(28) Kar, P. et al. Acceleration of Bursty Multiprotein Target Search Kinetics on DNA by Colocalisation. *Phys. Chem. Chem. Phys.* **2018**, *20*, 7931–7946.

(29) Condamin, S. et al. Random Walks and Brownian Motion: A Method of Computation for First-Passage Times and Related Quantities in Confined Geometries. *Phys. Rev. E* **2007**, *75*, 021111.

(30) Condamin, S. et al. First-Passage Times in Complex Scale-Invariant Media. *Nature* **2007**, *450*, 77–80.

(31) Bénichou, O. et al. Geometry-Controlled Kinetics. *Nat. Chem.* **2010**, *2*, 472–477.

(32) Meyer, B. et al. Universality Classes of First-Passage-Time Distribution in Confined Media. *Phys. Rev. E* **2011**, *83*, 051116.

(33) Grebenkov, D. S. et al. Effects of the Target Aspect Ratio and Intrinsic Reactivity Onto Diffusive Search in Bounded Domains. *New J. Phys.* **2017**, *19*, 103025.

(34) Mattos, T. G. et al. First Passages in Bounded Domains: When is the Mean First Passage Time Meaningful? *Phys. Rev. E* **2012**, *86*, 031143.

(35) Godec, A.; Metzler, R. Universal Proximity Effect in Target Search Kinetics in the Few-Encounter Limit. *Phys. Rev. X* **2016**, *6*, 041037.

(36) Godec, A.; Metzler, R. First Passage Time Distribution in Heterogeneity Controlled Kinetics: Going Beyond the Mean First Passage Time. *Sci. Rep.* **2016**, *6*, 20349.

(37) Grebenkov, D. S. et al. Towards a Full Quantitative Description of Single-Molecule Reaction Kinetics in Biological Cells. *Phys. Chem. Chem. Phys.* **2018**, *20*, 16393–16401.

(38) Grebenkov, D. S. et al. Strong Defocusing of Molecular Reaction Times Results from an Interplay of Geometry and Reaction Control. *Commun. Chem.* **2018**, *1*, 96.

(39) Grebenkov, D. S. et al. Full Distribution of First Exit Times in the Narrow Escape Problem. *New J. Phys.* **2019**, *21*, 122001.

(40) Golding, I.; Cox, E. C. Physical Nature of Bacterial Cytoplasm. *Phys. Rev. Lett.* **2006**, *96*, 098102.

(41) Park, S. et al. A Mini-Review of the Diffusion Dynamics of DNA-Binding Proteins: Experiments and Models. *J. Korean Phys. Soc.* **2021**, *78*, 408–426.

(42) English, B. P. et al. Single-Molecule Investigations of the Stringent Response Machinery in Living Bacterial Cells. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, E365–E373.

(43) Condamin, S. et al. Probing Microscopic Origins of Confined Subdiffusion by First-Passage Observables. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5675–5680.

(44) Weiss, M. et al. Anomalous Subdiffusion is a Measure for Cytoplasmic Crowding in Living Cells. *Biophys. J.* **2004**, *87*, 3518–3524.

(45) Wirtz, D. Particle-Tracking Microrheology of Living Cells: Principles and Applications. *Annu. Rev. Biophys.* **2009**, *38*, 301–326.

(46) Izeddin, I. et al. Single-Molecule Tracking in Live Cells Reveals Distinct Target-Search Strategies of Transcription Factors in the Nucleus. *Elife* **2014**, *3*, e02230.

(47) Kong, M. et al. Single-Molecule Imaging Reveals that Rad4 Employs a Dynamic DNA Damage Recognition Process. *Mol. Cell* **2016**, *64*, 376–387.

(48) Murugan, R. Theory of Site-Specific DNA-Protein Interactions in the Presence of Nucleosome Roadblocks. *Biophys. J.* **2018**, *114*, 2516–2529.

(49) Lomholt, M. A. et al. Subdiffusion and Weak Ergodicity Breaking in the Presence of a Reactive Boundary. *Phys. Rev. Lett.* **2007**, *98*, 200603.

(50) Metzler, R.; Klafter, J. The Restaurant at the End of the Random Walk: Recent Developments in the Description of Anomalous Transport by Fractional Dynamics. *J. Phys. A Math. Gen.* **2004**, *37*, R161.

(51) Liu, L. et al. Facilitated Diffusion of Transcription Factor Proteins With Anomalous Bulk Diffusion. *J. Phys. Chem. B* **2017**, *121*, 1284–1289.

(52) Krepel, D. et al. Mechanism of Facilitated Diffusion During a DNA Search in Crowded Environments. *J. Phys. Chem. B* **2016**, *120*, 11113–11122.

(53) Koslover, E. F. et al. Theoretical and Computational Modeling of Target-Site Search Kinetics In Vitro and In Vivo. *Biophys. J.* **2011**, *101*, 856–865.

(54) Esadze, A.; Stivers, J. T. Facilitated Diffusion Mechanisms in DNA Base Excision Repair and Transcriptional Activation. *Chem. Rev.* **2018**, *118*, 11298–11323.

(55) Monico, C. et al. Optical Methods to Study Protein-DNA Interactions In Vitro and in Living Cells at the Single-Molecule Level. *Int. J. Mol. Sci.* **2013**, *14*, 3961–3992.

(56) Lampo, T. J. et al. Cytoplasmic RNA-Protein Particles Exhibit Non-Gaussian Subdiffusive Behavior. *Biophys. J.* **2017**, *112*, 532–542.

(57) Dahirel, V. et al. Nonspecific DNA-Protein Interaction: Why Proteins can Diffuse Along DNA. *Phys. Rev. Lett.* **2009**, *102*, 228101.

(58) Ando, T.; Skolnick, J. Sliding of Proteins Non-Specifically Bound to DNA: Brownian Dynamics Studies with Coarse-Grained Protein and DNA Models. *PLoS Comput. Biol.* **2014**, *10*, e1003990.

(59) Marcovitz, A.; Levy, Y. Frustration in Protein–DNA Binding Influences Conformational Switching and Target Search Kinetics. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 17957–17962.

(60) Redding, S.; Greene, E. C. How do Proteins Locate Specific Targets in DNA? *Chem. Phys. Lett.* **2013**, *570*, 1–11.

(61) Cherstvy, A. et al. Protein-DNA Interactions: Reaching and Recognizing the Targets. *J. Phys. Chem. B* **2008**, *112*, 4741–4750.

(62) Das, R. K.; Kolomeisky, A. B. Facilitated Search of Proteins on DNA: Correlations are Important. *Phys. Chem. Chem. Phys.* **2010**, *12*, 2999–3004.

(63) Spitzer, J. J.; Poolman, B. Electrochemical Structure of the Crowded Cytoplasm. *Trends Biochem. Sci.* **2005**, *30*, 536–541.

(64) Wennerström, H. et al. Colloidal Stability of the Living Cell. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 10113–10121.

(65) Metzler, R.; Klafter, J. The Random Walk's Guide to Anomalous Diffusion: A Fractional Dynamics Approach. *Phys. Rep.* **2000**, *339*, 1–77.

(66) Barkai, E. et al. From Continuous Time Random Walks to the Fractional Fokker-Planck Equation. *Phys. Rev. E* **2000**, *61*, 132.

(67) Blainey, P. C. et al. Nonspecifically Bound Proteins Spin While Diffusing Along DNA. *Nat. Struct. Mol. Biol.* **2009**, *16*, 1224–1229.

(68) Lin, J. et al. TRF1 and TRF2 Use Different Mechanisms to Find Telomeric DNA but Share a Novel Mechanism to Search for Protein Partners at Telomeres. *Nucleic Acids Res.* **2014**, *42*, 2493–2504.

(69) Wieczór, M.; Czub, J. How Proteins Bind to DNA: Target Discrimination and Dynamic Sequence Search by the Telomeric Protein TRF1. *Nucleic Acids Res.* **2017**, *45*, 7643–7654.

(70) Bigman, L. S. et al. What are the Molecular Requirements for Protein Sliding Along DNA? *J. Phys. Chem. B* **2021**, *125*, 3119–3131.

(71) Grebenkov, D. S. Subdiffusion in a Bounded Domain with a Partially Absorbing-Reflecting Boundary. *Phys. Rev. E* **2010**, *81*, 021128.

(72) Bénichou, O. et al. Kinetics of Stochastically Gated Diffusion-Limited Reactions and Geometry of Random Walk Trajectories. *Phys. Rev. E* **2000**, *61*, 3388.

(73) Shoup, D. et al. Diffusion-Controlled Bimolecular Reaction Rates. The Effect of Rotational Diffusion and Orientation Constraints. *Biophys. J.* **1981**, *36*, 697–714.

(74) Grebenkov, D. S. Searching for Partially Reactive Sites: Analytical Results for Spherical Targets. *J. Chem. Phys.* **2010**, *132*.

(75) Normanno, D. et al. Probing the Target Search of DNA-Binding Proteins in Mammalian Cells using TetR as Model Searcher. *Nat. Commun.* **2015**, *6*, 7357.

(76) Stehfest, H. Algorithm 368: Numerical Inversion of Laplace Transforms [D5]. *Commun. ACM* **1970**, *13*, 47–49.

(77) Gorenflo, R. et al. Analytical Properties and Applications of the Wright Function. *Fract. Calc. Appl. Anal.* **1999**, *2*, 383–414.

(78) Guigas, G.; Weiss, M. Sampling the Cell with Anomalous Diffusion – The Discovery of Slowness. *Biophys. J.* **2008**, *94*, 90–94.

(79) Lua, R. C.; Grosberg, A. Y. First Passage Times and Asymmetry of DNA Translocation. *Phys. Rev. E* **2005**, *72*, 061918.

(80) Haubold, H. J. et al. Mittag-Leffler Functions and Their Applications. *J. Appl. Math.* **2011**, *2011*.

(81) Eaves, J. D.; Reichman, D. R. The Subdiffusive Targeting Problem. *J. Phys. Chem. B* **2008**, *112*, 4283–4289.

Chapter 5

Microscopic Mechanism of Macromolecular Crowder-Assisted DNA Capture and Translocation through Biological Nanopores

The content of this chapter and associated content in Appendix C are reprinted (adapted) with permission from (*J. Phys. Chem. B* 2023, 127, 26, 5850-5858). Copyright © 2023 American Chemical Society.

5.1 Introduction

The transport of bio-macromolecules across different membranes in cells is a vital life process.[1–3] These also involve the translocation of crucial charged polymers, such as DNA and RNA, through protein nanopores, which is often difficult to monitor in vivo.[4] Owing to the involvement of electrically charged macromolecules, Kasianowicz et al. have shown that the single-molecule detection of individually translocating single-stranded DNA (ssDNA) or RNA through an α-hemolysin (αHL) protein nanopore is possible under the influence of an applied electric field.[5] Further advancements in single-molecule techniques have made it possible to probe the translocation of individual polymers through biological as well as synthetic nanopores.[6–8] The unique charge distribution on each nucleotide provides a powerful method to determine genetic sequences of nucleic acids allowing the nanopores to act as sensors.[9–11] Base pair sequencing in DNA can be understood by measuring the changes in the ion current. Moreover, monitoring the blockades of the electrolytic ion current can provide information about the time required for a single polymer chain to traverse through the nanopore. In an attempt to investigate the mechanism of the translocation process, the efficiency and sensing abilities of the transmembrane pores can be improved in various ways,[12, 13] for example, by tweaking their structures,[14, 15] modulating the pH,[16, 17] changing the salt concentration,[18] etc.

The translocation process of a polymer through a nanopore is preceded by the rate-limiting step involving the capture of the polymer at the mouth of the pore.[7, 19] Since nanopores are also used for genome sequencing as individual nucleic acids pass through them, it is important to enhance the throughput of these sensors by increasing the capture rates of polymers within the pore.[20] Intuitively, increasing the strength of the applied electric field across the system could facilitate the capture of negatively charged nucleic acids but it will also speed up the translocation process such that sequencing becomes difficult. Multiple solutions have been proposed for this problem such as using optical tweezers to reduce the translocation speed,[21] controlling the capture and translocation by manipulating the applied electric field,[22, 23] increasing the solution flow rates to promote fast DNA capture,[24] ensuring asymmetric pH conditions on either side of the pore,[25] and introducing a salt gradient throughout the nanopores.[26, 27]

The desired results of fast polymer capture and slow translocation through nanopores can also be achieved by introducing macromolecular crowders in the bulk solution.[28] The motivation for creating a crowded environment in vitro is rooted in the need to

mimic real cellular conditions as about 10-40% of the cytoplasmic volume in biological cells is occupied by macromolecules.[29, 30] Due to their ubiquitous nature in the cell cytoplasm, macromolecular crowders are expected to influence the transport of desired chemical species across biological membranes. Larimi et al. probed the capture dynamics of the polypeptide Syn B2 within the biological αHL protein nanopore under the influence of macromolecular poly(ethylene glycol) PEG crowders existing at both the *cis* and *trans* sides of the pore.[31] Their results indicate that high concentrations of large-sized PEG crowders lead to stronger interactions of Syn B2 within the β-barrel of the αHL nanopore at the *trans* (entry) side. These interactions energetically favor an enhanced capture and a slow release of the same. Macromolecular crowding can also enhance the sensitivity and detection of nucleic acids and proteins in solid-state nanopores.[32, 33] Additionally, MD simulations show that PEG crowders etched to the inner walls of solid-state nanopores slowed down the translocation of amyloid-β (Aβ) fibrils.[34] The powerful role of PEG crowders in facilitating the capture of a ssDNA with an αHL nanopore along with a simultaneous decrease in their translocation across the nanopore was also reported by Yao et al.[35] The enhancement in the capture rate was attributed to the bulk crowding effect of PEG within the solution. On the other hand, slower translocation through the nanopore was explained based on an increase in solution viscosity and a decrease in the electrophoretic forces due to apparent nonspecific crowder-DNA interactions.[35]

Although PEG is mainly considered an inert crowding agent, it can interact with biological molecules and affect their stability.[36, 37] These large polymeric structures can traverse across membranes just like charged nucleic acids.[38] Multiple PEG molecules can also coexist within nanoscopic pores if their sizes are suitable enough.[39, 40] Moreover, it has been observed that while inside the αHL nanopore, PEG molecules are polycationic due to their successful coordination with positively charged counterions in the solution, which further enhances the binding of multiple PEG molecules within the nanopore.[41–43] This aspect of cooperative partitioning of PEGs has been completely neglected while explaining the experimental observations.[31, 35] To explain the effect of bulk macromolecular crowders on the DNA capture, its translocation, and accurate sequencing, one needs to have a molecular-level understanding of how the crowding effect of macromolecules (especially PEG) impacts the polymer capture and translocation rates.

This chapter presents a minimal discrete-state stochastic model to describe the capture of a single DNA molecule within αHL nanopores from the *cis* end and its subsequent translocation toward the *trans* side under the influence of the applied electric field in the

presence of PEG bulk crowders. Specifically, it is considered that these bulk crowders exist only on the *cis* side of the nanopore, which was also probed in a recent single-molecule study.[35] The theory is built on the hypothesis that individual PEG molecules can cooperatively enter within the αHL nanopore. Consequently, it is proposed that the presence of polycationic PEGs within the nanopore controls the capture of negatively charged DNA and its translocation due to electrostatic interactions. This minimal model accounts for these processes independently and provides exact analytical results. The resulting theoretical predictions are found to be in good agreement with existing experimental observations.

5.2 Theoretical Model

Let us consider a single DNA molecule on the *cis* side of an αHL nanopore, along with PEG crowders, translocating through the pore toward the *trans* side as schematically shown in Figure 5.1a. In this theoretical picture, it is proposed that the αHL protein nanopore can exist in two distinct macroscopic states − an empty state (E) without the DNA molecule within the pore and the blocked state (B) when the nanopore is occupied by the DNA. Consequently, as described in Figure 5.1a, the transition from state E to state B represents the capture of the single DNA molecule on the *cis* side at the mouth of the pore with the effective rate k_{on}. The translocation of the DNA molecule through the αHL nanopore toward the *trans* side leads to the transition from state B to state E with an effective rate k_{off}. As shown in Figure 5.1b, PEG molecules can enter and collide with the walls of the nanopore cavity irrespective of its macroscopic state. It is assumed that these crowder molecules can enter and leave the nanopore cavity from the *cis* side only. The main aim is to explore the effect of PEG crowders on the capture and translocation events depending on the number of PEGs partitioning within the nanopore.

Under the influence of an applied electric field, experimental studies have probed the capture and translocation events independently, i.e., they occur irreversibly.[31, 35] Thus, discrete-state stochastic models are introduced to describe these two processes separately. Figure 5.2 depicts the DNA capture process, whereas in Figure 5.3, the DNA translocation is presented. Using the schematic illustration of the DNA capture process in Figure 5.2a, corresponding discrete-state stochastic description in Figure 5.2b is developed. Here, each microscopic state n_E $(0 \leq n_E \leq N_E)$ represents the number of PEGs present within the αHL nanopore during the DNA capture process. Similarly, following the schematic picture in Figure 5.3a, each microscopic state n_B $(0 \leq n_B \leq N_B)$ in the discrete-state

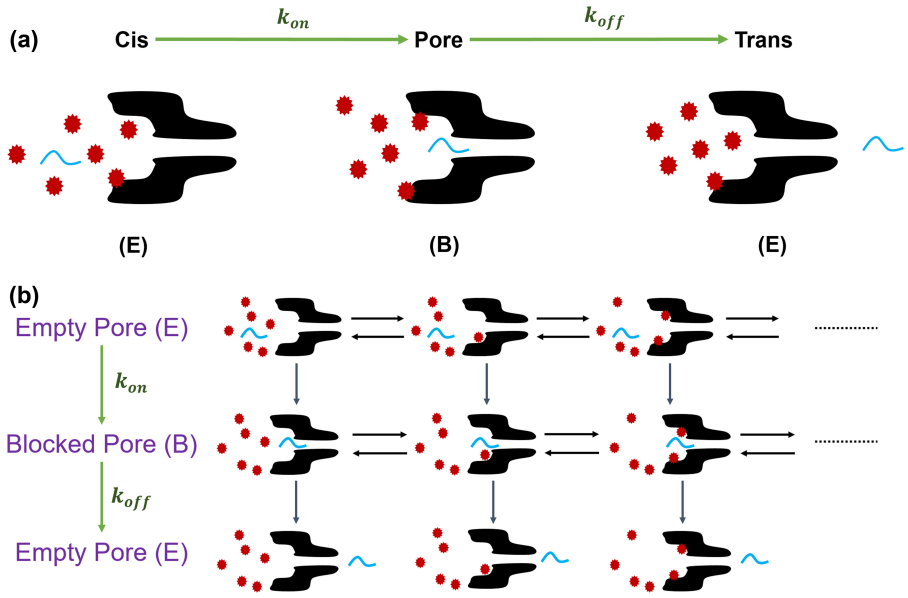


Figure 5.1: Schematic representation of (a) the capture and translocation of a single DNA (blue line) from the *cis* side of an αHL nanopore (black) to the *trans* side in the presence of PEG crowders (red) in the *cis* compartment only. (b) PEG molecules entering and leaving the nanopore cavity from the *cis* side irrespective of the position of the DNA molecule during the capture and translocation process.

stochastic model corresponds to the number of PEGs within the pore during the successful translocation of a DNA molecule (Figure 5.3b).

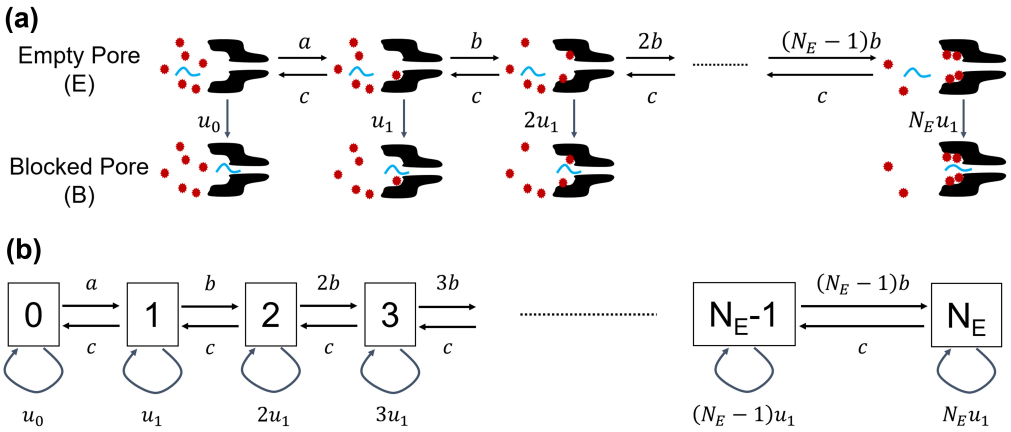


Figure 5.2: (a) Schematic representation of microscopic states during the DNA capture process in the presence of PEGs on the *cis* side of the pore. (b) Corresponding stochastic discrete-state description of the process. The rates $a = a_0[PEG]$ and $b = b_0[PEG]$ are pseudo-first-order rate constants.

The entry of the very first PEG molecule during the capture and translocation process occurs independently, whereas the entry of the subsequent PEGs is enhanced and occurs cooperatively due to the cation-induced conformational changes within the αHL nanopore.[41–43] To account for the conformational changes, the pseudo-first-order tran-

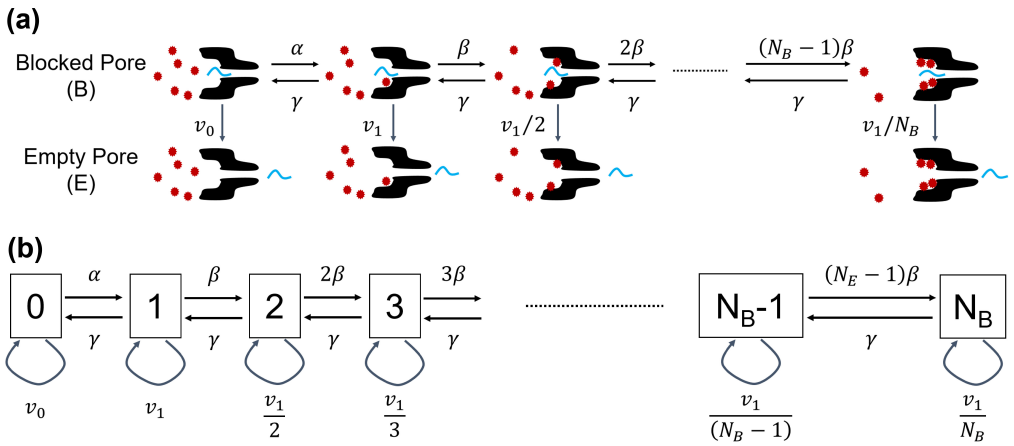


Figure 5.3: (a) Schematic representation of microscopic states during the DNA translocation process in the presence of PEGs on the *cis* side of the pore. (b) Corresponding stochastic discrete-state description of the process. The rates $\alpha = \alpha_0[PEG]$ and $\beta = \beta_0[PEG]$ are pseudo-first-order rate constants.

sition rate from the state n_E to $n_E + 1$ for $n_E = 0$ during capture (rate $a = a_0[PEG]$) or n_B to $n_B + 1$ for $n_B = 0$ during translocation (rate $\alpha = \alpha_0[PEG]$) is considered to be different from $n_E \geq 1$ or $n_B \geq 1$. After the addition of the first crowder within the pore, the cooperative entry of other crowders occurs at a different rate depending on the number of PEGs existing within the pore due to their enhanced binding reported previously.[41, 42] Motivated by the early studies by Adair and Klotz in understanding the receptor-ligand binding ensemble kinetics,[44, 45] this theoretical analysis considers that the cooperative addition of PEGs occurs in a sequential manner. Thus, the partitioning rate of a new crowder in states $n_E \geq 1$ or $n_B \geq 1$ is a product of the number of PEGs within the pore and a rate constant expressed via $n_E b$ or $n_B \beta$ such that $b = b_0[PEG]$ and $\beta = \beta_0[PEG]$ are pseudo-first-order rate constants. Such cooperative partitioning introduces positive charges on PEGs present in the nanopore cavity.[41]

Further, the irreversible capture or translocation from the state $n_E = 0$ or $n_B = 0$ happens through the rate u_0 or v_0, respectively. For $n_E \geq 1$ or $n_B \geq 1$, these irreversible rates are considered to be governed by the attractive electrostatic interactions between the negatively charged DNA and pseudocationic PEG crowders present in the cavity of the nanopore. These interactions can increase the DNA capture rates and retard the DNA translocation toward the *trans* side, thereby increasing the dwell times, both of which have been reported experimentally.[31, 35] Strong evidence for faster capture was also provided by the introduction of positive charges on the *cis* end of the αHL nanopore.[46] The presence of polycationic PEGs inside the nanopore cavity, therefore, not only leads to a cooperative addition of more PEG molecules but also promotes the cooperative capture

of DNA if PEG molecules are present within the pore. Additionally, an increase in the number of PEGs inside the nanopore will amplify the positive charge, which will strengthen the effect of electrostatic interactions during the DNA capture or translocation process. To account for these interactions in this description, it is considered that the irreversible capture happens with a rate $n_E \times u_1$ for states $n_E \geq 1$, and for $n_B \geq 1$, the DNA translocates with a rate of v_1/n_B. In the stochastic discrete-state description for the DNA capture (Figure 5.2b) or translocation (Figure 5.3b) from any state n_E or n_B, the system resets back to the same state as soon as this irreversible transition occurs. Physically, this means that the number of PEGs within the nanopore remains unaltered before and after an irreversible transition (see Figures 5.2a and 5.3a). From all crowder-occupied states $n_E \geq 1$ (or $n_B \geq 1$), a single PEG can independently leave the nanopore toward the *cis* side by performing a transition from n_E to $n_E - 1$ or n_B to $n_B - 1$ with a rate constant c or γ, respectively.

Since the stochastic discrete-state models described for DNA capture in Figure 5.2b and translocation in Figure 5.3b are similar in representation, one can combine these two separate but similar models and adopt a general discrete-state stochastic description as shown in Figure 5.4. Here, the microscopic state n implies that the nanopore is occupied by n PEGs, where state N corresponds to the maximum PEGs within the pore. For the capture stage, $N = N_E$ and $N = N_B$ during translocation. If $n = 0$, the addition of the first PEG within the nanopore happens through the rate constant x. For the state $n \geq 1$, the transition from n to $n + 1$ occurs with the rate constant ny to account for the enhanced binding of multiple PEGs within the pore due to cation-induced conformational changes.[41, 42] Further, from any state $n \geq 1$, a single crowder leaves the mouth of the nanopore to the *cis* side via the rate constant z. During the DNA capture process, $n = n_E$, $x = a$, $y = b$, and $z = c$ (see Figure 5.2b), whereas during DNA translocation, $n = n_B$, $x = \alpha$, $y = \beta$, and $z = \gamma$ (see Figure 5.3b). Additionally, in this general description, the irreversible DNA capture or translocation occurs with a rate constant w_n. For states $n = 0$ and $n \geq 1$, the rates $w_0 = u_0$ and $w_n = n_E u_1$ during DNA capture, whereas $w_0 = v_0$ and $w_n = v_1/n_B$ during DNA translocation. This minimal discrete-state stochastic description in Figure 5.4 is sufficient to explain the irreversible processes of DNA capture and translocation independently. It is emphasized that the cooperative addition of polycationic PEGs inside the nanopore in this theoretical description and the consequent electrostatic DNA-PEG interactions control the dynamics of the independently occurring DNA capture and translocation events.

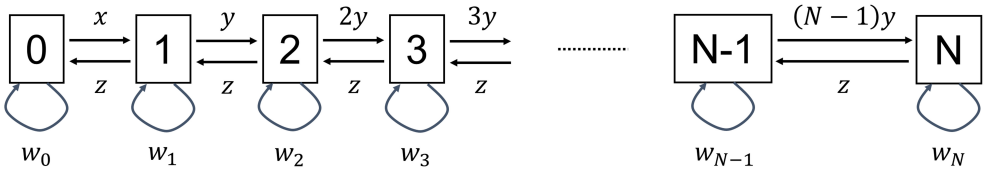


Figure 5.4: General discrete-state stochastic description for DNA capture or translocation in the presence of PEG crowders present on the *cis* side of the nanopore.

One can define a function $F_n(t)$ as the first-passage probability density function for DNA capture or translocation to occur at time t if the nanopore was occupied by n crowders at $t = 0$. Additionally, the function $F_T(t) = \delta(t)$ is defined as the boundary condition, which physically means that the desired irreversible transition is accomplished immediately at $t = 0$. The temporal evolution of the first-passage probabilities can be described through a set of backward master equations. For $n = 0$, one writes

$$\frac{dF_0(t)}{dt} = xF_1(t) + w_0F_T(t) - (x + w_0)F_0(t) \tag{5.1a}$$

and when $n \in [1, N-1]$,

$$\frac{dF_n(t)}{dt} = nyF_{n+1}(t) + w_nF_T(t) + zF_{n-1}(t) - (ny + w_n + z)F_n(t) \tag{5.1b}$$

For the microscopic state N,

$$\frac{dF_N(t)}{dt} = w_NF_T(t) + zF_{N-1}(t) - (w_N + z)F_N.(t) \tag{5.1c}$$

The aim is to obtain the analytical expressions of these first-passage probability distributions during DNA capture and translocation events. All of the dynamic properties such as the mean capture time, the mean translocation time (or dwell time), and their higher moments can be determined through the well-known relation $\langle \tau_m \rangle = \int_0^\infty t_m F_n(t)dt$. To simplify the analysis, it is assumed that the rate constant $z = 1\text{s}^{-1}$. The solutions for Eq. 5.1 are determined after performing their Laplace transform using $\tilde{F}_n(s) = \int_0^\infty F_n(t)e^{-st}dt$, and eventually, one obtains a general form of the function $\tilde{F}_n(s)$ as discussed in the Appendix C. From any microscopic state n, the mean first-passage time for DNA capture or translocation is provided by

$$\begin{aligned} \langle T_n \rangle &= -\frac{\partial \tilde{F}_n(s)}{\partial s}\Big|_{s=0} \\ &= \frac{\tilde{A}_n(0) + \sum_{i=1}^{N-n} \tilde{A}_{n+i}(0)\tilde{g}_i(0)}{\tilde{B}_n(0)} \end{aligned} \tag{5.2}$$

The second moment of $F_n(t)$ will help in understanding the role of underlying stochastic fluctuations within the system,[47] arising due to n PEG bulk crowders existing within the protein nanopore in state n during the DNA capture or translocation. It is determined using

$$\begin{aligned}\langle T_n^2 \rangle &= -\frac{\partial^2 \tilde{F}_n(s)}{\partial s^2}\Big|_{s=0} \\ &= \frac{2\left[\left\{\tilde{h}'_n(0)\tilde{A}_n(0) - \tilde{A}'_n(0)\right\} + \sum_{i=1}^{N-n} \tilde{g}_i(0)\left\{\tilde{A}_{n+i}(0)\sum_{k=0}^{i}\tilde{h}'_{n+k}(0) - \tilde{A}'_{n+i}(0)\right\}\right]}{\tilde{B}_n(0)}\end{aligned} \tag{5.3}$$

where $\tilde{h}_n(s) = \ln \tilde{B}_n(s)$ and all other functions are defined in the Appendix C. The values of n, x, y, and z will depend upon the DNA capture or translocation process. Equations 5.2 and 5.3 describe the dynamic properties of the DNA capture or translocation initiating from any state n. However, these irreversible transitions can start from any of the discrete states as presented in Figure 5.4. So, to obtain the average time for DNA capture or translocation process, the quantities that are measured in single-molecule experiments, additional calculations are required as explained below.

Let us define $P_n(t)$ as the probability for the nanopore to exist in state n at time t during DNA capture or translocation. The evolution of the probabilities $P_n(t)$ is expressed through a set of forward master equations presented in the Appendix C. Since PEGs are already present in the bulk solution on the *cis* side before the monitoring of DNA capture or translocation begins, it is assumed that the system achieves steady-state distribution ($t \to \infty$) quickly. By utilizing the normalization condition, i.e., $\sum_{i=0}^{i=N} P_i = 1$, one can solve for the steady-state probabilities (details in the Appendix C) for each microscopic state n, and eventually obtain

$$P_n = \begin{cases} \frac{1}{1+x\sum_{i=0}^{N-1} i!y^i} & n = 0 \\ \frac{xy^{n-1}(n-1)!}{1+x\sum_{i=0}^{N-1} i!y^i} & 1 \le n \le N \end{cases} \tag{5.4}$$

Using the above expressions of the steady-state probabilities, one can also determine the relative probability of the PEG-occupied nanopore to be

$$\Pi = \frac{\sum_{n=1}^{n=N} P_n}{\sum_{n=0}^{n=N} P_n} = \frac{x \sum_{i=1}^{N} (i-1)! y^{i-1}}{1 + x \sum_{i=0}^{N-1} i! y^i} \tag{5.5a}$$

Due to the presence of the pseudo-first-order rate constants, the above equation may also be expressed as a function of PEG concentration such as

$$\Pi = \frac{x_0 \sum_{i=1}^{N} (i-1)! y_0^{i-1} [PEG]^i}{1 + x_0 \sum_{i=0}^{N-1} i! y_0^i [PEG]^i} \tag{5.5b}$$

This expression for the relative probability of the crowder-occupied nanopore is qualitatively similar to the Adair-Klotz equation for the fractional occupancy of ligand-bound complex described for the ensemble kinetics of cooperative receptor-ligand interactions.[45]

To proceed with the calculations, it is required to determine the respective weights for the DNA capture or translocation to start from each microscopic state n using these probabilities defined in Eq. 5.4. According to the discrete-state description in Figure 5.4, a new event will start from any state n only as a consequence of reset, i.e., if the previous cycle was accomplished through rate w_n while the nanopore was occupied by n PEGs. The probability to reach state n to start a new event is a product of rate w_n and the steady-state probability P_n. Thus, the desired weighted contribution is the relative probability of starting a new event from state n normalized to all possible starting situations, which is represented by

$$f_n = \frac{w_n P_n}{\sum_{i=0}^{i=N} w_n P_n} \tag{5.6}$$

Finally, the experimental equivalents of the average time for DNA capture or translocation and their second moments can be determined using Eqs. 5.2-5.4 and 5.6. By averaging over all possible initial states, the mean time is evaluated according to

$$\begin{aligned}\langle\tau\rangle &= \sum_{n=0}^{N} f_n\langle T_n\rangle \\ &= \frac{1 + x\sum\limits_{i=0}^{N-1} i!y^i}{w_0 + x\sum\limits_{i=1}^{N}(i-1)!w_n y^{i-1}}.\end{aligned} \tag{5.7}$$

An alternative approach for evaluating the mean time in Eq. 5.7 from the steady-state probability distributions is discussed in the Appendix C. Using Eqs. 5.3 and 5.6, a similar analysis can be performed to determine the mean-squared times and one finds

$$\langle\tau^2\rangle = \sum_{n=0}^{N} f_n\langle T_n^2\rangle. \tag{5.8}$$

Since higher moments help provide information about underlying fluctuations within a system,[47] the well-known statistical quantity called randomness parameter is employed to quantify these stochastic effects[48] defined by

$$R = \frac{\langle\tau^2\rangle - \langle\tau\rangle^2}{\langle\tau\rangle^2} \tag{5.9}$$

where $\langle\tau\rangle$ and $\langle\tau^2\rangle$ are obtained using Eqs. 5.7 and 5.8, respectively. For single-step reactions involving no fluctuations, the randomness parameter is unity. R deviates from unity in the presence of stochastic fluctuations within any system.[48] Calculation of R can also provide information on the underlying chemical kinetic scheme. In this study, the average time for DNA capture and translocation is denoted by $\langle\tau\rangle_{on}$ and $\langle\tau\rangle_{off}$, respectively, and the corresponding randomness parameters via R_{on} or R_{off}, respectively, constitute the main results of this work.

5.3 Results and Discussion

The theoretical results obtained through this minimal discrete-state stochastic description can be directly applied to single-molecule experiments. One can determine numerous important quantities which can influence the physical conditions required to develop accurate and efficient biological sensors. Let us consider a recent single-molecule study, where the capture and translocation process of DNA molecules was explored under the influence of

an applied electric field in the presence of PEG polymers of different molecular weights and concentrations.[35] The DNA capture and translocation from the *cis* side to the *trans* side were monitored when PEG 4k crowders in varying concentrations from 0 to 40% w/v were placed on the *cis* side of the nanopore only. Multiple quantities such as the interevent frequency (f), the dwell time which is the average DNA translocation time, the effective rates k_{on} and k_{off}, the equilibrium binding constant (K_b), and the Gibbs free energy (ΔG) for the DNA capture and translocation through the nanopore were determined.

By using Eq. 5.6 and the respective constants corresponding to the DNA capture process, the interevent frequency (f) which is the reciprocal of the average capture time of the DNA at the mouth of the protein nanopore is analytically calculated as

$$f = \langle \tau \rangle_{on}^{-1} = \frac{u_0 + u_1 a_0 [PEG] \sum_{i=1}^{N_E} i! \big(b_0 [PEG]\big)^{i-1}}{1 + a_0 [PEG] \sum_{i=0}^{N_E - 1} i! \big(b_0 [PEG]\big)^{i}}. \tag{5.10}$$

Here, the PEG concentration in % w/v units is

$$[PEG] = \frac{1000}{M_{PEG}} \times \frac{C_{PEG}\%}{100\%} \tag{5.11}$$

such that $M_{PEG} = 4000$ g/mol is the molar mass of the PEG 4k crowders and C_{PEG} is the % w/v concentration used. After substituting Eq. 5.11 in Eq. 5.10, a good agreement with the experimentally obtained interevent frequencies at varying PEG 4k concentrations is observed when N_E lies between 7 and 10. This means that the *cis* end of the nanopore can accommodate about 7-10 molecules of PEG 4k (introduced cooperatively) in all possible conformations and arrangements when a successful DNA capture occurs in a cooperative manner. Figure 5.5a shows the comparison of this minimal stochastic theory using $N_E = 7$ with experimentally obtained interevent frequency. The best-fit value is obtained at $u_0 < u_1$, which confirms that the DNA capture within a crowder-occupied nanopore is accelerated due to strong electrostatic forces of attraction, compared to the situation where no PEG molecules are within the pore, thereby supporting this work. At $[DNA] = 100$ nM,[35] this theory estimates that $k_{on} = \big(\langle \tau \rangle_{on} [DNA]\big)^{-1} = f \times 10^7$ M^{-1} s^{-1}. Using Eq. 5.7 and the respective constants, one can also calculate the dwell time, which is the average time for the DNA translocation process

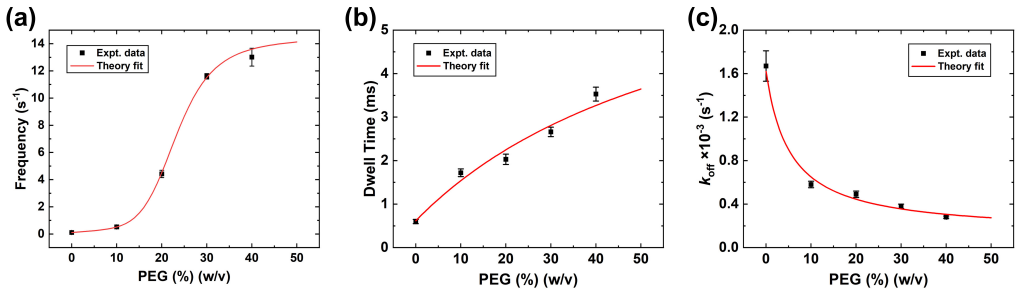


Figure 5.5: (a) Interevent frequency (f), (b) DNA translocation time ($\langle\tau\rangle_{off}$), and (c) DNA translocation rate (k_{off}) as a function of the concentration of PEG 4k crowders on the *cis* side of the pore. Black symbols correspond to experimental data extracted from ref.[35] whereas red curves are theoretical predictions. The best-fit parameters $N_E = 7$, $a_0 = 4.55\ \mathrm{s}^{-1}\ \mathrm{M}^{-1}$, b0 $= 6.97\ \mathrm{s}^{-1}\ \mathrm{M}^{-1}$, u0 $= 0.11\ \mathrm{s}^{-1}$, u1 $= 2.11\ \mathrm{s}^{-1}$, $N_B = 1$, $\alpha_0 = 74.61\ \mathrm{s}^{-1}\ \mathrm{M}^{-1}$, $v_0 = 1.6189\times10^3\ \mathrm{s}^{-1}$, and $v_1 = 0.1316\times10^3\ \mathrm{s}^{-1}$ were used for theoretical curves.

$$\langle\tau\rangle_{off} = \frac{1+\alpha_0[PEG]\sum_{i=0}^{N_B-1} i!\big(\beta_0[PEG]\big)^i}{v_0+v_1\alpha_0[PEG]\sum_{i=1}^{N_B} i!\big(\beta_0[PEG]\big)^{i-1}} \tag{5.12a}$$

After substituting Eq. 5.11 in the above equation, it was found that when $N_B = 1$, the resultant analytical expression provides an excellent agreement with the experimentally obtained dwell times at different PEG 4k concentrations. For $N_B = 1$, Eq. 5.12a is reduced to

$$\langle\tau\rangle_{off} = \frac{1+\alpha_0[PEG]}{v_0+v_1\alpha_0[PEG]}. \tag{5.12b}$$

Figure 5.5b shows the comparison of experimentally obtained translocation times with Eq. 5.12b. Physically, the value $N_B = 1$ represents that no more than one molecule of PEG 4k should exist in the nanopore during the DNA translocation to the *trans* side. Further, the rate constant of the DNA translocation can be determined from the average translocation time, $k_{off} = \langle\tau\rangle_{off}^{-1}$, and Figure 5.5c shows that it is in quantitative agreement with the experimental observations.

The binding constant, $K_b = k_{on}/k_{off}$, determined from the DNA capture and translocation rates can be represented as $K_b = f\langle\tau\rangle_{off}[DNA]^{-1}$. One can also determine the Gibbs free energy for the whole process using $\Delta G = -RT\ln K_b$. The theoretical values of K_b and ΔG are calculated by using the best-fit parameters for the DNA capture and translocation events mentioned in Figure 5.5. Excellent quantitative agreements with experimental data exist, as shown in Figure 5.6, which provides additional support to the proposed theoretical approach.

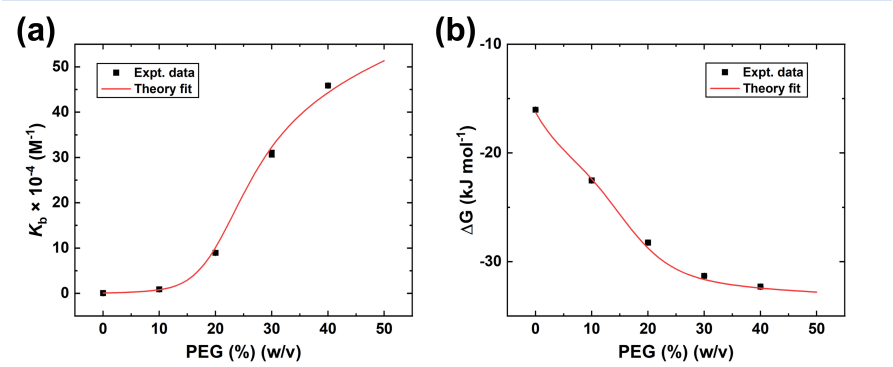


Figure 5.6: (a) Binding constant (K_b) and (b) Gibbs free energy (ΔG) as a function of PEG 4k concentration on the *cis* side of the pore. Black symbols correspond to experimental data extracted from ref.[35], whereas red curves are theoretical predictions.

Figure 5.7 presents the randomness parameter determined from Eq. 5.9 for the DNA capture and translocation events using the respective best-fit parameters. R shows a non-monotonic behavior with respect to PEG 4k concentration for both events. In the absence of any bulk crowders, $x = 0$, and according to the proposed discrete-state model in Figure 5.4, the nanopore will always remain in only one microscopic state, i.e., $n = 0$. Consequently, $\langle\tau\rangle_{x=0} = \langle\tau\rangle_{n=0} = w_0^{-1}$ and $\langle\tau^2\rangle_{x=0} = \langle\tau^2\rangle_{n=0} = 2w_0^{-2}$, which results in $R = 1$. Thus, in the absence of PEG crowders, $[PEG] = 0$, there are no fluctuations within the system. When $[PEG] > 0$, the randomness parameter gradually increases from unity. All microscopic states are accessible by the nanopore at moderate PEG 4k concentrations, which lead to stochastic fluctuations,[48] because irreversible transitions of DNA capture or translocation can occur from any state. At high PEG 4k concentrations, when $x \to \infty$ and $y \to \infty$, the protein nanopore is mainly found in only one state, i.e., $n = N$ according to the steady-state probabilities defined in Eq. 5.4. As a result, $\langle\tau\rangle_{x,y\to\infty} = \langle\tau\rangle_N \simeq (Nw_1)^{-1}$ and $\langle\tau^2\rangle_{x,y\to\infty} = \langle\tau^2\rangle_N \simeq 2(Nw_1)^{-2}$ and R is unity. This indicates that at very high crowder concentrations, there are no stochastic fluctuations within the system. Therefore, by controlling the concentration of bulk crowders, one can tune the fluctuations during DNA capture or translocation.

5.4 Conclusions

This chapter presented a new theoretical approach to probe the molecular mechanism of DNA capture and translocation process through an αHL nanopore, under the influence of an applied electric field, with macromolecular PEG crowders present only on the *cis* side of the nanopore. This study developed a novel discrete-state stochastic description that

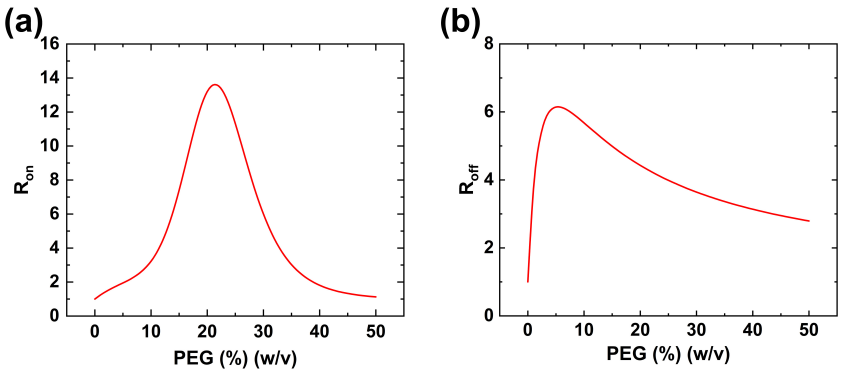


Figure 5.7: Randomness parameter determined using Eq. 5.9 for DNA (a) capture (R_{on}) and (b) translocation (R_{off}) as a function of PEG 4k concentration on the *cis* side of the pore.

allows for exact analytical solutions using stationary-state and first-passage probability distribution calculations. The observed enhanced binding of polycationic multiple PEG molecules within the αHL nanopore[41] forms the basis of this theory. It is considered that individual PEG crowders cooperatively enter within the nanopore cavity, which introduces a polycationic charge on these polymers inside the nanopore. Consequently, it is argued that the apparent electrostatic interactions existing between negatively charged DNA and polycationic PEG crowders within the nanopore result in the experimentally observed increase in the DNA capture rate[31, 35] and the translocation time.[35] This theory successfully explains the opposing nature of these events. Using obtained analytical results, one can even predict the maximum number of crowder molecules that a protein nanopore can accommodate within its cavity. In addition, the randomness parameter determined from this work is a measure of stochastic fluctuations in the system, which non-monotonically depends on the PEG concentration. A high crowder concentration can minimize fluctuations along with enhanced capture rates and translocation times desired for high-throughput accurate nanopore sensing. Excellent quantitative agreements with experimental data strongly support this theoretical approach. It is believed that the presented theoretical approach to model cooperative binding at the microscopic level is the first of its kind and can be adopted in other studies as well.

Though this simple minimal discrete-state stochastic description serves as an initial step to understanding the role of bulk crowders on the DNA capture and translocation events, there are assumptions in this model. The cooperative binding of a new PEG molecule is considered to occur sequentially as a simple linear function of the number of existing PEGs inside the nanopore cavity. However, the increase in the transition rates can follow other functional forms, such as exponential functions. The choice of such functions

on the translocation dynamics remains to be tested. This work considered PEG crowders on the *cis* side of the nanopore only, which allowed the construction of a minimal discrete-state model to probe the impact of these crowders on both the capture and translocation events. However, experiments have considered them being simultaneously present on the *trans* side as well.[31, 35] It is assumed that only one type of PEG molecule exists within the solution, whereas multiple types of PEGs could be present in the solution,[39] reflecting the heterogeneous nature of the cytoplasm. Experimentally, in such cases, it has been reported that crowder size can also affect these processes.[31, 35] It will be important to generalize this simple approach to accommodate all of these factors while still being able to solve it explicitly.

Bibliography

(1) Lodish, H. F. et al., *Molecular Cell Biology*; WH Freeman New York: 2006.

(2) Phillips, R. et al., *Physical Biology of the Cell*; Garland Science: 2012.

(3) Agarraberes, F. A.; Dice, J. F. Protein Translocation Across Membranes. *Biochim. Biophys. Acta, Biomembr.* **2001**, *1513*, 1–24.

(4) Muthukumar, M. Mechanism of DNA Transport Through Pores. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 435–450.

(5) Kasianowicz, J. J. et al. Characterization of Individual Polynucleotide Molecules using a Membrane Channel. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13770–13773.

(6) Deamer, D. W.; Branton, D. Characterization of Nucleic Acids by Nanopore Analysis. *Acc. Chem. Res.* **2002**, *35*, 817–825.

(7) Chen, P. et al. Probing Single DNA Molecule Transport Using Fabricated Nanopores. *Nano Lett.* **2004**, *4*, 2293–2298.

(8) Bates, M. et al. Dynamics of DNA Molecules in a Membrane Channel Probed by Active Control Techniques. *Biophys. J.* **2003**, *84*, 2366–2372.

(9) Branton, D. et al. The Potential and Challenges of Nanopore Sequencing. *Nat. Biotechnol.* **2008**, *26*, 1146–1153.

(10) Zhou, W. et al. Molecular Insights into Distinct Detection Properties of α-Hemolysin, MspA, CsgG, and Aerolysin Nanopore Sensors. *J. Phys. Chem. B* **2020**, *124*, 1611–1618.

(11) Mayer, S. F. et al. Biological Nanopores for Single-Molecule Sensing. *iScience* **2022**, *25*.

(12) Ayub, M.; Bayley, H. Engineered Transmembrane Pores. *Curr. Opin. Chem. Biol.* **2016**, *34*, 117–126.

(13) Shi, W. et al. Nanopore Sensing. *Anal. Chem.* **2017**, *89*, 157–188.

(14) Soskine, M. et al. An Engineered ClyA Nanopore Detects Folded Target Proteins by Selective External Association and Pore Entry. *Nano Lett.* **2012**, *12*, 4895–4900.

(15) Mohammad, M. M. et al. Engineering a Rigid Protein Tunnel for Biomolecular Detection. *J. Am. Chem. Soc.* **2012**, *134*, 9521–9531.

(16) De Zoysa, R. S. et al. Translocation of Single-Stranded DNA Through the α-Hemolysin Protein Nanopore in Acidic Solutions. *Electrophoresis* **2011**, *32*, 3034–3041.

(17) Ghosh, B.; Chaudhury, S. Translocation Dynamics of an Asymmetrically Charged Polymer through a Pore under the Influence of Different pH Conditions. *J. Phys. Chem. B* **2019**, *123*, 4318–4323.

(18) Smeets, R. M. et al. Salt Dependence of Ion Transport and DNA Translocation through Solid-State Nanopores. *Nano Lett.* **2006**, *6*, 89–95.

(19) Wong, C. T. A.; Muthukumar, M. Polymer Capture by Electro-Osmotic Flow of Oppositely Charged Nanopores. *J. Chem. Phys.* **2007**, *126*.

(20) He, Y. et al. DNA Capture in Nanopores for Genome Sequencing: Challenges and Opportunities. *J. Mater. Chem.* **2012**, *22*, 13423–13427.

(21) Trepagnier, E. H. et al. Controlling DNA Capture and Propagation through Artificial Nanopores. *Nano Lett.* **2007**, *7*, 2824–2830.

(22) Mirsaidov, U. et al. Slowing the Translocation of Double-Stranded DNA using a Nanopore Smaller than the Double Helix. *Nanotechnology* **2010**, *21*, 395501.

(23) He, Y. et al. Gate Manipulation of DNA Capture into Nanopores. *ACS Nano* **2011**, *5*, 8391–8397.

(24) Sohi, A. N. et al. DNA Capture by Nanopore Sensors under Flow. *Anal. Chem.* **2020**, *92*, 8108–8116.

(25) Wang, Y. et al. Remote Activation of a Nanopore for High-Performance Genetic Detection using a pH Taxis-Mimicking Mechanism. *Anal. Chem.* **2017**, *89*, 13039–13043.

(26) Wanunu, M. et al. Electrostatic Focusing of Unlabelled DNA into Nanoscale Pores using a Salt Gradient. *Nat. Nanotechnol.* **2010**, *5*, 160–165.

(27) Leong, I. W. et al. Salt Gradient Control of Translocation Dynamics in a Solid-State Nanopore. *Anal. Chem.* **2021**, *93*, 16700–16708.

(28) Singh, S. L. et al. Polymer Translocation and Nanopore Sequencing: A Review of Advances and Challenges. *Int. J. Mol. Sci.* **2023**, *24*, 6153.

(29) Zimmerman, S. B.; Trach, S. O. Estimation of Macromolecule Concentrations and Excluded Volume Effects for the Cytoplasm of Escherichia Coli. *J. Mol. Biol.* **1991**, *222*, 599–620.

(30) Ellis, R. J. Macromolecular Crowding: Obvious but Underappreciated. *Trends Biochem. Sci.* **2001**, *26*, 597–604.

(31) Larimi, M. G. et al. Interactions of a Polypeptide with a Protein Nanopore under Crowding Conditions. *ACS Nano* **2019**, *13*, 4469–4477.

(32) Chau, C. C. et al. Macromolecular Crowding Enhances the Detection of DNA and Proteins by a Solid-State Nanopore. *Nano Lett.* **2020**, *20*, 5553–5561.

(33) Marcuccio, F. et al. Mechanistic Study of the Conductance and Enhanced Single-Molecule Detection in a Polymer–Electrolyte Nanopore. *ACS Nanosci. Au* **2023**, *3*, 172–181.

(34) Meyer, N. et al. Detection of Amyloid-β Fibrils using Track-Etched Nanopores: Effect of Geometry and Crowding. *ACS Sensors* **2021**, *6*, 3733–3743.

(35) Yao, F. et al. Crowding-Induced DNA Translocation through a Protein Nanopore. *Anal. Chem.* **2020**, *92*, 3827–3833.

(36) Senske, M. et al. Protein Stabilization by Macromolecular Crowding Through Enthalpy Rather than Entropy. *J. Am. Chem. Soc.* **2014**, *136*, 9036–9041.

(37) Nakano, S.-i. et al. Effects of Molecular Crowding on the Structures, Interactions, and Functions of Nucleic Acids. *Chem. Rev.* **2014**, *114*, 2733–2758.

(38) Bezrukov, S. M. et al. Dynamics and Free Energy of Polymers Partitioning into a Nanoscale Pore. *Macromolecules* **1996**, *29*, 8517–8522.

(39) Gurnev, P. A. et al. Poly(ethylene glycol)s in Semidilute Regime: Radius of Gyration in the Bulk and Partitioning into a Nanopore. *Macromolecules* **2017**, *50*, 2477–2483.

(40) Movileanu, L. et al. Partitioning of Individual Flexible Polymers into a Nanoscopic Protein Pore. *Biophys. J.* **2003**, *85*, 897–910.

(41) Rodrigues, C. G. et al. Mechanism of KCl Enhancement in Detection of Nonionic Polymers by Nanopore Sensors. *Biophys. J.* **2008**, *95*, 5186–5192.

(42) Reiner, J. E. et al. Theory for Polymer Analysis using Nanopore-Based Single-Molecule Mass Spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 12080–12085.

(43) Balijepalli, A. et al. Theory of Polymer–Nanopore Interactions Refined using Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2013**, *135*, 7064–7072.

(44) Adair, G. S. et al. The Hemoglobin System: VI. The Oxygen Dissociation Curve of Hemoglobin. *J. Biol. Chem.* **1925**, *63*, 529–545.

(45) Klotz, I. M. Ligand-Receptor Complexes: Origin and Development of the Concept. *J. Biol. Chem.* **2004**, *279*, 1–12.

(46) Maglia, G. et al. Enhanced Translocation of Single DNA Molecules through α-Hemolysin Nanopores by Manipulation of Internal Charge. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 19720–19725.

(47) Singh, D. et al. Theoretical Tools to Quantify Stochastic Fluctuations in Single-Molecule Catalysis by Enzymes and Nanoparticles. *ACS Omega* **2022**, *7*, 47587–47600.

(48) Kou, S. et al. Single-Molecule Michaelis–Menten Equations. *J. Phys. Chem. B* **2005**, *109*, 19068–19081.

Chapter 6

Macromolecular Crowding Facilitates ssDNA Capture within Biological Nanopores: Role of Size Variation and Solution Heterogeneity

The content of this chapter and associated content in Appendix D are reprinted (adapted) with permission from (*J. Phys. Chem. B* 2024, 128, 8, 1876-1883). Copyright © 2024 American Chemical Society.

6.1 Introduction

Cellular membranes allow the selective transport of several biomolecules and ions, constituting a fundamental life process.[1, 2] Even though a challenging process to observe in vivo, essential charged polymers like single-stranded DNA (ssDNA) and RNA are transported across membranes through protein nanopores.[3, 4] Before the translocation of these polymers, they are first captured at the mouth of the nanopores which forms the rate-limiting step of the transport process.[5, 6] For the first time, Kasianowicz et al. demonstrated the detection of individual ssDNA or RNA molecules, driven by an applied electric field, as they pass through an α-hemolysin (αHL) protein nanopore.[7] Owing to the unique charge distribution of each nucleotide, monitoring the changes in the ion current during their transport paved the way for utilising biological and synthetic nanopores for genome sequencing.[5, 8–11]

Multiple attempts have been made over the years to improve the accuracy of these nanopore sensors by modulating the charge distribution of the system to achieve efficient characterization of the resulting current profiles.[12–16] Irrespective of the proposed solutions, it is desirable that these charged polymers reside in the nanopores for sufficient time for precise detection of each charged nucleotide. Simultaneously, it is also required that the capture of the polymers is fast to ensure a high throughput of sensors.[17] An enhanced capture rate may be achieved by several methods including manipulation of the electric field,[18] establishing salt gradients across the nanopores[19, 20] and controlling the flow rates of the solution.[21] The effect of macromolecular crowding[22] on ssDNA capture and translocation through the nanopore may be explored using single-molecule analytical techniques. Freedman et al. have shown that the inter-event frequency (or capture rate) between successive translocation processes is enhanced over time due to the apparent crowding of proteins around the nanopore.[23] The presence of large-size poly(ethylene glycol) (PEG) polymers in high concentrations within the solutions favor strong interactions (fast capture) between the polypeptide Syn B2 and the β-barrel of the αHL nanopore.[24] By introducing PEG crowders of varying size and concentrations on the *cis* side of αHL nanopore, Yao et al. have probed the capture and translocation of ssDNA using single-molecule experiments.[25] It was observed that high concentrations of optimum-size crowders favor fast ssDNA capture with a simultaneous slower translocation through the nanopore. Using an exactly solvable discrete-state stochastic framework, the previous chapter showed that the cooperative partitioning of individual polycationic

PEGs (owing to their coordination with positive counter-ions) within the nanopore and their strong electrostatic interactions with ssDNA may be responsible for the experimental observations by Yao et al. However, this theoretical analysis was limited to PEGs of a known molecular weight in varying concentrations and the role of their size was not explored.

The size (or molecular weight) of crowders existing in the solution also plays a crucial role in polymer sequencing. The detection of proteins by solid-state nanopores shows a significant enhancement in the presence of larger PEG polymers in solutions.[26, 27] Single-molecule studies by Yao et al. suggest that at fixed concentrations, the capture rates show a non-monotonic behavior with increasing PEG sizes whereas the translocation times remain largely independent of the molecular weights of these crowders.[25] In support of these observations, 2D Langevin dynamics simulation studies show that the translocation times remain unaltered to the crowder sizes but they show a non-monotonic dependence under the influence of active forces.[28] While keeping the area fractions of crowders the same on either side of the nanopore, Chen and Luo have theoretically shown that the translocation of polymer proceeds in the direction where the crowding agents are relatively bigger due to their large conformational entropy.[29]

Although most experimental studies consider homogenous crowded environments, in reality, the cellular matrix is highly complex. Several bio-macromolecules of various sizes occupy about 10-40 % of the cytoplasm.[22, 30] Thus, it is essential to explore the effect of such heterogeneous crowded conditions on the polymer capture and translocation through nanopores. A simple way to consider heterogeneity in solutions is to introduce a binary system. Multiple studies have probed the interactions of a binary mixture of PEG polymers with biological nanopores. Conductance-based measurements have shown enhanced PEG partitioning in biological nanopores for a binary mixture following the 'polymer-pushing-polymer' effect.[31, 32] According to this effect, large-size PEGs selectively push relatively smaller polymers within the nanopore.[32] The experimental observations that PEG 3400 polymers occupy merely 1/25th of the αHL nanopore if they exist in a 1:1 wt/wt mixture with PEG 200[33] further stresses the impact of inhomogeneity in solutions and the existence of 'polymer-pushing-polymer' concept. Unfortunately, the role of such binary mixtures of PEG molecules as crowders and their influence on the capture and translocation of charged polymers such as ssDNA and RNA through nanopores has not been investigated.

In this study, the previous chapter's minimal discrete-state stochastic framework[34] is extended to include the effect of varying crowder sizes on the capture of a ssDNA molecule at the *cis* side of a single αHL nanopore, under the influence of the applied electric field. This study is restricted to capture events only. Additionally, this work also probes the influence of heterogeneous crowding on this capture process by considering a simple binary mixture of PEG molecules. The exact analytical results in the presence of homogeneous crowded conditions are compared with existing single-molecule studies.[25] Obtained theoretical results on the influence of heterogeneous crowding are supported by suitable physical-chemical arguments.

6.2 Theoretical Model

6.2.1 Effect of Macromolecular Crowder Size on ssDNA Capture

Let us consider a ssDNA that is to be captured at the *cis* side of an αHL nanopore under the influence of an electric field. The *cis* compartment also consists of a known concentration of macromolecular PEG crowder molecules having the molecular weight MW. It is assumed that individual PEG polymers can sequentially partition within the nanopore during the ssDNA capture, in a cooperative manner.[34] If a single polymer occupies volume V_{PEG} of the αHL nanopore's total volume V_{pore}, one can define the maximum number of crowders (N) required to completely fill the pore using the relation

$$N = \lfloor R_V \rfloor = \lfloor V_{pore}/V_{PEG} \rfloor. \tag{6.1}$$

Here, $1/R_V$ denotes the volume fraction occupied by a single PEG crowder. The conventional notation of the floor function '$\lfloor x \rfloor$' is used to represent the integer part of any real number x. The number N varies according to the size (or MW) of the polymeric crowder. Considering the αHL nanopore as a cylinder of length 10 nm with an average radius of 1.2 nm across the membrane,[8] the V_{pore} is estimated to be 45.24 nm^3 during the capture process. PEG molecules also assume an almost cylindrical geometry within the nanopore due to induced conformational changes after binding with solvent cations.[35] From Ref.[35], the volume occupied by a single PEG consisting of m ($\sim \lfloor MW/44 \rfloor$) monomer units is $V_{PEG} = 67.425 \times m\text{Å}^3$. Table 6.1 presents the number N determined using Eq. 6.1 for

various PEG MWs used in a recent single molecule experiment during the ssDNA capture and translocation events.[25]

Table 6.1: Estimation of maximum PEG crowders N that can occupy an αHL nanopore during a ssDNA capture. The ratio R_V is evaluated using Eq. (1). As mentioned in the main text, the floor function '$\lfloor x \rfloor$' is used to define the integer part of any real number x.

PEG MW (Da)	Monomer units $m = \lfloor \frac{MW}{44} \rfloor$	V_{PEG} (nm^3)	R_V (Eq. 6.1)	$N = \lfloor R_V \rfloor$
200	4	0.270	167.55	167
400	9	0.607	74.53	74
1000	22	1.483	30.51	30
4000	90	6.068	7.46	7
8000	181	12.204	3.71	3
10,000	227	15.306	2.96	2

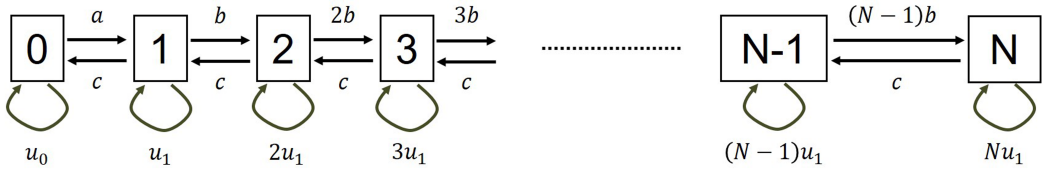


Figure 6.1: General discrete-state stochastic description for ssDNA capture in the presence of PEG crowders present on the *cis* side of the nanopore.

To study the influence of varying molecular weights (or sizes) of crowders on the capture rate of the ssDNA, a discrete-state stochastic description is implemented.[34] Figure 6.1 shows the microscopic state $n = 0, 1, \ldots, N$ where n PEG polymers of a fixed MW exist within the αHL pore. In state $n = 0$, the first PEG enters the nanopore independently with the pseudo-first-order rate $a = a_0[PEG]$. As proposed in the previous chapter,[34] the entry of the first PEG molecule facilitates the cooperative partitioning of subsequent PEGs within the nanopore due to cation-induced conformational changes within the αHL nanopore.[35–37] Assuming that the newly partitioned crowder occupies the volume V_{PEG} quickly, the subsequent addition of another PEG from state n to $n+1$ ($n > 0$) occurs sequentially in a cooperative manner with rate nb where $b = b_0[PEG]$.[34] To explore the role of crowder size variation, the second order rate constants a_0 and b_0 are crowder size dependent parameters and can be expressed as $a_0 = a_0'/R_V$ and $b_0 = b_0'/R_V$, respectively. The parameter $1/R_V$ accounts for the volume fraction that a new PEG will occupy within the nanopore and thereby incorporates the effect of the crowder size. Additional parameters a_0' and b_0' depend on other system properties such as the applied electric field, solvent viscosity, etc. but these are assumed to be constant

throughout this theoretical analysis. As discussed in Chapter 5, the entry rate of the crowder from state $n \geq 1$ is also influenced by n PEGs that already exist in the nanopore where each occupies a fixed volume $1/R_V$. Thus, an additional factor $n \times (1/R_V)$ is included in the rate expression for $n > 0$ which signifies that the existing volume fraction contributes to the cooperative addition of a new crowder molecule. Consequently, the modified PEG entry rates from states $n = 0$ and $n > 0$ are $a = a_0'[PEG]/R_V$ and $nb = nb_0'[PEG]/R_V^2$, respectively. A single crowder is independently removed from the nanopore with a universal rate c which leads to a transition from state n to $n-1$. Therefore, a single PEG's entry-exit process in the nanopore is guided only by the respective entry rates. In the absence of crowders, the ssDNA capture rate from state 0 is u_0.[34] For $n > 0$, the capture event will indirectly depend on the crowder-occupied volume in the nanopore since it is determined by the strength of electrostatic interactions between the negatively charged ssDNA and polycationic PEG polymers present within the nanopore. Thus, $u_{n\neq 0} = nu_1$ where $u_1 = u_1'/R_V$ such that u_1' is independent of the crowder size. The modified PEG capture rates from states $n > 0$ is $u_n = nu_1'/R_V$ where the quantity n/R_V reflects the volume fraction occupied by n PEGs contributing to the underlying attractive interactions leading to ssDNA capture. Following the results of the previous chapter,[34] the MW dependent average capture time of a ssDNA is written as

$$\langle \tau \rangle = \frac{1 + a_0[PEG] \sum_{i=0}^{N-1} i! \left(\frac{b_0[PEG]}{R_V} \right)^i}{u_0 + u_1 a_0[PEG] \sum_{i=1}^{N} i! \left(\frac{b_0[PEG]}{R_V} \right)^{i-1}} \tag{6.2}$$

where $a_0 = a_0'/R_V$, $b_0 = b_0'/R_V$ and $u_1 = u_1'/R_V$ are the crowder size dependent rate parameters when the homogeneous crowded conditions are introduced only on the *cis* side of the nanopore. Note that the microscopic average capture rate (or inter-event frequency, f) can be determined from $\langle \tau \rangle^{-1}$.[34] The size-dependent contribution of PEG crowders in Eq. 6.2 arrives from R_V defined in Eq. 6.1. It is one of the main results of this work and will be used to probe the role of a heterogeneous mixture of crowders in impacting the ssDNA capture process in the next section.

6.2.2 Effect of Heterogeneity in Crowded Solutions on ssDNA Capture

The cellular matrix is highly complex with multiple co-existing macromolecular crowders.[22, 30] Thus, it is critical to probe the influence of solution heterogeneity on the

capture of ssDNA within biological nanopores for their successful translocation. The simplest way to introduce inhomogeneity in this system is through a binary mixture of crowders. Let us consider two types of PEGs, labelled as PEG_1 and PEG_2 polymers with their respective molecular weights, MW_1 and MW_2, constituting the mixture such that the total monomeric concentration (in w/v %) of the solution is known. In further discussion, the subscripts 1 and 2 are used to define the properties of PEG_1 and PEG_2, respectively.

This section will also focus on the ssDNA capture process only and a similar theoretical model will be considered as introduced in the previous section. The *cis* compartment consists of a binary mixture of PEGs and individual crowder molecules are capable of sequentially partitioning within the nanopore during the ssDNA capture until they occupy the total volume V_{pore}. Let us define a state (n_1, n_2) indicating that n_1 and n_2 number of PEG_1 and PEG_2 crowders, respectively, are simultaneously present inside the nanopore: see Figure 6.2. Accordingly, the state (0,0) corresponds to the situation when the nanopore is free from PEG crowders. If $n_1, n_2 > 0$, the volume occupied by crowders should not exceed V_{pore} which is ensured by determining the fraction occupied by crowders in the state (n_1, n_2) as

$$\phi(n_1, n_2) = \frac{n_1}{R_{V,1}} + \frac{n_2}{R_{V,2}} \tag{6.3a}$$

where the ratios $R_{V,1}$ and $R_{V,2}$ are evaluated using Eq. 6.1. Since V_{pore} remains constant, $\phi(n_1, n_2)$ will be determined by the composition of the binary mixture in the heterogeneous solution. This suggests that the state (n_1, n_2) will exist only if $\phi(n_1, n_2) \leq 1$. Thus, to proceed with the theoretical analysis, a Heaviside step function is introduced

$$\Theta\big(x = 1 - \phi(n_1, n_2)\big) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{6.3b}$$

to make sure the discrete-state stochastic network is appropriately constructed considering a binary mixture. This term $\Theta(x)$ reflects the possibility of state (n_1, n_2). Using Table 6.1, one can define a maximum of N_1 (or N_2) crowders of PEG_1 (or PEG_2) that can be accommodated within the nanopore completely resulting in the state $(N_1, 0)$ (or $(0, N_2)$). Further, Eq. 6.3 ensures that state (N_1, N_2) will not exist due to the rigid structure of the nanopore with a fixed volume V_{pore} that is available for the crowders to occupy: see Figure 6.2.

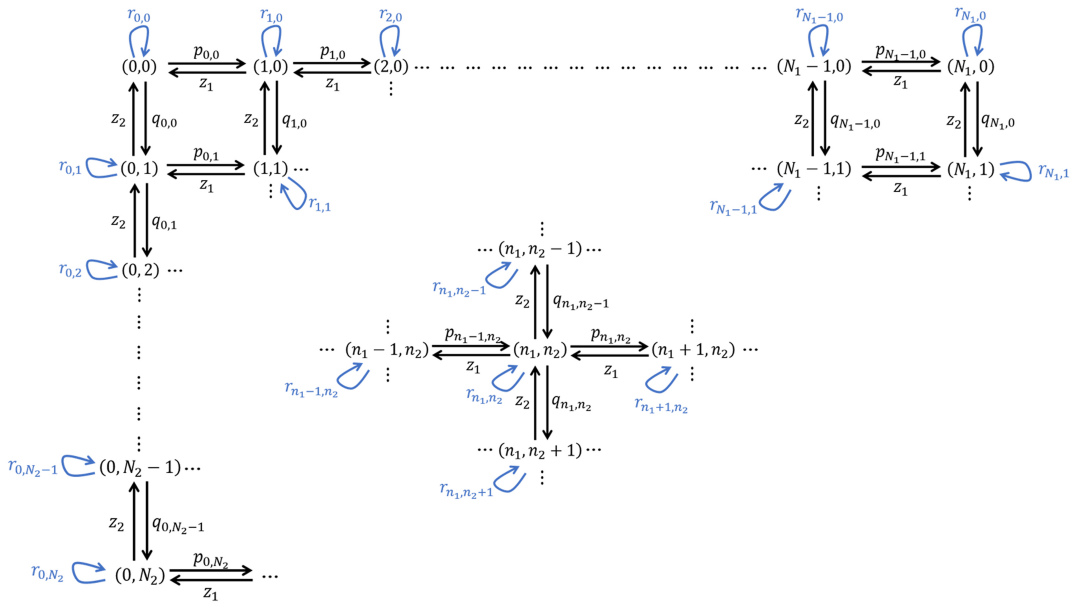


Figure 6.2: A 2D discrete-state stochastic network to describe the ssDNA capture in the nanopore influenced by a binary mixture of different PEG polymer sizes. Different transitions occurring from state (n_1, n_2) and their respective rates are discussed in the main text.

From the state (n_1, n_2), a single PEG_1 addition within the nanopore occurs with a rate p_{n_1,n_2} leading to state $(n_1 + 1, n_2)$. Similarly, a single polymer of PEG_2 partitions into the nanopore with rate q_{n_1,n_2} leading to the microscopic state $(n_1, n_2 + 1)$. Utilizing Eq. 6.3 and the rates defined in the previous discussion, p_{n_1,n_2} or q_{n_1,n_2} can be written as

$$p_{n_1,n_2} = \begin{cases} \Theta(1-\phi(1,0)) \times \frac{a_0'[PEG]_1}{R_{V,1}} & n_1 = n_2 = 0 \\ \Theta(1-\phi(n_1+1,n_2)) \times \phi(n_1,n_2) \times \frac{b_0'[PEG]_1}{R_{V,1}} & \text{otherwise} \end{cases} \tag{6.4a}$$

$$q_{n_1,n_2} = \begin{cases} \Theta(1-\phi(0,1)) \times \frac{a_0'[PEG]_2}{R_{V,2}} & n_1 = n_2 = 0 \\ \Theta(1-\phi(n_1,n_2+1)) \times \phi(n_1,n_2) \times \frac{b_0'[PEG]_2}{R_{V,2}} & \text{otherwise} \end{cases} \tag{6.4b}$$

A single PEG_1 or PEG_2 crowder independently leaves the nanopore from state (n_1, n_2) and transitions to state $(n_1 - 1, n_2)$ or $(n_1, n_2 - 1)$, respectively, with rate

$$z_1 = z_2 = c \times \Theta\big(1-\phi(n_1,n_2)\big). \tag{6.4c}$$

Additionally, the ssDNA capture event occurs from state (n_1, n_2) via rate r_{n_1,n_2} if it follows

$$r_{n_1,n_2} = \Theta\big(1-\phi(n_1,n_2)\big) \times \phi(n_1,n_2) \times u_1' \tag{6.4d}$$

while the number of crowders within the nanopore remains unchanged. All these

transitions are included in the two-dimensional (2D) discrete-state stochastic network presented in Figure 6.2. The Heaviside step function in Eq. 6.4 ensures that the transitions will occur exclusively from the states that exist according to Eq. 6.3.

Using this 2D model, one can determine the ssDNA's capture rate (or the inter-event frequency) in the presence of a binary mixture of PEG crowders on the *cis* compartment of the nanopore. Let us now define $P_{n_1,n_2}(t)$ as the probability for the system to exist in state (n_1, n_2) at time t. The evolution of these probabilities can be described by a set of forward master equations presented in detail within the Appendix D. Assuming that the system acquires a stationary state quickly, the solutions of P_{n_1,n_2} are obtained when $t \to \infty$, and one arrives at the inter-event frequency, f, from the net flux (J) using

$$f = J = \sum_{n_2} \sum_{n_1} r_{n_1,n_2} P_{n_1,n_2} \tag{6.5}$$

The details of the calculations are presented in the Appendix D. The average capture time is evaluated from the inverse of the inter-event frequency.

6.3 Results and Discussion

The capture rate of a ssDNA is calculated in a homogeneously crowded solution, i.e., when there is a single type of PEG crowder of a known MW and concentration on the *cis* side of the protein nanopore. This minimal discrete-state stochastic description finds direct application in existing single-molecule studies. The universal parameters a_0',b_0',u_0 and u_1' in Eq. 6.2 can be easily determined by fitting experimentally available data for single-molecule ssDNA capture measurements while the *cis* side of αHL nanopore contains PEG 4k crowders in varying concentrations[25]: see Fig. 6.3(a). The resultant fitting parameters, the values of N from Table 6.1 and Eq. 6.2 are then used to evaluate the inter-event frequency at fixed crowder concentration while their sizes vary. For 20 w/v %, these theoretical results show a good quantitative agreement when compared with the experimentally available data[25] as shown in Fig. 6.3(b).

A PEG crowder molecule of a given size (or MW) occupies a fixed volume within the nanopore and also, coordinates with a certain average number of positively charged counter-ions (usually K^+ ions).[36, 37] Consequently, this theory proposed that the capture event is guided by the strength of underlying electrostatic interactions between negatively

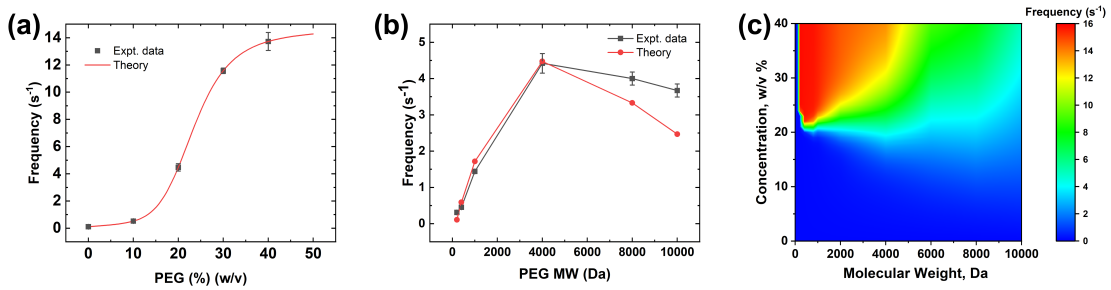


Figure 6.3: Inter-event frequency, $f = \langle \tau \rangle^{-1}$, of the ssDNA capture within the αHL nanopore as a function of (a) % w/v concentration of PEG 4k crowders and (b) MW of PEG crowders at 20 w/v % existing on the *cis* side of the nanopore. Black symbols correspond to data from single molecule experiments[25] and the red symbols are theoretically evaluated microscopic capture rates from Eq. 6.2. (c) Contour plot of the inter-event frequency for all ranges. The R_V and N values used in analytical calculations are listed in Table 6.1 for the respective MWs of PEG. The best-fit parameters used are $a_0' = 33.98$ M^{-1} s^{-1}, $b_0' = 384.49$ M^{-1} s^{-1}, $u_0 = 0.11$ s^{-1} and $u_1' = 15.95$ s^{-1}.

charged ssDNA and polycationic PEGs present in the nanopore.[34] This capture rate may be enhanced in two possible ways. (i) The overall positive charge distribution of individual crowders should be high which can be achieved by using large-size (high MW) PEGs that bind to more counter-cations. (ii) The extent of cooperative partitioning of crowders should be high which allows multiple PEGs to occupy the nanopore, i.e., the number $N \gg 1$ as defined in Eq. 6.1. These two factors act together and control the microscopic capture of ssDNA within the αHL pore. Consequently, as shown in Fig. 6.3(b), the capture rate initially increases with the MW but after PEG 4k, it begins to decrease with a further increase in crowder MW because the cooperative entrance of crowders within the nanopore weakens due to the limited volume V_{pore}. However, a deviation from the experimental values is observed for PEG 8k and PEG 10k crowders since this minimal model is built on the assumption that a single crowder fully occupies the nanopore and the possibility of partial partitioning into the pore is ignored. But in real conditions, a part of the crowder can be within the nanopore to enhance ssDNA capture and the other part of the same crowder may be hanging outside the pore. The presented theoretical model fails to account for such partial contributions which may result in an underestimation of the capture rates for large-sized crowders where only a few of them can accommodate the nanopore, see Table 6.1.

In Fig. 6.3(c), the inter-event frequency is presented as a function of the concentration and MWs of the PEG polymers using a contour plot. Following previous studies, the capture rate of the ssDNA increases with the concentration of PEG crowders due to a significant enhancement in cooperative partitioning of crowders.[25, 34] However, with an

increase in the size of crowders, the extent of this cooperativity gradually diminishes due to a decline in the required number of crowders that occupy a fixed volume of nanopore (see Table 6.1). This leads to a decrease in the inter-event frequency as the crowder MW increases.

Next, this work explores how the capture process is affected in complex crowded conditions introduced through a binary mixture of PEGs. According to Eq. 6.3, the number of possible microstates in the 2D discrete-state stochastic network (see Fig. 6.2) will be determined by the mixture components. Therefore, obtaining a general expression of the inter-event frequency is complicated and the capture rates are evaluated numerically in a binary mixture of crowders formed between PEG 400 (smaller size) and PEG 1k, PEG 4k and PEG 8k (comparatively bigger size) polymers. For any binary mixture, its composition is defined by a weight fraction, $\chi = \frac{[\text{big}]}{[\text{small}]+[\text{big}]}$, where the square brackets correspond to the respective w/v % concentrations. χ is altered by changing the concentration of the bigger crowder while the total monomeric concentration of the mixture, the denominator, remains constant. Fig. 6.4(a) shows a contour plot of inter-event frequencies for variable ranges of χ and the total monomeric concentration of the crowded solution of PEG 400 and PEG 4k mixture.

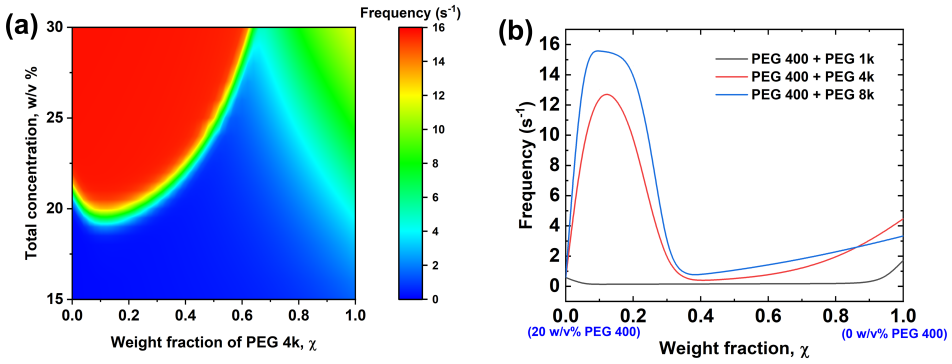


Figure 6.4: (a) Contour plot of inter-event frequencies influenced by PEG 400 and PEG 4k binary mixture plotted with respect to the total monomeric concentration of PEGs and the weight fraction, χ. (b) Frequencies of ssDNA capture event as a function of χ at total monomeric concentration 20 w/v % for a binary mixture formed between PEG 400 and PEG1k (black), PEG 4k (red) and PEG 8k (blue) polymers. The results are determined from Eq. 6.5 by numerical calculations. R_V and N values used in analytical calculations are listed in Table 6.1. The best-fit parameters used are $a_0' = 33.98$ M^{-1} s^{-1}, $b_0' = 384.49$ M^{-1} s^{-1}, $u_0 = 0.11$ s^{-1} and $u_1' = 15.95$ s^{-1}.

For all weight fractions, see Fig. 6.4(a), the inter-event frequency increases with the total monomeric concentration of the system due to enhancement in the cooperative partitioning of PEGs within the nanopore.[34] In moderately crowded conditions with fixed

total concentrations ($\sim$ 20 w/v %), a non-monotonic behaviour in the capture rates is observed depending on the value of χ. At $\chi = 0$, the capture rate corresponds to the situation when only small-size crowders (PEG 400) exist in the solution. On increasing χ slightly, i.e., on the addition of larger MW crowders, one observes a sharp rise in the capture rates. This occurs due to the selective partitioning of the smaller crowders within the nanopore compared to the bigger ones as a result of the 'polymer-pushing-polymer' concept.[31, 32] Consequently, it enhances the electrostatic interaction strength that leads to faster ssDNA capture. On further addition of high MW crowders by replacing the smaller ones (increasing χ), the 'polymer-pushing-polymer' effect weakens and the inter-event frequency declines due to weak partitioning of crowders. Such a decrease in PEG partitioning in binary mixtures has also been reported in conductance measurements from a previous experimental study.[33] It was argued that this reduction is due to the entropic penalty to move bigger PEGs inside the nanopore. When the mixture is mostly comprised of high MW crowders and they begin to enter the pore as well ($\chi \sim 1$), the positive charge distribution on the crowders increases which enhances the electrostatic interactions between the ssDNA and the PEGs. This leads to an increase in the capture rates.

The components of the binary mixture also influence the resultant capture rates. The 'polymer-pushing-polymer' effect will be weak if the size difference between the components of the binary mixture is small. Relatively large size crowder molecules will exert a stronger push on smaller-size crowders leading to higher capture rates. As shown in Figure 6.4(b), at 20 w/v % compositions, the capture rates at $\chi < 0.5$ is lower for PEG400 + PEG 1k mixture (black curve) compared to PEG 400 + PEG 4k (red line) and PEG 400 + PEG 8k mixture (blue line). One can also identify the critical composition required to observe this 'polymer-pushing-polymer' effect, as shown in Fig. D.1 in Appendix C. With PEG 400 as the smaller size crowder in this binary mixture and the total monomeric concentration fixed at 20 w/v %, the numerical results suggest that this effect is evident when the size of the bigger crowder is at least PEG 2500. After determining the hydrodynamic radii of the PEG 400 and PEG 2500 using the well-known scaling law $R_H = 0.145 \times MW^{0.571}$ Å,[38] the critical size ratio was found to be $\sim$ 2.85 for the 'polymer-pushing-polymer' effect to be observable.

Although there are no experimental or computational observations of the role of binary mixture of crowders on polymer capture, the obtained results for inter-event frequencies influenced by inhomogeneous crowding are consistent with the physical arguments that arise from the well-established 'polymer-pushing-polymer' concept, providing strong support to

this theoretical approach.

6.4 Conclusions

The unique charge distribution on polynucleotides allows biological nanopores to assist in base-pair sequencing. For a high throughput of these sensors, it is desired that the polymer capture occurs rapidly so that the frequency of translocation events increases. By adopting a minimal discrete-state stochastic framework, the previous chapter probed the capture and translocation process of a ssDNA through αHL nanopore in the presence of PEG crowders of variable concentrations.[34] In the current work, this novel theoretical approach is extended to include the role of size variation in affecting ssDNA capture. Obtained analytical results indicate that a fast capture is ensured by strong electrostatic interactions between cation-bound PEGs within the nanopore and the negatively charged ssDNA, which can be achieved in two ways. First, increasing the extent of cooperative partitioning of these crowders would strengthen the underlying electrostatic interactions. Second, introducing sufficiently large-size crowders to increase cation-coordinated interactions. In dilute crowding conditions, a weak entry rate of PEGs within the protein nanopore results in smaller capture rates. A non-monotonic behavior of the interevent frequency with respect to crowder sizes is explained based on relatively lower cooperative partitioning of very high MW crowders within a fixed volume of a nanopore. Existing single-molecule experiments also support the resulting theoretical findings.

To probe the effect of more realistic heterogeneous cytoplasmic conditions on the ssDNA capture process, the theoretical analysis was extended by introducing a binary mixture of PEG crowders in different compositions. In sufficiently crowded environments, the capture rates are enhanced as smaller-size crowders are replaced by high MW crowders. This could arise from the 'polymer-pushing-polymer' concept[32, 33] wherein bigger polymers push smaller ones within the nanopore, thereby, leading to selective partitioning of small-size crowders into the pore. This effect eventually weakens as the proportion of larger crowders increases. Eventually, as larger PEGs also begin partitioning within the nanopore, an enhancement in capture is observed due to an increase in the positive charge distribution on these bigger pseudo-cationic crowders. The 'polymer-pushing-polymer' effect is also affected by the components of this binary mixture. Similar non-monotonic partitioning at a fixed concentration of binary mixtures of PEG crowders and their dependence on the crowder fraction has also been reported experimentally.[33]

It is also essential to critically evaluate this theoretical approach. This theory is developed by assuming the PEG crowders exist only on the *cis* side of the nanopore, but in real situations, they are not restricted to being on one side of the nanopore alone. Single-molecule experiments have also been performed by placing crowders on the *trans* side as well.[24, 25] Extensive simulation studies on the translocation processes have introduced crowded conditions on both sides of the nanopore.[28, 29] This minimal model completely neglects the possibility of partial partitioning of PEG crowders within the nanopore and assumes that an entire crowder chain occupies the nanopore. Since the capture rates are directly related to the PEG molecules existing inside the nanopore, the analytical results from this theoretical model are underestimated especially in the case of larger crowders. Although the capture is field-driven, this theoretical picture assumes that the applied electric field during this capture process remains constant. The presence of macromolecular crowders changes the viscosity of the solution which may also affect the capture and translocation of the ssDNA[25] but this work has neglected the contribution of any such viscosity effects exclusively. While kinetic Monte Carlo simulations suggest that the capture rates may decrease on reaching near the capture radius,[39] this theory assumes that these rates are independent of the distance between the respective molecules and the nanopore's mouth. Regardless, it is believed that incorporating such realistic descriptions in this minimal theoretical model may quantitatively modify the results but the overall microscopic mechanism may not be affected.

In conclusion, the presented minimal stochastic discrete-state description captures the complete dynamics of ssDNA capture at αHL protein nanopore, influenced by homogenously crowded conditions. The proposed theoretical approach, additionally, provides specific predictions for the capture rates of ssDNA when the solution is crowded with a heterogeneous mixture of macromolecules. The findings of this work offer significant insights into the microscopic mechanisms of polymer capture within the nanopores, ultimately facilitating accurate genome sequencing within crowded solutions.

Bibliography

(1) Lodish, H. F. et al., *Molecular Cell Biology*; WH Freeman New York: 2006.

(2) Phillips, R. et al., *Physical Biology of the Cell*; Garland Science: 2012.

(3) Agarraberes, F. A.; Dice, J. F. Protein Translocation Across Membranes. *Biochim. Biophys. Acta, Biomembr.* **2001**, *1513*, 1–24.

(4) Muthukumar, M. Mechanism of DNA Transport Through Pores. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 435–450.

(5) Chen, P. et al. Probing Single DNA Molecule Transport Using Fabricated Nanopores. *Nano Lett.* **2004**, *4*, 2293–2298.

(6) Wong, C. T. A.; Muthukumar, M. Polymer Capture by Electro-Osmotic Flow of Oppositely Charged Nanopores. *J. Chem. Phys.* **2007**, *126*.

(7) Kasianowicz, J. J. et al. Characterization of Individual Polynucleotide Molecules using a Membrane Channel. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13770–13773.

(8) Deamer, D. W.; Branton, D. Characterization of Nucleic Acids by Nanopore Analysis. *Acc. Chem. Res.* **2002**, *35*, 817–825.

(9) Branton, D. et al. The Potential and Challenges of Nanopore Sequencing. *Nat. Biotechnol.* **2008**, *26*, 1146–1153.

(10) Zhou, W. et al. Molecular Insights into Distinct Detection Properties of α-Hemolysin, MspA, CsgG, and Aerolysin Nanopore Sensors. *J. Phys. Chem. B* **2020**, *124*, 1611–1618.

(11) Mayer, S. F. et al. Biological Nanopores for Single-Molecule Sensing. *iScience* **2022**, *25*.

(12) Ayub, M.; Bayley, H. Engineered Transmembrane Pores. *Curr. Opin. Chem. Biol.* **2016**, *34*, 117–126.

(13) Shi, W. et al. Nanopore Sensing. *Anal. Chem.* **2017**, *89*, 157–188.

(14) Mohammad, M. M. et al. Engineering a Rigid Protein Tunnel for Biomolecular Detection. *J. Am. Chem. Soc.* **2012**, *134*, 9521–9531.

(15) De Zoysa, R. S. et al. Translocation of Single-Stranded DNA Through the α-Hemolysin Protein Nanopore in Acidic Solutions. *Electrophoresis* **2011**, *32*, 3034–3041.

(16) Smeets, R. M. et al. Salt Dependence of Ion Transport and DNA Translocation through Solid-State Nanopores. *Nano Lett.* **2006**, *6*, 89–95.

(17) He, Y. et al. DNA Capture in Nanopores for Genome Sequencing: Challenges and Opportunities. *J. Mater. Chem.* **2012**, *22*, 13423–13427.

(18) He, Y. et al. Gate Manipulation of DNA Capture into Nanopores. *ACS Nano* **2011**, *5*, 8391–8397.

(19) Wanunu, M. et al. Electrostatic Focusing of Unlabelled DNA into Nanoscale Pores using a Salt Gradient. *Nat. Nanotechnol.* **2010**, *5*, 160–165.

(20) Leong, I. W. et al. Salt Gradient Control of Translocation Dynamics in a Solid-State Nanopore. *Anal. Chem.* **2021**, *93*, 16700–16708.

(21) Sohi, A. N. et al. DNA Capture by Nanopore Sensors under Flow. *Anal. Chem.* **2020**, *92*, 8108–8116.

(22) Ellis, R. J. Macromolecular Crowding: Obvious but Underappreciated. *Trends Biochem. Sci.* **2001**, *26*, 597–604.

(23) Freedman, K. J. et al. Nonequilibrium Capture Rates Induce Protein Accumulation and Enhanced Adsorption to Solid-State Nanopores. *ACS Nano* **2014**, *8*, 12238–12249.

(24) Larimi, M. G. et al. Interactions of a Polypeptide with a Protein Nanopore under Crowding Conditions. *ACS Nano* **2019**, *13*, 4469–4477.

(25) Yao, F. et al. Crowding-Induced DNA Translocation through a Protein Nanopore. *Anal. Chem.* **2020**, *92*, 3827–3833.

(26) Chau, C. C. et al. Macromolecular Crowding Enhances the Detection of DNA and Proteins by a Solid-State Nanopore. *Nano Lett.* **2020**, *20*, 5553–5561.

(27) Chau, C. et al. Probing RNA conformations using a Polymer–Electrolyte Solid-State Nanopore. *ACS Nano* **2022**, *16*, 20075–20085.

(28) Tan, F. et al. Effects of Active Crowder Size and Activity – Crowding Coupling on Polymer Translocation. *Soft Matter* **2021**, *17*, 1940–1954.

(29) Chen, Y.; Luo, K. Dynamics of Polymer Translocation through a Nanopore Induced by Different Sizes of Crowding Agents. *J. Chem. Phys.* **2013**, *138*.

(30) Zimmerman, S. B.; Trach, S. O. Estimation of Macromolecule Concentrations and Excluded Volume Effects for the Cytoplasm of Escherichia Coli. *J. Mol. Biol.* **1991**, *222*, 599–620.

(31) Aksoyoglu, M. A. et al. Size-Dependent Forced PEG Partitioning into Channels: VDAC, OmpC, and α-Hemolysin. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 9003–9008.

(32) Podgornik, R. et al. Polymers Pushing Polymers: Polymer Mixtures in Thermodynamic Equilibrium with a Pore. *Macromolecules* **2012**, *45*, 8921–8928.

(33) Gurnev, P. A. et al. Poly(ethylene glycol)s in Semidilute Regime: Radius of Gyration in the Bulk and Partitioning into a Nanopore. *Macromolecules* **2017**, *50*, 2477–2483.

(34) Punia, B.; Chaudhury, S. Microscopic Mechanism of Macromolecular Crowder-Assisted DNA Capture and Translocation through Biological Nanopores. *J. Phys. Chem. B* **2023**, *127*, 5850–5858.

(35) Reiner, J. E. et al. Theory for Polymer Analysis using Nanopore-Based Single-Molecule Mass Spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 12080–12085.

(36) Rodrigues, C. G. et al. Mechanism of KCl Enhancement in Detection of Nonionic Polymers by Nanopore Sensors. *Biophys. J.* **2008**, *95*, 5186–5192.

(37) Balijepalli, A. et al. Theory of Polymer–Nanopore Interactions Refined using Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2013**, *135*, 7064–7072.

(38) Devanand, K.; Selser, J. Asymptotic Behavior and Long-Range Interactions in Aqueous Solutions of Poly(ethylene oxide). *Macromolecules* **1991**, *24*, 5943–5947.

(39) Qiao, L. et al. An Efficient Kinetic Monte Carlo to Study Analyte Capture by a Nanopore: Transients, Boundary Conditions and Time-Dependent Fields. *Phys. Chem. Chem. Phys.* **2021**, *23*, 1489–1499.

Chapter 7

Understanding the Reaction Dynamics on Heterogeneous Catalysts Using a Simple Stochastic Approach

The content of this chapter and associated content in Appendix E are reprinted (adapted) with permission from (*J. Phys. Chem. Lett.* 2021, 12, 49, 11802-11810). Copyright © 2021 American Chemical Society.

7.1 Introduction

It is impossible to imagine modern industrial processes or scientific research activities without use of catalysts that accelerate the required chemical processes.[1–3] Despite the importance of catalysis and many years of intensive studies, however, the molecular mechanisms of catalytic phenomena still remain not well understood.[3, 4] A significant progress in development of experimental methods for probing the catalytic processes has been achieved in recent years. This is mostly due to the developments of various single-molecule techniques that allowed researchers to uncover the molecular properties of catalytic systems, which are hidden in the ensemble-averaged bulk measurements.[5, 6]

The single-molecule fluorescence microscopy experiments investigated the chemical transformations of non-fluorescent reactants into fluorescent products when single nanoparticles (NP) have been used as catalysts.[5, 7] This allowed researchers to achieve excellent spatial and temporal resolutions. The stochastic fluorescent bursts associated with the product formation events were attributed to single catalytic turnovers, and the distributions of waiting times for such events have been measured. The real-time observations of single catalytic turnovers on a single NP catalyst showed that there are wide distributions in the effective rates for the product formation and dissociation, and that there are also strong fluctuations in the time-dependent chemical activity.[5, 7–9] It has been argued that in a single NP, the intrinsic structural heterogeneity among the different active sites is responsible for such time-dependent fluctuations in the rates of catalytic product formation and dissociation. In addition, it was found that the temporal dependence of the catalytic activities is determined by the size of nanoparticles which can affect the dynamic surface restructuring.[10, 11] The study of catalysis using single-molecule fluorescence microscopy has been also successful in revealing the kinetic mechanisms for several specific enzymatic systems.[12, 13] Furthermore, these experimental methods have been used recently to investigate the single-molecule kinetics of chemical reactions on gold nanoparticles, uncovering hidden kinetic intermediates that are masked in ensemble-averaged studies.[14]

Single-molecule experimental studies of chemical processes on catalysts have collected a large amount of quantitative information, which stimulated the development of various theoretical methods.[6] The earliest theoretical attempt was based on a Langmuir-Hinshelwood approach that assumes the binding/unbinding equilibrium between the substrates and the catalyst and views all active sites as a single "effective" catalytic site.[15]

However, this approach could not explain the experimentally observed size-dependent catalytic activity, and due to its mean-field nature it also failed to account for various stochastic effects.[11] A more advanced theoretical method based on first-passage analysis of chemical dynamics,[16, 17] which could partially take into account the number of active sites and the related stochastic effects, have been proposed later.[18, 19] But this method was also too simplified in many aspects since it did not consider the details of the underlying chemical reactions or the number of intermediates for the chemical processes at each active site. Recently, a new general theoretical framework was developed for investigating the dynamics of chemical reactions with intermediate states on catalytic particles with multiple active sites.[20] It utilized a discrete-state chemical kinetic approach that takes into account the stochasticity of individual chemical reactions at each catalytic site on a single NP.[20, 21] The main idea of this method is to follow the dynamics of only those active sites that are just one step before making the final product. This allows to connect better with experimental observations and it also significantly decreases the complexity of mathematical calculations of dynamic properties since the originally very complicated multi-state system can now be described as a much simpler one-dimensional sequence of effective chemical kinetic "states."[20]

Using this discrete-state stochastic approach, it was explicitly shown that the mean reaction times are inversely proportional to the number of active sites irrespective of the details of underlying chemical reactions.[20] It was also found that the higher moments of reaction times are affected by the details of the chemical reactions at each active site. However, a comprehensive description of the reaction dynamics on multi-site catalyst has been obtained only for the simplified situations with $N = 1$, 2 and 3 active sites, while in real systems even on the smallest nanoparticles the number of active sites is significantly larger (probably $N \sim 10 - 1000$). This observation limits the applicability of the discrete-state stochastic framework for understanding the mechanisms of catalytic processes.

In this chapter, a simple discrete-state stochastic model of the catalytic chemical reactions on particles with multiple active sites is developed, which can be solved analytically for any number of active sites and for all ranges of relevant parameters. It allows for a comprehensive explicit description of the reaction dynamics on heterogeneous catalysts and to make the connections between the molecular processes and the observed dynamic features. This analysis explicitly shows the dependence of the dynamics of catalyzed chemical reactions on the number of active sites, number of intermediate chemical species and on the topology of underlying chemical reactions. It is argued that the calculated dynamic

properties present quantitative bounds for dynamic properties measured in experiments, clarifying many aspects of molecular mechanisms for chemical reactions on catalysts.

7.2 Theoretical Model

To obtain a comprehensive description of dynamic properties of catalyzed chemical reactions, consider a single nanoparticle with N active sites that might catalyze only one specific chemical reaction. Each active site is independent of each other, and for every chemical reaction there are M intermediate states, as schematically shown in Fig. 7.1a. In this process, the substrate molecule S binds to the catalytic site (labeled as C in Fig. 7.1a) to create the complex CS_1 that sequentially transitions through the M intermediate states CS_j $(1 \leq j \leq M)$. From the last complex CS_M, the final product P can be achieved that also releases the catalytic site for a new chemical process. All these transitions are irreversible and assumed to be happening with the same rate u (which sets the time scale in the system). Please note that the first transition (binding of the substrate) is different from other transitions and its rate is concentration-dependent, but for convenience it is assumed that all transitions rates are the same.

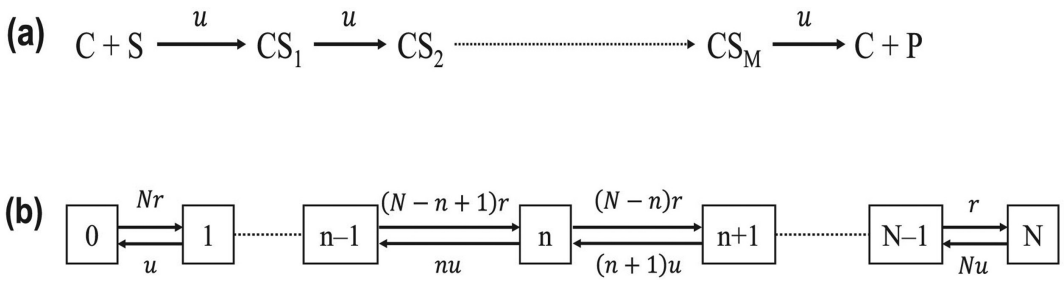


Figure 7.1: Schematic representation of chemical processes on heterogeneous catalysts with N identical active sites. (a) A sequential chemical reaction with M intermediates. (b) Overall effective chemical kinetic scheme for the whole catalytic particle. Each discrete state n corresponds to the state of the system with n $(0 \leq n \leq N)$ active sites in the chemical conformation (CS_M) just before making the product. For transitions between the states, the effective rate constant is $r = \frac{u}{M}$.

Using the idea of the mapping multiple chemical reactions catalyzed on N active sites of the single NP on a sequence of effective discrete states,[20] the effective chemical kinetic scheme for the whole catalyst with the sequential reaction is presented in Fig. 7.1b. The chemical processes on the single catalyst with N active sites can be viewed as a sequence of stochastic transitions between $N + 1$ discrete states. Each state n $(0 \leq n \leq N)$ corresponds to the effective chemical kinetic state of the system with exactly n active sites

in the conformation CS_M. This happens after M intermediate transitions at the given site are already accomplished ($CS_j \rightarrow CS_{j+1}$), and the system is ready for the final transition to make the product molecule at the given site (Fig. 7.1a). From the state n, the final product can be made with the effective rate nu because all active sites are independent. This also moves the system back into the state $n-1$ since the active site where the product was made is not in the CS_M conformation anymore. The effective forward transition ($n \rightarrow n+1$) in this kinetic scheme is taking place with a rate $(N-n)r = (N-n)u/M$ (Fig. 7.1b), where the effective rate constant r for transitions between discrete states (Fig. 7.1b) is given by

$$r = \frac{u}{M}. \tag{7.1}$$

This effective forward rate r can be explained using the following arguments. In this simplified model, all intermediate chemical states are equally probable, and the transition to CS_M will occur only from the chemical state CS_{M-1}. There are M possible intermediate chemical states that did not reach yet the state CS_M (i.e., $C, CS_1, CS_2, ..., CS_{M-1}$). This means that the probability to find the system in the chemical state CS_{M-1} from which the transition will happen is $1/M$. At the same time, for the effective state n there are $(N-n)$ active sites that did not yet reach the conformation CS_M.

This simplified theoretical model provides a minimal description of chemical reactions on catalysts with multiple active sites. It neglects the backward transitions, but in a real system, they are expected to be smaller than the forward transitions, making this simple theory reasonable enough from the experimental point of view. For simplicity, it is also assumed that $u = 1$ since it only determines the time scale in the system.

To analyze the dynamic properties of the system, a powerful method of first-passage probability densities is employed that was successfully utilized in analyzing various problems in chemistry, physics and biology. [22–25] This is also stimulated by experimental observations that count the catalytic cycle as soon as the product molecule is created for the first time at the given active site. For this purpose, one can define a function $F_n(t)$ as a probability density function to complete the catalytic cycle at time t starting at the state n at time $t = 0$. The temporal evolution of these first-passage probability densities is governed by a set of backward master equations[20, 22]

$$
\begin{aligned}
\frac{dF_0(t)}{dt} &= \frac{N}{M}F_1(t) - \frac{N}{M}F_0(t);\\
\frac{dF_1(t)}{dt} &= \frac{N-1}{M}F_2(t) + F_P(t) - \left(\frac{N-1}{M}+1\right)F_1(t);\\
&..\\
\frac{dF_n(t)}{dt} &= \frac{N-n}{M}F_{n+1}(t) + nF_P(t) - \left(\frac{N-n}{M}+n\right)F_n(t);\\
\frac{dF_N(t)}{dt} &= NF_P(t) - NF_N(t);
\end{aligned}
\tag{7.2}
$$

where $F_P(t)$ is for the product state with $F_P(t) = \delta(t)$. This physically means that if the system starts in this state, the process is immediately accomplished.[20]

This set of backward master equations can be analyzed in the Laplace space $[\widetilde{F_n}(s) = \int_0^\infty e^{-st}F_n(t)dt]$ with the initial condition $F_n(t=0) = 0$, which yields

$$
\begin{aligned}
(s+\frac{N}{M})\widetilde{F_0}(s) &= \frac{N}{M}\widetilde{F_1}(s);\\
(s+\frac{N-1}{M}+1)\widetilde{F_1}(s) &= 1 + \frac{N-1}{M}\widetilde{F_2}(s);\\
&..\\
(s+\frac{N-n}{M}+n)\widetilde{F_n}(s) &= n + \frac{N-n}{M}\widetilde{F_{n+1}}(s);\\
(s+N)\widetilde{F_N}(s) &= N
\end{aligned}
\tag{7.3}
$$

Solving these equations leads to the explicit expression for the first-passage probability density functions in a compact form:

$$
\widetilde{F_n}(s) = \frac{n + \sum_{i=1}^{N-n}(n+i)\prod_{j=1}^{i}\frac{\frac{N-n-(j-1)}{M}}{s+n+j+\frac{N-n-j}{M}}}{s+n+\frac{N-n}{M}}. \tag{7.4}
$$

Alternatively, the same result can be written as

$$
\widetilde{F_n}(s) = \frac{(n+\frac{N-n}{M}) - s\sum_{i=0}^{N-n-1}\prod_{j=0}^{i}\frac{\frac{N-n-j}{M}}{s+n+j+1+\frac{N-n-j-1}{M}}}{s+n+\frac{N-n}{M}}. \tag{7.5}
$$

These expressions for the first-passage probability density functions are crucial because they allow for a full dynamic description of chemical reactions on multi-site catalysts for a general set of parameters and conditions.

7.3 Results and Discussion

By inverting the explicit Laplace transformation expressions (see Eq. (7.5)), one can obtain the distributions of first-passage times to make a product molecule starting from any effective chemical state n. The results for the simplest model with $N = 2$ and $M = 1$ are presented in Fig. 7.2 to illustrate this approach. One can see that the distributions of first-passage times depend on the starting state of the system. Distributions that start initially from the state $n = 0$ ($F_0(t)$) are always non-monotonic because to reach the final product the system must first pass several intermediate states. However, if the process starts from the other states ($F_{n>0}(t)$), then the distributions should have a more complex shape, starting from a finite value at $t = 0$: see Figure 7.2.

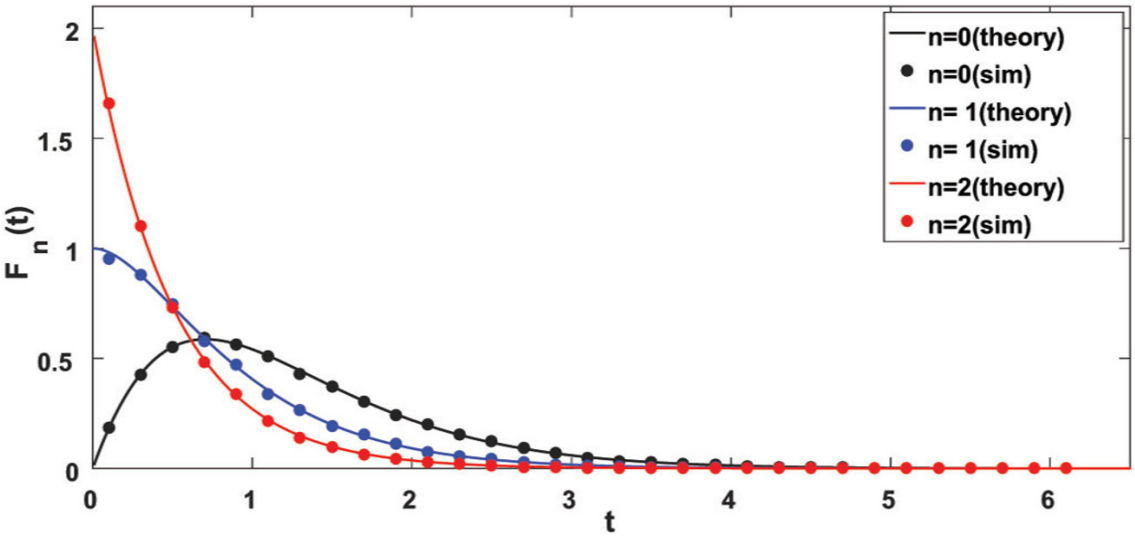


Figure 7.2: Chemical reaction time distributions for a system with $N = 2$ and $M = 1$, showing results for $n = 0$ (black), $n = 1$ (blue) and $n = 2$ (red). The solid lines are the theoretical predictions obtained by taking inverse Laplace transforms and the symbols are from Monte Carlo simulations.

Using this theoretical model, more specific dynamic information on underlying molecular mechanisms can be obtained. For example, one can explicitly estimate the mean reaction time (catalytic turnover times) and second moment of the reaction time. But these calculations involve several steps. Expanding $\widetilde{F_n}(s)$, one obtains

$$\widetilde{F_n}(s) \simeq 1 - s\langle T_n\rangle + \frac{1}{2}\langle T_n^2\rangle s^2 - .... \tag{7.6}$$

where $\langle T_n\rangle$ is the mean first-passage time and $\langle T_n^2\rangle$ is the second moment of the first-passage time if starting from the state n. Then from Eq.(7.5) it can be shown that

$$\langle T_n \rangle = -\left(\frac{\partial \widetilde{F_n}(s)}{\partial s} \right)_{s=0} = \frac{1 + \frac{N-n}{M} \langle T_{n+1} \rangle}{\frac{N-n}{M} + n}, \tag{7.7}$$

and to explicitly obtain

$$\langle T_n \rangle = \frac{1 + \sum_{k=0}^{N-n-1} \prod_{j=0}^{k} \frac{\frac{N-n-j}{M}}{n+j+1+\frac{N-n-j-1}{M}}}{n + \frac{N-n}{M}}. \tag{7.8}$$

One could also obtain $\langle T_N \rangle = \frac{1}{N}$, which can be easily explained. In the state N, all active sites are in the conformation CS_M just before making the product. The probability that the reaction will be accomplished from any active site is equal to $1/N$ and the time for this transition is 1 (because it is assumed that $u = 1$).

Similar analysis can be done for second moment of the first-passage times, leading to

$$\begin{aligned} \langle T_n^2 \rangle = \left(\frac{\partial^2 \widetilde{F_n}(s)}{\partial s^2} \right)_{s=0} = 2 \Bigg[& \frac{1 + \sum_{k=0}^{N-n-1} \prod_{j=0}^{k} \frac{\frac{N-n-j}{M}}{n+j+1+\frac{N-n-j-1}{M}}}{(n + \frac{N-n}{M})^2} \\ & + \frac{\sum_{k=0}^{N-n-1} \left(\prod_{j=0}^{k} \frac{\frac{N-n-j}{M}}{n+j+1+\frac{N-n-j-1}{M}} \sum_{j=0}^{k} \frac{1}{n+j+1+\frac{N-n-j-1}{M}} \right)}{(n + \frac{N-n}{M})} \Bigg]. \end{aligned} \tag{7.9}$$

Here one can also show that $\langle T_N^2 \rangle = \frac{2}{N^2}$. Note, however, that to obtain mean reaction times and second moment of the reaction times, the quantities that are measured in single-molecule experiments, additional calculations should be done, as explained below.[20]

Assuming that the system is already in the stationary state, P_n is defined as the steady-state probability of finding the system in the effective chemical kinetic state n. Then because the number of the discrete states in the system is finite and the steady state between these effective states will be reached at $t \to \infty$, the following relations can be written for these probabilities (see Fig. 7.1b),[20]

$$\begin{aligned} \frac{N}{M} P_0 &= P_1; \\ \frac{N-1}{M} P_1 &= 2P_2; \\ &\ldots. \\ \frac{1}{M} P_{N-1} &= N P_N; \end{aligned} \tag{7.10}$$

Considering the normalization condition, $\sum_{n=0}^{N} P_n = 1$, these equations can be easily solved to obtain the steady-state probabilities[20]

$$P_n(N) = \frac{N!}{(N-n)!n!}\frac{x^n}{(1+x)^N}, \tag{7.11}$$

where $x = \frac{1}{M}$.

This work aims to evaluate the mean reaction (catalytic turnover) times for chemical reactions described in Fig. 7.1a that are taking place on the catalysts with N active sites, $\langle\tau\rangle_N$. It is convenient to think about these processes in the way the single-molecule fluorescence experiments are done.[5, 7, 11, 14, 15, 26, 27] Every time the product molecule is formed, a spike in the fluorescence appears. Then the reaction times are the time intervals between two consecutive spikes, and the mean reaction times are measured by averaging over all pairs of intervals between consecutive spikes. In the language of this work's effective chemical-kinetic model (Fig. 7.1b), the mean reaction time is the properly weighted average over first-passage times from all the states n except the state $n = N$. This is because in the state $n = N$ all active sites are in the conformation CS_M just before making the product and none of the active sites had just produced the molecule P (had a spike in the fluorescent signal), and the reaction cannot start in this kinetic state. As a result, the mean reaction time can be written as

$$\begin{aligned}\langle\tau\rangle_N &= f_0\langle T_0\rangle + f_1\langle T_1\rangle + f_2\langle T_2\rangle + ... + f_{N-1}\langle T_{N-1}\rangle \\ &= \sum_{n=0}^{N-1} f_n\langle T_n\rangle = \frac{u+r}{Nru} = \frac{M+1}{Nu},\end{aligned} \tag{7.12}$$

where $r = u/M$. This important result predicts that the mean reaction times on such heterogeneous catalysts are proportional to the number of intermediates and inversely proportional to the number of active sites.

The coefficient f_n, which is the relative contribution of different effective states n into the reaction time, can be found from

$$f_n = \frac{(N-n)P_n(N)}{\sum_{j=0}^{N}(N-j)P_j(N)}. \tag{7.13}$$

This result can be explained in the following way. In the state n, there are $(N-n)$ active sites which are not in the conformation CS_M, and the catalytic cycle might start at

any of them. Then the coefficient f_n is the relative probability for the chemical reaction to start in the state n normalized over all possible starting situations. These relative probabilities can be estimated by substituting Eq. 7.11 into Eq. 7.13, producing

$$f_n = \frac{(N-1)!}{(N-n-1)!n!}\frac{x^{\mathrm{n}}}{(1+x)^{N-1}} = P_n(N-1). \tag{7.14}$$

with $f_N = 0$, as expected.

In Fig. 7.3a, the results for the mean reaction times are presented as a function of the number of active sites N for the fixed number of intermediates M. As expected, the mean reaction times decrease with N because more chemical reactions can take place at more active sites, decreasing the time intervals between successful fluorescence events. Figure 7.3b shows the mean reaction times as a function of the number of intermediates M for the fixed number of active sites N. Here the mean reaction times increase with the number of intermediates because more intermediate transitions must happen before the product is made. A linear dependence on M is observed because all chemical transitions have the same rates and only the forward steps are considered in these chemical reactions.

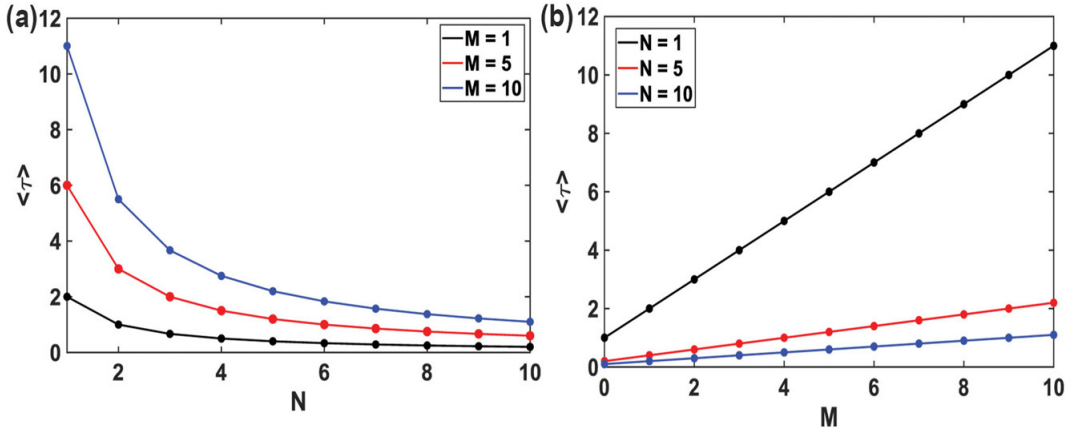


Figure 7.3: Mean reaction times as a function of (a) the number of active sites N for $M = 1$ (black line), $M = 5$ (red line) and $M = 10$ (blue line) intermediate chemical conformations; and (b) the number of intermediates M for $N = 1$ (black line), $N = 5$ (red line) and $N = 10$ (blue line) active sites. The solid lines are the theoretical predictions and the symbols are from Monte Carlo simulations.

Similar analysis can be done for second moment of the reaction times, and one can write

$$\langle \tau^2 \rangle_N = \sum_{n=0}^{N-1} f_n \langle T_n^2 \rangle. \tag{7.15}$$

Together, Eqs. 7.11 − 7.15 provide exact analytical description for mean reaction times and for second moment of reaction times for arbitrary sets of parameters.

To understand the molecular mechanisms of complex chemical and biological processes, one frequently utilizes a dimensionless parameter known as the randomness parameter, R.[22, 28] This is the quantity that measures the degree of stochastic fluctuations and it provides the information on the underlying chemical kinetic scheme.[22] The smaller the deviation of this parameter from unity, the smaller the degree of stochastic fluctuations in the system. In addition, $1/R$ gives the bound for the number of rate-limiting chemical states for the sequential kinetic schemes. There are several closely related definitions of this parameter. This work employs the following one,

$$R_{N,M} = \frac{\langle \tau^2 \rangle_N - \langle \tau \rangle_N^2}{\langle \tau \rangle_N^2}, \tag{7.16}$$

which views the randomness as the normalized variance of the reaction times.

Fig. 7.4 presents the results of the explicit calculations for the randomness parameters as a function of the number of intermediate chemical states M for variable number of active sites N. In all situations, the dependence is non-monotonic, starting from $R = 1$ for a non-catalytic reaction having $M = 0$, going through the minimal value at $M = 1$ and then approaching again to unity for $M \to \infty$. While the general non-monotonic behavior is observed in all conditions, increasing the number of active sites decreases the deviation of the randomness from one, which agrees with the reported experimental trends.[11] The most surprising observation here is that the maximal degree of stochastic fluctuations, measured via the normalized variances of the reaction times, is achieved for the systems with chemical reactions having only one intermediate state ($M = 1$), and adding more intermediates ($M > 1$) decreases the noise, in contrast to naive expectations.

These results can be explained by this study's simple theoretical model. For $M = 0$, which corresponds to a non-catalytic reaction in the effective chemical kinetic scheme (Fig. 7.1b) only one state $n = N$ is possible, i.e., all active sites are always in the state just before making the product. In this case,

$$\langle \tau \rangle_N = \langle T_N \rangle = \frac{1}{N}, \quad \langle \tau^2 \rangle_N = \langle T_N^2 \rangle = \frac{2}{N^2}, \tag{7.17}$$

which after substituting into Eq. (7.16) gives $R(N, M = 0) = 1$. The randomness param-

eter is equal to 1 because there is only one effective chemical state in the system. The product molecules can be created in one step from any of the N active sites. In the limit of $M \gg 1$, the system is also mainly found in only one effective chemical state, but this time it corresponds to $n = 0$ (see Fig. 7.1b) since it is almost impossible for any active site to reach the conformation CS_M before the production of molecules P. This again leads to $R(N, M \to \infty) = 1$. For any intermediate values of M, several effective chemical states are possible and the randomness parameter is lower than one, leading to the non-monotonic behavior shown in Fig. 7.4. But the system mostly prefers to be found in the states closer to $n = N$ for any $M > 1$. The situation for $M = 1$ corresponds to some kind of the intermediate situation. It is more probable to find the system in the state $n \simeq N/2$, but all other states in the system can also be reached. Because the larger number of chemical states is typically explored for $M = 1$, the randomness parameter deviates in the maximal way from unity, indicating the maximal level of stochastic fluctuations in the reaction times for this case.

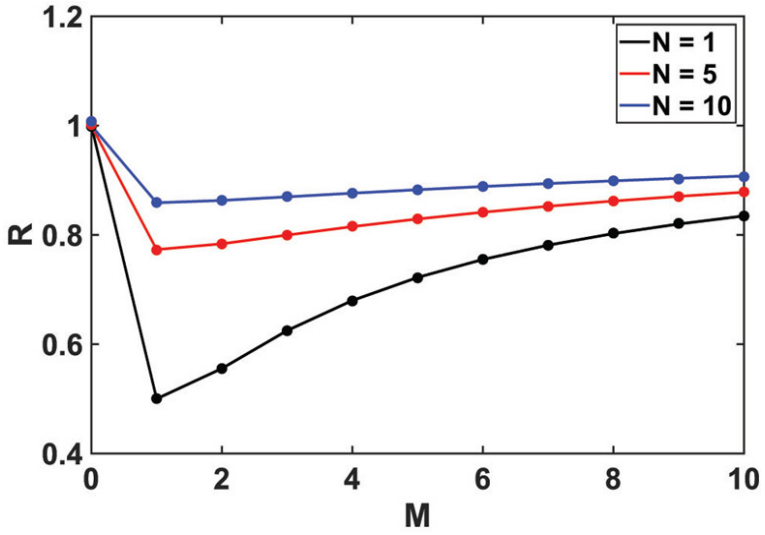


Figure 7.4: Randomness parameter as a function of the number of intermediate chemical states M for $N = 1$ (black line), $N = 5$ (red line) and $N = 10$ (blue line) active sites. The solid lines are the theoretical predictions and the symbols are from Monte Carlo simulations.

The dependence of the randomness parameter on the number of active sites is illustrated in Fig. 7.5. This parameter is always an increasing function of N with $R(N, M) \geq 0.5$ and $R(N \to \infty, M) \to 1$. One can notice that for very large number of active sites ($N \sim 100$) the randomness parameter becomes independent of the number of intermediates M. This suggests that experimental measurements of the randomness can only distinguish the mechanisms of underlying chemical reactions if catalysts have a relatively small number of active sites so that the stochastic effects might show up.

These predictions are expected to be valid even for more general mechanisms of catalyzed chemical reactions.

These observations can be explained because this theoretical approach gives the analytical solutions for all dynamic properties. For $N = 1$ active site, only two chemical states, $n = 0$ and $n = 1$, are possible. Then one can obtain from Eqs. 7.8 and 7.9,

$$\langle T_0 \rangle = (M + 1), \quad \langle T_1 \rangle = 1, \tag{7.18}$$

and

$$\langle T_0^2 \rangle = 2(M^2 + M + 1), \quad \langle T_1^2 \rangle = 2, \tag{7.19}$$

From Eq. (7.13), it can be shown that $f_0 = 1$, and this gives

$$\langle \tau \rangle_1 = \langle T_0 \rangle = (M + 1), \quad \langle \tau^2 \rangle_1 = \langle T_0^2 \rangle = 2(M^2 + M + 1). \tag{7.20}$$

Substituting this into Eq. (7.16) finally leads to

$$R(1, M) = \frac{M^2 + 1}{(M + 1)^2}. \tag{7.21}$$

Thus, for $N = 1$, the randomness parameter is lowest for small number of intermediates, e.g., $R(1, 1) = 0.5$, $R(1, 2) = 0.55$ and $R(1, 3) = 0.625$. For large number of intermediates, $M \gg 1$, the randomness parameter approaches unity, as explained above.

In the limit of large number of active sites ($N \to \infty$), it can be argued that the stationary distributions of chemical states in this effective scheme is peaking around a specific value n^* which can be found from the condition that

$$\frac{\partial P_n(N)}{\partial n} = 0 \tag{7.22}$$

at $n = n^*$. Explicit calculations presented in the Appendix E show that for $N \gg 1$

$$n^* \simeq \frac{N}{M + 1}, \quad P_{n*}(N \to \infty) \simeq 1. \tag{7.23}$$

Then, as shown in the Appendix E, the mean reaction times and the second moment of the reaction times are given by

$$\langle \tau \rangle_N \simeq \frac{M+1}{N}, \quad \langle \tau^2 \rangle_N \simeq 2\left(\frac{M+1}{N}\right)^2. \tag{7.24}$$

Applying these expressions in Eq. (7.16), eventually leads to $R(N \to \infty, M) \to 1$.

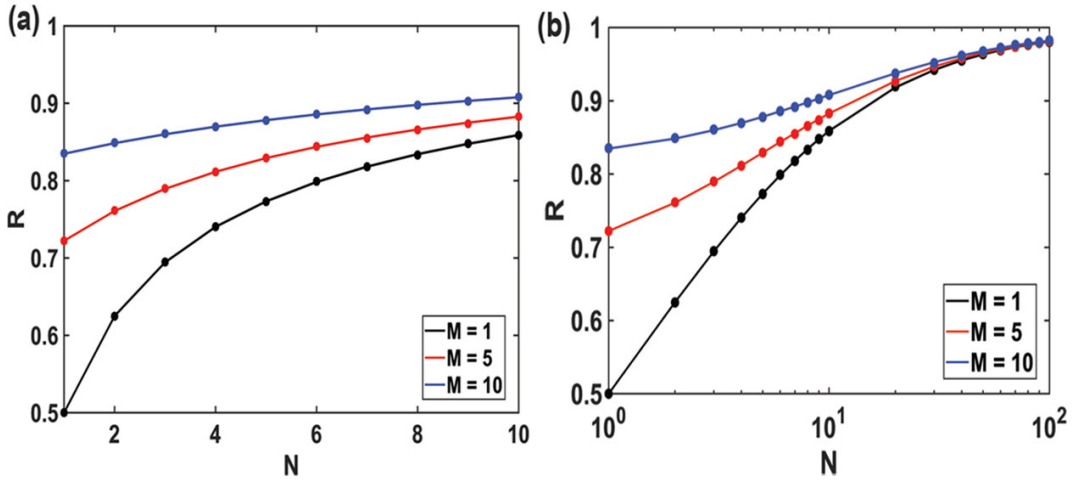


Figure 7.5: Randomness parameter as a function of the number of active sites for $M = 1$ (black line), $M = 5$ (red line) and $M = 10$ (blue line). a) Results are shown for $N \leq 10$; b) Results are shown for $N \leq 100$. The solid lines are the theoretical predictions and the symbols are from Monte Carlo simulations.

The main advantage of this theoretical approach is that all dynamic properties can be explicitly evaluated at all conditions, thus providing an explaination for the physics of the observed phenomena. At the same time, real chemical reactions on catalysts differ from the assumptions made in this model: all transitions are reversible and the rates are not uniform. In addition, more complex topology of chemical reaction pathways can also be observed. This limits the direct application of the simplest model that is presented above for analysis of experimental observations. However, one can argue that the explicit results can be used as bounds in determining the important dynamic features of catalytic processes, providing some important microscopic information. Because the presented simple model neglected the reverse chemical transitions and assumed that all rates are the same, the calculated catalytic turnover times give the lowest bound to the actual mean reaction times. Adding the backward steps and making the rates non-uniform should only slow down the processes in the system. Similarly, the calculated randomness parameter provides the upper bound to the parameter that can be measured in experiments because in real systems more stochasticity is expected.

To illustrate the potential application of this theoretical approach, consider a recent experimental study on nanoparticle catalysis.[14] In this work, single-molecule fluorescence microscopy has been utilized to investigate the oxidative deacetylation of organic molecules by hydrogen peroxide on gold nanoparticles. It was found that there is a single intermediate in this process, which corresponds to $M = 1$ in this work's notations. Using the reported distributions of reactions times, it may be estimated that the lowest randomness parameter value in this system is $R \sim 0.87$. Then, this theory predicts that in the simplest model this would correspond to $N \sim 11$. This means that in those gold nanoparticles there are, at least, 11 active sites where the reaction is taking place. These arguments show how this theoretical method can be applied for a better understanding of real catalytic processes.

Another advantage of the current theoretical approach is that it can be extended to more complex chemical reactions. To illustrate this and to understand the role of more complex chemical mechanisms that differ from the simplest sequential scheme in Fig. 7.1a, consider a chemical reaction that involves reversible branching transitions at all intermediate chemical states. [21, 22] At each intermediate state CS_j $(j = 0, 1, ..., M)$ of the chemical reaction, a reversible binding transition to the state CS_j^* can occur with v_1 and v_2 as the forward and backward transition rates, respectively (see Fig. 7.6a). Note that $j = 0$ case $(CS_0 \equiv C)$ describes the empty active site without any substrate and also assume that it might be in one of two configurations (CS_0 or CS_0^*). The substrate can only bind when the active site in the conformation CS_0 (Fig. 7.6a). For any active site, the conformation CS_M (before making the final product) can be achieved via a series of sequential forward steps that can be occasionally slowed down by reversible branching transitions. One can again build an effective chemical kinetic model with N sequential states, as shown in Fig. 7.6b, but with different effective rate constants $u' = u + v_1$ and r. As explained in the Appendix E, the effective rate constant r is given by

$$r = \frac{u + v_1}{M + (M + 1)v_1/v_2}. \tag{7.25}$$

This means that the same analysis that is developed for the sequential model can be directly applied for the model with reversible branched transitions by proper rescaling of the expressions. Thus, one can obtain an explicit description of dynamic properties for catalysis of branched chemical reactions. For example, for the mean reaction times for the branching chemical reactions it can be shown that

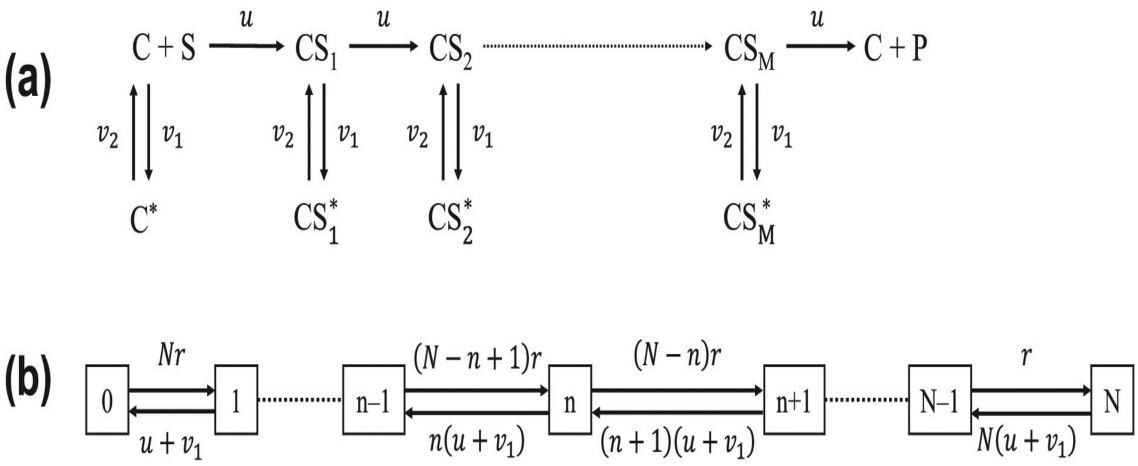


Figure 7.6: Schematic representations of chemical processes on heterogeneous catalysts with N identical active sites for (a) a branched chemical reaction with M intermediates. (b) Overall effective chemical kinetic scheme for the whole catalyst with the following effective rate constant $r = \frac{u+v_1}{M+(M+1)v_1/v_2}$.

$$< \tau >_N = \frac{(M+1)(1+v_1/v_2)}{Nu}. \tag{7.26}$$

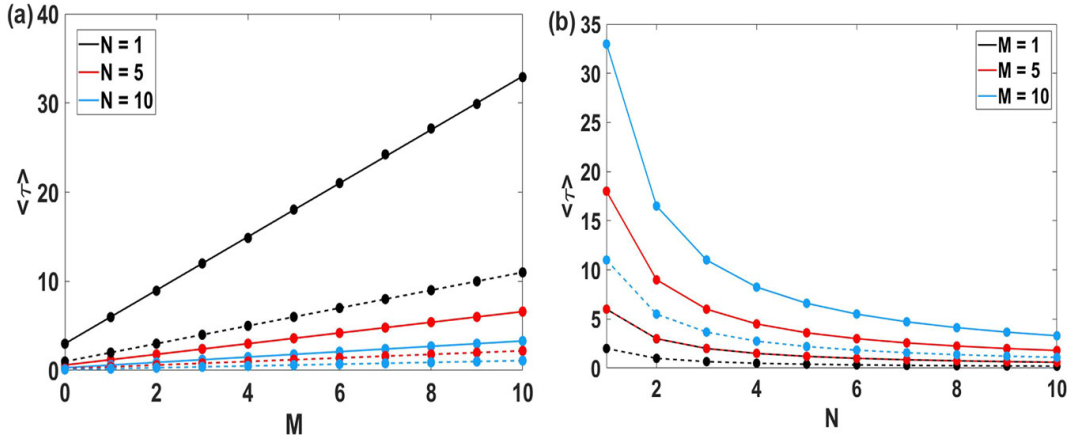


Figure 7.7: Mean reaction times as a function of (a) the number of intermediate chemical states M for $N = 1$ (black line), $N = 5$ (red line) and $N = 10$ (blue line) active sites; and (b) the number of active sites N for $M = 1$ (black line), $M = 5$ (red line) and $M = 10$ (blue line) intermediate chemical conformations

for reaction schemes involving branching. The branching transition rates used for calculations are $v_1 = 2\ s^{-1}$ and $v_2 = 1\ s^{-1}$. The solid lines are theoretical predictions for the branching model, the dashed lines are theoretical predictions for the sequential model, and the symbols are from Monte Carlo simulations of corresponding processes.

Fig. 7.7a shows the mean reaction times as a function of the number of intermediates M for the fixed number of active sites N and as a function of the number of active sites N for the fixed number of intermediates M as illustrated in Fig. 7b. The behavior of the mean reaction times is similar for both sequential and branching models. But for the fixed values of N and M, the mean reaction times are longer in the system with

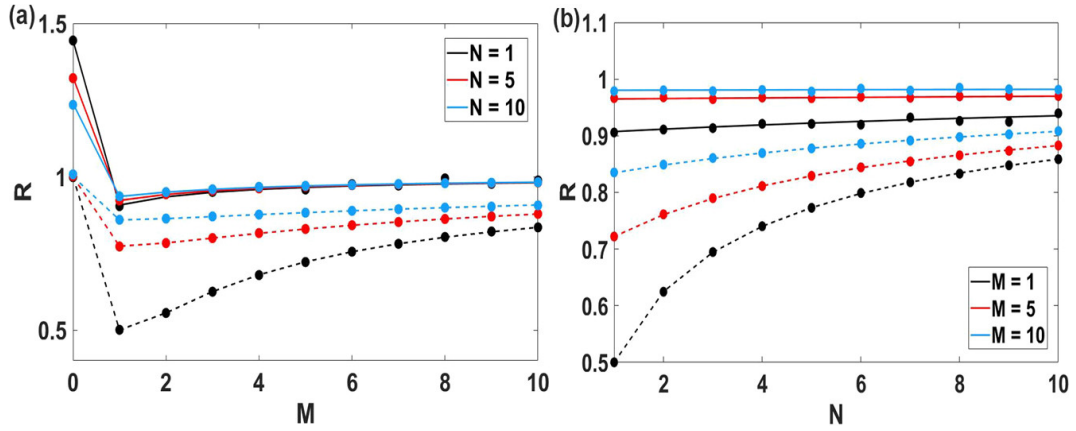


Figure 7.8: Randomness parameter as a function of (a) the number of intermediates M for $N = 1$ (black line), $N = 5$ (red line) and $N = 10$ (blue line) active sites; and (b) the number of active sites N for $M = 1$ (black line), $M = 5$ (red line) and $M = 10$ (blue line) intermediate chemical conformations for reaction schemes involving branching. The branching transition rates used for calculations are $v_1 = 2\ \mathrm{s}^{-1}$ and $v_2 = 1\ \mathrm{s}^{-1}$. The solid lines are theoretical predictions for the branching model, the dashed lines are theoretical predictions for the sequential model, and the symbols are from Monte Carlo simulations of corresponding processes.

branching reactions, as expected. This happens because the system can stay longer in those branched intermediate states (CS_j^*) that delay the overall process of reaching the conformation CS_M. As shown in the Appendix E, using the first-passage time formalism, the higher moments and the randomness parameter can be explicitly calculated for this branching model.

Figure 7.8 shows the dependence of the randomness parameter R on the number of intermediates and the number of active sites for sequential and branching chemical reactions. One can see that the randomness parameter is larger for the reactions with branching at the intermediate states. This corresponds to smaller stochastic fluctuations in the system (R is larger) for the system with the branched reactions. At first, this result seemed to be counter-intuitive since in the branched model larger number of chemical states is explored, and one could expect stronger stochastic fluctuations. To understand this, one might recall that the effective transition rate r (see Eq. 7.25) is smaller for the model with branches, at least for the parameters that this work is using. This suggests that the system with branching transitions fluctuates less between the effective discrete states (Fig. 7.6b) than the system that follows the sequential mechanism (Fig. 7.1b). This is the reason for smaller fluctuations in the mean reaction times for the system with branching transitions.

7.4 Conclusions

This study developed a simple theoretical approach to analyze the reaction dynamics of chemical reactions on heterogeneous catalysts with multiple active sites. The theoretical method is based on a discrete-state stochastic description that allows for an explicit evaluation of dynamic properties of the system via stationary-state and first-passage probability density calculations. This work investigated two different chemical reaction mechanisms occurring at each active site of the nanocatalyst. Specific attention was on analyzing the mean reaction times and the randomness parameter that describes the degree of stochastic fluctuations in the underlying chemical processes. It is found that the mean reaction times are proportional to the number of intermediate states M and inversely proportional to the number of active sites N. The effect of varying the parameters N and M is more complex for higher moments of the reaction times. It is also found that the randomness parameter, which is a measure of stochastic fluctuations in the system, non-monotonically depends on the number of intermediates, while it is always increasing with the number of active sites. This work predicted and explained why the maximal stochastic noise is observed for chemical reactions with only one intermediate state. In addition, when the chemical reactions involve reversible off-pathway branching transitions, the same qualitative behavior as for the model without branches is observed. But both the mean reaction times and the randomness are higher. These observations are explained by the fact that in the branched model more chemical states are visited, increasing the time needed to complete the chemical reaction at the given site. But the increased randomness for the model with branches correspond to smaller stochastic fluctuations which is explained by smaller transition rates between the effective chemical kinetic states of the system. It is also suggested that the explicit results for dynamic properties in this theoretical approach can be viewed as bounds for the dynamic properties of the real catalytic systems. Furthermore, the obtained scalings of dynamic properties as the function of number of active sites or number of intermediates are expected to hold for more general chemical reaction mechanisms. It is also shown explicitly how the presented theoretical analysis can be applied to obtain the important microscopic information for the underlying chemical processes.

While the presented theoretical model can be fully solved analytically, giving a better understanding of the molecular mechanisms of chemical reactions on catalysts, there are many simplifying assumptions in this model that restrict its direct application for analyzing the experimental data. The model neglects the backward chemical transitions, and

it also assumes that all chemical transition rates are the same. It will be important to generalize this simple approach to make it more realistic while still being able to solve it explicitly since this provides a better understanding of the microscopic picture in catalytic reactions. Testing the scaling predictions with various experimental methods will be also an important step in clarifying the complex mechanisms of chemical reactions on heterogeneous catalysts.

Bibliography

(1) Somorjai, G. A.; Li, Y., *Introduction to Surface Chemistry and Catalysis*; John Wiley & Sons: New York, 2010.

(2) Ross, J. R., *Heterogeneous Catalysis: Fundamentals and Applications*; Elsevier: Amsterdam: The Netherlands, 2011.

(3) Friend, C. M.; Xu, B. Heterogeneous Catalysis: A Central Science for a Sustainable Future. *Acc. Chem. Res.* **2017**, *50*, 517–521.

(4) Zhang, Z. et al. Ensembles of Metastable States Govern Heterogeneous Catalysis on Dynamic Interfaces. *Acc. Chem. Res.* **2020**, *53*, 447–458.

(5) Chen, P. et al. Spatiotemporal Catalytic Dynamics within Single Nanocatalysts Revealed by Single-Molecule Microscopy. *Chem. Soc. Rev.* **2014**, *43*, 1107–1117.

(6) Makarov, D. E., *Single Molecule Science: Physical Principles and Models*; CRC Press: 2015.

(7) Xu, W. et al. Single-Molecule Nanocatalysis Reveals Heterogeneous Reaction Pathways and Catalytic Dynamics. *Nat. Mater.* **2008**, *7*, 992–996.

(8) Xu, W. et al. Probing the Catalytic Activity and Heterogeneity of Au-Nanoparticles at the Single-Molecule Level. *Phys. Chem. Chem. Phys.* **2009**, *11*, 2767–2778.

(9) Ochoa, M. A. et al. Single Turnover Measurements of Nanoparticle Catalysis Analyzed with Dwell Time Correlation Functions and Constrained Mean Dwell Times. *J. Phys. Chem. C* **2013**, *117*, 19074–19081.

(10) Burda, C. et al. Chemistry and Properties of Nanocrystals of Different Shapes. *Chem. Rev.* **2005**, *105*, 1025–1102.

(11) Zhou, X. et al. Size-Dependent Catalytic Activity and Dynamics of Gold Nanoparticles at the Single-Molecule Level. *J. Am. Chem. Soc.* **2010**, *132*, 138–146.

(12) Lu, H. et al. Single-Molecule Enzymatic Dynamics. *Science* **1998**, *282*, 1877–1882.

(13) English, B. P. et al. Ever-Fluctuating Single Enzyme Molecules: Michaelis Menten Equation Revisited. *Nat. Chem. Biol.* **2006**, *2*, 87–92.

(14) Shen, H. et al. Single-Molecule Kinetics Reveals a Hidden Surface Reaction Intermediate in Single-Nanoparticle Catalysis. *J. Phys. Chem. C* **2014**, *118*, 26902–26911.

(15) Xu, W. et al. Single-Molecule Kinetic Theory of Heterogeneous and Enzyme Catalysis. *J. Phys. Chem. C* **2009**, *113*, 2393–2404.

(16) Chaudhury, S. et al. Universality of Poisson Indicator and Fano Factor of Transport Event Statistics in Ion Channels and Enzyme Kinetics. *J. Phys. Chem. B* **2013**, *117*, 503–509.

(17) Chaudhury, S. Poisson Indicator and Fano Factor for Probing Dynamic Disorder in Single-Molecule Enzyme Inhibition Kinetics. *J. Phys. Chem. B* **2014**, *118*, 10405–10412.

(18) Das, A.; Chaudhury, S. Modeling the Heterogeneous Catalytic Activity of a Single Nanoparticle Using a First Passage Time Distribution Formalism. *Chem. Phys. Lett.* **2015**, *641*, 193–198.

(19) Singh, D.; Chaudhury, S. A Stochastic Theoretical Approach to Study the Size-Dependent Catalytic Activity of a Metal Nanoparticle at the Single Molecule Level. *Phys. Chem. Chem. Phys.* **2017**, *19*, 8889–8895.

(20) Chaudhury, S. et al. Theoretical Investigations of the Dynamics of Chemical Reactions on Nanocatalysts with Multiple Active Sites. *J. Phys. Chem. Lett.* **2020**, *11*, 2330–2335.

(21) Kolomeisky, A. B.; Fisher, M. E. Periodic Sequential Kinetic Models with Jumping, Branching and Deaths. *Phys. A: Stat. Mech. Appl.* **2000**, *279*, 1–20.

(22) Kolomeisky, A. B., *Motor Proteins and Molecular Motors*; CRC press: New York, 2015.

(23) Metzler, R. et al., *First-Passage Phenomena and their Applications*; World Scientific: 2014; Vol. 35.

(24) Van Kampen, N. G., *Stochastic Processes in Physics and Chemistry*; Elsevier: 1992; Vol. 1.

(25) Redner, S., *A Guide to First-Passage Processes*; Cambridge university press: 2001.

(26) Xu, W. et al. Chemistry and Properties of Nanocrystals of Different Shapes. *Chem. Rev.* **2009**, *2*, 911–922.

(27) Ye, R. et al. Analogy between Enzyme and Nanoparticle Catalysis: A Single-Molecule Perspective. *ACS Catal.* **2019**, *9*, 1985–1992.

(28) Yang, S. et al. Quantitative Interpretation of the Randomness in Single Enzyme Turnover Times. *Biophys. J* **2011**, *101*, 519–524.

Chapter 8

Microscopic Mechanisms of Cooperative Communications within Single Nanocatalysts

The content of this chapter and associated content in Appendix F are reprinted (adapted) with permission from (*Proc. Natl. Acad. Sci. USA*, 2022, 119, e2115135119). Copyright © 2022 licensed under CC BY-NC-ND.

8.1 Introduction

Catalysis is a set of experimental approaches that are crucial for fundamental research as well as for multiple technological applications.[1, 2] In these methods, generally slow chemical reactions are significantly accelerated by adding small amounts of specific compounds, called catalysts, that are not consumed in these processes. Although catalytic phenomena have been known and intensively studied for a long time, many aspects of these complex physical-chemical processes are still not well understood.[2–4] It is especially difficult to fully characterize the heterogeneous catalysis when reactants, products, and catalysts operate in distinct thermodynamic phases.[5]

In recent years, nanoparticles have become important heterogeneous catalysts with multiple applications ranging from synthesis of medically important compounds to energy storage.[3, 6, 7] The intrinsic heterogeneity of nanocatalysts, however, complicates understanding the mechanisms of chemical processes in these systems using ensemble bulk measurements. More microscopic information can be obtained from single-molecule experimental approaches.[8–14] These studies showed that there is a wide distribution in the rates of product formations and dissociations from the nanocatalysts. In addition, there are strong temporal fluctuations in chemical reaction rates that depend on the size of the nanoparticles.[15, 16] However, the most surprising observation in nanocatalysts is an apparent communication between different catalytic sites.[16] Artificially similar phenomena have been observed in enzymes when catalyzed reactions occurring at different sites communicate cooperatively over distances of up to a few nanometers.[17, 18] It has been argued that this is a result of allosteric processes due to local structural modifications near the active site that might propagate along the protein structures. However, in nanocatalysts, the catalytic communication definitely cannot follow the same mechanisms because these catalysts are rigid solid bodies with fixed structures. In addition, communications in nanocatalysts have been observed at much larger distances reaching hundreds of nanometers.[16, 18]

To understand the microscopic mechanisms of catalytic cooperativity, single-molecule fluorescence localization microscopy has been utilized to evaluate the correlations between the subsequent time-dependent reactions occurring at different locations for single Pd- or Au-based nanocatalysts.[16] In these experiments, several catalyzed reactions have been monitored in real time with high temporal and spatial resolutions on nanoparticles consisting of metal cores covered by mesoporous silica shells. By dividing the nanocatalysts

into segments, the correlations between reaction times at different active sites have been measured. Surprisingly, it was found that spatially close active sites communicate with each other; the reaction that starts at one of them would have a higher probability to be followed soon by the same reaction at the neighboring active site. Interestingly, such correlations extend over distances $\sim$ 100 to 600 nm, persisting for times $\sim$ 10 to 100 s. Another striking observation was an apparent universality of intraparticle catalytic communications; for a given nanocatalyst and chemical reaction, the correlation lengths and times are independent of segmentation size, laser intensity, and catalytic activity as well as the reactants concentrations. Further experiments uncovered that communications between different active sites on nanocatalysts involve positively charged messenger species, most probably charged holes on the surfaces of the nanoparticles.[16] This was proven by applying the external electric fields in the direction opposing or supporting the migration flux of these particles along the catalytic nanorods.

Although experiments clearly connected the intraparticle catalytic communications with the motion of positively charged holes, many fundamental questions on the microscopic mechanisms of cooperativity in nanocatalysts remain unanswered. More specifically, why and how are catalytic communications accomplished by charged holes? What drives them between the different active sites? What is the microscopic origin of universality of the catalytic communications? How do we quantify the correlation lengths, memory times, and the effect of external electric fields?

To answer these questions, this chapter discusses a theoretical framework to quantitatively describe the catalytic communications phenomena from the microscopic point of view. The main idea is that the probability for the given active site to fire up depends on the local concentration of charged holes, and each catalyzed chemical reaction produces additional holes at this location. This creates a gradient of charged holes that drives them into the neighboring segments, temporarily increasing the local concentrations of holes there and stimulating new catalyzed reactions at closely located active sites. The memory times correspond to lifetimes of charged holes in the nanocatalysts, while the communication lengths are determined by the average distances traveled by the charged holes before they disappear. Using this hypothesis, this study develops a discrete-state stochastic model of the catalytic cooperativity in nanocatalysts that can be exactly solved, providing quantitative estimates for correlation lengths, memory times, and the effect of external electric fields. The theoretical predictions fully explain all experimental observations and they also allow one to make experimentally testable predictions.

8.2 Theoretical Model

Consider a single nanocatalyst particle at which chemical reactions of only one type are taking place at active sites as illustrated in Fig. 8.1a. These processes are redox reactions that involve the transfer of charged species. Since the charged holes are important for catalytic properties of active sites[1, 4, 19–22], it is hypothesized that the probability for the redox chemical reaction to fire up at the given site might be a function of the local concentration of charged holes.[22, 23]. In addition, it is assumed that each reaction produces new charged holes with a rate k (Fig. 8.1a). This leads to the increased local concentration of charged holes, and they start to diffuse away to other segments with lower local concentrations. This migration process is taking place with a diffusion rate u. Due to the presence of charged traps and other mechanisms[24–27], however, the lifetimes of these charged messengers are finite, and they disappear with a rate r (Fig. 8.1a). The combination of diffusion and death creates a region of the increased local density of charged holes around the active site where these messengers were initially produced. This means that there is an increased probability for the reaction to occur on the active sites in the spatially close region to the original active site. It is postulated that the size of such region corresponds to the correlation length, while the mean lifetime of the charged holes determines the memory time of the catalytic communications. This is the main idea of this theoretical approach.

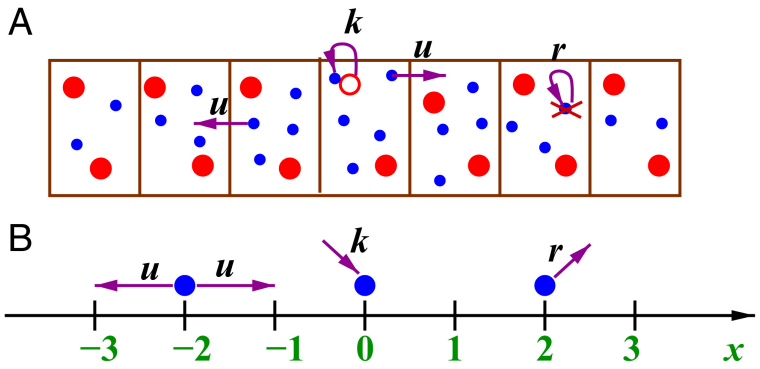


Figure 8.1: (a) A schematic view of a catalytic nanorod with multiple active sites (red filled circles). The nanocatalyst is divided into equal-sized segments. The probability for the reaction to happen at a given site is proportional to the local concentration of charged holes (blue circles). When the reaction is taking place (at the red unfilled circle), the holes are created with a rate k. Charged holes can diffuse along the nanocatalyst with a rate u, and eventually, they will disappear with a rate r. (B) A corresponding discrete-state stochastic model of the catalytic cooperativity. Lattice sites $x = 0, \pm 1, \pm 2, \ldots$ label different segments of the nanorod. The charged hole appears first at the segment $x = 0$ at time $t = 0$. It diffuses along the lattice with the rate u until it dies with rate r.

To quantify these theoretical arguments, a discrete-state stochastic model is introduced as presented in Fig. 8.1b. Since most investigated catalytic particles were nanorods, a one-dimensional description is adopted. The lattice describes the nanorod divided into segments, and the distance between neighboring lattice segments, l, corresponds to the size of the segment $\sim$ 100 nm in experimental settings.[16] Note that there might be multiple active sites in each segment. To simplify calculations, a single-particle view of the dynamics in the system is also adopted. At time $t = 0$, the charged hole appears at the given site of the segment n that is chosen as $x = 0$ in this lattice description with the rate k (Fig. 8.1b). This hole can diffuse along the lattice with the diffusion rate u, and it eventually disappears with the rate r (Fig. 8.1b). It is assumed that the produced hole can explore a large part of the nanorod before the production of another hole. In addition, the nanocatalysts are considered as very long ($|x| \to \infty$ in Fig. 8.1b), although the analysis can also be extended for the finite-sized nanorods, as shown in detail in Appendix F. The main goal of these calculations here is to show that there is a region of increased local concentration of charged holes around the active site where they were originally created after the chemical reaction fired up.

The function $P_n(t)\big(n = 0, \pm 1, \pm 2, \dots\big)$ is defined as the probability density function of finding the messenger at the segment n at time t if at $t = 0$, it was created at the segment $n = 0$. This function is proportional to the local concentration of charged holes. In addition, there might be a situation when the charged hole is not produced yet, while the earlier produced messenger already died. It is described by the probability density function $P_{off}(t)$. The temporal evolution of these probability density functions is governed by a set of master equations:

$$\begin{aligned} \frac{dP_n(t)}{dt} &= uP_{n-1} + uP_{n+1}(t) - (2u + r)P_n(t) \\ \frac{dP_0(t)}{dt} &= uP_{-1} + uP_{+1}(t) - (2u + r)P_0(t) \\ \frac{dP_{off}(t)}{dt} &= r \sum_{n=-\infty}^{+\infty} P_n(t) - kP_{off}(t). \end{aligned} \tag{8.1}$$

In addition, there exists a normalization condition:

$$P_{off}(t) + \sum_{n=-\infty}^{+\infty} P_n(t) = 1. \tag{8.2}$$

Assuming that the stationary distribution of finding the charged hole at the lattice sites

around the place where it was produced ($n = 0$ in Fig. 8.1b) can be achieved fast, one can calculate these steady-state probability density functions, as shown in detail in Appendix F:

$$P_{off} = \frac{r}{r+k} = \frac{a}{a+b} \tag{8.3}$$

$$P_n = By^{|n|}, \qquad n = \pm 1, \pm 2, \ldots \tag{8.4}$$

$$P_0 = \frac{2By + ab/(a+b)}{a+2}, \tag{8.5}$$

where $a \equiv r/u$ and $b \equiv k/u$, and other parameters are given by

$$y = 1 + \frac{a}{2} - \sqrt{\frac{a^2}{4} + a} \tag{8.6}$$

and

$$B = \frac{b(1-y)}{y(a+b)(3+a-y)}. \tag{8.7}$$

The calculated stationary distribution of probabilities to find the charged hole around the place where it was initially produced is presented in Fig. 8.2. One can see that there is an increased probability to find the messenger at the same place ($n = 0$), and the distribution exhibits the exponential decay as the distance from the origin increases. This defines a specific length scale λ (in units of the segment size l) associated with the decay:

$$\lambda = -\frac{1}{\ln y}. \tag{8.8}$$

The physical meaning of this length is that it corresponds to the size of the region with the increased local concentration of charged holes due to the initial production at the active site at $n = 0$. Consequently, this length can be associated with the correlation length of catalytic communications because as this study postulated, in the region $[-\lambda, \lambda]$ there will be an increased probability of firing up the chemical reaction at another active site.

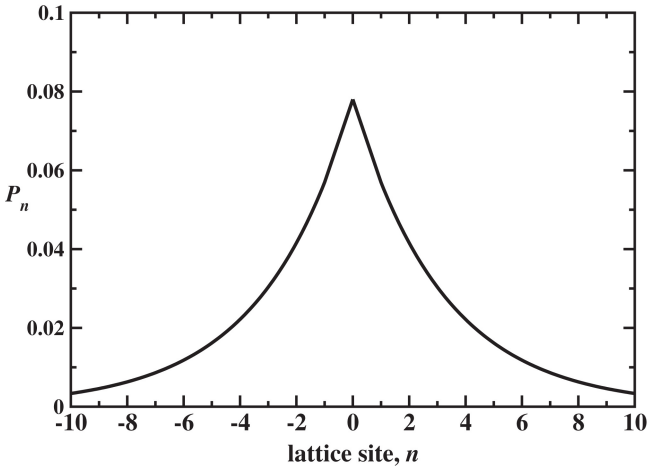


Figure 8.2: Stationary distribution of probabilities of finding the charged holes particles near the position where they were produced in the chemical reaction on site 0. Parameters used in calculations are $u = 10\ \mathrm{s}^{-1}$ and $r = k = 1\ \mathrm{s}^{-1}$.

If the lifetime of the charged hole is very short ($r \gg 1, a \gg 1$), one obtains $y \simeq 2/a$, and the correlation length of the catalytic communication will be very short, $\lambda \sim 1/\ln a$. In this case, the charged holes cannot create a significant region with the increased local density of messengers, and thus, weak or no catalytic communication is expected. In a more realistic scenario, the charged holes are quite stable ($r \to 1, a \to 1$), and one finds $y \simeq 1 - \sqrt{a}$ with the correlation length being equal to $\lambda \simeq 1/\sqrt{a}$. This means that the region of cooperativity in the catalysis on nanorods is solely determined by the balance between the diffusion and death rates of the messenger particles ($a = r/u$). Note that it does not depend on the charged hole production rate k.

8.3 Results and Discussion

8.3.1 Catalytic Communications in Metallic Nanorods

Now this theoretical approach can be applied for clarifying the microscopic mechanisms that lead to the catalytic communications phenomena observed recently on metal nanocatalysts.[16] In these experiments, various redox chemical reactions have been catalyzed on single Pd and Au nanorods. More specifically, the photo-induced disproportionation of resazurin (Rz) to generate fluorescent resorufin on Pd nanorods, the oxidative deacetylation of the amplex red (AR) into the fluorescent resorufin by H_2O_2, and the reductive deoxygenation of Rz to fluorescent resorufin by NH_2OH on Au nanorods have been comprehensively analyzed.[16]

The measured memory time τ is associated as the inverse death rate for the charged holes, $r = 1/\tau$. For measuring the correlations in reaction times, the nanorods were divided into segments of length $l \simeq 100$ nm, and the diffusion constants D of the messengers were also measured in these experiments.[16] This allows for the evaluation of the effective migration rate in this discrete-state stochastic model as $u = 2D/l^2$. Finally, the correlation lengths are calculated explicitly using Eq. 8.8 with $y \simeq 1 - \sqrt{a}$. The results of the theoretical calculations for Pd and Au single nanocatalysts and the comparison with experimental measurements are presented in Table 8.1. An excellent agreement between theoretical predictions and experimental values is observed for all investigated systems, strongly supporting this theoretical approach.

Table 8.1: Comparison of experimentally determined and theoretically calculated correlation lengths in the catalytic communication. Data are taken from Ref.[16] and the theoretical correlation lengths are calculated using Eq. 8.8.

	Pd nanorods	**Au nanorods (Rz disproportionation)**	**Au nanorods (AR deacetylation)**
Lifetimes, τ (s), experimental	27.5 ± 2.4	128 ± 32	168 ± 36
Diffusion constants, D (s^{-1}), experimental	0.92 ± 0.08	1.1 ± 0.4	1.2 ± 0.4
Correlation lengths, experimental (nm)	225 ± 20	516 ± 160	624 ± 180
Correlation lengths, theoretical (nm)	225 ± 20	531 ± 164	635 ± 175

In experimental studies, the correlation lengths and memory times in the catalytic communications were specifically obtained by analyzing a Pearson's cross-correlation coefficient between chemical reaction times for different events on different segments.[16] This parameter provides a quantitative measure of how the catalyzed chemical reactions at different sites and at different times are correlated with each other. A positive value of the Pearson coefficient was attributed to a sign of the existence of cooperative communications in the nanocatalysts. This discrete-state stochastic model is also convenient for calculating the Pearson's cross-correlation coefficient, providing an important test on the validity of the proposed theoretical approach. To make the comparison between experiments and theory more realistic, this work utilized Monte Carlo computer simulations of

the discrete-state stochastic model as explained in Appendix F. The results obtained from simulations are presented in Fig. 8.3. One can see that the calculated correlation agrees well with experimentally obtained values. Importantly, this analysis produces the same range of amplitudes for the Pearson's coefficient, and the decay is also clearly observed. This gives the additional support to this theoretical model.

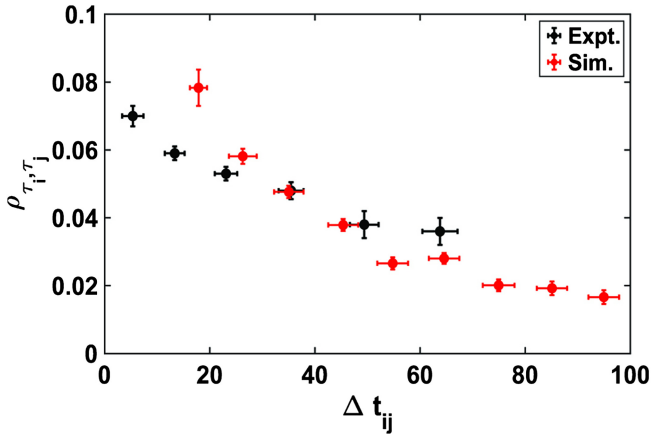


Figure 8.3: Experimental (Expt.; black symbols; from ref.[16]) and the simulated (Sim.; red symbols) Pearson's cross-correlation coefficients vs. average time separation ($\Delta t_{i,j}$) of temporally subsequent reactions at two different segments on the same Pd nanorod. Parameters used in theoretical calculations are $r = 0.036$ s^{-1}, $u = 0.184$ s^{-1}, $k = 0.0145$ s^{-1}, and $n = 700$. Details of the simulation procedure are given in Appendix F.

Experimental results exhibit relatively low amplitudes of correlations. The presented theoretical approach can explain these observations by noting that the experimentally observed migration rates are relatively slow (from experimental data, it can be estimated that $u = 0.184$ s^{-1} for Pd nanoparticles and $u = 0.23$ s^{-1} for Au nanoparticles), and the correlations are measured for active sites located at different segments at relatively long distances of ~ 100 nm. It is proposed then that the low amplitude of the reported correlations might be the result of the following factors: large distances between the active sites where the correlations are measured, low rates of the charged holes diffusion that lead to the limited influx to other active sites, and possibly, large densities of already present charged holes that are only weakly modified by the incoming migration fluxes of the holes. However, importantly, these arguments also suggest additional ways of testing the proposed theoretical model. It is predicted that the amplitude of correlations will increase if the correlations are to be measured for the active sites that are closer to each other and for increasing the flux of incoming charged holes to new active sites (via accelerating the diffusion rates or making more holes during the catalyzed chemical reactions).

8.3.2 The Effect of External Electric Field

To probe the origin of catalytic communications in single nanocatalysts, the correlations have been also measured under conditions when the external voltage has been applied.[16] It was found that the communication distance in the downfield direction increased while the same distance in the opposite direction decreased. In addition, these correlation lengths showed a cosinusoidal modulation as a function of the angle between the field and the axis of the nanorod. Furthermore, these distances exhibited a linear scaling with the increased applied voltage, at least for the strengths of external fields utilized in these experiments. These observations allowed researchers to conclude that the catalytic cooperativity is due to positively charged messenger species.[16] Since several quantitative results on the effect of external fields have been obtained, it is important to test if this theoretical model can explain these observations.

In this approach, under an external electric field, the effective diffusion rate of the charged holes to move between neighboring segments will be different depending on the angle between the nanorod and the applied voltage,

$$u^{\pm} = u \exp\left[\pm \frac{m\Delta\varepsilon \cos\phi}{k_B T} \right] \tag{8.9}$$

where ϕ is the angle, m is the number of holes that pass the border between segments per unit time, and $\Delta\varepsilon$ is the energy needed to move one charged messenger from one segment to the next one. It is given by

$$\Delta\varepsilon = \frac{q|V|l}{d}, \tag{8.10}$$

where q is the charge of a single hole, $|V|$ is the strength of the applied voltage, d is the distance between electrodes in the external voltage experiments, and l is the size of the segment. Thus, in the absence of the external field, the charged holes exhibit an unbiased diffusion motion, while the introduction of the applied voltage changes the character of the motion to be the biased diffusion. This is the main effect of the external field according to this work.

On average, charged holes have a lifetime $\tau = 1/r$ before they disappear. Then, the increase in the communication length in the direction favored by the external field in comparison with the opposite direction (in units of the segment length l) is given by

$$\frac{\Delta x_{0,EF}}{l} = (u^+ - u^-)\tau \simeq \frac{2u}{r}\frac{m\Delta\varepsilon\cos\phi}{k_BT}. \tag{8.11}$$

The last expression is an approximation because at typical experimental conditions $|V| \sim 1$ V, $l = 100$ nm, and $d \sim 1$ to 10 mm, it yields a very small energy cost of moving the single charged hole between the segments, $\Delta\varepsilon \sim 0.001$ to 0.01 k_BT. In addition, the parameter m (number of charged holes passing the border between two neighboring segments per unit time) is also expected to be not very large. This allows for the expansion of the exponent in Eq. 8.9, yielding the result in Eq. 8.11.

The main quantitative result given in Eq. 8.11 has three important predictions. First, it argues that the catalytic communication distances in the external electric field, $x_{0,EF} = x_0 + \Delta x_{0,EF}$ with x_0 being the voltage-free correlation length, will exhibit a cosine dependence in the orientation angle relative to the electric field direction. It is explicitly illustrated in Fig. 8.4, and excellent agreement with experimental observations is found. This theoretical picture naturally explains the cosine dependence, suggesting that this is the consequence of the relatively weak external voltages. For stronger fields or large fluxes between segments (larger m), it is predicted that the behavior would have a stronger exponential dependence on the cosine of the orientation angle (Eq. 8.9). This is a prediction that can be tested experimentally.

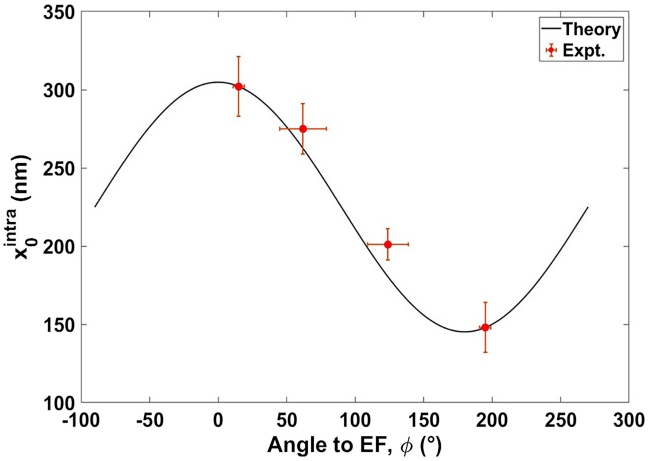


Figure 8.4: Dependence of the catalytic communication distance for temporally subsequent reactions at two different segments as a function of the orientation angle ϕ at the given voltage from the discrete-state stochastic model. Red symbols with error bars are the experimental data (Expt.) on Pd nanorods catalyzing the photo-induced disproportionation of the compound Rz as described in ref.[16]. Parameters used for calculations are $r = 0.036$ s^{-1}, $u = 0.184$ s^{-1}, $q = 1.6 \times 10^{-19}$ C, $l = 100$ nm, $d = 1$ mm, $m = 17.1$, and $V = 1.2$ V.

The second conclusion from Eq. 8.11 is that the change in the communication length,

$\Delta x_{0,EF}$, depends linearly on the applied voltage as suggested by Eq. 8.10. Figure 8.5 compares theoretical predictions with experimental observations for different voltages. Again, a linear trend in the dependence of $\Delta x_{0,EF}$ as a function of $|V|$ is clearly observed. These findings give additional support to this theoretical model, emphasizing the important role of charged holes and their dynamics.

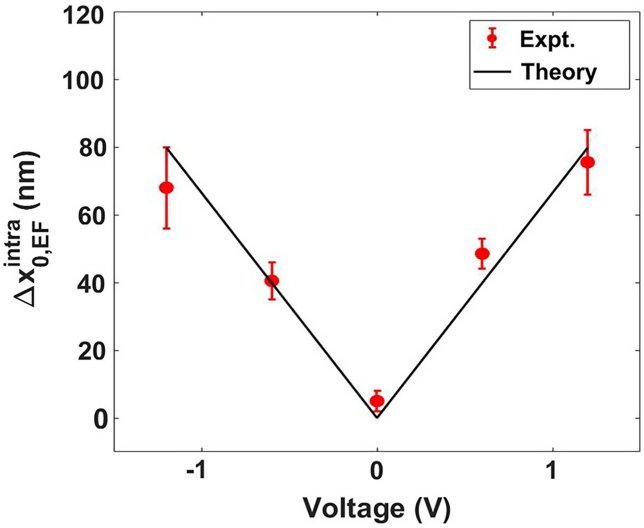


Figure 8.5: Dependence of the amplitude of the catalytic communication length variation as a function of the applied voltage calculated using the proposed theoretical model (solid black line). Red symbols with error bars correspond to the experimental data on Pd nanorods as discussed in ref.[16]. Parameters used for calculations are $r = 0.036$ s^{-1}, $u = 0.184$ s^{-1}, $q = 1.6 \times 10^{-19}$ C, $l = 100$ nm, $m = 17.1$, and $d = 1$ mm.

A third conclusion from Eq. 8.11 is that under the external field, the correlation length depends linearly on the diffusion rate and the lifetime, in contrast to the field-free case when the correlation length depends on a square root of the diffusion rate and the lifetime of the charged holes. It is important also to note that this theoretical model predicts that the memory times are independent of the external field because the death rate does not depend on the applied voltage, again in agreement with experimental observations.[16]

An important parameter in the transport of charged messengers is m, which gives the flux of the charged holes crossing from one segment to another per unit time. By comparing the theoretical predictions with experimental observations, one can explicitly determine this quantity for different systems. It can be shown that $m \simeq 17.1$ for Rz disproportionation on Pd nanorods, $m \simeq 4.6$ for AR deacetylation on Au nanorods, and $m \simeq 3.0$ for Rz deoxygenation on Au nanorods. In all these situations, these numbers are not large so that the expansion of the exponential term in Eq. 8.9 is justified, producing the cosine dependence of the modulation of the correlation length on the orientation angle.

It is also reasonable that the values of m for Au nanorods are similar for different reactions ($m \simeq 4.6$ and $m \simeq 3.0$). It will be interesting to compare these theoretical predictions with real experimental measurements of these quantities.

It is expected that the number of charged holes on the surface of the catalytic nanoparticle is probably quite large, and it is surprising that this theoretical analysis identifies only a few holes migrating between the segments per unit time. There are two possible reasons for such observations. First, it is possible that experiments could not determine all migrating holes due to the limitations in their detecting setups. Second, if the number of created holes after every catalyzed event is not large in comparison with the already existing number of charged holes at this location, then the gradient that drives the diffusion of holes will not be significant. It is believed that the second reason is more probable because it is also supported by the low amplitudes of the correlations reported from the experiments (Fig. 8.3).

8.4 Conclusions

This study presented a comprehensive theoretical approach to investigate the microscopic origin of intraparticle catalytic communications in metallic nanocatalysts. In this method, it is assumed that the properties of positively charged holes play a main role in correlating chemical reactions on different active sites of single catalytic nanoparticles. The main idea is that these reactions depend on the local concentration of charged holes, and these reactions also produce them, creating the local gradients that lead to the diffusive fluxes of these particles to different parts of the nanocatalyst. Because the probability of the reaction to occur at the given active site depends on the local concentration of the charged holes, the temporary increased concentration of charged holes stimulates the reactions in the spatially close region. Since the charged holes eventually disappear, the balance between the diffusion and the death processes defines the effective length scale of the catalytic communication, while the charged holes lifetimes determine the memory times of the catalytic cooperativity.

Using these theoretical arguments, a fully quantitative discrete-state stochastic model is developed that explicitly evaluates the lifetimes of the charged holes and the correlation lengths of the catalytic communications. It was also shown that this theoretical method could quantitatively explain the effect of external applied voltages on the catalytic com-

munication. In the presented theoretical picture, the external field modifies differently the effective diffusive rates in the field direction and against the field. The theoretical predictions are found to be in excellent agreement with all available experimental observations of catalytic communications in metal nanocatalysts.

The essence of this theoretical idea is that the nanocatalysts function via a mechanisms that is similar to branched chemical reactions.[28] These processes have been studied in macroscopic catalysts, and they can give rise to various interesting phenomena, such as traveling waves and other nonlinear effects.[29] It will be interesting to probe from this angle the catalytic communications in nanocatalysts.

While this theoretical approach can quantitatively describe all features of the catalytic communications, its most important contribution is the ability to explain the universality of such nanocatalyst cooperativity effects from the microscopic point of view.[16] It was found experimentally that the intraparticle catalytic cooperativity is independent of the experimental time resolutions, the segmentation size, the laser intensity, nanorod's catalytic activity, details of the fluorescence measurements, and the concentrations of reactants in catalyzed chemical reactions. All these properties of the systems are not influenced by the diffusion and the death rates of the charged holes that specify the microscopic picture of the catalytic communications. Thus, the correlations can be affected only by changing the dynamics and lifetimes of the charged messengers. For example, using the proposed theoretical method, one can argue how the increase in the temperature (T) might influence the catalytic communications. Since the death rate will increase with T, this would decrease the memory time. However, the effect on the correlation length would depend on what rate, diffusion or death, grows faster with the temperature. If the diffusion rate increases quicker, then larger communication distances are expected with the temperature increase. The opposite effect is expected when the death rate of the charged holes changes faster than the diffusion rate. It will be interesting to test these theoretical predictions, as well as others discussed in this work, in experimental studies.

The proposed microscopic picture of catalytic communications raises a question on how to improve the efficiency of heterogeneous catalytic processes. There are significant efforts in developing new artificial enzyme mimics of catalysts with controlled chemical properties, and metal nanocatalysts with cooperative correlations might be a good candidate for such developments.[30] Several specific suggestions to improve the catalytic efficiency can be made based on these theoretical arguments. It seems reasonable that the most efficient

strategy would be to utilize nanocatalysts where the distances between the active sites are shorter than the correlation lengths. This would ensure the increased activity in all the nanocatalyst system. One could also tune the catalytic efficiency by varying the density of charged holes and density of traps that kill them.[31] There are multiple experimental techniques to affect the concentrations of charged holes, including various photochemical and electrochemical methods. Furthermore, one could suggest that the nanocatalysts with more rough surfaces would have a larger amount of active sites, leading to more efficient catalysts. It is also interesting to note that these ideas on importance of charged holes in catalytic processes are related to recent proposals on the role of charged holes in organic electrosynthesis and photochemical water oxidation.[32, 33]

It is also important to discuss the limitations of the presented theoretical approach. The detailed microscopic mechanisms of how charged holes specifically stimulate the reaction on the active sites and how the additional charged holes are produced after such reactions are not discussed in this model.[4] This is because more detailed quantum-mechanical and electronic structure information is needed to clarify these issues, and they are beyond of the scope of this study. In addition, these details probably will not affect the dynamics and lifetimes of charged messengers, and thus, they will not influence the cooperativity observed in the metal nanocatalysts. However, the following semiphenomenological arguments might still be presented. The experimentally investigated processes are redox chemical reactions that have been carried out on nanocatalyst particles that consist of metal cores covered by silica shells.[16] It was argued that the metal surfaces are oxidized, allowing for the existence of localized charged holes on the surfaces (instead of delocalized holes in bulk metals). Considering that all the active sites are present on the surface only, one might assume that a high local concentration of charged holes is associated with the presence of active sites. This is because the incoming substrates might donate electrons to the nearby holes, leading to effective interactions that allow the catalyzed processes to occur. Moving the charges to the product molecules can also create more charged holes at this location. Of course, these arguments are quite speculative, and most probably, more complex processes are occurring in these systems. However, at the minimum, this theoretical method provides an effective quasiclassical mesoscopic description of catalytic communications that, due to its quantitative nature, might still provide some useful insights.

This theoretical model also considers an effectively one-dimensional picture, while real catalysts are three-dimensional objects. In addition, it is assumed that the stationary

dynamics are very quickly reached in the system, but this should be tested more carefully. Furthermore, the diffusion and death rates of the charged holes are assumed to be constant in time and uniform along the nanocatalyst, while it is more realistic to expect these transition rates to be spatially and temporary heterogeneous. It remains unclear how this heterogeneity would affect the catalytic communications effect. However, despite these limitations, this theoretical method provides a physically consistent microscopic picture that also allows for a comprehensive quantitative description of the correlations in chemical reactions catalyzed on single nanoparticles. Thus, this theoretical method clarifies important aspects of the fundamental mechanisms of the complex processes that are taking place during the heterogeneous catalysis.

Bibliography

(1) Somorjai, G. A.; Li, Y., *Introduction to Surface Chemistry and Catalysis*; John Wiley & Sons: New York, 2010.

(2) Ross, J. R., *Heterogeneous Catalysis: Fundamentals and Applications*; Elsevier: Amsterdam: The Netherlands, 2011.

(3) Astruc, D. Introduction: Nanoparticles in Catalysis. *Chem. Rev.* **2020**, *120*, 461–463.

(4) Zhang, Z. et al. Ensembles of Metastable States Govern Heterogeneous Catalysis on Dynamic Interfaces. *Acc. Chem. Res.* **2020**, *53*, 447–458.

(5) Fechete, I. et al. The Past, Present and Future of Heterogeneous Catalysis. *Catal. Today* **2012**, *189*, 2–27.

(6) Corma, A.; Garcia, H. Supported Gold Nanoparticles as Catalysts for Organic Reactions. *Chem. Soc. Rev.* **2008**, *37*, 2096–2126.

(7) Torres Galvis, H. M. et al. Supported Iron Nanoparticles as Catalysts for Sustainable Production of Lower Olefins. *Science* **2012**, *335*, 835–838.

(8) Moerner, W.; Fromm, D. P. Methods of Single-Molecule Fluorescence Spectroscopy and Microscopy. *Rev. Sci. Instrum.* **2003**, *74*, 3597–3619.

(9) Betzig, E.; Chichester, R. J. Single Molecules Observed by Near-Field Scanning Optical Microscopy. *Science* **1993**, *262*, 1422–1425.

(10) Chen, P. et al. Spatiotemporal Catalytic Dynamics within Single Nanocatalysts Revealed by Single-Molecule Microscopy. *Chem. Soc. Rev.* **2014**, *43*, 1107–1117.

(11) Xu, W. et al. Single-Molecule Nanocatalysis Reveals Heterogeneous Reaction Pathways and Catalytic Dynamics. *Nat. Mater.* **2008**, *7*, 992–996.

(12) Xu, W. et al. Probing the Catalytic Activity and Heterogeneity of Au-Nanoparticles at the Single-Molecule Level. *Phys. Chem. Chem. Phys.* **2009**, *11*, 2767–2778.

(13) Chen, T. et al. Single-Molecule Nanocatalysis Reveals Catalytic Activation Energy of Single Nanocatalysts. *J. Am. Chem. Soc.* **2016**, *138*, 12414–12421.

(14) Maris, J. E. et al. Single-Molecule Observation of Diffusion and Catalysis in Nanoporous Solids. *Adsorption* **2021**, *27*, 423–452.

(15) Burda, C. et al. Chemistry and Properties of Nanocrystals of Different Shapes. *Chem. Rev.* **2005**, *105*, 1025–1102.

(16) Zou, N. et al. Cooperative Communication within and between Single Nanocatalysts. *Nat. Chem.* **2018**, *10*, 607–614.

(17) Cui, Q.; Karplus, M. Allostery and Cooperativity Revisited. *Prot. Sci.* **2008**, *17*, 1295–1307.

(18) Ye, R. et al. Analogy between Enzyme and Nanoparticle Catalysis: A Single-Molecule Perspective. *ACS Catal.* **2019**, *9*, 1985–1992.

(19) Sapp, W. et al. Charge Transfer Mechanism in Titanium-Doped Microporous Silica for Photocatalytic Water-Splitting Applications. *Catalysts* **2016**, *6*, 34.

(20) Zhang, L. et al. Kinetics and Mechanisms of Charge Transfer Processes in Photocatalytic Systems: A Review. *J. Photochem. Photobiol. C: Photochem. Rev.* **2012**, *13*, 263–276.

(21) Jang, M.-H. et al. Synergetic Effect on Catalytic Activity and Charge Transfer in Pt-Pd Bimetallic Model Catalysts Prepared by Atomic Layer Deposition. *J. Chem. Phys.* **2020**, *152*.

(22) Luca, O. R. et al. Catalysis by Electrons and Holes: Formal Potential Scales and Preparative Organic Electrochemistry. *Org. Chem. Front.* **2015**, *2*, 823–848.

(23) Studer, A.; Curran, D. P. The electron is a catalyst. *Nat. Chem.* **2014**, *6*, 765–773.

(24) Bian, Z. et al. Superior Electron Transport and Photocatalytic Abilities of Metal-Nanoparticle-Loaded TiO2 Superstructures. *J. Phys. Chem. C* **2012**, *116*, 25444–25453.

(25) Chernyak, L. et al. Electron Beam-Induced Increase of Electron Diffusion Length in p-Type GaN and AlGaN/GaN Superlattices. *Appl. Phys. Lett.* **2000**, *77*, 875–877.

(26) Gonzalez-Vazquez, J. et al. Determination of the Electron Diffusion Length in Dye-Sensitized Solar Cells by Random Walk Simulation: Compensation Effects and Voltage Dependence. *J. Phys. Chem. C* **2010**, *114*, 8552–8558.

(27) Utterback, J. K. et al. Observation of Trapped-Hole Diffusion on the Surfaces of CdS Nanorods. *Nat. Chem.* **2016**, *8*, 1061–1066.

(28) Houston, P. L., *Chemical Kinetics and Reaction Dynamics*; Courier Corporation: 2012.

(29) Feinberg, M.; Terman, D. Traveling Composition Waves on Isothermal Catalyst Surfaces. *Arch. Rational Mech. Anal.* **1991**, *116*, 35–69.

(30) Huang, Y. et al. Nanozymes: Classification, Catalytic Mechanisms, Activity Regulation, and Applications. *Chem. Rev.* **2019**, *119*, 4357–4412.

(31) Rozen, J. et al. Increase in Oxide Hole Trap Density Associated with Nitrogen Incorporation at the SiO2/SiC Interface. *J. Appl. Phys.* **2008**, *103*.

(32) Francke, R.; Little, R. D. Electrons and Holes as Catalysts in Organic Electrosynthesis. *ChemElectroChem* **2019**, *6*, 4373–4382.

(33) Zhang, Y. et al. Modulating Multi-Hole Reaction Pathways for Photoelectrochemical Water Oxidation on Gold Nanocatalysts. *Energy Environ. Sci.* **2020**, *13*, 1501–1508.

Chapter 9

How Heterogeneity Affects Cooperative Communications within Single Nanocatalysts

The content of this chapter and associated content in Appendix G are reprinted (adapted) with permission from (*J. Phys. Chem. Lett.* 2023, 14, 36, 8227-8234). Copyright © 2023 American Chemical Society.

9.1 Introduction

Rates of chemical reactions are governed by the highest energetic barriers on the path from breaking old bonds and creating new ones. Catalysis is a powerful method of accelerating these chemical reactions by adding certain compounds that are not consumed in these reactions, called catalysts, that provide alternative faster free-energy pathways.[1] From living systems to industrial processes, catalytic phenomena have been extensively studied over the years, but the underlying microscopic mechanisms in many systems have not been fully clarified.[2, 3] In particular, there are multiple questions to understand heterogeneous catalysis where the reactants, catalysts, and products might exist in distinct thermodynamic phases.[4–6]

Recent advances in nanotechnology have led to the development of various types of nanoparticles as promising heterogeneous catalysts with potentially wide applications.[7, 8] Nanocatalysts might be synthesized in specific shapes and sizes that help to rationally tune their catalytic properties, increasing their efficiency.[9–11] Another advantage of nanoparticles is that catalytic processes can now be investigated with high temporal and spatial resolution, uncovering phenomena that could not be observed in bulk-solution experiments.[12–18] One of the most surprising recent observations obtained using single-molecule fluorescence microscopy is a discovery of catalytic communications between surface active sites in gold (Au) and palladium (Pd) nanoparticles.[19] Analysis of correlations between successive time-dependent product formation events occurring at different segments of nanoparticles revealed that chemical reactions at specific sites effectively stimulate chemical reactions at the neighboring sites, but the effect weakens with the distance and time. Such correlations might exist up to $\sim 10 - 100$ s and the corresponding correlation lengths are reaching $\sim 200 - 600$ nm.[19] These experiments also indicated that charged holes are most probably responsible for this phenomenon.

Single-particle experimental measurements found that the rates of product formation and product dissociation on nanocatalysts show a wide range of distribution indicating the existence of strong temporal fluctuations at the molecular level. Furthermore, these nanoparticles might undergo dynamic restructuring, exhibiting different chemical activity with changes in size.[20–22] Several experiments have also demonstrated changes in the catalytic efficiency of the nanocatalysts depending on their facet shapes.[23–25] These spatiotemporal fluctuations might be also connected with surface defects on nanoparticles that can disturb the arrangement of active sites on the lattice.[26, 27] As a result, the

catalytic reactivity patterns might differ significantly between apparently similar nanoparticles.[26] These findings indicate that nanocatalysts are strongly heterogeneous systems, and this should affect their catalytic properties.

To explain the microscopic mechanisms of catalytic communications in single nanoparticles, the previous chapter discussed a novel discrete-state stochastic framework to study catalysis of redox chemical reactions.[28] Single-molecule microscopy experiments have reported both intra- and inter-particle communications in catalytic reactions on Pd- or Au-based nanocatalysts through redox processes.[19] As experimental evidences suggest that a positively charged species is responsible for cooperative catalytic communication within single nanocatalysts by moving along oxide surfaces of nanoparticles, it was assumed that the probability for a chemical reaction to occur at the given site depends on the local concentration of positively charged holes,[29–31] and each chemical reaction also creates new charged holes. The idea was that the increased local concentration of charged holes forces them to move into the areas of the nanoparticle with smaller concentrations, and this stimulates chemical reactions at neighboring sites due to incoming charged particles increasing the concentrations of charged holes around these sites.[28] A quantitative model based on these ideas has been developed, allowing for explicit estimates of communication lifetimes and distances. Excellent agreement with all available experimental observations has been found. Importantly, this theoretical approach was able to explain not only the characteristic spatial and temporal lengths scales of catalytic communications and their dependence on external electric fields but also the universality and robustness of this phenomenon.[28] At the same time, this theoretical approach neglected a crucial aspect of heterogeneity of nanoparticles,[24–27, 32, 33] raising an important question of how inhomogeneous microscopic rates that control the dynamics of charged particles will affect the catalytic communications phenomenon in single nanoparticles.

9.2 Theoretical Model

This chapter presents a theoretical investigation of the role of heterogeneity in catalytic communications within single nanoparticles. By extending the original homogeneous stochastic approach (introduced in the previous chapter) to a more realistic heterogeneous description, the cooperativity in catalyzed chemical reactions on the single nanoparticle is evaluated using explicit calculations and Monte Carlo computer simulations. By specifically considering several heterogeneous distributions of transition rates that spatially vary

along the nanoparticles, communication lifetimes and correlation lengths are evaluated. The analytical calculations and computer simulations indicate that heterogeneity increases the strength of correlations while catalytic communication distances and lifetimes generally decrease. Theoretical arguments to explain these observations are also presented.

Consider the catalytic cooperativity processes in a single nanoparticle as shown in Fig. 9.1a. Since experiments were mostly performed on nanorods, this analysis also concentrates only on 1D nanocatalysts.[19] The single nanorod particle can be viewed as a lattice consisting of multiple segments, and each segment is assumed to have a length l nm, similar to what was done in the analysis of single-molecule fluorescent measurements of catalyzed chemical reactions.[19] The number of active sites in each segment might vary reflecting the heterogeneity of the system (Fig. 9.1a). It is assumed also that only one type of redox chemical reaction can take place on any active site of the nanocatalyst. In addition, since these reactions involve charged species such as electrons and holes, it is postulated that the probability of a chemical reaction occurring at the specific site depends on the local concentration of charged holes, and each reaction creates a new charged hole with a rate k (Fig. 9.1a). The charged holes can diffuse along the nanorod with equal probability in both directions and the migration rate out of the segment j to the segment $j+1$, as well as the rate to move from the segment $j+1$ to the segment j are equal to u_j: see Fig. 9.1a. The charged holes can also disappear with a death rate r_j from the segment j. Different position-dependent transition rates reflect the fact that due to structural inhomogeneities, each segment of the nanorod might have a different number of active sites and defects, leading to different migration rates u_j and death rates r_j. Now, a heterogeneous discrete-state stochastic model for catalytic communications can be easily developed as a generalization of the original homogeneous approach.[28] In this model, the average time for the created charged hole to stay in the system corresponds to the communication lifetime, while the average distance that this charged hole moves after creation and before disappearing is associated with the communication length scale.

To better understand the quantitative effects of heterogeneity on catalytic communications, this work considers in detail the simplest heterogeneous model with alternating transition rates, as illustrated in Fig. 9.1b. It is assumed that a new hole is produced on the nanorod with rate k at the segment labeled as state $n = 0$. From this segment, the hole can diffuse to adjacent segments ($n = \pm 1, \pm 2, ...$), and it might also eventually disappear.[28] Additionally, it is assumed that the nanorod is very long. In this approach, the forward transition rates from the odd-numbered segment to the even-numbered segment and the

backward transition rates from the even-numbered segment to the odd-numbered segment are given by u_1, while the forward transition rates from the even-numbered segment to the odd-numbered segment and the backward transition rates from the odd-numbered segment to the even-numbered segment are given by u_2: see Fig. 9.1b. In addition, charged holes might disappear from the odd-numbered segments with the rate r_1, while the death rate from the even-numbered segments is r_2. Since all charged holes in the system are independent of each other, the evolution of a single hole is monitored in this theoretical framework and this should provide a quantitative description of catalytic communications.

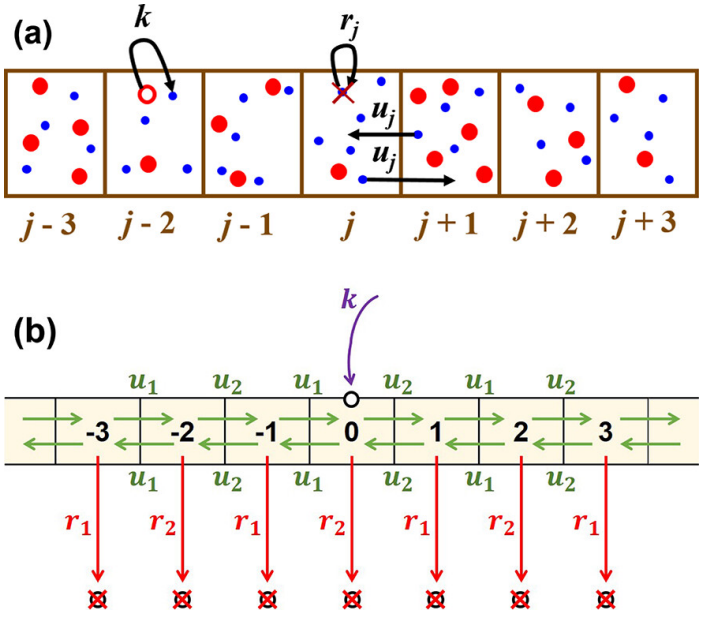


Figure 9.1: (a) A schematic description of a heterogeneous single nanorod where chemical reactions are catalyzed at active sites (red-filled circles), creating charged holes (blue circles). The nanorod is divided into multiple segments of equal length. A new charged hole appears with the rate k at a given active site (empty red circle). Charged holes can migrate out of the segment j to the segment $j+1$ or back from the segment $j+1$ to the segment j with the rate u_j, and they can also disappear from the segment j with the rate r_j. (b) A simplest alternating heterogeneous discrete-state stochastic model for catalytic communications in a single nanoparticle. There are two types of segments in the system. A new hole is produced at $n = 0$ with the rate k. The forward rate from the odd-numbered segment to the even-numbered segment and the backward rate for the same transition is equal to u_1. Similarly, the forward rate from the even-numbered segment to the odd-numbered segment and the backward rate for the same transition is equal to u_2. The death rate of the charged hole from an odd-numbered (even-numbered) segment is r_1 (r_2).

To quantify the catalytic communications in the heterogeneous model with alternating transition rates, $P_n(t)$ is defined as the probability to find a charged hole particle in the segment n at time t. Additionally, the function $P_{off}(t)$ describes the probability of the state when the hole is already disappeared and not yet produced on the nanorod. The

temporal evolution of these probabilities is described by a set of forward master equations,

$$\frac{dP_n(t)}{dt} = u_1 P_{n+1}(t) + u_2 P_{n-1}(t) - (u_1 + u_2 + r_1) P_n(t), \tag{9.1}$$

for n odd, while for even n

$$\frac{dP_n(t)}{dt} = u_2 P_{n+1}(t) + u_1 P_{n-1}(t) - (u_1 + u_2 + r_2) P_n(t). \tag{9.2}$$

At the segment where the charged hole is created ($n = 0$), the corresponding expression is

$$\frac{dP_0(t)}{dt} = k P_{off}(t) + u_2 P_1(t) + u_1 P_{-1}(t) - (u_1 + u_2 + r_2) P_0(t). \tag{9.3}$$

In addition, for the state of the system when the hole is not present on the lattice ($n = off$),

$$\frac{dP_{off}(t)}{dt} = r_1 \Sigma_{i=odd} P_i(t) + r_2 \Sigma_{i=even} P_i(t) - k P_{off}(t). \tag{9.4}$$

Moreover, there is a normalization condition, requiring that

$$P_{off}(t) + \Sigma_{i=odd} P_i(t) + \Sigma_{i=even} P_i(t) = 1. \tag{9.5}$$

It is reasonable to assume that the system will quickly reach the steady state after the charged hole is created at $n = 0$,[28] and then one can analytically calculate stationary properties of the system. As explained in detail in the Appendix G, the stationary probabilities to find the charged hole at the segment n of the nanorod can be written as

$$P_n \simeq B y^{|n|}, \quad n = 0, \pm 1, \pm 2, ..., \tag{9.6}$$

where

$$y = \sqrt{\frac{z}{2u_1 u_2} - \sqrt{\left(\frac{z}{2u_1 u_2}\right)^2 - 1}}, \tag{9.7}$$

and

$$z = 2u_1u_2 + r_1r_2 + (u_1 + u_2)(r_1 + r_2). \tag{9.8}$$

The parameters B in Eq. 9.6 differ for odd-numbered and even-numbered segments, and they are explicitly evaluated in the Appendix G.

9.3 Results and Discussion

The results of calculations for stationary probabilities (obtained analytically and via Monte Carlo computer simulations) for different heterogeneous situations are presented in Fig. 9.2. One can see that although alternating transition rates clearly influence the stationary probabilities, overall these quantities can be viewed as effectively decaying with the distance from the production site at segment $n = 0$. The spread of the distribution describes the region across which the hole is likely to exist before it disappears from the nanorod. Then, the effective characteristic length of this exponential decay (l is the size of each segment),

$$\lambda = -\frac{l}{\ln y}, \tag{9.9}$$

can be associated with the communication lengths in catalytic cooperativity.[28] This is because this length describes the region of increased local concentration due to the produced charged holes, and this increased concentration of charged holes leads to a higher probability for the reaction to occur for the active sites within this region. This exactly corresponds to the correlation length as was measured in experiments.[19] The results of analytical calculations and estimates from Monte Carlo simulations of communication lengths for different heterogeneity cases are also presented in Table 9.1. It is important to note that this communication length is independent of the hole creation rate k and is controlled only by the balance between the migration and death rates until it disappears from the lattice. As shown in Fig. G.2, the rate k affects the amplitude (maximum value) of the distribution whereas the decay region of this distribution remains almost unaltered.

This theoretical approach can estimate the lifetimes τ of the catalytic communication phenomena. This is the time after the new charged hole has been created at the given site by the chemical reaction until it disappears. This leads to

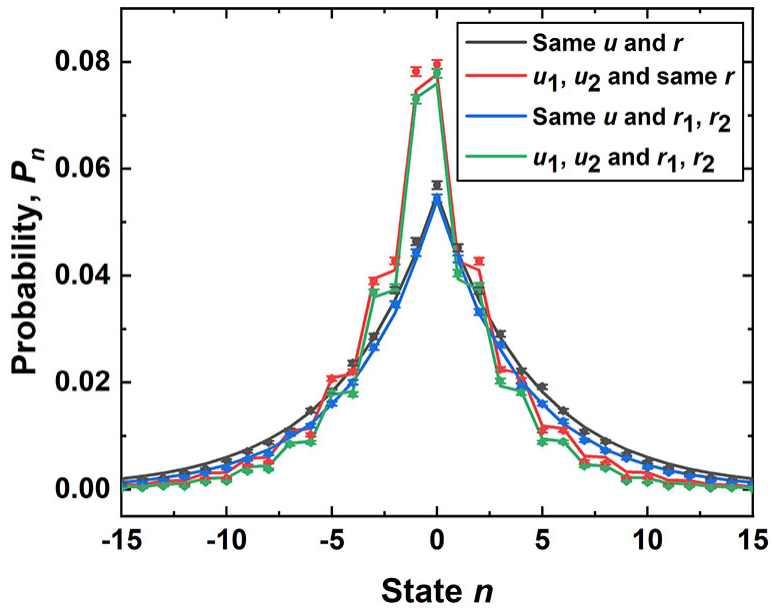


Figure 9.2: Stationary probability distributions of finding a charged hole at different segments on the nanorod after it is initially produced at the segment $n = 0$. The nanorod is assumed to be divided into segments of length $l = 50$ nm. The following parameters have also been used in calculations: $u_1 = 4u, u_2 = u/4, r_1 = r/2$ and $r_2 = 2r$, $k = r = 0.03636$ s^{-1}, and $u = 0.736$ s^{-1}. Solid curves represent the analytical results (assuming infinitely long nanorods), while the symbols correspond to the results from Monte Carlo computer simulations (with the nanorod of total length 2550 nm).

$$\frac{1}{\tau} = \frac{\sum_{j \neq off} r_j P_j}{\sum_{j \neq off} P_j} = \frac{k P_{off}}{1 - P_{off}}. \tag{9.10}$$

This expression reflects the fact that the lifetime can be estimated only when the charged hole is on the nanorod, and it is given by the inverse flux for the charged holes to be removed from the lattice After substituting the expression of P_{off} from Eq. G.9c in Appendix G in Eq. 9.10, it is found that $\tau = M_2/M_1$. These constants M_1 and M_2 are defined in Eq. G.8 in Appendix G and they are independent of the creation rate of the hole k. Thus, the mean lifetime of the hole does not depend on its creation rate. At the same time, because the lifetimes are independent of the creating rate k, one can propose a simpler two-state model to evaluate this quantity, as explained in Appendix G. It is shown here that

$$\tau \approx \frac{2u + r_1 + r_2}{u(r_1 + r_2) + r_1 r_2}. \tag{9.11}$$

This is an excellent approximation to the exact expression for the lifetime, and the reason it is an approximation is that the symmetry of the original model (with the production at the specific site) is slightly different from the simplified two-state model. Also, in the system

with the homogeneous death rates, $r_1 = r_2 = r$, one can obtain $\tau = 1/r$, in agreement with previous theoretical evaluations of catalytic communications for homogeneous systems.[28]

The lifetime of catalytic cooperativity is an important factor that affects catalytic processes in the system. If these lifetimes are very short (fast death rates), which corresponds to r_1 and r_2 being much larger than u_1 and u_2, one can obtain from Eqs. 9.7 and 9.8,

$$y \sim \sqrt{\frac{u_1 u_2}{r_1 r_2}}, \tag{9.12}$$

which leads to the following estimate of the communication length,

$$\lambda \sim \frac{l}{\ln\left(\sqrt{\frac{r_1 r_2}{u_1 u_2}}\right)} \to 0. \tag{9.13}$$

Since $r_1, r_2 \gg u_1, u_2$, these communication lengths are too short to support meaningful catalytic cooperativity at long distances. However, the analysis of experimental measurements suggests that probably a more realistic situation is when the diffusion rates are much faster than the death rates ($u_1, u_2 \gg r_1, r_2$).[28] In this limit, the properties of catalytic communications can be described as

$$\tau \simeq \frac{2}{r_1 + r_2}, \tag{9.14}$$

and

$$\lambda \simeq 2l\sqrt{\frac{u_1 u_2}{(r_1 + r_2)(u_1 + u_2)}}. \tag{9.15}$$

Now, considering this realistic case of fast migration rates, it is convenient to quantify the effect of heterogeneity. To do so, heterogeneity parameters α_u and α_r are introduced via the corresponding expressions for the migration and death rates

$$u_1 = \alpha_u u, \quad u_2 = \frac{u}{\alpha_u}; \quad r_1 = \alpha_r r, \quad r_2 = \frac{r}{\alpha_r}, \tag{9.16}$$

where rates u and r correspond to the purely homogeneous case. This is a convenient approach because by tuning the dimensionless control parameters α_u and α_r, the system

can be modified from the homogeneous case ($\alpha_u = \alpha_r = 1$) to any degree of heterogeneity. Please note that this is not a unique way of quantifying heterogeneity, but it can be shown explicitly that the main physics will not change for other scenarios of random distributions of transition rates.

The expressions for the lifetimes and communication distances can now be written as

$$\tau = \frac{1}{r}\frac{2\alpha_r}{\alpha_r^2 + 1}, \tag{9.17}$$

and

$$\lambda = 2l\sqrt{\frac{u}{r}}\sqrt{\frac{\alpha_u\alpha_r}{(1+\alpha_u^2)(1+\alpha_r^2)}}. \tag{9.18}$$

Table 9.1 presents these properties of catalytic cooperativity for alternating heterogeneous models for different distributions of transition rates. One can see that temporal memory (τ) depends only on heterogeneity in the death rates (α_r), but not on the inhomogeneity in the migration rates (α_u). Also, the lifetimes are shorter in comparison with homogeneous systems. At the same time, communication lengths (λ) depend on both heterogeneity parameters, and they also decrease for more inhomogeneous systems, although the effect is stronger for varying α_u.

To better illustrate the effect of heterogeneity, dimensionless parameters are introduced such as

$$R(\tau) = \frac{\tau(\alpha_r)}{\tau(\alpha_r = 1)} = \frac{2\alpha_r}{1+\alpha_r^2}; \tag{9.19}$$

and

$$R(\lambda) = \frac{\lambda(\alpha_r, \alpha_u)}{\lambda(\alpha_r = \alpha_u = 1)} = 2\sqrt{\frac{\alpha_u\alpha_r}{(1+\alpha_u^2)(1+\alpha_r^2)}}. \tag{9.20}$$

These quantities specify how the properties of catalytic communications change for heterogeneous systems in comparison with homogeneous systems. The results for specific calculations for these ratios are presented in figure 9.3. As one can see, the longest communication lifetimes and distances can be achieved only for homogeneous systems, and

Table 9.1: Analytically calculated and computer simulation results of the catalytic communication lifetimes (τ) and distances (λ). The nanorod is assumed to be divided into segments of length $l = 50$ nm. The following parameters have been used in calculations: $k = r = 0.03636$ s^{-1}, $u = 0.736$ s^{-1}, $\alpha_u = 4$ and $\alpha_r = 2$. The simulation procedure is explained in the in Appendix G. The data in brackets are taken from Ref.[19] corresponding to the experimentally measured ranges for catalytic communications on Pd nanorods.

Case	Temporal memory τ (s)		Communication distance λ (nm)	
	Theory	**Simulations**	**Theory**	**Simulations**
$\alpha_u = \alpha_r = 1$	27.5 $(27.5 \pm 2.4)^a$	27.50 ± 0.27	224.96 $(225 \pm 20)^a$	223.96 ± 2.21
$\alpha_u = 4,\ \alpha_r = 1$	27.5	27.07 ± 0.27	154.32	153.54 ± 1.53
$\alpha_u = 1,\ \alpha_r = 2$	22	21.76 ± 0.22	201.21	197.06 ± 1.99
$\alpha_u = 4,\ \alpha_r = 2$	22	21.91 ± 0.22	138.03	133.60 ± 1.44

increasing the degree of heterogeneity will always lead to shorter communication lifetimes and distances. Thus, heterogeneity reduces the range of catalytic cooperativity in both spatial and temporal domains.

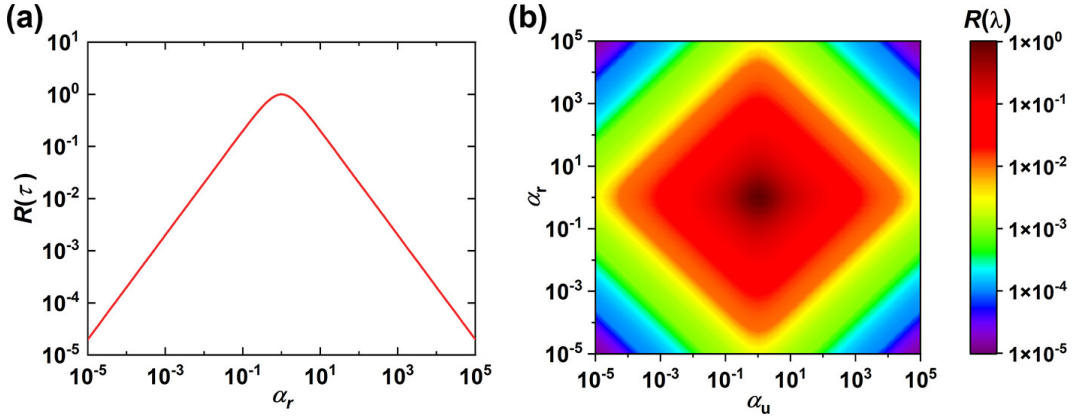


Figure 9.3: (a) Relative communication lifetimes $R(\tau)$ (see Eq. 9.18) as a function of the heterogeneity parameter α_r; and (b) relative communication distances $R(\lambda)$ (see Eq. 9.19) as a function of heterogeneity parameters α_u and α_r.

The effect of inhomogeneous transition rates can also be seen in the stationary probabilities of finding the charged hole at different segments as illustrated in Figure 9.2. Modifying the originally homogeneous model ($u_1 = u_2$, $r_1 = r_2$, black curve) by adding heterogeneity only to the death rates ($u_1 = u_2$, $r_1 \neq r_2$, blue curve) changes the distribu-

tion P_n only slightly, but one can already see a narrower distribution for the stationary probabilities. However, adding heterogeneity to the migration rates ($u_1 \neq u_2$, $r_1 = r_2$, red curve; and $u_1 \neq u_2$, $r_1 \neq r_2$, green curve) have a much stronger effect: it increases the probability in the region near the origin and makes the whole distribution more narrow further away from the origin (see Fig. 9.2). This actually means that heterogeneity decreases the catalytic communication length because the region of the increased local concentration of charged holes is now smaller in comparison with the homogeneous system. Additionally, the steady-state distributions become asymmetric on introducing inhomogeneous migration rates. By interchanging the values of the rates u_1 and u_2 (see Fig. G.3), the asymmetry shifts to the opposite direction and it is concluded that slow migration rates from any segment n causes the hole to stay on that segment longer which increases its probability P_n.

There is another way to evaluate the properties of catalytic communications. In single-molecule experiments, these properties have been determined by analyzing Pearson's cross-correlation coefficient (PCC or ρ_{τ_i,τ_j}).[19] This quantity exhibited an exponential decay with respect to the average time separation ($\Delta t_{i,j}$) between successive product formation events as well as the distance separation ($\Delta x_{i,j}$) of any two segments i and j on the single nanorod. The respective decay constants correspond to catalytic communications lifetimes and distances, respectively. This theoretical approach uses Monte Carlo computer simulations to mimic experimental measurements in order to evaluate the PCC values. The estimates for communication lifetimes and distances obtained from these computer simulations are shown in Figure 9.4. It is found that the amplitude of correlations is larger for the systems with inhomogeneous transition rates in comparison with the homogeneous situations, in agreement with the previous findings.

It is important to emphasize that this theoretical analysis of the role of heterogeneity in catalytic communications on single nanorods concentrated so far on the simplest model with alternating migration and death rates (see Fig. 9.1b). However, it is much more realistic to encounter more random distributions of transition rates in nanoparticles. This situation is illustrated in Fig. 9.5a. In addition, the analysis of an alternating model suggests that heterogeneity increases the strength of catalytic cooperativity while shortening communication lengths and lifetimes. But it is unclear if this is a general result or specific to the alternating rates model. To test these theoretical predictions, consider a model where all transition rates are random. More specifically, using Monte Carlo computer simulations catalytic communication phenomenon is analyzed for the situations when each

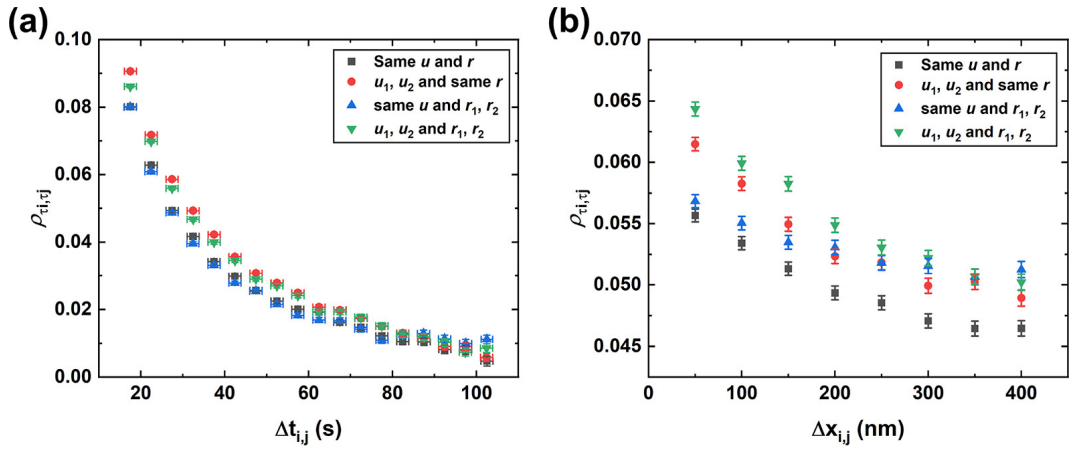


Figure 9.4: Pearson's cross-correlation coefficients as a function of (a) average time separation, $\Delta t_{i,j}$; and as a function of (b) average distance separation, $\Delta x_{i,j}$, for any segments i and j on the single nanorod. A 1200 nm nanorod is utilized in computer simulations, and it is divided into segments of length $l = 50$ nm. The following parameters have been also used in calculations: $k = r = 0.03636$ s^{-1}, $u = 0.736$ s^{-1}, $\alpha_u = 4$ and $\alpha_r = 2$. The detailed procedure to determine Pearson's cross-correlation coefficients from Monte Carlo computer simulations is presented in the Appendix G.

rate is taken from different Gaussian distributions. Figure 9.5b and Table 9.2 present the simulation results for stationary probabilities of created charged holes and for corresponding communication lifetimes and distances. One can again observe that heterogeneity increases the amplitude of catalytic cooperativity near the site where the charged holes are produced. It also leads to decreased communication lifetimes and distances. These results fully agree with the explicit calculations for alternating heterogeneous models.

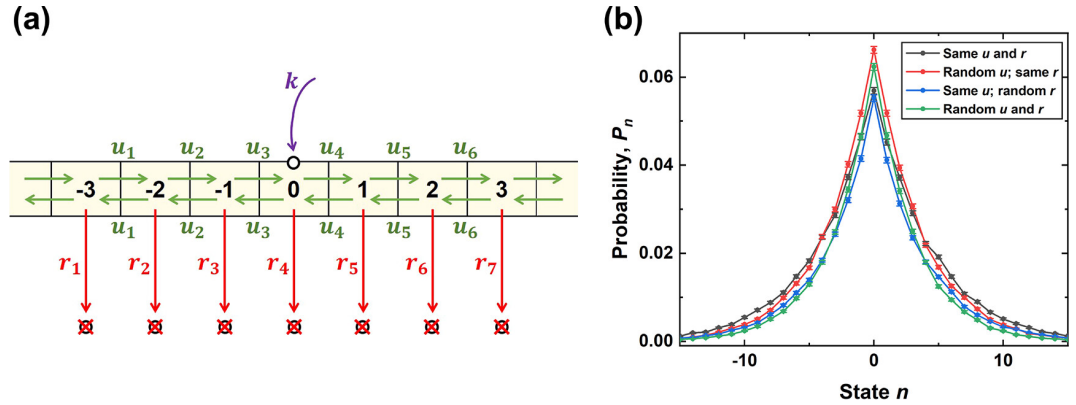


Figure 9.5: (a) A discrete-state stochastic model of catalytic communications with random transition rates. A new hole is produced at $n = 0$ with a rate k and it diffuses on the nanorod with random migration rates u_i $(i = 1, 2, ...)$ until it disappears with a death rate r_i $(i = 1, 2, ...)$. (b) Monte Carlo computer simulation results for the steady-state probability distributions of finding a hole at different segments of the nanorod after it is initially produced at the segment $n = 0$. All transition rates are randomly chosen from respective Gaussian distributions with the average values of death rates $r = 0.03636s^{-1}$ and migration rates $u = 0.736s^{-1}$. The nanorod is assumed to be divided into segments of length $l = 50$ nm, and the hole is created with the rate $k = 0.03636s^{-1}$.

Table 9.2: Catalytic communications lifetimes, τ, and distances, λ, determined from Monte Carlo computer simulations with random transition rates. All transition rates are randomly chosen from respective Gaussian distributions with the average value of death rates $r = 0.03636s^{-1}$ and migration rates $u = 0.736s^{-1}$. Furthermore, the standard deviations about the mean values are $\sigma_r = 0.05s^{-1}$ and $\sigma_u = 0.5s^{-1}$, respectively. The creation rate of a hole is considered to be $k = 0.03636s^{-1}$.

Case	**Temporal memory** τ **(s)**	**Communication distance** λ **(nm)**
Same u **and** r	27.50 ± 0.27	223.96 ± 2.21
Random u **and same** r	27.47 ± 0.27	186.34 ± 1.86
Same u **and random** r	18.76 ± 0.19	182.46 ± 1.87
Random u **and** r	18.87 ± 0.20	160.98 ± 1.62

The theoretical analysis that explored analytical calculations and computer simulations indicates that, generally, heterogeneity in transition rates influences catalytic communications by increasing the amplitude of correlations and decreasing the communication lifetimes and distances, as compared to the homogeneous distribution of transition rates. The advantage of this theoretical method is that one can present microscopic arguments to explain these observations.

First, the properties of catalytic communications for a heterogeneous system with some random distributions of death rates r_j and migration rates u_j are compared with a homogeneous system with uniform averaged death rates $r = \langle r_j \rangle$ and averaged migration rates $u = \langle u_j \rangle$. Charged holes can disappear from any segment of the nanorod, yielding the following death flux

$$J_{death} \simeq \sum_j r_j P_j. \tag{9.21}$$

However, this flux is dominated by death events from the segments with the fastest death rates r_{max}. This means that the corresponding communication lifetime, $\tau_{hetero} = 1/J_{death} \sim 1/r_{max}$ is shorter than the corresponding time for the homogeneous system, $\tau_{homo} = 1/r$ because $r_{max} \geq r$. Thus, it is generally expected to have $\tau_{hetero} \leq \tau_{homo}$.

For the diffusion of charged holes along the nanorod, it is expected that they will be found preferentially at the segments with the smallest migration rates, u_{min}, out of them. Then, the average communication distance can be well approximated by

$\lambda_{hetero} \sim \sqrt{u_{min}/r}$.[28] But this should lead to shorter communication distances than for the homogeneous system, $\lambda_{homo} \sim \sqrt{u/r}$ since generally $u_{min} \leq u$. Now, because in the heterogeneous model the charged holes after the creation at the given site, on average, will diffuse shorter distances, there will be an increased local concentration of charged holes which is larger than for the corresponding homogeneous case. This should lead to stronger catalytic cooperativity at these shorter distances since more charged holes are available to stimulate chemical reactions at neighboring active sites. The results of these theoretical calculations and computer simulations are fully consistent with these more microscopic arguments.

9.4 Conclusions

This chapter presented a theoretical investigation of the role of heterogeneity in the catalytic communication phenomena on single nanoparticles. Because of the variability of structural properties of nanocatalysts, in this theoretical approach, the heterogeneity is accounted for by considering random distributions of transition rates of charged holes that are controlling catalytic cooperativity. To reflect experimental observations of catalytic processes on nanorods, the analysis is done for effective 1D systems. Consequently, the original homogeneous discrete-state stochastic model could be extended to situations with inhomogeneous diffusion and death rates for charged holes. By specifically considering a simplified heterogeneous model, where transition rates alternate between odd and even segments of the nanorod, the properties of catalytic communications phenomena are evaluated via exact analytical calculations supplemented by Monte Carlo simulations. It is found that heterogeneity generally decreases the catalytic communications lifetimes and distances, although the effect of cooperativity becomes stronger. Similar trends have been also observed for general models with random distributions of transition rates, supporting the generality of the theoretical predictions on the nature of heterogeneity in catalytic communications. Furthermore, microscopic arguments to explain these observations have been also presented. These theoretical results suggest that specific details of local structures around the active sites (facet types, roughness, defects, etc.) might be the leading factors that can be tuned to control the efficiency of catalytic processes at these sites.

It is important to critically evaluate the presented theoretical approach. To model the variability of chemical and structural properties of single nanocatalysts, a specific "bond" heterogeneity has been assumed in this theoretical method. In other words, for

the specific bond that connects two neighboring segments of the nanorod, it is assumed the same migration rates, while these rates are different for other bonds. However, there is another scenario of how heterogeneity might be implemented, which can be labelled as a "segment" heterogeneity. In this case, one could assume the migration rates in any direction from the given segment to be the same, and the corresponding diffusion rates are different for other segments. It is expected that the quantitative values of catalytic communication properties will depend on different heterogeneity scenarios, but the general trends (increased correlation strength, decreased communication lifetimes, and distances) will be the same. In addition, it was assumed that the stationary dynamics are quickly reached in the system, but this might not be the case. Thus, it will be interesting to investigate transient phenomena in catalytic communication processes.

Also, this theoretical method postulated that catalyzed chemical reactions produce charged holes and that the probability for the active site to operate depends on the local concentration of holes. These statements, although plausible, have not been experimentally tested. In addition, although nanorods are three-dimensional (3D) systems, a minimal 1D description is sufficient to quantitatively probe catalytic cooperative communications within single nanorods. However, this 1D description is only valid for nanorods and for more complex nanoparticles (such as nanoplates), a 2D description may be explored. Furthermore, infinitely long nanorods have been assumed in the analytical calculations, but the real nanoparticles are finite and some finite-size effects might be present. Finally, although currently there are no experimental observations, the advantage of this method is that there are specific predictions, at least for general trends of catalytic communications, that can be utilized to probe the presented theoretical predictions. This should lead to a better understanding of the fundamental mechanisms of catalytic processes.

Bibliography

(1) Farrauto, R. et al., *Introduction to Catalysis and Industrial Catalytic Processes*; Wiley: 2016.

(2) Somorjai, G. A.; Li, Y., *Introduction to Surface Chemistry and Catalysis*; John Wiley & Sons: New York, 2010.

(3) Roduner, E. Understanding Catalysis. *Chem. Soc. Rev.* **2014**, *43*, 8226–8239.

(4) Ross, J. R., *Heterogeneous Catalysis: Fundamentals and Applications*; Elsevier: Amsterdam: The Netherlands, 2011.

(5) Fechete, I. et al. The Past, Present and Future of Heterogeneous Catalysis. *Catal. Today* **2012**, *189*, 2–27.

(6) Schlögl, R. Heterogeneous Catalysis. *Angew. Chem. Int. Ed.* **2015**, *54*, 3465–3520.

(7) Liu, L.; Corma, A. Metal Catalysts for Heterogeneous Catalysis: From Single Atoms to Nanoclusters and Nanoparticles. *Chem. Rev.* **2018**, *118*, 4981–5079.

(8) Astruc, D. Introduction: Nanoparticles in Catalysis. *Chem. Rev.* **2020**, *120*, 461–463.

(9) Tao, F.; Spivey, J., *Metal Nanoparticles for Catalysis: Advances and Applications*; ISSN; Royal Society of Chemistry: 2014.

(10) Roldan Cuenya, B. Metal Nanoparticle Catalysts Beginning to Shape-Up. *Acc. Chem. Res.* **2013**, *46*, 1682–1691.

(11) An, K.; Somorjai, G. A. Size and Shape Control of Metal Nanoparticles for Reaction Selectivity in Catalysis. *ChemCatChem* **2012**, *4*, 1512–1524.

(12) Janssen, K. P. et al. Single Molecule Methods for the Study of Catalysis: From Enzymes to Heterogeneous Catalysts. *Chem. Soc. Rev.* **2014**, *43*, 990–1006.

(13) Decan, M. R. et al. Copper Nanoparticle Heterogeneous Catalytic 'Click' Cycloaddition Confirmed by Single-Molecule Spectroscopy. *Nat. Commun.* **2014**, *5*, 1–8.

(14) Shen, H. et al. Single-Molecule Kinetics Reveals a Hidden Surface Reaction Intermediate in Single-Nanoparticle Catalysis. *J. Phys. Chem. C* **2014**, *118*, 26902–26911.

(15) Sambur, J. B.; Chen, P. Approaches to Single-Nanoparticle Catalysis. *Annu. Rev. Phys. Chem.* **2014**, *65*, 395–422.

(16) Chen, T. et al. Single-Molecule Nanocatalysis Reveals Catalytic Activation Energy of Single Nanocatalysts. *J. Am. Chem. Soc.* **2016**, *138*, 12414–12421.

(17) Qu, X. et al. Single-Molecule Nanocatalysis Reveals the Kinetics of the Synergistic Effect Based on Single-AuAg Bimetal Nanocatalysts. *J. Phys. Chem. Lett.* **2022**, *13*, 830–837.

(18) Chen, P. et al. Spatiotemporal Catalytic Dynamics within Single Nanocatalysts Revealed by Single-Molecule Microscopy. *Chem. Soc. Rev.* **2014**, *43*, 1107–1117.

(19) Zou, N. et al. Cooperative Communication within and between Single Nanocatalysts. *Nat. Chem.* **2018**, *10*, 607–614.

(20) Xu, W. et al. Single-Molecule Nanocatalysis Reveals Heterogeneous Reaction Pathways and Catalytic Dynamics. *Nat. Mater.* **2008**, *7*, 992–996.

(21) Zhou, X. et al. Size-Dependent Catalytic Activity and Dynamics of Gold Nanoparticles at the Single-Molecule Level. *J. Am. Chem. Soc.* **2010**, *132*, 138–146.

(22) Zhang, Y. et al. Unique Size-Dependent Nanocatalysis Revealed at the Single Atomically Precise Gold Cluster Level. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 10588–10593.

(23) Lee, K. et al. Catalytic Nanoparticles Being Facet-Controlled. *J. Mater. Chem. A* **2010**, *20*, 3791–3798.

(24) Chen, T. et al. Single-Molecule Nanocatalysis Reveals Facet-Dependent Catalytic Kinetics and Dynamics of Pallidium Nanoparticles. *ACS Catal.* **2017**, *7*, 2967–2972.

(25) Tachikawa, T. et al. Evidence for Crystal-Face-Dependent TiO2 Photocatalysis from Single-Molecule Imaging and Kinetic Analysis. *J. Am. Chem. Soc.* **2011**, *133*, 7197–7204.

(26) Zhou, X. et al. Quantitative Super-Resolution Imaging Uncovers Reactivity Patterns on Single Nanocatalysts. *Nat. Nanotechnol.* **2012**, *7*, 237–241.

(27) Andoy, N. M. et al. Single-Molecule Catalysis Mapping Quantifies Site-Specific Activity and Uncovers Radial Activity Gradient on Single 2D Nanocrystals. *J. Am. Chem. Soc.* **2013**, *135*, 1845–1852.

(28) Punia, B. et al. Microscopic Mechanisms of Cooperative Communications within Single Nanocatalysts. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119*, e2115135119.

(29) Furube, A. et al. Charge Carrier Dynamics of Standard TiO2 Catalysts Revealed by Femtosecond Diffuse Reflectance Spectroscopy. *J. Phys. Chem. B* **1999**, *103*, 3120–3127.

(30) Studer, A.; Curran, D. P. The electron is a catalyst. *Nat. Chem.* **2014**, *6*, 765–773.

(31) Hu, Q. et al. Nanomaterial Catalysts for Organic Photoredox Catalysis-Mechanistic Perspective. *Nanoscale* **2021**, *13*, 18044–18053.

(32) Xu, W. et al. Probing the Catalytic Activity and Heterogeneity of Au-Nanoparticles at the Single-Molecule Level. *Phys. Chem. Chem. Phys.* **2009**, *11*, 2767–2778.

(33) Chaudhury, S. et al. Dynamics of Chemical Reactions on Single Nanocatalysts with Heterogeneous Active Sites. *J. Chem. Phys.* **2023**, *158*, 074101.

Chapter 10

Conclusions and Future Directions

Natural events surrounding us seem to occur deterministically while following classical physics. However, on a closer inspection, it is revealed that these events are inherently stochastic at the microscopic scale due to the random collisions of individual interacting particles. Small-volume systems like biological cells are naturally stochastic and probing their underlying fluctuations becomes crucial for understanding biomolecular functions.[1–3] Specifically, chemical reactions and biological events occur as a consequence of stochastic collisions among individual reactant species and biomolecules, respectively.[3–5] This leads to variations in the microscopic timescales of each reaction event with multiple possibilities of the sequence of occurrences, guided by respective probabilities. Bulk measurement techniques tend to average the outcomes from participating molecules in similar conditions, thereby obscuring stochastic effects from each interacting molecule.[6, 7] On the contrary, single-molecule (SM) experiments have revolutionized science by allowing researchers to isolate and manipulate individual molecules, uncovering new discoveries and revealing the underlying mechanisms and complexities of stochastic processes.[8–10] The large quantitative data obtained in these experiments, including microscopic timescales of individual events, can produce probability distributions that assist in exploring the stochastic events in chemistry and biology. Interestingly, these distributions can also be derived using computer simulations and theoretical tools of statistical mechanics, allowing researchers to probe the molecular aspects of the processes and unveiling significant insights of the underlying microscopic mechanisms.

Biological cells are extremely complex structures where multiple processes occur simultaneously. Given that biomolecules account for about 40% of the intra-cellular volume,

they may act as crowding agents and influence biological events.[1, 2, 11] Although these macromolecules are widely present, they were mostly overlooked in previous research and have since gained significant attention. The present thesis theoretically investigated the influence of these bulk crowders on two crucial stochastic biological processes evolving over time. First, the search of DNA-binding proteins to their specific target sequences on the DNA and second, the transport of single-stranded DNA (ssDNA) across cellular membranes through biological nanopores. The underlying dynamics of these systems were closely examined through discrete-state stochastic modelling frameworks and performing first-passage time analysis. The exact analytical expressions for the average timescales of these events have been determined, which often corresponds to the macroscopic rates in bulk measurements. Comprehensive analysis reveals that the search of proteins is hindered by crowders due to an obstructed 3D diffusion in the cytoplasm. However, the presence of such macromolecules can be beneficial for genetic sequencing since they accelerate the capture of ssDNA within nanopores followed by a slowed translocation. Such theoretical frameworks and corresponding analysis, in this context, provide clarity about crowder-induced microscopic mechanisms that may be exploited by researchers in synthetic systems as well as in novel developments.

In addition to studying biological systems, this thesis also explored the underlying dynamics of chemical reactions catalysed by single metal nanoparticles that possess multiple active sites. Extensive experimental measurements have shown that such nanocatalytic systems are inherently stochastic and heterogeneous, possessing surface inhomogeneity and exhibiting dynamic temporal fluctuations, thus leading to dynamic disorder.[12, 13] The complexity of these systems can be accurately captured through discrete-state stochastic models. Moreover, this approach can successfully quantify inherent stochastic fluctuations by the determination of randomness parameter, providing crucial insights that are obscured in ensemble measurements. These theoretical modelling techniques may also be employed in probing the microscopic mechanisms of, rather surprising, intra-particle cooperative communications observed within single metal nanocatalysts via SM imaging techniques. This underscores the importance of SM experiments and theoretical statistical modelling techniques working in tandem, uncovering new findings and enhancing any system-of-interest at the microscopic scale.

Chemical and biological processes are inherently stochastic due to the random nature of molecular interactions, the influence of thermal fluctuations, conformational changes, variations in ionic and chemical concentrations, etc. at the microscopic scale. Partic-

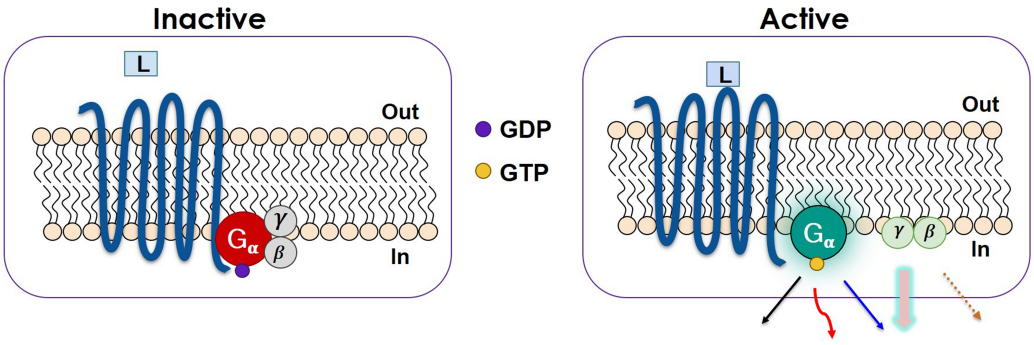


Figure 10.1: Schematic representation of ligand-induced GPCR signalling process. The seven-transmembrane (7TM) receptor lies across cellular membranes and is coupled to intracellular G-proteins that constitute Gα, Gβ, and Gγ subunits. In the inactive state, the ligand (L) binding has not yet occured outside the membrane and the Gα subunit (in red) of the G-protein is bound to GDP (in purple). The binding of ligand to the 7TM bundle renders the G-protein active due to the switching of GDP to GTP (in yellow) at the Gα subunit (in green) within the cell and the G$\beta\gamma$ dimer also dissociates. Eventually, this leads to cascade of multiple responses or signals (shown by various arrows) within the biological cells.

ularly, conformational fluctuations are fundamental to the function of various cellular components, including surface receptors. From reading this text to savouring a delicious cuisine, our bodies constantly respond to external stimuli through a complex network of biological receptors. G-protein Coupled Receptors (GPCRs) are a pivotal class of highly specific membrane receptors involved in transmitting extracellular signals into intracellular responses. These receptors play a crucial role in numerous physiological processes and are targets for a significant portion of therapeutic drugs.[14–16] In the 1950s, Earl Sutherland and Ted Rall[17] began ground-breaking research in this field that was later carried on by Alfred G. Gilman[18] and Martin Rodbell.[19] Their efforts helped to explain the fundamental process of GPCR signalling, which involves a receptor, a G protein that acts as a transducer, and an effector. GPCRs are encoded by over 1000 genes and form a seven-transmembrane (7TM) receptor grouped into six categories based on structural resemblance.[20] Small ligands associate with the receptor 7TM bundle, while peptides and proteins bind to the NH_2 terminal and/or extracellular loops.[14, 21] GPCRs are coupled to G-proteins which are heterotrimeric guanine nucleotide binding proteins with Gα, Gβ, and Gγ subunits.[22, 23] As shown in Fig. 10.1, the activity of GPCRs is regulated by the guanosine triphosphate (GTP) and its inactive counterpart guanosine diphosphate (GDP).[24] The binding of a ligand (or agonist) to the receptor leads to structural rearrangements, triggering a cascade of intracellular responses. The active GTP-bound Gα subunit dissociates from the receptor and G$\beta\gamma$ dimer, possibly partially, allowing it to

interact with downstream effectors like adenylyl cyclase enzyme, potassium and calcium channels. The Gα subunit's intrinsic GTPase activity hydrolyzes GTP to GDP, rendering it inactive and re-association of the G protein heterotrimer terminates the signal.[20, 24]

Despite their critical role, the underlying mechanisms of ligand activation and signal generation in GPCRs remain inadequately understood at the molecular level. Crystal structure examination and various SM studies have shown that extracellular ligand binding triggers a sequence of structural alterations that ultimately enhances the binding of G-proteins on the intracellular side.[25, 26] Alternatively, evidences also show that G-protein interaction may induce fluctuations to control the activity of the receptors as well as allosteric binding of ligands.[27, 28] These findings might suggest a two-way allosteric interaction between the respective intra- and inter-cellular binding sites. Regardless, the exact mechanisms by which these microscopic changes influence GPCR activation and consequent signal transduction are not fully understood. Interestingly, ligands can sometimes preferentially activate only a subset of the many possible signaling pathways, a phenomenon known as biased agonism.[29, 30] Understanding the kinetics of biased agonism holds immense potential for the development of next-generation therapeutics that selectively target specific pathways, thereby stimulating desired physiological responses.[31] Several FRET and bioluminescence resonance energy transfer (BRET) studies have also indicated the possibility of pre-coupled receptor-G protein complexes that produce basal signals, even without any agonist binding.[32–34] Additionally, SM tracking of G-proteins diffusing over cell surfaces in vivo reveal the existence of possible "hot spots" where these proteins accumulate and preferentially bind to receptors, resulting in their activation.[35] Such findings show how intrinsic complexity of cellular conditions may also impact G-protein activation and eventual signal production. Kenakin and coworkers have made significant contributions in pharmacology through theoretical modelling of these complex processes, elucidating the concepts of ligand efficacy, receptor activation, and biased agonism, but their approach has been mostly deterministic.[22, 30, 36–39] Despite advances in understanding GPCR function, several questions remain. How do underlying stochastic fluctuations translate into specific cellular responses? What role do they play in the temporal and spatial regulation of signalling? Additionally, the impact of cellular microenvironments and the presence of molecular crowding on GPCR behaviour are areas that require further investigation. Future research may employ discrete-state stochastic modelling techniques to explore these questions. Such models can provide detailed insights into the dynamic processes governing GPCR signalling, allowing for the prediction of sys-

tem behaviours under various conditions. Such efforts will enhance the understanding of GPCR function and support the development of targeted therapeutic strategies.

To conclude, discrete-state stochastic modelling techniques are effective in revealing intricacies of complex chemical and biological systems, observed at the microscopic scale. Analysing the temporal evolution of these models using first passage time distribution analyses underscore their ability to provide crucial insights that seem unattainable through ensemble measurements. Altogether, such techniques and efforts aim to offer significant contributions in both theoretical and experimental research, enhancing our comprehension of stochastic processes.

Bibliography

(1) Lodish, H. F. et al., *Molecular Cell Biology*; WH Freeman New York: 2006.

(2) Alberts, B. et al., *Molecular Biology of the Cell*; Garland Science, Taylor and Francis Group: 2014.

(3) Teimouri, H.; Kolomeisky, A. B. Power of Stochastic Kinetic Models: From Biological Signaling and Antibiotic Activities to T Cell Activation and Cancer Initiation Dynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2022**, *12*, e1612.

(4) Gillespie, D. T. Stochastic Simulation of Chemical Kinetics. *Annu. Rev. Phys. Chem.* **2007**, *58*, 35–55.

(5) Kumar, A. B.; Ramaswamy, R. Chemistry at the Nanoscale: When Every Reaction is a Discrete Event. *Resonance* **2018**, *23*, 23–40.

(6) Tamarat, P. et al. Ten Years of Single-Molecule Spectroscopy. *J. Phys. Chem. A* **1999**, *104*, 1–16.

(7) Singh, D. et al. Theoretical Tools to Quantify Stochastic Fluctuations in Single-Molecule Catalysis by Enzymes and Nanoparticles. *ACS Omega* **2022**, *7*, 47587–47600.

(8) Moerner, W.; Fromm, D. P. Methods of Single-Molecule Fluorescence Spectroscopy and Microscopy. *Rev. Sci. Instrum.* **2003**, *74*, 3597–3619.

(9) Sasmal, D. K. et al. Single-Molecule Fluorescence Resonance Energy Transfer in Molecular Biology. *Nanoscale* **2016**, *8*, 19928–19944.

(10) Bustamante, C. J. et al. Optical Tweezers in Single-Molecule Biophysics. *Nat. Rev. Methods Primers* **2021**, *1*, 25.

(11) Ellis, R. J. Macromolecular Crowding: Obvious but Underappreciated. *Trends Biochem. Sci.* **2001**, *26*, 597–604.

(12) Chen, P. et al. Spatiotemporal Catalytic Dynamics within Single Nanocatalysts Revealed by Single-Molecule Microscopy. *Chem. Soc. Rev.* **2014**, *43*, 1107–1117.

(13) Ye, R. et al. Analogy between Enzyme and Nanoparticle Catalysis: A Single-Molecule Perspective. *ACS Catal.* **2019**, *9*, 1985–1992.

(14) Lagerström, M. C.; Schiöth, H. B. Structural Diversity of G Protein-Coupled Receptors and Significance for Drug Discovery. *Nat. Rev. Drug Discov.* **2008**, *7*, 339–357.

(15) Calebiro, D.; Sungkaworn, T. Single-Molecule Imaging of GPCR Interactions. *Trends Pharmacol. Sci.* **2018**, *39*, 109–122.

(16) Chavez-Abiega, S. et al. Physical Biology of GPCR Signalling Dynamics Inferred from Fluorescence Spectroscopy and Imaging. *Curr. Opin. Struct. Biol.* **2019**, *55*, 204–211.

(17) Sutherland, E. W.; Rall, T. W. Fractionation and Characterization of a Cyclic Adenine Ribonucleotide Formed by Tissue Particles. *J. Biol. Chem.* **1958**, *232*, 1077–1091.

(18) Gilman, A. G. et al. G Proteins: Transducers of Receptor-Generated Signals. *Annu. Rev. Biochem.* **1987**, *56*, 615–649.

(19) Rodbell, M. The Role of Hormone Receptors and GTP-Regulatory Proteins in Membrane Transduction. *Nature* **1980**, *284*, 17–22.

(20) Calebiro, D. et al. G Protein-Coupled Receptor-G Protein Interactions: A Single-Molecule Perspective. *Physiol. Rev.* **2021**, *101*, 857–906.

(21) Ji, T. H. et al. G Protein-Coupled Receptors: I. Diversity of Receptor-Ligand Interactions. *J. Biol. Chem.* **1998**, *273*, 17299–17302.

(22) Neubig, R. R. et al. International Union of Pharmacology Committee on Receptor Nomenclature and Drug Classification. XXXVIII. Update on terms and symbols in quantitative pharmacology. *Pharmacol. Rev.* **2003**, *55*, 597–606.

(23) T Andresen, B. A Pharmacological Primer of Biased Agonism. *Endocr. Metab. Immune Disord. Drug Targets* **2011**, *11*, 92–98.

(24) Syrovatkina, V. et al. Regulation, Signaling, and Physiological Functions of G-Proteins. *J. Mol. Biol.* **2016**, *428*, 3850–3868.

(25) Rasmussen, S. G. et al. Crystal Structure of the β2 Adrenergic Receptor–Gs Protein Complex. *Nature* **2011**, *477*, 549–555.

(26) Quast, R. B.; Margeat, E. Studying GPCR Conformational Dynamics by Single Molecule Fluorescence. *Mol. Cell. Endocrinol.* **2019**, *493*, 110469.

(27) DeVree, B. T. et al. Allosteric Coupling from G Protein to the Agonist-Binding Pocket in GPCRs. *Nature* **2016**, *535*, 182–186.

(28) Mafi, A. et al. The Mechanism for Ligand Activation of the GPCR–G Protein Complex. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119*, e2110085119.

(29) Wootten, D. et al. Mechanisms of Signalling and Biased Agonism in G Protein-Coupled Receptors. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 638–653.

(30) Kenakin, T. Biased Receptor Signaling in Drug Discovery. *Pharmacol. Rev.* **2019**, *71*, 267–315.

(31) Onaran, H. O.; Costa, T. Conceptual and Experimental Issues in Biased Agonism. *Cell. Signal.* **2021**, *82*, 109955.

(32) Gales, C. et al. Real-Time Monitoring of Receptor and G-Protein Interactions in Living Cells. *Nat. Methods* **2005**, *2*, 177–184.

(33) Nobles, M. et al. Heterotrimeric G Proteins Precouple with G Protein-Coupled Receptors in Living Cells. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 18706–18711.

(34) Ayoub, M. A. et al. Real-Time Analysis of Agonist-Induced Activation of Protease-Activated Receptor 1/Gαi1 Protein Complex Measured by Bioluminescence Resonance Energy Transfer in Living Cells. *Mol. Pharmacol.* **2007**, *71*, 1329–1340.

(35) Sungkaworn, T. et al. Single-Molecule Imaging Reveals Receptor–G Protein Interactions at Cell Surface Hot Spots. *Nature* **2017**, *550*, 543–547.

(36) Weiss, J. M. et al. The Cubic Ternary Complex Receptor–Occupancy Model I. Model Description. *J. Theor. Biol.* **1996**, *178*, 151–167.

(37) Weiss, J. M. et al. The Cubic Ternary Complex Receptor–Occupancy Model II. Understanding Apparent Affinity. *J. Theor. Biol.* **1996**, *178*, 169–182.

(38) Weiss, J. M. et al. The Cubic Ternary Complex Receptor–Occupancy Model III. Resurrecting Efficacy. *J. Theor. Biol.* **1996**, *181*, 381–397.

(39) Kenakin, T. Theoretical Aspects of GPCR–Ligand Complex Pharmacology. *Chem. Rev.* **2017**, *117*, 4–20.

Appendix A

Solution of the backward master equations

This section provides the details of derivations for the target search time by obtaining the solution for the first-passage probability distribution function $\tilde{F}_0(s)$ from Eq. (3.2) in Chapter 3. Eq. (3.2a) is a recurrence relation and it can be solved for $\tilde{F}_n(s)$ for the unblocked binding sites of the DNA. A general solution is eventually obtained as

$$\tilde{F}_n(s) = \begin{cases} Ap^n + Bp^{-n} + K_1 & n \in [1, a] \\ Cp^{L+1-n} + Dp^{n-L-1} + K_1 & n \in [a + l_{ob} + 1, L] \end{cases} \tag{A.1a}$$

$$K_1 = \frac{k_{off}\tilde{F}_0(s)}{s + k_{off}} \tag{A.1b}$$

$$p = \frac{s + 2u + k_{off} - \sqrt{(s + 2u + k_{off})^2 - 4u^2}}{2u}. \tag{A.1c}$$

For the blocked sites of the DNA, $n \in [a + 1, a + l_{ob}]$ except $n = m$, the linear equations in Eqs. (3.2d) and (3.2e) in Chapter 3 form another recurrence relation for the function $\tilde{F}^c_n(s)$.

$$\begin{aligned}(s + 2u)(s + c_{off} + r_d)\tilde{F}^c_n(s) = u(s + c_{off} + r_d)[\tilde{F}^c_{n+1}(s) \\ + \tilde{F}^c_{n-1}(s)] + sc_{off}\tilde{F}_0(s).\end{aligned} \tag{A.2}$$

From Eq. (A.2), the form of $\tilde{F}^c_n(s)$ is obtained as

$$\tilde{F}_n^c(s) = \begin{cases} Eq^{n-a} + Fq^{a-n} + K_2 & n \in [a+1, m] \\ Gq^{a+l_{ob}+1-n} + Hq^{n-(a+l_{ob}+1)} + K_2 & n \in [m, a+l_{ob}] \end{cases} \tag{A.3a}$$

$$K_2 = \frac{c_{off}\tilde{F}_0(s)}{s + c_{off} + r_d} \tag{A.3b}$$

$$q = \frac{s + 2u - \sqrt{(s+2u)^2 - 4u^2}}{2u}. \tag{A.3c}$$

Using Eq. (3.2e) of Chapter 3, the general form of $\tilde{F}_n(s)$ is determined for the blocked DNA binding sites $n \in [a+1, a+l_{ob}]$ except $n = m$.

$$\tilde{F}_n(s) = \begin{cases} r_d^{-1}(s + c_{off} + r_d)[Eq^{n-a} + Fq^{a-n}] & n \in [a+1, m] \\ r_d^{-1}(s + c_{off} + r_d)[Gq^{a+l_{ob}+1-n} + Hq^{n-(a+l_{ob}+1)}] & n \in [m, a+l_{ob}] \end{cases} \tag{A.4}$$

It is now required to determine the unknown constants to proceed further for which the boundary conditions are used. At $n = 1$ and $n = L$, substituting Eq. (A.1) in Eqs. (3.2b) and (3.2c) gives $B = Ap$ and $D = Cp$, respectively. At $n = a$ and $n = a+1$, substituting Eqs. (A.1) and (A.4) in the Eqs. (3.2a) and (3.2d) respectively results in

$$A(p^{a+1} + p^{-a}) - r_d^{-1}(s + c_{off} + r_d)[Eq + Fq^{-1}] = -K_1 \tag{A.5}$$

$$-A(p^{a} + p^{1-a}) + r_d^{-1}(s + c_{off} + r_d)[E + F] = K_1 \tag{A.6}$$

Similarly, at $n = a + l_{ob}$ and $n = a + l_{ob} + 1$,

$$C(p^{b+1} + p^{-b}) - r_d^{-1}(s + c_{off} + r_d)[Gq + Hq^{-1}] = -K_1 \tag{A.7}$$

$$-C(p^{b} + p^{1-b}) + r_d^{-1}(s + c_{off} + r_d)[G + H] = K_1 \tag{A.8}$$

The last remaining boundary condition is obtained from Eq. (3.2g) which is the final condition of this system.

$$r_d^{-1}(s + c_{off} + r_d)[Eq^{1+x} + Fq^{-1-x}] = r_d^{-1}(s + c_{off} + r_d)[Gq^{1+y} + Hq^{-1-y}] = 1 \tag{A.9}$$

The boundary conditions from equations (A.5) - (A.9) are used to find the remaining 6 constants apart from B and D. Finally, after substituting the resultant forms of $\tilde{F}_n(s)$ and $\tilde{F}_n^c(s)$ in Eq. (3.2f), the analytical expression of $\tilde{F}_0(s)$ can be determined.

$$\tilde{F}_0(s) = \frac{(s+k_{off})\big[(s+c_{off}+r_d)k_{on}S_1(s) + c_{on}r_d\big(1+S_2(s)\big)\big]}{s^2 + s^2\phi_1 + s\phi_2 + k_{off}\big[k_{on}(s+c_{off}+r_d)S_3(s) + r_d c_{on}\big(l_{ob} - S_4(s)\big)\big]} \tag{A.10}$$

In the above equation, provided $l_{ob} = 1 + x + y$ and $L - l_{ob} = a + b$,

$$\phi_1 = k_{on}(L - l_{ob}) + k_{off} + c_{on}l_{ob} + r_d + c_{off} \tag{A.11}$$

$$\phi_2 = (c_{off} + r_d)\big[k_{on}(L - l_{ob}) + k_{off}\big] + (k_{off} + r_d)c_{on}l_{ob}. \tag{A.12}$$

All the auxiliary functions $S_i(s)$ in Eq. (A.10) are a sum of two terms where one is a function of the parameters a and x while the other is a function of b and y (see Fig. 3.1 in Chapter 3) such that $S_i(s) = S_{i,ax}(s) + S_{i,by}(s)$. The form of both the terms in this sum is same and the only difference is that the parameters a and x in $S_{i,ax}(s)$ are replaced by b and y respectively in $S_{i,by}(s)$. Consider $S_1(s)$ for instance, $S_1(s) = S_{1,ax}(s) + S_{1,by}(s)$.

$$\begin{aligned} S_1(s) = &\frac{p(p^{2a}-1)(q^2-1)q^x}{(p-1)\big[(1+p^{1+2a})(q^{2+2x}-1) - (p+p^{2a})(q^{1+2x}-q)\big]} \\ &+ \frac{p(p^{2b}-1)(q^2-1)q^y}{(p-1)\big[(1+p^{1+2b})(q^{2+2y}-1) - (p+p^{2b})(q^{1+2y}-q)\big]} \end{aligned} \tag{A.13}$$

Thus, for simplicity, only the $S_{i,ax}(s)$ part of all the $S_i(s)$ functions is mentioned in the entire discussion.

$$S_{1,ax}(s) = \frac{p(p^{2a}-1)(q^2-1)q^x}{(p-1)\big[(1+p^{1+2a})(q^{2+2x}-1) - (p+p^{2a})(q^{1+2x}-q)\big]} \tag{A.14a}$$

$$S_{2,ax}(s) = \frac{q(q^x-1)\big[(1+p^{1+2a})(q^{1+x}-1) - (p+p^{2a})(q^x-q)\big]}{(q-1)\big[(1+p^{1+2a})(q^{2+2x}-1) - (p+p^{2a})(q^{1+2x}-q)\big]} \tag{A.14b}$$

$$S_{3,ax}(s) = \frac{p(p^{2a}-1)(q-1)(q^{1+2x}+1)}{(p-1)\big[(1+p^{1+2a})(q^{2+2x}-1) - (p+p^{2a})(q^{1+2x}-q)\big]} \tag{A.14c}$$

$$S_{4,ax}(s) = \frac{(p-1)(p^{2a}-1)(q^{x+1}-1)(q^x-1)q}{(q-1)\big[(1+p^{1+2a})(q^{2+2x}-1) - (p+p^{2a})(q^{1+2x}-q)\big]}. \tag{A.14d}$$

The expressions for the function $S_{i,by}(s)$ will be same as that for all the corresponding index i of the function $S_{i,ax}(s)$. From Eq. (A.10), the average target search time can be

determined, if the protein starts its search from the bulk solution labelled with state '0', which is provided in Chapter 3.

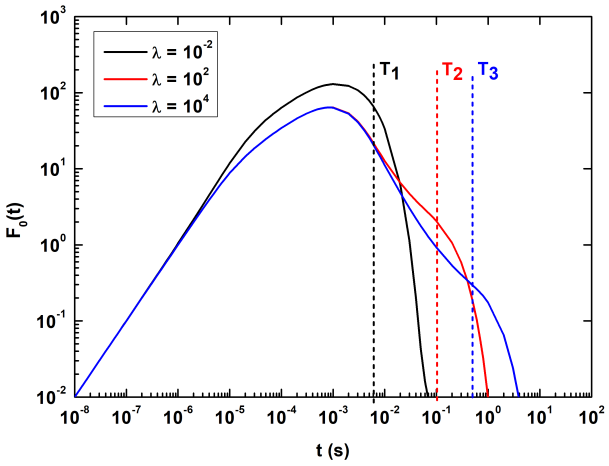


Figure A.1: First passage time distribution, $F_0(t)$, of finding the target on the DNA in the presence of protein-interacting bulk crowders provided the initial state of the protein was in the cellular matrix (state 0). The vertical dashed lines represent the mean value of the search time for the three search phases respectively at times T1 = 0.0063 s (for $\lambda = 10^{-2}$), T2 = 0.1057 s (for $\lambda = 10^2$) and T3 = 0.4962 s (for $\lambda = 10^4$). For each λ, the rate k_{off} was determined using the definition of scanning length. Fixed parameters used: $u = 10^5$ s^{-1},$k_{on} = 10^2$ s^{-1} per each DNA binding site, $c_{on} = 10^3$ s^{-1},$c_{off} = 1$ s^{-1}, $r_d = 10^3$ s^{-1}, $L = 1001$, $m = 501$, $l_{ob} = 101$ and $x = 50$.

First Passage Time Distribution (FPTD)

The FPTD of the search process provided the protein starts from the bulk solution, $F_0(t)$, is determined by carrying out a numerical inverse Laplace transform (using the Gaver-Stehfest method)[1] of $\tilde{F}_0(s)$. Figure A.1 provides the FPTDs for the three different search regimes, $\lambda = 10^{-2}, 10^2$ and 10^4. for a fixed set of parameters.

Details of Monte Carlo simulations

The DNA was modeled as a 1D lattice of L binding sites corresponding to states 1 to L. Another set of l_{ob} binding sites was considered that represented the crowded region that blocked l_{ob} number of DNA binding sites. If the protein is not bound to any of these binding sites, the protein was assumed to be in the bulk solution with the state '0'. Thus, a total of $L + l_{ob} + 1$ states exist in the system. For a particular situation of the crowded

DNA to be simulated, the rate constant parameters and the constants L, m, l_{ob}, a, and x (see Fig. 3.1 in Chapter 3) were defined.

For each target search process, the simulations were carried out using the Gillespie Algorithm[2, 3] where each transition of the protein was considered as a 'reaction' step. Here, a 'reaction' step refers to different transitions that the protein undergoes to be in any of the discrete states that are defined in the theoretical model. From Fig. 3.1, it can be observed that there are a total of 7 transitions (reaction steps) that the protein can make which are associated with a rate constant. The initial time for each search process was set as $t = 0$. Based on the Gillespie Algorithm, two random numbers were generated for each 'reaction' step which individually determined the next transition that the protein will undergo (depending on its current state) as well as the time step of the next transition. This process continued until the protein recognized the target site at state 'm' of the L binding sites. The time was recorded and a new run was started. Each search process was simulated over 2000 times followed by determining the average search time.

Bibliography

(1) Stehfest, H. Algorithm 368: Numerical Inversion of Laplace Transforms [D5]. *Commun. ACM* **1970**, *13*, 47–49.

(2) Gillespie, D. T. A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. *J. Comput. Phys.* **1976**, *22*, 403–434.

(3) Gillespie, D. T. Exact Stochastic Simulation of Coupled Chemical Reactions. *J. Phys. Chem.* **1977**, *81*, 2340–2361.

Appendix B

Solving FFPE to obtain Survival probability, $S(r,\theta;s)$ and determining the surface-averaged FPTD, $\tilde{F}(r;s)$

In this section, an approximate solution of the survival probability for a DBP, $\tilde{S}(r,\theta;s)$, from the FFPE in Eq. (4.4) is obtained under the self-consistent approximation (SCA) used previously[1, 2] for a similar system. First, the mixed boundary condition in Eq. (4.5) is replaced with the in-homogeneous condition $D_\alpha\big(\partial_r\tilde{S}(r,\theta;s)\big)_{r=\rho} = Q\Theta(\epsilon-\theta)$ where $\Theta(w)$ is the Heaviside step function and Q is the effective flux which needs to be determined by assuring the condition

$$D_\alpha\int_0^\epsilon\left(\frac{\partial\tilde{S}(r,\theta;s)}{\partial r}\right)_{r=\rho}d\theta=\kappa\int_0^\epsilon\tilde{S}(r=\rho,\theta;s)d\theta. \tag{B.1}$$

To begin, the general solution of Eq. (4.4) is obtained as

$$\tilde{S}(r,\theta;s)=\tilde{a}(r;s)+\sum_{n=0}^{\infty}c_n\ \tilde{b}_n(r;s)\ \cos(n\theta) \tag{B.2}$$

where the first term is the solution of the non-homogeneous problem, the constants c_n need to be determined and $\tilde{b}_n(r;s)$ is the solution of the differential equation

$$r^2\tilde{b}_n''(r;s)+r\tilde{b}_n'(r;s)-r^2\Big(\frac{n^2\pi^2}{L^2}+\frac{s^\alpha}{D_\alpha}\Big)\tilde{b}_n(r;s)=0. \tag{B.3}$$

By applying the appropriate boundary conditions, the solution of Eq. (B.3) is obtained as a linear combination of modified Bessel functions of the first, $I_\nu(x)$, and the second kind,

$K_\nu(x)$, as shown in Eq. (4.7) in Chapter 4. The solution of the non-homogeneous problem is obtained through Dirichlet boundary condition at $r = \rho$ such that

$$\tilde{a}(r;s) = \frac{1}{s}\left[1 - \frac{\tilde{b}_0(r;s)}{\tilde{b}_0(\rho;s)}\right]. \tag{B.4}$$

Following the method previously explained in detail for normally diffusing DBPs,[1] the effective flux Q and the constants c_n in terms of c_0, respectively, are derived. This independent constant c_0 is determined by using the condition described in Eq. (B.1) and eventually the expression for the survival probability, $\tilde{S}(r,\theta;s)$, is obtained. These details of the derivation can be readily followed from Ref. [1] as well and the final expressions of the quantities are

$$Q = \frac{\pi D_\alpha}{\epsilon\ \kappa}[\ \tilde{a}'(\rho;s)\ +\ c_0\ \tilde{b}_0'(\rho;s)\ ] \tag{B.5}$$

$$c_n = \frac{2\ Q\ \kappa}{\pi\ D_\alpha\ \tilde{b}_n'(\rho;s)}\frac{\sin(n\epsilon)}{n} \tag{B.6}$$

$$c_0 = \frac{1-\tilde{\eta}(\rho;s)}{s\ \tilde{b}_0(\rho;s)} \tag{B.7}$$

The parameter $\tilde{\eta}(\rho;s)$ is defined in Eq. (4.8) in Chapter 4. Finally, the survival probability in the Laplace domain is obtained as

$$\tilde{S}(r,\theta;s) = \frac{1}{s}\left[1 - \frac{\tilde{\eta}(\rho;s)}{\tilde{b}_0(\rho;s)}\left\{\tilde{b}_0(r;s) + 2\tilde{b}_0'(\rho;s)\sum_{n=1}^{\infty}\frac{\tilde{b}_n(r;s)}{\tilde{b}_n'(\rho;s)}\frac{sin(n\epsilon)}{n\epsilon}cos(n\theta)\right\}\right]. \tag{B.8}$$

From this expression of Survival probability, one can determine the FPTD, $\tilde{F}(r,\theta;s)$, which is described in Eq. (4.6). The surface-averaged FPTD, $\tilde{F}(r;s)$, represents the distribution for recognizing the target region for the first time provided the search of the DBP started from some fixed radial distance from the entire DNA molecule. This distribution is determined by integrating out the θ coordinate from Eq. (4.6),

$$\begin{aligned}\tilde{F}(r;s) &= \frac{1}{\theta}\int_0^{\theta}\tilde{F}(r,\theta;s)d\theta \\ &= \tilde{\eta}(\rho;s)\frac{\tilde{b}_0(r;s)}{\tilde{b}_0(\rho;s)}\end{aligned} \tag{B.9}$$

Using Eq. (4.9), one may also write $\tilde{F}(r;s) = \tilde{F}(r;s)_\epsilon + \tilde{\xi}(r;s)\frac{\tilde{b}_0'(\rho;s)}{\tilde{b}_0(\rho;s)}$. The most-probable time and the conditional MFPT in this case may be determined by using similar calculations as discussed in Chapter 4. If $r = \rho$, then $\tilde{F}(r = \rho; s) = \tilde{\eta}(\rho; s)$, i.e., the search of the subdiffusing DBP will begin from the surface of the DNA. Further, the surface-averaged FPTD can be useful in probing the 1D diffusion of the DBP while it searches for the target site along the DNA. This may be achieved by using the condition $R = r = \rho$ in the expression of $\tilde{F}(r;s)$.

Numerical simulations to determine the FPTD

To ensure the accuracy of the analytical result $\tilde{F}(r;s)_\epsilon$ in Eq. (4.9) which is obtained under SCA, numerical simulations were performed to solve to original FFPE for the FPTD $\tilde{F}(r,z;s)$. Here, note that the dimensionless notation for the z-coordinate is not being considered. By using the substitution present in Eq. (4.6a) within Eq. (4.4), one arrives at

$$\frac{\partial}{\partial r}\Big(r\frac{\partial}{\partial r}\Big)\tilde{F}(r,z;s) + \frac{\partial}{\partial z}\Big(r\frac{\partial}{\partial z}\Big)\tilde{F}(r,z;s) = r\frac{s^\alpha}{D_\alpha}\tilde{F}(r,z;s). \tag{B.10}$$

With the help of PDEtool in MATLAB, the solutions for $\tilde{F}(r,z;s)$ within appropriate boundary conditions can be obtained using Finite Elements Method at various values of the Laplace variable s.[1] A rectangular domain of dimensions $(\rho, R)\times(0, L)$ is constructed with reflective boundaries (Neumann boundary condition) applied everywhere except $(\rho)\times(0,\epsilon)$ where lies the partially absorbing boundary condition which is

$$\tilde{F}(\rho,z;s) - \frac{D_\alpha}{\kappa}\Big(\frac{\partial\tilde{F}(r,z;s)}{\partial r}\Big)_{r=\rho} = 1 \qquad (0 < z < \epsilon) \tag{B.11}$$

To determine the radially dependent FPTDs, $\tilde{F}(r;s)_\epsilon$, the solution first was linearly interpolated at a fixed radial distance followed by numerical integration according to Eq. (4.9). The results obtained through numerical calculations are in excellent agreement with the analytically determined FPTD within SCA (Fig. B.1). An agreement within the Laplace space assures an agreement within the time space as well because both the quantities are uniquely associated through Laplace transforms. Thus, to avoid time-consuming simulation analysis in the time domain, the FFPE was solved in the Laplace space along with other reasons discussed elsewhere.[1]

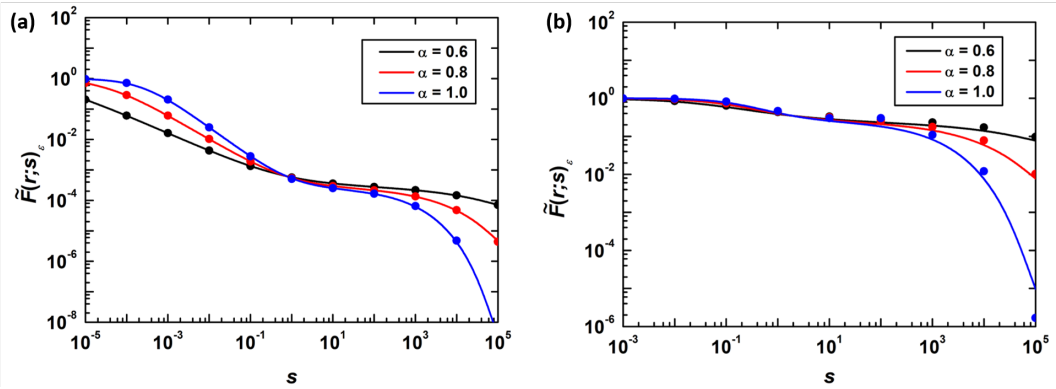


Figure B.1: The function $\tilde{F}(r;s)_\epsilon$ (solid lines) in Eq. (4.9) compared with the numerical solutions (filled circles) obtained as described in the text with respect to the dimensionless Laplace variable s at different values of α depicting the degree of sub-diffusion of the protein. Parameters chosen are: $R = 1$ μm, $L = \pi$ μm, $D_1 = 10$ μm^2 s^{-1}, $\rho = 2$ nm, $r = 0.1$ μm and $\epsilon = 0.2$ μm $\equiv 0.2$. For (a), $\kappa = 1$ μm s^{-1} and for (b), perfect reactions were assumed, $\kappa = \infty$.

Short time asymptotic expression of FPTDs - Most probable time

The short-time asymptotic of $F(r;t)_\epsilon$ may be obtained by applying the limit $s \to \infty$ in the Laplace space. By following the procedure detailed previously for normally diffusing DBP, the general expressions can be evaluated that are valid for any type of subdiffusing particle.[1] To find the short-time asymptote of $F(r;t)_\epsilon$, using Eq. (4.9) it is required to determine $\tilde{\xi}(r;s\to\infty)$ to eventually obtain $\tilde{\eta}(\rho;s\to\infty)$. The large-argument asymptotic expansion of $K_\nu(x\to\infty)$[3] can be written as

$$\frac{\tilde{b}_n(r;s)}{\tilde{b}'_n(\rho;s)} \simeq -\frac{K_0(rk_n)}{k_n K_1(\rho k_n)} \simeq -\frac{e^{-k_n(r-\rho)}}{k_n}\left[\sqrt{\frac{\rho}{r}} - \frac{3r+\rho}{8rk_n\sqrt{r\rho}} + ...\right] \tag{B.12a}$$

$$\frac{\tilde{b}'_0(\rho;s)}{\tilde{b}_0(\rho;s)} \simeq -k_0\left[1 + \frac{1}{2\rho k_0} - \frac{1}{8(\rho k_0)^2} + ...\right] \tag{B.12b}$$

Now, from the definition of $\tilde{\xi}(r;s)$ in Eq. (4.8b) and writing $1/k_n$ from Eq. (4.7b) as

$$\frac{1}{k_n} = \frac{1}{k_0}\left[1 - \frac{n^2\pi^2}{L^2k_n(k_n+k_0)}\right] \tag{B.13}$$

one finds,

$$\tilde{\xi}(r;s\to\infty) \simeq 2\sum_{n=1}^{\infty}\left(\frac{\sin(n\epsilon)}{n\epsilon}\right)^2 \times \frac{e^{-k_0(r-\rho)}}{k_0}\left[\sqrt{\frac{\rho}{r}} - \frac{1}{k_n}\left(\frac{n^2\pi^2\sqrt{\rho/r}}{L^2(k_n+k_0)} + \frac{3r+\rho}{8r\sqrt{r\rho}}\right) + ...\right] \tag{B.14}$$

To proceed, each term in Eq. (B.14) is considered separately. Using the identity $\sum_{n=1}^{\infty} \sin^2 n\epsilon/(n\epsilon)^2 = (\pi - \epsilon)/(2\epsilon)$, the first term of Eq. (B.14) becomes

$$\tilde{\xi}(r; s \to \infty)^{(1)} \simeq e^{-k_0(r-\rho)} \sqrt{\frac{\rho}{r}} \frac{\pi - \epsilon}{k_0 \epsilon}. \tag{B.15}$$

The second term is simplified to

$$\tilde{\xi}(r; s \to \infty)^{(2)} \simeq -\frac{2\pi^2}{k_0 \epsilon^2} e^{-k_0(r-\rho)} \sum_{n=1}^{\infty} \frac{\sin^2(n\epsilon)}{n^2\pi^2 + 2k_0^2 L^2}. \tag{B.16}$$

Considering $z = \sqrt{2} k_0 L$ and $x = \epsilon/\pi$, the identity

$$\sum_{n=1}^{\infty} \frac{\sin^2(nx)}{n^2\pi^2 + z^2} = \frac{1 - e^{-2xz}}{4z} \tag{B.17}$$

is used to eventually determine the second term of Eq. (B.14) which is

$$\begin{aligned} \tilde{\xi}(r; s \to \infty)^{(2)} &\simeq -\frac{\pi^2}{2\sqrt{2} k_0^2 \epsilon^2 L} \sqrt{\frac{\rho}{r}} e^{-k_0(r-\rho)} (1 - e^{-2\sqrt{2} k_0 L}) \\ &\simeq -\frac{\pi^2}{2\sqrt{2} k_0^2 \epsilon^2 L} \sqrt{\frac{\rho}{r}} e^{-k_0(r-\rho)} \end{aligned} \tag{B.18}$$

For the third term and other subsequent terms, the substitution $k_n \approx k_0$ is made as they are in higher orders of powers and they may even be neglected after the third term. Thus, after combining all the terms the function $\tilde{\xi}(r; s \to \infty)$ is finally obtained as

$$\tilde{\xi}(r; s \to \infty) \cong e^{-k_0(r-\rho)} \left[\sqrt{\frac{\rho}{r}} \frac{\pi - \epsilon}{k_0 \epsilon} - \frac{1}{k_0^2} \left\{ \frac{\pi^2 \sqrt{\rho/r}}{2\sqrt{2} k_0^2 \epsilon^2 L} + \frac{(3r + \rho)(\pi - \epsilon)}{8 r \epsilon \sqrt{r\rho}} \right\} + ... \right]. \tag{B.19}$$

After determining $\tilde{\xi}(r; s \to \infty)$, the function $\tilde{\eta}(\rho; s \to \infty)$ can now be obtained corresponding to perfect and imperfect reactions in the target search process after the subdiffusing DBP arrives at the target on the DNA. For the case of perfect reactions, i.e., $\kappa = \infty$,

$$\tilde{\eta}(\rho; s \to \infty) \simeq \frac{\epsilon}{\pi} + \frac{1}{2\sqrt{2} L k_0} + \frac{(\sqrt{2} L \epsilon + \rho\pi)}{8 L^2 \epsilon \rho k_0^2} + \mathcal{O}(k_0^{-3}) \tag{B.20}$$

whereas, for the case of imperfect reactions, i.e., $\kappa < \infty$,

$$\tilde{\eta}(\rho; s \to \infty) \simeq \frac{\epsilon\kappa}{D_\alpha \pi k_0} - \frac{\epsilon\kappa}{D_\alpha \pi k_0^2}\left(\frac{\kappa}{D_\alpha} + \frac{1}{2\rho}\right) + \mathcal{O}(k_0^{-3}). \tag{B.21}$$

Now, only $\tilde{b}_0(r; s \to \infty)$ is remaining to determine the function $\tilde{F}(r; s \to \infty)$. From the asymptotic expansions of $K_\nu(x)$ and $I_\nu(x)$, one finds

$$\begin{aligned}\tilde{b}_0(r; s \to \infty) &\cong \frac{\cosh\big(k_0(R-r)\big)}{\sqrt{k_0^2 rR}}\left[1 - \frac{(3r+R)}{8k_0 rR}\tanh\big(k_0(R-r)\big)\right] \\ &\cong \frac{\cosh\big(k_0(R-r)\big)}{\sqrt{k_0^2 rR}}.\end{aligned} \tag{B.22}$$

Thus, eventually leading to

$$\frac{\tilde{b}_0(r; s)}{\tilde{b}_0(\rho; s)} \cong \sqrt{\frac{\rho}{r}}\, e^{-k_0(r-\rho)}. \tag{B.23}$$

Finally, after combining Eqs. (B.12), (B.20), (B.21) and (B.23), the short time FPTDs for different cases of imperfect and perfect reactions in the target search process were determined. In the case of perfect reactions, $\kappa = \infty$,

$$\tilde{F}(r; s \to \infty)_\epsilon \simeq \sqrt{\frac{\rho}{r}} e^{-(r-\rho)\sqrt{s^\alpha/D_\alpha}}\left[1 + \frac{(\pi-\epsilon)(r-\rho)}{8\pi r\rho\sqrt{s^\alpha/D_\alpha}} + \mathcal{O}(s^{-\alpha})\right] \tag{B.24}$$

and in the case of imperfect reactions, $\kappa < \infty$,

$$\tilde{F}(r; s \to \infty)_\epsilon \simeq \kappa\sqrt{\frac{\rho}{rD_\alpha s^\alpha}} e^{-(r-\rho)\sqrt{s^\alpha/D_\alpha}} \tag{B.25}$$

The inverse Laplace transform of the above equations (B.24)-(B.24) is obtained in terms of Wright functions[4] described by $W_{\lambda,\mu}(x) = \sum_{n=0}^{\infty} x^n/[n!\Gamma(\mu + \lambda n)]$ and are provided in Eq. (4.11). To determine the most probable time, t_{mp}, the method of derivatives is adopted. The time derivative of Eq. (4.11b) after simplifications results in

$$W_{-\frac{\alpha}{2},\frac{\alpha}{2}-1}\left(\frac{\rho - r}{\sqrt{D_\alpha t^\alpha}}\right) = 0 \tag{B.26}$$

From the relation $W_{\lambda,\mu-1}(z) + (1-\mu)W_{\lambda,\mu}(z) = \lambda z W_{\lambda,\lambda+\mu}(z)$ known[5] and using the asymptotic expansion of the Wright functions[6], one can obtain t_{mp} provided in Eq. (4.12) within Chapter 4. The most-probable time increases with respect to the parameter α (Fig. B.2(a)) and the initial distance of the DBP (Fig. B.2(b)).

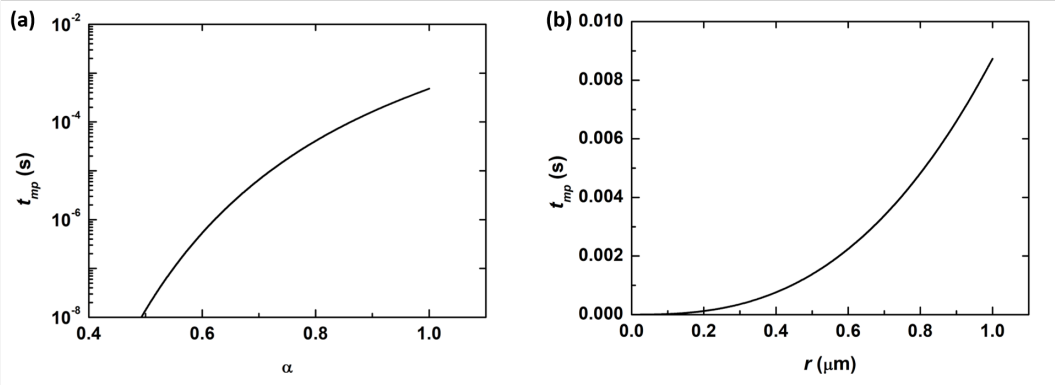


Figure B.2: The time t_{mp} in Eq. (4.12) with respect to (a) α and (b) the initial distance, r. Parameters used are: $D_1 = 10\ \mu\text{m}^2\ \text{s}^{-1}$ and $\rho = 0.2$ nm. For (a), $r = 0.1\ \mu$m and for (b), $\alpha = 0.75$.

Large time asymptotic expression of FPTDs - Conditional mean time

Since the MFPT may be analytically determined using the expression

$$\tilde{S}(r; s=0) = \left(\frac{1-\tilde{F}(r;s)}{s}\right)_{s=0}, \tag{B.27}$$

the large-time asymptotic of $\tilde{F}(r;s)_\epsilon$ is determined by applying the limit $s \to 0$. To begin with, from the definition of k_n in Eq. (4.7b), one may write the function $\tilde{\xi}(r; s \to 0) \cong \tilde{\xi}(r; s=0)$. Next, the small-argument asymptotic expansion of the modified Bessel functions[3] is used to obtain

$$\left(\frac{\tilde{b}_0(r;s)}{\tilde{b}_0(\rho;s)}\right)_{s\to 0} \simeq 1 - \frac{R^2 \ln(r/\rho)}{2D_\alpha} s^\alpha + \mathcal{O}(s^{2\alpha}) \tag{B.28a}$$

$$\left(\frac{\tilde{b}_0'(r;s)}{\tilde{b}_0(\rho;s)}\right)_{s\to 0} \simeq -\frac{(R^2-\rho^2)}{2\rho D_\alpha} s^\alpha + \mathcal{O}(s^{2\alpha}). \tag{B.28b}$$

Using the above mentioned approximations in Eq. (4.8), the long-time approximation is determined for

$$\tilde{\eta}(\rho; s \to 0) \simeq \left[1 + \left\{\frac{D_\alpha \pi}{\kappa\epsilon} + \tilde{\xi}(\rho; s=0)\right\} \frac{(R^2-\rho^2)}{2\rho D_\alpha} s^\alpha\right]^{-1} \tag{B.29}$$

Finally, the large-time expression of the FPTD $\tilde{F}(r;s)_\epsilon$ in the Laplace space is written as

$$\tilde{F}(r; s \to 0)_\epsilon \simeq \frac{1 + (C - B)s^\alpha}{1 + As^\alpha} \tag{B.30}$$

where, the constants A, B and C are defined Chapter 4's Eq. (4.15). Now, to obtain the MFPT for the case when the subdiffusing DBP ($\alpha < 1$) starts its search from a fixed *radial* distance away from the target, Eq. (B.27) is used to determine

$$\text{MFPT} = \tilde{S}(r; s = 0)_\epsilon \simeq \left(\frac{(A + B - C)s^{\alpha - 1}}{1 + As^\alpha} \right)_{s=0} = \infty. \tag{B.31}$$

It should also be noted that if $\alpha = 1$ for a normally diffusing DBP, the MFPT would have been a finite quantity $\text{MFPT}(\alpha = 1) = \tilde{S}(r; s = 0)_\epsilon = A + B - C$. The divergence of the MFPT for a single subdiffusing DBP is the reason this work obtains a conditional MFPT, T_c, instead. The inverse Laplace transform of $\tilde{F}(r; s \to 0)_\epsilon$ (which is provided in Eq. (4.14)) is performed using properties of Mittag-Leffler functions,[7, 8] defined by $E_{p,q}(y) = \sum_{n=0}^{\infty} y^n / \Gamma(q + pn)$. The behaviour of time T_c (Eq. (4.16) in Chapter 4) with respect to different parameters is presented in Fig. (B.3) and the corresponding discussions are given in Chapter 4.

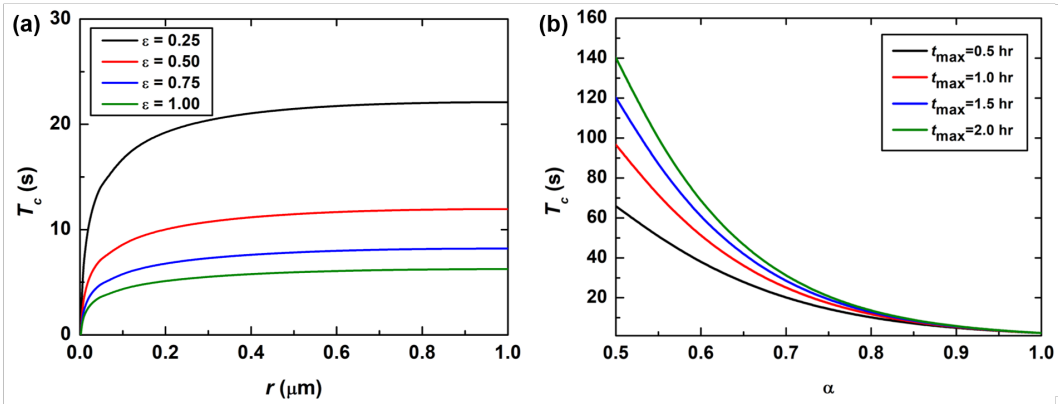


Figure B.3: The time T_c obtained via Eq. (4.16) represented as a function of (a) initial distance between the DBP and the target, r and (b) the anomaly exponent α. Other parameters used are: $R = 1$ μm, $L = \pi$ μm, $D_1 = 10$ μm^2 s^{-1}, $\kappa = \infty$ and $\rho = 2$ nm. For (a), $t_{max} = 3600$ s (1 hour) and $\alpha = 0.75$ while for (b), $\epsilon = 0.25$ μm $\equiv 0.25$ and $r = 0.1$ μm.

Application of the theory to real systems

The typical dimensions of the nucleus of U2OS cells are 10 μm $\times$ 10 μm $\times$ 6 μm.[9] Suppose that a single P-TEFb protein is initially situated 0.25 μm away from the target

site which is roughly about 1 μm in size. The apparent radius of the DNA is considered to be 2 nm.[10] Thus, according to the proposed theoretical model, $R = 5$ μm, $L = 6$ μm, $r = 0.25$ μm, $\epsilon = 1$ μm and $\rho = 2$ nm. Additionally, the target is considered to be perfectly absorbing, i.e., $\kappa = \infty$ for convenience. Since, it is more appropriate to consider the target site in the middle of the DNA, one may as well halve the values of L and ϵ. Moreover, the dimensionless form of the target size must be considered to use the obtained theoretical results and thus, $\epsilon \equiv \pi\epsilon/L = 0.52$. From Eq. (4.17), the anomaly exponent $\alpha = 0.61$ and $D_1 = 1.46$ μ m^2/s as discussed in Chapter 4. All these parameters are used in Eq. (4.9) to determine the distance dependent FPTD, $\tilde{F}(r;s)_\epsilon$, in the Laplace space. Numerical inversion of $\tilde{F}(r;s)_\epsilon$ is performed using Gaver-Stehfest method[11] and the results are shown in Fig. 4.4 within Chapter 4.

Using Eq. (4.12), the most-probable times are found to be $t_{mp}(\alpha = 0.61) = 3.54 \times 10^{-4}$ s and $t_{mp}(\alpha = 1) = 0.021$ s. The single particle tracking experiments were performed for about 45 minutes which means that the observation time becomes $t_{max} = 2700$ s and thus, by using Eq. (4.16) the conditional MFPTs for these two situations are evaluated to be $T_c(\alpha = 0.61) = 897.14$ s $\simeq 15$ minutes and $T_c(\alpha = 1) = 231.42$ s $\simeq 4$ minutes. It should also be noted that the numerical value of the mean times are quite high because this work considered the target search of a single DBP within the system and for multiple particles searching for the same target, the time would be less.[12, 13]

Bibliography

(1) Grebenkov, D. S. et al. Towards a Full Quantitative Description of Single-Molecule Reaction Kinetics in Biological Cells. *Phys. Chem. Chem. Phys.* **2018**, *20*, 16393–16401.

(2) Shoup, D. et al. Diffusion-Controlled Bimolecular Reaction Rates. The Effect of Rotational Diffusion and Orientation Constraints. *Biophys. J.* **1981**, *36*, 697–714.

(3) Abramowitz, M.; Stegun, I. A., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*; US Government printing office: 1948; Vol. 55.

(4) Gorenflo, R. et al. Analytical Properties and Applications of the Wright Function. *Fract. Calc. Appl. Anal.* **1999**, *2*, 383–414.

(5) El-Shahed, M.; Salem, A. An Extension of Wright Function and its Properties. *J. Math.* **2015**, *2015*.

(6) Luchko, Y. Asymptotics of Zeros of the Wright Function. *Journal for Analysis and its Applications* **2000**, *19*, 583–595.

(7) Haubold, H. J. et al. Mittag-Leffler Functions and Their Applications. *J. Appl. Math.* **2011**, *2011*.

(8) Apelblat, A. Differentiation of the Mittag-Leffler Functions With Respect to Parameters in the Laplace Transform Approach. *Mathematics* **2020**, *8*, 657.

(9) Izeddin, I. et al. Single-Molecule Tracking in Live Cells Reveals Distinct Target-Search Strategies of Transcription Factors in the Nucleus. *Elife* **2014**, *3*, e02230.

(10) Grebenkov, D. S. et al. Effects of the Target Aspect Ratio and Intrinsic Reactivity Onto Diffusive Search in Bounded Domains. *New J. Phys.* **2017**, *19*, 103025.

(11) Stehfest, H. Algorithm 368: Numerical Inversion of Laplace Transforms [D5]. *Commun. ACM* **1970**, *13*, 47–49.

(12) Grebenkov, D. S. Subdiffusion in a Bounded Domain with a Partially Absorbing-Reflecting Boundary. *Phys. Rev. E* **2010**, *81*, 021128.

(13) Martens, K. J. et al. Visualisation of dCas9 Target Search In Vivo using an Open-Microscopy Framework. *Nat. Commun.* **2019**, *10*, 3552.

Appendix C

Determination of the first passage time probability distribution functions

In this section, the calculations for obtaining the general equation of the first passage probability distribution functions for the stochastic discrete-state model presented in Fig. 5.4 in Chapter 5 are discussed. At $z = 1\ \mathrm{s}^{-1}$, the Laplace transform of Eq. (5.1) is written as

$$(s + x + w_0)\tilde{F}_0(s) = x\tilde{F}_1(s) + w_0 \tag{C.1a}$$

for the state $n = 0$ and when $n \in [1, N-1]$,

$$(s + ny + w_n + 1)\tilde{F}_n(s) = ny\tilde{F}_{n+1}(s) + \tilde{F}_{n-1}(s) + w_n. \tag{C.1b}$$

If the nanopore is occupied by the maximum PEG crowders possible, $n = N$, then the Laplace transform of Eq. (5.1c) can be written as

$$(s + w_N + 1)\tilde{F}_N(s) = \tilde{F}_{N-1}(s) + w_N. \tag{C.1c}$$

It should be noted that the Laplace transform of the boundary condition is $\tilde{F}_T(s) = 1$. The linear equations above are solved using Thomas Algorithm to solve tridiagonal matrices and the general expression is determined to be

$$\tilde{F}_n(s) = 1 - \frac{s}{\tilde{B}_n(s)} \times \left[\tilde{A}_n(s) + \sum_{i=1}^{N-n} \tilde{A}_{n+i}(s)\tilde{g}_i(s)\right]. \tag{C.2a}$$

Here,

$$\tilde{A}_n(s) = 1 + \sum_{k=1}^{n} \Pi_{l=1}^{k} \frac{1}{\tilde{B}_{n-l}(s)} \tag{C.2b}$$

$$\tilde{g}_i(s) = \Pi_{j=1}^{i} \frac{\delta_{(0,n+j-1)}x + (n+j-1)y}{\tilde{B}_{n+j}(s)} \tag{C.2c}$$

$$\tilde{B}_n(s) = q_0(s) + \mathcal{K}_{i=1}^{n} \frac{p_i}{q_i(s)}. \tag{C.2d}$$

In the above set of equations, $\delta_{(m_1,m_2)}$ represents Kronecker delta and $\mathcal{K}_{i=1}^{n}$ is a continued fraction operator with $p_i = -[\delta_{(0,n-i)}x + (n-i)y]$ and $q_i(s) = s + u_{n-i} + \big(1 - \delta_{(0,n-i)}\big) + [\delta_{(0,n-i)}x + (n-i)y] \times \big(1 - \delta_{(N,n-i)}\big)$. The function $\tilde{F}_n(s)$ defined in Eq. (C.2) is the first passage probability distribution for the DNA capture or translocation process, presented in Fig. 5.4, to occur for the first time if the nanopore was initially occupied by n molecules of PEG from the *cis* side.

Steady-state probabilities of the system

For the kinetic scheme Fig. 5.4 let us define $P_n(t)$ as the probability for a protein nanopore, in either an empty or blocked state, to be in a microscopic state n such that the pore is occupied by n PEG molecules at time t. The evolution of probabilities, at $z = 1$ s^{-1}, is described through a set of forward master equations as

$$\frac{dP_n(t)}{dt} = (n-1)yP_{n-1}(t) + P_{n+1}(t) - (ny+1)P_n(t) \tag{C.3a}$$

for $2 \leq n \leq N-1$. For the microscopic state $n = 0$,

$$\frac{dP_0(t)}{dt} = P_1(t) - xP_0(t) \tag{C.3b}$$

whereas, for $n = 1$, the forward master equation is written as

$$\frac{dP_1(t)}{dt} = xP_0(t) + P_2(t) - (y+1)P_1(t). \tag{C.3c}$$

Finally, when the nanopore is occupied by maximum PEG crowders,

$$\frac{dP_N(t)}{dt} = (N-1)yP_{N-1}(t) - P_N(t). \tag{C.3d}$$

When the system achieves steady-state, the probabilities will become independent of time and $dP_n(t)/dt = 0$. Thus, Eq. (C.3) may be rewritten in the steady-state as

$$(ny+1)P_n = (n-1)yP_{n-1} + P_{n+1} \tag{C.4a}$$

for $2 \leq n \leq N-1$. If the nanopore is not occupied by any PEG molecules, $n = 0$,

$$xP_0 = P_1 - \tag{C.4b}$$

whereas, for $n = 1$, the steady-state equation is written as

$$(y+1)P_1 = xP_0 + P_2. \tag{C.4c}$$

Finally, when $n = N$,

$$P_N = (N-1)yP_{N-1}. \tag{C.4d}$$

After using the normalisation condition, i.e., $\sum_{i=0}^{N} P_i = 1$, the steady-state probabilities for each state can be obtained and they are presented in Eq. (5.4).

Determination of the mean DNA capture or translocation time from steady-state probability distributions

The total flux for the irreversible transition of DNA capture or translocation to occur is obtained from the steady-state probabilities of all states as

$$J = \sum_{n=0}^{N} w_n P_n \tag{C.5a}$$

Using the probabilities in the steady-state defined in Eq. (5.4),

$$J = \frac{w_0 + x\sum_{i=1}^{N}(i-1)!w_n y^{i-1}}{1 + x\sum_{i=0}^{N-1} i!y^i}. \tag{C.5b}$$

From the total flux of the system in Eq. (C.5b), the mean time for the irreversible transition is determined from its reciprocal which is also presented in Eq. (5.7) within Chapter 5.

Appendix D

Stationary-state solutions for the binary mixture

Let $P_{n_1,n_2}(t)$ be the probability that the system exists in the state (n_1, n_2) at time t. The evolution of probabilities can be described by a set of forward master equations. For a general state (n_1, n_2),

$$\begin{aligned}\frac{dP_{n_1,n_2}(t)}{dt} &= \Big(p_{n_1-1,n_2}P_{n_1-1,n2}(t) + z_1P_{n_1+1,n_2}(t)\Big) \\ &+ \Big(q_{n_1,n_2-1}P_{n_1,n2-1}(t) + z_2P_{n_1,n_2+1}(t)\Big) \\ &- \Big(p_{n_1,n_2} + q_{n_1,n_2} + 2z_1\Big)P_{n_1,n_2}(t)\end{aligned} \tag{D.1a}$$

such that $n_1 \in (0, N_1)$ and $n_2 \in (0, N_2)$. At the edges and the corner states of the 2D network described in Fig. 6.2, the equations are slightly different. When $n_1 = 0$ and $n_1 = N_1$, respectively, one obtains

$$\begin{aligned}\frac{dP_{0,n_2}(t)}{dt} &= z_1P_{1,n_2}(t) + \Big(q_{0,n_2-1}P_{0,n_2-1}(t) + z_2P_{0,n_2+1}(t)\Big) \\ &- \Big(p_{0,n_2} + q_{0,n_2} + z_2\Big)P_{0,n_2}(t)\end{aligned} \tag{D.1b}$$

$$\begin{aligned}\frac{dP_{N_1,n_2}(t)}{dt} &= p_{N_1-1,n_2}P_{N_1-1,n_2}(t) + \Big(q_{N_1,n_2-1}P_{N_1,n_2-1}(t) + z_2P_{N_1,n_2+1}(t)\Big) \\ &- \Big(q_{N_1,n_2} + 2z_1\Big)P_{N_1,n_2}(t)\end{aligned} \tag{D.1c}$$

whereas if $n_2 = 0$ and $n_2 = N_2$, respectively, one writes

$$\frac{dP_{n_1,0}(t)}{dt} = z_2 P_{n_1,1}(t) + \Big(p_{n_1-1,0}P_{n_1-1,0}(t) + z_1 P_{n_1+1,0}(t)\Big) - \Big(p_{n_1,0} + q_{n_1,0} + z_1\Big)P_{n_1,0}(t) \tag{D.1d}$$

$$\frac{dP_{n_1,N_2}(t)}{dt} = q_{n_1,N_2-1}P_{n_1,N_2-1}(t) + \Big(p_{n_1-1,N_2}P_{n_1-1,N_2}(t) + z_1 P_{n_1+1,N_2}(t)\Big) - \Big(p_{n_1,N_2} + 2z_1\Big)P_{n_1,N_2}(t). \tag{D.1e}$$

For the corner states $(0,0)$, $(N_1,0)$ and $(0,N_2)$, respectively, the forward master equations are written as

$$\frac{dP_{0,0}(t)}{dt} = z_1 P_{1,0}(t) + z_2 P_{0,1}(t) - (p_{0,0} + q_{0,0})P_{0,0}(t) \tag{D.1f}$$

$$\frac{dP_{N_1,0}(t)}{dt} = p_{N_1-1,0}P_{N_1-1,0}(t) + z_2 P_{N_1,1}(t) - (z_1 + q_{N_1,0})P_{N_1,0}(t) \tag{D.1g}$$

$$\frac{dP_{0,N_2}(t)}{dt} = z_1 P_{1,N_2}(t) + q_{0,N_2-1}P_{0,N_2-1}(t) - (z_1 + p_{0,N_2})P_{0,N_2}(t). \tag{D.1h}$$

Note that the state $(N1, N2)$ will never exist according to Eq. (6.3a) in Chapter 6 since N_1 and N_2 are maximum crowders of type 1 and 2 PEG polymers, respectively, to accommodate the nanopore. At all times, the probabilities $P_{n_1,n_2}(t)$ are normalized such that

$$\sum_{n_2}\sum_{n_1} P_{n_1,n_2}(t) = 1. \tag{D.2}$$

Assuming the system reaches stationary state quickly, the solutions of P_{n_1,n_2} are evaluated. When the time $t \to \infty$, one can rewrite the forward master equations in Eq. (D.1) as follows. For a general state (n_1, n_2),

$$\Big(p_{n_1,n_2} + q_{n_1,n_2} + 2z_1\Big)P_{n_1,n_2} = \Big(p_{n_1-1,n_2}P_{n_1-1,n2} + z_1 P_{n_1+1,n_2}\Big) + \Big(q_{n_1,n_2-1}P_{n_1,n2-1} + z_2 P_{n_1,n_2+1}\Big) \tag{D.3a}$$

such that $n_1 \in (0, N_1)$ and $n_2 \in (0, N_2)$. When $n_1 = 0$ and $n_1 = N_1$, respectively,

$$\Big(p_{0,n_2} + q_{0,n_2} + z_2\Big)P_{0,n_2} = z_1 P_{1,n_2} + \Big(q_{0,n_2-1}P_{0,n_2-1} + z_2 P_{0,n_2+1}\Big) \tag{D.3b}$$

$$\Big(q_{N_1,n_2} + 2z_1\Big)P_{N_1,n_2} = p_{N_1-1,n_2}P_{N_1-1,n_2} + \Big(q_{N_1,n_2-1}P_{N_1,n_2-1} + z_2 P_{N_1,n_2+1}\Big) \tag{D.3c}$$

whereas if $n_2 = 0$ and $n_2 = N_2$, respectively,

$$\Big(p_{n_1,0} + q_{n_1,0} + z_1\Big)P_{n_1,0} = z_2 P_{n_1,1} + \Big(p_{n_1-1,0}P_{n_1-1,0} + z_1 P_{n_1+1,0}\Big) \tag{D.3d}$$

$$\Big(p_{n_1,N_2} + 2z_1\Big)P_{n_1,N_2} = q_{n_1,N_2-1}P_{n_1,N_2-1} + \Big(p_{n_1-1,N_2}P_{n_1-1,N_2} + z_1 P_{n_1+1,N_2}\Big). \tag{D.3e}$$

For the corner states $(0,0)$, $(N_1,0)$ and $(0,N_2)$, respectively, the forward master equations in the steady-state are rewritten as

$$(p_{0,0} + q_{0,0})P_{0,0} = z_1 P_{1,0} + z_2 P_{0,1} \tag{D.3f}$$

$$(z_1 + q_{N_1,0})P_{N_1,0} = p_{N_1-1,0}P_{N_1-1,0} + z_2 P_{N_1,1} \tag{D.3g}$$

$$(z_1 + p_{0,N_2})P_{0,N_2} = z_1 P_{1,N_2} + q_{0,N_2-1}P_{0,N_2-1}. \tag{D.3h}$$

Finally, by utilising the normalisation condition provided in Eq. (D.2) and solving the set of linear equations in Eq. (D.3), the steady-state probabilities for all discrete states in the system are determined.

Inter-event frequency for different compositions of binary mixtures

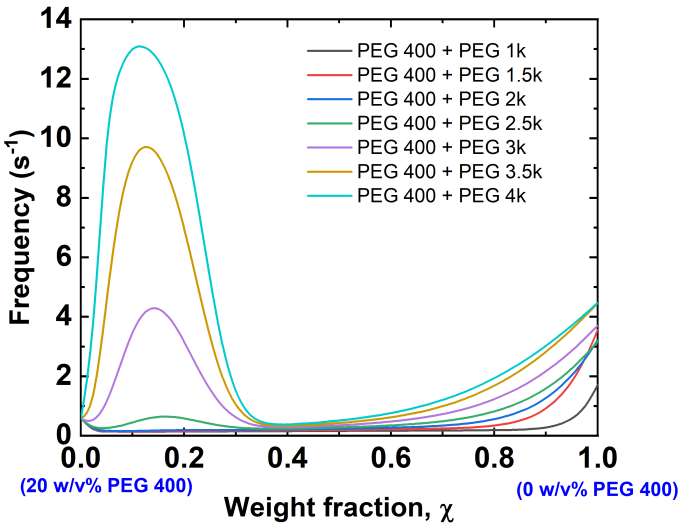


Figure D.1: Inter-event frequencies of ssDNA capture event as a function of χ at total monomeric concentration 20 w/v % for a binary mixture. The higher molecular weight crowder varies between PEG1k to PEG 4k polymers with PEG 400 being the smaller crowder in the mixture. The results are determined using Eq. (6.5) in Chapter 6 by numerical calculations. R_V and N values used in analytical calculations are listed in Table 6.1 within Chapter 6. Additional parameters are $a_0' = 33.98$ M^{-1} s^{-1},$b_0' = 384.49$ M^{-1} s^{-1}, $u_0 = 0.11$ s^{-1} and $u_1' = 15.95$ s^{-1}.

Appendix E

Randomness parameter for the sequential model in the limit of very large number of active sites

To understand the dynamics in the system in the limit of large number of active sites ($N \gg 1$), the most probable chemical state in the stationary regime may be determined. For this purpose, it is required to find the maximum of the stationary probability distribution in Eq. (7.10) in Chapter 7. This can be done with the help of Stirling approximation and considering $\log P_n(N)$ instead of $P_n(N)$, which gives us

$$\log P_n(N) \simeq N \log N - (N-n)\log(N-n) - n \log n + n \log x - N \log(1+x). \quad \text{(E.1)}$$

The most probable state is at $n = n^*$, which is determined from

$$\frac{\partial \log P_n(N)}{\partial n} = 0, \quad \text{(E.2)}$$

yielding

$$n^* \simeq \frac{N}{M+1}. \quad \text{(E.3)}$$

Then it can be shown that $P_{n^*}(N) \simeq 1$ and $f_{n^*} \simeq 1$, which means that the distribution is sharply peaked at $n = n^*$. Applying these results to Eqs. (7.7) and (7.8) in Chapter 7, one can obtain

$$< \tau >_N \simeq \frac{M+1}{N}, \quad < \tau^2 >_N \simeq 2\left(\frac{M+1}{N}\right)^2, \quad \text{(E.4)}$$

which eventually leads to $R_{N\to\infty} \simeq 1$.

Mean reaction time and randomness parameter for the branching model

let us define $P(CS_j)$ and $P(CS_j^*)$ as the probabilities to be found in the states CS_j and CS_j^* respectively. The evolution of the system can be described by the following master equations

$$\begin{aligned}\frac{dP(CS_j)}{dt} &= uP(CS_{j-1}) + v_2P(CS_j^*) - (u+v_1)P(CS_j);\\ \frac{dP(CS_j^*)}{dt} &= v_1P(CS_j) - v_2P(CS_j^*);\end{aligned} \tag{E.5}$$

with the normalization condition, $\sum_{j=0}^{M}(P(CS_j) + P(CS_j^*)) = 1$. Here j goes from 0 to M. $CS_0 = C$ and $CS_0^* = C^*$. At infinite time, $t \to \infty$, the system reaches a steady state and the steady state probabilities are given by

$$\begin{aligned}P(CS_j) &= \frac{1}{(M+1)(1+\frac{v_1}{v_2})};\\ P(CS_j^*) &= \frac{\frac{v_1}{v_2}}{(M+1)(1+\frac{v_1}{v_2})};\end{aligned} \tag{E.6}$$

The stationary state flux to form the product molecules P is then given by

$$J = uP(CS_M) = \frac{u}{(M+1)(1+\frac{v_1}{v_2})} \tag{E.7}$$

Let us consider that this chemical reaction takes place on each active of a catalyst with N active sites. For this system, an effective kinetic model as shown in Figure 7.6(b) is developed. At $t \to \infty$, the following relations can be written for the stationary state probabilities

$$\begin{aligned}NrP_0 &= (u+v_1)P_1;\\ (N-1)rP_1 &= 2(u+v_1)P_2;\\ &....\\ (N-n)rP_n &= (n+1)(u+v_1)P_{n+1};\end{aligned} \tag{E.8}$$

where P_n is the stationary probability of the state n (n=0,1,2,3,...N). Considering the normalization condition, $\sum_{n=0}^{N} P_n = 1$, these equations can be easily solved to obtain the steady-state probabilities

$$P_n(N) = \frac{N!}{(N-n)!n!} \frac{x^{\mathrm{n}}}{(1+x)^N}, \tag{E.9}$$

where $x = \frac{r}{u+v_1}$. The overall flux can be shown as

$$J_N = uP_1 + 2uP_2 + ... + NuP_N = \frac{Nru}{r+u+v_1} \tag{E.10}$$

where

$$r = \frac{u+v_1}{M + (M+1)\frac{v_1}{v_2}}. \tag{E.11}$$

The method of first-passage probabilities is employed to calculate the higher moments. Consider the first passage probability density function $F_n(t)$ which expresses the probability to form a product for the first time starting from the state 'n' at time $t = 0$. The Laplace transform of the set of backward master equations corresponding to each state is given by

$$\begin{aligned}
(s+Nr)F_0(s) &= NrF_1(s); \\
(s+(N-n)r+n(u+v_1))F_n(s) &= (N-n)rF_{n+1}(s) + nv_1F_{n-1}(s) + nu; n \in [1, N-1] \\
(s+N(u+v_1))F_N(s) &= Nv_1F_{N-1}(s) + Nu;
\end{aligned} \tag{E.12}$$

This set of equations are solved to obtain $F_n(s)$ by using the Thomas Algorithm to solve tridiagonal matrices which results in

$$F_n(s) = \frac{S_n(s) + (N-n)r - sA_n(s)}{C_n(s)} \tag{E.13}$$

where $A_n(s) = \sum_{i=0}^{N-n-1} \prod_{j=0}^{i} \frac{(N-n-j)r}{C_{n+j+1}(s)}(1 + \sum_{k=0}^{n+i} \prod_{l=0}^{k} \frac{(n+i+1-l)v_1}{C_{n+i-l}(s)})$, $S_n(s) = \sum_{i=0}^{n-1}(n-i)u \prod_{j=1}^{i} \frac{(n-j+1)v_1}{C_{n-j}(s)}$.

Here $C_n(s) = b_0 + \mathbf{K}_{i=1}^{n} \frac{a_i}{b_i}$ where $\mathbf{K}_{i=1}^{n}$ is a continued fraction operator with $a_i = -(n+1-i)v_1(N-(n-i))r$ and $b_i = s + (N-(n-i))r + (n-i)(u+v_1)$. $C_n(s)$ can be

expressed in terms of $S_n(s)$ via the relation $C_n(s) = S_n(s) + (N-n)r + s(1+B_n(s))$ with $B_n(s) = \sum_{p=0}^{n-1} \prod_{q=0}^{p} \frac{(n-q)v_1}{C_{n-q-1}(s)}$. Using the above relations, the average time of product formation for each individual state (mean first passage time) can be obtained which can be calculated by taking the first derivative of $F_n(s)$.

$$< T_n > = (\frac{\partial \widetilde{F_n}(s)}{\partial s})_{s=0} = \frac{1 + A_n(0) + B_n(0)}{C_n(0)} \tag{E.14}$$

As discussed in Chapter 7, a probability factor f_n is introduced to evaluate the mean reaction (catalytic turnover) times for chemical reactions described in Fig. 6a that are taking place on the catalyst with N active sites

$$< \tau >_N = \sum_{n=0}^{N-1} f_n < T_n > = \frac{r + u + v_1}{Nru} \tag{E.15}$$

where

$$f_n = \frac{r^n (u+v_1)^{N-n-1} (\frac{N-n}{n})}{(r+u+v_1)^{N-1}} * \frac{N!}{(N-n)!n!} = P_n(N-1) \tag{E.16}$$

The mean square first passage time can be similarly derived and is given by

$$< T_n^2 > = (\frac{\partial^2 \widetilde{F_n}(s)}{\partial s^2})_{s=0} = \frac{2[C_n^{'}(0) < T_n > -B_n^{'}(0) - A_n^{'}(0)]}{C_n(0)} \tag{E.17}$$

The terms $A_n^{'}(s), B_n^{'}(s)$, and $C_n^{'}(s)$ are the first derivatives. Following the same method, one can proceed to calculate the second moment of the reaction time

$$< \tau^2 >_N = \sum_{n=0}^{N-1} f_n < T_n^2 > \tag{E.18}$$

From Eq. (E.15) and (E.18), the randomness parameter can be calculated as $R = \frac{<\tau^2>_N - <\tau>_N^2}{<\tau>_N^2}$.

Appendix F

Specific calculations of stationary probabilities

Substituting Eq. (8.3) in the second line of Eq. (8.1) and rearranging gives $P_0(t)$

$$P_0(t) = \frac{u\big(P_{-1}(t) + P_1(t)\big) + \frac{rk}{r+k}}{2u + r} \tag{F.1}$$

To obtain a general form of $P_l(t)$, the homogeneous recurrence relation given by Eq. (8.1) is solved. To proceed, the characteristic equation of Eq. (8.1) in written in terms of a variable x.

$$ux^{l-1} + ux^{l+1} - (2u + r)x^l = 0 \tag{F.2}$$

This simplifies into a quadratic equation

$$u + ux^2 - (2u + r)x = 0 \tag{F.3}$$

and the roots are given by

$$x_1 = \frac{2u + r - \sqrt{(2u + r)^2 - 4u^2}}{2u} \tag{F.4}$$

$$x_2 = \frac{2u + r + \sqrt{(2u + r)^2 - 4u^2}}{2u} \tag{F.5}$$

and $x_2 = x_1^{-1}$. The probability density function can be represented as a linear combination of these two roots. The general form of the probability densities $P_l(t)$ for different regions on the nanorod with respect to 0^{th} state is

$$P_l(t) = \begin{cases} Ax^l + Bx^{-1} & l \in [-L, -1] \\ Cx^l + Dx^{-1} & l \in [1, L] \end{cases} \tag{F.6}$$

where $x \equiv x_1$. The constants A, B, C and D are determined by applying the boundary conditions and the normalisation condition. At $l = -L$ and L, the following results are obtained

$$A = Bx^{2L+1} \tag{F.7}$$

$$D = Cx^{2L+1} \tag{F.8}$$

At $l = 1$,

$$(2u + r)P_1(t) = uP_0(t) + uP_2(t) \tag{F.9}$$

$$(2u + r)P_1(t) \ -uP_2(t) = uP_0(t) \tag{F.10}$$

At $l = -1$,

$$(2u + r)P_{-1}(t) = uP_0(t) + uP_{-2}(t) \tag{F.11}$$

$$(2u + r)P_1(t) \ -uP_{-2}(t) = uP_0(t) \tag{F.12}$$

Dividing Eq. (F.9) by Eq. (F.11) leads to the following equation

$$(2u + r)P_1(t) - uP_2(t) = (2u + r)P_{-1}(t) - uP_{-2}(t). \tag{F.13}$$

Substituting the form of $P_l(t)$ provided by Eq. (F.6) in the Eq. (F.13) leads to $B = C$. Using this relation and Eq. (F.1), the steady state probability densities can be expressed as

$$P_l(t) = \begin{cases} B(x^{2L+1+l} + x^l) & l \in [-L, -1] \\ \frac{2uB(x+x^{2L}) + \frac{rk}{r+k}}{2u+r} & l = 0 \\ B(x^l + x^{2L+1-l}) & l \in [1, L] \end{cases} . \tag{F.14}$$

Calculation of Pearson coefficient for catalytic communication within single Pd nanorod

To probe whether catalytic reactions occurring at different sites can communicate with each other within a single nanocatalyst, one can compute the Pearson's cross-correlation coefficient the reaction time τ_i of a catalytic event at a segment i and the time, τ_j, of the immediate subsequent catalytic event taking place at another segment j. Based on the proposed hole migration model in Chapter 8, Monte Carlo simulations are carried out to obtain the Pearson's cross correlation coefficient of temporal sequence and spatial locations of the individual catalytic events for different nanoparticle catalysed reaction systems. Here the example of Pd nanorods catalysing the photo-induced resazurin disproportionation is considered. The simulations monitored the migration of multiple charged holes across different segments on the nanorod. For the nanorod of length $L = 1200$ nm, 12 segments each of length 100 nm were considered. The nanorods in the simulations were treated as 1-D objects.The hopping rate constant for each hole, $u = 0.184$ s^{-1} and the rate of destruction of a hole, $r = 0.03636$ s^{-1} are derived from the experimental study[1] and was used as input parameters for the simulations. The rate constant for the creation of a single hole k and the number of charged holes N that need to be monitored at any instant of time are the undetermined quantities which were set arbitrarily to obtain the cross correlation coefficient. The time sequences of hole formation events from each segment were collected and then the correlation coefficient was studied with respect to the average distance separation (Δx_{ij}) between the segments as well as the average time separation (Δt_{ij}) of the subsequent hole formation events. For the latter, the average time separations were binned together and the corresponding correlation coefficients were then averaged within the bins. This cross correlation coefficient is a quantitative measure of how any catalytic event at one location is correlated with that of a subsequent event at a different location on the same nanorod.

Bibliography

(1) Zou, N. et al. Cooperative Communication within and between Single Nanocatalysts. *Nat. Chem.* **2018**, *10*, 607–614.

Appendix G

Determination of steady-state probabilities

In this section, the calculations to determine the stationary-state probabilities to find a charged hole on the nanorod lattice or in the 'off' state are presented. At $t \to \infty$, the system reaches steady-state and one can re-write the set of forward master equations in Eq. (9.1 − 9.4) in Chapter 9. For odd n,

$$(u_1 + u_2 + r_1)P_n = u_1 P_{n+1} + u_2 P_{n-1} \tag{G.1}$$

whereas, for even n one gets

$$(u_1 + u_2 + r_2)P_n = u_2 P_{n+1} + u_1 P_{n-1}. \tag{G.2}$$

At the origin segment $n = 0$ where the charged hole is produced,

$$(u_1 + u_2 + r_2)P_0 = kP_{off} + u_2 P_1 + u_1 P_{-1} \tag{G.3}$$

and for the state $n = off$, the steady-state equation is

$$kP_{off} = r_1 \Sigma_{i=odd} P_i + r_2 \Sigma_{i=even} P_i. \tag{G.4}$$

The solution of recurrences in Eqs. (G.1) and (G.2) may be obtained by introducing

general solutions of the form

$$P_n = \begin{cases} B_1 y^n & ; n = 1, 3, 5, ... \\ B_3 y^{-n} & ; n = -1, -3, -5, ... \\ B_2 y^n & ; n = 2, 4, 6, ... \\ B_4 y^{-n} & ; n = -2, -4, -6, ... \end{cases}. \tag{G.5}$$

After substituting Eq. (G.2) in Eq. (G.1) and using Eq. (G.5) afterwards, one can obtain

$$y = \sqrt{\frac{z - \sqrt{z^2 - 4u_1^2 u_2^2}}{2u_1 u_2}} \tag{G.6}$$

where $z = 2u_1u_2 + r_1r_2 + (u_1 + u_2)(r_1 + r_2)$. Now, it is required to determine the four constants in Eq. (G.5) to arrive at the steady-state probabilities. At states $n = \pm 1$, the recurrences in Eqs. (G.1) and (G.2) result in $B_4 = B_2$ and $B_3 = cB_1$ and after dividing Eq. (G.1) by Eq. (G.2), it can be shown that $B_2 = \phi B_1$. Here, the parameters c and ϕ are defined by

$$c = (u_1 + u_2 y^2)(u_1 y^2 + u_2)^{-1} \tag{G.7a}$$

$$\phi = \sqrt{\frac{c(u_1 + u_2 + r_1)}{(u_1 + u_2 + r_2)}}. \tag{G.7b}$$

The constant B_1 is determined by using Eqs. (G.1 − G.4) and the normalisation condition provided in Eq. (9.5) within Chapter 9.

$$B_1 = \frac{k(u_1 + u_2)(1 - y^2)}{y(M_1 + kM_2)} \tag{G.8a}$$

where,

$$M_1 = r_2(cu_1 + u_2)(1 - y^2) + (u_1 + u_2 + r_2)(r_1(1 + c) + 2r_2\phi y) \tag{G.8b}$$

$$M_2 = (cu_1 + u_2)(1 - y^2) + (u_1 + u_2 + r_1)(1 + c) + 2\phi y(u_1 + u_2 + r_2). \tag{G.8c}$$

Finally, after using all the above relations, one can obtain the steady-state probabilities

to find a charged hole at any lattice segment n or in the 'off' state as

$$P_n = \begin{cases} B_1 y^n & ; n = 1, 3, 5, ... \\ cB_1 y^{-n} & ; n = -1, -3, -5, ... \\ \phi B_1 y^{|n|} & ; n = \pm 2, \pm 4, \pm 6, ... \end{cases} \tag{G.9a}$$

$$P_0 = \frac{kM_3}{M_1 + kM_2} \tag{G.9b}$$

$$P_{off} = \frac{M_1}{M_1 + kM_2}. \tag{G.9c}$$

In the above equation, the constants c, ϕ, B_1, M_1 and M_2 are defined in Eqs. (G.7 − G.8) and additionally, the parameter $M_3 = (cu_1 + u_2)(1 - y^2) + r_1(1 + c) + 2r_2\phi y$.

Monte Carlo simulations to determine the steady-state probabilities, temporal memory and communication distances

The presented theoretical model is developed by monitoring the evolution of a single charged hole on the entire nanrod lattice. Thus, to determine the steady-state probabilities, the communication lifetimes (τ) and distances (λ), the migration of only one hole on a one-dimensional (1D) lattice is simulated. A 2550 nm long nanorod is introduced which is divided into segments of 50 nm each. Depending on the situation being simulated, define the distribution of migration rates, u_i, and death rates, r_i, which are governed by the segment involved, i. It is considered that a new hole is always produced at the middle segment, labeled as state $n = 0$, with rate k. This set-up ensures the theoretical assumption of very long nanorods such that $n \to |\infty|$. At $t = 0$, the hole is not produced yet and is in state $n =$ 'off'. The Gillespie algorithm[1, 2] is used to monitor the production of the hole and its evolution until is disappears from the lattice. After the hole is produced, it performs a 1D random walk on the nanorod and it disappears from any segment according to the migration and death rates associated with the segment. For the determination of the steady-state probabilities, each Monte Carlo step is run for a finite time (t_f) after which the state of the hole is noted. After all the runs are complete, the relative probabilities of finding the hole in each state are evaluated. To determine the communication lifetimes and distances, each run is stopped with the disappearance of the hole from the nanorod after it was produced for the first time. Then, the time spent by

the hole on the lattice and the distance it travelled from the state $n = 0$ are noted. All simulations were performed for 10^5 Monte Carlo steps.

Multiple-particle Monte Carlo simulations for Pearson's cross-correlation coefficient (PCC) analysis

Herein, a 1200 nm long Pd nanorod is simulated according to the experimental observations[3, 4] and divided it into 24 segments of 50 nm each. The migration rates u_i and death rates r_i were defined along the nanorod according to different situations described in Chapter 9. For multiple-particle simulations, it is assumed that there are about 600 active sites present on the nanorod which can produce holes during each redox reaction. Please note that this number was arbitrarily decided. At any time during a Monte Carlo step, any one from the four possible events (analogous to chemical reactions) can happen on each segment. These events are - a new hole being produced, one of the holes from the segment diffusing to either of the adjacent segments (except for the segments at the ends of the nanorod) or one of the holes disappearing from the segment. Gillespie algorithm[1, 2] is used to decide the events that occur from a segment of the nanorod at the successive time steps. Since the single-molecule experiments were monitored until about 8000 product-formations happened on each nanorod,[3] each Monte Carlo step also continued until 8000 holes were produced on the nanorod lattice. Note that the theory assumes that a new hole is produced from each chemical redox reaction. During each run, the time and the segment at which every new hole is produced on the nanorod is noted. 400 Monte Carlo steps were simulated, which physically represents a situation where 400 nanorods are being monitored independently. After the simulations were complete, the procedure described in Appendix F is followed to perform the Pearson's cross-correlation analysis. The final results are presented in Fig. 9.4 in Chapter 9.

Communication lifetime from a simple two-state model

Since the lifetime of the cooperative communications are independent of their creation rates, one can consider a minimal two-state model for a single hole migrating between two states with the same rate $u = u_1 + u_2$: see Fig. G.1. Additionally, the hole disappears

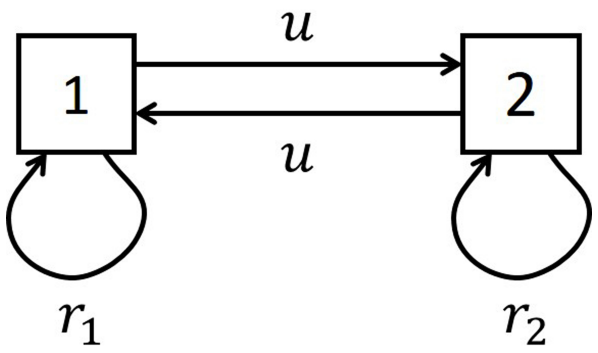


Figure G.1: A simple two-state model for the migration of a single hole until it disappears from the lattice. The migration rate between the state is same, u, while the death rates are different.

from states 1 and 2 with rates r_1 and r_2, respectively. The residence time (τ_n) for the states $n = 1, 2$ can be written as

$$\tau_1 = \frac{1}{u + r_1} \tag{G.10a}$$

$$\tau_2 = \frac{1}{u + r_2}. \tag{G.10b}$$

Since the residence times of any state are proportional to their stationary-state probabilities P_n, one can write

$$P_1 = c\tau_1 \tag{G.11a}$$

$$P_2 = c\tau_2. \tag{G.11b}$$

Further, the normalisation condition must hold for all times using which one can determine the proportionality constant c as

$$c = \frac{(u + r_1)(u + r_2)}{2u + r_1 + r_2}. \tag{G.12}$$

The steady-state probabilities in Eq. (G.11) are now evaluated using the above result to obtain

$$P_1 = \frac{(u + r_2)}{2u + r_1 + r_2} \tag{G.13a}$$

$$P_2 = \frac{(u + r_1)}{2u + r_1 + r_2}. \tag{G.13b}$$

Now, the lifetime (τ) of the hole for this simple two-state model is determined from the flux of the system

$$\tau = \frac{1}{J} = \frac{1}{r_1 P_1 + r_2 P_2} = \frac{2u + r_1 + r_2}{u(r_1 + r_2) + 2r_1 r_2}. \tag{G.14}$$

Additional figures

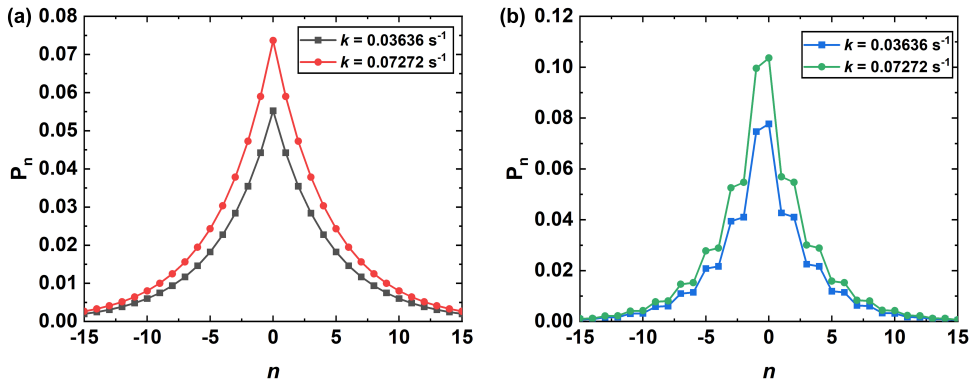


Figure G.2: Steady-state probability distributions of finding a charged hole on the nanorod after it is produced at the segment $n = 0$. The death rate for all segments is kept constant at $r = 0.03636s^{-1}$ whereas the migration rates are (a) $u_1 = u_2 = u = 0.736s^{-1}$ and (b) $u_1 = 4u$ and $u_2 = u/4$. The spread of the distribution governs the region across which the hole is likely to exist before it disappears from the nanorod. This region is controlled by the migration and the death rates of the hole on the lattice until it disappears. On the contrary, the creation rate of the hole affects the amplitude (maximum value) of the distribution and the decay region of the distribution remains almost unaltered. Thus, the communication distance provided in Chapter 9's Eq. (9.9) is independent of the rate k.

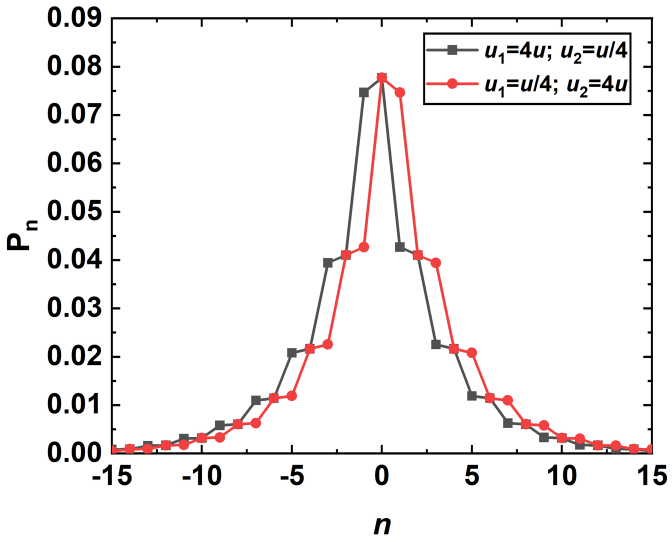


Figure G.3: Steady-state probability distributions of finding a charged hole on the nanorod after it is produced at the segment $n = 0$ for different distribution of migration rates on the nanorod. The rate constants used are $k = r = 0.03636s^{-1}$and $u = 0.736s^{-1}$. Slow migration rates from any segment n cause the hole to stay on that segment longer which increases its probability P_n.

Bibliography

(1) Gillespie, D. T. A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. *J. Comput. Phys.* **1976**, *22*, 403–434.

(2) Gillespie, D. T. Exact Stochastic Simulation of Coupled Chemical Reactions. *J. Phys. Chem.* **1977**, *81*, 2340–2361.

(3) Zou, N. et al. Cooperative Communication within and between Single Nanocatalysts. *Nat. Chem.* **2018**, *10*, 607–614.

(4) Punia, B. et al. Microscopic Mechanisms of Cooperative Communications within Single Nanocatalysts. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119*, e2115135119.

LIST OF PUBLICATIONS

1. Macromolecular Crowding Facilitates ssDNA Capture within Biological Nanopores: Role of Size Variation and Solution Heterogeneity, **B. Punia** and S. Chaudhury*, *J. Phys. Chem. B*, **2024**, 128, 8, 1876–1883

2. How Heterogeneity Affects Cooperative Communications within Single Nanocatalysts, **B. Punia**, S. Chaudhury* and A. B. Kolomeisky*, *J. Phys. Chem. Lett.*, **2023**, 14, 36, 8227–8234

3. Microscopic Mechanism of Macromolecular Crowder-Assisted DNA Capture and Translocation through Biological Nanopores, **B. Punia** and S. Chaudhury*, *J. Phys. Chem. B*, **2023**, 127, 26, 5850–5858

4. Theoretical Tools to Quantify Stochastic Fluctuations in Single-Molecule Catalysis by Enzymes and Nanoparticles, D. Singh, **B. Punia** and S. Chaudhury*, *ACS Omega*, **2022**, 7, 51, 47587–47600 (Invited Review)

5. Theoretical Insights into the Full Description of DNA Target Search by Subdiffusing Proteins, **B. Punia** and S. Chaudhury*, *Phys. Chem. Chem. Phys.*, **2022**, 24, 29074-29083

6. Influence of Nonspecific Interactions between Proteins and in vivo Cytoplasmic Crowders in Facilitated Diffusion of Proteins: Theoretical Insights, **B. Punia** and S. Chaudhury*, *J. Phys. Chem. B*, **2022**, 126, 16, 3037–3047

7. Microscopic Mechanisms of Cooperative Communications within Single Nanocatalysts, **B. Punia**, S. Chaudhury* and A. B. Kolomeisky*, *Proc. Natl. Acad. Sci. USA*, **2022**, 119, e2115135119

8. Understanding the Reaction Dynamics on Heterogeneous Catalysts using a Simple Stochastic Approach, **B. Punia**, S. Chaudhury* and A. B. Kolomeisky*, *J. Phys. Chem. Lett.*, **2021**, 12, 11802-11810

9. Theoretical Study of the Role of Dynamic Bulk Crowders in the Protein Search for Targets on DNA, **B. Punia** and S. Chaudhury*, *J. Stat. Mech.: Theory Exp*, **2021**, 073502

RIGHTS AND PERMISSIONS



CCC Marketplace

This is a License Agreement between Bhawakshi Punia ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the Marketplace Permissions General Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below.

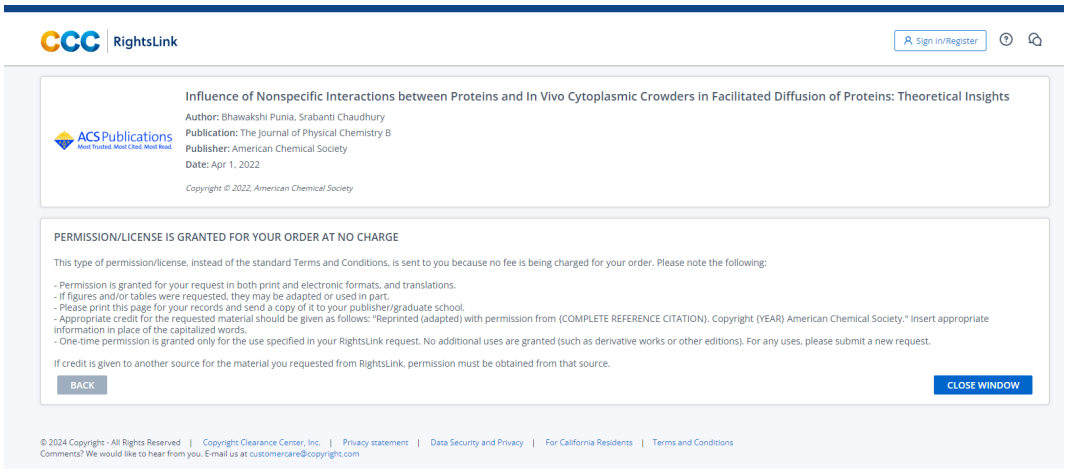
All payments must be made in full to CCC in accordance with the Marketplace Permissions General Terms and Conditions below.

Order Date	24-Jun-2024	Type of Use	Republish in a thesis/dissertation
Order License ID	1497504-1	Publisher	IOP Publishing
ISSN	1742-5468	Portion	Chapter/article

LICENSED CONTENT

Publication Title	Journal of Statistical Mechanics : Theory & Experiment	Rightsholder	IOP Publishing, Ltd
Article Title	Theoretical study of the role of dynamic bulk crowders in the protein search for targets on DNA	Publication Type	e-Journal
		Start Page	073502
		Issue	7
Author / Editor	Institute of Physics Publishing., International School for Advanced Studies (Trieste, Italy)	Volume	2021
		URL	http://iopscience.iop.org/1742-5468/
Date	01/01/2004		
Language	English		
Country	United Kingdom of Great Britain and Northern Ireland		

Chapter 2



CCC RightsLink

Sign in/Register

ACS Publications

Influence of Nonspecific Interactions between Proteins and In Vivo Cytoplasmic Crowders in Facilitated Diffusion of Proteins: Theoretical Insights

Author: Bhawakshi Punia, Srabanti Chaudhury
Publication: The Journal of Physical Chemistry B
Publisher: American Chemical Society
Date: Apr 1, 2022

Copyright © 2022, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION). Copyright {YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

BACK CLOSE WINDOW

© 2024 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Data Security and Privacy | For California Residents | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Chapter 3

IISER PUNE

Bhawakshi Punia <punia.bhawakshi@students.iiserpune.ac.in

RE: Issuing Copyrights
1 message

CONTRACTS-COPYRIGHT (shared) <Contracts-Copyright@rsc.org> Fri, Jun 7, 2024 at 7:27 PM
To: "punia.bhawakshi@students.iiserpune.ac.in" <punia.bhawakshi@students.iiserpune.ac.in>

Dear Bhawakshi Punia

The Royal Society of Chemistry (RSC) hereby grants permission for the use of your paper(s) specified below in the printed and microfilm version of your thesis. You may also make available the PDF version of your paper(s) that the RSC sent to the corresponding author(s) of your paper(s) upon publication of the paper(s) in the following ways: in your thesis via any website that your university may have for the deposition of theses, via your university's Intranet or via your own personal website. We are however unable to grant you permission to include the PDF version of the paper(s) on its own in your institutional repository. The Royal Society of Chemistry is a signatory to the STM Guidelines on Permissions (available on request).

Please note that if the material specified below or any part of it appears with credit or acknowledgement to a third party then you must also secure permission from that third party before reproducing that material.

Please ensure that the thesis includes the correct acknowledgement (see http://rsc.li/permissions for details) and a link is included to the paper on the Royal Society of Chemistry's website.

Please also ensure that your co-authors are aware that you are including the paper in your thesis.

Best wishes

Becky

Becky Roberts
Contracts & Copyright Executive, Sales Operations

Royal Society of Chemistry

www.rsc.org

ROYAL SOCIETY OF CHEMISTRY

From: do-not-reply@rsc.org <do-not-reply@rsc.org>
Sent: Friday, June 7, 2024 1:36 PM
To: CONTRACTS-COPYRIGHT (shared) <Contracts-Copyright@rsc.org>
Subject: Issuing Copyrights

PLEASE DO NOT REPLY DIRECTLY TO THIS EMAIL AS THE INBOX IS NOT MONITORED.
The email address below can be used to reply to the person who submitted this form.

Name: Bhawakshi Punia
Email address (please reply to this address): punia.bhawakshi@students.iiserpune.ac.in
Message: Dear Team,
Kindly issue the copyrights for my paper published in Phys. Chem. Chem. Phys. journal. I want to use the full article in my thesis. Please find the details for the same:
Theoretical insights into the full description of DNA target search by subdiffusing proteins, B. Punia and S. Chaudhury*, Phys. Chem. Chem. Phys., 2022, 24, 29074-29083
(https://pubs.rsc.org/en/content/articlelanding/2022/cp/d2cp04934a)

Thank you
Bhawakshi Punia

Chapter 4

CCC RightsLink

Sign in/Register

ACS Publications
Most Trusted. Most Cited. Most Read.

Microscopic Mechanism of Macromolecular Crowder-Assisted DNA Capture and Translocation through Biological Nanopores

Author: Bhawakshi Punia, Srabanti Chaudhury
Publication: The Journal of Physical Chemistry B
Publisher: American Chemical Society
Date: Jul 1, 2023

Copyright © 2023, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

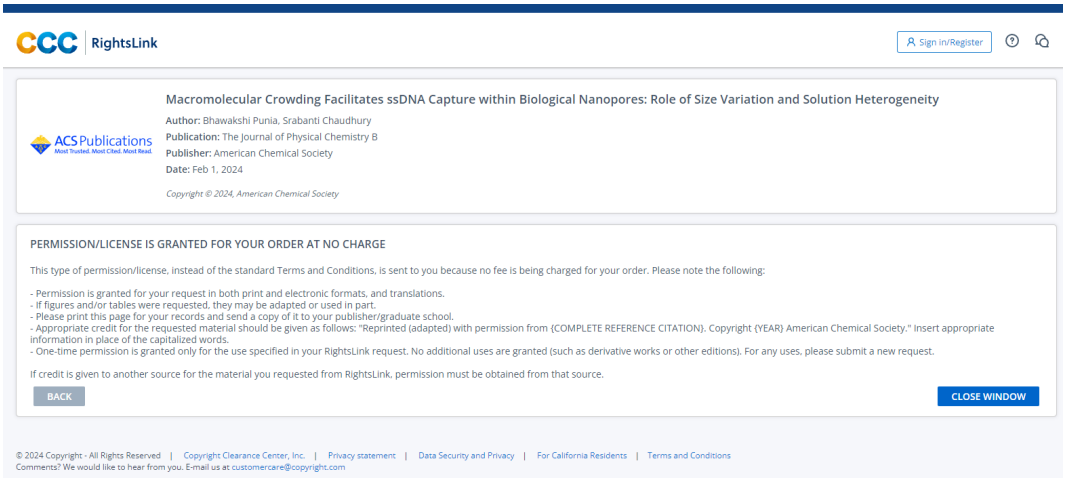
- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

BACK CLOSE WINDOW

© 2024 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Data Security and Privacy | For California Residents | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Chapter 5



CCC RightsLink

Sign in/Register

ACS Publications
Most Trusted. Most Cited. Most Read.

Macromolecular Crowding Facilitates ssDNA Capture within Biological Nanopores: Role of Size Variation and Solution Heterogeneity

Author: Bhawakshi Punia, Srabanti Chaudhury
Publication: The Journal of Physical Chemistry B
Publisher: American Chemical Society
Date: Feb 1, 2024

Copyright © 2024, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

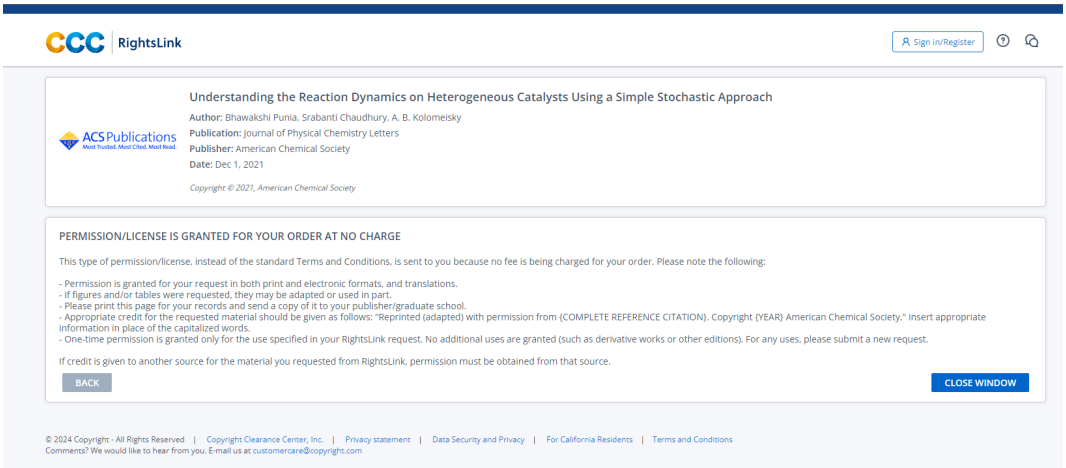
- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

BACK CLOSE WINDOW

© 2024 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Data Security and Privacy | For California Residents | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Chapter 6



CCC RightsLink

Sign in/Register

ACS Publications
Most Trusted. Most Cited. Most Read.

Understanding the Reaction Dynamics on Heterogeneous Catalysts Using a Simple Stochastic Approach

Author: Bhawakshi Punia, Srabanti Chaudhury, A. B. Kolomeisky
Publication: Journal of Physical Chemistry Letters
Publisher: American Chemical Society
Date: Dec 1, 2021

Copyright © 2021, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

BACK CLOSE WINDOW

© 2024 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Data Security and Privacy | For California Residents | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Chapter 7

IISER PUNE

Bhawakshi Punia <punia.bhawakshi@students.iiserpune.ac.in>

Issuing Copyrights

PNAS Permissions <pnaspermissions@nas.edu> Wed, Jun 12, 2024 at 7:11 PM
To: "punia.bhawakshi@students.iiserpune.ac.in" <punia.bhawakshi@students.iiserpune.ac.in>

Dear Bhawakshi,

Thank you for your message. Please note that your PNAS article published under a CC BY-NC-ND license, and as the author, you hold copyright. You may refer to https://creativecommons.org/licenses/by-nc-nd/4.0/ for a summary of terms and conditions.

Best regards,
William Choi for
Diane Sullenberger
PNAS Executive Editor

---------- Original Message ----------
From: Bhawakshi Punia [punia.bhawakshi@students.iiserpune.ac.in]
Sent: 6/7/2024, 8:50 AM
To: pnaspermissions@nas.edu
Subject: Issuing Copyrights

Dear Team,

My name is Bhawakshi Punia and I am a doctoral student at Indian Institute of Science Education and Research Pune, India.

Kindly issue the copyrights for my paper published in PNAS. I want to use the full article in my thesis. Please find the details for the same:

Microscopic mechanisms of cooperative communications within single nanocatalysts, B. Punia, S. Chaudhury* and A. B. Kolomeisky*, Proc. Natl. Acad. Sci. USA, *2022*, 119, e2115135119

(https://www.pnas.org/doi/full/10.1073/pnas.2115135119)

Regards,
Bhawakshi Punia

Chapter 8

CCC RightsLink

Sign in/Register

ACS Publications
Most Trusted. Most Cited. Most Read.

How Heterogeneity Affects Cooperative Communications within Single Nanocatalysts

Author: Bhawakshi Punia, Srabanti Chaudhury, Anatoly Kolomeisky
Publication: Journal of Physical Chemistry Letters
Publisher: American Chemical Society
Date: Sep 1, 2023

Copyright © 2023, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

BACK CLOSE WINDOW

© 2024 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Data Security and Privacy | For California Residents | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Chapter 9