APPROVAL

Name: Laleh Samii
Degree: Doctor of Philosophy
Title of Thesis: Studies of Synthetic Molecular Motors
Examinining Committee: Dr. J. Steven Dodge, Associate Professor (Chair)

_________________________________
Dr. Nancy R. Forde, Senior Supervisor, Associate Professor

_________________________________
Dr. Martin J. Zuckermann, Supervisor, Adjunct Professor

_________________________________
Dr. Eldon Emberly, Supervisor, Associate Professor

_________________________________
Dr. Dipankar Sen, Supervisor, Professor
Department of Molecular Biology and Biochemistry

_________________________________
Dr. Peter J. Unrau, Internal Examiner, Associate Professor
Department of Molecular Biology and Biochemistry

_________________________________
Dr. Zev Bryant, External Examiner, Assistant Professor
Schools of Medicine and Engineering, Stanford University

Date Approved: April 18, 2013
Partial Copyright Licence

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the “Institutional Repository” link of the SFU Library website (www.lib.sfu.ca) at http://summit/sfu.ca and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author’s written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

While licensing SFU to permit the above uses, the author retains copyright in the thesis, project or extended essays, including the right to change the work for subsequent purposes, including editing and publishing the work in whole or in part, and licensing other parties, as the author may desire.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, British Columbia, Canada

revised Fall 2011
Abstract

To better understand the underlying principles by which biological motors operate, recent work has focused both on understanding their operational principles, and on designing new molecular motors *ab initio*. Here, by studying and designing motors which use Brownian motion and track asymmetry to bias the direction of motion, I gained insight into the underlying principles by which such motors operate. "Molecular spiders" [JACS. 128, 12693 (2006), Nature 465, 206 (2010)] are one example of synthetic biomolecular walkers able to generate biased motion by coupling the chemical asymmetry arising from substrate binding and cleavage to bias their mechanical stepping. These DNA-based motors diffuse to their substrate track where productive binding between a molecular spider’s DNAzyme leg and a ssDNA substrate facilitates cleavage of the substrate. Once cleaved, the decreased binding affinity between the DNAzyme and resulting product allows the motor to diffuse along the track and form new interactions with uncleaved substrate molecules. To investigate the origin of biased motion of molecular spiders, I have performed Monte Carlo simulations. Using my simulations, I also investigated their performance as molecular motors, and determined how to optimize their motor properties by modifying tunable experimental parameters in spider design. These studies assisted us in the design and construction of a novel protein-based synthetic motor, the "Lawnmower", which uses a burnt–bridges type of mechanism, the same as spiders, to autonomously and diffusively move forward. The lawnmower has trypsin proteases as blades, linked to a quantum dot hub, that interact with a one-dimensional peptide substrate track via binding to and cleavage of the substrates. Experimentally, it is confirmed with kinetic assays that our lawnmower is an active motor and that there are an average number of 8 blades on each motor. I also outlined the synthesis and characterization of a highly modified DNA-peptide construct, which acts as the track for the lawnmower. For this, I employed PCR to generate a densely labeled DNA and click chemistry for peptide conjugation to the functionalized DNA. As an additional motors-related project, I present the synthesis of a long one-dimensional DNA track with periodically repeating
elements that provide specific binding sites for the "Tumbleweed" molecular motor [HFSP J. 3, 204 (2009)].
To my grandpa (Javad) and my uncle (GholamNabi) for teaching me a different way of looking at life.
Acknowledgments

I feel most fortunate to have had the opportunity of studying in the Department of Physics at Simon Fraser University with such friendly environment and amazing people. A great number of people have helped and inspired me during my PhD. I would like to take this opportunity to thank them.

First and foremost, I would like to thank Dr. Nancy Forde, my supervisor, who has supported me throughout my PhD work with her guidance, encouragement and (most importantly) patience. I greatly appreciate her thoughtful advices for research beside her genuinely caring for her students. I would also like to thank Dr. Martin Zuckermann, my co-supervisor, who was abundantly helpful and offered invaluable support and guidance. Thank you Martin for sharing with me your extensive knowledge of science, culture, language, history and of course the history of my own country; Iran.

I would like to thank Dr. Dipankar Sen, Dr. Eldon Emberly and Dr. Michael Plischke, for their invaluable ideas in the past few years. I would also like to thank the supportive staff of Physics Department, past and present.

I would also like to thank the great members of Forde lab, past and present, for helpful discussions, reading my thesis and for being such amazing people. I would like to especially acknowledge Suzana and Andrew for "teaching a physicist" many basics and techniques of molecular biology and biochemistry. I would like to thank HFSP-funded molecular motors design team for all the helpful discussions and feedback throughout doing these projects.

I am very lucky to have met so many great people at SFU. Special thanks to my dear friends, Fereydoun, Sara, Hamed, Azadeh, Alireza, Dorna, Kamyar, Mohammad for always being there to support me when I needed it. I would also like to thank my Vancouver family (Mohsen, Shiva, Ardalan and Arghavan) for all their love and encouragement.

Last, but not least, I thank my family for their never-ending support and encouragement. To maman, baba, Leila, Hessam, Hossein: I really, truly cannot thank you enough for everything all of you have done for me throughout the years. Despite being on the other side of the planet the entire
time I’ve been in Vancouver, you were always there for me. Thank you.
# Contents

<table>
<thead>
<tr>
<th>Approval</th>
<th>ii</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partial Copyright Licence</td>
<td>iii</td>
</tr>
<tr>
<td>Abstract</td>
<td>iv</td>
</tr>
<tr>
<td>Dedication</td>
<td>vi</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>vii</td>
</tr>
<tr>
<td>Contents</td>
<td>ix</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xiii</td>
</tr>
</tbody>
</table>

## 1 Introduction

1.1 Physical Properties of Molecular Motors ........................................ 2
1.2 Brownian Ratchet ................................................................. 6
1.3 Synthetic DNA-Based Molecular Motors ........................................... 10
1.4 Observation Techniques .......................................................... 13
   1.4.1 DNA Curtains and Racks ..................................................... 13
   1.4.2 DNA Tightropes ............................................................... 15
1.5 Overview of Synthetic Molecular Motors in Our Group ......................... 15

## 2 Molecular Spiders

2.1 Introduction ................................................................................. 17
2.2 Studying the Biased Motion of HOH and IW Bipedal Spiders .................... 23
   2.2.1 Modeling ................................................................................. 23
   2.2.2 Computational Method .......................................................... 26
   2.2.3 Results and Discussion ....................................................... 28
   2.2.4 Summary and Concluding Remarks .......................................... 44
2.3 Effect of Tuneable Experimental Parameters ...................................... 46
   2.3.1 Methods ............................................................................... 47
## CONTENTS

2.3.2 Results ................................................................. 51
2.4 Conclusions ............................................................ 69

3 Lawnmower: an Autonomous Molecular Motor 71
  3.1 Introduction .............................................................. 71
  3.2 Lawnmower Design ...................................................... 74
    3.2.1 Lawnmower Hub .................................................. 74
    3.2.2 Lawnmower Linker ............................................... 76
    3.2.3 Lawnmower Blades .............................................. 77
  3.3 Lawnmower Construction ............................................. 78
  3.4 Lawnmower Track Design ........................................... 79
  3.5 Construction of Lawnmower Track ................................... 85
    3.5.1 Construction of Alkyne-DNA ................................... 85
    3.5.2 Construction of Peptide-DNA ................................ 93
    3.5.3 Experiments to Improve Peptide-DNA Solubility ......... 97
    3.5.4 Characterization of Peptide-DNA Using Atomic Force Microscopy 103
  3.6 Conclusions and Future Directions ................................ 108

4 Design and Construction of a DNA Track for the Tumbleweed Motor 111
  4.1 Introduction ............................................................. 111
  4.2 Design of TW DNA Track ............................................. 114
    4.2.1 Design Parameters ............................................. 114
  4.3 DNA Track Construction ............................................. 115
    4.3.1 Biotin-DNA TW Track Construction Using Lambda DNA ... 117
    4.3.2 Biotin-Digoxigenin DNA TW Track Construction Using Bacterial Artificial Chromosomes .......... 122
  4.4 Construction of the Fiducial Marker ................................ 128
  4.5 Conclusions ............................................................ 132

5 Conclusions and Future Directions 134

Appendix A Rate Constants for Molecular Spiders 139
  A.1 Diffusion Rate Constant ........................................... 139
  A.2 Calculation of the First-order $k_{on}$ Rate .................... 140
  A.3 Calculation of $k_c$ ............................................... 141

Appendix B Gillespie Algorithm 143

Appendix C Drift of Diffusing Spiders on a $P-S$ Track 145

Appendix D Construction of Lawnmower 148
  D.1 Production of Free Thiols ........................................ 148
  D.2 Biotinylation with Biotin-PEG2-Maleimide ...................... 151
  D.3 Maleimide Functionalization of QDs ......................... 151
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>D.4 Conjugation of Trypsin to Maleimide QDs</td>
<td>152</td>
</tr>
<tr>
<td>D.5 Purification of Trypsin Conjugated QDs</td>
<td>153</td>
</tr>
<tr>
<td>D.6 Fluorescence Assays</td>
<td>153</td>
</tr>
<tr>
<td>D.7 Method of Lawnmower Quantification</td>
<td>153</td>
</tr>
<tr>
<td>Appendix E Construction of Azide-PEG4-Peptide</td>
<td>155</td>
</tr>
<tr>
<td>Appendix F PCR Thermocycles to Generate Alkyne-DNA</td>
<td>157</td>
</tr>
<tr>
<td>F.1 Products with Pwo Polymerase</td>
<td>157</td>
</tr>
<tr>
<td>F.2 1 kbp and 3 kbp Products with KOD XL Polymerase</td>
<td>158</td>
</tr>
<tr>
<td>Bibliography</td>
<td>159</td>
</tr>
</tbody>
</table>
### List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
</tr>
</thead>
</table>
| 2.1   | Transition rate constants for interactions of a leg with a substrate or a product.  
|       | a Estimated from [86, 107]. b In some of the simulations, $k_{on}$ is changed to 2.4 and 1.6 s$^{-1}$. c In some of the simulations, $k_{off,S}$ is changed to 0.0035 s$^{-1}$. d In some of the simulations, $k_{c}$ is changed to 0.0055 s$^{-1}$ and $k_{-c}$ is kept at $k_{c} / 100$. e Reference [86]. |
| 2.2   | Average number of legs bound to the track for bound spiders with $n=2$, 3 and 4 and $S=3$ (average calculated over the first $t = 800$ min). |
| 3.1   | Effect of experimentally adjustable parameters on motor properties of molecular spiders. a Studied only for a special case of quadrupedal spiders with $S=3$. b Studied only for a special case of HOH bipedal spiders. Rate constants refer to the kinetic scheme in Figure 2.2. |
| 3.2   | Sequence of the primers used in PCR amplification to produce 3 kbp DNA products. |
| 3.3   | Sequences of the primers used in PCR amplification to produce 352 bp (PspOMI-400) and 952 bp (PspOMI-1000) DNA products. |
| 3.4   | Sequences of the new EagI site carrying primer (EagI primer-N) along with PspOMI site carrying primers to produce 371 bp (PspOMI-400N) and 968 bp (PspOMI-1000N) DNA PCR products. |
| D.1   | The calculated moles of trypsin and QD in each fraction. The average number of trypsins per QD was calculated from the ratio of moles of trypsin : moles of QD in each fraction. (Figure courtesy of Dr. Suzana Kovacic.) |
| F.1   | Thermocycle for generation of 400 bp and 1 kbp products using Pwo polymerase. EagI primers are listed in Tables 3.2 and 3.4, PspOMI primers are in Tables 3.3 and 3.4 (400 bp). |
| F.2   | Thermocycle for generation of 1 kbp alkyne-DNA using KOD XL polymerase and primers in Table 3.4. |
| F.3   | Thermocycle for generation of 3 kbp alkyne-DNA using KOD XL polymerase and primers in Table 3.2. |
List of Figures

1.1 Examples of linear motor proteins from the three major superfamilies: (a) Myosin (b) Cytoplasmic dynein. (c) Kinesin. Representations based on atomic resolution structures [27]. Blue: motor domains; purple: cargo attachment sites; green: associated motor units. Adapted with permission from [27]. ................. 3

1.2 (a) Illustration of the cycle of changes in myosin shape during cross-bridge cycling [58] (states 1, 2, 3, and 4) (see text for details). (b) Ratchet model for myosin. In the attached state, the myosin head is located in a local minimum of the potential of interaction between myosin and an actin filament (potential $V_A(x)$)(1). After an ATP is bound by the myosin head, the head detaches from the actin filament and undergoes free Brownian motion (2 on $V_D(x)$). After ATP hydrolysis (3), and because of the asymmetry of the potential once the myosin head attaches to the actin filament, the particle is more likely to be found in a region of negative slope on the potential $V_A(x)$. The particle then either returns to its starting location on $V_A(x)$ or moves to the next local minimum of the potential (from 4 to 1) (Figure is adapted from [59]; used with permission). ......................... 7

1.3 An example of a DNA nanowalker that executes stepwise movement along its linear track. Shown here are the initial binding of one foot to the track (via attachment strand A1), forward binding of the second foot (via attachment strand A2), and detachment of the trailing foot (via detachment strand D1). See text for details. Figure is taken with permission from [77]. ......................... 12

1.4 Techniques for extending DNA-based motor tracks. When flow is applied, DNA molecules are aligned with respect to one another in (a) DNA curtains and (b) can be tethered also at their ends in DNA racks. (Adapted with permission from [90] and [91], respectively). (c) A cartoon of DNA tightropes formed between micron-sized beads. Adapted with permission from [99]. ......................... 14

2.1 (a) Schematic of a molecular spider with a streptavidin hub and 8-17 DNA enzyme legs. (b) Digestion of a ssDNA substrate, $S$, by a deoxyribozyme leg and production of $P_1$ and $P_2$ products (adapted with permission from [87]). ......................... 18

2.2 Kinetic model and possible pathways of leg-track interactions. The values of all rate constants are given in Table 2.1. ......................... 24

2.3 Stepping mechanisms of IW and HOH spiders. ......................... 26
2.4 Sample trajectories of an IW spider and of a HOH spider. 29
2.5 Evolution of position probability distributions $P(x_{cm})$ on (a, d) $P$ tracks, (b, e) $S$ tracks, (c, f) $P$-$S$ tracks for HOH (left column) and IW (right column) bipedal spiders. In all cases, the initial number of spider trajectories is $10^6$. $P(x_{cm})$ at each subsequent time is calculated only using trajectories that still have a spider bound to the track at that time. For clarity, HOH spider distributions are smoothed by combining probabilities of successive site occupation. This is necessary because the HOH spider spends significantly more time in states where both feet are bound to the track (half-integral $x_{cm}$) compared with states in which only one foot is bound (integral $x_{cm}$). 30
2.6 Graph of $\langle \Delta x_{cm} \rangle$ versus time for IW and HOH spiders on $P$-$S$ tracks. $\langle \Delta x_{cm} \rangle$ is the ensemble-averaged center-of-mass displacement of spiders with respect to their initial position at the middle of the track. It is clear that a population of HOH spiders exhibits much greater speed than IW spiders. 31
2.7 The randomness parameter, $r$, is greater for IW than HOH spiders and increases with time for both types of spiders. 32
2.8 Percentage of trajectories with bound spiders as a function of time, for HOH spiders on $S$ tracks, on $P$-$S$ tracks, on $P$ tracks and on uncleavable $S$ tracks, and for IW spiders on $P$-$S$ tracks. Spiders exhibit the longest dwell times on uncleavable $S$ tracks and detach most rapidly from $P$ tracks. 33
2.9 (a) Graph of $\langle \Delta x_{cm} \rangle$ versus time for different classes of HOH spiders and the predictions of the diffusive model described in Appendix C. Spiders capable of cleavage and detachment show the greatest average displacement per time. (b) A plot of $\log(\Delta x_{cm})$ versus $\log(t)$ shows that the bias of spiders with neither cleavage nor detachment is diffusive, with a slope of this plot of 0.5. The other two classes of spiders show enhanced biased motion is seen. (c) Graph of variance versus time for different classes of HOH spiders show similar behaviour for HOH spiders with cleavage and with/without detachment. 34
2.10 Graphs of randomness for (a) HOH and (b) IW spiders. For both types of spider with cleavage, removing detachment results in increased values of $r$, while $r$ reduces substantially when cleavage is also disallowed. In all cases $r$ increases with time. 35
2.11 Ratio of spiders on $S$ to $P$ side of the track for HOH spiders with neither cleavage nor detachment increases rapidly, then plateaus. 36
2.12 Probability distributions of the number of cleavages performed by each spider prior to its detachment from a track. HOH spiders exhibit greater processivity than IW spiders, and their processivity is significantly enhanced on $S$ tracks. Lines through the points are guides to the eye. 38
2.13 Probability distributions of separation between sequential cleavage events for IW and HOH spiders on $P$-$S$ tracks, showing tighter mechanochemical coupling for IW spiders. Lines through the points are guides to the eyes. 40
2.14 Graphs of velocity versus force for IW and HOH spiders on $P$-$S$ tracks with (a) $k_{on} = 20$ s$^{-1}$ and (b) $k_{on} = 2.4$ s$^{-1}$. Decreasing the binding rate increases the stall forces, which are very similar for HOH and IW spiders. Lines through the points are guides to the eye. 41
2.15 Graph of $<\Delta x_{cm}>$ versus time for HOH spiders on P-S tracks with $k_{on}=20\text{ s}^{-1}$ and $k_{on}=2.4\text{ s}^{-1}$. Decreasing the binding rate increases the speed of HOH spiders. $<\Delta x_{cm}>$ is the ensemble-averaged center-of-mass displacement of spiders with respect to their initial position at the middle of the track. .......................... 42

2.16 Experimental dissociation curves of spiders from either an uncleavable substrate matrix or a product matrix. Spiders were named with X.NA, where X denotes the number of legs and N denotes the number of adenosines in the foot region. Figure is taken with permission from [86]. ..................................................... 47

2.17 Initial configurations of bipedal, tripedal and quadrupedal spiders on a 1D product-substrate track of $10^5$ sites. Open circles indicate product sites, while substrate sites are denoted by closed circles. All three spiders shown here have a maximum span of $S=3$. .......................................................... 48

2.18 (a) Fraction of spider trajectories with a spider bound to the track as a function of time for spiders with $n=2$, 3 and 4 legs and maximum span of $S=3$, showing that spiders with greater leg number stay longer on the track. Values are calculated at each 100 s time interval. (b) Probability distributions of the number of cleavages performed by each type of spider prior to its detachment from a track for spiders with $n=2$ and $n=3$ and prior to a preset time ($t=10^7$ seconds) for spiders with $n=4$. The results indicate that processivity increases with leg number. The inset shows an expanded view of the number of cleavages for bipedal spiders. Lines through the points are guides to the eyes. .......................................................... 52

2.19 (a) Dependence of biased motion on number of legs shows that the mean velocity of spiders decreases with increasing number of legs. $<x_{cm}>$ is the ensemble-averaged center-of-mass displacement of spiders with respect to their initial position at the centre of the track. $<x_{cm}>$ is calculated every 100 s by averaging over only those spiders remaining bound to the track at that time. The inset shows an expanded view of the first 80 minutes. Error bars represent the standard deviation of spider positions, calculated every 50 min. (b) Average number of legs bound to the track for spiders with $n=2$, 3 and 4 and $S=3$ (shown over the first $t=200$ min) indicates that this value remains constant at a value very close to the number of spider legs. .......................... 53

2.20 Average minimum (blue), detachment (orange) and maximum (white) center-of-mass positions of each type of spider, determined from all sites sampled prior to detachment ($n=2$, 3) or prior to a preset long time limit ($t=10^7$ seconds; $n=4$). Error bars represent the standard deviations of positions sampled by each population of spider. Zero on the $<x_{cm}>$ axis represents the initial position of the spiders at the middle of the track, and negative and positive values on the axis represent the initial P and S sides of the P-S track, respectively. The results indicate that run length increases with the number of legs, that spiders travel over a much wider expanse of the track than simply from their initial position to their detachment positions, and that there is significant dispersion in dynamics among different spiders of the population. .................................................. 56
2.21 (a) Fraction of spider trajectories with a spider bound to the track as a function of time showing that the binding time of spiders to the track increases with span. Values were calculated at each 100 s time interval. (b) Probability distributions of the number of cleavages performed by $t = 5 \times 10^6$ seconds by quadrupedal spiders with $S = 3, 4$ and 5. Spiders with larger span cleave more substrates in this time period. Lines through the points are guides to the eyes. 

2.22 Dependence of biased motion on maximum span shows greater mean velocity for quadrupedal spiders with larger maximum span. $<x_{cm}>$ is the ensemble-averaged center-of-mass displacement of spiders with respect to their initial position at the middle of the track. $<x_{cm}>$ is calculated every 100 s by averaging over only those spiders remaining bound to the track at that time. Error bars represent the standard deviation of spider positions, calculated every 50 min. 

2.23 Dependence of biased motion of tripedal and bipedal spiders on maximum span shows greater binding time (a and b), processivity (c and d) and mean velocity (e and f) for spiders with larger maximum span. $<x_{cm}>$ is the ensemble-averaged center-of-mass displacement of spiders with respect to their initial position at the middle of the track. $<x_{cm}>$ is calculated by averaging over only those spiders remaining bound to the track at that time. 

2.24 Dependence of biased motion on the dissociation rate from substrates shows a greater mean velocity of spiders with a greater ratio of $k_c = 0.055 \text{ s}^{-1}$ to $k_{off,S}$. These results are for quadrupedal spiders with $S = 3$. $<x_{cm}>$ is the ensemble-averaged center-of-mass displacement of spiders with respect to their initial position at the middle of the track. $<x_{cm}>$ is calculated every 100 s by averaging over only those spiders remaining bound to the track at that time. Error bars represent the standard deviation of spider positions, calculated every 50 min. 

2.25 The mechanochemical coupling $\xi(t)$ versus time for spiders with $n = 3$ and 4 legs and $S = 3$ at (a) zero force and $F = 0.01 \text{ pN}$ (b) $F = 0.03 \text{ pN}$ exhibits time and force dependence. Mechanochemical coupling of quadrupedal spiders is greater than tripedal spiders and the coupling ratio decreases faster with time at higher applied forces. (c) Long-time behaviour of mechanochemical coupling at zero force shows a constant coupling. Coupling is calculated over $\Delta t = 100 \text{ s}$ time intervals. 

2.26 Graphs of efficiency versus force for spiders with $n = 2, 3$ and 4 legs and $S = 3$ for the first and second (inset) $\Delta t = 100 \text{ s}$ time intervals. Quadrupedal spiders are the most efficient of these molecular spiders and reach stall at greatest applied force. Lines through the points are guides to the eyes. 

2.27 Graph of maximum efficiency versus time for spiders with $n = 2, 3, 4$ and $S = 3$. For each type of spider, maximal efficiency occurs in the first time interval and decreases to zero at longer times. Quadrupedal spiders are the most efficient. Lines through the points are guides to the eyes. 

2.28 Graph of efficiency versus force shows the increased thermodynamic efficiency of spiders with $n = 4$ legs and $S = 3$ with $k_{off,P} = 1 \text{ s}^{-1}$ for the first $\Delta t = 100 \text{ s}$ time interval. Lines through the points are guides to the eyes.
2.29 Improvement of biased motion of quadrupedal spiders by increasing maximum span (to $S = 5$) and decreasing $k_{off}\,S$ (from $0.035\,s^{-1}$ to $0.0035\,s^{-1}$) to compensate for the decrease in speed when $n$ increase the number of spider legs from $n = 2$ to $n = 4$. $<x_{cm}>$ is the ensemble-averaged center-of-mass displacement of spiders with respect to their initial position at the middle of the track. $<x_{cm}>$ is calculated every 100 s by averaging over only those spiders remaining bound to the track at that time. 67

3.1 Schematic of lawnmower design and its lawn. Blades of lawnmower interact with substrate lawn and cut it at specific sites. The marker is designed as an indicator to monitor the activity of the motor. 72

3.2 Schematic illustration of the increase in the number of potentially interacting lawnmower blades (triangles) with substrates when going from a streptavidin hub (bottom, right), to a QD hub (bottom, left) and to a micron-size-bead hub (top). Figures are drawn to scale (a 5 nm streptavidin, a 20 nm QD and a 1.27 $\mu$m bead) assuming proteases are densely linked to the hub and substrates are located every 5 nm on a track (the minimum inter-substrate distance required for adjacent binding of protease blades (average diameter of protein $\sim 5$ nm) [151]). 75

3.3 Selected linkers of (a) 1 nm Sulfo-SMCC: Sulfosuccinimidyl-4-$\left(\text{N}-\text{maleimidomethyl}\right)$ cyclohexane-1-carboxylate (Pierce) and (b) 10 nm (Maleimide-dPEG24-NHS ester (Quanta Biodesign) to be used in the construction of lawnmower. 77

3.4 Monitoring the fluorescence signal of both QD (blue points) and from a trypsin kinetic assay (red bars) for the collected fractions from (a) reaction, trypsin-conjugated QD and (b) control, unconjugated trypsin and QD. QD and trypsin-conjugated QD elute earlier from this size-exclusion column while unconjugated trypsins elute in later fractions. (Figure courtesy of Dr. Suzana Kovacic.) 80

3.5 (a) Structure of the peptide substrate with fluorophore and quencher. The sequence is MCA-Ala-Pro-Ala-Phe-Phe-Arg-Leu-Lys(Dnp)-NH$_2$. (b) Schematic of azide-peptide and alkyne-DNA before click reaction. Azide-peptide is formed by the reaction of NHS-PEG4-azide with the amine group of the lysine side chain, and alkynes are incorporated into DNA using PCR. (c) Click chemistry between alkyne-DNA and azide-peptide results in the formation of the peptide-DNA lawnmower track. MCA and Dnp are the fluorophore and quencher on the peptide. 82

3.6 The copper-catalyzed cycloaddition of an azide and an alkyne is an example of click chemistry. 84

3.7 Structure of C8-Alkyne-dUTP used in the PCR to introduce alkyne groups to DNA. 86

3.8 EtBr-stained 1% agarose gel of PCR products of pUC19 with primers shown in Table 3.2, Taq polymerase and varying ratios of Alkyne-dUTP and dTTP as indicated. Products are seen only when less modified nucleotide is used. 87
3.9 EtBr-stained 1% agarose gel of (a) 400 bp and (b) 1 kbp PCR products of control DNA (lanes 2) and alkyne-DNA (lanes 3) by Pwo polymerase. Comparing the intensities of each control with its alkyne-DNA, it shows the efficient incorporation of alkyne in shorter 400 bp PCR products but less efficient incorporation in 1 kbp lengths. Lanes 1: DNA ladder with numbers on the left indicating selected fragment lengths in base pairs. (c) Absorption spectra of control and alkyne DNA (400 bp) confirms the formation of alkyne-DNA due to the altered shape of the absorption spectrum [171].

3.10 (a) EtBr-stained 1% agarose gel of ligation products from EagI-digested 1 kbp alkyne-DNA and PspOMI-digested control 1 kbp DNA (lane 2) shows formation of 2 kbp ligation products. (b) Lane 2: gel-purified 2 kbp band from agarose gel in (a). EagI (lane 3) and PspOMI (lane 4) digestion of the 2 kbp ligation products shows the dominant homodimerization of control DNA in the ligation reaction. Lanes 1: DNA ladder with numbers on the left indicating selected fragment lengths in base pairs.

3.11 EtBr-stained 1% agarose gel of ligation products of EagI-digested alkyne-DNA and dephosphorylated PspOMI-digested control DNA (lane 1), post EagI (lane 2) and post PspOMI (lane 3) digestion of ligation products. Digestion of ligation products with either enzymes does not increase the intensity of the 1 kbp band significantly.

3.12 The PCR products of control (lane 2) and alkyne-DNA reactions (lane 3): (a) 400 bp with Pwo polymerase and (b) 1 kbp with KOD XL polymerase, run in 1% agarose gels. c) PCR products for 3 kbp control (lanes 2 and 4) and alkyne-DNA (lanes 3 and 5) show the successful production of alkyne DNA when KOD XL polymerase is used (lane 3) but not when Pwo is used (lanes 4 and 5). Lane 1 is 1 kbp DNA ladder (Fermentas) in all pictures.

3.13 Alkyne modification does not interfere with digestion and ligation reactions with newly designed primers. (a) Lanes 4 and 5 show the successful self-ligation of 371 bp alkyne DNA after an EagI digest and a PspOMI digest, respectively, which each generate 738bp fragments. Self- ligations of EagI- and PspOMI-digested control DNA (no alkyne modification) are shown in lanes 2 and 3, respectively. (b) Lanes 4 and 5 show the successful self-ligation of 968 bp alkyne DNA after an EagI digest and a PspOMI digest, respectively, which each generate 1936 bp fragments. Self- ligations of PspOMI- and EagI-digested control DNA (no alkyne modification) are shown in lanes 2 and 3, respectively. Lanes 1 contain a 1 kbp ladder, with numbers on the left indicating selected fragment lengths in base pairs.

3.14 (a) Absorption spectrum of 6.5 µM peptide in 10 mM Tris buffer (pH=8.5) shows absorption peaks of MCA (fluorophore) and Dnp (quencher) in the peptide at about 345 and 420 nm, respectively. (b) Presence of the same absorption peaks along with that for DNA at 260 nm in the peptide-DNA reaction spectrum confirms the conjugation of peptides to alkyne-DNA. Control for which no conjugation was expected shows only the alkyne-DNA absorption spectrum. Inset: enlargement of the region around the peptide features.
3.15 An EtBr stained 1% agarose gel of highly labelled 1 kbp alkyne DNA with azide-peptide. Lane 1: 1 kbp DNA ladder, Lane 2: control and lane 3: reaction. Control DNA ran normally in the gel while the DNA band corresponding to highly labelled peptide-DNA did not appear to migrate in the gel but generated a dark band in the well (lane 3).

3.16 (a) Absorption spectra of peptide-DNA in 100% DMSO show the precipitation of peptide-DNA over time. (b) Absorption spectra of unreacted azide-PEG4-peptide remaining in the ethanol supernatant following precipitation, in control and reaction samples. (Courtesy of Dr. Suzana Kovacic.)

3.17 Absorption of peptide-DNA in 30% DMSO decreases as a function of time, indicating the peptide-DNA is falling out of solution. (Figure courtesy of Dr. Suzana Kovacic.)

3.18 Increase in the MCA fluorescence signal for the peptide-DNA sample shows that peptide substrate in the highly labeled 1 kbp peptide-DNA can be digested by trypsin in 30% DMSO. (Figure courtesy of Dr. Suzana Kovacic)

3.19 (a) Absorption spectrum of peptide-DNA ligation product compared to its control confirms the formation of peptide-DNA. (b) An EtBr-stained 1% agarose gel of the click reaction products from the control (lane 2) and peptide-DNA reactions (lane 3). The 1, 4 and 5 kbp DNA bands corresponding to unligated control DNA, DNA handle and ligation product, respectively are observed in the control (lane 2). In lane 3 the 1 and 5 kbp bands including alkyne-DNA disappeared after the click reaction while the unlabelled 4 kbp band remains visible at the expected position. Again, a dark band appears in the well of the peptide-DNA reaction (lane 3). Lane 1 is 1 kbp DNA ladder.

3.20 Digestion of the peptide-DNA with long handle by an excess of trypsin (in 10mM Tris pH 8.5), along with 100, 250 and 400 pmol peptide-azide controls. The fluorescence intensity of the long peptide DNA digested with trypsin approached that of 100 pmol azide-peptide.

3.21 AFM images of the 1 kbp DNA controls: (a) plain DNA, and (b) alkyne-DNA. AFM image analysis estimated average contour lengths of 294±18 nm and 288±26 nm and persistence lengths of 50± 5 nm and 53±8 nm for the samples in (a) and (b), respectively. These results are calculated based on analysis for N=118 and N= 92 control DNA and alkyne-DNA, respectively. (Figures are courtesy of Dr. Guillaume Lamour.)

3.22 AFM images of the 1 kbp low-labeled peptide-DNA in 50%ethanol : 50% chloroform and the height analysis of the observed structures. There are fiber-like structures on the surface with longer lengths than the expected ∼300 nm peptide-DNA length. (Figure courtesy of Dr. Guillaume Lamour.)

3.23 A schematic of the lawnmower on a stretched DNA in a rack setup (not to scale). Digestion of the peptide substrate by lawnmower blades, along with motion of the lawnmowers, can be monitored using TIRF microscopy.
4.1 A model of the conceived tumbleweed motor. The TW consists of three unique DNA-binding proteins called repressor proteins, which each bind with high affinity to a specific sequence of DNA in the presence of their respective small molecule ligands in solution, attached to the "arms" of a Y-shaped coiled-coil protein hub. (Structure file for this image created by Drs. Richard Sessions and Elizabeth Bromley [190]).

4.2 Illustration of the basic concept of the tumbleweed and its stepping mechanism. Figure from Ref. [179]. $R_A$, $R_B$ and $R_C$ are repressors that each act as a "foot" of the TW that can be made to bind and unbind from its respective well-defined recognition sequence A, B or C on the DNA track, depending on the concentration of its specific ligand (a, b and c) present in solution. See text for details of the TW stepping. The circular hub represents the rigid Y-shaped coiled-coil protein hub and flexible linkers are the ankle joints.

4.3 Schematic of the DNA track with attached TW motor in a DNA curtains setup. The track is anchored at one end to a surface, which has been made inert by means of a lipid bilayer, using a biotin/streptavidin interaction (left top) [97]. Several DNA tracks of about 14 $\mu$m length can be simultaneously stretched into a DNA "curtain" and allow observation of the movement of many TW motors at the same time. The DNA track consists of several repeats of the ABC motif of recognition sequences along which the tripedal TW motor steps (top right: $R_A$, $R_B$, and $R_C$ indicate the three repressor "feet", step size $\sim$11 nm). Figure from [192].

4.4 Positions of recognition sites of RsrII and EagI restriction enzymes in the pK$_8$ plasmid. RsrII and EagI digest pK$_8$ 1708 bp (580 nm) downstream and 1025 bp (348 nm) upstream of the K$_8$ unit, respectively.

4.5 Schematic of the ligation reaction between the pK$_8$ and Lambda DNA fragments (not to scale). The EagI overhang (blue sequence) of the pK$_8$ plasmid fragment ligates to the PspOMI overhang (purple sequence) of the Lambda DNA fragment. The sequences of the RsrII overhang (pK$_8$ fragment) and the 12 nucleotides of Lambda DNA overhang are coloured in red and brown, respectively.

4.6 EtBr-stained digestion products of (a) pK$_8$ with RsrII and EagI restriction enzymes, 3650 and 609 bp, in a 1% agarose gel and (b) Lambda DNA with PspOMI, 38412 and 10090bp, in a 0.7% LMP agarose gel. (c) EtBr-stained ligation product of 3650 and 38412 bp fragments from plasmid and Lambda DNA, respectively, in a 0.7% LMP agarose gel. The longer dark band should be the desired product. (d) EcoRV digestion of the K$_8$-Lambda fragment ligation product from part (c) shows that the ligation is successful, as evidenced by the extra high-molecular weight band in lane 3 compared with the digest of Lambda DNA alone (lane 2). The first lane in each gel shows the DNA ladder, with fragment lengths given in base pairs.
4.7 Biotinylation of the K₈-Lambda DNA construct, using ligation of short biotinylated duplex DNA, is successful. (a) EtBr-stained gel showing the expected molecular weight of the construct (open arrow). (b) Construct transferred to a nitrocellulose membrane, probed with streptavidin horseradish peroxidase, and visualized by chemiluminescence (ECL Plus, Pierce). Streptavidin binding coincides with the predicted size of the construct (filled arrow), indicating the presence of a biotin label on this high-molecular-weight DNA.

4.8 (a) Verification of formation of pLSNFplus plasmids by digestion with HindIII restriction enzyme, running in a 0.7% agarose gel and staining with EtBr. Lane 1 is the DNA ladder, lanes 2 and 3 show the expected pattern and an unexpected band, respectively, in digestion products of miniprepped plasmids from different grown colonies of cells. Cells corresponding to plasmids in lane 2 were maxiprepped for use in the next steps. (b) AvrII digestion products of pK₈ plasmid and pLSNFplus plasmid (lanes 3 and 4, respectively) were run in a 1% agarose gel and stained with EtBr. Lane 1 is the DNA ladder and lane 2 is the 3164 bp fragment from AvrII digestion of pK₈ after gel purification. This fragment was inserted into AvrII-linearized pLSNFplus plasmid (lane 4) to construct the long TW track.

4.9 (a) Digestion products of the pLSNFTW plasmid with AvrII, AflII and NcoI restriction enzymes were run in a 0.7% agarose gel and stained with EtBr and compared with (b) their predicted corresponding patterns on the virtual gel from Geneious software [193]. (c) The predicted digestion patterns of the pLSNFplus plasmid on a virtual gel with the same set of enzymes. The digestion patterns on the EtBr-stained gel coincide with the expected patterns of the virtual gel (b), confirming the formation of the expected plasmid. The first lane of each gel contains a DNA ladder.

4.10 The resultant pLSNFTW plasmid. AvrII restriction sites flank the inserted pK₈. The 8 kbp fragment between HindIII restriction sites contains the original BAC plasmid. AscI and XbaI restriction sites are used to generate the required length of handles for DNA tightropes and also overhangs for biotin and digoxigenin end-labelling of the TW track, respectively.

4.11 EtBr-stained gel and nitrocellulose membrane transfer of TW track construct after (a) digoxigenin and (b) biotin end labelling. Lane 1: 1kbp DNA ladder; lane 2: negative control DNA fragment; lane 3: positive control DNA fragment; and lane 4: the labeled TW track. Positive control is biotin or digoxigenin end-labeled DNA fragment with the same size as TW track with previously confirmed labelling. The negative control is the same length of DNA without label. Nitrocellulose membranes were probed with streptavidin or anti-digoxigenin horseradish peroxidase for biotin or digoxigenin, respectively, and visualized by chemiluminescence (ECL Plus, Pierce). Streptavidin and anti-digoxigenin binding coincides with the predicted size of the (A) positive control and (B) TW track construct for both labellings, indicating the presence of biotin and digoxigenin in this high-molecular-weight DNA. Work performed with assistance of Tristan Hansen.

4.12 Schematic of the (ABC)₈ TW track with biotin and digoxigenin end labels and an internal fluorescent fiducial marker (not to scale).
4.13 (a) EtBr-stained 1% agarose gel showing controls, lanes 2 and 4, and reactions, lanes 3 and 5, for low and high carboxyrhodamine 110-labelled 1 kbp alkyne-DNA. The black blob in the upper right is the excess of azide-fluorophore in the highly-labeled control and reaction. The highly labeled fluor-DNA does not appear, lane 5. Lane 1: 1 kbp DNA ladder. (b) Scan of the gel (before EtBR staining) at 532 nm excitation and 555/20 nm emission shows the reaction product between carboxyrhodamine 110-PEG3-azide and 1 kbp alkyne-labelled DNA (lane 3) and unreacted fluorophores above the wells.  

4.14 Absorption scans confirm the fluorescent labelling of the 1000 bp alkyne-labeled DNA by the click reaction. Controls are unlabelled DNA reacted with azide-fluor under identical conditions. The 260 nm absorption peak corresponds to DNA, and that at ~500 nm to carboxyrhodamine 110.  

4.15 TIRF microscopy images of the low and high labelled 1 kbp DNA and controls confirms the fluorescent labelling. The brightness of the images of the high labeled samples shows the higher degree of labelling for these samples. Images are recorded with identical exposure ($\lambda_{exc}=488\pm2$ nm) and all images have the same settings for brightness/contrast. Images are courtesy of Cassandra Nimann, PhD student, Lund University.  

C.1 Evolution of position probability distributions $P(x_{cm})$ (a) for HOH spiders without cleavage and without detachment and (b) calculated using the model presented in this Appendix.  

C.2 Graph of variance versus time for HOH bipedal spiders on a $P$ track and on an $S$ track. Diffusion coefficients of spiders on $P$ and $S$ tracks are calculated from the slopes.  

D.1 Activity of proteases after each chemical treatment step toward construction of lawnmower blades investigated using a fluorescence assay. Trypsin showed the highest preserved activity after different chemical modifications. Note: Proteinase K showed a very high activity but was eliminated due to its non-specific proteolytic activity. (Figure courtesy of Dr. Suzana Kovacic.)  

D.2 Western blot analysis of biotinylated TCEP– and Traut– treated proteases. Protease controls are unbiotinylated proteases which act as negative control. The dark intense bands in the biotinylated proteases are indications of biotin labelling of these samples. (Figure courtesy of Dr. Suzana Kovacic.)  

E.1 Electron spray ionization mass spectroscopy of peptide-azide shows the high intensity peak at 1732.85, which coincides with the expected mass of 1733 g/mol for peptide-azide. (Figure courtesy of Dr. Suzana Kovacic an obtained with an ESI ion source on an Agilent Time-of-Flight LC/MS mass spectrometer.)
Chapter 1

Introduction

The cellular environment is a crowded and noisy medium where molecules are randomly colliding with each other. Inside the cell, cellular activities are performed by proteins at a scale where thermal fluctuations predominate. Recent advances in experimental micro- and nanoscale techniques have changed the view of a cell being simply a confined "chemical reactor" [1]. Instead, experiments have revealed many facts about the structures and precise functions of components of cell machinery and demonstrated that a cell is equipped with a variety of molecular machines to perform complex tasks on time scales much shorter than could be accomplished by random diffusion. For example, a basic calculation of the time required for transport of a protein along the 1 m long axon based on diffusion shows that this time would be on the order of 60 years [2], which is clearly much longer than needed. Conventional and newly developed microscopy techniques have illuminated how mechanical tasks such as cell division and intracellular transport in living cells are performed using a large collection of protein molecular motors [3, 4, 5, 6].

Biomolecular motors are predominantly multi-subunit proteins, which transduce chemical energy into mechanical work. They are central to many cellular processes and functions [3, 5, 6], including DNA replication and repair [7], protein synthesis [8], cell division [9, 10, 11], intracellular transport [11, 12], cell motility [9, 13, 14], and ion pumping [15]. It is therefore a major goal in structural molecular biology to develop a clear biochemical and biophysical understanding of the general operational principles of such motors. Experimental studies of biological motors using techniques such as protein-structure analysis, kinetic studies, genetic modification, and single-molecule experiments [16, 17, 18, 19, 20] have provided the basis for several models of biological motors and machines [3, 21, 22, 23, 24]. These studies shed light on the details of motor operation, perfor-
However, it is also of great importance to establish a solid, fundamental understanding of the basic construction and operational principles of generic protein motors. Here, following Richard Feynman’s quote of "What I cannot create, I do not understand" [25], my goal is to test and improve our current understanding of fundamental, operational principles of molecular motors by designing, constructing and characterizing fully synthetic, protein-based motors. In this, my work is part of an international collaboration to work together theoretically on design and experimentally on construction of two protein-based molecular motors: the Tumbleweed and the Lawnmower. The translocation of these molecular motors requires specific tracks on which they can walk. Therefore, construction of the relevant molecular-scale track along with an appropriate detection method is of a great importance to our goal. The function of these motors on their designed tracks plus my contribution to these projects are going to be explained in this thesis.

In the following, an overview of natural molecular motors and their physical properties is presented (Section 1.1). Then, in Section 1.2, the "Brownian Ratchet" as a general model of molecular motor function is introduced. Following this, I describe synthetic DNA-based molecular motors in Section 1.3. (To the best of my knowledge, no synthetic protein motors have been constructed by others.) In Section 1.4, some of the relevant single-molecule techniques to study the function of molecular motors are explained. The choice among these techniques impacts design of tracks for the synthetic motors. In Section 1.5, a description of the synthetic motors conceived by our international team is presented along with an overview of my contributions, which are the basis of this thesis.

1.1 Physical Properties of Molecular Motors

A motor is a device that converts various energy forms such as chemical, electrical or thermal energy into mechanical work. Biological molecular motors are a molecular subclass of these machines that convert different forms of chemical energy, such as hydrolysis of ATP and concentration gradients of ions, into mechanical work. These microscopic engines are similar to heat engines, as they are both cyclic machines. However, unlike heat engines in which work is produced from a thermodynamic cycle that transfers heat from a hot to a cold reservoir, biological molecular motors perform work via biochemical cycles, such as ATP hydrolysis.

Biological molecular motors can be divided into two classes: rotary motors such as $F_0F_1$-
ATPase [18] and the bacterial flagellar motor [26]; and linear motors such as the three superfamilies of cytoskeletal motors, myosins, kinesins and dyneins, which participate in a wide range of processes that occur in cells such as mitosis, cell division, organelle transport and organelle synthesis [5, 6]. Here, due to their similarities to the synthetic molecular motors of our interest, we focus on the linear molecular motors.

Figure 1.1 shows structures of these three classes of molecular walkers with their cargo- and track-binding ends specified. Kinesin, dynein and myosin are walkers that can travel short or long distances along the cell’s "highways" (microtubules or actin filaments) to perform different tasks. Kinesin and dynein family members move along microtubules while the members of the myosin superfamily move along actin filaments. There is a great diversity in the properties of these motors.
even within one family, and they vary considerably in their structures and detailed operations [6].

In spite of differences in their cellular roles and structures, these motor proteins share several important characteristics. First, in contrast to macroscopic machines, they operate under conditions of low Reynolds number where inertia and momentum are irrelevant and viscous forces and random thermal motion strongly influence the dynamics [29, 30]. Thermal, nondeterministic motion is thus an important aspect of the dynamics of motor proteins [31]. Second, they convert the free energy input of ATP hydrolysis into mechanical work. In spite of such noisy environments, using the energy derived from hydrolysis ATP molecules (∼ 80 pN·nm per ATP at physiological conditions [32]) molecular motors perform force-generating tasks with efficiencies higher that of macroscopic motors (around 50%, for these linear motors [1]). Third, their motion is restricted to migration along rigid, essentially one-dimensional tracks.

Some physical characteristics of conventional kinesin and myosin such as directionality, processivity, mechanochemical coupling and stall force are presented in the following.

Directionality is defined as migration of the molecular walker preferentially or exclusively towards one end of a molecular track. Microtubule and actin filaments are polymer tracks with polarity so that they both have "+" and "-" ends. Along with many other members of these superfamilies, conventional kinesin and myosin move toward the plus end of a microtubule and an actin filament, respectively [6, 22, 27, 33]. However, there are examples in each family that move directionally toward the minus end of the microtubule or actin filaments, respectively [6].

Processivity is defined as the ability of a molecular walker to remain attached to its track during its operation, i.e. to migrate along the track for more than a single motor cycle. Since molecular walkers are constantly jostled by thermal fluctuations, mechanistically this is not trivial to achieve. Experimental studies showed that conventional kinesin steps with steps of size 8 nm in a hand-over-hand manner along a microtubule [16, 34, 35]. Kinesin can step processively hundreds of steps on a microtubule with a velocity of ∼1 μm/s [16, 36, 37]. In contrast, conventional myosins (responsible for contraction of muscle) detach completely from actin at the end of each hydrolysis cycle and do not move processively [17, 19]. Therefore, not all of the naturally occurring molecular motors operate in a processive manner, but non-processive motors can perform work in large ensembles. As an example, there are about 250 conventional myosin molecules in every thick filament [38] and they work together so that the actin and myosin do not drift apart in the muscle contraction process.

One fundamental concept in molecular motor studies is the coupling between chemical input energy (the hydrolysis of ATP) and directed motion. This concept can be defined through the
mechanochemical coupling ratio, which relates the number of ATP molecules hydrolysed to the number of mechanical cycles (e.g. steps for kinesin) during a single enzymatic cycle. For motors with a defined step size, one practical method to calculate mechanochemical coupling is by using the following definition,

\[ \xi = \frac{\langle v \rangle}{\delta \langle r \rangle}, \]  

where \( \langle v \rangle \) is the mean velocity, \( \langle r \rangle \) the mean chemical reaction rate, and \( \delta \) the step size [23]. For tightly coupled molecular motors \( \xi \) will be close to unity, while for the loosely coupled molecular motors in which a significant number of chemical steps do not lead to motion (or lead to partial steps), \( \xi \) will be less than 1. Both kinesin and myosin hydrolyse one molecule of ATP for each step and hence are categorized as tightly coupled molecular motors [39, 40]. In these motors the process of ATP hydrolysis induces conformational changes in the motor structure so that it triggers the forward stepping of the motor by generation of a power stroke. It should be noted that, depending on the type of motor, the power stroke can be generated at nucleotide binding, nucleotide hydrolysis, or phosphate release [22, 41, 42]. As an example, ATP binding by the leading, microtubule-bound head of kinesin results in conformational changes in the neck linker between the two legs so that it throws the freely diffusing tailing head forward by 16 nm. Therefore, ATP binding drives the trailing moving forward and biases the motion [23, 43]. Due to the high coupling and correlation between consumption of chemical energy and performance of mechanical work, biological molecular motors are categorized as highly efficient motors (especially when compared to macroscopic motors).

Experimentally, thermodynamic efficiency of biological molecular motors is calculated by applying a force opposing forward motion, e.g. by using single-molecule techniques such as a laser trap [44] or atomic force microscope [45]. As the motor works against an external conservative force, it is possible then to measure the thermodynamic efficiency by calculating the ratio of the work done by the motor to the energy input (e.g. energy of ATP hydrolysis) using the following formula,

\[ \eta_T = \frac{F \delta}{-\Delta G} = \frac{F \langle v \rangle}{-\Delta G \langle r \rangle}, \]  

where \( F \) is the applied conservative force and \( \Delta G \) is the free energy of ATP hydrolysis. This equation is valid for tightly coupled molecular motors in which one ATP is hydrolyzed per one molecular motor step [23, 46]. The calculated efficiencies of kinesin and myosin are about 50-60\% [6, 23].

The stall force is defined as the external force under which the mean velocity of the molecular
motor is zero. This is measured using the single-molecule techniques mentioned above, here by obtaining velocity-force curves. The stall force of a molecular motor represents the average maximum force that the motor itself can generate to move forward during its mechanical cycle. The average calculated stall forces of kinesin and myosin motors are 7 and 5 pN, respectively [6, 23].

Studies of molecular motors draw on a diversity of fields from biochemistry to mathematics. Due to great progress in experimental studies of molecular motors and quantitative data, physicists and mathematicians have presented various theoretical models to understand the dynamics of molecular motors, taking into account that thermal fluctuations are not negligible [21, 30, 47, 48, 49].

In a more general approach, the physics community has established the conditions under which a molecular-scale object exposed to substantial random (thermal) motion can perform directed motion in the absence of a macroscopic chemical or force gradient [50]. In this context, molecular motors are classified as Brownian motors [51, 52], that is, systems that combine local asymmetry, thermal noise, and a deviation from thermal equilibrium to induce directed motion in the absence of macroscopic forces or gradients.

The correlation between thermal fluctuations and Brownian motor theory in the function of molecular motors is well described by Astumian in his quote that "any microscopic machine must either work with Brownian motion or fight against it, and the former seems to be the preferable choice" [53].

The operational principles of simple Brownian motors, including the constructive role of thermal noise, as well as their relationship to biological motors are now well understood [21, 29]. The simplest model of Brownian motors is the "Brownian ratchet" [21, 51, 54, 55], in which periodic or random application of an asymmetric, spatially periodic potential to freely diffusing particles results in generation of a net current. Due to its applicability to the function of many molecular motors, in the following section a basic model of a flashing ratchet is explained.

1.2 Brownian Ratchet

In general, a Brownian ratchet (inspired by Smoluchowski’s and Feynman’s famous ratchet and pawl [56, 57]) is a model that is useful for investigating the general characteristics of diffusion, systematic asymmetry, and non-equilibrium processes in motor motors [47, 50, 52]. Here, I use conventional myosin (involved in muscle contraction) as the biological context for describing how a Brownian ratchet performs mechanical work.
Figure 1.2: (a) Illustration of the cycle of changes in myosin shape during cross-bridge cycling [58] (states 1, 2, 3, and 4) (see text for details). (b) Ratchet model for myosin. In the attached state, the myosin head is located in a local minimum of the potential of interaction between myosin and an actin filament (potential $V_A(x)$). After an ATP is bound by the myosin head, the head detaches from the actin filament and undergoes free Brownian motion (2 on $V_D(x)$). After ATP hydrolysis (3), and because of the asymmetry of the potential once the myosin head attaches to the actin filament, the particle is more likely to be found in a region of negative slope on the potential $V_A(x)$. The particle then either returns to its starting location on $V_A(x)$ or moves to the next local minimum of the potential (from 4 to 1) (Figure is adapted from [59]; used with permission).
It is known that the basis of muscle contraction is a longitudinal movement of myosin and actin filaments relative to each other known as "actin and myosin cross-bridge cycling" [58]. As described previously, this interaction is coupled with ATP hydrolysis in the myosin heads. Figure 1.2(a) illustrates the "cross-bridge cycling". In the absence of ATP, the myosin head is strongly bound to actin filament (state (1)). However, when a myosin head binds ATP, its affinity to actin strongly decreases and results in the detachment of the myosin head from the actin filament (state (2)). Subsequent ATP hydrolysis is accompanied by conformational changes in the myosin molecule, then binding to an actin filament in the ADP.P\(_i\) bound form (states (3) and (4)). At this step, the myosin-actin binding is weak and a slight conformational change occurs on myosin that promotes the release of inorganic phosphate (P\(_i\)). Following this binding and release of P\(_i\), which reinforces the myosin-actin interaction, the orientation of the myosin head relative to actin is significantly altered, triggering the "power stroke" and leading to movement of the actin filament along the myosin filament. This force-generating step on the actin filament brings the head back to its original conformation. As myosin regains its original conformation, the ADP is released, and the myosin head remains tightly bound to the filament at a new position from where it started, thereby bringing the chemical cycle back to the beginning with translocation of the motor relative to its track. The myosin head again binds ATP, dissociates from actin, and the cycle repeats again and again. As a result of this cycle, a "sliding" of myosin and actin filaments occurs relative to one another, which is the basis for muscle contraction.

Based on this scenario one can assume two states, attached and detached, for a myosin head interacting with an actin filament [59]. In the detached state, the myosin head is assumed to undergo free diffusion away from the actin filament. Therefore, it experiences a flat potential. However, in the attached state, the motion of myosin head depends on the myosin-actin interaction. Due to the asymmetry of the actin filament and its geometric periodicity, I can assign an asymmetric and periodic potential to this state of the system where the bound myosin head is located at the minimum of the potential.

In Brownian ratchet theory, molecular motors are described as a particle switching between two states of a potential. The myosin-attached state is similar to having the asymmetric potential in the "on" state. So in the detached state, the asymmetric potential is off and the myosin-actin system is similar to a freely diffusing particle in the constant flat potential, where the probability distribution of the position of the particle follows a Gaussian function and spreads symmetrically with time. Following some waiting time, because of asymmetry of the potential, the particle is now more
likely to be located in the region corresponding to the negative slope in the asymmetric potential. Thus when this potential is switched on (Fig. 1.2(b)), an average forward movement of the particle results.

Based on this model, the three components of a Brownian ratchet necessary for achieving transport are a randomizing element (thermal energy), asymmetry of the potential in the dimension in which motion occurs, and free energy input. The randomizing element spreads the particle(s), asymmetry in the potential shape results in directionality, and free energy input cycles the potential, thus taking the system out of equilibrium and performing work on particle(s) to confine them.

Although here myosin function was used as a model to describe the Brownian ratchet model, similarities among biological molecular motors mean that this model is applicable to others such as kinesin and dynein [54, 60]. Similarities with these are hydrolysis of ATP as the free energy source, thermal energy as the randomizing element, and the periodicity and polarity of the microtubule track to produce the required asymmetry.

The above ratchet model is the simplest model introduced to study molecular motors. There are many variations of the Brownian ratchet to fit the dynamics of specific molecular motors and to include experimentally observed power strokes and tight mechanochemical coupling. Some of these variations consider an asymmetric periodic potential subject to a time-correlated force [61], a periodic potential subject to a two-valued fluctuating potential [39], or a three-state asymmetric potential [62].

Besides the above examples, there is a class of biological molecular motors whose function can be explained by a different Brownian ratchet model, called "Burnt-Bridges Brownian ratchet". In this class, the mechanism of motion is again biased diffusion, here with the bias dependent on the cleavage of the track by the motor. One example of these motors is MMP-1, a type of collagenase [63, 64, 65, 66], which is the biological inspiration for our lawnmower molecular motor. MMP-1 is a protein enzyme that moves processively on collagen fibrils [63, 66]. Collagen fibrils are self-assembled rod-like structures consisting of aligned collagen molecules, with orderly positioned cleavage sites that can be recognized and cleaved by MMP-1. As first proposed by Saffarian et al. [63], MMP-1 undergoes ATP-independent biased diffusion along collagen fibrils by digesting its cleavage sites, thereby following the burnt-bridges Brownian ratchet mechanism. Based on this mechanism, cleavage of the collagen track by the enzyme passing the site acts as a roadblock so that it prevents the backward motion and biases the enzyme diffusion toward the forward direction. Therefore, based on this model, MMP-1 is a Brownian ratchet that is able to rectify diffusion by
coupling its motion to collagen proteolysis.

There are other examples of systems that use a burnt-bridges mechanism for their function. One example, which has been studied in prokaryotic cells, is Par-mediated chromosome segregation in *C. crescentus* bacteria [67]. It is possible that the burnt-bridges mechanism might be applicable to the function of other protein enzymes such as exonucleases [68], which function by cleaving nucleotides from the end of a nucleic acid chain. It is important to note that the burnt-bridges mechanism does not necessarily exclude consumption of ATP, which can be involved in the function of the both of the above examples.

### 1.3 Synthetic DNA-Based Molecular Motors

Inspired by biological molecular motors stepping along filaments and by the physics of Brownian motors, several approaches to synthetic molecular motors have been developed. These include DNA-based nano-walkers and synthetic chemical machines in which motion can be controlled by molecular structures [28, 69]. An overview of chemically synthesized machines such as the "molecular gear" and "molecular beaker" can be found in a review by Kay *et al.* [69].

Due to its excellent molecular recognition capability and predictable secondary structures, DNA has been exploited for building nano structures [70, 71], machines [72, 73, 74] and walkers [28, 75]. DNA walkers provided the first experimental proof of the feasibility of molecular walkers. This approach, first demonstrated by a DNA walker capable of performing steps along a DNA track in a non-autonomous fashion, fuelled by single stranded DNA [76], was followed by both non-autonomous [77, 78, 79, 80] and autonomous molecular motor designs [81, 82, 83, 84, 85, 86, 87]. One example of an autonomous molecular motor is the "molecular spider", which uses enzymatic reactions to power directional motion [86, 87] (see Chapter 2 for more details).

As an example of their operation, the functional principles of the first non-autonomous DNA-based walkers are explained. The main components of this class of walkers are generally as follows: 1) a walker (usually bipedal); 2) a track; 3) attachment fuel strands; and 4) detachment fuel strands. The number of attachment and detachment fuel strands depends on the DNA walker and track design.

Figure 1.3 shows the walker designed by Shin and Pierce [77] which I use as an example to explain the working principle of these motors. This DNA walker uses four single-stranded attachment strands, A1 to A4, and four detachment strands, D1 to D4, which are complementary to the
attachment strands. Each leg of the walker consists of a single-stranded DNA molecule, and these are joined in a duplex in the "body" of the walker. The track is double-stranded DNA branched with ssDNAs, all on the same side of the track and approximately 5 nm apart. Application of the attachment fuel A1, which is partly complementary to a leg and partly complementary to a track branch, specifically anchors the walker to a branch by inducing hybridization of the corresponding leg and branch with the A1 fuel. Subsequent application of A2 results in the walker's second leg binding to the track. From this state, the trailing leg is released by supplying the D1 strand. As shown in the figure, the attachment fuels are designed so that besides the complementary part to the track and the leg, they leave an unhybridized overhang to be used as a "toe-hold" by the detachment strand to release the bound legs. D1 thus nucleates with the perfectly complementary A1 strand at its 10-base overhang and then undergoes a strand displacement reaction to produce duplex waste and free the walker leg for the next step. Duplex replacement is driven by the free-energy difference between the two duplexes. The sequential application of appropriate fuels results in the stepping of the walker along the track. The majority of DNA-based walkers function based on this principle, and their successful operation has been demonstrated [28, 76, 77, 78, 79, 80, 88].

Other artificial DNA walkers, including molecular spiders [86, 87], rely on a "burnt-bridges" strategy to perform directional motion. With the goal of producing autonomous motion, in this strategy the molecular motor achieves directionality by chemically altering each traversed binding site on its track (either by an enzymatic reaction or by "covering" it with a strongly bound fuel) [82, 86, 87, 88, 89].

Both non-autonomous and autonomous DNA based walkers are designed ideally to avoid backward stepping so as to bias binding of a diffusing foot to a forward binding site. In other words, this reduced probability of backward binding acts as a ratchet, rectifying diffusion to produce a forward motion of the walker.

These nucleic-acid-based approaches to artificial molecular motors are very promising because relatively straightforward design rules can be used to construct nanoscale devices by deterministic self-organization. At the same time, however, the properties and design rules of RNA and DNA motors may be different from those of protein motors, limiting their usefulness as model systems for biological motors. Also as proteins have more diversity in their properties compared to DNA and RNA, they provide various building blocks in order to be used to introduce desirable specificities and complexities to a synthetic molecular motor. Because of these properties, and the ubiquity of protein-based motors in nature, the focus of our group is to design protein-based molecular motors.
Figure 1.3: An example of a DNA nanowalker that executes stepwise movement along its linear track. Shown here are the initial binding of one foot to the track (via attachment strand A1), forward binding of the second foot (via attachment strand A2), and detachment of the trailing foot (via detachment strand D1). See text for details. Figure is taken with permission from [77].
In addition some proteins are able to adopt different conformations. This property makes them potential candidates for future designs of molecular motors incorporating a power stroke, in which it would be necessary to couple the conformational change with the energy cycle of the motor.

Another aspect important to note is that for the above autonomous or non-autonomous molecular motors, the track design and the nature of the motor-track binding interaction play essential roles in the directional motion of molecular motors. In general, for a molecular motor a track is an extended structure with a periodic array of motor binding sites. Therefore, synthesis of an appropriate track is critical to the success of a motor’s function. Hence, I also design and construct molecular tracks for our synthetic protein-based molecular motors. Beside its ability to promote motor directionality, a track must be designed to be used in an appropriate setup for observation and characterization of motor function. In the following an overview of the relevant techniques as potential observation methods (being implemented in our collaborator Heiner Linke’s lab) is presented.

1.4 Observation Techniques

1.4.1 DNA Curtains and Racks

DNA curtains and racks (shown schematically in Fig. 1.4(a) and (b), respectively) were first introduced by Eric Greene’s group as high-throughput tools for single-molecule optical imaging [90, 91]. The great advantage of these methods is the simultaneous detection of up to hundreds of individual DNA molecules within a single field of view. In these techniques, which use flow to stretch DNA, experiments have been performed in TIRF microscope-mounted microfluidic chambers to study protein-DNA interactions at the single-molecule level. Using these techniques, they visualized the dynamics (such as target-search mechanisms) of a broad range of DNA-binding proteins that function in homologous recombination and mismatch repair [92, 93, 94, 95].

To make DNA curtains, first, the fused silica surface substrate is patterned using e-beam lithography with a barrier perpendicular to the planned direction of flow. The substrate is then passivated with a lipid bilayer containing a small percentage of lipids labeled with biotin. These subsequently react with neutravidin. Biotinylated Lambda DNA (\(\sim 16 \mu m\)) can then be attached to the surface by binding to neutravidin (which can bind up to four biotins). Applying flow, the diffusive lipid bilayer reorganizes and lipid-tethered DNA molecules are aligned at the edge of the barrier. The DNA molecules extend and stretch out over the barrier in the flow direction and, because they are
extended parallel to the surface, remain confined within the evanescent field generated by total-
internal reflection of visible light (Fig. 1.4(a)).

To keep the DNA extended in a curtain, flow needs to be applied continuously. To study molec-
ular motor dynamics on a track, the presence of flow could affect the motion due to the associated
drag force. Even for small objects such as DNA-binding proteins, changes in their behaviour have
been observed in the presence of flow in DNA curtains [96, 97]. As the drag force scales with hydro-
dynamic radius, multidomain molecular motors would experience a greater force than the proteins
studied in these references.

To avoid effects of flow during the measurements, Gorman et al. developed DNA racks, which
are DNA curtains pinned down at their distal end to a coated metallic barrier [91]. The DNA used in
racks is doubly end-labeled with biotin and digoxigenin, and the metallic barrier is coated with anti-
digoxigenin. When the biotin-tethered DNA extends under flow (as in curtains), the digoxigenin
end can be bound by anti-digoxigenin on the metallic barrier. The aligned and anchored DNA
maintains extended form in the vicinity of the surface after flow stops (Fig. 1.4(b)). Both DNA
curtains and racks have been used in studying the dynamics of many DNA-interacting systems
[91, 92, 93, 94, 95, 96, 97, 98].
1.4.2 DNA Tightropes

Another approach to studying the dynamics of protein-DNA interactions in real time and at the single-molecule level involves DNA tightropes. Similar to DNA curtains, DNA tightropes do not require flow during visualization.

Here, DNA is extended and suspended between micron-sized beads above a surface [99, 100]. In this flow-based technique, a PEGylated glass substrate with positively charged polylysine beads attached is prepared in a flow chamber. Flushing the 16 µm long Lambda DNA through the chamber results in non-specific binding to the beads of negatively charged DNA, which remains stretched under flow, potentially binding to a second bead to become pinned in a stretched configuration. This DNA tightrope set-up can generate many extended DNA molecules bridging between each pair of beads.

In DNA tightropes, DNA is on the order of one micrometer away from the surface. This is too far for TIRF microscopy. Instead, the labeled DNA and proteins are observed using Oblique Angle Fluorescence (OAF) microscopy, which gives control over observation at the height of the beads [99, 100]. This approach reduces the background from fluorescently labelled molecules attached nonspecifically to the substrate surface.

Compared to DNA curtains/racks, making DNA tightropes is a simpler method that does not require nano-fabrication or patterning techniques, though it does not provide control over the spacing of the stretched DNA molecules or their directionality. If DNA tightropes form under flow, the DNA remains extended in the absence of flow, permitting visualization under a no-flow condition. For all three techniques, long DNA, on the order of ten micrometers, is needed to permit its initial extension by flow.

1.5 Overview of Synthetic Molecular Motors in Our Group

Here I describe our research team’s approach to synthetic protein-based motors, which has as its goal the establishment of a system that will allow us to perform controlled experiments on the relationship between structure and function in protein motors. This task is challenging, but lays a foundation for taking advantage of the great mechanical, kinetic, and structural versatility and specificity afforded by Nature’s material of choice for nanoscale machines. Our group uses molecular biology and biochemistry to combine non-motor protein domains in a way to produce motor
function. Since the design details are known precisely, the resulting motor designs are amenable to detailed numerical and analytical modelling, even though their complexity may ultimately approach that of a biological system. Models of these motors can then be compared to results from single-molecule experiments. This novel bottom-up route would allow us to test and further our understanding of structure-function relationships in natural motors.

Toward this goal, as part of an international collaboration, our group has taken the lead in the design and construction of a linear autonomous molecular motor I dub the "lawnmower". The lawnmower is a protein-based molecular motor designed to use the burnt-bridges strategy to produce asymmetry for ratchet-like directional motion. It uses enzymatic reactions between its protease feet and peptide substrates on a designed track to direct motor motion. While our long-term goal of synthesizing artificial molecular motors in assembling a molecular scale device capable of performing task such as programable transport, our short-term goal is to construct a functional molecular motor, which is able to move directionally and autonomously along a specifically designed track. To characterize its performance, the motion must occur over length scales and timescales detectable with the available single molecule technologies. For $\sim 10$ nm steps considered here, our initial goal would be to detect 10 directed processive steps occurring in $\leq 60$ min. My work combines simulations on the related system of "molecular spiders" (Chapter 2), to provide an understanding of operational principles, with experimental design and construction of the lawnmower and, of particular focus here, my work on its track (Chapter 3).

The second motor featured in this thesis is the "tumbleweed", a linear but non-autonomous molecular motor. The tumbleweed is designed to be a three-legged self-assembled protein complex that moves by cyclically ligand-gated, rectified diffusion along a DNA track, using three discrete ligand-dependent DNA-binding domains to control binding and unbinding events. Each research group in our collaboration has a specific role in achieving this project, and my contribution is construction of its DNA track. Tumbleweed function is described in Chapter 4, where I also outline my work constructing its DNA track to be amenable to visualization of motor activity.

Chapter 5 contains conclusions and suggestions for future directions.
Chapter 2

Molecular Spiders

2.1 Introduction

One type of autonomous synthetic walker whose performance has been studied as a function of experimental parameters is the molecular spider [86, 87]. Molecular spiders are nanoscale synthetic DNA-based molecular walkers capable of generating autonomous biased motion by interaction with their track. In this chapter, a theoretical study of biased motion of molecular spiders is explained, while referring to the models and results of other studies. One of our main motivations to perform simulations of molecular spiders was to explore effects of different design parameters on the biased motion of our synthetic protein-based molecular motor "lawnmower" by studying a similar model system. Application of my simulation results to narrow down the choices of feasible design elements for synthesizing lawnmower is described in Chapter 3.

A molecular spider consists of a streptavidin hub with attached legs, each of which is an 8-17 deoxyribozyme capable of binding to, cleaving, and releasing from a specific substrate (Fig. 2.1). The substrate is a single-stranded DNA (ssDNA) to which the DNAzyme leg can hybridize, and contains a single ribonucleotide bond at the cleavage site. Following hybridization, the substrate is cleaved to the resulting products ($P_1, P_2$) in the presence of a divalent cation, which acts as a cofactor of the reaction.

The directional motion of molecular spiders has been studied experimentally with both bulk and single-molecule techniques [86, 87]. Pei et al. assembled molecular spiders from one or two streptavidin hubs linking flexibly a controllable number (2, 3, 4 or 6) of catalytic legs. In their first experiments, ssDNA substrate was distributed throughout a quasi-two-dimensional matrix (of 100-
200 nm depth), and a large number of spiders was dispersed initially throughout this matrix [86]. By monitoring via surface-plasmon resonance (SPR), the activity of the spiders was inferred from the reduction in substrate DNA bound to the surface (the rate of $P_2$ product release while $P_1$ remains bound) after supplying $Zn^{2+}$ to the environment. Using this technique the authors were able to learn about the rate of the reaction and dissociation rate of spiders to solution. These rates changed for different spider assemblies depending on the number of legs and on the binding strength of the legs to $S$ and $P_1$ (where in all cases, binding to $S$ is energetically more favourable than binding to $P_1$). The results suggested that the spiders exhibited biased motion, moving preferentially toward an uncleaved substrate rather than remaining in an already-cleaved patch of product DNA, and were processive in the sense that they were able to remain bound to the matrix for multiple catalytic turnovers [86]. The mechanism proposed to account for their observations involves a spider cleaving a patch of substrate by cycles of binding, cleavage and release by each leg, where processive biased motion results from (a) enhanced binding to substrate compared with product; and (b) maintaining at least one leg of the spider bound to the matrix at all times. Here, I define processivity as the number of cleavage events undertaken by a spider before its complete dissociation from the surface. The preferential binding to substrates implies a biased diffusion mechanism for spider motility: after initially clearing a patch of substrate (converting it, via cleavage, from $S \rightarrow P_1$), a spider will...
preferentially move towards an uncleaved region, where it gains in binding energy and can release free energy upon substrate hydrolysis. This study also investigated the dependence of mean velocity and processivity on tunable experimental parameters, demonstrating, for example, that increasing the number of legs or the binding strength of legs to substrates led to increased spider processivity.

In the single-molecule study of molecular spiders, a single spider with three deoxyribozyme legs was released from a "start" site at one end of a quasi-one-dimensional (1D) substrate track presented on a DNA origami surface [87]. The molecular spider then performed a series of experimentally designed actions such as "follow", "turn" and eventually reached the "stop" site. Statistical analysis of atomic force microscopy (AFM) images from parallel experiments showed that, as time increased, the population of spiders diminished at the "start" site, while an increasing number of spiders reached the "stop" site at the other end of the track. Tracking of single spiders by total-internal-reflection fluorescence microscopy (TIRFM) also confirmed the directional motion of the spiders from "start" to "stop". The calculated average speed of the spiders (1-6 nm/min) is faster than the diffusion rate of molecular spiders from the "start" to "stop" site on an all-product track, evidence that the motion is biased. Analysis of the experimental results showed that the ability of spiders to reach the "stop" site was limited by detachment from the track and backtracking of the spiders, and that the probability of spiders reaching the stop site decreased with increasing track length.

In addition to the discussed experiments, analytical and computational models have also studied the biased motion of molecular spiders on a 1D track [101, 102, 103], and their dynamics on a 2D plane [104]. In two or three dimensions, the biased motility would be hard to observe, appearing as a weakly self-avoiding walk, but in one dimension, the preference to bind S sites should give rise to a more clearly observable biased random walk. Thus, theoretical treatments of the spider are easier to undertake and analyse in one dimension.

The motility of various molecular spiders on a 1D periodic lattice track was investigated initially in two theoretical articles by Antal et al. [101, 102]. Here, I highlight the points most relevant to the experimental case of the symmetric bipedal spider (i.e., two identical legs) and to my modelling. In [101], motility of spiders with multiple legs on a 1D lattice was examined by analytical solution of the appropriate Master Equation given particular rules for leg stepping. For bipedal steppers with maximum leg separation of two lattice sites and nearest-neighbor stepping ($m \leftrightarrow m \pm 1$, where each leg can step with equal transition probability only to its unoccupied nearest neighbor site(s)), they found a diffusion constant of the spider of $D = 1/4$ and no biased motion. I classified this stepping
mechanism as an unbiased inchworm (IW), which is one of the mechanisms used in my studies (see Section 2.2.1). An alternative mode of bipedal motility considered was the "quick" spider, which always has its two legs on nearest-neighbor lattice sites and moves by putting one leg either in front of or behind the other with equal probability. For this stepping mechanism, they also found no bias, and showed $D = 1$, not surprising, given that the centre-of-mass motion of this spider per step is twice that of the IW ($\Delta x = 1$ versus $\Delta x = 1/2$, respectively). I classified this stepping mechanism as unbiased hand-over-hand (HOH), which is the second of the mechanisms used in my studies. This initial work of Antal et al. considered legs only to bind and unbind from a homogenous lattice, whereas the experimental spider has the ability to alter the lattice sites (cleavage from $S$ to $P_1$) upon binding. Since legs interact differently with $S$ and $P_1$ sites, this cleavage of the track introduces a memory effect into the system, making the problem non-Markovian.

Further work by Antal and Krapivsky [102] extended their analysis of spider motility on a 1D periodic track by including memory effects. In this model, each lattice site can be either an unvisited site representing a substrate, $S$, or a visited site representing the product, $P_1$. A spider leg on a visited ($P_1$) site can jump to a neighboring site with rate 1 whereas a spider leg on an unvisited ($S$) site is assigned a rate $r < 1$ of jumping to a neighboring site. The parameter $r$ implicitly includes cleavage of the substrate and release of products. After the spider jumps, the unvisited site becomes a visited site, which gives the motion a non-Markovian character. The authors point out that, after some time, the visited sites form an "island" in a sea of unvisited sites. It is shown analytically, that memory effects, $r < 1$, lead to faster visitation of $N$ new sites for spiders with 2 or more legs compared with the no-memory case ($r \sim 1$). This biased diffusion agrees with the interpretation of experimental molecular spider motion given by Pei et al. [86].

In a more recent study of dynamics of molecular spiders, Semenov et al. used kinetic Monte Carlo (MC) simulations based on the memory model described in the previous paragraph (the "AK model") to explore the mechanism of diffusive transport by molecular spiders [103]. Comparing the behaviour of bipedal spiders on a 1D track with and without memory effects (hopping with rate $r$ from a newly visited site), they showed how faster visitation of $N$ new sites occurs when spider stepping takes more time ($r < 1$) from newly visited sites. They concluded that in spite of stochastic hopping, this difference in the rate leads to an effective bias toward unvisited sites and generalized their study to multipedal spiders.

Using diffusion constant analysis, Semenov et al. observed that, in the long time limit, the asymptotic behaviour of spiders is diffusive for all values of $r$. This long-term behaviour agrees with
the analytical predictions of the AK model [102]. Semenov et al. [103] reformulated the AK model and used kinetic MC simulations to study the transient behaviour of molecular spiders as a function of time on a 1D substrate track for $r < 1$. By analyzing the mean-square displacement of bipedal spiders as a function of time $<x^2(t)> = 2Dt^\alpha$, they found different behaviour of molecular spiders as a function of time. Their fundings suggest there is a sub-diffusive initial regime ($\alpha < 1$) in which times are short enough that the average number of cleavages per spider is less than 1. As time passes and spiders have cleaved a small region into products and defined a boundary between products and substrates, there is an effective outward bias and spiders are in the super-diffusive regime ($\alpha > 1$). Eventually spiders move to the diffusive regime in which they move with $<x^2(t)> \propto t$ ($\alpha = 1$).

Based on this finding they categorized the spiders into two states: D (diffusive state on all-product regions) and B (boundary state between product and substrate regions) and discussed the transient behaviour of spiders based on these two metastates. They concluded that if the super-diffusive state lasts sufficiently long, these spiders can perform useful work in finite-time experiments. They also showed that spiders with smallest $r$ ($r=0.005$), and hence largest difference between substrate and product interactions, show the strongest bias and go significantly farther on the track.

More recently, this biased motion of spiders by decreasing $r$ has been studied by Antal and Krapivsky for spiders on a 2D plane using both diffusion and first passage time analysis [104]. They found a smaller rate of hopping, $r$, to lead to larger diffusion constants and to increase the average number of visited sites. Using simulations, they also studied the effect of number of legs and maximum span on the diffusion coefficient and number of visited sites for spiders without memory ($r = 1$). They found that, generally, increasing the number of legs makes spiders less motile while increasing the maximum span results in an increase in diffusion coefficient. Unique among these theoretical studies, they considered spiders capable of detaching completely from substrate, and found analytically that the time until spiders detach grows exponentially with the number of legs.

While providing insight into some of the mechanisms responsible for spider motility, these theoretical treatments have not considered other aspects of the spider mechanism that may be important for understanding the operational principles of these as motors. For example, detailed kinetics of substrate binding, cleavage, and product release are only implicitly contained in the choice of unbinding probabilities, but since these are values that can be manipulated experimentally [86], a detailed kinetic model of the spider cycle could provide a framework for optimizing desired aspects of motor performance. Also, except for a brief study of detachment of spiders without memory from their track [104], spiders were treated as infinitely processive, i.e., the models required at least
one leg of the spider to be bound at all times. The loss of spiders from the matrix was observed experimentally [86, 87], implying that considerations of processivity are necessary to fully characterize performance. Finally, the ability of the spider to undergo biased motility in the absence of load implies that it should be able to do work against an applied load. For synthetic molecular motors to be useful transport devices, this ability to perform useful work is critical, making it important to quantify stall force. To date, the thermodynamic efficiency of no synthetic walker has been experimentally measured, although for one design it has been theoretically estimated [85].

Stochastic effects arising from competing chemical kinetic pathways are not easily incorporated in an analytical framework, hence, I used Monte Carlo (MC) simulations with experimentally derived rate constants to study the origin of the observed biased motion and characterize the expected motor properties of molecular spiders. In Section 2.2, I first explain the details of my kinetic model, the transition rate constants and the computational method used to simulate the trajectories of molecular spiders. Using the simulated trajectories of spiders on 1D tracks, I first analyze the bipedal spiders’ performance as motors by characterizing properties such as bias (directionality), speed, randomness parameter, processivity, mechanochemical coupling, and efficiency (via estimation of the stall force). I compare these results for the two simplest mechanisms of bipedal motility, namely, inchworm (IW) and hand-over-hand (HOH), to determine how these two types of stepping mechanism contribute to motor performance. While the experiments that stimulated my work examined spiders with two and more legs, initially, I limited my study to bipedal spiders to allow investigation of how stepping mechanism impacts motor performance. I also determined how motor characteristics of bipedal spiders change when the rate of binding or cleavage is altered. I conclude Section 2.2 by explaining how my results shed light on the mechanism responsible for biased motion of molecular spiders and make predictions of their ability to perform work against an externally applied load.

In Section 2.3, I investigate the effect of experimentally tunable parameters on motor properties of multipedal molecular spiders. As shown experimentally, the performance of spiders, in particular the time they stay attached to a track, can be improved by increasing the number of legs and by optimizing relevant design parameters. Because parameter space is easier to explore with simulations, I apply my model to characterize how experimental designable parameters, such as number of spider legs, length of the legs (which relates to the maximum span between binding sites) and kinetics of leg release from a substrate or product site, can be tuned to optimize aspects of motor performance, such as velocity, processivity and efficiency. In addition, to study thermodynamic efficiency, a pa-
rameter not investigated experimentally, I introduce a generalized expression that is a function of both force and time, and comment on the need for this interpretation. It should be noted that, except for the case of "quick spiders" [101], all of the aforementioned theoretical studies considered the nearest-neighbor stepping of spider legs, while in my model of Section 2.3, a freely diffusing spider leg is allowed to step over other bound legs.

Finally to conclude this chapter, in Section 2.4, I summarize my results of the two sections, compare the efficiency of molecular spiders with biological molecular motors and explain how the differences in the mechanism of their biased motions might lead to their considerable differences in efficiency.

The methods, results and discussions of Section 2.2 for bipedal spiders and Section 2.3 for multipedal spiders were adapted from our publications [105] and [106], respectively.

2.2 Studying the Biased Motion of HOH and IW Bipedal Spiders

In order to study the dynamics of bipedal spiders, I established a simple kinetic model for the interaction of a deoxyribozyme leg with its substrate based on previous studies of (deoxy)ribozyme kinetics [107, 108]. Here, I explain the details of my model and computational method of studying the biased motion of bipedal spiders. Then, I define the studied properties of bipedal spiders and how I calculated each parameter. I present my results and compare the investigated properties for HOH and IW stepping of bipedal spiders.

2.2.1 Modeling

Kinetic Model

Figure 2.2 shows the possible kinetic pathways for a deoxyribozyme leg in its interactions with a substrate and a product. The possible states of each leg are described as follows. $E + S$ represents an unbound enzymatic leg ($E$) and a surface-bound substrate ($S$) (state 1). In $ES$, the substrate-bound state, an active enzyme-substrate complex forms through base-pairing of the leg and substrate (state 2). From $ES$, a leg can either cleave the substrate to form a complex of enzyme and products, $EP_1P_2$ (state 3), or dissociate to give an unbound leg and a substrate ($E + S$). From $EP_1P_2$, the complex can either religate to $ES$ (state 2) or proceed with dissociation of the leg from the products, $E + P_1 + P_2$ (state 4). Because $P_2$ diffuses into solution after its release, resulting in a negligible
concentration of free $P_2$ [86, 87], I assumed the dissociation of $P_2$ to be irreversible. After release of the leg from the products, the leg can rebind either to the remaining surface-bound product $EP_1$ (state 5), or to a new neighbor site, which could be either $S$ (state 2) or $P_1$ (state 5).

In my simulations, the transition rates between states were specified by the rate constants in Table 2.1. Product binding and unbinding rates $k_{on}$ and $k_{off,P}$ were experimentally determined and $k_c$ was estimated from values provided for the bipedal spider NICK-2.4A [86] (Appendix A.1). I assumed a leg to bind at the same rate to a substrate as to a product. I converted from the measured second-order $k_{on}$ in $M^{-1}s^{-1}$ to an effective first-order $k_{on}$ by using an estimated local concentration of a free leg when one spider leg was bound (Appendix A.1). The remaining rate constants were estimated from kinetics of the hammerhead HH10 ribozyme [107]. Besides the time scales for biochemical reactions, I must also consider the time scale for diffusion, which enables the spider to explore possible binding sites. By approximating the bipedal NICK-2.4A structure as a sphere of radius 6 nm, I estimated its translational diffusion constant in water. I found that the spider can diffuse over an average distance of 9 nm between adjacent binding sites [86] orders of magnitude faster ($\sim 10^{-6}$ s) than it can bind to substrate (0.05 s) (Appendix A.1). Thus, I ignored the time scale of diffusion in my simulations and, in the absence of applied force, I assumed that a spider leg can bind with equal probability to all allowed binding sites.
CHAPTER 2. MOLECULAR SPIDERS

<table>
<thead>
<tr>
<th>Rate constant</th>
<th>Physical meaning</th>
<th>Value(s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{on}$</td>
<td>Substrate binding</td>
<td>$20^{a,b}$</td>
</tr>
<tr>
<td>$k_{off,S}$</td>
<td>Substrate unbinding</td>
<td>$0.035^{a,c}$</td>
</tr>
<tr>
<td>$k_c$</td>
<td>Cleavage</td>
<td>$0.055^{a,d}$</td>
</tr>
<tr>
<td>$k_{-c}$</td>
<td>Religation</td>
<td>$0.0005^{a,d}$</td>
</tr>
<tr>
<td>$k_r$</td>
<td>Release of product</td>
<td>$0.046^a$</td>
</tr>
<tr>
<td>$k_{on}$</td>
<td>Product binding</td>
<td>$20^{a,b}$</td>
</tr>
<tr>
<td>$k_{off,P}$</td>
<td>Product unbinding</td>
<td>$0.14^e$</td>
</tr>
</tbody>
</table>

Table 2.1: Transition rate constants for interactions of a leg with a substrate or a product. $^a$Estimated from [86, 107]. $^b$In some of the simulations, $k_{on}$ is changed to 2.4 and 1.6 s$^{-1}$. $^c$In some of the simulations, $k_{off,S}$ is changed to 0.0035 s$^{-1}$. $^d$In some of the simulations, $k_c$ is changed to 0.0055 s$^{-1}$ and $k_{-c}$ is kept at $k_c / 100$. $^e$Reference [86].

1D Track

In my model, I replaced the ssDNA substrate track with a 1D lattice of 1000 sites, which allowed for a straightforward assessment of biased motility. The distance between neighboring sites was $\Delta x = 1$ unit of length, a dimensionless parameter in the simulations that corresponds to the estimated intersubstrate distance of 9 nm [86]. Here, I considered three different tracks, denoted $S$, $P$ and $P$-$S$. The $S$ ($P$) track represents a lattice of all substrate (product) sites while the $P$-$S$ track represents a 1D lattice where the left side of the track (sites 1-500) has only product sites and the right half of the track (501-1000) has only substrate sites. The $P$-$S$ track was chosen for ease of examining bias of spider motion.

Stepping Mechanisms

I modelled the bipedal spiders as two identical physically coupled legs moving on a 1D track. The spider legs were not allowed to occupy the same lattice site, i.e. they interacted via exclusion. Here, I considered separately two possible mechanisms of spider stepping, namely, inchworm (IW) and hand-over-hand (HOH) (Fig. 2.3), to determine how these two types of stepping mechanism contribute to motor performance. For the IW, binding of a leg was permitted either to its original site (resulting in a center-of-mass change of $\delta = 0$) or to its nearest neighbor ($\delta = 0.5$ or $-0.5$). In the HOH spider, rebinding of a leg to its original site was again permitted or it could tumble over the other leg and bind to the next accessible site ($\delta = 1$ or $-1$).
2.2.2 Computational Method

Kinetics

For a given track, an IW or HOH stepping mechanism, and the chemical kinetic model for each leg (Fig. 2.2), I performed MC simulations of spider trajectories. Each of my simulations started with one spider placed at the middle of a track, such that the initial site coordinates of the legs were taken to be (500, 501). In this situation, the initial biochemical states of the legs are (5, 5), (2, 2) and (5, 2) on \( P \), \( S \), and \( P-S \) tracks, respectively (see Fig. 2.2). The transitions between biochemical states of my kinetic model (Fig. 2.2) are Markovian stochastic processes which can be numerically simulated using the Gillespie algorithm [114] (Appendix B). At each step of the calculation, each of the legs was in one of five biochemical states, with possible kinetic transitions specified by my model. The Gillespie algorithm takes into account all the possible reactions \( R_i \) from a given state of the system with their specific rate constants \( k_i \). For example for a bipedal spider, when the spider legs were in states (5, 2), the first leg could only go to state 4 and the second leg could go to state 1 or 3, giving a total of three possible reactions \( R_i \). By using the Gillespie algorithm, my simulations determined at each step the outcome of this stochastic process: which of these possible reactions occurred and how long it took. The state of the system and the clock were updated and the Gillespie algorithm was applied to this new state to determine the next transition of the system.

When cleavage of a substrate occurred at a given site, the state of the site was changed from \( S \) to \( P \), introducing a memory effect into the dynamics of the spider. In my simulations, I made this site change after the irreversible product release step (rate constant \( k_r \)). To explore the role of memory in the bias of the spiders, in some of my simulations I suppressed this site change (i.e., the original state of the track was maintained), but otherwise retained the same kinetic scheme.
My model explicitly included binding and unbinding of each spider leg to the track, since a leg was unbound from the track when it was in state 1 or 4 (Fig. 2.2). If the Gillespie algorithm next selected an unbinding reaction for the other leg, the spider detached from the track. In most cases, I stopped the simulation at this point and considered that the spider diffused into solution. However, in some of my simulations of bipedal spiders in this section, I made the spiders infinitely processive by "holding" a detached spider at its last center-of-mass position until a leg rebound.

In order to analyze trajectories for a large ensemble of bipedal spiders, I recorded \(10^6\) spider trajectories for each set of parameters tested. For spiders that could detach from the track, a trajectory ended upon detachment. For infinitely processive spiders, however, the trajectory ended when a preset long-time limit was reached. At each kinetic step in all simulations, I recorded and updated the elapsed time, the biochemical state and physical coordinate of each leg, and the state of the track.

**Application of Force**

I studied the force-velocity relationship for the IW and HOH spiders by applying a force opposing the biased motion of the spiders on the track. In principle, force could act on any of the kinetic steps of binding, cleavage, release, and diffusion [115, 116, 117]. However, without knowledge of the force dependence of ribozyme kinetics, I made the simplest assumption that force acts only to bias the choice of binding site by an unbound leg (i.e., it acts on the diffusive, translocation step). Thus, if leg binding was the reaction chosen by the Gillespie algorithm, the choice of binding site was weighted by the force-dependent probability of binding to the forward (+) site over the rearward (-) site

\[
\frac{P_+}{P_-} = \exp\left(-\frac{F\delta}{k_BT}\right),
\]

(2.1)

where \(P_+ + P_- = 1\). Here, \(F\) is a rearward force representing a load applied to the spider and \(\delta\) represents the step size or separation between the center-of-mass positions of the spider in each of the two possible binding sites: \(\delta_{HOH} = 9\) nm and \(\delta_{IW} = 4.5\) nm. Note that here I used real, rather than dimensionless, values for the separations, so that forces were also obtained in real units. In the case of no external force, \(P_+/P_- = 1\), i.e., this reduces to equal probability of choosing one of the two allowed binding sites.
CHAPTER 2. MOLECULAR SPIDERS

Calculation of Bias and Randomness

The ensemble-averaged velocity \( \langle v \rangle \) of the spiders was calculated from the slope of \( \langle \Delta x_{cm} \rangle \) versus time, where \( \Delta x_{cm} \) was the distance of each spider’s center of mass from its starting position of \( x_{cm} = 500.5 \). The averages were calculated over all spiders remaining bound to their tracks at the specified time. The time window for the average velocity calculation started at 300 s, a time chosen to be far enough from the initial conditions that transient behavior was excluded, and ended at the latest time at which 1000 trajectories had a spider bound to the track. The time window thus changed for different track, stepping, and kinetic conditions.

In order to quantify the stochastic variability of biased motion, I calculated a dimensionless quantity known as the randomness parameter \( r \) given by

\[
 r = \frac{Var(x_{cm})}{\langle \Delta x_{cm} \rangle \delta},
\]

(2.2)

\( Var(x_{cm}) \) is the variance of spider positions at time \( t \) and is related to an effective diffusion constant by

\[
 Var(x_{cm}) = \langle \Delta x_{cm}^2 \rangle - \langle \Delta x_{cm} \rangle^2 = 2D_{eff}t.
\]

(2.3)

The randomness parameter has been derived and used in the analysis of motor transport by several authors [118, 119, 120, 121] and, apart from a factor of 2, its inverse is the Péclet number of hydrodynamics as defined in Ref. [122]. Equation (2.2) shows that the larger the value of \( r \), the more diffusion predominates over directed stepping.

2.2.3 Results and Discussion

Effect of Track Properties on Biased Motion

Sample trajectories of IW and HOH spiders are shown in Fig. 2.4 in order to provide an illustration of the motion of individual spiders on a \( P-S \) track.

The directional motion of a population of spiders was studied by ensemble averaging over \( 10^6 \) of these trajectories for IW and for HOH spiders. I calculated the probability \( P(x_{cm}) \) of finding the centre of mass of a spider at position \( x \) on the track at different times for both IW and HOH spiders. It is important to note that since detachment of spiders from the track could occur, at each time point \( P(x_{cm}) \) was calculated by averaging only over the spiders remaining bound to the track.
at that time. Figures 2.5(a)-(c) show the evolution of $P(x_{cm})$ with time for HOH spiders on $P$, $S$ and $P$-$S$ tracks, while Figures 2.5(d)-(f) show the same for IW spiders.

Spiders on a pure $P$ track exhibited symmetric Gaussian probability distributions about the centre ($x = 500.5$) of the track, which broaden as time increases (Figs. 2.5(a),(d)). This is a characteristic of unbiased diffusion and confirms that there is no biased motion for the spiders on a pure $P$ track, as one would expect.

In contrast to this purely diffusive motion, spiders on an $S$ track clearly exhibited outward bias from the centre of the track, as seen in Figs. 2.5(b) and (e). The $P(x_{cm})$ distributions were bimodal, due to the lack of initial asymmetry in the system: the initial cleavage happens randomly either with the left or the right leg of each spider. Modification of this (left or right) track site from $S$ to $P$ imposed an asymmetry on the system and spider motion then appeared to propagate outward in the direction of this initial cleavage site.

The initial symmetry of the track and subsequent symmetry of the $P$ states produced by the population of spiders as a function of time made it challenging to analyze the bias on a purely $S$ track. Thus, I introduced track asymmetry into my simulations to make it easier to observe and quantify the bias of spiders in response to different stepping and kinetic conditions. Evolution of $P(x_{cm})$ on a half-$P$ / half-$S$ ($P$-$S$) track (see Section 2.2.1) clearly exhibited the expected directional motion towards the $S$ side of the track, for both HOH spiders (Fig. 2.5(c)) and for IW spiders (Fig. 2.5(f)).
Figure 2.5: Evolution of position probability distributions $P(x_{cm})$ on (a, d) $P$ tracks, (b, e) $S$ tracks, (c, f) $P$-$S$ tracks for HOH (left column) and IW (right column) bipedal spiders. In all cases, the initial number of spider trajectories is $10^6$. $P(x_{cm})$ at each subsequent time is calculated only using trajectories that still have a spider bound to the track at that time. For clarity, HOH spider distributions are smoothed by combining probabilities of successive site occupation. This is necessary because the HOH spider spends significantly more time in states where both feet are bound to the track (half-integral $x_{cm}$) compared with states in which only one foot is bound (integral $x_{cm}$).
CHAPTER 2. MOLECULAR SPIDERS

Figure 2.6: Graph of $\langle \Delta x_{cm} \rangle$ versus time for IW and HOH spiders on $P$-$S$ tracks. $\langle \Delta x_{cm} \rangle$ is the ensemble-averaged center-of-mass displacement of spiders with respect to their initial position at the middle of the track. It is clear that a population of HOH spiders exhibits much greater speed than IW spiders.

I used this $P$-$S$ track for the rest of my studies, as it simplifies analysis of biased motion.

Mechanism of Biased Motion

There is a variety of possible mechanisms that may contribute to biased motion of the population of spiders: system memory effects introduced by modification of the track from $S$ to $P$; differences in dwell time on $S$ and $P$ sites; and preferential loss of spiders on the $P$ side of the track. In this section, I present and discuss the results of my simulations examining the contributions of these different mechanisms.

In Fig. 2.6 I show the average displacement of the center of mass, $\langle \Delta x_{cm} \rangle$, of IW and HOH spiders as a function of time. In my calculations, $\Delta x_{cm}$ was defined as the displacement of the spider’s center of mass from its initial position ($x_{cm} = 500.5$). The plots show that the HOH stepping mechanism leads to an approximately four times greater speed than the IW mechanism, as seen also in Figs. 2.5(c) and 2.5(f). About half of this difference can be ascribed to the factor-of-two difference in step sizes ($\delta_{IW} = 0.5$ and $\delta_{HOH} = 1$). This still leaves a factor of approximately two difference in speed, suggesting that the stepping mechanism has a role in determining the magnitude
Figure 2.7: The randomness parameter, $r$, is greater for IW than HOH spiders and increases with time for both types of spiders.

To examine the extent of stochastic broadening of the biased motion of each type of spider I calculated their randomness parameters $r$, using Eq. (2.2). I found that $r$ increases with time for both IW and HOH spiders, and that $r$ was greater for IW spiders than for HOH spiders (Fig. 2.7). For IW spiders, both $\langle \Delta x_{cm} \rangle / Var(x_{cm})$ and step size were smaller than for HOH spiders.

My kinetic model provided an asymmetry between $S$ and $P$ sites via the average dwell time of a leg on a substrate (25 s) versus product (7 s). The shorter dwell time of a leg on a $P$ versus $S$ site means that it is more probable (per unit time) that a leg dissociates from the $P$ side of the track. Thus, one must consider that part of the observed averaged biased motion could be the result of preferential detachment of spiders from the $P$ side of the track (see Fig. 2.8). To determine the influence of detachment of spiders from the track on the observed biased motion of the bound population, I simulated HOH spiders on a $P$-$S$ track following the same kinetic scheme as shown in Fig. 2.2 but that were kept from leaving the track by holding a detached spider at its last center-of-mass position until a leg rebinded. I called this case "spiders with cleavage and without detachment". It is important to note that this was a purely artificial construct, designed to probe the role of detachment, which was not achieved in experiments, where spiders had a finite binding time to the substrate matrix.
Figure 2.8: Percentage of trajectories with bound spiders as a function of time, for HOH spiders on S tracks, on P-S tracks, on P tracks and on uncleavable S tracks, and for IW spiders on P-S tracks. Spiders exhibit the longest dwell times on uncleavable S tracks and detach most rapidly from P tracks.

I saw that the average displacement of HOH spiders decreased when detachment was not permitted (Fig. 2.9(a)). However, disallowing detachment resulted in a significant increase in $r$ for HOH spiders (Fig. 2.10). This was predominantly due to their decreased speed, since the variance was almost identical for spiders with and without detachment (Fig. 2.9). These results indicated that detachment contributes significantly to the biased motion of the population, and thus, that the observed bias for spiders with cleavage and detachment is in part a result of spiders detaching from the track preferentially from the P side.

It should also be noted that in this case of cleavage without detachment, $r$ was again smaller for HOH spiders than for IW spiders (Fig. 2.10). Previous theoretical treatment has assumed track cleavage to play a major role in promoting biased motion [102]. In this model, cleavage of an S to a P site was deterministic, based solely on the binding of a spider leg (i.e., cleavage was 100% efficient). In such a case, the movement of the P-S boundary could be used as an indicator of spider population speed. In my kinetic model, by contrast, binding of a leg to an S site was not necessarily
Figure 2.9: (a) Graph of $<\Delta x_{cm}>$ versus time for different classes of HOH spiders and the predictions of the diffusive model described in Appendix C. Spiders capable of cleavage and detachment show the greatest average displacement per time. (b) A plot of $\log <\Delta x_{cm}>$ versus $\log(t)$ shows that the bias of spiders with neither cleavage nor detachment is diffusive, with a slope of this plot of 0.5. The other two classes of spiders show enhanced biased motion is seen. (c) Graph of variance versus time for different classes of HOH spiders show similar behaviour for HOH spiders with cleavage and with/without detachment.
followed by cleavage: detachment without cleavage was allowed. This made it less obvious how strongly track cleavage contributed to the population bias.

I addressed the question of how much track cleavage (i.e. non-Markovian or memory effects) contributed to the observed bias of the spider population by examining HOH spiders that could neither cleave nor detach. Here, the spiders again followed the kinetic scheme shown in Fig. 2.2 but they were not allowed to detach from or cleave substrates on the track. Therefore, the state of the track was not changed from substrate to product when a transition from state 3 to state 4 occurred, and spiders were forced to stay on the track, as above. It is worth noting that for both of these artificial spider models, with cleavage and without detachment and with neither cleavage nor detachment, it was important to keep the same kinetic scheme as the original spiders, so as not to change the timescale of the leg-track interaction.

Shown in Fig. 2.9(a) is a graph of $<\Delta x_{cm}>$ versus $t$ for the HOH spiders with neither cleavage nor detachment, demonstrating that the bias is further reduced but is still present when both cleavage and detachment were removed. These results support the hypothesis that track cleavage (spider "memory") improves the directionality of spider motion. It is, however, surprising that a population of spiders undergoing only effective diffusion on asymmetric tracks was capable of biased motion. What can be responsible for this? A third possible factor contributing to bias is the free energy preference for binding to $S$ versus $P$, manifest in my kinetic simulations by the leg dwell times on
Figure 2.11: Ratio of spiders on S to P side of the track for HOH spiders with neither cleavage nor detachment increases rapidly, then plateaus.

S and P sites. This hypothesis suggested that I should observe spiders with neither cleavage nor detachment to preferentially populate the S side of a P-S track. Indeed, my simulations showed that the ratio of spiders on S : P sides of the tracks increases from an initial value of 1 to a plateau of 6.4 after approximately 3000 seconds (Fig. 2.11). A statistical mechanical calculation, using partition functions to describe states with free energies of a bipedal spider binding zero, one or two legs to the S side versus the P side of the track, predicted a much larger ratio of 15. Clearly, this statistical argument, based on thermodynamics, overestimates the preference for S-site occupation compared with my simulations. It is, however, important to note that my artificially constrained spiders "without detachment" sample a reduced phase space compared with experimentally realizable spiders, and hence may not be able to perform a proper sampling of states required by this Boltzmann treatment.

Alternatively, I considered the contributions of effective diffusion, where the populations on P tracks or on uncleavable S tracks evolve with \( D_P > D_S \) due to the shorter dwell times on P sites (Fig. C.2). A heuristic argument can be made that the free-energy bias towards S-state occupation is counteracted by an effective diffusional drift towards the P side of the track. Such a population drift towards the P side is predicted in some treatments of state-dependent diffusion [122, 123],
which calculated $\langle \Delta x_{cm} \rangle$ to evolve with $\sqrt{t}$ towards the region of greater diffusion constant. In Appendix C, I describe a diffusive model that assumed a constant population ratio of $S : P$ spiders and simulation-derived diffusion constants $D_P$ and $D_S$. My model gave the result that $\langle \Delta x_{cm} \rangle \propto \sqrt{t}$, indicating an increasing displacement with time but with a bias to the $S$ side of the track, the region of lower effective diffusion constant, see Appendix C. The predicted $\langle \Delta x_{cm} \rangle$ from the model is plotted in Fig. 2.9(a) and shows excellent agreement with the observed bias of spiders from my simulations. The biased motion of this artificial model system of bipedal spiders with neither cleavage nor detachment was somewhat surprising, illustrating the ability of the Gillespie-based Monte Carlo approach to uncover dynamics that might not be apparent from separate treatments of thermodynamics and diffusion.

Like the other classes of spider, spiders with neither cleavage nor detachment exhibited an increasing randomness with time (Fig. 2.10). However, the removal of cleavage from spiders unable to detach resulted in a drastically reduced $r$. It is surprising that the least "random" of my spiders are those that have no memory and are capable only of effective diffusive motion. Although the velocity was lower for those spiders than others, their substantially reduced variance resulted in their smaller value of $r$ by Eq. (2.2) (data not shown). This is likely a result of increased population on the permanently $S$ side of the track (see above), where diffusion is much smaller than on $P$ sites.

In the remainder of this section, I describe the molecular motor properties of the bipedal spiders capable of both cleavage and detachment, since this is the most realistic of the models considered here.

**Processivity**

Next, I investigated the influence of the stepping mechanism on processivity for both IW and HOH spiders. By definition [124, 125], processivity of a motor is the number of catalytic cycles (cleavage events) before detachment from its track. High processivity is one of the desired features when designing a molecular motor.

By analogy with experiment [86], I first calculated the percentage of spiders remaining bound to their respective tracks as a function of time (Fig. 2.8). These results showed most rapid detachment from a $P$ track and slowest detachment from an uncleavable $S$ track (in agreement with experiments), and faster loss of IW than HOH spiders. To simulate an uncleavable $S$ track [86], I allowed only transitions between states (1) and (2) (Fig. 2.2). Experimentally, this was achieved by substi-
Figure 2.12: Probability distributions of the number of cleavages performed by each spider prior to its detachment from a track. HOH spiders exhibit greater processivity than IW spiders, and their processivity is significantly enhanced on S tracks. Lines through the points are guides to the eye.

I investigated processivity by calculating the probability distributions of the number of cleavage events for IW and HOH spiders before detachment from the track (Fig. 2.12). These results showed HOH spiders to be more processive than IW spiders. The longer binding time and increased processivity of HOH compared to IW spiders can be rationalized by analysing their respective stepping mechanisms. Consider the initial configuration of a spider at sites (500, 501) bound to sites (P, S). Due to the shorter dwell time, a leg bound to the P site is more likely to unbind than a leg bound to the S site. In the case of the IW spider, the available binding sites for the mobile leg are the nearest (500) and next-nearest neighbours (499) of the bound leg, i.e., both on the P side of the track. However, in the case of the HOH spider, the reachable binding sites are the nearest neighbours of the bound leg, in this case a P site (500) and an S site (502). The HOH spider, therefore, has an earlier opportunity to bind to a new S site, which increased its binding time to the track. Additionally, the increased accessibility of S sites to the HOH spiders increased the probability of a cleavage event.
In agreement with the proposed role of binding to $S$ in enhancing processivity, I found that HOH spiders performed significantly more cleavage events prior to detachment when they were initially bound in the centre of an all-$S$ track (Fig. 2.12). The much increased processivity in this case is the behaviour that would be expected of the experimental spiders, sparsely distributed within a substrate matrix [86].

**Mechanochemical Coupling**

Another key parameter of molecular motors is mechanochemical coupling, which indicates the correlation between mechanical steps and chemical cycles. In order to characterize the mechanochemical coupling of IW and HOH spiders I calculated the spatial correlation of sequential cleavages (Fig. 2.13). This calculation showed that 81% of IW spiders sequentially cleaved adjacent $S$ sites ($\Delta x = +1$), while only 34% of HOH spiders exhibited sequential cleavage of neighbouring $S$ sites in the direction of population bias. Thus the data showed a tight coupling between mechanical and chemical cycles of IW spiders but a weaker coupling for HOH spiders. From the same figure, it is clear that there was also a considerable probability of HOH spiders cleaving the rearward neighbour ($\Delta x = -1$) and next-nearest forward neighbour ($\Delta x = +2$), while these probabilities were very small for IW spiders. I conclude that the ability of HOH spiders to "tumble" along the track provided easier access to further sites, which while increasing their speed, reduced their mechanochemical coupling.

Having tightly coupled mechanochemistry serves two important roles in the function of spiders as molecular motors. First, tighter coupling between cleavage and motion results in more efficient use of the fuel for stepping, with fewer unproductive cleavage cycles that do not couple to motion. Importantly, tighter coupling also means that bound substrates are more likely to be cleaved, which in turn reduces the likelihood that patches of uncleaved substrate could remain behind the spider to act as backward "traps" to the desired forward bias. This relationship between cleavage probability and mechanochemical coupling was further illustrated by the reduced coupling found when the cleavage rate was lowered by an order of magnitude in our simulations. (Experimentally, this could be achieved by lowering the concentration of $Zn^{2+}$, a necessary cofactor for cleavage.) In this case of $k_c = 0.0055 s^{-1}$, I found from my simulations that the percentage of spiders that sequentially cleaved adjacent sites was reduced to 38% and 21% for IW and HOH spiders, respectively. This example illustrates the role that individual, experimentally tunable rate constants play in optimizing
specific aspects of motor performance. I explore this point in Section 2.3.2 further.

**External Force**

Both to gauge the spiders’ performance relative to their biological counterparts, and to evaluate their possible work output as synthetic motors, I investigated the force-velocity relationship and determined the stall force (the force at which the average velocity drops to zero) for both IW and HOH spiders. To do this, I incorporated the typical intersubstrate distance of experiments, $\Delta x = 9$ nm [86], to determine velocity (using Eq. (2.1)). For each applied force used in my simulations, I calculated the spiders’ average velocity from the slope of $<\Delta x_{cm}>$ versus $t$. The resultant plots of velocity versus force for IW and HOH spiders are shown in Fig. 2.14(a). For both types of spider, the velocity decreased in response to applied force before reaching stall conditions at approximately 0.045 pN for both IW and HOH. For forces greater than the stall force the spiders preferentially walked backwards towards the $P$ side of the track.

Figure 2.14 shows that the magnitude of the slope of $v$ versus $F$ for HOH spiders is greater than for IW spiders. This is consistent with the effect of a given load to increase the rearward bias for
HOH spiders ($\delta_{HOH} = 9 \text{ nm}$) compared with IW spiders ($\delta_{IW} = 4.5 \text{ nm}$) (see Eq. (2.1)). However, the velocity of HOH spiders at $F = 0$ is larger than for IW spiders. Hence it is reasonable to expect that the graphs of $v$ versus $F$ for IW and HOH spiders should intersect as observed in Fig. 2.14. Where these curves intersect, however, as well as which stepping mechanism should give a larger stall force, depends on the specific biochemical kinetics and cannot be derived analytically; rather these types of issues can be addressed only through simulations such as the ones reported here.

These values of the stall force are much lower than for biological motors such as kinesin [44]. The stall force of the spider can be increased, as seen in Fig. 2.14(b) for simulations in which I have decreased the effective binding rate by an order of magnitude, from $k_{on} = 20 \text{ s}^{-1}$ to $k_{on} = 2.4 \text{ s}^{-1}$. The larger stall force might be due predominantly to the increased average speed at zero force for $k_{on} = 2.4 \text{ s}^{-1}$ (Fig. 2.15), which arose from enhanced detachment of spiders from the $P$ side of the track.

The thermodynamic efficiency of these spiders was estimated from the free energy bias of the reaction and the applied force. The general definition of the thermodynamic efficiency for a molecular motor, $\eta_{th}$, is the ratio of the work done by the motor against a conservative external force, $F$, to its input energy, $\Delta G$, at each step of motor.
I first considered the case where substrate cleavage provides the free energy to bias motility. As a value for the free energy change, I took that of phosphate bond hydrolysis of ATP under standard conditions, \( \Delta G^0 = -30 \text{ kJ/mol} \) [126]. This is a slightly greater value than the free energy change due to the hydrolysis of a DNA phosphodiester bond (\( \Delta G^0 = -22 \text{ kJ/mol} \) [127]). Because of the very low local concentration of \( P_2 \) in experiments, the magnitude of the Gibbs free energy change of \( S \rightarrow P_1 + P_2 \) would be much greater than \( \Delta G^0 \). In fact, in the limit of \( [P_2] \rightarrow 0 \) (my simulations), the reaction was irreversible and no free energy change could be defined. For the sake of argument, I considered \( K_{eq} = 1 \), for which the thermodynamic efficiency at stall is \( \eta_{th} = F_{stall} \delta / |\Delta G^0| \). Here \( F_{stall} \) was the stall force determined for each motor (Fig. 2.14) with a step size of \( \delta \). For the HOH spider, this gave \( \eta_{th,HOH} = 0.08 \) while for the IW spider, \( \eta_{th,IW} = 0.04 \). It is clear that the
spider does not efficiently convert free energy released from substrate cleavage into useful work. It should be noted that even with the smaller free energy released from cleavage of a phosphodiester bond, the thermodynamic efficiency is still very small for these spiders ($\eta_{th,HOH} = 0.1$ and $\eta_{th,IW} = 0.05$).

Additionally, I considered the bias for spider motility to arise from the free energy preference to bind $S$ versus $P$ and I then determined the force required to remove this bias. This treatment follows similar reasoning used in reference [85], although in my kinetic model, a leg binds to $S$ and to $P$ sites at the same rate, while dissociating more rapidly from a $P$ site. To determine the free energy difference between binding to $S$ and to $P$ (i.e., between states (2) and (5) in Fig. 2.2), I used the second-order binding rate constant $k_{on}$ [86] and the first-order unbinding rate constants $k_{off,S}$ and $k_{off,P}$ (Table 2.1) to calculate $K_{eq}$ for leg binding to $S$ and to $P$, from which $\Delta G_{P-S}$ was calculated. This resulted in a free energy preference for a leg to bind a forward $S$ site over a rearward $P$ site of $|\Delta G_{P-S}| = 3.5$ kJ/mol. The elimination of this energetic preference for leg binding would require the application of a force against this bias: $F_{stall,P-S} = |\Delta G_{P-S}|/\delta$. For the two types of stepping mechanism considered here, this argument suggested $F_{stall,P-S,HOH} = 0.6$ pN ($\eta_{P-S,HOH} = 0.17$) and $F_{stall,P-S,IW} = 1.2$ pN ($\eta_{P-S,IW} = 0.12$). Both this treatment and the one presented in the previous paragraph suggested that the HOH spider is more efficient than the IW spider; again, it was clear that the spider is not efficient at transducing an energetic bias into mechanical work. It is important to note that the efficiency of a molecular motor and its mechanochemical coupling are independent parameters and are not necessarily related [1], so although the IW spider has high coupling (Fig. 2.13), it is not necessary that this correlate with a high efficiency.

Consideration of the possible leg configurations showed how the above calculations placed only upper limits on the stall force for the spider. The calculations assumed that an unbound leg has the choice between binding a forward $S$ site or a rearward $P$ site. If $S$-site cleavage correlated 100% with leg binding, an assumption made in a previous theoretical treatment [102], the track would retain a distinct boundary between $P$ sites and $S$ sites that would move progressively forward with time and whose movement would relate to spider speed. My kinetic model, by contrast, allowed for leg dissociation from an $S$ site without cleavage (transition represented by $k_{off,S}$ in Fig. 2.2), making it possible for the spider to step forward without cleaving $S$ from the initially imposed $P-S$ boundary, to have its legs bound later to $(S,S)$ sites and, upon further cleavage, leave an “island” of $S$ sites behind it. Thus, choices of binding possibilities for a leg are not only between a rearward
CHAPTER 2. MOLECULAR SPIDERS

$P$ and forward $S$ \((P, S)\) as considered in the previous paragraph, but also \((S, P), (S, S)\) and even \((P, P)\), for spiders that have stepped diffusively into the product zone. The symmetric binding choices \((S, S)\) and \((P, P)\) are stalled at zero force, while the choice of \((S, P)\) sites generate reverse bias even in the absence of load. The treatment of the previous paragraph, which considered stalling the spider from a \((P, S)\) choice, thus provided an upper limit to the maximum stall force possible for my kinetic model.

It is surprising to observe that HOH and IW spiders exhibited roughly the same stall force, in spite of their different step sizes. HOH versus IW behaviour has previously been theoretically investigated for a different model system, and there, HOH and IW stepping gave rise to the same stall force [128]. However, the model assumed the same binding site separation for both types of motion, whereas here, the difference in step size would lead one naïvely to expect that the stall force for HOH should be a factor of two lower than for IW. Why is this not the case? If I consider the idealized case of a spider always located at a $P$-$S$ boundary, I will see that detachment of the trailing foot from a $P$ site is more probable than detachment from an $S$ site. For an HOH spider, rebinding of this foot is forward-biased [choice of \((P, S)\) binding sites], while for an IW spider, rebinding of this foot is not biased in either direction [choice of \((P, P)\) binding sites]. In this idealized scenario, therefore, force must work against the forward bias of both feet of an HOH spider, while it needed to work only against the forward bias of the leading foot of an IW spider. This provides a possible rationale for the surprising observation of equal stall forces for both stepping mechanisms in our work.

2.2.4 Summary and Concluding Remarks

In this study, I introduced a kinetic model to investigate the biased motion of a population of bipedal spiders. Using realistic experimental rate constants, I determined that spider bias arises from three contributing factors: (1) preferential occupation of substrates compared with products, due to the free energy difference in leg binding; (2) faster dissociation from products, enhancing the ratio of legs bound to substrate versus to product; and (3) cleavage of substrate, which converts substrate to the weaker leg-binding product.

Our investigations of speed, processivity, mechanochemical coupling and stall force (efficiency) revealed that bipedal spiders act as molecular motors, albeit only weakly. These characteristics were studied for a large population of spiders, and it is clear from trajectories of single spiders
that each individual may not exhibit directed motion and hence would not be useful as a practical molecular motor. Thus, single-molecule studies of bipedal spiders would require averaging over many observed trajectories to ascertain their performance and directional bias.

I studied two distinct stepping mechanisms in order to better understand their individual contributions to motor performance. Comparing these two stepping mechanisms, I found that the increased accessibility to new substrate sites for HOH spiders appears to be responsible for their greater speed and processivity, while the greater coupling and efficiency found for IW spiders suggest they may use the trailing leg as a clamp to maintain the spider in better register with the $P-S$ boundary, or "Boundary" state, when compared with HOH spiders that can "tumble" promiscuously past this boundary into a "Diffusive" state [103].

My calculations for the randomness parameter, $r$, of spiders on a $P-S$ track showed that in all cases $r$ is considerably larger than one, implying that the spider motor is more diffusive than directional. In addition, $r$ increases as a function of time mostly due to the result that the gradient of $\langle \Delta x_{cm} \rangle$ decreases as a function of time, though in some cases the increase in $r$ is not very significant. This increase in the randomness parameter is in agreement with the observed asymptotic diffusive behaviour in other theoretical work on molecular spiders [101, 102, 103]. Furthermore, the value of $r$ was found to be higher for IW spiders than for HOH spiders by a factor of 2-3, a factor similar to their difference in step size. In contrast, previous studies of these two types of stepping mechanism found a value of $r$ close to one, which was lower for IW motors ($r \sim 0.7$) than for HOH motors ($r \sim 0.85$), indicating a balance between diffusion and directional bias [128]. However, the stepping probabilities used in that work strongly forward-biased their motors. It is therefore not surprising that they found values of $r$ which are considerably smaller than our values. It would also be of interest to calculate $r$ as a function of load force for our model [118, 128].

The experiments on molecular spiders were performed on an initially symmetric all-substrate surface [86], while I have done most of my simulations on a $P-S$ track. This initially imposed asymmetry helped to simplify my analysis of biased motion. However, my studies of binding time and processivity have also examined performance on a symmetric $S$ track (see Section 2.2.3), and clearly showed the influence of track construction on these parameters.

My simulation results demonstrated that the bipedal spider performs far below the standards set by biological motors such as conventional kinesin. Kinesin 1 is well known as a processive molecular motor [129], which in vitro can undergo $\sim 150$ HOH steps of 8 nm before detaching from a microtubule [130, 131]. The processivity of IW and HOH spiders on an all-$S$ track is an order
of magnitude lower (Fig. 2.12). Kinesin furthermore exhibits tight coupling between its mechanical stepping and chemical cycles, such that its velocity is proportional to the rate of hydrolysis with \textit{in vitro} speeds of \( v \approx 1 \, \mu\text{m/s} \) \[3\]. The combination of much faster hydrolysis and more efficient coupling of this chemical reaction to forward stepping explains why the speed of kinesin is orders of magnitude greater than HOH and IW spiders, which have roughly the same step size. HOH and IW spiders exhibited weak coupling between cleavage and directional motion, and cleavage could frequently be followed by backwards steps, particularly for HOH stepping.

The experimental spider is not limited to an IW or HOH stepping mechanism but is likely to be constrained in leg binding only by the maximum span between legs, \( S \) \[101\]. For \( S = 1 \), only the HOH stepping mechanism was allowed. For \( S = 2 \), motion can consist of IW and HOH stepping, along with larger HOH-type steps that include step sizes of \( \delta = 2 \). In their experimental study of molecular spiders, Pei \textit{et al.} sought to increase processivity in two ways: (1) by increasing the binding strength of legs to substrate and product; and (2) by increasing the number of spider legs \[86\]. They quoted few results for the bipedal spider, however, I can see by comparison of detachment rates of two-legged spiders in my model (Fig. 2.8 for spider on an uncleavable \( S \) track) with their reported loss of four- and six-legged spiders from the matrix (Fig. 2.16; reproduced from Ref. \[86\]) that the shape of my spider loss curve is similar to the experimental results and that the timescale of detachment from my model is shorter than reported for four-legged spiders, as expected \[86\]. Experimental results suggested that adding more legs to the spiders improves processivity but decreases their speed \[86\].

In order to study a more realistic model of spiders, in the next section, I describe a modified stepping mechanism in which all steps possible steps are allowed, subject to a maximum span between legs. I investigated the effect of experimental parameters on the biased motion of these molecular spiders.

### 2.3 Effect of Tuneable Experimental Parameters

Experimentally, it is possible to tune design parameters of molecular spiders such as number of legs, length of each leg and chemical kinetics \[86\]. I studied through simulations the effects of these same parameters on several motor properties of molecular spiders: velocity, processivity, mechanochemical coupling and thermodynamic efficiency, and probed their force dependence.
2.3.1 Methods

Simulating Spider Dynamics

Here, I describe the biased motion of multipedal spiders with \( n \) legs \( (n = 2, 3, 4) \), Fig. 2.17, by modelling their dynamics on a 1D track using MC simulations. The \( n \)-legged spider was simulated as \( n \) identical physically coupled legs moving on an asymmetric (P-S) 1D track (Section 2.2.1). As the spider stepped on the track, it chemically interacted with the sites on the track, via binding, cleavage and release of a substrate site, and via binding and release of a product site as explained in Section 2.2.1 (Fig. 2.2).

As in the previous section, spider legs were independent but here were not restricted to either an IW or HOH stepping mechanism. Instead, legs were governed only by two constraints: (1) Only one spider leg can occupy a given site; and (2) All legs of the spider are constrained to remain within a fixed maximum span \( S \) (Fig. 2.17). I used the introduced kinetic model (Fig. 2.2) for
CHAPTER 2. MOLECULAR SPIDERS

Figure 2.17: Initial configurations of bipedal, tripedal and quadrupedal spiders on a 1D product-substrate track of $10^5$ sites. Open circles indicate product sites, while substrate sites are denoted by closed circles. All three spiders shown here have a maximum span of $S=3$.

the interaction of a spider’s deoxyribozyme leg with its substrate and implemented MC simulations using the Gillespie algorithm as before (see Section 2.2.2) to simulate the kinetic model numerically. In contrast with my previous study of bipedal spiders (Section 2.2), here, I counted leg binding to each possible site as a separate chemical transition having the same transition rate $k_{on}$. For example, if there were two possible sites available for binding of an unbound leg, the Gillespie algorithm took into account two binding reactions for this leg. In my previous model, I considered binding of an unbound leg to all possible sites as only one chemical event, then used a random number to choose one of the possible sites for the leg to bind. The new method captured the feature that an increased number of possible binding sites should increase the chance of an unbound leg to bind to any site.

Using this improved Gillespie-based MC approach for simulating dynamics of spiders on a 1D $P$-$S$ track, $10^6$ trajectories were recorded for each type of spider (varying number of legs, $n$, and span, $S$) and analyzed to study its motor properties. I recorded the center-of-mass displacement of each spider as a function of time, $x_{cm}(t)$, from its initial position at the middle of the track at $t = 0$ (see Fig 2.17). A trajectory ended upon detachment of the spider from the track or when a pre-set time limit was reached (for quadrupedal spiders which were found to be highly processive). At
each kinetic step in all simulations, I also updated and recorded the biochemical state and physical coordinate of each leg and the state of the track.

**Track**

The track used here was a 1D lattice of 100,000 sites. This track was sufficiently long that no spiders in my simulations reached its ends \( i.e. \), it can be considered infinitely long on the timescale of my simulations. At the start of the simulations, the left side of the track (sites 1-50,000) consisted of product sites and the right half of the track (50,001-100,000) consisted of substrate sites. My study of bipedal spiders showed that this initial asymmetry of the track makes it easier to observe and quantify the biased motion of spiders. Again, the distance between neighboring sites on the track was \( \Delta x = 1 \) unit of length, a dimensionless parameter in the simulations that corresponds to the estimated experimental intersubstrate distance of 9 nm [86].

**Spider Efficiency**

In the single-molecule experimental study of molecular spiders, the authors suggested that molecular spiders are not as efficient and powerful as naturally occurring protein-based walkers, though they measured neither the force-dependent motion nor output power of the spiders [87]. By employing force-dependent kinetics, my previous study of bipedal spiders provided an estimate for their very low efficiency (Section 2.2.3). There have, however, been no quantitative studies of the efficiencies of multipedal molecular spiders and their dependence on the number of spider legs.

The general definition of the thermodynamic efficiency for a molecular motor, \( \eta_{th} \), is the ratio of the work done by the motor against a conservative external force, \( F \), to its input energy, \( \Delta G \), at each step of motor (Eq. (2.4)). For molecular spiders, the input energy is provided by the cleavage of substrate on the track. To study the efficiency, here, I applied a range of conservative forces opposite to the direction of the observed bias and calculated the work done by the molecular spiders at each value of force. As before, I made the simplest assumption that force acts to change the binding rate of an unbound leg to each site [132] by an amount dependent on \( \delta \), the change in the spider’s center-of-mass position before and after leg binding to that specific site (\( i.e. \), it acts on the diffusive, translocation step):

\[
k_{on}(F) = k_{on}(F = 0) \exp\left(\frac{-F\delta}{k_B T}\right).
\] (2.5)
Here, $k_{on}(F = 0)$ is 20 s$^{-1}$ (Table 2.1), $F$ is a rearward force representing a load applied to the spider, $k_B$ is Boltzmann’s constant and $T = 298 K$ is the absolute temperature. The $\delta$-dependence resulted in different steps from a given leg configuration occurring with different probabilities, i.e., all possible steps were no longer equally probable in the presence of an applied force.

Because of the possibility of different step sizes, $\delta$ was not uniquely defined for each multipedal spider. This precluded calculation of the randomness parameter as performed for bipedal spiders as in Section 2.2.3. In my calculations of efficiency, I therefore replaced step size, $\delta$, with the average center-of-mass displacement of the spiders:

$$<\Delta x_{cm}(F, t)>_{\Delta t} = <x_{cm}(F, t + \Delta t) - x_{cm}(F, t)>,$$

over a time interval of $\Delta t$ for each applied force, $F$. This is equivalent to previous analyses of kinesin which included back-steps when calculating efficiency [132, 133]. Furthermore, there was not necessarily a one-to-one correspondence between cleavage and stepping. Thus, in order to characterize the coupling between cleavage of substrate sites and stepping of the molecular spiders, I also determined the average number of substrate cleavages per spider, $<N(F, t)>_{\Delta t}$ in the same time interval. It was important to determine both $<\Delta x_{cm}>$ and $<N>$ in this way, because both of these parameters may in principle be time- and force-dependent. Then, for each type of spider under an applied force, $F$, and for each time interval of $\Delta t$, I calculated the mechanochemical coupling, $\xi$, as the ratio of average center-of-mass displacement, $<\Delta x_{cm}(F, t)>_{\Delta t}$, to average number of cleavages per spider, $<N(F, t)>_{\Delta t}$:

$$\xi_{\Delta t}(F, t) = \frac{<\Delta x_{cm}(F, t)>_{\Delta t}}{<N(F, t)>_{\Delta t}}.$$  \hspace{1cm} (2.6)

Note that this is a different definition of coupling than I used in Section 2.2.3.

I then calculated the thermodynamic efficiency of the molecular spiders based on the following definition:

$$\eta_{\Delta t}(F, t) = \frac{F <\Delta x_{cm}(F, t)>_{\Delta t}}{\Delta G <N(F, t)>_{\Delta t}} = \frac{F}{\Delta G} \xi_{\Delta t}(F, t).$$ \hspace{1cm} (2.7)

Again, I assumed that the free energy of cleavage of a phosphate bond in the ssDNA substrate is equivalent to the standard free energy of phosphate bond hydrolysis of ATP under standard conditions, $\Delta G^\circ = -30 \text{ kJmol}^{-1}$ [126], so that the total free energy change in the time interval $\Delta t$ is equal to $\Delta G^\circ <N(F, t)>_{\Delta t}$. Note that Eq. (2.7) is equivalent to the definition of efficiency commonly used for cyclical engines, where one would take $\Delta t$ to be the time for a complete cycle. Because the spider is not a purely cyclical motor (alternative kinetic pathways can also lead to motion), $\Delta t$ is not uniquely defined by a natural time scale of the spider. Instead, below I fixed a value of $\Delta t$
=100 s in order to evaluate the efficiency over consecutive 100 s time intervals and explored how spiders’ efficiency evolves over time.

2.3.2 Results

Effect of Number of Legs

To study the effect of leg number on the properties of molecular spiders, I analyzed the simulated trajectories of spiders with different number of legs ($n = 2, 3, 4$) and with the same maximum span of $S = 3$. This limited the step size of these multipedal spiders to the same maximum. It is useful to keep in mind that I defined the parameter $S$ based on the spacing between the sites on the track, so that $S = 3$ spans 4 adjacent sites on the track (Fig. 2.17).

Figure 2.18(a) shows the fraction of spider trajectories with a bound spider on the track as a function of time for bipedal, tripedal and quadrupedal spiders. As the number of legs was decreased, the spiders detached from the track more rapidly so that for example only half of bipedal spiders remained bound at 50 minutes. The increase in track binding time with number of legs can be understood in two ways: (1) When a spider was bound to the track by just one leg (the configuration of the spider before complete detachment) having more legs increased the probability of a rebinding event occurring rather than an unbinding event. (2) Spiders with more legs had a greater number of bound configurations, which reduced the probability of being in a single-leg-bound configuration.

For an $n$-legged spider with a maximum span of $S$ between the legs, the number of bound configurations of spiders, $C(n, S)$ is:

$$C(n, S) = \sum_{n_b=1}^{n} \binom{S}{S - (n_b - 1)}$$ (2.8)

where $n_b$ is the number of legs of an $n$-legged spider bound to the track, $1 \leq n_b \leq n$. In this formula, a bound configuration is defined based on the relative positions of the bound legs to the track without considering the state of the site ($S$ or $P$) to which each leg is bound. For example, when just one leg of a multipedal spider was bound to the track ($n_b = 1$), this counted as one configuration whether it was bound to an $S$ or a $P$ site. For a bipedal spider with span $S = 3$, there are 3 possible configurations in which both legs can be bound to the track, with legs at sites $(i, i + 1)$, $(i, i + 2)$ and $(i, i + 3)$. For spiders with $n = 2, 3$ or 4, $C(n, 3)$ is 4, 7 or 8, respectively, while for each of these cases there is just one configuration with only one leg bound to the track.
Clearly the probability of a spider being in the single-leg-bound state, \( \frac{1}{C(n,S)} \), decreases with an increase in the number of legs.

I measured the processivity of multipedal spiders with \( S = 3 \) by plotting the distribution of the number of cleavages performed by each type of spider prior to its detachment from a \( P-S \) track. As shown in Fig. 2.18(b), the processivity of the spiders increased as the number of legs increased from 2 to 4. This figure also shows that the number of cleavages by bipedal spiders is more than an order of magnitude smaller than by the tripedal and quadrupedal spiders. It should be noted that quadrupedal spiders were exceptionally processive, which is consistent with almost negligible detachment from both product- and substrate-covered matrices in experimental studies of molecular spiders [86]. Thus, for these spiders, I stopped the simulations when a preset long-time limit (\( t = 10^7 \) seconds) was reached rather than when the spiders detached from the track. This time was longer than the binding time for any of the bipedal and tripedal spiders. Because of this
Figure 2.19: (a) Dependence of biased motion on number of legs shows that the mean velocity of spiders decreases with increasing number of legs. \( \langle x_{cm} \rangle \) is the ensemble-averaged center-of-mass displacement of spiders with respect to their initial position at the centre of the track. \( \langle x_{cm} \rangle \) is calculated every 100 s by averaging over only those spiders remaining bound to the track at that time. The inset shows an expanded view of the first 80 minutes. Error bars represent the standard deviation of spider positions, calculated every 50 min. (b) Average number of legs bound to the track for spiders with \( n = 2, 3 \) and 4 and \( S=3 \) (shown over the first \( t=200 \) min) indicates that this value remains constant at a value very close to the number of spider legs.

choice of cut-off time the probability distribution of cleavages was peaked towards a lower number of cleavages than it would be for a longer time limit (or for the true processivity of these quadrupedal spiders).

To study how the number of legs affects the mean velocity of spiders, I calculated the average displacement of the spider population, \( \langle x_{cm} \rangle \), as a function of time. Figure 2.19 shows \( \langle x_{cm} \rangle \) versus time for bipedal, tripedal and quadrupedal spiders. It is important to note that, as in Section 2.2, \( \langle x_{cm} \rangle \) was calculated at each time point by averaging over only the spiders remaining bound to the track at that time.

As shown in Fig. 2.19, the mean velocity of spiders decreased as the number of legs was increased, in particular when increasing the number of legs from 2 to 3. In Section 2.2.3, I showed that detachment of spiders from the track significantly affected the mean velocity of the bound spider population by preferentially eliminating spiders from the product side of the track (the 3D-loss pathway discussed below). I therefore attributed much of the difference in mean velocity between
Average number of bound legs

<table>
<thead>
<tr>
<th>( n )</th>
<th>Average number of bound legs</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3.97</td>
</tr>
<tr>
<td>3</td>
<td>2.99</td>
</tr>
<tr>
<td>2</td>
<td>1.99</td>
</tr>
</tbody>
</table>

Table 2.2: Average number of legs bound to the track for bound spiders with \( n=2, 3 \) and 4 and \( S=3 \) (average calculated over the first \( t = 800 \) min).

bipedal and tripedal spiders seen here to the much greater rate of loss of bipedal spiders from the track (Fig. 2.18(a)). The greater number of bound configurations for tripedal spiders, \( C(3, 3) = 7 \), compared to bipedal spiders, \( C(2, 3) = 4 \), also suggests a lower detachment probability of tripedal spiders, while the closer values of \( C(3, 3) = 7 \) and \( C(4, 3) = 8 \) for tripedal and quadrupedal spiders, can help to account for the smaller difference between the mean velocities of these spiders. Interestingly, my simulations showed that all types of spiders are predominantly found with all legs bound to the track, independent of time (Fig. 2.19(b) and Table 2.2).

In order to explain the decrease in the mean velocity of the tripedal and quadrupedal spiders, I examined the results of Fig. 2.19 by making use of the concept of an ideal spider motor. An ideal spider motor, first placed at the initial \( P-S \) boundary, would cleave substrates in such a way that it would move this boundary forward with its motion, remaining pinned to the well defined \( P-S \) junction with all product sites to its left and all substrate sites to its right. This ideal spider would retain at all times the asymmetry necessary to produce biased motion. However, the chemical kinetics of spiders, as used in my simulations, included the possibility of "nonideal" steps (e.g. release of \( S \) without cleavage; stepping over a site on the track), resulting in spiders later being found at a variety of possible interfaces: \( P-P \), \( S-S \), \( S-P \) and \( P-S \). The \( P-P \) interface, which results in diffusive stepping, was accessed when spiders stepped either to the product side of the track or to a newly generated product site in the substrate side of the track. Similarly diffusive behavior (though slower) occurred for spiders at an \( S-S \) interface. In contrast, for spiders with legs at an \( S-P \) or \( P-S \) interface, release of the \( P \) site was more probable than release of the \( S \) site, resulting in preferential backward or forward stepping, respectively. I therefore refer to the configuration of spiders at the \( P-S \) boundary as the "ideal configuration" for forward-biased motion and motor function.

My studies of bipedal spiders showed that the peak of the probability distribution of spider positions on a \( P-S \) track moves directionally towards the \( S \) side of the track and that the distribution
spreads with time (Fig. 2.5). The broadening of the probability distribution was consistent with
diffusive spreading, suggesting that spiders (a) move diffusively; and (b) are lost from the ideal \( P-S \)
boundary along the track. I termed this loss from the ideal \( P-S \) boundary, to states as described
in the previous paragraph, one-dimensional (1D) loss. This is akin to the loss of spiders from the
boundary (B) to the diffusive (D) state described by others [103]. An additional pathway by which
spiders can be lost from the ideal \( P-S \) boundary is by detachment of all legs, resulting in loss into
solution. I termed this pathway 3D loss. As found in my previous studies (Section 2.2.3), 3D loss
contributes substantially to bipedal spider dynamics, as does 1D loss. The results present here in
Fig. 2.18 suggested that 3D loss is not a significant pathway in the dynamics of spiders with more
legs. My hypothesis was therefore that the decrease in the mean velocity of the spiders shown in
Fig. 2.19 is due to the decrease in the number of spiders in ideal configurations as time increases, a
result of 1D loss of bipedal, tripedal and quadrupedal spiders and of 3D loss of bipedal spiders.

To illustrate the nonideality of multipedal spiders, and the contributions of 1D loss to their dy-
namics, I calculated run lengths and maximum and minimum positions reached by spiders. The
results are shown in Fig. 2.20, and demonstrate that having more legs allows the spiders the op-
portunity to travel further, on average, along their tracks before detachment (to 22, 400 and >1118
sites beyond the initial \( P-S \) boundary for \( n = 2, 3 \) and 4, respectively). The contributions from 1D
loss are exemplified in two ways: (1) In general, spiders venture into the product side of the track
\(<x_{cm}>_{\text{min}} < 0 \) in Fig. 2.20); and (2) In general, spiders reach much further onto the \( S \) side than
their detachment positions, demonstrating significant back-stepping and 1D loss from the farthest
attained cleavage site (contrast their maximum and detachment positions). The 1D loss channel is
significant in their dynamics, irrespective of the number of legs. The stochastic nature of spider
dynamics was clearly apparent in the large standard deviations of maximum, minimum and detach-
ment positions, illustrating why I stress average spider properties in this work.

I now comment on my calculated results in the light of experiments. Calculated from the data in
Fig. 2.19, the approximately constant slope of \(<x_{cm}>\) versus time for bipedal spiders corresponds
to a mean velocity of \(~ 4 \text{ nm/min} \). The mean velocity of the tripedal and quadrupedal spider popu-
lations decreased with time but started at a similar value as the bipedal spiders (inset of Fig. 2.19).
Experimentally, the speed of tripedal spiders was measured to be between 1-6 \text{ nm/min} on a 90-nm-
long quasi-1D track of substrate [87], comparable to the initial mean velocity of my tripedal spiders
over these distances . Finally, a comparison of the results of Fig. 2.20 with experiments showed that
the calculated average run length of \(~3.6 \mu \text{m} \) for tripedal spiders is very much larger than the 90
Figure 2.20: Average minimum (blue), detachment (orange) and maximum (white) center-of-mass positions of each type of spider, determined from all sites sampled prior to detachment (n=2, 3) or prior to a preset long time limit (t=10^7 seconds; n=4). Error bars represent the standard deviations of positions sampled by each population of spider. Zero on the $<x_{cm}>$ axis represents the initial position of the spiders at the middle of the track, and negative and positive values on the axis represent the initial $P$ and $S$ sides of the $P$-$S$ track, respectively. The results indicate that run length increases with the number of legs, that spiders travel over a much wider expanse of the track than simply from their initial position to their detachment positions, and that there is significant dispersion in dynamics among different spiders of the population.

Effect of Maximum Span

Another adjustable parameter is the maximum span, $S$, of the legs, which is experimentally achievable by using linkers of different lengths between the streptavidin body of the spider and its catalytic DNA legs. I studied the effect of this parameter on the performance of molecular spiders by simulating quadrupedal spiders for different maximum spans of $S=3, 4$ and $5$. Figure 2.21(a) shows that spiders with greater span remained bound to the track for longer times and that all of these spiders remained bound for many orders of magnitude longer than experimentally relevant timescales. By decreasing the binding rate from $k_{on} = 20$ s$^{-1}$ to $k_{on} = 1.6$ s$^{-1}$, I was able to reduce this binding time by orders of magnitude (to $t_{1/2} \approx 800$ minutes for quadrupedal spiders with $S = 3$) but quadrupedal spiders still remained bound to the track for exceptionally long times.
Figure 2.21: (a) Fraction of spider trajectories with a spider bound to the track as a function of time showing that the binding time of spiders to the track increases with span. Values were calculated at each 100 s time interval. (b) Probability distributions of the number of cleavages performed by $t = 5 \times 10^6$ seconds by quadrupedal spiders with $S = 3, 4$ and 5. Spiders with larger span cleave more substrates in this time period. Lines through the points are guides to the eyes.

Increasing the span of the legs provides access to more sites on the track for each unbound spider leg. This increases the probability of a rebinding event for the unbound legs and results in a decrease in the overall detachment rate of spiders. My studies also show an increased number of cleavages for larger values of the maximum span but overall, the changes induced by larger span are rather small (Fig. 2.21(b)). As before, the shape of the cleavage probability distribution depends on the choice of the preset time cut-off.

In order to examining the effect of span on the mean velocity of quadrupedal spiders, I plotted $<x_{cm}>$ versus time for quadrupedal spiders with $S = 3, 4$ and 5 (Fig. 2.22). I found that the biased motion of quadrupedal spiders with greater spans is faster. With increased span, similar improvements in binding time, processivity and mean velocity are observed for bipedal spiders with $S = 1, 2, 3$ and tripedal spiders with $S = 2, 3, 4$ (Fig. 2.23). Based on Fig. 2.22, another explanation of the largest number of cleavages by quadrupedal spiders with $S = 5$ (Fig. 2.21(b)) is their access to the most substrate sites as they travel furthest into the substrate side of the track.
CHAPTER 2. MOLECULAR SPIDERS

Figure 2.22: Dependence of biased motion on maximum span shows greater mean velocity for quadrupedal spiders with larger maximum span. $\langle x_{cm} \rangle$ is the ensemble-averaged center-of-mass displacement of spiders with respect to their initial position at the middle of the track. $\langle x_{cm} \rangle$ is calculated every 100 s by averaging over only those spiders remaining bound to the track at that time. Error bars represent the standard deviation of spider positions, calculated every 50 min.

Effect of Substrate Unbinding Rate

The first experimental study of molecular spiders showed that it is possible to change some kinetic rates of interaction, such as the unbinding rate of a leg from its substrate or product, by modifying the sequence of the leg [86]. In the simulations to this point, I used a rate of leg unbinding from a substrate, $k_{off,S} = 0.035 \text{ s}^{-1}$ of the same order as the cleavage rate (Table 2.1). I next decreased the unbinding rate by an order of magnitude to $k_{off,S} = 0.0035 \text{ s}^{-1}$, so as to increase the relative probability of substrate cleavage. As a test case, I studied the properties of quadrupedal spiders with $S = 3$.

Figure 2.24 presents plots of $\langle x_{cm} \rangle$ versus time for spiders with these two different values of $k_{off,S}$. The results show that the mean velocity of spiders is greater for $k_{off,S} = 0.0035 \text{ s}^{-1}$. Because $k_{off,S}$ decreases relative to $k_c$, this increases the probability of a cleavage event rather than an unbinding event from a substrate site (state 2), which results in: 1) decreasing the likelihood of multiple binding/unbinding events to a substrate site before its cleavage; 2) decreasing the chance of leaving an uncleaved substrate site behind that can act like a trap for the spider at later times;
Figure 2.23: Dependence of biased motion of tripedal and bipedal spiders on maximum span shows greater binding time (a and b), processivity (c and d) and mean velocity (e and f) for spiders with larger maximum span. $\langle x_{cm} \rangle$ is the ensemble-averaged center-of-mass displacement of spiders with respect to their initial position at the middle of the track. $\langle x_{cm} \rangle$ is calculated by averaging over only those spiders remaining bound to the track at that time.
Figure 2.24: Dependence of biased motion on the dissociation rate from substrates shows a greater mean velocity of spiders with a greater ratio of $k_c = 0.055 \text{ s}^{-1}$ to $k_{off,S}$. These results are for quadrupedal spiders with $S = 3$. $\langle x_{cm} \rangle$ is the ensemble-averaged center-of-mass displacement of spiders with respect to their initial position at the middle of the track. $\langle x_{cm} \rangle$ is calculated every 100 s by averaging over only those spiders remaining bound to the track at that time. Error bars represent the standard deviation of spider positions, calculated every 50 min.

and 3) increasing the probability that the spider is maintained at a $P$-$S$ boundary. These effects increase the spiders’ biased motion, resulting in the greater mean velocity when $k_{off,S} = 0.0035 \text{ s}^{-1}$. Spiders did not detach from the track by 800 minutes for either off rate (data not shown), therefore detachment of spiders (3D loss) does not contribute to this increased mean velocity.

Mechanochemical Coupling

Using Eq. (2.6), I studied $\xi(t)$, the coupling between mechanical steps and the chemical cycle of substrate cleavage, as a function of force applied to oppose the biased motion of multipedal spiders. Figure 2.25(a) shows the mechanochemical coupling of tripedal and quadrupedal spiders as a function of time at zero force and at a very small applied force of $F = 0.01 \text{ pN}$. As time increased, the tripedal and quadrupedal spiders exhibited approximately constant coupling ratios of 0.52 and 0.57, respectively, at no applied force. This means that a spider’s centre of mass moves on average one site on the track forward for each two substrates cleaved. However, at $F = 0.01 \text{ pN}$, the coupling ratio of both types decreased significantly with time and reached zero before 140 min. This
Figure 2.25: The mechanochemical coupling $\xi(t)$ versus time for spiders with $n = 3$ and 4 legs and $S = 3$ at (a) zero force and $F = 0.01$ pN (b) $F = 0.03$ pN exhibits time and force dependence. Mechanochemical coupling of quadrupedal spiders is greater than tripedal spiders and the coupling ratio decreases faster with time at higher applied forces. (c) Long-time behaviour of mechanochemical coupling at zero force shows a constant coupling. Coupling is calculated over $\Delta t = 100$ s time intervals.
applied force, while providing an extremely small bias compared with thermal energies \( F \Delta x \approx 0.1 \) pN.nm \( \approx 0.025 \ k_B T \) if the step size is \( \Delta x = 1 \) unit = 9 nm), was sufficient to qualitatively change the dynamics of the spiders. The time-dependent coupling is due to the increased back-stepping of the spiders in the presence of an applied force, resulting in more rapid 1D loss of the ideal \( P-S \) configuration. The coupling ratio decreased faster with time at higher applied forces, so that it reached zero at increasingly smaller times (Fig. 2.25(b)). In contrast, the constant coupling ratio of these spiders at \( F = 0 \) lasted to the end of my simulations (800 minutes; Fig. 2.25(c)).

These results indicated that the mechanochemical coupling of molecular spiders exhibited both force and time dependence. The results also showed greater coupling ratios for quadrupedal spiders compared to the tripedal spiders – independent of applied force – and that the mechanochemical coupling decreased more slowly for quadrupedal spiders for \( F > 0 \) compared to tripedal spiders. I attribute this to the increased number of spider legs: as the number of legs increases, the spider has more legs bound to the track (on average), which act like a brake against its backward motion under an external applied force, reducing the extent of 1D loss.

The result that the coupling ratios are less than unity indicates that the chemical cycles and mechanical steps are not tightly coupled for the molecular spiders even at no applied force. One reason for this loose coupling is the lack of a power stroke, which results in a high frequency of back-steps for molecular spiders. In comparison, for kinesin, with a coupling ratio of one step (8 nm) per ATP hydrolysis cycle [35, 118], a combination of gating (coordinated stepping of the two heads) and a possible power stroke provides the required asymmetry to bias binding to the forward site under most applied forces [118, 134]. For molecular spiders, binding of an unbound leg to the track happens only after a diffusive search. Therefore, in spite of the initially imposed asymmetric state of spiders’ legs at the beginning of my simulations, at longer times legs can be bound to any possible configuration of product and substrate sites, of which the ideal \( P-S \) configuration, optimal for forward bias, is but one.

**Efficiency**

To study how efficiently molecular spiders use the chemical energy of substrate cleavage to produce biased motion, I calculated the thermodynamic efficiency, \( \eta_{\Delta t} \), of bipedal, tripedal and quadrupedal spiders with \( S = 3 \). As discussed in Section 2.3.1 and shown in Eq. (2.7), \( \eta_{\Delta t} \) can be in principle both time- and force-dependent. Therefore, I calculated \( \eta_{\Delta t} \) as a function of force, \( F \), during time
Figure 2.26: Graphs of efficiency versus force for spiders with \( n = 2, 3 \) and 4 legs and \( S = 3 \) for the first and second (inset) \( \Delta t = 100 \text{ s} \) time intervals. Quadrupedal spiders are the most efficient of these molecular spiders and reach stall at greatest applied force. Lines through the points are guides to the eyes.

Intervals of \( \Delta t = 100 \text{ s} \) of my simulations. Figure 2.26 shows the graph of \( \eta_{\Delta t}(F) \) versus \( F \) for the first 100 s time interval. For all spiders, the average efficiency increased initially with \( F \) to its maximum, then decreased gradually as \( F \) continued to increase, eventually reaching stall, at which the spiders no longer move on average. At forces greater than the stall force, spiders moved backward in the direction of applied force so that the average center-of-mass displacement \( <\Delta x_{cm}>_{\Delta t} \) was negative. Based on Eq. (2.7), this resulted in negative values of efficiency (Fig. 2.26). The quadrupedal spiders exhibited both highest maximum efficiency and highest stall force, while the bipedal spiders had the lowest efficiency and stall force.

The dependence of efficiency on leg number can be understood, as for coupling, by the ability of bipedal spiders to tumble backward on the track in the direction of the external applied force, compared with the braking action of multiple bound legs of tripedal and quadrupedal spiders. The movement of bipedal spiders is thus more sensitive to an external applied force, the extent of 1D loss is increased, and they walk backward at smaller forces. (Applied force does not change the degree of 3D loss in my kinetic scheme.)
Figure 2.27: Graph of maximum efficiency versus time for spiders with \( n = 2, 3, 4 \) and \( S = 3 \). For each type of spider, maximal efficiency occurs in the first time interval and decreases to zero at longer times. Quadrupedal spiders are the most efficient. Lines through the points are guides to the eyes.

At later time intervals, I found a similar parabolic-like dependence of efficiency on force and efficiency that increased with number of legs (Fig. 2.26 inset). As expected from Eq. (2.7), the coupling is linear with force over this range of forces, but becomes increasingly nonlinear as it decreases below zero (data not shown).

I used the graphs of efficiency versus applied force to calculate the maximum efficiency of each type of spider in every time interval (Fig. 2.27). This analysis showed that the maximum efficiency of the spiders decreases with time for all three species of spider. The very fast initial decrease of maximum efficiency as a function of time is a result of the initial condition of the spiders, where their legs were bound to the middle of the track at a well defined product-substrate boundary. Such an asymmetric configuration is ideal for biased motion. At later times, the spiders are no longer pinned to the \( P-S \) boundary and the fraction of spiders in the ideal, asymmetric state decreases through 1D and 3D loss.

Recall that these efficiencies were calculated using the input energy of phosphate bond hydrolysis of ATP under standard conditions, \( \Delta G^\circ = -30 \text{ kJmol}^{-1} \) [126]. These graphs would rescale to slightly larger efficiencies if I use instead the hydrolysis energy of a DNA phosphodiester bond (\( \Delta G^\circ = -22 \text{ kJmol}^{-1} \) [127]) in my calculations. Using \( \Delta G^\circ \) of a phosphodiester bond, for the first time interval of 100 s, the calculated maximum efficiencies would increase (from 0.006, 0.004 and...
Figure 2.28: Graph of efficiency versus force shows the increased thermodynamic efficiency of spiders with $n = 4$ legs and $S = 3$ with $k_{off,P} = 1 \text{ s}^{-1}$ for the first $\Delta t = 100 \text{ s}$ time interval. Lines through the points are guides to the eyes.

By increasing $k_{off-P}$, thereby further biasing spiders’ binding to $S$ sites, I was able to increase the efficiency of spiders (and their mean velocity). For example, a modest increase in $k_{off-P}$ from 0.14 $\text{s}^{-1}$ to 1 $\text{s}^{-1}$ roughly doubled the maximum efficiency, to $\eta = 1.2\%$ for quadrupedal spiders (Fig. 2.28), but also increased 3D loss of spiders from the track. A substantial increase to $k_{off-P} = 100 \text{ s}^{-1}$ further increased efficiency, but this resulted in a very rapid 3D loss of spiders, not a viable mechanism for increasing motor performance.

The shape of my force-dependent efficiency curves (Fig. 2.26) is qualitatively similar to that observed for kinesin when back-steps are included in its efficiency analysis [132, 133]. I note, however, that the efficiency of spiders is time-dependent while for kinesin this is not expected to be the case. This is because of the dependence of spider activity on the nature of the track at the position of the spider ($i.e.$, whether it is at a $P-S$ boundary or elsewhere). Furthermore, kinesin operates at much higher forces than molecular spiders.
Summary and Concluding Remarks

In this section I performed Monte Carlo simulations of multipedal molecular spider dynamics based on a chemical kinetics model of interaction of spider legs with the track. My goal was to investigate the dependence of molecular spider properties on tunable experimental parameters such as number of legs, maximum span between legs and interaction kinetics. The results of my simulations for multipedal spiders, $n = 2, 3$ and $4$, with the same maximum span of $S = 3$ indicated that the binding time and processivity of spiders increased as the number of legs was increased from $n = 2$ to $4$. This result is in agreement with experimentally observed increases in binding time and processivity of molecular spiders with an increase in the number of legs [86].

My results also showed that the improvement in binding time and processivity of spiders came with a decrease in mean velocity, where the quadrupedal spiders exhibited the slowest directional motion. However, by increasing the maximum span of quadrupedal spiders ($S = 3, 4$ and $5$) I found that not only the binding time and processivity but also the mean velocity increased with increasing $S$. Thus, I could partially recover the loss of mean velocity with larger $n$ by increasing the spiders’ span. These results are in agreement with theoretical predictions that the motility of spiders should decrease for spiders with more legs and increases with increased maximum span, though their treatment neglected cleavage [104].

I found that an increase in the mean velocity of quadrupedal spiders could also be achieved by modifying the ratio of $k_{off,S}$ to $k_c$ so that, when bound to a substrate, the probability of cleavage relative to dissociation was increased. This occurred when I decreased $k_{off,S}$ from $0.035 \text{ s}^{-1}$ to $0.0035 \text{ s}^{-1}$ while keeping $k_c = 0.055 \text{ s}^{-1}$ and the other rate constants the same. Although our quadrupedal spiders with $S = 3$ and $k_{off,S} = 0.0035 \text{ s}^{-1}$ could not achieve the mean velocity of bipedal spiders, they exhibited both increased mean velocity and increased processivity with respect to quadrupedal spiders with $k_{off,S} = 0.035 \text{ s}^{-1}$. In contrast to the experimental studies of molecular spiders in which "any increase in processivity comes at the cost of lower speed" [87], I have presented feasible experimental means by which the mean velocity of molecular spiders can be increased along with processivity. In additional studies of quadrupedal spiders, I have found that they can reach and even exceed the bipedal spiders’ mean velocity when $S = 5$ (Fig. 2.29).

It is important to note that the faster mean velocity of quadrupedal spiders compared to bipedal spiders was transient, so that by 200 minutes the mean velocity of quadrupedal spiders was lower than bipedal spiders. However, at 200 minutes, less than 5% of bipedal spiders but almost all
quadrupedal spiders were bound to the track, so the effect of detachment of spiders from the track (3D loss) dominated the observed greater mean velocity of bipedal spiders.

My investigations of mechanochemical coupling showed that the coupling ratios for tripedal and quadrupedal spiders are time-independent in the absence of force. In the presence of an applied force, however, these ratios become time-dependent, such that the spiders’ mechanochemical coupling decreases with time, more rapidly for tripedal than quadrupedal spiders. To my knowledge, a time-dependent mechanochemical coupling under an external applied force has not been observed or proposed for any biological or synthetic molecular motor. This behavior is a consequence of the design of the spider, which permits the loss of spiders from the ideal $P-S$ configuration, either through 1D (diffusive) or 3D (detachment) loss, and thereby introduces a dependence on the position of the spider into its motor properties. The 1D loss pathway is responsible for the time-dependent coupling elucidated in my study. It is likely that populations of motors operating under similar principles, such as has been proposed for the biological example of collagenase [63], will exhibit similar time dependence.

The importance of 1D loss to spider velocity has been clearly elucidated in a recent theoretical study [103]. In that work, the loss of spiders from the $P-S$ boundary state to a diffusive state resulted
in spiders whose motion was purely diffusive in the long-time limit. My simulations also display such transient ballistic motion, though I do not see a complete transition to diffusive behavior in the finite times sampled in my simulations. 1D loss from the boundary state plays a central role in the mechanism of spider motion, even in the context of my more complex kinetic scheme. For example, kinetic pathways allowed in my simulations, and in experiments, include the ability of a leg to release from substrate without cleavage into product (resulting in spiders being able to step over and ahead of an existing $P$-$S$ boundary) and the detachment of spiders from the track (3D loss into solution), yet I also see the gradual transition of spider population away from the ideal $P$-$S$ boundary state and a decrease in motor performance with time. Simulations with my more realistic kinetic scheme reveal complexity to the dynamics, particularly of bipedal spiders, that are not seen in the idealized model [103], and allow us to make more specific predictions about timescales and motor properties of the experimental systems.

Finally, I studied the thermodynamic efficiency of multipedal spiders under a range of applied conservative forces. With the exception of my previous studies of the efficiency of inchworm and hand-over-hand bipedal spiders (Section 2.2.3), no theoretical or experimental investigations of spider efficiency have been performed. To study the efficiency of molecular spiders, I introduced a definition of efficiency in Eq. (2.7) that allows us to take time-dependent performance into account. This definition is applicable to any molecular motor working against a conservative force including motors with a varying coupling ratio. For molecular motors with a fixed coupling ratio, efficiency is clearly a function of force. However for these motors the efficiency at different applied external forces is predictable, while for spiders, this is not the case.

My results for the efficiency of bipedal, tripedal and quadrupedal spiders showed that all three species exhibited force dependence, where the efficiency increased with force to a maximum and then decreased to zero at the stall force. My calculation of the force-dependent efficiency of multipedal spiders as a function of time showed the quadrupedal spiders to be the most efficient spiders that I studied. Moreover, they reach stall conditions at the largest forces. My simulations provide testable predictions of this performance characteristic of these synthetic walkers. They also show that, while molecular spiders are weak molecular motors on the average, they are capable of utilizing some of the chemical energy of substrate cleavage to bias their motion for extended periods of time. The diffusive nature of their stepping mechanism results in increasing population loss from the (mobile) $P$-$S$ boundary with time (1D loss), and thus they must be considered as transient molecular motors (albeit transient over timescales far longer than most experimental timescales).
2.4 Conclusions

My goal was to understand the origin of biased motion of molecular spiders and how experimentally designable parameters of molecular spiders can be tuned to optimize motor performance. My simulation results suggest that the observed biased motion is a result of the difference between the dwell time of the interaction of a spider’s leg with a substrate and product, and that cleavage of the substrate improves the directionality by producing a mobile $P-S$ boundary. I also learned from this study that improving measures such as velocity and processivity can be achieved by tuning more than one experimental parameter simultaneously, otherwise some spider properties might improve at the expense of others. For example, spiders with $n = 4$ legs showed increased processivity and efficiency but decreased average velocity compared to bipedal spiders. This is different from the behavior of systems with several molecular motors pulling a load along a track, where each "leg" is an individual motor [135, 136, 137, 138]. Adapting the system design to one in which an individual spider leg can step processively [82] may provide an alternative strategy to optimize performance. In addition, the decrease in speed with increasing number of legs can be contrasted with the transport of particles by multiple kinesins, in which speed was found to be independent of number of motors; in both examples, however, processivity increases with the number of motor units [139].

My studies of molecular spiders show that they are rather weak molecular motors. Even the most efficient spiders I have studied, quadrupedal spiders ($\eta_{\text{max}} = 1.2\%$, Section 2.3.2), possess an efficiency much smaller than the efficiency of tightly coupled biological molecular walkers like kinesin ($\eta = 40 - 60\%$) [34, 119, 140]. Although this maximum efficiency increases to 1.7% when considering the input energy of $\Delta G^\circ = -22 \text{kJmol}^{-1}$ [127] for cleavage of a phosphodiester bond, it is still much smaller than the efficiency of molecular motors such as kinesin. The efficiency of molecular spiders is also significantly lower than that of an idealized "flashing ratchet" ($\eta \leq 5\%$) [141], but is on the same order as the efficiency of other proposed synthetic molecular motors [142, 143, 144]. My studies showed that spiders lack mechanochemical coupling between their legs. In biological molecular motors this coordinates leg binding and release [145]. Spiders also lack a power stroke to intrinsically bias the motion of a leg in the forward direction [134]. The large randomness parameter found for spiders in my simulations ($r > 10$, Section 2.2.3) is further evidence of the dominance of diffusion over biased motion, especially when compared with $r \leq 1.5$ for kinesin [118]. The fuel source for spiders and kinesin is similar, as both hydrolyse a phosphate bond. However, even with a similar step size, kinesin is capable of working against loads of approximately 7 pN [1], much
higher than my predicted stall force of all types of studied spiders.

This work is related to studies of burnt-bridges motors which have served as theoretical models for the biological motor collagenase [63, 64, 146]. Simulations of these motors suggest that, like the molecular spider, they are weak motors with low efficiency. Unlike burnt-bridges motors, the legs of molecular spiders retain the ability to bind to the cleaved products, as seen through my use of $k_{on,P} > 0$, and are able to move in either direction from the cleavage site, even if $k_{on,P} = 0$. My spider model is more similar to a model for collagenase that allows the motor to diffuse to either side of a bond following cleavage [65].

I hope that the approach used here will serve as an experimentally relevant tool to assist in the design of synthetic motors with specific performance goals such as speed, processivity or efficiency. By using as input experimentally controllable parameters such as rates of binding, cleavage and dissociation and number of spider legs for the molecular spider, my model creates the opportunity for in silico optimization of motor performance.

My new approach to calculating the efficiency of molecular spiders provides a tool for characterizing their performance as motors for a given set of experimental conditions. These in silico approaches to motor design can then be applied in the laboratory, for applications as diverse as transport, drug delivery and chemical synthesis [86, 147].

To extend my model from a 1D track to a quasi-2D matrix is beyond the scope of my current study, but it is certainly feasible using the same Monte Carlo treatment of chemical kinetics for each leg, while generalizing the stepping mechanism.
Chapter 3

Lawnmower: an Autonomous Molecular Motor

3.1 Introduction

In this chapter, I present the design and construction of a novel synthetic motor, the "lawnmower", and its track. The lawnmower is a synthetic protein-based molecular motor constructed from building blocks that lack motor function, and is designed to use cleavage of its substrate to rectify diffusion. It functions based on the same operational principles as molecular spiders [86, 87, 105, 106] (Chapter 2).

As an illustration of the concept (Fig. 3.1), consider a surface covered with a "lawn" of substrates that each contain a recognition sequence for a specific enzyme. Placed on the lawn, the lawnmower will "cut the lawn" locally and produce a patch of digested lawn. The lawnmower should thus consist of a hub covered with enzymatic blades that cleave the substrates. The motion will be autonomous and biased if (i) the lawnmower has a significantly higher affinity to undigested substrate sites than to digested sites, and if (ii) the motor always moves off the digested sites.

The lawnmower molecular motor concept was inspired by human interstitial collagenase MMP-1, which cleaves collagen fibrils at specific sites that are about 300 nm apart and is thought to move by one-dimensional diffusion along the fibril between sites [63, 66]. MMP-1 is powered by proteolysis, and its motion is proposed to be biased because it literally destroys its track, forcing itself to move in one direction [63]. This is similar to a burnt-bridges Brownian ratchet in which
biased diffusion depends on the state of the substrate track: asymmetry can be created by track cleavage, which prevents backward diffusion [59, 148, 149].

For the lawnmower, the burnt-bridges concept would be realized not by destroying the track (as in the case of collagenase), but by leaving behind a product. This is akin to the behaviour of molecular spiders in which a DNAzyme digests the substrates and leaves shorter ssDNA products [86]. Similar to molecular spiders [87], the direction of biased motion could be controlled by patterning a lawnmower track on a surface.

Since our goal here is to synthesize a protein-based molecular motor, we use protein enzymes as blades and thus protein or peptide substrates as the lawn of the lawnmower. To construct the lawnmower, there are different classes of proteases and associated substrates that can be used as lawnmower blades and lawn, respectively. Also, the central hub of the lawnmower can be selected from a variety of possible candidates such as microspheres, quantum dots (QDs) or even streptavidin molecules as used in the construction of molecular spiders [86, 87]. In addition, to attach the proteases specifically to the central hub, different linkers and chemistries can be utilized.

Each component of the lawnmower plays an important role in determining its chemical and physical properties, such as kinetics of interactions, flexibility of the blades to search for a substrate, and density of blades on the hub, and therefore, impacts lawnmower function. Thus, to design the lawnmower, we face a vast potential parameter space from which lawnmower components should be selected to result in generating biased motion. Exploring the parameter space can be done both
In Chapter 2, I presented an algorithm based on the interaction of a molecular spider with its substrate. Using this algorithm, I investigated the effects of experimental parameters on the directional motion of molecular spiders. Here, I use my findings and results from molecular spider simulations to guide the selection of lawnmower building blocks. Table 3.1 shows a summary of the simulation results for molecular spiders, which will be used as a reference for discussions in this chapter about the choice of lawnmower components.

In this chapter, I outline the selection of molecular components of the lawnmower (the hub, linker and blades) (Section 3.2). Then, in Section 3.3, I describe the experimental construction of the lawnmower. Because the experimental work on the lawnmower itself was primarily conducted by research associate Dr. Suzana Kovacic, only highlights necessary to understand the motor design and activity are presented in this thesis. The choice of lawnmower substrate is explained in Section 3.4, together with criteria for detecting its cleavage. In Section 3.5, my work making and characterizing the lawn is presented.

The lawn for the lawnmower can be designed to be a 1D road, a 2D plane or a 3D matrix of substrates. Intuitively, it is easiest to characterize directionality on a 1D track of substrates. One way to design a 1D lawn is assembling it on linear DNA, which serves as its structural backbone.

DNA presents significant advantages as a building block for nanostructured materials, as specific sequences of DNA can be manipulated easily using a large set of biochemical and molecular biology tools and protocols. In addition to its biological role of encoding genetic information, DNA also possesses unique physical properties that make it amenable for use as a template or material for nanostructures. These include its mechanical stiffness [150] and its ability to self-assemble driven by the specific recognition of complementary bases. In Section 3.5, I explain how DNA as a 1D

<table>
<thead>
<tr>
<th>Physical property</th>
<th>More legs</th>
<th>Greater span</th>
<th>Smaller $k_{on}$</th>
<th>Larger $k_{off,P}$</th>
<th>Smaller $k_{off,S}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binding time</td>
<td>increase</td>
<td>increase</td>
<td>no change</td>
<td>decrease</td>
<td>decrease</td>
</tr>
<tr>
<td>Processivity</td>
<td>increase</td>
<td>increase</td>
<td>slight increase</td>
<td>not studied</td>
<td>not studied</td>
</tr>
<tr>
<td>Speed</td>
<td>decrease</td>
<td>increase</td>
<td>increase</td>
<td>increase</td>
<td>increase</td>
</tr>
</tbody>
</table>

Table 3.1: Effect of experimentally adjustable parameters on motor properties of molecular spiders. 

"Studied only for a special case of quadrupedal spiders with $S=3$. "Studied only for a special case of HOH bipedal spiders. Rate constants refer to the kinetic scheme in Figure 2.2."
molecule is used as the backbone for peptide tracks of the lawnmower. In that section, I also show how the lawnmower track is designed so that it can be manipulated for later application in techniques such as DNA tightropes [99, 100] and DNA curtains and racks [90, 91] (Section 1.4).

3.2 Lawnmower Design

To assemble the lawnmower, first its main components were selected (Fig. 3.1), which are: 1) the motor hub: the base of the motor; 2) the blades: functional parts of the motor; and 3) the linkers: connectors of the proteases to the central hub.

3.2.1 Lawnmower Hub

The hub of the lawnmower is the central part of the motor to which proteases are attached via linkers. Depending on the desired size of the lawnmower, a micron-sized bead, a nanoscale quantum dot (QD) or a streptavidin protein can be used as the lawnmower hub. Each of these candidates provides different binding capacities for linkers and so for the proteases. Also, from an experimental point of view each has its own advantages and disadvantages, as discussed in the rest of this section.

Streptavidin has four natural binding sites for attachment of biotinylated linkers. Increasing the size of the hub further to a nanoscale QD and then to a micron-scale bead increases the surface area by four orders of magnitude. To use these beads as hubs, their surfaces must be chemically functionalized to permit attachment of linkers to the blades. The greater surface area of a micron-scale bead and its larger radius of curvature increases the number of protease blades that can simultaneously interact with substrates on the track. This is illustrated schematically in Fig. 3.2.

I investigated previously the effect of leg/blade number on directional motion by varying the number of legs of molecular spiders (Table 3.1). In simulations of multipedal molecular spiders, increasing the number of legs from \( n = 2 \) to 4 resulted in a great improvement in binding time and processivity but at the same time reduced the speed drastically.

In the lawnmower experiments, we would expect that a densely covered micron-size bead hub results in a lawnmower with a large number of legs interacting simultaneously with substrates (Fig. 3.2). This would slow down the lawnmower dramatically so that it gets stuck on the track on the minutes time-scale of the experiments. Therefore, in spite of the great advantage of a micron-sized bead whose position can be tracked easily using bright-field imaging, this option is eliminated because of its anticipated very slow speed. Also to avoid presumably rapid dissociation
Figure 3.2: Schematic illustration of the increase in the number of potentially interacting lawnmower blades (triangles) with substrates when going from a streptavidin hub (bottom, right), to a QD hub (bottom, left) and to a micron-size-bead hub (top). Figures are drawn to scale (a 5 nm streptavidin, a 20 nm QD and a 1.27 µm bead) assuming proteases are densely linked to the hub and substrates are located every 5 nm on a track (the minimum inter-substrate distance required for adjacent binding of protease blades (average diameter of protein ∼5 nm) [151]).

of streptavidin-based lawnmower on a 1D track of substrates (see Section 3.4 for track details), streptavidin is also eliminated. It should be noted that, except if there are very long flexible linkers between the proteases and the hub, when the lawnmower is close to a 1D track, the steric hindrance of the streptavidin hub would not allow all four lawnmower legs to interact with the track at the same time (Fig. 3.2). Also from an experimental point of view a streptavidin hub requires fluorescent labelling in order that it can be visualized and its displacement measured. This is not the case if the streptavidin is replaced with a QD.

As depicted in Fig. 3.2, it is likely that a nanoscale QD hub provides a mid-range number of interacting legs, which is expected to increase the binding time of a QD-lawnmower compared to a streptavidin-lawnmower while increasing speed compared to a micron-bead lawnmower.

Also, experimentally, QDs offer significant observation advantages. Core/shell semiconductor QDs are extremely efficient materials for generating fluorescence when excited [152, 153]. They have very high emission quantum yields, sharp emission spectra, chemical stability and photostability. Their intrinsic brightness is often many times that observed for traditional organic fluorophores, and their photostability is many orders of magnitude greater. Therefore the fluorescent signal from a
QD can be utilized to detect its displacement without any further chemical modification of the hub. In addition, the fluorescence wavelength of QDs is easily selectable throughout the whole visible spectrum [154, 155]. The only drawback is the blinking of the QD, which refers to its discontinuous emission of light [156, 157], though this is possible to address through appropriate chemical synthesis or control of the QD [158, 159].

A variety of quantum dots with different fluorescence wavelengths and surface chemistries are commercially available. Functional groups (such as carboxyl and amino) can be utilized as handles to couple the biomolecules of interest, here the linkers between the protease and the QD, to the surface of the QD. From the commercially available QDs, we selected a 20 nm amine-derivatized quantum dot with 600 nm emission wavelength (Invitrogen). Based on the information provided by the manufacturer, there should be 10-12 reactive amine groups accessible for surface chemistry on the surface of this QD.

### 3.2.2 Lawnmower Linker

In order to attach the proteases to the QD, a heterobifunctional molecular linker is used to couple each protease to the QD.

The effect of the linker’s length on biased motion was studied for molecular spiders (maximum span, Table 3.1). I found that longer linkers, via their greater diffusional search volume, provide access to further sites, larger step sizes on the track, and increased biased motion of molecular spiders. Linkers furthermore reduce geometrical and steric constraints of proteases by allowing them to more freely locate the substrate. By attaching proteases specifically via linkers, nonspecific binding of proteases to QDs can be eliminated, which might otherwise lead to dissociation of proteases from the QD surface during experiments.

Based on the surface chemistry of the QDs and chemical methods designed to be used for cross-linking of the proteases, two heterobifunctional linkers with the approximate lengths of 1 nm (Sulfo-SMCC: Sulfosuccinimidyl-4-((N-maleimidomethyl)cyclohexane-1-carboxylate; Pierce) and 10 nm (Maleimide-dPEG24-NHS ester; Quanta Biodesign) were selected (Fig. 3.3). These will permit experimental verification of the function of linker length in future experiments.
3.2.3 Lawnmower Blades

To bind the protease blades to the hub, we exploit functional groups available on the protease surfaces to covalently bind to the linker.

Practically, two of the protein chemical targets that are commonly used for cross-linking and chemical modification techniques are primary amines (-NH$_2$) and sulfhydryls (-SH). In proteins, a sulfhydryl group is present in the side chain of each cysteine (Cys, C) residue, while primary amines exist at the N-terminus of each polypeptide chain and in the side chain of lysine (Lys, K) residues.

In order to target -NH$_2$ or -SH functional groups, Lys or Cys must be in the amino acid sequence of the protease. Also, these groups must be positioned close to the surface of the tertiary structure of the protease to be accessible to cross-linking agents. This can be determined by checking the tertiary structure of proteases using the Protein Data Bank [160]. In addition, the targeted Cys and Lys should not be part of the active site of the protease. Otherwise, cross-linking might result in deactivation of the protease. Also, the chemical treatments of these groups should not affect the structure of the protease. Often, disulphide bonds (-S-S-) formed between two cysteines stabilize...
the structure of a protein. These must be reduced to sulfhydryls prior to cross-linking, which may result in the disruption of protease structure or function.

Considering the above concerns, we decided to select protease blades by screening the activity of a group of proteases from different families following cross-linking to primary amines or sulfhydryls. The protease with highest preserved activity would then be chosen for construction of the lawnmower.

While a more elegant approach to cross-linking, and to tuning of kinetic rate constants (e.g. Table 3.1), would involve recombinant expression and mutagenesis, this lies beyond the scope of the present work.

3.3 Lawnmower Construction

Construction of the lawnmower was undertaken by Dr. Suzana Kovacic. Details of her synthetic work are provided in Appendix D for reference. Here, I highlight the results demonstrating that a functional trypsin-based lawnmower is made. This motivates my work on its track in the following sections.

To construct the lawnmower, first proteases from different families of enzymes were selected based on availability. The effect of cross-linking on the proteases’ activities was investigated following different chemical treatments (Fig. D.1). Trypsin was selected for the lawnmower blades due to its preserved activity following coupling to the linkers via Lys residues.

When considering trypsin for the blades of the lawnmower, a comparison should be made between its kinetic cycle and that of the deoxyribozyme legs used in the spiders. A literature review on trypsin kinetics indicated that, compared to spiders’ kinetics, trypsin’s catalytic rate is much faster [161, 162]. This might decrease the interaction time of lawnmower blades with their substrates and therefore the processivity. However, it has also been shown that the rate of product release is much smaller for trypsin [163], so that product release will increase this interaction time and improve the processivity. In addition, trypsin kinetic rates are highly substrate-dependent [161, 162]: its catalytic rate varies by at least three orders of magnitude for substrates of differing amino acid sequence [161, 162]. The dependence of trypsin kinetics on the sequence of the substrate sequence gives us the potential in the future to modify the interaction time of lawnmower blades with their substrates by screening different substrates.

The lawnmower was constructed by using heterobifunctional Sulfo-SMCC linkers of length 1
nm (Fig. 3.3(a)) to crosslink generated thiols on trypsins to amine-fictionalized QDs. (The shorter length linker was chosen for these initial experiments for cost efficiency.)

Lawnmowers were purified from unbound trypsins using size exclusion chromatography (SEC). In SEC, particles of different sizes elute through a stationary phase at different rates. This results in the separation of a solution of particles based on size. Provided that all the particles are loaded simultaneously or near-simultaneously onto the column, particles of the same size should elute together. Therefore, it was expected that the trypsin-conjugated QDs and QDs, in reaction and control, respectively, elute faster than unconjugated trypsins. Fluorescence both from QDs and generated by trypsin cleavage of a fluorogenic substrate, as an indicator of trypsin activity, were measured in the SEC-eluted fractions.

Figure 3.4 shows the fluorescence signal from QDs and from trypsin cleavage of the fluorogenic substrate, for the collected fractions from reaction and control. Control here was a mix of untreated trypsins with unmodified QDs. As expected, in both samples, the QD fluorescence signal was detected in the initial eluted fractions. The fluorescence signal from digestion of the fluorogenic substrate by trypsin was negligible in the control, for earlier elutions, which indicated that there was no trypsin in these eluates. However, in the early fractions from the reaction, this signal co-eluted with QDs. We attribute this co-elution to trypsins conjugated to QDs. In both reaction and control, there was a fluorescence activity peak at later elutions which corresponding to unconjugated trypsins.

These measurements confirm the construction of an active lawnmower. Using the fluorescence intensity of QDs and rates of cleavage measured for the lawnmower (Fig. 3.4(a)), and comparing these with the intensities of known amounts of QDs and rates of fluorogenic substrate cleavage by known amounts of trypsins, the number of conjugated trypsins per QD is estimated to be eight (Appendix D.7).

### 3.4 Lawnmower Track Design

As discussed in Section 3.3, the lawnmower uses trypsin as its blades. Therefore, the substrate needs to be a protein or a peptide that is cleaved by trypsin. Trypsin digests a peptide chain after Lysine (Lys) or Arginine (Arg) residues except when followed by a Proline (Pro) residue. Therefore, the substrate peptide has to include at least one of these cleavage sites in its sequence. To observe the digestion of the substrate and correlate that to the displacement of the lawnmower QD hub, a
Figure 3.4: Monitoring the fluorescence signal of both QD (blue points) and from a trypsin kinetic assay (red bars) for the collected fractions from (a) reaction, trypsin-conjugated QD and (b) control, unconjugated trypsins and QD. QD and trypsin-conjugated QD elute earlier from this size-exclusion column while unconjugated trypsins elute in later fractions. (Figure courtesy of Dr. Suzana Kovacic.)
marker is desirable (Fig. 3.1). This can be done through different fluorescence labelling methods. In one method, substrates can be tagged fluorescently so that disappearance of the fluorophores after digestion indicates the lawnmower activity. Another approach utilizes an internally quenched fluorogenic substrate which becomes fluorescent after cleavage. It is easier to detect appearance of a single-molecule fluorescence signal in the second method compared to its disappearance in the first one. In the first method, reduction in fluorescence can happen due to photobleaching even without substrate digestion by trypsin. To avoid any possible product inhibition of protease blades, after digestion of a substrate, a simple linear peptide rather than a globular protein substrate was chosen. This would result in a reduced local concentration of products after digestion reaction.

Among commercially available peptide substrates for protease kinetic assays, we found an internally quenched fluorogenic peptide substrate which is digested by trypsin. The peptide is designed with a methoxycoumarin (MCA) fluorophore at one end and a dinitrophenyl (Dnp) quencher at the other end. The peptide sequence with quencher and fluorophore is MCA-Ala-Pro-Ala-Lys-Phe-Phe-Arg-Leu-Lys(Dnp)NH2 (Fig.3.5(a)). There are three potential trypsin cleavage sites on this substrate, though the Arg site is most likely cleaved in the context of the track as the Lys are modified (by fluorophore and by attachment to the DNA, below). Digestion of this substrate by trypsin results in the release of Dnp and generation of a fluorescence. This is an appropriate candidate to be used in the substrate lawn of the lawnmower.

To construct the lawnmower track, the next decision was the dimensionality of the track. As explained in Section 3.1, we wished to build the lawnmower track on a DNA backbone. The challenge became how to attach the peptide substrates to a dsDNA molecule.
Figure 3.5: (a) Structure of the peptide substrate with fluorophore and quencher. The sequence is MCA-Ala-Pro-Ala-Lys-Phe-Phe-Arg-Leu-Lys(Dnp)-NH2. (b) Schematic of azide-peptide and alkyne-DNA before click reaction. Azide-peptide is formed by the reaction of NHS-PEG4-azide with the amine group of the lysine side chain, and alkynes are incorporated into DNA using PCR. (c) Click chemistry between alkyne-DNA and azide-peptide results in the formation of the peptide-DNA lawnmower track. MCA and Dnp are the fluorophore and quencher on the peptide.
One powerful approach that has found many applications in selective DNA labelling is click chemistry [165, 166]. In click chemistry, the 1,3-dipolar cycloaddition of azides to alkynes covalently forms a very stable triazole unit (Fig. 3.6). Such processes proceed rapidly to completion and also tend to be highly selective for a single product. Click chemistry has recently found a tremendous number of novel applications in labelling a wide range of biomolecules [167, 168, 169, 170].

An advantage of click chemistry for construction of the lawnmower track is that azides and alkynes can be attached to peptides and nucleic acids without greatly disturbing their biophysical properties [165].

Click chemistry can be applied in the synthesis of the lawnmower track by introduction of an azide moiety into the peptide and incorporating alkyne groups into DNA (Fig. 3.5(b)). To introduce the azide moiety into the peptide substrate, the primary amine of the lysine side chain can be reacted with NHS-azide to covalently link the azide to the peptide (Appendix E). Then, incubation of alkyne-functionalized DNA with azide derivatized-peptide under appropriate conditions should result in the high yield formation of a peptide-labeled DNA track (Fig. 3.5(c)). Our concerns, here, are steric interactions arising from the proximal DNA that might hinder digestion of the peptide substrate by trypsin blades. The highly charged DNA backbone could also affect trypsin binding and activity. We therefore decided to introduce a spacer between the peptide and DNA to increase their separation.

NHS-PEG4-azide (in which PEG4 is the spacer) was selected for azide labelling of the peptide substrate. The PEG chain provides flexibility and distance between the DNA backbone and the peptide substrate. Both of these effects, presumably, would improve the trypsin-substrate binding efficiency. It is likely that linking to the amine group on Lys can reduce or eliminate trypsin cleavage at this site due to interference with chemical recognition of the Lys by trypsin. However, since the Arg residue is unaltered, this peptide can still function as a cleavable lawnmower substrate.

There are examples in the literature showing that construction of a densely internally labeled DNA can be achieved through polymerase chain reaction (PCR) amplification by incorporating unnatural triphosphates carrying the desired functional groups [171, 172]. The types of chemical modifications tolerated by DNA polymerases are limited, and it is frequently unpredictable which modifications will be tolerated [173, 174, 175].

In work most relevant to ours, Gierlich et al. [171] attempted to incorporate 5-alkyne- and 5-azide-modified pyrimidine triphosphate into 300, 900 bp and 2 kbp fragments of DNA using PCR amplification with different polymerases. The optimization of their PCR conditions showed that
the stability of the resulting strands required the alteration of PCR protocols. They indicated that DNA polymerases of family B, like Deep Vent exo\(^{-}\), Pwo and KOD XL, are superior compared to other polymerases when efficient incorporation of the modified triphosphates is desired. Their experimental results also showed that PCR is more efficient for making shorter strands of DNA (300 and 900 bp) in the presence of alkyne-containing triphosphate while it requires a very precise control of the annealing temperature to create 2 kbp PCR products.

The approximate size of trypsin is about 5 nm and the pitch of DNA is \(\sim3.5\) nm. Considering these sizes, the ideal track for the lawnmower would have spacing of at least 7 nm (two DNA turns) between substrate "blades of grass", presenting the substrates on the same side of the track and sparse enough for simultaneous nearest-neighbour binding of different lawnmower legs. Larger distances between substrates are also possible, but would require long linkers between trypsins and the QD to enable multivalent binding.

Assuming a 7 nm spacing, a track of 900 bp alkyne-DNA (\(\sim300\) nm) that can be produced using PCR conditions [171] provides about 44 substrates, likely enough to study processivity of the lawnmower.

Therefore the only other concern is extension of the lawnmower track for its application in flow-based DNA stretching techniques (Section 1.4). For this goal, ligation of the 900 bp alkyne-DNA to long DNA handles is necessary. This was considered in the design of PCR primers to contain proper restriction sites for later digestion/ligation reactions of 900 bp alkyne-DNA products. The restriction sites can also be used to extend the alkyne-DNA region of the track if necessary. It is possible that the presence of alkyne in the restriction sites could affect digestion and/or ligation. Therefore, EagI and PspOMI restriction enzymes with restriction sequences of 5’CGGCCG3’ and
5’GGGCCC3’, respectively, were chosen since these restriction sequences lack alkyne-modified nucleotides following PCR amplification. Cleavage at the EagI and PspOMI restriction site ends can be followed by ligation of the track to long dsDNA "handles" for manipulation with techniques such as DNA racks or tightropes (Section 1.4). Another advantage of these two restriction sites is that they are compatible with each other allowing end-to-end ligation of shorter fragments to elongate the alkyne-DNA track if desired.

There are advantages in using the described method to construct the lawnmower track. 1) It is a scalable method of constructing a one-dimensional nanoscale track that does not require nanofabrication technology. 2) Yield of each synthesis step can be checked using common methods of biochemistry and molecular biology. Formation of alkyne-DNA can be verified by clicking the DNA to azide-fluor and probing for the fluorescence signal coincident with DNA bands in an agarose gel. Alkyne-DNA PCR/ligation/digestion products can be run in agarose gels to confirm the formation of the expected DNA products. Conjugation of peptides to DNA can be examined using UV-Vis spectroscopy after purification of the click reaction products from initial components. 3) The average spacing between peptide substrates can be controlled and adjusted. Controlling the density of the peptide substrates on the DNA backbone can be done by varying either the alkyne concentration in the PCR reaction or the azide-peptide concentration in the click reaction with alkyne-DNA. The substrate spacing is an important parameter for the activity of the lawnmower. If the peptides are spaced too closely, the chance of leaving undigested substrates behind increases, so that these sites can act as traps to keep the lawnmower behind the optimal product-substrate boundary. The biggest disadvantage to this method is that there is no control over producing an evenly spaced peptide-DNA track that also has all peptides on the same side of the track.

### 3.5 Construction of Lawnmower Track

#### 3.5.1 Construction of Alkyne-DNA

First, I tried to produce alkyne-DNA PCR products with different labelling density of alkyne. To generate alkyne-DNA, in the PCR reaction the deoxythymidine triphosphate, dTTP, was replaced in different amounts of 0, 20, 50, 80 and 100% by alkyne-dUTP [C8-Alkyne-dUTP (5-(octa-1,7-diylnyl)-2′-deoxyuridine 5′-triphosphate)] (Jena Bioscience) (Fig. 3.7). The pUC19 plasmid (NEB), 2686 bp, was used as the template. Right and left primers (Integrated DNA Technologies, IDT)
CHAPTER 3. LAWNMOWER: AN AUTONOMOUS MOLECULAR MOTOR

Table 3.2: Sequence of the primers used in PCR amplification to produce 3 kbp DNA products.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>EagI primer</td>
<td>5’ATG CTT CGG CCG TAT GCG GTG TGA AA3’</td>
</tr>
<tr>
<td>PspOMI-3000</td>
<td>5’TCA TAT GGG CCC TGG TGC TGC ACT CTC AG3’</td>
</tr>
</tbody>
</table>

Table 3.2: Sequence of the primers used in PCR amplification to produce 3 kbp DNA products.

named EagI primer and PspOMI-3000 (Table 3.2), were designed to carry the EagI, CGGCCG, and PspOMI, GGGCCC, recognition sites, respectively. These primers are expected to generate linear 2710 bp (~3 kbp) DNA products. Based on a general rule in designing primers with restriction sites, 6 basepairs were added to the 5’ ends of the restriction sites to ensure efficient binding by the enzymes (coloured in red in Table 3.2). Following these are the recognition sites of EagI and PspOMI restriction enzymes (coloured in green in Table 3.2), and the rest of the sequences were designed to basepair with their complementary sequences in the pUC19 template.

Figure 3.8 shows the PCR products from a reaction with Taq polymerase (Invitrogen), run in a 1% agarose gel and stained with ethidium bromide (EtBr). There is trace of products in 0, 20 and 50% reactions with alkyne-dUTP but not in the 80 and 100% ones. A possible explanation for these results is that there is no alkyne-dUTP product in the reactions and the appearance of DNA bands and their increasing intensity with increased percentage of the dTTP is due to production of unlabelled DNA. The same results and trends were observed when the Taq polymerase was replaced with Taq native (Invitrogen) or Phusion polymerase (Thermoscientific) (data not shown).

Following these results, we found the work by Gierlich et al. [171] on synthesis of short alkyne-DNA constructs using family B polymerases. To take advantage of their findings, the PCR primers were redesigned to generate shorter products. To do this, the EagI primer was kept the same (Table 3.2), while the PspOMI primer was replaced with newly designed primers named PspOMI-400
CHAPTER 3. LAWNMOWER: AN AUTONOMOUS MOLECULAR MOTOR

Figure 3.8: EtBr-stained 1% agarose gel of PCR products of pUC19 with primers shown in Table 3.2, Taq polymerase and varying ratios of Alkyne-dUTP and dTTP as indicated. Products are seen only when less modified nucleotide is used.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PspOMI-400</td>
<td>5’TCT TAT GGG CCC TGT GGA ATT GTG AGC G3’</td>
</tr>
<tr>
<td>PspOMI-1000</td>
<td>5’TGA TAA GGG CCC AGC TTG GAG CGA AC3’</td>
</tr>
</tbody>
</table>

Table 3.3: Sequences of the primers used in PCR amplification to produce 352 bp (PspOMI-400) and 952 bp (PspOMI-1000) DNA products.

and PspOMI-1000 to generate 352 bp ("400 bp") and 952 bp ("1 kb") products, respectively (Table 3.3). To guarantee alkyne incorporation in successful PCR reactions (thermocycles given in Appendix F.1), they were performed with dTTP fully replaced with alkyne-dUTP. Following Gierlich et al. [171], I also used Pwo polymerase (Roche Applied Science).

Figures 3.9(a) and (b) show the EtBr-stained PCR products for control- and alkyne-DNA after running in a 1% agarose gel. The results show the successful PCR amplification of the 400 bp alkyne-DNA as the band intensity of this product is about the same as that for control DNA. For 1 kb products, alkyne-DNA was produced but not as efficiently as the control. The observed slightly slower migration of alkyne-labelled DNA (Fig. 3.9(a)) in an agarose gel relative to control DNA has been reported previously [171].

Both alkyne-DNA and control DNA were PCR purified (PCR purification kit, QIAGEN) and their absorption spectra were measured using a UV-Vis spectrophotometer (1700 UV-Vis Spectrometer, Shimadzu). Absorption spectra of 400 bp products are shown in Fig. 3.9(c). The use of an alkyne group in the linker of the modified nucleotide (Fig. 3.7) extends the π conjugation of
Figure 3.9: EtBr-stained 1% agarose gel of (a) 400 bp and (b) 1 kbp PCR products of control DNA (lanes 2) and alkyne-DNA (lanes 3) by Pwo polymerase. Comparing the intensities of each control with its alkyne-DNA, it shows the efficient incorporation of alkyne in shorter 400 bp PCR products but less efficient incorporation in 1 kbp lengths. Lanes 1: DNA ladder with numbers on the left indicating selected fragment lengths in base pairs. (c) Absorption spectra of control and alkyne DNA (400 bp) confirms the formation of alkyne-DNA due to the altered shape of the absorption spectrum [171].
pyrimidine ring and results in new absorbance in the the 300–320 nm region and a small shift from the usual 260 nm DNA absorption maximum, as noted previously [171]. The appearance of this absorption shoulder along with DNA absorption at ∼260 nm in the spectrum of the alkyne-DNA confirms the successful incorporation of alkyne into DNA.

To prepare the track to be stretched in any of the flow-based techniques requires ligation to long DNA handles, here implemented using restriction sites. Therefore, the abilities of restriction enzymes to digest the ends of the alkyne-DNA and of ligases to ligate the digested fragments need to be investigated.

In an experiment on 1 kbp products, the alkyne- and control DNA PCR products were digested separately with EagI (NEB, NEBuffer 3) and PspOMI (NEB, NEBuffer 4), respectively, at 37°C overnight. After purification of digestion reactions (PCR purification kit, QIAGEN), products were ligated with a 1:1 molar ratio of alkyne-DNA to control DNA using T4 ligase (Invitrogen) in T4 DNA ligase buffer at room temperature overnight. The ligation products were run in a 1% agarose gel and stained with EtBr (Fig. 3.10(a)). The results demonstrate successful ligation in the appearance of a 2 kbp band.

To clarify whether the 2 kbp band arises from homodimerization or heterodimerization in the ligation products, I did the following experiments. First, the 2 kbp band was excised and gel purified (gel purification kit, QIAGEN), then digested with EagI (NEB, NEBuffer 3) and PspOMI (NEB, NEBuffer 4) in two separate reactions at 37°C overnight, then run in a 1% agarose gel (Fig. 3.10(b)). The results show the PspOMI digestion of a significant fraction of 2 kbp products while incubation with EagI has no noticeable effect. From the band intensity following PspOMI digestion (lane 4 of Fig. 3.10(b)), we concluded that the ligation products comprised mostly homodimers of 1 kbp control DNA. The 2 kbp residual band in the PspOMI digestion reaction presumably contains mostly heterodimers (and possibly a small amount of alkyne-DNA homodimers).

There are different possible explanations for the above observations. If ligation is inhibited by alkyne groups, then ligation would be more likely to proceed with control DNA homodimers. Also the homodimerization of alkyne-DNA might be a rare event due to the difficulty of basepairing two fragments of DNA each with "side chains". Another reason could be that, due to the presence of C8-alkynes around the restriction sites, the EagI enzyme is not able to digest the 1 kbp alkyne-DNA in the digestion reaction prior to ligation to make the sticky ends available.

To avoid the ligation of homodimers of control DNA and to increase the chance of heterodimer formation, control DNA homodimerization was eliminated by dephosphorylation of the PspOMI-
digested control DNA. To do this, after digestion of control DNA with PspOMI (NEB, NEBuffer 4 at 37°C overnight), it was dephosphorylated using Shrimp Alkaline Phosphatase (SAP) (Fermentas, in SAP buffer at 37°C for 1 hour). The reaction was stopped by heating at 65°C for 15 minutes. Products were purified (PCR purification kit, QIAGEN). The ligation reaction was then carried out with a 1:1 molar ratio of alkyne to control DNA using T4 ligase (Invitrogen) in T4 DNA ligase buffer at room temperature, overnight. Ligases were removed (PCR purification kit, QIAGEN), and overnight digestion of ligation products by EagI (NEB, NEBuffer 3) and PspOMI (NEB, NEBuffer 4) enzymes was performed in parallel at 37°C overnight. Figure 3.11 shows the ligation products (lane 1) along with digestion reaction products with PspOMI (lane 2) and EagI (lane 3) restriction enzymes, run in a 1% agarose gel and stained with EtBr. The lower intensity 2 kbp ligation product band compared to the unligated products (lane 1) indicates the low efficiency of the ligation reaction under these conditions.

PspOMI digestion of ligation products is not expected to change either of the bands’ intensities as there should be no control DNA homodimerization. However, an increase in the intensity of the 1 kbp band (and decrease in the 2 kbp band) is expected by EagI digestion if any alkyne-DNA homodimers were present. Comparing lanes 2 and 3 in the gel there is not a significant difference between the intensities of the 2 kbp and 1 kbp bands. Also the band intensities of lanes 2 and 3 are
very similar to those in lane 1. This is likely due to the interference of alkyne groups with digestion and/or ligation reactions.

Due to the observed low efficiency of ligation (higher intensity of the unligated products compared to ligated products in Lane 1, Fig. 3.11), I investigated the digestion efficiency, as the prior step of the ligation reaction, in the presence of alkyne groups. To do this, the primers were redesigned so that the restriction sites were much farther from the incorporated alkyne-dUTP. Table 3.4 lists the new EagI and PspOMI primers. In the PCR, any adenine (A) on the template pairs with an alkyne-dUTP. Therefore, in new primers, no adenine was included in the 6 initial nucleotides (red letters). In addition, the complementary sequence of the primer to the template was selected so that the first adenine lies at least 5 nucleotides 3’ of the restriction site. The new primers are expected to generate 371 and 968 bp amplicons from PCR. I will refer to these also as 400 bp and 1 kbp products, respectively, and use these sequences henceforth.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>EagI primer-N</td>
<td>5’TTG CTT CGG CCG TTG TAC TGA GAG TGC ACC3’</td>
</tr>
<tr>
<td>PspOMI-400N</td>
<td>5’TCT TGT GGG CCC TGT GGA ATT GTG AGC AC3’</td>
</tr>
<tr>
<td>PspOMI-1000N</td>
<td>5’TGT CTT GGG CCC TTT GGA GCG AAC GAC3’</td>
</tr>
</tbody>
</table>

Table 3.4: Sequences of the new EagI site carrying primer (EagI primer-N) along with PspOMI site carrying primers to produce 371 bp (PspOMI-400N) and 968 bp (PspOMI-1000N) DNA PCR products.

PCR reactions with these new primers (IDT) were performed, substituting dTTP with alkyne-dUTP. Due to the low efficiency of the previous PCR reaction with Pwo polymerase to produce
CHAPTER 3. LAWNMOWER: AN AUTONOMOUS MOLECULAR MOTOR

Figure 3.12: The PCR products of control (lane 2) and alkyne-DNA reactions (lane 3): (a) 400 bp with Pwo polymerase and (b) 1 kbp with KOD XL polymerase, run in 1% agarose gels. c) PCR products for 3 kbp control (lanes 2 and 4) and alkyne-DNA (lanes 3 and 5) show the successful production of alkyne DNA when KOD XL polymerase is used (lane 3) but not when Pwo is used (lanes 4 and 5). Lane 1 is 1 kbp DNA ladder (Fermentas) in all pictures.

1 kbp alkyne-DNA (Fig. 3.9(b)), another member of family B polymerases, KOD XL polymerase (Novagen), was tested for amplification of the longer DNA fragment (Appendix F.2).

Figures 3.12(a) and (b) show 400 bp and 1 kbp PCR products, respectively, using the new primers, and run in a 1% agarose gel and stained with EtBr. These results demonstrate successful, efficient incorporation of the alkyne-dUTP into 400 bp and 1 kbp DNA when Pwo and KOD XL polymerase were used, respectively. It should be noted that this formation of alkyne-incorporated DNA was also confirmed by click reaction between alkyne-DNA and azide-flour, using the fluorescence signal as a probe. Detailed discussions on preparation and results of this experiment and characterizations of the fluor-labeled DNA can be found in Section 4.4.

To investigate the digestion and ligation efficiency of these new DNA products, after purification of the PCR products (PCR purification kit, QIAGEN), the 400 bp and 1 kbp control and alkyne-DNA were digested separately with EagI (NEB, NEBuffer 3) and PspOMI (NEB, NEBuffer 4) restriction enzymes at 37°C overnight. After removal of short fragments and restriction enzymes (PCR purification kit, QIAGEN), in another strategy, each reaction was self-ligated using T4 DNA ligase (Invitrogen, T4 DNA ligase buffer, at room temperature overnight). Figure 3.13 shows the successful self-ligation of alkyne-DNA after both EagI and PspOMI digests. The formation of new higher molecular weight bands for alkyne-DNA with intensities comparable to controls confirms the successful ligation and formation of alkyne-DNA homodimers. The formation of homodimers of alkyne-DNA validates the approach of removing alkyne groups from the vicinity of the EagI and PspOMI restriction sites.

Encouraged by this success, the incorporation of alkynes into a 3 kbp fragment of DNA was
Figure 3.13: Alkyne modification does not interfere with digestion and ligation reactions with newly designed primers. (a) Lanes 4 and 5 show the successful self-ligation of 371 bp alkyne DNA after an EagI digest and a PspOMI digest, respectively, which each generate 738bp fragments. Self-ligations of EagI- and PspOMI-digested control DNA (no alkyne modification) are shown in lanes 2 and 3, respectively. (b) Lanes 4 and 5 show the successful self-ligation of 968 bp alkyne DNA after an EagI digest and a PspOMI digest, respectively, which each generate 1936 bp fragments. Self-ligations of PspOMI- and EagI-digested control DNA (no alkyne modification) are shown in lanes 2 and 3, respectively. Lanes 1 contain a 1 kbp ladder, with numbers on the left indicating selected fragment lengths in base pairs.

undertaken. For this, PCR was performed with pUC19 as template, with EagI and PspOMI-3000 primers (Table 3.2) and with Pwo and KOD XL polymerases. Fig. 3.12(c) (lanes 2 and 3) shows success in generating the expected 3 kbp alkyne-DNA products with KOD XL polymerase (Appendix F.2). Unlike with KOD XL polymerase, no alkyne-DNA product was generated under these PCR conditions when Pwo was used.

Formation of this longer alkyne-DNA is a great achievement for the construction of the lawn-mower track, as it eliminates the need for ligation of shorter alkyne-DNA fragments to make a longer track, if such is needed. (The primer sequences would likely need to be modified for ligation to long DNA handles for manipulation.) Also, this is a great experimental success in internal dense functionalization of long DNA. To our knowledge this is the longest fragment of alkyne-labelled DNA to be generated.

3.5.2 Construction of Peptide-DNA

Following preparation of alkyne-DNA and azide-peptide (prepared by Dr. Suzana Kovacic; Appendix E), peptide-DNA was generated by click reactions between the 1 kbp alkyne-labelled DNA and azide-labelled peptide. To do this, 1 kbp alkyne-DNA (PCR purified in 10 mM Tris, pH=8.5) was reacted with azide-peptide (methanol/0.1% acetic acid) in a 1:1 molar ratio of alkyne:azide. The
reaction was catalyzed by the addition of copper (II) sulfate (0.5 mM) and ascorbic acid (0.5 mM) in the presence of 0.5 mM TBTA (tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl] amine) in 47.5% DMSO and 10% methanol. It was incubated in the dark at room temperature for 2 hours with rotary mixing. A control reaction, in which the alkyne-DNA was similarly treated with peptide (no azide) was prepared in parallel. After 2 hours of incubation a loose yellow precipitate could be observed in the reaction tube but not the control tube suggesting that peptide-DNA falls out of solution as a yellow pellet. Both reaction and control products were purified by overnight ethanol precipitation. Ethanol precipitated samples were resuspended in 10 mM Tris, pH=8.5 buffer.

In order to confirm the formation of highly labeled peptide-DNA, the absorption spectrum of the peptide-DNA and its control were measured (Fig. 3.14(b)). The presence of two peaks at 345 nm and 420 nm (Fig. 3.14(b), inset) due to methoxycoumarin and dinitrophenyl (Fig. 3.14(a), spectrum of peptide-azide) [176], respectively, as well as the 260 nm DNA absorption peak confirms the incorporation of the peptide onto DNA. The absorbance scan of the control shows only the absorption spectrum of alkyne-DNA.

In general, absorption readings can be used to determine DNA and peptide concentrations. Therefore, one might consider estimating the amount of peptide labelling per DNA using these absorption peaks. However, here, experimental complications make this quantitative calculation impossible. For example, click chemistry components that may precipitate also absorb near 260 nm (the DNA peak). The copper(I)-stabilizing ligand tris(benzyltriazolylmethyl)amine (TBTA) [177] is used to prevent oxidative damage to preserve DNA from damage caused by copper. TBTA precipitates along with DNA in the ethanol precipitation of click and control reactions and absorbs at 260 nm. Thus, it is not possible to measure DNA concentration using absorption at 260 nm. In addition to this, due to the insolubility of the peptide-DNA and therefore its aggregation, a scattering peak is observed that superimposes on the peptide-DNA absorbance spectrum and precludes accurate estimation of peptide concentration using the 345 nm and 420 nm absorption peaks.

It is important to note that, assuming a complete degree of 1:1 labelling, 500 peptides per DNA would produce a lawnmower track with peptide density far greater than that desired for lawnmower motility. This density is equivalent to one peptide per 2 bp or 0.68 nm. As mentioned previously (Section 3.4), a peptide density of about 1 per 20 bp is desired. However, to characterize the track we focused on the densely labeled DNA so as to maximize signal from the peptide fluorophores.

To compare the mobility of this highly labeled peptide-DNA construct with control DNA, both were run in a 1% agarose gel and stained with EtBr (Fig. 3.15). The result showed the expected
Figure 3.14: (a) Absorption spectrum of 6.5 μM peptide in 10 mM Tris buffer (pH=8.5) shows absorption peaks of MCA (fluorophore) and Dnp (quencher) in the peptide at about 345 and 420 nm, respectively. (b) Presence of the same absorption peaks along with that for DNA at 260 nm in the peptide-DNA reaction spectrum confirms the conjugation of peptides to alkyne-DNA. Control for which no conjugation was expected shows only the alkyne-DNA absorption spectrum. Inset: enlargement of the region around the peptide features.
There are different possibilities to explain the disappearance of the highly labeled peptide-DNA in the agarose gel. One hypothesis is that the peptides on the heavily labeled DNA might interfere with binding of EtBr to the DNA backbone. This has been previously reported for highly Cy-dye labeled DNA [178]. Another possibility could be the stochastic peptide labelling of alkyne-DNA so that there are different number of peptides on each DNA molecule. This would result in production of peptide-DNA molecules with a distribution of molecular weights which each run differently on a gel. Thus, the labeled DNA products will appear as a smear, rather than a unique band on a gel. Poor solubility could be another reason for the observed lack of a mobile band of highly labeled 1kbpeptide-DNA. In different trials of this experiment, a dark band was always observed in the well of agarose gels. This could be the highly labeled 1 kbp peptide-DNA, if it aggregates and thus cannot enter the pores of the gel.

In addition to the observed dark band in agarose gel wells, the hypothesis of poor solubility of the peptide-DNA strengthened when our attempts to purify the peptide-DNA by size exclusion
chromatography and spin filtration under aqueous conditions failed. Besides the issues it causes for characterizing the peptide-DNA, decreased solubility would also affect the digestion of peptide substrates by trypsin as they would not be easily accessible for the digestion reaction. In the following, our experimental efforts to improve peptide-DNA solubility and its characterization are explained.

3.5.3 Experiments to Improve Peptide-DNA Solubility

Solubility of Peptide-DNA in Organic Solvent

One strategy to overcome the solubility problem involves dissolving the peptide-DNA in an organic solvent. To try this, after ethanol precipitation of the click reaction product as above, the peptide-DNA was resuspended in 100% DMSO. Absorption of peptide-DNA was measured and monitored as a function of time (Fig. 3.16(a)) to investigate if peptide-DNA precipitates out of this solution. The measurements showed a decrease in the intensity of DNA and peptide peaks after 5 minutes, indicating that peptide-DNA falls out of solution.

Attempts to determine the peptide-labelling efficiency by comparing relative absorbances of the DNA and peptide peaks proved difficult. Again, this is in part due to a scattering peak superimposed on the peptide-DNA absorbance scan, presumably due to aggregation of peptide-DNA. In addition, the DMSO absorbance at 260 nm makes quantification of DNA difficult. Reaction efficiency was therefore estimated by comparing the amount of unreacted azide-PEG4-peptide remaining in the ethanol supernatant following precipitation of the reaction sample to that remaining in the control sample (Fig. 3.16(b)). Of the 500 alkyne groups in a 1 kbp fragment of alkyne-DNA, approximately 300 are estimated to have reacted with azide-PEG4-peptide.

One of our aims in characterizing the peptide-DNA construct was to investigate if DNA and peptide each retained functionality in the peptide-DNA hybrid construct. To demonstrate peptide functionality, we examined the ability of trypsin to hydrolyze densely labeled peptide-DNA.

The ability of free trypsin to proteolyze peptide-DNA was evaluated by adding trypsin to peptide-DNA and monitoring proteolysis by the increase in methoxycoumarin fluorescence. The reaction was performed in 30% DMSO in 10 mM Tris pH 8.5. Here, peptide-DNA solubility was further compromised by the use of 30% DMSO (Fig. 3.17) but was necessary to retain trypsin activity. The amount of trypsin added was sufficient to immediately proteolyze control samples of free azide-PEG4-peptide. Surprisingly, trypsin was able to hydrolyze peptide-DNA, although at a slower rate than the free azide-PEG4-peptide (Fig. 3.18).
Figure 3.16: (a) Absorption spectra of peptide-DNA in 100% DMSO show the precipitation of peptide-DNA over time. (b) Absorption spectra of unreacted azide-PEG4-peptide remaining in the ethanol supernatant following precipitation, in control and reaction samples. (Courtesy of Dr. Suzana Kovacic.)
Analysis of these results shows the degree of labelling of about 80 peptide per DNA. Assuming no loss of DNA throughout the reactions (∼5 pmol), 400 pmol peptide in the reaction is equivalent to about 80 peptides per DNA. This degree of labelling is far below stoichiometric levels (∼500 peptides). It is also lower than our previous absorbance-based estimate of 300 peptide/DNA. Therefore, there may be either a significant loss of DNA in ethanol precipitation, or more peptides bound to DNA in the click reacted sample which are not accessible to trypsin. Poor digestion/accessibility could arise from poor solubility, the close proximity of the peptide to the DNA molecule and/or the high density of peptide packing on the DNA molecule. Although digestion of peptide was observed in these peptide-DNA samples, it could be that still longer incubations would generate more product, if peptides attached are simply cleaved with a much slower reaction rate.

Although solubility remains an issue, this very promising result indicates the ability of trypsin to digest the peptide-DNA click reaction products. Encouraged by trypsin digestion results, another experiment to improve the peptide-DNA solubility was performed, which is explained in the

Figure 3.17: Absorption of peptide-DNA in 30% DMSO decreases as a function of time, indicating the peptide-DNA is falling out of solution. (Figure courtesy of Dr. Suzana Kovacic.)
CHAPTER 3. LAWNMOWER: AN AUTONOMOUS MOLECULAR MOTOR

Figure 3.18: Increase in the MCA fluorescence signal for the peptide-DNA sample shows that peptide substrate in the highly labeled 1 kbp peptide-DNA can be digested by trypsin in 30% DMSO. (Figure courtesy of Dr. Suzana Kovacic)

following section.

Increase the DNA Characteristics of Peptide-DNA

In another attempt to overcome the solubility problem, I ligated a 3987 bp (∼4 kbp) DNA handle to the 1 kbp peptide-DNA via one of its flanking restriction sites. I reasoned that this might provide sufficient DNA character to the construct to increase its solubility in aqueous solution.

The 4 kbp linear DNA was first produced by digestion of 4179 bp pK₈ plasmid (an available plasmid in our lab) with EagI and FspI restriction enzymes. To do this, first, pK₈ plasmid (an available plasmid in our lab; see Section 4.3) was digested with EagI (NEB, NEBuffer 3, 37°C overnight). After purification (PCR purification kit, QIAGEN), it was digested with FspI (NEB, NEBuffer 4, 37°C overnight). These reactions produce two fragments (3987 and 192 bp) from which the longer band was purified by running the products in agarose gel electrophoresis, excising the ∼4 kbp band from the gel, and purifying the DNA from the gel (gel purification kit, QIAGEN). The purified 4 kbp fragment is blunt-ended on one end (FspI digest) and carries an EagI overhang
on the other end. This fragment was dephosphorylated by using SAP (Fermentas, in SAP buffer at 37°C for 1 hour), then PCR purified (PCR purification kit, QIAGEN) to be used in the subsequent ligation step. In parallel, the 1 kbp alkyne-DNA was digested with PspOMI enzyme (NEB, NEBuffer 4 at 37°C overnight). A ligation reaction between 4 kbp DNA handle and digested 1 kbp alkyne-DNA was performed with T4 ligase (Invitrogen, T4 DNA ligase buffer) overnight at room temperature. In order to eliminate self-ligation of PspOMI-digested alkyne-DNA products, the ligation was performed in the presence of PspOMI restriction enzyme in the reaction. The ligation product was PCR purified (PCR purification kit, QIAGEN). The click reaction between azide-peptide and ligation product was performed with a 1:1 molar ratio of azide-peptide to alkyne groups on the alkyne-DNA. A control ligation was also performed with peptide (without azide) and ligation product in the presence of all other click chemistry components. After 2 hours of incubation, samples were ethanol precipitated, resuspended in 10 mM Tris-HCl, and their UV-Vis absorption spectra were measured (1700 UV-Vis Spectrometer, Shimadzu) (Fig. 3.19(a)). As before the peptide-DNA formation was confirmed by the presence of two peaks at 345 nm and 420 nm (methoxycoumarin and dinitrophenyl, respectively) as well as the 260 nm peak of DNA. The absorbance scan of the control reaction shows only the 260 nm peak due to DNA.

The control and reaction products, then, were run in a 1% agarose gel (Fig. 3.19(b)). The control lane (lane 2) showed three bands corresponding to 1 kbp alkyne-DNA, 4 kbp DNA handle and the 5 kbp ligation product. To our surprise, both bands containing alkyne-DNA (1 kbp and 5 kbp) disappeared following reaction with azide-peptide (lane 3). Disappearance of the 1 kbp peptide-DNA band was expected based on previous experiments. However, it was hoped that the 4 kbp handle would improve the solubility, permitting migration into the gel and staining by EtBr due to its intact DNA handle. A dark band was seen in the well of this gel again.

To investigate trypsin digestion of this construct, the long handle peptide-DNA product was reacted with trypsin and probed for fluorescence signal from MCA fluorophore. To do this, in separate reactions, to 80 µl each of peptide-DNA and its control in 10mM Tris pH 8.5, 50 µg of trypsin was added. As controls, 100, 250 and 400 pmol azide-PEG4-peptide were treated with the same amount of trypsin. The reactions were mixed on a 96-well plate and excited at 360 nm while monitoring for fluorescence at 400 nm (BioTek Synergy fluorescence plate reader). The signal from digestion of known amounts of peptide was used as a reference for later comparison to estimate the amount of peptide in the click reaction product. Figure 3.20(a) shows the increase in the fluorescence signal of all peptide samples including the peptide-DNA click products. The
peptide-DNA shows a very weak signal, although greater than control, which merges toward the 100 pmol azide-peptide control sample.

Assuming no loss of DNA throughout the reactions (~1.6 pmol), 100 pmol peptide in the reaction is equivalent to about 63 peptides per DNA. Here again, this degree of labelling is far below stoichiometric levels (~500 peptides), which could be the result of poor digestion/accessibility arising from the close proximity of the peptide to the DNA molecule, the high density of peptide packing on the DNA molecule, and/or the insolubility of the construct as discussed above for digestion of peptide-DNA in 30% DMSO.

The presence of fluorophore and quencher absorption peaks in peptide-DNA absorption spectra, and the observed increase in the fluorescence signal after its digestion with trypsin are promising results toward the construction of lawnmower track. One confirms conjugation of peptide to DNA.
Figure 3.20: Digestion of the peptide-DNA with long handle by an excess of trypsin (in 10mM Tris pH 8.5), along with 100, 250 and 400 pmol peptide-azide controls. The fluorescence intensity of the long peptide DNA digested with trypsin approached that of 100 pmol azide-peptide.

and the other confirms the ability of trypsin to digest the peptide attached to DNA. In spite of these results, the problem of poor solubility of peptide-DNA remains to be solved before lawnmower assays become a reality.

As a final experimental approach to characterize the peptide-DNA, we used Atomic Force Microscopy (AFM). We hoped that this would provide the necessary tool to observe the click chemistry products at the single molecule level and directly demonstrate formation of the peptide-DNA construct. In addition, we hoped that observations would shed light on the insoluble products (their formation and structure i.e. aggregates or networks) to help in trouble-shooting solubility issues.

3.5.4 Characterization of Peptide-DNA Using Atomic Force Microscopy

Here, our approach was to use Atomic Force Microscopy (AFM) imaging to characterize the click reaction products using information about DNA contour and persistence lengths and height. We expected that conjugation of peptide to DNA would result in an increase in the height of DNA when
it is deposited on the surface compared to unconjugated DNA controls. Details of experiments are as follows.

While I did the sample preparations, AFM imaging (tapping mode, Asylum Research) in air and shape fluctuation analysis of images were performed on samples by Dr. Guillaume Lamour in Dr. Hongbin Li’s Laboratory at University of British Columbia.

To generate the imaging controls, 1 kbp control DNA was deposited on mica substrates. To do this, the DNA needed to be in an appropriate buffer ("AFM buffer"), which contains 10 mM Tris, 2 mM NaCl and 2 mM MgCl$_2$ at pH=8. The 1 kbp control DNA (420 ng/$\mu$l) was then diluted 10$^4$ fold in the AFM buffer. 95 $\mu$l of the sample was deposited onto a mica substrate and incubated for 10 minutes. Because DNA and mica are both negatively charged, it is necessary to modify the mica surface or add counterions to allow binding. Experimentally, the counterion method was performed by adsorbing DNA onto the mica in the presence of Ni$^{2+}$ cations. Therefore, 5 $\mu$l of 50 mM NiCl$_2$ was added, gently mixed with AFM buffer on the mica surface and incubated for 1 minute. Then, the sample was rinsed with pure water and dried by flow of nitrogen.

Figure 3.21(a) shows an AFM image of the control DNA. Analysis of these images resulted in an average contour length of 294$\pm$18 nm and a persistence length of 50$\pm$5 nm for DNA ($N$=118).

Following this experiment, as another control, 1 kbp alkyne DNA was deposited on mica using the same protocol and imaged (Figure 3.21(b)). From AFM image analysis, the contour length of the alkyne DNA was calculated to be on average 288$\pm$26 nm and the persistence length was 53$\pm$8 nm ($N$= 92).

The estimated contour lengths of both controls are slightly shorter than expected (329 nm for 968 bp). This likely arises from the manual nature of end selection in the analysis for which an error of $\pm$ 30 nm is added [Dr. Guillaume Lamour, personal communication].

The persistence length of DNA is about 50 nm [150], as found for both DNA samples. We had anticipated that incorporation of alkyne groups might alter the flexibility of DNA due to steric and hydrophobic effects. However, dense alkyne incorporation does not change the measured persistence length.

Encouraged by successful analysis of the controls, the same deposition protocol was used to deposit the click reaction products on the mica surface. Unlike controls, the AFM images of these samples did not show any trace of DNA or DNA-like structure on the surface. This may be a result of poor solubility of peptide-DNA and its likely removal from the surface in washing and drying steps.
Figure 3.21: AFM images of the 1 kbp DNA controls: (a) plain DNA, and (b) alkyne-DNA. AFM image analysis estimated average contour lengths of 294±18 nm and 288±26 nm and persistence lengths of 50±5 nm and 53±8 nm for the samples in (a) and (b), respectively. These results are calculated based on analysis for $N=118$ and $N=92$ control DNA and alkyne-DNA, respectively. (Figures are courtesy of Dr. Guillaume Lamour.)
A possible strategy to improve deposition of peptide-DNA would be, rather than dilution of the click reaction product in the AFM buffer, to deposit it on mica directly while it is in DMSO and let the sample dry. The problem with this approach is that DMSO evaporates very slowly at atmospheric pressure, and handling the mica substrate in vacuum to evaporate DMSO is not easy to do and may remove the peptide-DNA.

Suggested by protein experts in our group, we explored resuspending the click reaction products after ethanol precipitation in chloroform or in a mixture of 50% ethanol : 50% chloroform and directly depositing the samples onto mica. Chloroform is also an organic solvent and vaporizes very quickly at atmospheric pressure.

To try this, two samples each of low- and high-labeled peptide which were prepared already in DMSO were ethanol precipitated. The low-labeled sample was prepared by click reaction between 79 pmol of peptide-azide and 1.5 pmol of alkyne-DNA in a 30 µl reaction volume following the previously explained steps for preparation of high-labeled peptide-DNA. It is expected that on average 50 peptides react with the 1 kbp DNA. After ethanol precipitation, from each low- and high-labeled samples one was resuspended in 50% ethanol : 50% chloroform and one in chloroform. (All procedures involving chloroform were performed in a fume hood.) Our control experiments depositing just the solvents (chloroform and 50% ethanol : 50% chloroform) on mica substrates showed that chloroform reacts with mica and produces structures on the surface. Therefore, mica was replaced with glass substrates for preparation of these samples. The resuspended click reaction products in each solvent were deposited directly on different glass substrates. After complete evaporation of solvents and a final drying step using a flow of nitrogen, the AFM measurements were performed. Imaging the samples, nothing was observed on the surface of any sample except the low-labeled peptide DNA in 50% ethanol : 50% chloroform (Figure 3.22).

Unexpectedly, long fiber-like structures were observed on the substrate surface in this sample which are considerably longer (∼1 µm) than the 300 nm expected length of peptide-DNA. It would be interesting to investigate the observed structures and characterize them further. However, this is beyond the scope of this thesis.

In the next section, some suggestions to pursue characterization of the peptide-DNA click products and to overcome the solubility issues to construct lawnmower track in DNA curtains setup are provided.
Figure 3.22: AFM images of the 1 kbp low-labeled peptide-DNA in 50% ethanol : 50% chloroform and the height analysis of the observed structures. There are fiber-like structures on the surface with longer lengths than the expected ∼300 nm peptide-DNA length. (Figure courtesy of Dr. Guillaume Lamour.)
3.6 Conclusions and Future Directions

In this chapter, I introduced the lawnmower as a protein-based autonomous molecular motor. As explained, the lawnmower is designed to use digestion of its substrate to bias its diffusion. Here, using results of our investigation of biased motion of molecular spiders (Chapter 2), lawnmower molecular components were selected so that they lead us toward the construction an autonomous molecular motor.

Using these components, a lawnmower was constructed on a 20 nm QD hub with trypsins covalently attached using 1 nm linkers. Using fluorescence assays, the results to demonstrate the successful construction of an active lawnmower were presented. It was estimated that lawnmowers have on average eight active blades.

Tracks for the lawnmower were then designed, to be made of peptide substrates arranged on DNA as a one dimensional track. To construct the lawnmower track, click chemistry was employed to covalently attach azide-labeled peptides onto alkyne-functionalized DNA. The size of alkyne-DNA used for track experiments was 1 kbp, however, we were also able to successfully construct a longer 3 kbp internally labeled alkyne-DNA using PCR with KOD XL polymerase. Here, internally quenched peptides were selected as trypsin substrates so that the digestion reaction generates fluorescence signal, useful for future detection of lawnmower activity.

Successful conjugation of peptides to 1 kbp alkyne-DNA track was shown using absorption spectroscopy. In addition, it was shown that free trypsin is able to digest the peptide substrate when it is attached to a DNA backbone, although with a slower rate than digesting free azide-peptide substrate. These are very promising results to follow up with construction of lawnmower track using the explained method.

For the final goal of observing biased motion of the lawnmower, its activity must be checked on the peptide’s substrate when it is attached to the DNA backbone. At the present stage, we showed that the lawnmower can cleave free substrate, and free trypsin is able to digest the peptide-DNA track. The quantities and cost of reagents preclude bulk measurements of lawnmower activity on its peptide-DNA lawn. Therefore, single-molecule efforts are currently underway to investigate the digestibility of the peptide-DNA substrate by the lawnmower. This is planned to be done by our collaborators in Lund via TIRF microscopy, and monitoring the QDs along with appearance of the MCA fluorescence signals as lawnmowers interact with their tracks, initially nonspecifically deposited on glass.
In spite of the above achievements, our attempts to characterize the constructed peptide-DNA either by running in a gel or by the single-molecule imaging technique of AFM did not help us to learn about the structure of the formed product. We think this could be due to the poor solubility of the peptide-DNA. Two strategies to improve the solubility of the peptide-DNA either through using organic solvents or attachment of long DNA handles to peptide-DNA were employed, neither of which significantly improved solubility.

However, to be used in DNA racks the peptide-DNA should be extended to 10-15 µm length [90, 91]. This is a much longer handle than what we used in the DNA-peptide experiments with DNA handle. In addition, it should be noted that for the lawnmower function, a lower density of peptide-DNA is eventually required. The presence of such long handles and a lower density of peptides on the alkyne-DNA fragment may resolve the problem of solubility.

In addition, to ligate the lawnmower track to long handles, it may be necessary to perform the click chemistry reaction between peptide-azide and alkyne-DNA after ligation of alkyne-DNA to handles. From our findings on the effect of alkynes on the efficiency of digestion and ligation reactions, I think a larger side chain on DNA, such as peptides, would interfere with these enzymes’ functions. Therefore, one can envision performing click reactions even after tethering the two ends of the long DNA fragment with embedded 1 kbp alkyne-DNA in DNA racks. I hope this would prevent collapse of the peptide-DNA out of solution due to poor solubility.

If the suggested directions result in a soluble, well characterized peptide-DNA track, after adjustment of peptide density, the activity of the lawnmower in the DNA rack setup can be investigated. Figure 3.23 shows a cartoon of a lawnmower in a DNA rack setup. The lawnmower activity on stretched 1D tracks in proximity to the surface can be monitored using TIRF microscopy. A region of the DNA handles of the track can be fluorescently labeled so that the QD hub displacement is measured with respect to a reference. (See Section 4.4 for construction of such a fiducial marker.)

The current lawnmower uses 1 nm flexible linkers. As discussed in Section 3.2.2 it is likely advantageous to use the longer 10 nm PEG linkers in the ultimate lawnmower. Due to the higher cost of these linkers, the current lawnmower pre-production prototype used the 1 nm linker in establishing these assays, however it should be straightforward to repeat this synthesis with the 10 nm linker using the same chemistry, for observation of the lawnmower function on 1D tracks.
Figure 3.23: A schematic of the lawnmower on a stretched DNA in a rack setup (not to scale). Digestion of the peptide substrate by lawnmower blades, along with motion of the lawnmowers, can be monitored using TIRF microscopy.
Chapter 4

Design and Construction of a DNA Track for the Tumbleweed Motor

4.1 Introduction

In this chapter, the design and construction of a periodic one-dimensional DNA track for a novel artificial molecular motor, the tumbleweed (TW) [179, 180], is described.

The TW is a synthetic protein motor designed to step unidirectionally along a DNA track. The TW comprises a coil-coiled Y shape central hub connecting three different DNA repressor proteins: \( R_A \) (methionine repressor MetJ (Q44K) [181, 182]), \( R_B \) (tryptophan repressor TrpR [183, 184]), and \( R_C \) (purine repressor PurR [185, 186]) (Fig. 4.1). Each repressor acts as a "foot" of the motor that can be made to bind and unbind from its well-defined recognition sequence on a DNA track, depending on the concentration of its specific ligand present in solution [187, 188, 189]. Ligands (a: S-adenosyl methionine, b: L-tryptophan, and c: hypoxanthine or guanine) induce the binding of \( R_A \), \( R_B \) and \( R_C \) to their corresponding DNA binding sites. Specific target sites (represented by A, B, C) that are the repressors’ binding sites should be spaced and ordered appropriately along a synthesized DNA track. (a, b, c) represent ligands for binding of repressors A, B and C respectively to their binding sites.

The TW diffusively steps "hand-over-hand" along the track by cyclically binding its three motor feet (Fig. 4.2). Directed, processive tumbleweed-like motion is expected to be induced by externally supplying ligands (a, b, c) in the temporal order [a, b], [b, c], [c, a], such that at all times at least one
Figure 4.1: A model of the conceived tumbleweed motor. The TW consists of three unique DNA-binding proteins called repressor proteins, which each bind with high affinity to a specific sequence of DNA in the presence of their respective small molecule ligands in solution, attached to the "arms" of a Y-shaped coiled-coil protein hub. (Structure file for this image created by Drs. Richard Sessions and Elizabeth Bromley [190]).

foot is bound to the track. For instance, if the motor is initially bound to the DNA track with foot $R_A$, in the presence of ligand pulse $[a, b]$, $R_B$ can bind to the adjacent site B (due to the chosen sizes of track and motor). In this state $R_C$ is unbound and undergoes confined diffusion. During the subsequent pulse $[b, c]$, $R_A$ unbinds from site A as it is no longer ligand-bound to the DNA. Meanwhile, $R_B$ stays bound to site B while $R_C$ eventually locates and binds to the downstream site C, as dictated by the directionality of the track, relative size of the motor and presence of associated ligand c in the media. Then, ligand pulse $[b,c]$ is replaced with $[c,a]$ so that $R_B$ detaches from the track while $R_C$ remains bound. In this state, $R_A$ diffusively searches for its binding site and eventually binds specifically to the DNA. Introduction of pulse $[a,b]$ results in detachment of $R_C$ while $R_A$ remains bound and brings the system back to the original state for the next ABC binding cycle down the track.
The Tumbleweed: towards a synthetic protein—based molecular motor derived from an engineered, self-assembled protein complex designed to move by cyclically ligand-gated, rectified, processive tumbleweedlike motion is expected to be achieved via a self-assembled motor composed of three motifs, cooperatively designed to occupy a DNA track. This motor will be a self-assembled protein complex comprising a rigid Y-shaped coiled-coil protein hub and flexible linkers as the ankle joints, with a DNA track that is self-assembled and functional. The motor will utilize the DNA track as a track specificity and with appropriate flexibility. This is followed by remarks on design criteria for the DNA track.

SYNTHESIS ROAD MAP

First, we identify "feet"—ligand-gated, DNA-binding protein domains—that interact specifically and controllably with novel domains that specify temporal order [a, b, c], [b, c], [c, a], such that at all times at least one domain is bound to the track. For instance, if the motor is configured and with appropriate flexibility. This is followed by remarks on design criteria for the DNA track.

To illustrate that the synthesis of an artificial molecular motor is feasible with the current state of the art, we develop a road map for the experimental realization of a TW motor. Specifically, we discuss the design of a "hub" and foot modules, and the assembly of these subunits into a functional motor. We also present a computational approach that combines protein domains that possess the ability to bind the track selectively, and combine them with novel domains that specify temporal order [a, b, c], [b, c], [c, a], such that at all times at least one domain is bound to the track. For instance, if the motor is configured and with appropriate flexibility.

In summary, the key design features of the TW are three discrete protein domains that operate cooperatively to move along the DNA track. These domains are repressors that each act as a "foot" of the TW that can be made to bind and unbind from its respective well-defined recognition sequence A, B or C on the DNA track, depending on the concentration of its specific ligand (a, b and c) present in solution. See text for details of the TW stepping.

Figure 1: Illustration of the basic concept of the Tumbleweed. Figure from Ref. [179]. R_A, R_B and R_C are repressors that each act as a "foot" of the TW that can be made to bind and unbind from its respective well-defined recognition sequence A, B or C on the DNA track, depending on the concentration of its specific ligand (a, b and c) present in solution. See text for details of the TW stepping. The circular hub represents the rigid Y-shaped coiled-coil protein hub and flexible linkers are the ankle joints.

Figure 4.2: Illustration of the basic concept of the Tumbleweed and its stepping mechanism. Figure from Ref. [179]. R_A, R_B and R_C are repressors that each act as a "foot" of the TW that can be made to bind and unbind from its respective well-defined recognition sequence A, B or C on the DNA track, depending on the concentration of its specific ligand (a, b and c) present in solution. See text for details of the TW stepping. The circular hub represents the rigid Y-shaped coiled-coil protein hub and flexible linkers are the ankle joints.
CHAPTER 4. DESIGN AND CONSTRUCTION OF A DNA TRACK FOR THE TW

The tumbleweed molecular motor was inspired by classes of linear molecular motors which travel unidirectionally along tracks. Myosins, kinesins and dyneins are three classes of linear motors in the cell. As explained in Chapter 1, myosins walk on actin filaments, while kinesins and dyneins walk in opposite directions on microtubules. These molecular motors are powered with a single type of fuel molecule, ATP. In contrast, the stepping cycle of the TW requires three different chemical fuels and walks on a DNA track.

Unlike the lawnmower, which is an autonomous molecular motor, the TW is gated through externally controlled and timed supply and removal of ligands from solution. Because its repressor proteins each bind to a specific DNA sequence only in the presence of their associated ligand, they make perfect candidates for the feet of this motor, since the spatial and temporal coordination of binding can be controlled by DNA design and the local chemical environment.

As part of an international collaboration, our group is responsible for the design and construction of the TW DNA track. Containing a stretch of repetitive A, B, C binding sequence, it will eventually be used in a DNA curtains/racks or DNA tightropes setup (Section 1.4). In this chapter, I first discuss the importance of the design parameters of the TW track (Section 4.2). Then the experimental methods for constructing a long DNA track containing a periodic sequence of repressor binding sites are presented (Section 4.3). To follow the progress of the TW motor along the track, the presence of a fiducial marker as a reference along the track is required. In Section 4.4, experimental methods to construct a fluorescent fiducial marker are described.

4.2 Design of TW DNA Track

4.2.1 Design Parameters

As shown in Fig. 4.1, the key functional components of the TW are three distinct ligand-dependent DNA-binding repressors (R_A, R_B, R_C). Each repressor protein "foot" binds with high affinity to a unique double-stranded DNA (dsDNA) recognition sequence in the presence of its specific ligand (a, b or c, respectively). Thus, the main concern in designing the TW track is to include the three unique repressor motifs in the dsDNA track. In addition, in order to demonstrate motor motion over significant distances, the track needs to be highly repetitive so that processive motor stepping can be achieved.

The feet of the TW are attached to the hub by means of flexible linkers or "legs", such that the
motor can reach the next recognition site on the track and take steps of appropriate size (Fig. 4.1). On the size scale of the TW, DNA is not a featureless molecule but presents two distinct sides (grooves) that twist helically around the length of the molecule with an $\sim$10.4 bp pitch [191]. Ideally, one should ensure that all three recognition sites present on the same side of the DNA molecule to avoid possible steric hindrance caused by the motor "corkscrewing" around the DNA track instead of stepping along it. This is especially a concern if the track is on or close to a surface as is typically the case for single-molecule tracking experiments [97].

The proper sequences, spacing, and spatial orientation of the recognition sites on the track are thus important requirements for the design of the DNA track.

Each repressor recognition sequence is approximately the length of one helical turn of DNA ($\sim$3.5 nm) [179]. To reduce steric interactions between bound repressors, the track is designed such that there is an inactive sequence spacer, approximately the length of two helical turns, between the active binding sequences. Each binding sequence can then be considered a unique binding site separated by approximately 11 nm from adjacent steps. To build asymmetry into the DNA track, the three unique binding sites, A, B, and C, are arranged in repeating sequences of A_B_C_A_B_C... , where "_" represent the inactive spacer sequences.

Experimentally, depending on the observation technique the end(s) of the DNA construct need to be modified so that it can be anchored in a stretched configuration and form a linear track along which the movements of the TW motor can be observed. Figure 4.3 shows a schematic of the TW and its track in a DNA curtains setup, in which the biotinylated DNA track is stretched out under flow. A further challenge is presented by the unavoidable drift of the sample on the microscope during the single-molecule visualization of tumbleweed dynamics. As such, the DNA track should incorporate a fluorescent fiducial marker to provide a readout of this drift. This allows us to measure the position of the TW relative to the track, thus removing any effect of the movement of the whole sample during the experiment. Lastly, the track should be designed in a way so that it can be constructed and modified easily and quickly.

### 4.3 DNA Track Construction

Considering the above parameters, primer sequences carrying the A, B and C recognition sites and the spacing primers were developed by Dr. Gerhard Blab using a genetic algorithm [192]. A, B, and C correspond to the repressor binding sites for MetJ, TrpR, and DtxR, respectively. Plasmids,
named pK$_N$, containing different (ABC)$_N$=K$_N$, N-repeating units (N=1, 2, 4, 8, 16 and 32) of the ordered recognition sequences A, B, and C, were constructed in high copy-number bacterial pYIC plasmids [192]. The 11 nm step size of the TW motor translates into a repeating ABC unit which is 103 bp long. Here, I focus on my work which involved the incorporation of the K$_8$ sequence into a longer DNA construct.

For the initial experiments of construction of a long TW track, the pK$_8$ plasmid was utilized, as it was determined to be stable in the chosen bacterial system and at the same time contains enough repeats of the ABC unit to enable observation of multiple steps by the TW motor. We propose to use DNA curtains (Section 1.4) [90] and to stretch many DNA molecules in parallel and observe movement of TW motors along the DNA tracks (Fig. 4.3). The K$_8$ motif, however, is too short to be stretched as a DNA curtain, as this typically requires DNA which is more than 10 µm long [90].

Two strategies were tested to elongate the repeating segment of the pK$_8$ plasmid, K$_8$, to make it applicable in DNA curtains. Based on the specifications provided by our collaborators at Lund University, for experimental implementation in DNA curtains, the K$_8$ unit should be located about
CHAPTER 4. DESIGN AND CONSTRUCTION OF A DNA TRACK FOR THE TW

Figure 4.4: Positions of recognition sites of RsrII and EagI restriction enzymes in the pK₈ plasmid. RsrII and EagI digest pK₈ 1708 bp (580 nm) downstream and 1025 bp (348 nm) upstream of the K₈ unit, respectively.

300-500 nm from the biotinylated end of the DNA, and the total length of the DNA in which the TW track is embedded needs to be ≥ 10 µm. In the first method the K₈ track was lengthened from about 280 nm (824 bp; each ABC unit is 103 bp long) to 14 µm by attaching long DNA handles derived from Lambda DNA to the end of the track. In the other method, I took advantage of bacteria to do elongation and amplification of a K₈-containing plasmid to produce a 12 µm (35.8 kbp) DNA fragment. Details of experiments performed for these two strategies are explained in the following subsections.

4.3.1 Biotin-DNA TW Track Construction Using Lambda DNA

pK₈ is a 4179 bp (~1.4 µm) plasmid. Therefore, the plasmid itself can be used to provide the 300-500 nm spacer between the K₈ unit and the biotin label. Figure 4.4 shows the position of a unique RsrII cleavage site on the pK₈ plasmid with respect to the K₈ unit. The RsrII enzyme cuts the plasmid 1707 bp (580 nm) from the (ABC)₈ unit and produces the required spacer. The overhang generated by RsrII digestion was subsequently biotinylated, as described later in this section.
CHAPTER 4. DESIGN AND CONSTRUCTION OF A DNA TRACK FOR THE TW

Figure 4.5: Schematic of the ligation reaction between the pK₈ and Lambda DNA fragments (not to scale). The EagI overhang (blue sequence) of the pK₈ plasmid fragment ligates to the PspOMI overhang (purple sequence) of the Lambda DNA fragment. The sequences of the RsrII overhang (pK₈ fragment) and the 12 nucleotides of Lambda DNA overhang are coloured in red and brown, respectively.

To lengthen the DNA track, here Lambda DNA was used. Lambda DNA is approximately 16 µm long (48502 bp), contains a number of restriction sites which facilitate the insertion of the TW DNA track, and, importantly, does not contain repressor binding sites which could compete for binding with the TW motor. Using Geneious software [193] the sequences of Lambda DNA and the pK₈ plasmid were investigated to find a restriction site in each that produce complementary overhangs. The chosen restriction enzymes should not digest the pK₈ plasmid in the repetitive ABC binding region. Also they should not cut the Lambda DNA into many small fragments but leave a reasonably long fragment that can be ligated as a handle to the TW track. I found that EagI digestion of pK₈ (Fig. 4.4) and PspOMI digestion of Lambda DNA could be used to generate complementary overhangs (Fig. 4.5).

Following these investigations in Geneious, experimentally, the pK₈ plasmid was first digested with RsrII restriction enzyme (NEB, NEBuffer 4 at 37°C overnight). After purification (PCR purification kit, QIAGEN), the linearized plasmid was digested with EagI restriction enzyme (NEB, NEBuffer 3, 37°C overnight). Digestion of the plasmid with RsrII and EagI generated two DNA fragments (3650 and 609 bp in length) of which the longer one was the desired fragment. Thus following digestion, the products were run in a 1% agarose gel (Fig. 4.6(a)) and the 3650 bp band was excised and purified (gel extraction kit, QIAGEN).

The complementary overhang to the EagI sticky end was generated by PspOMI cleavage (NEB, NEBuffer 4 at 37°C overnight) of Lambda DNA. This generated two fragments (38412 and 10090 bp), of which the 38 kbp fragment was used. Figure 4.6(b) shows the digestion products of Lambda DNA by PspOMI after running in agarose gel (low melting point (LMP), Sigma). The higher molecular weight band in the gel was excised and purified (GELase, Epicentre) for the next steps.

Following purification, the desired fragments from the above digestions were ligated using T4
Figure 4.6: EtBr-stained digestion products of (a) pK₈ with RsrII and EagI restriction enzymes, 3650 and 609 bp, in a 1% agarose gel and (b) Lambda DNA with PspOMI, 38412 and 10090bp, in a 0.7% LMP agarose gel. (c) EtBr-stained ligation product of 3650 and 38412 bp fragments from plasmid and Lambda DNA, respectively, in a 0.7% LMP agarose gel. The longer dark band should be the desired product. (d) EcoRV digestion of the K₈-Lambda fragment ligation product from part (c) shows that the ligation is successful, as evidenced by the extra high-molecular weight band in lane 3 compared with the digest of Lambda DNA alone (lane 2). The first lane in each gel shows the DNA ladder, with fragment lengths given in base pairs.
DNA ligase (Invitrogen, overnight at room temperature) (Fig. 4.5). The products of the ligation were run in a 0.7% LMP agarose gel (Fig. 4.6(c)), from which the longer band was excised and purified. The poor length resolution of the gel for these long fragments makes it hard to assess if ligation occurred as desired.

To investigate the formation of the desired construct, the ligation product was then digested with EcoRV restriction enzyme (NEB, NEBuffer 3 at 37°C for 2 hours), and digestion products were compared with those of Lambda DNA as a control. Comparison of the digestion patterns of the construct and of control Lambda DNA shows the existence of a 7 kbp molecular weight band in the digestion pattern of the construct which does not exist in the Lambda DNA pattern (Fig. 4.6(d)). The presence of this band was predicted in Geneious software [193] as a characteristic band confirming formation of the expected product. Thus, formation of a DNA construct that is 41982 bp long, with an RsrII generated overhang at one end and the 12-base overhang of Lambda DNA, 5’-GGGCGGCGACCT-3’, at the other end (Fig. 4.5) has been accomplished.

To immobilize the DNA in curtains, the DNA should possess a biotin tag at one end. Biotin is used to attach the DNA to a lipid bilayer via streptavidin in the DNA curtains setup (Fig. 4.3). Here, a biotin tag was introduced into the construct by ligation to the 5’-GTC overhang, produced by initial digestion of the pK8 plasmid with RsrII (Fig. 4.5). To the 5’-GAC overhang was ligated a short sticky-ended duplex DNA made from oligonucleotide 5’-biotin-GCC GTA GTC TCG TCT A-3’ and its complement 5’-GTC TAG ACG AGA CTA CGG C-3’ (both from IDT). To make the DNA duplex from these oligos, following IDT protocol, they were first each dissolved to a final concentration of 60 µM in DuplexBuffer (100mM Potassium Acetate, 30 mM HEPES, pH 7.5). Then, they were mixed together in a 1:1 molar ratio and heated to 94°C in a water bath and gradually cooled to room temperature on the bench.

It should be noted that the resolution of conventional agarose gel electrophoresis is insufficient to determine the success of this ligation nor to separate the 41982 bp construct from unligated 38412 bp Lambda fragment in the previous ligation step prior to biotinylation. However, as in this step only the construct with an RsrII-generated sticky end reacts with biotinylated short DNA and therefore, in the future, with a lipid bilayer via streptavidin, the unligated Lambda fragments will flush away in the DNA curtains setup. Thus, here, I simply ensured that some of the long tracks were successfully biotinylated.

To probe for the presence of the biotin tag on the construct, the sample was run on a 0.7% TBE agarose gel to separate the construct from any unligated short oligonucleotides, then transferred to
Figure 4.7: Biotinylation of the K$_8$-Lambda DNA construct, using ligation of short biotinylated duplex DNA, is successful. (a) EtBr-stained gel showing the expected molecular weight of the construct (open arrow). (b) Construct transferred to a nitrocellulose membrane, probed with streptavidin horseradish peroxidase, and visualized by chemiluminescence (ECL Plus, Pierce). Streptavidin binding coincides with the predicted size of the construct (filled arrow), indicating the presence of a biotin label on this high-molecular-weight DNA.

The constructed DNA track is about 15 $\mu$m long and can be used in the DNA curtains setup. Also it can easily tagged with digoxigenin on the other end by using the Lambda DNA sticky end (Fig. 4.5) to be applicable in the DNA racks setup (Section 1.4) as well. This can be done using, for example, digoxigenin-labeled dNTPs and a DNA polymerase such as Klenow exo$^-$ to incorporate digoxigenin into the complement to the 12-base overhang of the Lambda fragment.

In spite of the successful construction of biotinylated TW track, this method involved multiple digestion, ligation and purification steps, which resulted in a low-yield production of this long DNA. This is a serious issue for the larger-scale production of the TW track to be used in DNA curtain experiments. In order to overcome this problem, a more efficient alternative method in which I took advantage of bacterial machinery to produce a large amount of elongated TW track was undertaken. This method gives the flexibility to produce TW track to be used in either DNA curtains/racks or
DNA tightropes, and is described in the following section.

4.3.2 Biotin-Digoxigenin DNA TW Track Construction Using Bacterial Artificial Chromosomes

As mentioned already, the DNA fragment carrying the TW track is required to be at least 10 \( \mu m \) in length. To employ bacteria to amplify a long TW track, the repetitive K<sub>S</sub> track with long extension(s) needs to be inserted into a plasmid first and then transformed into bacteria for amplification. Due to the large size of insert (~30-40 kbp), selecting the proper plasmid and bacterial cell line to be able to tolerate this is challenging. Most general cloning plasmids can carry a DNA insert up to around 15-20 kbp in size [194]. Inserts in excess of this place constraints on proper replication of the plasmids (particularly for high-copy-number vectors) and can cause problems with insert stability. However, cosmids and bacterial artificial chromosome (BAC) systems have been developed for the cloning of large DNA inserts averaging 40 kbp and 130 kbp (range: 90-300 kbp), respectively [194].

In this section, I describe how I utilized an 8 kbp BAC plasmid to incorporate a long (12 \( \mu m \), 35.8 kbp) K<sub>S</sub>-containing DNA track.

The BAC plasmid used here was linearized by the manufacturer so that it carries HindIII sticky ends. Therefore, in a different strategy compared to the previous section, first a linear 33 kbp DNA fragment with HindIII overhangs was inserted into the BAC plasmid. This fragment was used to just increase the size of the plasmid, and later functions as handles for the K<sub>S</sub> track. The sequence of the plasmid along with the inserted fragment was checked to confirm that there are no binding sites for the repressor feet of the TW molecular motor. Plasmids were constructed, transformed, purified and assayed as follows.

The CopyControl BAC Cloning Kit (Epicentre) with linearized and dephosphorylated 8128 bp pCC1BAC cloning-ready vectors with HindIII overhangs was used to extend and amplify the TW track. Following the manufacturer’s protocol, first, the 33 kbp positive control DNA fragment with HindIII sticky ends, called Control Insert DNA (provided in the kit), was inserted into pCC1BAC using provided Fast-Link DNA ligase. To do this 100 ng of the Control Insert DNA was combined with 25 ng of pCC1BAC cloning-ready vectors. The volume was adjusted to 87 \( \mu l \) with sterile water, then mixed gently. The solution was incubated at 55°C for 10 min and cooled at room temperature for 15 minutes. To the cooled solution, 10 \( \mu l \) of the 10X Fast-Link ligation buffer, ATP (final concentration of 1 mM) and 2 \( \mu l \) of Fast-Link DNA ligase (4 U) were added. The ligation reaction
was incubated at 16°C overnight and stopped by heating the reaction at 65°C for 15 minutes. The reaction products were ethanol precipitated by first mixing with 10μl of 3M Sodium Acetate. Then, 2.5X volume of 100% ethanol was added to the mixture. The solution was centrifuged for 30 minutes at 13000 rpm. The supernatant was removed and 500 ml of 70% ethanol was added followed by centrifugation at 13000 rpm for 5 minutes. After carefully removing the supernatant, the pellet was resuspended in 10μl water.

The resulting product was transformed into TransforMax EPI300 Electrocompetent *E. Coli* cells (Epicentre) using electroporation. These cells are specially engineered for use with the BAC cloning systems and qualified for use with all clone sizes up to at least 140 kbp. First, 50μl of the TransforMax EPI300 Electrocompetent *E. Coli* cells were mixed with 10μl of the purified ligation components in a pre-chilled Eppendorf tube on ice. The mixture was transferred to a pre-chilled 1 mm electroporation cuvette (Gene Pulser/MicroPulser Cuvettes, BioRad). The cuvette was placed into the electroporator (MicroPulser Electroporator, BioRad) and an electric pulse of 1.8 kV was applied for 4.5 ms. Immediately after electroporation, 950μl of room temperature LB medium was added to the cuvette and mixed. Cells were transferred then to 15 ml tubes and incubated at 37°C for one hour while shaking vigorously. This step is to allow the cells to recover and express the antibiotic resistance marker. Following the 1 hour incubation, cells were plated and grown overnight at 37°C on LB-agar plates supplemented with 12.5 μg/ml chloramphenicol. Single colonies were selected and grown overnight in liquid LB-chloramphenicol medium.

Following this step, the induction of the CopyControl BAC clones to high-copy number was performed using an induction solution provided by the manufacturer. The induction solution induces expression of a mutant gene contained in the TransforMax EPI300 cells which, results in initiation of replication from the high-copy origin of replication engineered in the BAC plasmid and subsequent amplification of the clones. In this step, CopyControl BAC clones can be amplified to 10 – 20 copies per cell. Experimentally, 500μl of the overnight culture was mixed with 4.5 ml of freshly made LB-chloramphenicol (12.5 μg/ml). Then, 5μl of the 1000X concentrated CopyControl Induction Solution (supplied with TransforMax EPI300 *E. coli*, Epicentre) was added and the copy number amplification step was done for 5 hours at 37°C under vigorous shaking and aeration.

Plasmids were purified using a miniprep kit (QIAGEN) with the volumes of the reagents doubled and 70°C 10mM Tris-HCl buffer used in last step for eluting DNA. The purified plasmid DNA was then digested with HindIII restriction enzyme (NEB, NEBuffer 2 at 37°C for 2 hours) to investigate and confirm the existence of the expected insert in the grown cells (Fig. 4.8(a)). Plasmids with two
bands, an 8 kbp BAC plasmid and a high molecular weight insert band (lane 2 in Fig. 4.8(a)), in the HindIII digestion reaction are then selected and used as the cloning vector in the next step. In the following, "pLSNFplus" plasmid refers to this new 41 kbp plasmid including BAC and its insert.

To insert the Ks unit into the pLSNFplus plasmid, AvrII digestion and subsequent ligation of pKs and pLSNFplus plasmids was performed. The pLSNFplus and pKs plasmids were each digested with AvrII (NEB, NEBuffer 4 at 37°C overnight) in separate tubes (Fig. 4.8(b)). The pKs digestion products (3164 and 1015 bp fragments) were dephosphorylated using Shrimp Alkaline Phosphatase (Fermentas) and the 3164 bp fragment which carries the TW track was purified using a gel purification kit (QIAGEN) (Fig. 4.8(b)). This fragment was ligated to the linearized pLSNFplus plasmid using Fast-link DNA ligase (Epicentre) as described above. The ligation products were transformed into TransforMax EPI300 Electrocompetent E. coli and the cells grown as described above. The plasmids were extracted and purified using a mini prep kit (QIAGEN). The presence

Figure 4.8: (a) Verification of formation of pLSNFplus plasmids by digestion with HindIII restriction enzyme, running in a 0.7% agarose gel and staining with EtBr. Lane 1 is the DNA ladder, lanes 2 and 3 show the expected pattern and an unexpected band, respectively, in digestion products of miniprepped plasmids from different grown colonies of cells. Cells corresponding to plasmids in lane 2 were maxiprepped for use in the next steps. (b) AvrII digestion products of pKs plasmid and pLSNFplus plasmid (lanes 3 and 4, respectively) were run in a 1% agarose gel and stained with EtBr. Lane 1 is the DNA ladder and lane 2 is the 3164 bp fragment from AvrII digestion of pKs after gel purification. This fragment was inserted into AvrII-linearized pLSNFplus plasmid (lane 4) to construct the long TW track.
of the TW track in plasmids was investigated by digestion using AvrII (NEB, NEBuffer 4), AflII (NEB, NEBuffer 4) and NcoI (NEB, NEBuffer 3) restriction enzymes. The digestion reactions were performed at 37°C for 2 hours. The presence of the TW track was then confirmed by comparing the digestion patterns with the expected digestion patterns in virtual gels obtained from Geneious software [193] for both resultant plasmid, named pLSNFTW, and control pLSNFplus (Fig. 4.9). Colonies containing the pLSNFTW plasmid were maxi-prepped (QIAGEN) for large scale isolation of the plasmid, to be used in the next steps of labelling.

As both the BAC plasmid and the TransforMax EPI300 E.Coli cell line can tolerate long inserts, the great advantage of this method is that amplification of the TW track along with its handle is done using bacterial machinery. This avoids the low yield of product DNA following the many digestion/ligation steps required in the previous method (Section 4.3.1). The main drawback to this BAC method is that the BAC plasmid is a low-copy number plasmid and produces only 15-20 copies per cell. Therefore, large-scale production of the TW track requires handling a large amount of media. In spite of this, this method is still far more efficient than the previous approach.

The resulting pLSNFTW product contains the K₈ unit embedded in a 44.2 kbp plasmid that
Figure 4.10: The resultant pLSNFTW plasmid. AvrII restriction sites flank the inserted pKs. The 8 kbp fragment between HindIII restriction sites contains the original BAC plasmid. Ascl and XbaI restriction sites are used to generate the required length of handles for DNA tightropes and also overhangs for biotin and digoxigenin end-labelling of the TW track, respectively.

can be digested and linearized with restriction enzymes at different places (Fig. 4.10). By selecting restriction enzymes which digest at appropriate position(s), one can make the TW track based on the required length scale of handles to be applicable in DNA curtains/racks or DNA tightropes. Also, either one or both ends of the track can be labeled for further use in any of these DNA stretching techniques.

Here, end-labelling of the track with biotin and digoxigenin was done using a DNA polymerase and biotin/digoxigenin- labeled dNTPs. To do this, the plasmid sequence was examined using Geneious software [193] and XbaI and Ascl restriction enzymes were selected to linearize the plasmid with appropriate overhangs to be used in the labelling process.

The pLSNFTW plasmid was first digested with XbaI (NEB, at 37°C in NEBuffer 2 overnight) into 39 and 5 kbp fragments (Fig. 4.10) with 5’-CTAG-3’ overhangs. 1 µl of Klenow exo− DNA
polymers (5 U/µl, NEB) was used to fill in the overhangs (final concentration of 0.03 mM digoxigenin-dUTP (Roche), dGTP, dCTP and dATP (Invitrogen)) in a final volume of 240 µl to result in digoxigenin end-labelling of the DNA.

Labelling was confirmed by running the samples in agarose gel electrophoresis, blotting DNA bands and probing with anti-digoxigenin (Fig. 4.11(a)). To do this, the sample was run on a 0.7% TBE agarose gel to separate the construct from other fragments, then transferred to a nitrocellulose membrane (Trans-Blot, Bio-Rad). The membrane was washed and blocked (with 2% BSA in 1X DPBST + 0.1% Tween) and probed for digoxigenin with horseradish peroxidase-conjugated anti-digoxigenin (Jackson ImmunoResearch) according to the manufacturer’s protocol (Fig. 4.11(a)). The bands in lanes A and B are coincide with those in lanes 3 (positive control) and 4 (digoxigenin-labeled track) in the agarose gel which confirmed the successful digoxigenin labelling of the elongated TW track. The positive control here is a fragment of DNA the same size as the TW track, whose labelling had been previously confirmed.

Following ethanol precipitation, the digoxigenin-labelled construct was digested with Ascl restriction enzyme (NEB, overnight at 37°C in NEBuffer 4) to produce a 5'-GCGC overhang. The
DNA was then labeled with biotin using biotin-dCTP (Invitrogen) in the presence of Klenow exo− DNA polymerase (5U/µl, NEB) following the protocol described above. The product was run on a 0.7% TBE agarose gel to separate the construct from other fragments, then transferred to a nitrocellulose membrane (Trans-Blot, Bio-Rad). The membrane was washed and blocked (with 2% BSA in 1X DPBST + 0.1% Tween) and probed for biotin with horseradish peroxidase-conjugated streptavidin according to the manufacturer’s protocol (Fig. 4.11(b)). The coincidence of bands in lanes A and B with those in lanes 3 (positive control) and 4 (biotin-labeled track) in the agarose gel confirmed the successful biotinylation of the elongated TW track.

In total these digestions generated four DNA fragments of 35819, 5306, 2703 and 436 bp (Fig. 4.10). After confirming the end labelling, the high-molecular-weight band of 35819 bp, containing (ABC)₈, was excised after running the labelling products in a 0.7% LMP agarose gel. Following the manufacturer’s protocol, the excised gel was digested with GELase enzyme (Pierce) to yield the desired long DNA in solution.

### 4.4 Construction of the Fiducial Marker

The TW motor itself will be labelled to visualize the motor dynamics. However, unless the track is also fluorescently labelled, it will be difficult to establish that the TW motor is stepping along the track rather than simply appearing to move because of drift of the sample. A fluorescent marker is thus planned to be incorporated into the DNA track. This acts as a fiducial marker (Fig. 4.12) to provide a fixed reference point against which the movement of the TW motor can be monitored.

In my experiments, the fiducial marker is a 1 kbp fluorescently labelled DNA which will be inserted into the TW track using its flanking restriction sites. It is generated by PCR and subsequent click chemistry (Section 3.5.2), incorporating fluorophores rather than peptides.
Here, I report the construction of a fiducial marker consisting of a highly fluorescent 1 kbp region of DNA. This provides a marker with appropriate brightness to be detected with any of the single-molecule fluorescence microscopy techniques. Fluorescent DNA was generated by click chemistry between alkyne-labelled DNA and azide-labelled carboxyrhodamine-110 as follows. 1 pmol of 1 kbp alkyne-DNA (10 mM Tris (pH=8.5), Section 3.5.1) was reacted in two different reaction conditions with 10 and 4500 pmols azide-fluor 488 (carboxyrhodamine 110-PEG3-azide, Jena Bioscience). Quantitative incorporation of the fluorophore would result in one fluorophore per 100 and 2 bp (full incorporation) of sequence, respectively, for the mentioned reaction conditions. It should be noted that on average there are 500 alkynes on a 1 kbp DNA, therefore the 4500 pmols azide-fluor was far in excess of the required amount for one-to-one labelling of each alkyne on DNA. Reactions were catalyzed by the addition of copper (II) sulphate (0.5 mM) and ascorbic acid (0.5 mM) in the presence of 0.5 mM TBTA and 50% DMSO in total volumes of 20 and 25 µl for low and high azide-fluor reaction, respectively. TBTA was added to protect the DNA from Cu(I)-mediated DNA strand breaks [177]. The reaction was incubated at room temperature for 2 hours with rotary mixing. Control reactions, in which the DNA contained all natural bases and no alkynes, were similarly treated for each (high and low carboxyrhodamine 110-PEG3-azide) reaction condition. The reaction products were recovered by ethanol precipitation. Labelling of the alkyne DNA for the low labelling condition was verified by running the sample on a 1% agarose gel, scanning the gel at 532 nm and monitoring emission at 555 nm (555/20 nm band-pass filter) using the Typhoon gel scanner (Typhoon 9410 Gel and Blot Imager) (Fig. 4.13).

When the fluorophore was added in excess relative to alkyne (4500 equivalents per dsDNA molecule), the resulting highly fluorescent DNA exhibited unusual physicochemical properties. The DNA did not stain with EtBr and/or did not migrate as expected by agarose gel electrophoresis (Fig. 4.13). Similar changes were previously described for DNA densely labeled with Cyanine fluorophores [178]. As seen in Fig. 4.13, however, when the amount of fluorophore was reduced to 10 mole equivalents per mole equivalent DNA, the fluorophore-modified DNA migrates as expected on an agarose gel and can be detected by both EtBr staining and fluorescence. Further analysis of the fluorescently labelled DNA using UV-Visible absorption spectroscopy (1700 UV-Vis Spectrometer, Shimadzu) confirmed the labelling of DNA in both high and low labelling reaction conditions. Figure 4.14 shows the absorption spectra of both samples and their controls after ethanol precipitation and resuspension in 10 mM Tris-HCl buffer. The absorbance peak at 516 nm in the reaction samples corresponds to carboxyrhodamine-110 absorption. It should
be noted that in free solution carboxyrhodamine 110 absorbs at $\lambda_{abs,max} = 501$ nm and emits at $\lambda_{em,max} = 525$ nm. Our measurements of absorption and emission of the fluor-DNA showed that, in the presence of DNA the red-shift to $\lambda_{abs,max} = 516$ nm and $\lambda_{em,max} \sim$ unchanged (data not shown). The fluorescent labelling of the alkyne-DNA was also confirmed by looking at both samples and their controls using TIRF microscopy (Fig. 4.15).

It is not possible to quantify the amount of DNA labelling using the UV-Visible absorption peaks of DNA and fluorophore due to the coincidence of the Cu/TBTA absorption with DNA at $\lambda_{abs} = 260$ nm. As explained in the previous chapter, Cu/TBTA precipitates along with DNA in ethanol precipitation.

In an attempt to quantify the number of fluorophores per DNA, the fluorescence intensity of the unbound fluorophores that remained in the supernatant after ethanol precipitation of the click reaction products was measured for all samples. Due to the excess of fluorophores in the supernatant of both reaction and control of highly labelled samples, it was not possible to estimate the degree
Figure 4.14: Absorption scans confirm the fluorescent labelling of the 1000 bp alkyne-labeled DNA by the click reaction. Controls are unlabelled DNA reacted with azide-fluor under identical conditions. The 260 nm absorption peak corresponds to DNA, and that at $\sim$500 nm to carboxyrhodamine 110.

of labelling for these click-reacted samples. However, an average number of 9 fluorophores per DNA was calculated for low-labeled fluor-DNA. This is approximately equivalent to the degree of labelling anticipated from molar ratios of reactions (10 fluorophores/DNA).

It is possible that fluorophore-modified DNA will be more difficult to incorporate at fiducial marker sites than alkyne-modified DNA due to the greater steric bulk and altered local chemisery from the fluorophore, which may hinder digestion and/or ligation. Should this prove the case, alkyne-modified DNA will be ligated into the fiducial marker sites followed by click chemistry to introduce the fluorophore label. Alkyne-modified DNA has been successfully ligated in tandem demonstrating that the restriction enzymes EagI and PspOMI and T4 DNA ligase retain the ability to digest and ligate alkyne-modified DNA (Section 3.5.1).
4.5 Conclusions

In this chapter, I presented the stepping principles of TW, which is a protein-based motor that diffusively walks, or "tumbles", along a DNA track by cyclically binding and unbinding its "feet" to specific binding sites on the track. Based on the structure of the motor, the desired parameters for designing a repetitive DNA track with binding sites for the feet of TW molecular motor were explained.

To observe and characterize stepping of the TW under the application of ligand pulses, the constructed (ABC)$_8$ TW track needed to be extended in order to be stretched out to act as a 1D DNA track. Two potential techniques that are under preparation in our collaborator Prof. Heiner Linke’s laboratory are DNA curtains and DNA tightropes. The combination of DNA extension and fluorescence microscopy will allow the stepping of fluorescently labelled TW to be visualized on the track.

I described in this chapter two methods for extending an 8-repeat (ABC) track from $\sim$280 nm (824 bp) to tens of micrometres, which is the required length scale for DNA in these two stretching techniques. In the more efficient method, we used BAC plasmid and a special strain of E. coli to successfully construct and amplify a long DNA containing the TW track sequence in BAC plasmids. This method provides flexibility in construction of a TW track with appropriate lengths of handles for different stretching techniques.

To be able to account for drift during measurements and also have a reference with respect to which to observe the motion of TW, the importance of a fiducial marker was explained. The fiducial
marker, here a 1 kbp fluorescently labeled DNA, was designed to be incorporated beside the K₈ track. This fiducial marker can be inserted into the TW track using restriction sites designed at its two ends. The stepping of the TW can then be monitored by the relative motion of the fluorescent signal of the motor with respect to the fixed fluorescent fiducial marker in the TW track.
Chapter 5

Conclusions and Future Directions

In this thesis, I introduced two artificial protein-based molecular motors, dubbed "Lawnmower" and "Tumbleweed" (TW) that are inspired by biological molecular motors. The lawnmower is designed to be a multi-bladed motor that uses digestion of its substrate "lawn" to produce autonomous directional motion in a burnt-bridges strategy. The TW is a concept for a non-autonomous motor which is designed to move unidirectionally along a linear DNA track by externally controlled cyclic application of its required fuels. While the main focus of the thesis was on a systematic approach toward design and construction of the lawnmower and its track, I also explained an efficient technique to construct a long DNA track for the TW.

To guide the design of the lawnmower, I undertook investigations of a related nucleic-acid-based walker, the "molecular spider". Using molecular spiders as a model system allowed me to establish an algorithm, based on a realistic model of interaction of a spider’s leg with its substrate track, to explore how molecular spider performance depends on its design details.

First, from simulations of simple bipedal spider dynamics on an asymmetric $P$-$S$ track, I learned that biased motion arises from the preferred binding of spider legs to substrates, faster dissociation from products, and cleavage of substrates. For an ideal tightly coupled molecular spider, the combination of these effects produces a mobile $P$-$S$ boundary with molecular spider legs remaining at the $(P$-$S)$ boundary all the time and moving forward with the boundary. However, the diffusive nature of molecular spider stepping permits the loss of spider population from the ideal $P$-$S$ configuration so that spiders can be found in any of the $S$-$P$, $P$-$P$, and $S$-$S$ configurations as well. This 1D loss (diffusive) is a consequence of the design of the spider, which also resulted in a 3D loss (detachment) of bipedal spiders from the track.
Following these results for bipedal spiders, I used the algorithm to investigate how tuneable experimental parameters such as number of legs \((n)\), maximum span between legs \((S)\), and interaction kinetics influence binding time, processivity, speed and efficiency of multipedal molecular spiders. Briefly, from simulations of multipedal spider \((n = 2, 3 \text{ and } 4)\) dynamics, I found that an increase in the number of spider legs increases the spiders’ processivity and binding time but at the same time decreases spiders’ mean velocity. Here, the slowest speed was observed for quadrupedal spiders with \(S = 3\). However, I could partially recover the reduction in mean velocity by increasing the maximum span between quadrupedal spiders’ legs from \(S = 3\) to \(S = 5\). In addition, an increase in the mean velocity of quadrupedal spiders was achieved by decreasing the ratio of \(k_{off,S}\) to \(k_c\) (\(k_c\) remained constant so that, when bound to a substrate, the probability of cleavage relative to dissociation was increased). Interestingly, with the adjusted kinetic rates, quadrupedal spiders with \(S=5\) could reach and exceed bipedal spiders’ mean velocity. These results indicated that improving measures such as velocity and processivity can be achieved by tuning more than one experimental parameter simultaneously, otherwise some spider properties might improve at the expense of others.

Simulations of multipedal spiders also showed an average time-dependent behaviour in their displacement and other properties such as mechanochemical coupling and efficiency. I related this behaviour to the 1D loss pathway of molecular spiders, which is the result of the diffusive nature of their stepping, lack of power stroke and their loose mechanochemical coupling. This time dependence and lack of a unique step size led us to introduce a definition of mechanochemical coupling and efficiency (Eqs. (2.6) and (2.7)) to be applicable to molecular spiders.

These studies showed a time-independent mechanochemical coupling for tripedal and quadrupedal spiders in the absence of force. In the presence of an applied force (against their directional motion), however, these ratios become time-dependent such that the spiders’ mechanochemical coupling decreased with time, more rapidly for tripedal than quadrupedal spiders. Calculation of efficiencies of multipedal spiders \((n=2, 3 \text{ and } 4, S=3)\) showed that these are very weak motors and among them quadrupedal spiders are the most efficient ones \((\eta_{\text{max}} = 0.7\%)\). In addition, spider efficiency decreased with time and approached zero over time. Therefore, we concluded that spiders must be considered as transient molecular motors.

Following the simulations, we applied these findings to the design of the lawnmower and selection of its components to optimize the likelihood of detecting and characterizing biased motion.

To construct the lawnmower, trypsin was selected as the blades due to its preserved activity after performing the required chemical functionalization steps. A 20 nm amine-modified QD was chosen
for the lawnmower hub. Besides its advantages in observation, it is likely that the QD hub provides enough binding sites for linking the trypsin blades so that they keep the lawnmower bound to the track while performing progressive directional motion.

The first generation of the lawnmower was constructed by covalent attachment of thiolated trypsins to maleimide-modified QDs using 1 nm heterobifunctional linkers. Formation of an active lawnmower was confirmed using trypsin activity assay for trypsin conjugated QDs. The lawnmowers were quantified as (on average) eight-bladed motors.

As a requirement for the function of lawnmower, the lawnmower track was constructed as a 1D peptide track, for which the backbone was a 1 kbp DNA fragment. To be able to correlate directional motion of lawnmower (using fluorescence signal from its QD hub) to chemical reaction, an internally quenched peptide substrate was chosen as the lawnmower substrate. The track was constructed by using click chemistry to covalently conjugate the azide-modified peptides to alkyne-incorporated DNA. For this goal, the densely labelled 1 kbp alkyne-DNA backbone was generated using PCR by replacing dTTP with C8-alkyne-dUTP in the PCR reaction. It should be noted that, using PCR, I was able to generate alkyne-DNA as long as 3 kbp with full incorporation of alkynes into DNA, which is the longest ever reported (to my knowledge).

The clicked peptide-azide to alkyne-DNA product was analyzed, and conjugation of peptide to DNA was confirmed using UV-Vis spectroscopy. Interestingly, using trypsin digestion assay, it was shown that trypsin is able to digest the DNA-conjugated peptide although with a slower rate. This is very promising and confirms the preserved functionality of the peptide even in close proximity to the DNA molecule. Due to the poor solubility of the peptide-DNA, quantification of the product and its precise analysis were not possible. Attempts to improve the peptide-DNA solubility (such as using an inorganic solvent or attachment of long DNA handles) did not enhance the solubility of the peptide-DNA. In spite of solubility challenges, the current results, particularly the digestibility of peptide-DNA, are very promising.

Ultimately, the function of the lawnmower will be monitored on a peptide-DNA track with a lower density of labelling. Also, for application in observation techniques, such as DNA racks and DNA tightropes, it must be embedded in a 10-15 \( \mu \)m long DNA fragment. Therefore, it is possible that the decreased density of labelling along with significantly enhanced DNA properties of the construct (1 kbp peptide-DNA vs \( \sim \) 30 kbp long DNA handles) will improve the solubility of the motor track. In addition, for these observation techniques, DNA is stretched by tethering its ends using affinity labels. I propose, therefore to first stretch and tether the long DNA (including the
inserted 1kbp alkyne-DNA) in the setup, then perform click chemistry between the peptide-azide and alkyne-DNA. I hypothesized that the stretched DNA will keep the embedded construct from collapsing, so that it remains in stretched form with DNA handles.

Overcoming the solubility of the peptide-DNA construct, the next step would be examining digestibility of the peptides on the DNA track with trypsin blades of lawnmower. If solubility can be resolved by stretching the long DNA followed by peptide click reaction (as suggested above), this examination can be combined with investigating lawnmower directional and processive motion. The processive motion of the lawnmower would be investigated by monitoring the appearance of fluorescence signal from digestion of peptide substrates concomitantly with tracking the location of fluorescence from the QD hub.

I also described a strategy to construct an extended 8-repeat (ABC) DNA track for the TW molecular motor using a BAC plasmid. The strategy allowed me to produce a long track (∼10-15 µm) efficiently and provided the flexibility to modify it with appropriate length of handles for different stretching techniques. Also construction of a fiducial marker as a reference to monitor the dynamics of the TW along with the drift of the track during measurements was presented.

After end-labeling of the TW track, we plan to use it in DNA racks first to examine the binding of fluorescently labelled repressors in the presence of their corresponding ligands. Following these tests and successful construction of the TW motor, the processivity and directionality of the TW can be investigated by monitoring TW stepping via its introduced fluorescence signal relative to the inserted fiducial marker along the track.

Performing simulations using my proposed modelling algorithm to study molecular spiders played a constructive role in providing insight for the design of the lawnmower. In general, this algorithm can be used as an efficient computational tool to investigate properties of different types of synthetic molecular motors. It is applicable to both autonomous and non-autonomous molecular motors with known kinetics of interactions with their substrate track and/or their fuel. In related work by Martin Zuckermann, application of this algorithm to the TW and inchworm (a molecular motor with a power stroke) is underway. This algorithm provides a tool to study and to understand how experimentally designable parameters of a molecular motor can be tuned to optimize motor properties and performance. More generally, my approach to calculating the efficiency of molecular spiders provides a tool for characterizing their performance as motors for a given set of experimental conditions. I hope that the approach used here will serve as an experimentally relevant tool to assist in the design of synthetic motors with specific performance measures such as speed,
processivity, or efficiency and creates the opportunity for in silico optimization of motor performance. These in silico approaches to motor design can be applied in the laboratory for applications as diverse as transport, drug delivery, and chemical synthesis.

My experimental work on the construction of a 1D densely labeled peptide-DNA track can be employed as a simple method of construction of DNA-based nanostructures and templates. Compared to DNA origami, a powerful method to make DNA-based templates, it does not require much groundwork in earlier steps prior to template construction such as precise design of the many required "staple-strands" and investigation of assembly algorithms. DNA origami-based structures are rigid and allow modifications only at specifically engineered sites on the DNA superstructure, an advantage for low-density presentation but which restricts the maximum density of labelling. In contrast, a nanostructure based on linear DNA not only would possess greater configurational flexibility but also would permit greater versatility with respect to modifications since these can be introduced throughout the length of the DNA.

A biomaterial based on a hybrid structure of linear DNA conjugated to various biomolecules along the DNA strand has the potential to provide orthogonal information streams to independently direct assembly not only in different ways but potentially at different times or under different conditions. A hybrid biomaterial based on DNA and peptide, for instance, allows the DNA to encode for assembly while the peptide encodes for interactions with proteins or small molecules.

The combined system of lawnmowers with their peptide-DNA tracks may also have applications as a new type of "active material" [195]. These are materials consisting of a polymer network whose elastic properties are controlled by enzymatic activity (e.g actin filament network with myosin motors). Due to similarities between cell structure and mechanical function with these active networks, they are developed both to help understand the mechanics of the cell and as novel materials.
Appendix A

Rate Constants for Molecular Spiders

A.1 Diffusion Rate Constant

Here, I first calculate the approximate size of a molecular spider, then use this to estimate a centre-of-mass diffusion time for spider steps, using the example of the bipedal spider NICK-2.4A [86].

The structure of a catalytic unit (spider’s leg) is [86]:

5’-Biotin-(TEG)-SP(18)-SP(18)-CTCTTTCTCCGAGCC GGTCGAAATAGTGA(1-5).

Here SP(18) is an 18-atom hexaethylene glycol spacer and TEG is tetraethylene glycol. Pei et al. varied the sequence of the "foot" region of legs to include from one to five adenines as shown by A(1-5) in the above sequence.

The number of nucleotides in each deoxyribozyme leg is 28 to 32. Considering the distance between nucleotides (0.34 nm) in ssDNA, the length of this part is approximately 10 nm. The approximate length of an ethylene glycol is 4 Å (estimated from information provided by Pierce about SM(PEG)_n crosslinkers [109]). Therefore the total length of a spider’s leg is equal to 16.6 nm. Considering the persistence length of $L_p = 2$ nm for ssDNA [110] and the calculated contour length of $L_c = 16.6$ nm, the radius of gyration of a spider’s leg is [111],

$$R_g = \sqrt{\frac{L_p L_c}{3}} = 3.3 \text{ nm.} \quad (A.1)$$

The crystal structure of streptavidin is known and shows an approximate diameter of 5 nm [112]. Having the knowledge of all of these parts, we replace a bipedal molecular spider with an
sphere of radius 6 nm (size of a leg, 3.3 nm, plus the radius of streptavidin, 2.5 nm). Then, we use
the Stokes-Einstein relation [113] for the diffusion constant of a sphere with radius \( R \), in a medium
of viscosity \( \eta \) (which is an aqueous solution here), to estimate the diffusion constant of a molecular
spider:

\[
D = \frac{k_B T}{6\pi \eta R} = \frac{4 \times 10^{-21} \, J}{6 \times 3.14 \times 10^{-3} \, kg/m \cdot sec \times 6 \times 10^{-9} \, m} = 3.5 \times 10^{-11} \, m^2/s. \tag{A.2}
\]

In order to estimate the timescale of diffusion of the spider on my 1D track we calculate the
average time required to diffuse 9 nm in 1D:

\[
t = \frac{<x^2>}{2D} = \frac{81 \times 10^{-18} \, m^2}{2 \times 3.5 \times 10^{-11} \, m^2/s} = 11 \times 10^{-7} \, s \approx 10^{-6} \, s \tag{A.3}
\]

Thus,

\[
k_{diffusion} = \frac{1}{t} = 10^6 \, s^{-1} \tag{A.4}
\]

A.2 Calculation of the First-order \( k_{on} \) Rate

Based on the information of reference [86] that concentration of substrate is 3 mM and the ratio of
substrate to quadrupedal spider is 3800:1, we calculate the first-order \( k_{on} \) rate in units of \( (s^{-1}) \). We
first calculate the concentration of legs, \([\text{leg}]\), as follows.

\[
[\text{leg}] = \frac{3mM(\text{concentration of substrate}) \times 4(\text{number of spider legs})}{3800 \text{ substrate/spider}} = 3 \, \mu M \tag{A.5}
\]
The second-order rate constant for binding of a leg is $k_{on} = 7.9 \times 10^5 \text{ (M}^{-1}\text{s}^{-1})$ [86], therefore

$$k_{on}(s^{-1}) = k_{on} (M^{-1}s^{-1}) \times \text{Conc. (M)}$$

$$= 7.9 \times 10^5 (M^{-1}s^{-1}) \times 3 \times 10^{-6} (M)$$

$$\approx 2.4 \text{ (s}^{-1})$$  \hspace{1cm} \text{(A.6)}

These above calculations approximate the case of freely diffusing molecular spiders.

When an unbound leg of a spider is going to bind to a new site, however, the spider is already bound to the surface so we need to do a correction. In order to do this for a bipedal spider, we first need to calculate the local concentration of a freely diffusing spider’s leg when the other leg is bound. Based on [86], the maximum span of a spider is 25 nm. Therefore, assuming one leg is bound, the other leg is able to explore a spherical volume with radius of 25 nm which is $V_{\text{sphere}} = \frac{4}{3} \pi r^3 = 6.5 \times 10^{-23} \text{ nm}^3$. The concentration of the freely diffusing leg, $C_{\text{leg}}$, in molar (M) is

$$C_{\text{leg}} = \frac{n(\text{mols})}{V(\text{litre})} = \frac{N(\text{free legs})}{N_A V_{\text{sphere}}} = 2.5 \times 10^{-5} M$$  \hspace{1cm} \text{(A.7)}

where $N_A$ is Avogadro’s number and $N(\text{free legs}) = 1$ for a bipedal spider. Using this concentration and the second-order rate of a leg binding $k_{on} = 7.9 \times 10^5 \text{ (M}^{-1}\text{s}^{-1})$, the first-order rate of binding is

$$k_{on}(s^{-1}) = k_{on} (M^{-1}s^{-1}) C_{\text{leg}}(M) = 7.9 \times 10^5 \times 2.5 \times 10^{-5} = 20 \text{ s}^{-1}$$  \hspace{1cm} \text{(A.8)}

My simulations predominantly use this value of $k_{on} = 20 \text{ s}^{-1}$ (Table 2.1), but we also test $k_{on} = 2.4 \text{ s}^{-1}$ (Section 2.2.3).

**A.3 Calculation of $k_c$**

Pei et al. measured the substrate turnover rate by bipedal spiders with 4A to be $k_{\text{cat}} = 1.5 \text{ min}^{-1}$ [86]. This differs from my $k_c$, Table 2.1, because it includes contribution from $k_c$ and $k_r$. 
\[ \frac{1}{k_{\text{cat}}} = \frac{1}{k_c} + \frac{1}{k_r} \]

\[ \frac{1}{1.5\text{(min}^{-1}) \frac{1}{60}\text{(min s}^{-1})} = \frac{1}{k_c} + \frac{1}{0.046\text{(s}^{-1})} \]  

(A.9)

using \(k_r\) from [86]. Thus, we have taken \(k_c = 0.055 \text{ s}^{-1}\) for my kinetic model.
Appendix B

Gillespie Algorithm

The Gillespie Algorithm is a mathematical method for stochastic simulation of chemical reactions [114]. As explained in Section 2.2.2, we use this method to simulate the interaction of an spider leg with its substrate based on the kinetic model shown in Fig. 2.2 and explained in Section 2.2.1. Here, a general explanation of the Gillespie Algorithm is presented.

Consider $M$ possible reactions $R_i$ from a given state of the system, each having a rate constant $k_i(k_1, k_2, \ldots k_M)$.

We will assume here no degeneracy, as is the case for spiders, as each of the legs is separately identified and has rates associated with it.

What we need to do is to determine which of the $M$ reactions will occur first, $R_j$, and how long it will take, $\tau$.

Consider the joint probability density function

$$P(\tau, j|x, t) = k_j(x) \exp(-k_0(x)\tau)$$

representing the probability, given a state of the system $X(t) = x$, that the next reaction in the system will occur in the infinitesimal time interval $[t + \tau, t + \tau + dt)$, and will be an $R_j$ reaction. Here

$$k_0(x) = \sum_{i=1}^{M} k_i(x)$$

is the sum of all rate constants.

From this, we determine $j$ and $\tau$: Draw two random numbers, $r_1$ and $r_2$ from a uniform proba-
bility distribution between 0 and 1. The length of time for a reaction to occur is given by

\[ \tau = \frac{1}{k_0} \ln \left( \frac{1}{r_1} \right) \] (B.3)

and \( j \) is the smallest integer satisfying

\[ \sum_{i=1}^{j} k_i(x) > r_2k_0(x) \] (B.4)
Appendix C

Drift of Diffusing Spiders on a $P$-$S$ Track

The simulation results for HOH spiders with neither cleavage nor detachment on a $P$-$S$ track showed that $\langle \Delta x_{cm} \rangle$ increases with time (see Fig. 2.6(b)). My simulations also found that the spider population evolved to a fixed ratio of the number of spiders on the $S$ part of the track to the number of spiders on the $P$ side of the track. In this context, Fig. C.1(a) shows the evolution of $P(x_{cm})$ versus $x$ for HOH spiders with neither cleavage nor dissociation on a $P$-$S$ track.

In this case, the spiders do not alter the track, and the boundary between $S$ and $P$ sites remains at $x = 500.5$. We therefore expect spiders to diffuse with the diffusion constant, $D_P$, for spiders on a pure $P$-track for $x < 500$, and with the diffusion constant, $D_S$, for spiders on a pure $S$-track without cleavage for $x \geq 501$. In both cases there is no detachment. Furthermore, both $D_P$ and $D_S$ can be calculated using Eq. (2.3) and the simulation results for bipedal spiders on all-$P$ and all-$S$ (uncleavable) tracks (Fig. C.2).

In this appendix we present an analytic model that can account for the findings of Fig C.1(a). For my calculations, we assume that the $P$-$S$ boundary is at $x = 0$, that spiders diffuse outwards from this barrier with diffusion constant $D_S$ to the right and diffusion constant $D_P$ to the left. On the basis of this assumption, the distribution functions, $P(x, t)$, of spiders diffusing from this boundary are a half-Gaussian on each side of the boundary and are given by:
APPENDIX C. DRIFT OF DIFFUSING SPIDERS ON A P-S TRACK

Figure C.1: Evolution of position probability distributions \( P(x_{cm}) \) (a) for HOH spiders without cleavage and without detachment and (b) calculated using the model presented in this Appendix.

\[
P(x, t) = \begin{cases} 
\frac{2}{\sqrt{\pi D_P t}} \exp \left( -\frac{x^2}{4 D_P t} \right) & x < 0 \\
\frac{2}{\sqrt{\pi D_S t}} \exp \left( -\frac{x^2}{4 D_S t} \right) & x > 0 
\end{cases} 
\]  
(C.1)

Substituting the values of \( D_P = 0.067 \text{ s}^{-1} \) and \( D_S = 0.021 \text{ s}^{-1} \) found from simulations into Eq. (C.1), we obtain \( P(x, t) \) for each side of P-S track (Fig. C.2(b)).

Having the distribution function, \(<x_{cm}>\) can then be calculated for each side of the track.

\[
<x_{cm,S}> = \int_0^\infty x \ P(x, t) \, dx 
= \sqrt{\frac{2}{\pi}} \sqrt{2D_S t} \frac{1}{2} 
\]  
(C.2)

\[
<x_{cm,P}> = \int_{-\infty}^0 x \ P(x, t) \, dx 
= -\sqrt{\frac{2}{\pi}} \sqrt{2D_P t} \frac{1}{2} 
\]  
(C.3)
Finally, we calculate $\langle x_{\text{cm}} \rangle$ for the whole system as follows:

$$\langle x_{\text{cm}} \rangle = \frac{N_S \langle x_{\text{cm},S} \rangle + N_P \langle x_{\text{cm},P} \rangle}{N_S + N_P} \quad \text{(C.4)}$$

Here $N_S$ and $N_P$ are numbers of spiders and $\langle x_{\text{cm},S} \rangle$ and $\langle x_{\text{cm},P} \rangle$ are mean of center-of-mass displacements from origin ($x = 500.5$) for spiders on the $S$ and $P$ sides of the track, respectively. Substituting equations (C.2) and (C.3) and also the calculated ratios of $N_S/N_P = 6.37$ and values of $D_P$ and $D_S$ into Eq. (C.4), we obtain:

$$\langle x_{\text{cm}} \rangle = 0.18 \sqrt{t} \quad \text{(C.5)}$$

My analytic calculations therefore predict that $\langle x_{\text{cm}} \rangle$ increases with the square root of time. The calculated values of $\langle x_{\text{cm}} \rangle$ versus time from equation (C.5) are plotted in Fig. 2.9(a) and show that this simple analytic model is in good agreement with the simulation results.
Appendix D

Construction of Lawnmower

Protocols are provided by Dr. Suzana Kovacic.

D.1 Production of Free Thiols

The following proteases were initially selected as possible candidates for lawnmower construction: thrombin (bovine, Haematologic Technologies), elastase (porcine pancreas, Calbiochem), pepsin, trypsin and proteinase K (all from Sigma). These proteases were chosen because they were readily available and had cysteine residues.

Two chemical treatments to test the effect of bio-functionalization of these proteases were performed in parallel. In the first method, proteases were reduced with immobilized TCEP, Tris[2-carboxyethyl] phosphine, disulfide reducing gel (50% slurry in ultra pure water, Pierce) to produce the free thiols to be used for cross-linking to the QD hub. Therefore, to an average of 0.017 mg of each protease in 1X DPBS 5mM EDTA (50 ul) was added an equal volume of TCEP slurry resin. Each sample was incubated at room temperature for 30 min with rotary mixing. Thereafter, TCEP gel was removed by centrifuging at 6000 rpm for 3 min and supernatant containing the reduced proteases was recovered. Controls followed the same procedure but TCEP was eliminated.

Alternatively, thiol groups were introduced to the primary amines of these proteases using Traut’s reagent (Pierce). To each sample, Traut’s reagent was added in 2X molar excess over primary amines in 0.1M phosphate buffer (pH=8, total volume 50 µl). Samples were incubated at room temperature for one hour. Excess of glycine was added to quench the reaction.

Proteases treated with both methods were then cleaned up by Zeba desalting columns (7K
MWCO, Pierce) following the manufacturer’s protocol and assayed for protease activity (Fig. D.1) using a fluorescence assay (Section D.6).
Figure D.1: Activity of proteases after each chemical treatment step toward construction of lawnmower blades investigated using a fluorescence assay. Trypsin showed the highest preserved activity after different chemical modifications. Note: Proteinase K showed a very high activity but was eliminated due to its non-specific proteolytic activity. (Figure courtesy of Dr. Suzana Kovacic.)
D.2 Biotinylation with Biotin-PEG2-Maleimide

To confirm that free thiols were available for further modification, proteases prepared using either of the two above methods were reacted with biotin-PEG2-maleimide (Pierce). This cross-linker was chosen as it contains a thiol-reactive maleimide group as will be used for cross-linking the proteases to the QDs. The biotin tag allows for indirect confirmation by Western blotting of the presence of free thiols on the protease and the successful maleimide-thiol reaction.

A 20 mM stock solution of biotin-PEG2-maleimide was prepared in DMSO. One equivalent biotin-PEG2-maleimide was added to one equivalent protease (10 \( \mu \)g) at pH~7. pH was adjusted as necessary to pH=7 by addition of 0.2 M sodium phosphate (pH~6). Reaction samples and controls were incubated for 2 hours at room temperature. To probe for the presence of the biotin tag, loading buffer was added to each sample and heated at 100\(^\circ\)C for 5 min, then run in a 12% SDS-PolyAcrylamide gel. The protein bands were transferred to a PVDF membrane (Trans-Blot, Bio-Rad). The membrane was washed and blocked with 2% BSA in 1xDPBST+0.1% Tween and probed according to the manufacturer’s protocol with horseradish peroxidase-conjugated streptavidin (Jackson ImmunoResearch) (Fig. D.2). Assaying for activity following biotinylation of the proteases was performed again using a fluorescence assay (Fig. D.1) (Section D.6).

D.3 Maleimide Functionalization of QDs

To 80 pmol of amine-derivatized QDs (Invitrogen), 367 pmol Sulfo-SMCC (sulfosuccinimidyl 4-[N-maleimidomethyl] cyclohexane-1-carboxylate) (Pierce) was added in the presence of DPBS and 5mM EDTA. Sulfo-SMCC is a heterobifunctional cross-linker that contains both a N-hydroxy-Succinimide (NHS) ester and a maleimide to allow covalent conjugation of amine- and sulfhydryl-containing molecules. NHS esters react with primary amines to form amide bonds, while maleimides react with sulfhydryl groups to form stable thioether bonds. This reaction changes the surface chemistry of the QDs from amine to maleimide to permit subsequent reaction with free thiols on the protease. A control reaction with 80 pmol QDs and no Sulfo-SMCC was also prepared. Samples were incubated for 1 hr at room temperature with rotary mixing. The reactions were stopped by adding 5.3 \( \mu \)mol glycine and incubating for 15 min. The maleimide-conjugated QDs were purified using Zeba desalting columns (7K MWCO, Pierce) following the manufacturer’s protocol.
Figure D.2: Western blot analysis of biotinylated TCEP– and Traut– treated proteases. Protease controls are unbiotinylated proteases which act as negative control. The dark intense bands in the biotinylated proteases are indications of biotin labelling of these samples. (Figure courtesy of Dr. Suzana Kovacic.)

**D.4 Conjugation of Trypsin to Maleimide QDs**

Immediately after purifying QD-maleimide, the lawnmower was constructed by incubating 0.25 mg trypsin with reduced thiols (10 nmol) with 40 pmol maleimide-QDs. The pH of the reaction was adjusted to 7.2 with 10X DPBS. The reaction was incubated for 2 hours at room temperature with rotary mixing. As a control, the same amounts of trypsin and QDs (both without any chemical modification) were mixed and incubated in the same conditions. To stop the reactions, 2 nmol of β-mercaptoethanol were added to all samples and incubated for 30 min. Samples were concentrated using Vivaspin 500 (Vivaproducts) 50K MWCO spin filters at 1000 rpm for 4 min.
D.5 Purification of Trypsin Conjugated QDs

Two 4.5 ml Sephacryl S300 (GE Healthcare Life Sciences) size-exclusion gravity columns were prepared in 5 ml disposable columns (Pierce) and equilibrated with DPBS. Following loading of the samples (about 20 µl), columns were washed with 1.5 ml PBS before fractions were collected (each fraction volume is 5 drops ~72 µl).

D.6 Fluorescence Assays

The fluorescence assay to examine the activity of the QD-linked proteases in each fraction was done using EnzChek Protease Assay Kits (red fluorescence, Molecular Probes). QD-protease was added to BODIPY-casein substrate (Molecular Probe) in a 96-well plate following the manufacturer’s protocols. Protease activity was monitored using a fluorimeter with Ex=590/20 nm and Em=645/40 filter settings.

QD fluorescence was also monitored in each fraction with Ex=460/50 nm and Em=595/35 nm.

D.7 Method of Lawnmower Quantification

The number of active trypsins on a single QD was calculated as follows. First, the QD fluorescence in each lawnmower fraction that had eluted from the column was normalized against a QD sample of known concentration in the same buffer. This gives the moles of QD in each fraction. Then the trypsin activity in each fraction was measured (as above). The trypsin activity in each sample was compared against the trypsin activity in a control sample where the amount of trypsin was known. This gives the moles of trypsin in each fraction. The number of trypsins per QD is calculated from the ratio of moles of trypsin : moles of QD in each fraction (Table D.1).

It should be noted that there was some observed residual trypsin activity in all fractions. Therefore, in fractions with not much QD present, the ratio of moles of trypsin : moles of QD did not result in a meaningful number as a low trypsin activity was divided by a small number for QD fluorescence (there was no active lawnmower in those fractions). Therefore, all these ratios were ignored except for those fractions where the presence of lawnmower was obvious based on QD fluorescence. The average ratio of those fractions was calculated as the number of trypsins per QD which is about 8.
### Table D.1: The calculated moles of trypsin and QD in each fraction. The average number of trypsins per QD was calculated from the ratio of moles of trypsin : moles of QD in each fraction. (Figure courtesy of Dr. Suzana Kovacic.)

<table>
<thead>
<tr>
<th>Fraction number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mole trypsin</td>
<td>6.20E-12</td>
<td>9.05E-12</td>
<td>1.23E-11</td>
<td>1.98E-11</td>
<td>2.23E-11</td>
<td>2.05E-11</td>
<td>1.37E-11</td>
<td>1.70E-11</td>
<td>1.40E-11</td>
<td>9.15E-12</td>
</tr>
<tr>
<td>Mole QD</td>
<td>2.78E-14</td>
<td>2.32E-13</td>
<td>8.43E-13</td>
<td>2.88E-12</td>
<td>2.97E-12</td>
<td>2.30E-12</td>
<td>2.03E-12</td>
<td>1.67E-12</td>
<td>1.32E-12</td>
<td>9.47E-13</td>
</tr>
<tr>
<td>#trypsin/QD</td>
<td>2.23E+02</td>
<td>3.91E+01</td>
<td>1.45E+01</td>
<td>6.87E+00</td>
<td>7.48E+00</td>
<td>8.88E+00</td>
<td>6.74E+00</td>
<td>1.01E+01</td>
<td>1.06E+01</td>
<td>9.66E+00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fraction number</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mole trypsin</td>
<td>7.04E-12</td>
<td>7.08E-12</td>
<td>1.04E-11</td>
<td>8.97E-12</td>
<td>9.13E-12</td>
<td>7.96E-12</td>
<td>7.51E-12</td>
<td>7.31E-12</td>
<td>7.92E-12</td>
<td>1.04E-11</td>
</tr>
<tr>
<td>Mole QD</td>
<td>7.06E-13</td>
<td>5.84E-13</td>
<td>4.89E-13</td>
<td>3.17E-13</td>
<td>2.34E-13</td>
<td>2.15E-13</td>
<td>1.83E-13</td>
<td>1.37E-13</td>
<td>1.19E-13</td>
<td>8.99E-14</td>
</tr>
<tr>
<td>#trypsin/QD</td>
<td>9.96E+00</td>
<td>1.21E+01</td>
<td>2.12E+01</td>
<td>2.83E+01</td>
<td>3.90E+01</td>
<td>3.70E+01</td>
<td>4.11E+01</td>
<td>5.33E+01</td>
<td>6.68E+01</td>
<td>1.16E+02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fraction number</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
<th>26</th>
<th>27</th>
<th>28</th>
<th>29</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mole trypsin</td>
<td>1.45E-11</td>
<td>1.58E-11</td>
<td>1.67E-11</td>
<td>2.27E-11</td>
<td>2.40E-11</td>
<td>2.57E-11</td>
<td>2.58E-11</td>
<td>3.10E-11</td>
<td>2.97E-11</td>
</tr>
<tr>
<td>Mole QD</td>
<td>6.66E-14</td>
<td>5.80E-14</td>
<td>5.66E-14</td>
<td>2.62E-14</td>
<td>2.30E-14</td>
<td>3.21E-14</td>
<td>1.99E-14</td>
<td>1.97E-14</td>
<td>1.34E-14</td>
</tr>
<tr>
<td>#trypsin/QD</td>
<td>2.18E+02</td>
<td>2.71E+02</td>
<td>2.96E+02</td>
<td>8.66E+02</td>
<td>1.05E+03</td>
<td>8.00E+02</td>
<td>1.29E+03</td>
<td>1.58E+03</td>
<td>2.22E+03</td>
</tr>
</tbody>
</table>
Appendix E

Construction of Azide-PEG4-Peptide

Protocol provided by Dr. Suzana Kovacic.

One equivalent of the peptide, MOCAc-Ala-Pro-Ala-Lys-Phe-Phe-Arg-Leu-Lys(Dnp)-NH2 ((7-Methoxycoumarin-4-yl)acetyl- L-alanyl- L-prolyl- L-alanyl- L-lysyl- L-phenylalanyls- L-phenylalanyls- L-arginyl- L-leucyl- Nε- (2,4-dinitrophenyl)- L-lysine amide, Peptides International), Fig. 3.5 was reacted with 10 equivalents of NHS-PEG4-azide (15-Azido-4,7,10,13-tetraoxa-pentadecanoic acid succinimidy1 ester, Jena Bioscience) in DMSO in the presence of N,N-diisopropylethylamine (Sigma). The reaction proceeded overnight at room temperature in the dark with rotary mixing. The product was purified by first acidifying the reaction mixture to 0.1% trifluoroacetic acid (TFA), then the peptide product was adsorbed to a pre-equilibrated C18 pipette tip (100ul, Pierce). The C18 tip was successively washed with 0.1% TFA/5% acetonitrile in water, 10% methanol/0.1% acetic acid and 40% methanol/0.1% acetic acid. Azide-PEG4-peptide was eluted from the resin with 100% methanol/0.1% acetic acid. The mass of the product was confirmed by electron spray ionization mass spectroscopy (Fig. E.1).
APPENDIX E. CONSTRUCTION OF AZIDE-PEG4-PEPTIDE

Figure E.1: Electron spray ionization mass spectroscopy of peptide-azide shows the high intensity peak at 1732.85, which coincides with the expected mass of 1733 g/mol for peptide-azide. (Figure courtesy of Dr. Suzana Kovacic an obtained with an ESI ion source on an Agilent Time-of-Flight LC/MS mass spectrometer.)
Appendix F

PCR Thermocycles to Generate Alkyne-DNA

F.1 Products with Pwo Polymerase

<table>
<thead>
<tr>
<th></th>
<th>Temperature</th>
<th>Time</th>
<th>Cycle Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>94°C</td>
<td>2 minutes</td>
<td>1X</td>
</tr>
<tr>
<td>Denaturation</td>
<td>94°C</td>
<td>30 seconds</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td>64°C</td>
<td>15 seconds</td>
<td>30X</td>
</tr>
<tr>
<td>Elongation</td>
<td>72°C</td>
<td>45 seconds</td>
<td></td>
</tr>
<tr>
<td>Final elongation</td>
<td>72°C</td>
<td>10 minutes</td>
<td>1X</td>
</tr>
</tbody>
</table>

Table F.1: Thermocycle for generation of 400 bp and 1 kbp products using Pwo polymerase. EagI primers are listed in Tables 3.2 and 3.4, PspOMI primers are in Tables 3.3 and 3.4 (400 bp).
F.2 1 kbp and 3 kbp Products with KOD XL Polymerase

<table>
<thead>
<tr>
<th></th>
<th>Temperature</th>
<th>Time</th>
<th>Cycle Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>94°C</td>
<td>2 minutes</td>
<td>1X</td>
</tr>
<tr>
<td>Denaturation</td>
<td>94°C</td>
<td>30 seconds</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td>64°C</td>
<td>15 seconds</td>
<td>30X</td>
</tr>
<tr>
<td>Elongation</td>
<td>72°C</td>
<td>45 seconds</td>
<td></td>
</tr>
<tr>
<td>Final elongation</td>
<td>72°C</td>
<td>10 minutes</td>
<td>1X</td>
</tr>
</tbody>
</table>

Table F.2: Thermocycle for generation of 1 kbp alkyne-DNA using KOD XL polymerase and primers in Table 3.4.

<table>
<thead>
<tr>
<th></th>
<th>Temperature</th>
<th>Time</th>
<th>Cycle Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>94°C</td>
<td>2 minutes</td>
<td>1X</td>
</tr>
<tr>
<td>Denaturation</td>
<td>94°C</td>
<td>30 seconds</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td>64°C</td>
<td>15 seconds</td>
<td>30X</td>
</tr>
<tr>
<td>Elongation</td>
<td>72°C</td>
<td>60 seconds</td>
<td></td>
</tr>
<tr>
<td>Final elongation</td>
<td>72°C</td>
<td>10 minutes</td>
<td>1X</td>
</tr>
</tbody>
</table>

Table F.3: Thermocycle for generation of 3 kbp alkyne-DNA using KOD XL polymerase and primers in Table 3.2.
Bibliography


BIBLIOGRAPHY


