UNIVERSITY OF CALIFORNIA,
IRVINE

Characteristics and Emergent Behaviors of Multisite Intrinsically Disordered Proteins

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Mathematical, Computational, and Systems Biology

by

Lara C Clemens

Dissertation Committee:
Associate Professor Jun Allard, Chair
Professor Lee Bardwell
Professor Qing Nie
Assistant Professor Elizabeth Read

2019

DEDICATION

To my family.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Jun Allard, for his continuous support over the last five years. His passion for science and dedication to his graduate students is a constant inspiration. I am grateful for all of the knowledge he has imparted, consciously and not, to help me become a better scientist and mentor.

I would like to thank Omer Dushek and Misha Kutuzov for many enjoyable and educational conversations, and for their patience and guidance. I'd also like to thank the entire Dushek Lab and MIG for being so welcoming each time I visited.

I would like to thank Margot Quinlan and Aanand Patel for their curiosity and for asking 'But what about this?'. I'd also like to thank the MathBioU/MathExpLR students I've worked with over the past two summers — Lucy, Katherine, Kathryn, Poorvi, and Joanna — for their enthusiasm in helping answer that question.

I'd like to thank the CCBS administrative staff for making all the details of graduate school run smoothly for me.

I would like to thank my labmates — Matt, Abdon, Kathryn, Rob, Sohyeon, Trini, and Derek — for their support, advice, conversations, and friendship.

I would like to thank my parents for everything they've done over the years. I am especially grateful for their constant encouragement and support for every decision I've made. I would like to thank my brothers, Sean and Ryan, for leading by example.

I'd like to thank Matt, Kerrigan, Tessa, Joe, and Abdon for all our adventures big and small.

I'd finally like to thank my MCSB family for making grad school as awesome as it can be. Their passion, encouragement, collaboration, and indeed sometimes commiseration, sustained me through all the highs and lows of graduate school.

This work was supported in part by NSF CAREER Award DMS1454739 to Jun Allard, NSF-Simons Center grant NSF DMS1763272, and Simons Foundation grant 594598 to Qing Nie.

Figures reproduced with permission from the copyright holders.

STATEMENT OF CO-AUTHORSHIP

Jun Allard[1,2,3,4] has been involved in the design, development, and analysis of all projects. In all projects, I have been involved in model development, performing simulations, and analyzing results.

In Chapter 2, a model of entropic multisite binding, Omer Dushek[5] helped design the research and provided critical feedback on model development and analysis.

In Chapter 3, a model of formin-mediated actin polymerization, Margot Quinlan[6] and Aanand Patel[6] assisted in model design and provided experimental data on formin polymerization rates.

In Chapter 4, measuring molecular reach, Omer Dushek[5] designed the research and Mikhail Kutuzov[5] conducted and analyzed the experimental data.

[1]Department of Mathematics, University of California, Irvine
[2]Department of Physics, University of California, Irvine
[3]Center for Complex Biological Systems, University of California, Irvine
[4]Center for Multiscale Cell Fate, University of California, Irvine
[5]Dunn School of Pathology, University of Oxford
[6]Department of Chemistry and Biochemistry, University of California Los Angeles

CURRICULUM VITAE

Lara C Clemens

Education

University of California, Irvine, Irvine, CA 2014 - current

- Ph.D., Mathematical, Computational, and Systems Biology
- Advisor: Jun Allard
- Thesis: Emergent characteristics and behaviors of multisite intrinsically disordered proteins

St. Lawrence University, Canton, NY 2010 - 2014

- B.S., Mathematics, Biochemistry

Publications

Zhang Y, **Clemens L**, Goyette J, Allard J, Dushek O, Isaacson SA. (2019). The influence of molecular reach and diffusivity on the efficacy of membrane-confined reactions, *Biophysical Journal*, 117:1189-1201. doi: `https://doi.org/10.1016/j.bpj.2019.08.023`

Bryant, D, **Clemens, L**, Allard J. (2017). Computational simulation of formin-mediated actin polymerization predicts homologue-dependent mechanosensitivity. *Cytoskeleton*. 74:29-39. doi: `https://doi.org/10.1002/cm.21344`

Mukhopadhyay H, de Wet B, **Clemens L**, Maini PK, Allard J, van der Merwe PA, Dushek O. (2016). Multisite phosphorylation of the T cell receptor zeta-chain enhances the potency but not the switch-like response. *Biophysical Journal*. 110:1896-1906. doi: `https://doi.org/10.1016/j.bpj.2016.03.024`

Articles in Preparation

Clemens L, Dushek O, Allard J. Modification of disordered multisite signaling molecules can create cooperativity and emergent sequential binding.

Grants and Fellowships

Center Fellow, NSF-Simons Foundation Center for Multi-scale Cell Fate Sept - Dec 2019
University of California, Irvine

Mathematical and Computational Biology Fellowship 2014 - 2015
University of California, Irvine

Undergraduate Researcher, Research in Industrial Projects for Students, Hong Kong Jun - Aug 2013
St. Lawrence University

Honors and Awards

Best Poster in Subgroup, IUPAB/EBSA Meeting July 2017
Best Poster Honorable Mention, Mechanobiology Symposium at UC Irvine July 2018

Invited Talks

UC Irvine MathBioU/MathExpLR, Irvine, CA, Jul 2019
St. Lawrence University Mathematics Q-Club Seminar, Canton, NY, Sept 2018
BIRS Workshop: Mathematics of the Cell, Banff, Canada, Aug 2018

Posters

Mechanobiology Symposium Irvine, CA, Oct 2019
Annual Cell Fate Symposium Irvine, CA, Oct 2019
Early Career Research Cell Fate Symposium Irvine, CA, Mar 2019
American Society for Cell Biology Annual Meeting San Diego, CA, Dec 2018
Mechanobiology Symposium Irvine, CA, July 2018
Biophysical Society Annual Meeting San Francisco, CA, Feb 2018
IUPAB/EBSA Meeting Edinburgh, UK, July 2017

Biophysical Society Annual Meeting — New Orleans, LA, Feb 2017
Joint Mathematics Meetings — Baltimore, MD, Jan 2014

Teaching Experience

University of California, Irvine
- **Instructor**, Math., Comp., and Systems Biol. Bootcamp — Sept 2019
- **Research Manager and Mentor** MathBioU, MathExpLR — Jul - Aug 2019
- **Teaching Assistant**, Calculus — Jan - Mar 2019
- **Research Manager and Mentor** MathBioU, MathExpLR — Jul - Aug 2018
- **Teaching Assistant**, MCBU — Jul 2017
- **Teaching Assistant**, Math., Comp., and Systems Biol. Bootcamp — Sept 2016, 2018
- **Teaching Assistant**, Calculus for Life Sciences — Sept 2015 - Jun 2016
- **Teaching Assistant**, CCBS Short Course — Jan 2016
- **Teaching Assistant**, COSMOS — Jun - Jul 2015
- **Instructor**, TechTrek — Jul 2015

St. Lawrence University
- **Peer Tutor**, Quantitative Resource Center — Jan 2012 - May 2014
- **Teaching Assistant**, Group Theory — Jan 2014 - May 2014

Additional Experience

Science Communications Course, Jan-Mar 2019

Opioid West Hackathon Participant, Oct 2018
Chosen as one of twenty five teams to participate in hackathon.

Business of Science Certificate Program, Oct 2018

Interdisciplinary Careers Panelist, CCBS Short Course May 2018

Mentoring Excellence Program, Apr - Jun 2015

ABSTRACT OF THE DISSERTATION

Characteristics and Emergent Behaviors of Multisite Intrinsically Disordered Proteins

By

Lara C Clemens

Doctor of Philosophy in Mathematical, Computational, and Systems Biology

University of California, Irvine, 2019

Associate Professor Jun Allard, Chair

Proteins perform diverse jobs around the cell, including transporting nutrients and transmitting signals across membranes. For a long time, the traditional understanding of proteins has been that a protein's structure determines its function. However, recent research shows that more than 40% of human proteins contain intrinsically disordered domains, or pieces of protein lacking a stable structure. These domains have been found in diverse biological systems, including associated with the cytoskeleton, immune receptor tails, and as linkers between structured domains. Given their abundance and variety, it has become critical to understand how disordered domains behave. Here we investigate how characteristics of disordered proteins — including length, binding site location, long- and short- lived interactions with other domains — influence their binding kinetics. Specifically, we consider three disordered proteins in different biological systems: (1) multiply phosphorylated T cell receptor cytoplasmic tails, (2) formins — a family of actin polymerization mediators, and (3) PD-1 tethered T cell signaling inhibition. For the T cell receptor cytoplasmic tails and formins, we use a freely-jointed chain model to simulate interactions between multisite disordered domains (E.g. T Cell Receptor, formin) and a binding partner (e.g. kinase, profilin-actin) represented as an idealized spherical domain. In each case, we demonstrate that disorder naturally leads to site-specific binding rates in multisite domains. For T cell receptor tails, we also demonstrate how physical changes due to post-translational modifications can give

rise to positive cooperativity, negative cooperativity, and ultrasensitivity. We next show how variation between formin homologues, such as dimerization state, length, and binding site locations, can lead to changes in actin polymerization rate. Lastly, we analyze surface plasmon resonance experiments of the phosphatase SHP-1 dephosphorylating its tether, PD-1. We couple worm-like chain models with differential equations to extract binding kinetics, catalytic rates, and molecular reach parameters from the data. We use this data to quantify how changes in the lengths of disordered domains influence the rate and reach of reactions. Together, these projects demonstrate the diversity and complexity of roles disordered proteins can play in cellular processes.

Chapter 1

Introduction

Proteins are the workhorses of the cell. Cells in the human body build proteins to perform jobs around the cell. These jobs are incredibly diverse, including transporting nutrients within the cell or across the cell membrane, transmitting signals from outside the cell to the nucleus, and even helping to build new proteins. The understanding of proteins for more than one hundred years has been that a protein's role in the cell is determined by its structure [1], much like a wrench can only loosen bolts of a specific size. However, the past several decades have collected evidence for a paradigm shift: not all functional proteins have a specific structure. Proteins which do not form a stable, specific structure, but instead take on a wide ensemble of conformations, are termed 'intrinsically disordered' proteins. Disordered proteins, or disordered domains of larger proteins complexes, have been found in a variety of contexts including associated with the cytoskeleton (e.g. Tau, MAP7), immune receptor tails (e.g. CD28, CD3ζ, PD-1), and as linkers between structured domains (e.g. SHP-1, ZAP70) [2–5].

Studying disordered domains has presented many challenges. Some traditional methods of understanding proteins experimentally, such as X-ray crystallography, do not extend well to

disordered domains. Experimental methods such as NMR spectroscopy, small X-ray scattering, and circular dichroism, can allude to the disordered nature of a domain, but elucidating their function — particularly how it relates to their disorder — remains problematic.

Recent investigations have turned to modeling to computationally explore how disordered proteins affect their biological systems. Many tools exist to model proteins, including varying levels of molecular detail from describing only kinetics with no spatial detail to including every atom in the system. Each of these has been applied to disordered proteins with varying degrees of success.

Differential equation models of proteins often describe protein-protein interactions or proteins moving through different conformational or chemical states using rate constants and concentrations. These types of equations are particularly useful for looking at long timescale behaviors describing interactions over a finite number of states. For example, Mukhopadhyay et al. use a system of differential equations describing the phosphorylation states of the disordered CD3ζ domain of the T Cell Receptor to demonstrate cooperative rate enhancement between phosphorylation events is needed to match experimental data [6]. These models can bring about important understanding of protein networks and contributions of disordered proteins to these networks, especially for understanding long timescale behaviors (seconds to minutes or more). However, these models lump effects from the physical features of the protein, such as transient steric occlusion by the disordered domain, into a single rate constant value which must be inferred from data.

On the opposite scale of molecular detail, all-atom molecular dynamics (MD) simulations describe each electrostatic interaction between atoms of the protein with itself and its surrounding environment including solvent, such as the cytoplasm, or the cell membrane. This type of model produces real-time simulations describing protein interactions, which can offer important insights into enzymatic mechanisms or how mutations will influence protein function. However, computing interactions between every atom and the atoms in its surroundings

is computationally expensive. Simplifying these models, for example by coarse-graining the solvent, can improve computational time, but simulations of milliseconds are still challenging to attain. Disordered proteins explore their conformational space on the time scale of microseconds to milliseconds [7]. Molecular dynamics simulations have been used to investigate behaviors of disordered proteins before [8], but often the timescales investigated are shorter than that of a full conformational sweep and obtaining an accurate sampling of conformational space is challenging [9]. All of these challenges are enhanced for when studying longer disordered domains such as those found in immune receptors (e.g. PD-1, CD3ζ; 30-120 amino acids) or on microtubules (e.g. tau; >100 amino acids).

To explore how the conformational flexibility of disordered proteins affects their interactions with structured domains, such as kinases, modelers have also used ideas from polymer physics to develop 'mesoscale' models, i.e. models of intermediate molecular detail. There are several cases where new biological understanding has been obtained by modeling disordered domains as random walks or worm-like chains (WLC). These types of models simplify disordered domains down to two main parameters: the length and the flexibility of the domain. More complicated features, such as angular restrictions, are typically ignored. Previous work has used this framework to generate qualitative and quantitative insights into systems with disordered regions. For example, Van Valen et al use a worm-like chain model to demonstrate how the length of a tether between two structured domains influences their effective concentration, such as in the auto-inhibition of WASP [10]. These models offer a way to access micro- to milli- second timescales while still accounting for some physical details such the ensemble of configurations a disordered protein can take on.

In the following chapters, we employ polymer models and differential equations to examine disordered proteins in various contexts. In Chapters 2 and 3, we represent each disordered domain as a freely-jointed chain, with a specific number and length of segments. We simulate interactions between disordered domains and a binding partner, e.g. kinase, represented as an

idealized spherical domain. At this level of molecular detail, we are able to efficiently explore how fluctuations in the configurations of the disordered domain affect the kinetics of the reactions. Additionally, we can simulate how factors influencing the configuration space (e.g. disordered-to-ordered transitions, tensile forces) will change the reaction. In Chapter 4, we couple worm-like chain models with differential equations to extract biophysical parameters from experimental data. At this scale, we are able to match time-series data on the timescale of seconds to minutes and quantify how changes in the lengths of disordered domains influence the rate and reach of reactions. The biological contexts and model extensions for each chapter is detailed in the following paragraphs.

One of the main receptors in the adaptive immune system, the T Cell Receptor, contains six disordered domains in its cytoplasmic region. One of their most important features is their multiple phosphorylation sites. Phosphorylation of these sites is one of the first steps in T Cell activation and developing an immune response. Previous work suggests that the T Cell Receptor disordered domains can exhibit cooperative enhancement of phosphorylation [6]. In Chapter 2, we simulate the domains of the T Cell Receptor as freely jointed chains to explore how phosphorylation could influence future binding events. We explore the disordered domains in three contexts: (1) phosphorylations create disordered-to-ordered transitions, (2) phosphorylation-dependent membrane association, and (3) simultaneous multiple binding of kinases to phosphorylated sites. In each case, we show how nonlinear signaling behavior, such as cooperativity and ultrasensitivity, can be achieved with post-translational modifications to a disordered domain.

In Chapter 3, we examine a family of disordered proteins called formins, which modulate actin polymerization rates. Formin associates with the growing end of actin, processing as the filament grows. The disordered domains of formin execute a capture-and-deliver process: they perform a search to capture actin monomers and, once bound, deliver them to the growing end of actin. There are fifteen different formin homologues in humans, each

with different characteristics, such as different lengths or number and location of binding sites. Notably, some formins exist dimerized at both the N-terminus and C-terminus of the disordered arms whereas others are only dimerized at the C-terminus. In this chapter, we explore how dimerization at both ends of the disordered domain affects the ability of formin to polymerize actin compared to those only dimerized at the C-terminus. This analysis offers insights as to what effect biological variation in formins has on actin polymerization.

The length of disordered domains has been shown to influence reactions in many systems, including auto inhibition of WASP, formin-mediated actin polymerization, and dephosphorylation of immune receptor tails [7, 10–12]. One example of this last case is seen with the phosphatase SHP-1. SHP-1 can bind to phosphorylated receptor tails, such as PD-1, and dephosphorylate any phosphorylated tails within reach (e.g. CD28). In this reaction, it is clear that the length of the tethers involved will influence how far away the two receptors can be and still permit an interaction. However, the size of the enzyme will also contribute to this. We call the length describing the size of the accessible region as the 'molecular reach' of the reaction. The molecular reach of reaction is made up of the reach of the tethers and of the enzyme. In Chapter 4, we use differential equations to fit experimental data describing these tethered dephosphorylation reactions in order to extract the molecular reach of reaction. We fit reactions of SHP-1 over many different synthetic disordered peptides of different lengths in order to determine the reach of SHP-1 alone. Using this, we then determine the reach for several immune receptors including PD-1, CD28, and SLAM. Understanding how the reach of the enzyme and the reach of the tethers influence the reaction gives deeper intuition as to how disordered tethers influence immune networks.

Chapter 2

Entropic Multisite Binding

2.1 Introduction

Traditional perceptions of protein function centered around the tenet 'structure determines function'. Contrary to this, a growing body of work reveals functional intrinsically disordered proteins (IDP), or proteins which take on an entire ensemble of conformations. IDPs are in fact abundant in the proteome: more than 40% of human proteins contain intrinsically disordered regions longer than 30 amino acids [13, 14]. Disordered domains appear in many contexts including as linkers between globular domains (e.g. Src) [15], associated with the cytoskeleton (e.g. formin) [16, 17], or in signaling networks (e.g. CD3ζ, PD-1) [18, 19]. These regions often include sites for binding and post-translational modification, indicating purpose beyond as passive tethers [15, 20, 21]. Additionally, the length of the domains themselves has been suggested to influence their interactions - affecting binding kinetics and catalytic abilities [7, 10–12].

An example of significant interest is the disordered domains of the T Cell Receptor (TCR). T cells play a critical role in the adaptive immune response by recognizing and discriminating

between stimuli (i.e. antigens). The T cell receptor binds antigens, transmitting information through the cell to mount a fast, sensitive, and specific response. Upon T cell activation, an extensive signaling cascade leads to cytokine production, proliferation, differentiation, and target killing [22, 23].

The T cell receptor is one of the most complicated immune receptors, composed of eight domains complexed together: TCR$\alpha\beta$ and CD3$\delta\epsilon$, CD3$\gamma\epsilon$, CD3$\zeta\zeta$. Of these, the subunits CD3ζ, CD3ϵ, CD3δ, CD3γ all contain long intrinsically disordered cytoplasmic tails. Among the six disordered domains, there are twenty tyrosines paired in immunoreceptor tyrosine-based activation motifs (ITAMs). Most immune receptors contain only a single ITAM, consistent with CD3δ, CD3γ, and CD3ϵ, however each CD3ζ has three ITAMs [24]. After antigen engagement, the kinase Lck phosphorylates the TCR marking one of the first intracellular steps of T cell signaling [24].

Over the course of T cell activation, the disordered domains of the T cell receptor have many interaction partners. Prior to TCR triggering, constitutively active kinases (Lck, Fyn) and phosphatases (CD45) compete to phosphorylate and dephosphorylate the cytoplasmic domains. Such a competition runs the risk of introducing false signaling, where T cell activation occurs without proper stimulus. Guarding against this, the CD3ζ and CD3ϵ domains associate with the inner leaflet of the cell membrane, effectively sequestering the tyrosines from phosphorylation [25–27]. After ligation by an antigen, the cytoplasmic tails no longer associate with the cell membrane and Src-family kinases (e.g. Lck) phosphorylate the T cell receptor [25, 28–30]. Phosphorylation of the ITAMs allows a second kinase, ZAP70, to tandemly bind to the phospho-tyrosines and proceed to phosphorylate downstream molecules such as LAT and SLP-76 [5].

Despite lack of structure, these disordered domains display nonlinear behaviors, including cooperative phosphorylation [6]. For structured proteins, cooperative effects often emerge

from conformational changes [31]. For disordered domains, this begs the question: can nonlinear behaviors arise as emergent properties from disordered domains?

Modeling disordered domains presents interesting challenges due to their large conformational space. Representative samplings of conformation space take on the order of milliseconds, a timescale that is computationally taxing to reach with current atomistic molecular dynamic methods. Alternative modeling methods borrow ideas from polymer physics: representing disordered domains as random walks or worm-like chains [10, 11, 32, 33]. These models have offered qualitative and quantitative understanding of disordered regions in a number of biological contexts. For example, Van Valen et al use a worm-like chain to model tethered receptor-ligand kinetics for N-WASP promotion of actin polymerization, demonstrating a tether length dependence for signaling [10]. Other systems where polymer models have contributed insights include immune receptor tails, the formin family of actin associated proteins, and the flexible kinesin neck linker [6, 7, 11, 34]. This progress is despite the simplicity of polymer models: they ignore excluded volume effects and electrostatic interactions of the chain. To use these methods for systems like the TCR cytoplasmic domains, we would need to consider interactions between not only the disordered domains, but also the membrane and relevant ligands. Given the simplicity of a polymer model, what features of the interaction between a disordered region and a globular domain, membrane or other disordered domain can be captured by entropy calculations of freely jointed chains interacting with rigid spheres and planes?

Here we simulate a disordered domain as a freely-jointed chain in an implicit solvent interacting with a spherical ligand. We investigate how properties of disordered domains influence interactions with ligands, taking as a case study the cytoplasmic domains of the TCR interacting with associated kinases, Lck and ZAP70. We explore how these disordered domains behave in various contexts: (1) disordered-to-ordered transitions upon phosphorylation, (2) phosphorylation-dependent membrane association, and (3) simultaneous multiple binding of

ZAP70 to the full TCR (Fig. 2.1). Membrane interactions with the disordered domains are modeled two different ways: In models (1) and (3), the membrane is treated as an impenetrable wall, while in model (2) the membrane is a soft barrier dependent on amino acid properties. From these simulations, we calculate effective binding rates of kinases to the TCR. We find that the entropic fluctuations of disordered domains naturally impose differential binding rates between binding sites, which can lead to preferential sequential binding. We also show that under different assumptions of post-translational modification or binding effects, disordered domains can exhibit emergent cooperativity (positive and negative) and ultrasensitivity.

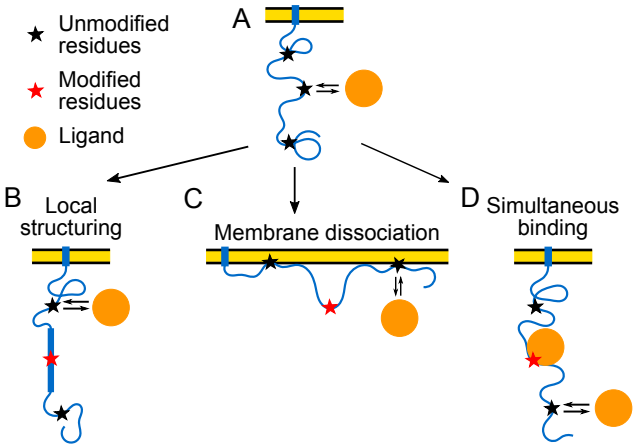


Figure 2.1: **Possible consequences of membrane-bound disordered protein interacting with ligand.** (A) Interaction of ligand (orange circle) with membrane-bound disordered protein with multiple binding sites (black stars) where (B) binding causes local stiffening near post-translational modification (e.g. phosphorylation; red stars), (C) phosphorylation reduces membrane-association of the polymer or (D) the ligand remains bound while more ligands attempt to bind simultaneously.

2.2 Methods

2.2.1 Polymer model of disordered protein

We create a generic model of a disordered protein using a simplified θ-solvent freely-jointed chain (FJC) from polymer physics. This model requires only specifying a number of rods (N) and a length per rod (Kuhn length, δ). The FJC consists of N rigid rods of length δ which are allowed to perform a random walk where the only constraints are the chain length and connections. The ligand is modeled as an idealized sphere which may interact with the FJC.

We compute quasi-equilibrium statistics of the chain and its ligands using a Monte Carlo (Metropolis) Algorithm. At each proposed configuration, the Metropolis algorithm computes the energy of the system and accepts or rejects based on energetic constraints. In the simplest model with only the freely-jointed chain, all configurations are equally energetically favorable and the FJC explores its configuration space through randomized movements. In models with a membrane, a hard constraint is included at $z = 0$ such that configurations passing through the hard membrane are rejected.

We model the intrinsically disordered cytoplasmic domains of the TCR, CD3ζ,ϵ,δ,γ, as freely jointed chains with 113, 55, 47, and 45 segments respectively. Here, we represent each segment as an amino acid (Kuhn length = 0.3 nm). We consider a spherical estimate for the kinases Lck and Zap-70, with radii 2.1 nm (7 Kuhn lengths) and 2.7 nm (9 Kuhn lengths) respectively. Estimates for Lck and Zap-70 sizes are measured from PDB files (3LCK and 2OQ1 respectively) using PyMol. For simulations of the full TCR, the relative location of the membrane-anchor for each cytoplasmic domain is estimated from PDB 6JXR with PyMol.

2.2.2 Calculation of ligand binding rates from occlusion statistics

We model the polymer in the canonical ensemble, i.e., equilibrium. Steric occlusion of binding sites by the rest of the chain therefore gives rise to a change in K_D. From detailed balance, we have $K_D \equiv k_{\text{off}}/k_{\text{on}}$, which allows us to compute the change in K_D between two states, e.g., fully phosphorylated compared to dephosphorylated, as follows:

$$\frac{k_{on}}{k_{off}} = \exp\left(\frac{-\Delta G}{k_B T}\right) \tag{2.1}$$

$$\frac{K_{D_1}}{K_{D_2}} = \exp\left(\frac{\Delta G_1 - \Delta G_2}{k_B T}\right) \tag{2.2}$$

$$= \exp\left(\frac{(E_1 - TS_1) - (E_2 - TS_2)}{k_B T}\right) \tag{2.3}$$

$$= \exp\left(\frac{(k_B \ln(\frac{\Omega_2 P_2}{\Omega_2})) - (k_B \ln(\frac{\Omega_1 P_1}{\Omega_1})}{k_B}\right) \tag{2.4}$$

$$\frac{K_{D_1}}{K_{D_2}} = \frac{P_2}{P_1} \tag{2.5}$$

where $G_j = E_j - TS_j$ is the free energy of binding in a given state, $S_j = k_B \ln W$ is the entropy of binding, where W is the number of microstates and P_j is the probability in the canonical ensemble that the configuration allows for binding. We let Ω_j be the total number of microstates, with $\Omega_j P_j$ being the microstates available when the ligand is bound for a given polymer state (e.g., dephosphorylated). We also note $\Delta E_1 = \Delta E_2$ since the change in energy due to ligand binding is the same, regardless of conformation.

We define P_{occ} as the probability that the region of space needed by the kinase domain is occupied by some of the polymer, or another steric barrier. Thus, $P_{occ_j} = 1 - P_j$. Our problem of interest is now reduced to computing the occlusion probability.

$$\frac{K_{D_1}}{K_{D_2}} = \frac{1 - P_{occ_2}}{1 - P_{occ_1}}. \tag{2.6}$$

Steric occlusion by the polymer impacts the ability of a ligand to localize to the binding site. Although entropic forces could also impact unbinding of the polymer, we assume this influence to be negligible compared to the change in k_{on}. Therefore, for the body of this work, we assume that the change in K_D stems from a change in k_{on}.

The occlusion probability is computed as the probability a ligand is able to bind to an oriented sphere tangentially attached to the polymer. We define 'able to bind' as when the specified sphere is empty of both other polymer segments and any other constraints (e.g., membrane). In each conformation, the location of the segment end points and ligands are checked for chain-ligand intersections. If a membrane is present, a second constraint checks if the ligand sphere crosses below the half-space surface designated at $z = 0$. Given that the ligand sphere is large compared to the Kuhn length, we assume the probability of tangential occlusion (where a segment has end points outside the sphere but part of the segment lies within) is negligible compared to end point occlusion.

2.2.3 Simulation of multi-step processes from individual rates

We use a six-step Gillespie algorithm to make two calculations: 1) the probability of a specific sequence of irreversible (de)phosphorylation, and 2) the sequence-weighted average binding rate to transition between phosphostates (e.g. from one to two total phosphorylations). A matrix of binding rates is created from the Metropolis simulations, including the binding probabilities to each site in each (de)phosphorylation state. Each new state is chosen as the minimum of a exponentially distributed random variable based on the binding rates to the remaining sites in that state. At the end of each run of the six-step process, we record the event sequence and the times to transition between each step. Probabilities of each sequence are computed based on the total iterations of the algorithm. For simplicity, only the probabilities for the two perfectly ordered sequences are considered. The sequence-

weighted average binding rates are calculated as the inverse of the average transition time for a specific step.

We use a reversible Gillespie algorithm to create the dose-response curves. The rate matrix contains binding rates of a kinase to a disordered domain in a given phosphorylation state. Under constant dephosphorylation, a constant rate is set for the dephosphorylation rate of each site in each phosphorylation state. For sterically-limited dephosphorylation, the binding rates of a phosphatase binding to phosphorylated sites in a disordered domain as determined from the Metropolis algorithm are used for the reverse reaction rates. The phosphatase is assumed to be of equivalent size to the kinase. The effective ratio of kinase to phosphatase is controlled by multiplying the reverse rates by a constant value. Each step in the algorithm is chosen as the minimum of a exponentially distributed random variable based on the binding rates in that state. The average number of phosphorylated sites is calculated once the system reaches a steady state.

2.2.4 Models for consequences of site modification

Local stiffening

Local stiffening is simulated as a limitation on which joints of the freely jointed chain are allowed to rotate. For each modified residue, some number of joints are 'frozen' in a straight conformation. The section of frozen segments acts as a single long segment, still rotating when neighboring segments rotate. This method of local stiffening leaves the modified binding site in a primarily available configuration. The number of segments frozen per modification, or degree of stiffening, is varied.

Similarly, dephosphorylation is modeled as 'unfreezing' nearby segments. The initial polymer state is fully phosphorylated with the parallel amount of local stiffening. For simplicity, the

phosphatase is assumed to be the same size as the ligand Lck. For dephosphorylation, the degree of stiffening is limited to five segments on each side of the binding site to avoid the ambiguity of overlapping regions of influence from different binding sites.

Membrane affinity

The polymer-membrane association is modeled as a potential acting on each segment of a freely-jointed chain. Each potential only acts in the direction of the membrane (defined as $z = 0$). We consider three groups of residues: tyrosines, phosphotyrosines, and basic residues.

Basic residues are needed to create membrane-association [25] and are therefore modeled as having an attractive interaction with the cell membrane. To match an approximately Gaussian distribution of both the tyrosines and polymer along the membrane edge, a parabolic-constant piecewise potential is used for the basic residues with depth E_{B0} [8] (Eq. 2.7).

$$E_{\text{Basic}}(z_i) = \begin{cases} k_{PC} z_i^2 - E_{B0} & z_i < \sqrt{\frac{E_{B0}}{k_{PC}}} \\ 0 & z_i \geq \sqrt{\frac{E_{B0}}{k_{PC}}} \end{cases} \tag{2.7}$$

Tyrosine phosphorylation is sufficient to dissociate the polymer from the membrane [25]. We therefore model phosphorylated tyrosines as having a repulsive interaction with the membrane. Since phosphotyrosines interacting with a membrane are equivalent to charged particles interacting with a surface, the membrane-phosphate interaction is modeled as an exponential distribution (Eq. 2.9).

The Debye length, z_{Debye}, describes the length scale over which electrostatic interactions are felt in a medium (Eq. 2.8).

$$z_{\text{Debye}} = \sqrt{\frac{1}{4\pi l_B c_\infty}} \tag{2.8}$$

where l_B is the Bjerrum length ($\sim 0.7\,\text{nm}$ in water) and c_∞ is the concentration of ions in the cytoplasm ($\sim 150 - 200\mu M$) [35]. From this, we can estimate the Debye length for the cytoplasm as approximately 1 nm or 3 Kuhn lengths.

We therefore set the range of membrane interactions to be 1 nm. Specifically, $z_{Debye} \sim 1\,\text{nm}$ and $\sqrt{E_{B0}/k_{PC}} \sim 1\,\text{nm}$.

$$E_{Y_p}(z_i) = E_{P0} e^{-z_i/z_{\text{Debye}}} \tag{2.9}$$

The membrane still acts as a steric barrier to the remaining amino acids. A soft wall constraint is applied to the unphosphorylated tyrosines and remaining amino acids (Eq. 2.10). This imposes an energetic penalty on amino acids entering the membrane. Given that amino acids are fairly small compared to lipids, we assume that the ability of an amino acid to penetrate the cell membrane is dependent on hydrophobicity more than steric influences. The average hydrophobicity of CD3ζ is estimated based on a hydrophobicity scale [36] taking into account the amino acid composition of CD3ζ. From this, k_S is set to $0.05\text{k}_\text{B}\text{T}$ for all simulations of CD3ζ.

$$E_S(z_j) = \begin{cases} k_S z_j^2 & z_j < 0 \\ 0 & z_j \geq 0 \end{cases} \tag{2.10}$$

Multiple simultaneous binding

To simulate simultaneous binding, each bound ligand is included as an idealized sphere tangentially bound to a polymer. The bound ligands rotate with the polymer, but maintain their orientation tangent to their binding site. The polymers may not occupy the same space as the bound ligands, nor may bound ligands occupy the same space as other bound ligands. When a membrane is present, the bound ligands cannot penetrate the membrane. For these simulations, the bound and unbound ligands are the same size. The unbound, incoming ligand is simulated as an idealized ghost sphere tangent to a single segment. In this modification, a binding site is unoccluded in a given conformation if the specified sphere is empty of polymer segments, membrane, and bound ligands.

2.3 Results

2.3.1 I. Local stiffening

One way in which disordered protein can participate in signaling cascades is to undergo a disorder-to-order transition [20, 37] upon post-translational modification (e.g., tyrosine phosphorylation), becoming locally stiff, thereby changing the kinetics of future modifications. This is shown schematically in Fig. 2.1B.

Previous experimental-modeling work [6] reconstituted the early TCR signaling cascade ectopically *in vivo*. The results suggest that the tyrosines on the TCR ζ chain undergo phosphorylations at different rates. Specifically, the most consistent model had phosphorylation rates that were enhanced by previous phosphorylations. Can this rate enhancement be explained by phosphorylation-induced local stiffening?

Disorder leads to entropic binding rate differences between phosphorylation sites

We first calculate the binding rate to each of the six tyrosines on an unphosphorylated, membrane-bound CD3ζ-chain. We find that the binding rate to the tyrosine closest to the membrane is about 3-fold lower than the binding rate of the tyrosine furthest from the membrane, for an enzyme approximately the size of Lck (Fig. 2.2). Intuitively, the membrane creates a steric barrier. Binding sites proximal to the membrane experience many configurations where a large enzyme will not have room to bind. Sites further along the disordered domain have a larger volume they are able to explore further away from the membrane, increasing the likelihood of configurations where an enzyme could bind.

Rate differences can also persist without the presence of a membrane. We simulate an unphosphorylated, cytosolic CD3ζ-chain to explore how the presence of a membrane influences these rate differences. We find that binding sites near the ends of the disordered domain have binding rates 2- to 3-fold higher than binding rates towards the middle (data not shown). In this case, binding site occlusion occurs solely from the rest of the disordered domain. The ends of the disordered domain can more easily explore space away from the rest of the domain, where there is enough unoccupied volume for the enzyme to bind. In comparison, the binding sites closest to the middle experience occlusion from both ends of the domain, leading to lower binding rates.

Each tyrosine of CD3ζ experiences a different environment based on its location along the disordered domain. The entropic fluctuations of the disordered domain naturally creates different binding rates to each phosphorylation site.

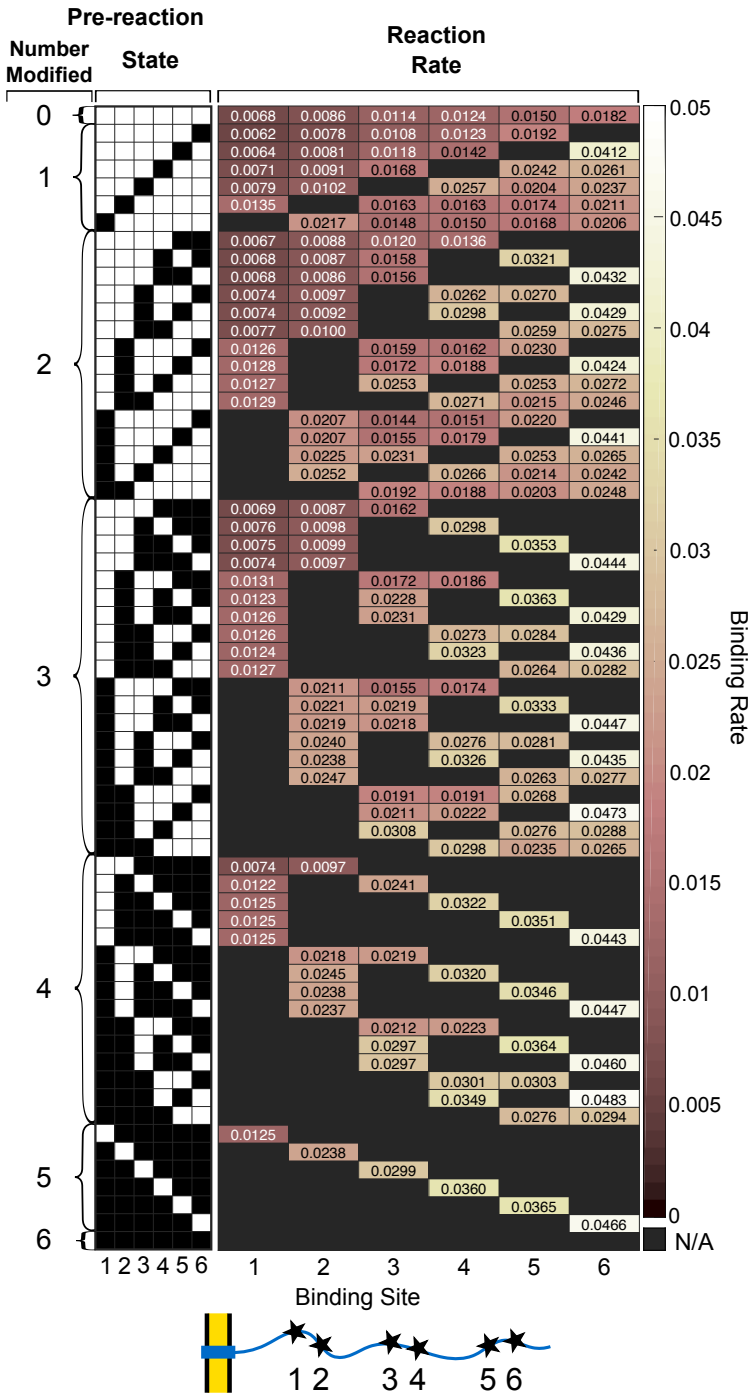


Figure 2.2: **Disorder leads to entropic binding rate differences between phosphorylation sites.** Reaction rates of a kinase binding to the six binding sites of membrane bound CD3ζ in a given phosphorylation state, assuming each phosphorylation event locally stiffens 11 amino acids (dark red: low binding rate; white: high binding rate; black: phosphorylated site). Pre-reaction number modified and state indicate the number and configuration of phosphorylated sites on CD3ζ before the next binding event (white: unphosphorylated site; black: phosphorylated site). The six binding sites have different rates prior to the first phosphorylation event. Cartoon indicates location of binding sites along CD3ζ relative to the membrane.

Rate differences imply emergent preference of phosphorylation sequence

There are six different binding sites on the CD3ζ chain, offering 720 (6 factorial) possible sequences for phosphorylation or dephosphorylation, assuming each process is irreversible. Using a Gillespie algorithm in conjunction with the binding rates found above, we compute the probability of (de)phosphorylating in each sequence. For simplicity, we compare only two: where (de)phosphorylation occurs membrane proximal-to-distal (123456) and membrane distal-to-proximal (654321). For membrane bound CD3ζ, there is an approximately 10-fold increase in the probability of binding membrane distal-to-proximal compared to membrane proximal-to-distal for both the kinase and phosphatase (Fig. 2.3 A.i,B.i). Intuitively, binding sites close to the membrane are sterically occluded by both the polymer and membrane whereas binding sites distal to the membrane are primarily influenced by the polymer and experience less occlusion by the membrane. This indicates that disordered domains in the presence of a membrane can have an emergent preference of phosphorylation sequence.

Simulations of a cytosolic CD3ζ show reduced preference for phosphorylating membrane distal-to-proximal, supporting intuition. However, phosphorylating membrane distal-to-proximal is still 2-fold more likely than phosphorylating proximal-to-distal (Fig. 2.3 A.ii,B.ii). When we re-simulate cytosolic CD3ζ, but now with its six tyrosines spaced evenly along the amino acid sequence, there is no preference for phosphorylating membrane distal-to-proximal compared to proximal-to-distal (Fig. 2.3A.iii,B.iii). This can be understood by comparing the locations of the tyrosines in CD3ζ. In CD3ζ, the membrane distal tyrosine is twelve amino acids away from the C-terminal while the membrane proximal tyrosine is twenty-one amino acids away from the membrane. For cytosolic CD3ζ, the extra amino acids between the first tyrosine and the membrane proximal end increases the steric hindrance that binding site experiences. These results show that changing binding site locations by as little as ten amino acids is sufficient to induce a preferential binding sequence.

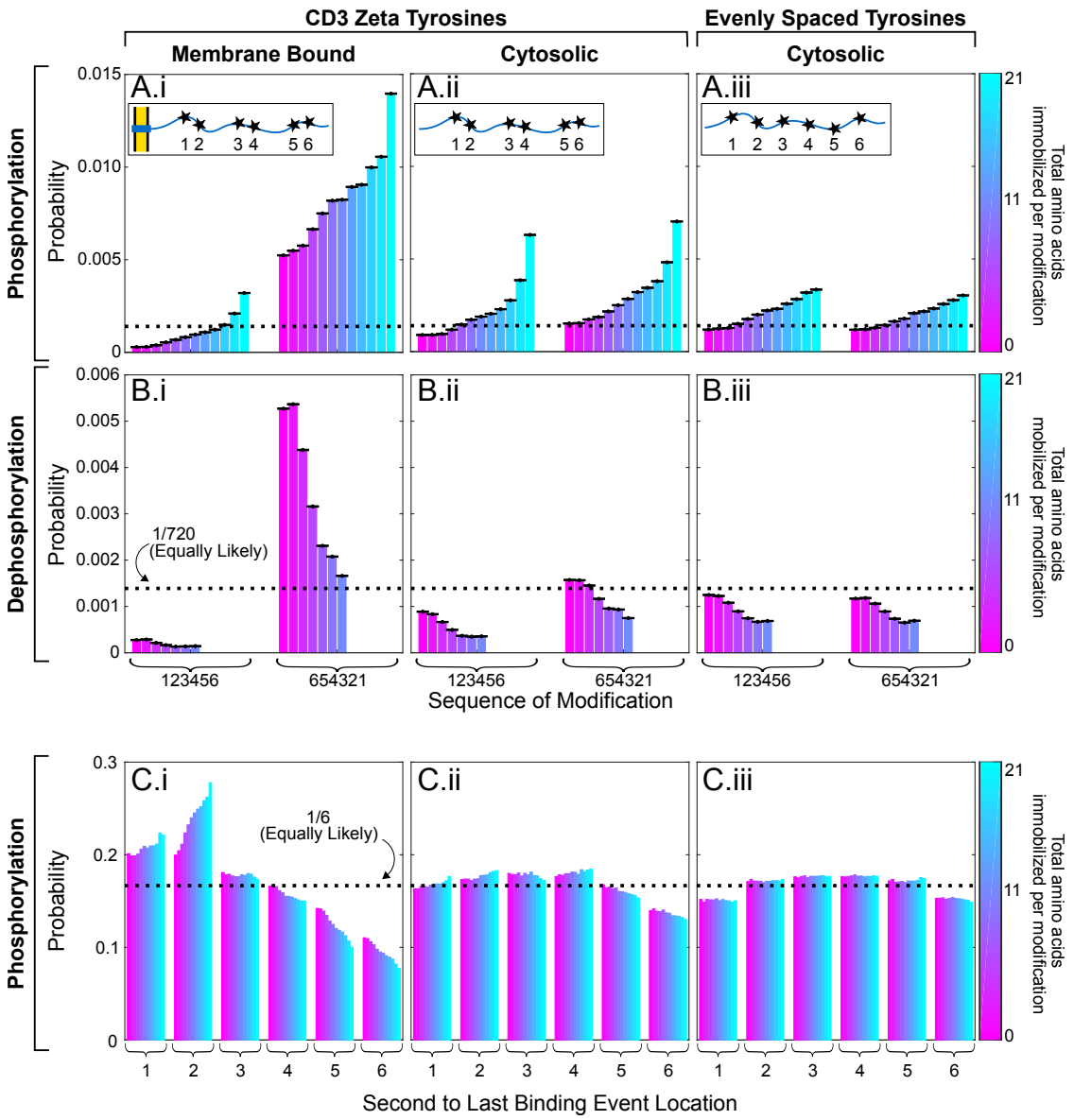


Figure 2.3: **Binding rate differences imply emergent preference of (de)phosphorylation sequence.** Probability of (A) phosphorylating or (B) dephosphorylating membrane proximal-to-distal (123456) compared to membrane distal-to-proximal (654321) for different ranges of local stiffening/unstiffening per binding event, and (C) probability of the second to last binding event occurring at a specific location (i) with a membrane, assuming sites spaced like tyrosines of CD3ζ, (ii) without a membrane, assuming sites spaced like tyrosines of CD3ζ, and (iii) without a membrane, assuming sites spaced evenly along the length of CD3ζ. Probability if all events are equally likely (A,B: 1/6!; C: 1/6) indicated with black dotted line. Error bars (A,B) represent standard error of the mean, treating each Gillespie run as an individual Bernoulli trial. Phosphorylation and dephosphorylation are more likely to occur membrane distal-to-proximal compared to proximal-to-distal for both the membrane bound and cytosolic CD3ζ chain. For cytosolic CD3ζ with evenly spaced tyrosines, both (de)phosphorylation sequences are approximately equally probable.

Local stiffening modulates binding rates to multi-site disordered domain

We introduce local stiffening upon phosphorylation to examine how disordered-to-ordered transitions influence kinase binding rates. We compute the average binding rate of a kinase to the membrane-bound CD3ζ domain in its unphosphorylated state. Here, each average binding rate is weighted by the most likely (de)phosphorylation sequences. Without disordered-to-ordered transitions, the average kinase binding rate decreases as more phosphorylations occur. Since there is a natural preference to phosphorylate membrane distal-to-proximal, it is most likely that the membrane-proximal tyrosine will be the last tyrosine phosphorylated. This tyrosine has a lower binding rate, bringing down the average binding rate of the last phosphorylation compared to the first event, when the faster, membrane-distal sites dominate.

However, if phosphorylation introduces enough local stiffening, then a single phosphorylation event creates an overall increase in the average binding rate of the kinase to another tyrosine. This effect increases with total phosphorylations and degree of local stiffening per phosphorylation. Specifically, if each phosphorylation locally stiffens 11 amino acids ($\sim$1/12 of the polymer length) then the sixth kinase binding event will occur almost two times times faster than the first (Fig. 2.4A). This phenomenon can be intuitively understood as local stiffening reduces the degrees of freedom of the polymer. This causes the polymer to be more elongated on average and sample more configurations where the remaining tyrosines are kinase-accessible. Through this mechanism, disordered-to-ordered transitions can increase the binding rate of a kinase to the remaining unphosphorylated tyrosines.

Interestingly, there is a dip in the average binding rate between the fourth and fifth binding event. For membrane-bound CD3ζ, the fifth binding event is most likely to occur at the second tyrosine with the first tyrosine as next most likely (Fig. 2.3Ci). Therefore, the average binding rate is most heavily influenced by the binding rate at the second tyrosine. The

dip in binding rate seen there marks a compeitition between disordered-to-ordered transitions making all sites more accessible to phosphorylation versus the polymer and membrane occlusion. Compared to the occluding effect of the membrane at the second tyrosine, the reduction in chain occlusion created by disordered-to-ordered transitions of the other tyrosine is not enough to increase binding rate of the fifth event over the fourth. However, for the sixth event, the additional, nearby stiffening increases the binding rate to the last tyrosine enough to increase the average binding rate to above previous levels (Fig. 2.4A). Further explorations of this competition indicate that depending on the magnitude of local stiffening, the location of the binding sites and in particular the grouping of tyrosines into ITAMs, can create significant changes in the average binding rates (Fig. 2.6)

While local stiffening of phosphorylated tyrosines enhances the binding rate of kinases to unphosphorylated tyrosines, it also enhances the binding rate of phosphatases to phosphorylated tyrosines. We therefore explore how dephosphorylation (or loss of local stiffening) impacts the binding rate of phosphatases to the domain. If each phosphorylation locally stiffens 11 amino acids, then each dephosphorylation similarly relaxes 11 amino acids. For this degree of local stiffening, a fully phosphorylated CD3ζ chain would have more than half of its length stiffened. As dephosphorylation occurs, entropic flexibility is returned to the polymer. By the sixth dephosphorylation event, the average phosphatase binding rate is decreased by half (Fig. 2.4B). This shows that ordered-to-disordered transitions can also modulate binding rates, decreasing the binding rate of a phosphatase to the remaining phosphorylated tyrosines.

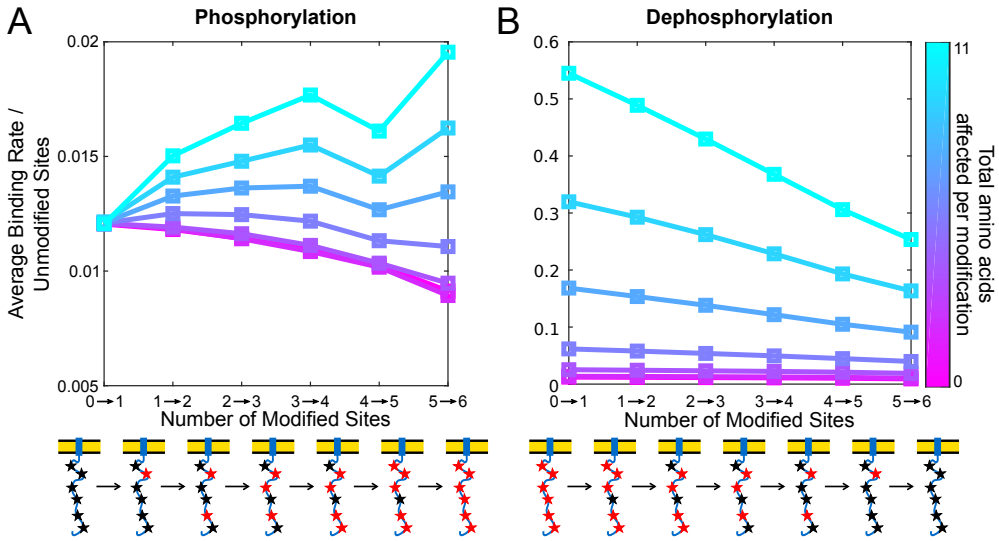


Figure 2.4: **Local stiffening modulates binding rates to multi-site disordered domain.** (A) Average binding rates of a kinase binding to CD3ζ at different phosphorylation states for varying range of local stiffening (pink: no residues are stiffened; blue: 11 residues are stiffened, 5 on each side of site). The kinase average binding rate increases with each phosphorylation and also with local stiffening range. (B) Average binding rates of a phosphatase binding to CD3ζ at different dephosphorylation states for varying ranges of local un-stiffening per dephosphorylation event. When local stiffening occurs, the phosphatase average binding rate decreases with each phosphorylation and also with local stiffening level. For both (A) and (B), schematic below axis shows example configuration for each phosphorylation state. Unphosphorylated residues represented by black stars, phosphorylated residues are red stars. Kinase and phosphatase radius is 2.1nm. Rates are normalized to the free-space binding rate.

Binding rate cooperativity creates ultrasensitivity, even in reversible symmetric phosphorylation/dephosphorylation cycles

We explore two reversible systems: 1) constant dephosphorylation and 2) steric dephosphorylation. In each case, a large kinase attempts to phosphorylate the tyrosines and its binding rate is affected by steric influences from the chain and the membrane.

In the first system, we assume the phosphatase is of negligible size. Therefore, dephosphorylation is not impacted by the sterics of the chain or membrane and instead occurs at a constant rate at each site. Using the site-specific binding rates determined above for membrane bound CD3ζ, we calculate the steady state fraction of sites phosphorylated through a Gillespie algorithm. When no local stiffening occurs, reversible phosphorylation occurs with approximately Michaelis-Menten kinetics with a Hill coefficient of 1. However, as the number of amino acids stiffened per phosphorylation increases, the curves diverge from Michaelis-Menten kinetics. Qualitatively, an increase in the slope of the curves can been seen as the degree of local stiffening increases (Fig. 2.5A). This change can be quantified as an effective Hill coefficient, here defined as the maximum logarithmic slope of each curve. This value increases with the degree of local stiffening per phosphorylation, attaining a Hill coefficient of 1.8 when approximately 1/12 of the chain is stiffened per phosphorylation event. Therefore, the binding rate cooperativity introduced by local disordered-to-ordered transitions can lead to ultrasensitivity in a signaling network.

In the second system, we assume the phosphatase is large and therefore sterically influenced in the same manner as the kinase. Without local stiffening, this reversible system also shows approximately Michaelis-Menten behavior with Hill coefficient 1. Local disordered-to-ordered transitions creates a slight increase in the slope of the dose-response curves. At 1/12 chain local stiffening per phosphorylation, the Hill coefficient is 1.3 (Fig. 2.5B). This suggests that even under steric dephosphorylation, a reversible system with local disordered-

to-ordered transitions is still capable of creating mild ultrasensitivity, dependent on the degree of local stiffening.

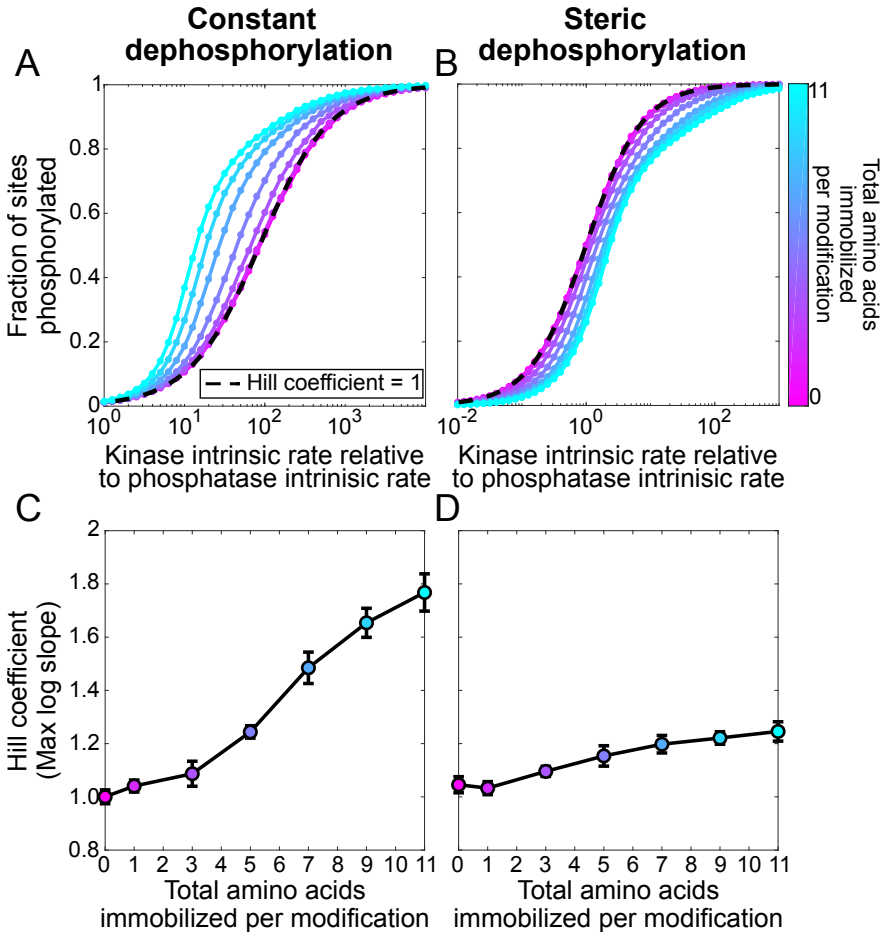


Figure 2.5: **Emergent cooperativity from binding rate enhancement makes a good switch, even in symmetric phosphorylation/dephosphorylation cycles.** (A,B) Fraction of sites phosphorylated as a function of the kinase-to-phosphatase activity ratio (measured by the ratio of free-space rates) for different ranges of local stiffening (colorbar), assuming a phosphatase with (A) negligible size or (B) radius of 2.1 nm, equal to the kinase. Black dashed line indicates a linear dose-response, i.e., with Hill coefficient 1. (C,D) Hill coefficients for different ranges of local stiffening per phosphorylation event, assuming a phosphatase of (C) negligible size or (D) radius of 2.1 nm. Hill coefficients calculated from maximum log-log slope of the dose response curves. Error bars indicate root-mean-square error from a cubic polynomial fit to slope. In both cases, the Hill coefficient increases with the range of local stiffening, even when dephosphorylation is assumed to be sterically hindered (D).

2.3.2 Supplemental to local stiffening

Local stiffening modulates sequence-dependent binding rates to multi-site disordered domain

We explore how varying the local stiffening affects the sequence-weighted average binding rates for a cytoplasmic CD3ζ. For maximal local stiffening, i.e. all residues becoming stiff upon first phosphorylation, the binding rate is 1 for subsequent phosphorylations (Fig. 2.6A). This indicates the kinase is uninhibited and always able to access the binding sites. This is expected, since here local stiffening puts the binding site in a primarily available configuration.

Interestingly, at intermediate values of local stiffening (e.g. 19-67 residues stiffened per phosphorylation), the average binding rate does not consistently increase but oscillates (Fig. 2.6A). This phenomenon can be understood by considering the distribution of tyrosines along CD3ζ. The six tyrosines are grouped in pairs, signaling motifs called ITAMs. The tyrosines in each pair are signficantly closer to each other than to the tyrosines of the next ITAM. This uneven distribution of tyrosines leads to uneven response to local stiffening upon phosphorylation. After the first phosphorylation occurs, the induced local stiffening reduces the flexibility of the domain and increases the likelihood of nearby sites being accessible to the kinase. For sites within the local stiffening region, this is intuitive since the stiffening will place them in a primarily available configuration, only restricted by the fluctuations of the unstiffened sections of the domain.

For sites outside the local stiffening region, this effect is lessened with distance. Sections closest to the binding site are most likely to cause occlusion, since they are restricted to be consistently nearby. Therefore, tyrosines close to the local stiffening site will experience the most influence from the local stiffening. How does this cause the average binding rate

to decrease from the second to the third event? Since tyrosines are paired, the first phosphorylation will create local stiffening at one tyrosine of a pair. The second tyrosine of the pair will experience a increase in entropic availability — significantly more than the next closest tyrosine in another ITAM. This means the second event will most likely occur at the second tyrosine of the pair, where the binding rate is now much faster. However, after this second event has occurred, the tyrosines in other ITAMs are far enough away to not benefit from the local stiffening to the same degree. Therefore, the average binding rate to the next ITAM is considerably slower. When local stiffening is large enough to stiffen residues up to the next ITAM, the oscillation is lost and the average binding rate consistently increases. We can see that when we simulate CD3ζ with evenly spaced tyrosines, this feature is also lost (Fig. 2.6B).

Local stiffening modulates sequence-independent binding rates to multi-site disordered domain

For comparison, we consider the average binding rate of a kinase to the membrane-bound CD3ζ if all possible phosphorylation sequences were equally likely. For example, assuming any two tyrosines are phosphorylated on the domain with no preference for which ones, what rate will the third phosphorylation occur on average? Consistent with intuition, without disordered-to-ordered transitions, the average kinase binding rate remains constant regardless of how many phosphorylations occur. However, if phosphorylation introduces local stiffening, then a single phosphorylation event creates an increase in the average binding rate of the kinase to another tyrosine. This effect increases with total phosphorylations and degree of local stiffening per phosphorylation. Specifically, if each phosphorylation locally stiffens 11 amino acids ($\sim$1/12 of the polymer length) then the sixth kinase binding event will occur three times faster than the first (Fig. 2.7A). Without weighting by most likely phosphorylation states, the average binding rate smoothly increases with number of phos-

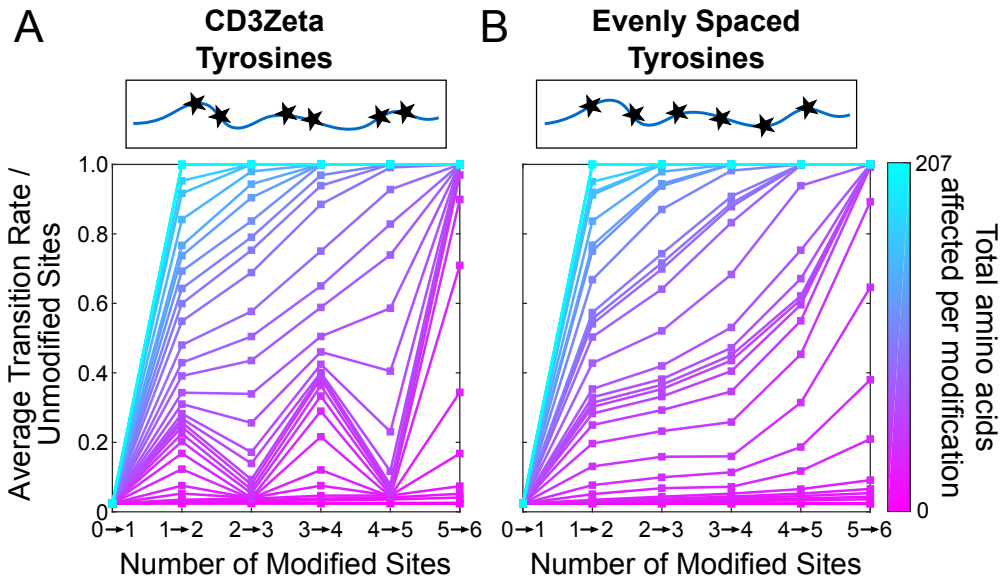


Figure 2.6: **Large local stiffening of multi-site disordered domain demonstrates effect of tyrosine location.** Sequence-dependent average binding rates of a kinase binding to cytosolic CD3ζ (A) and cytosolic CD3ζ with evenly spaced tyrosines (B) at different phosphorylation states for varying range of local stiffening (pink: no residues are stiffened; blue: all residues are stiffened after first phosphorylation). For small and large magnitudes of local stiffening, the average binding rate increases with each phosphorylation. For intermediate magnitudes of local stiffening, cytosolic CD3ζ exhibits oscillations in the average binding rate with an overall increase (A) while cytosolic CD3ζ with evenly spaced tyrosines increases without oscillations (B). In both cases at maximal local stiffening per phosphorylation, the entire domain is stiff leading to uninhibited kinase binding. Schematic below axis shows example configuration for each phosphorylation state. Unphosphorylated residues represented by black stars, phosphorylated residues are red stars. Kinase and phosphatase radius is 2.1nm. Rates are normalized to the free-space binding rate.

phorylations. As local stiffening rate increases, the magnitude of cooperativity increases as well, up to the limit where all residues are phosphorylated upon first phosphorylation (Fig. 2.8).

For dephosphorylation, removing the dephosphorylation-sequence weighting from the average binding rate has no effect on the trends, and only a small influence on the binding rates themselves (Fig. 2.7B). Here, dephosphorylation removes local stiffening from a site. When dephosphorylation occurs, there is an increase in entropic flexibility, which will have the most effect on the binding rate to nearby phospho-tyrosines. However, the nearby phospho-tyrosines remain locally stiff and primarily accessible to the phosphatase. The small decrease in accessibility is not significant enough to overcome the differences in accessibility imposed by the membrane. Therefore, at for the degree of local de-stiffening we consider, dephosphorylation rates are not heavily influenced by weighting by most likely sequence.

Emergent cooperativity from binding rate enhancement makes a good switch but a weak threshold.

We investigate two characteristics of the dose-respone curves: 1) switch-like behavior and 2) threshold behavior. The switch-like behavior of a dose-response curve describes how quickly it increases from a less-phosphorylated state to a more-phosphorylated state and can be used as a measure of cooperativity. Mathematically, the switch-like behavior is the maximum logarithmic slope of each curve. The threshold behavior of the curve describes how high of a dose is required before a response occurs. We calculate this value as $log_{10}(81)/log_{10}(\mathrm{EC90/EC10})$ for each curve. We calculate the switch-like and threshold behavior of the dose-response curves under constant and steric dephosphorylation.

For constant dephosphorylation, the switch like behavior increases with the number of amino acids stiffened per phosphorylation event. When 1/12 chain is stiffened per event, the switch

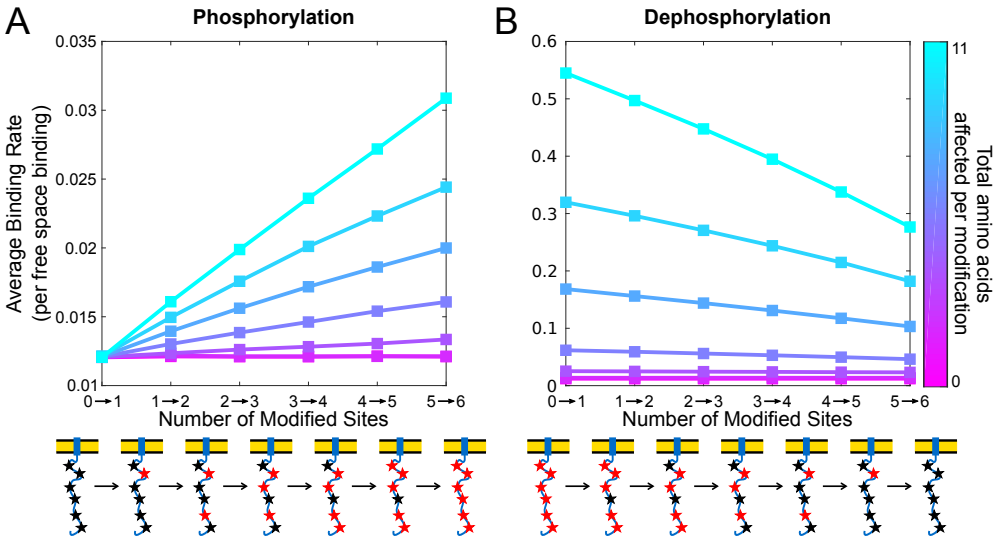


Figure 2.7: **Local stiffening modulates sequence-independent binding rates to multi-site disordered domain.** (A) Sequence-independent average binding rates of a kinase binding to CD3ζ at different phosphorylation states for varying range of local stiffening (pink: no residues are stiffened; blue: 11 residues are stiffened, 5 on each side of site). The kinase average binding rate increases with each phosphorylation and also with local stiffening range. (B) Sequence-independent average binding rates of a phosphatase binding to CD3ζ at different dephosphorylation states for varying ranges of local un-stiffening per dephosphorylation event. When local stiffening occurs, the phosphatase average binding rate decreases with each phosphorylation and also with local stiffening level. For both (A) and (B), schematic below axis shows example configuration for each phosphorylation state. Unphosphorylated residues represented by black stars, phosphorylated residues are red stars. Kinase and phosphatase radius is 2.1nm. Rates are normalized to the free-space binding rate.

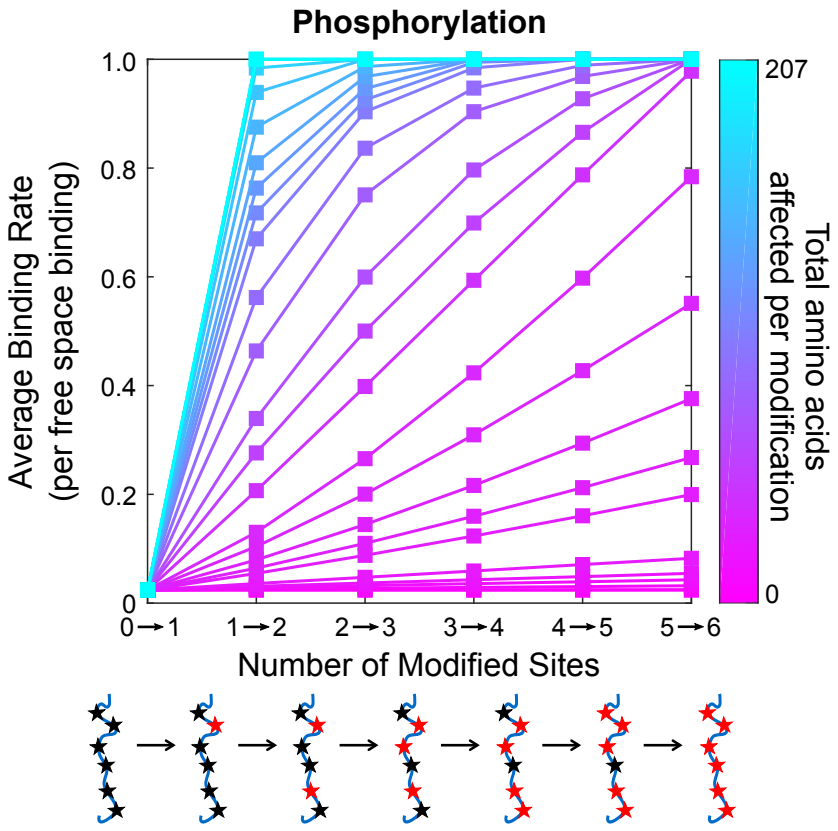


Figure 2.8: **Maximal local stiffening of multi-site disordered domain leads to uninhibited binding.** Sequence-independent average binding rates of a kinase binding to cytosolic CD3ζ at different phosphorylation states for varying range of local stiffening (pink: no residues are stiffened; blue: all residues are stiffened after first phosphorylation). The average binding rate increases with each phosphorylation and also with local stiffening range. At maximal local stiffening per phosphorylation, the entire domain is stiff leading to uninhibited kinase binding. Schematic below axis shows example configuration for each phosphorylation state. Unphosphorylated residues represented by black stars, phosphorylated residues are red stars. Kinase and phosphatase radius is 2.1nm. Rates are normalized to the free-space binding rate.

hill coefficient is 1.8, suggesting moderate switch-like behavior, while the threshold hill coefficient is less than 1.2, suggesting a poor threshold (Fig. 2.9 C). For steric dephosphorylation, the same increasing trend persists, but at 1/12 chain local stiffening, the switch hill coefficient is only 1.3 and the threshold hill coefficient is about 0.8 (Fig. 2.9 D). Local stiffening therefore creates cooperativity, even under reversible phosphorylation, but makes a poor threshold.

2.3.3 II. Membrane association

Recent evidence [26, 38] suggests that, before receptor triggering, the chains of the TCR are unphosphorylated and membrane-bound (protecting them from phosphorylation by kinases), only dissociating from the membrane after triggering, at which time they can become phosphorylated. But TCR phosphorylation is thought to be one of the first steps in triggering (and indeed, for antigen discrimination, the TCR must be involved in the first step) [39]. How are the first chains phosphorylated?

A possible hypothesis to resolve this chicken-and-egg puzzle is that, prior to triggering, the chain is biased towards the membrane but spends a small-yet-significant time in the cytoplasm, accessible to kinases. This is shown schematically in Fig. 2.1C. Then, upon initial phosphorylation, the bias is shifted towards the cytoplasm, allowing further phosphorylation. This hypothesis suggests a delicate balance between membrane affinity and phosphorylation. In this section, we ask, if it is true that phosphorylation controls membrane association, what are the quantitative constraints on the interaction strengths between the chain, the membrane and the kinase?

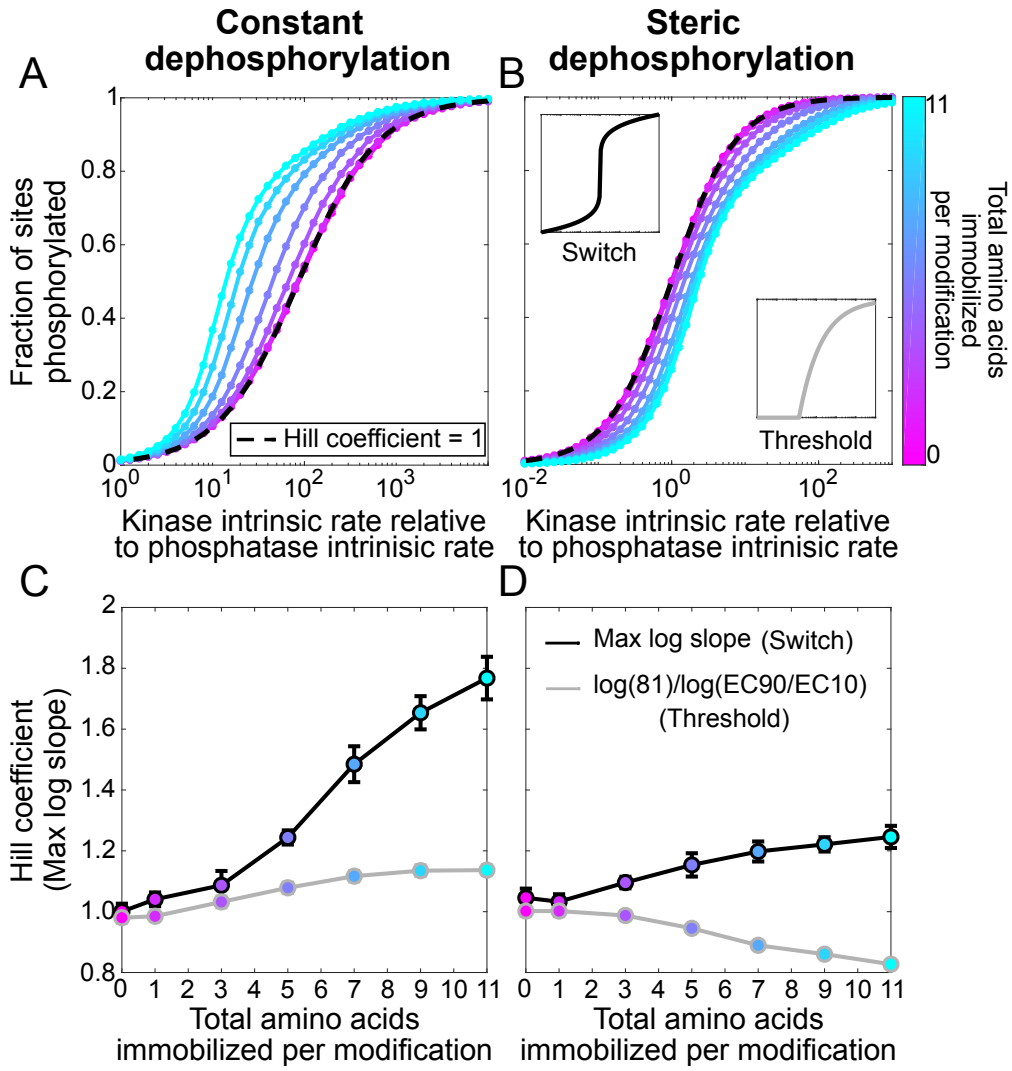


Figure 2.9: **Emergent cooperativity from binding rate enhancement makes a good switch, even in symmetric phosphorylation/dephosphorylation cycles.** (A,B) Fraction of sites phosphorylated as a function of the kinase-to-phosphatase activity ratio (measured by the ratio of free-space rates) for different ranges of local stiffening (colorbar), assuming a phosphatase with (A) negligible size or (B) radius of 2.1 nm, equal to the kinase. Black dashed line indicates a linear dose-response, i.e., with Hill coefficient 1. (C,D) Hill coefficients for different ranges of local stiffening per phosphorylation event, assuming a phosphatase of (C) negligible size or (D) radius of 2.1 nm. Hill coefficients calculated from maximum log-log slope (black line) or log(81)/log(EC90/EC10) (gray line) of the dose response curves. Error bars for max log slope indicate root-mean-square error from a cubic polynomial fit to slope. Error bars for log(81)/log(EC90/EC10) indicate standard deviation of hill coefficients from bootstrap sampling from dose-response curve.

Simplified electrostatic membrane interaction model explains both basic residue effects and phosphorylation effects

Experimental data suggests the CD3ζ and CD3ϵ chains are membrane associated prior to TCR triggering and that this association has two prominent features: 1) the basic residues in CD3ζ,ϵ are required for membrane association, 2) fully phosphorylated CD3ζ,ϵ does not associate with the membrane (Fig. 2.10 A). We create a simplified electrostatic interaction model, where phosphorylated tyrosines feel a repulsion from the cell membrane and basic residues feel an attraction (Fig. 2.10 B). Other amino acids, including unphosphorylated tyrosines, experience the membrane as a soft wall potential.

Each potential is parameterized to reproduce the two experimental phenomena. Specifically, we tune the strength of the basic residue potential, E_{B0}, based on the 'confinement' of the residues. Confinement of a tyrosine is here defined as when its positional variance is below $((1/e)(\text{range(variance)})+\text{min(variance)})$. Then the minimum E_{B0} such that the basic residues cause the polymer to be membrane associated is approximately 0.5 k_BT (Fig. 2.10 C, inset).

We next tune the strength of the phosphorylation potential, E_{P0}, such that tyrosine phosphorylation compensates for the basic residues (i.e. to match the null distribution where $E_{B0} = 0$) (Fig. 2.10 D). Given that there are 23 basic residues (n_B) and only 6 tyrosines (n_Y) in the CD3ζ chain, one would expect to need $E_{P0} \geq n_B/n_Y * E_{B0}$ in order for full phosphorylation to be enough to counteract the membrane association created by the basic residues. For CD3ζ, this is consistent with the values found from simulation, where $E_{B0} \sim 0.5k_BT$ and $E_{P0} \sim 2.8k_BT$ (Fig. 2.11). However, simulations of CD3ϵ indicate that $E_{P0} \sim 2k_BT$ is sufficient for its two tyrosines to compensate for the basic residues (Fig. 2.10 D). This contradicts the equation, which for CD3ϵ predicts a minimum E_{P0} of $3.5k_BT$.

We can explain this conflict by examining the distribution of basic residues compared to tyrosines. For CD3ζ, there are both tyrosines and basic residues along the full length of the domain. In CD3ϵ, the two tyrosines are clumped at the membrane-distal end of the domain, with fewer basic residues nearby. This difference in relative distributions of basic residues and tyrosines can account for the discrepancy between the prediction from the formula compared to the simulation. Tyrosines are primarily influenced by nearby basic residues and therefore only require enough strength to overcome local effects.

Phosphorylation-modulated membrane association allows some early phosphorylation, and accelerates late phosphorylation

For the minimum confinement of tyrosines to the membrane in CD3ϵ, we require the basic residues to feel at least a $0.5 k_B T$ attractive potential. In order for the two phosphorylated residues to be sufficient to eliminate the membrane association, each phosphate must experience a minimum of $2 k_B T$ repulsive interaction with the membrane. However, a stronger repulsive interaction could exist and replicate the same phenomena. To examine how phosphorylation influences binding kinetics, we simulate a range of strengths for the phosphate-membrane interaction assuming a basic residue-membrane interaction strength of $0.5 k_B T$.

When we compute the binding rates for kinases/phosphatases to the CD3ϵ chain for different levels of phosphorylation, we see that when the phosphate-membrane interaction is nonexistent or weak ($E_{P0} = 0, 1 k_B T$) (de)phosphorylation has minimal effect on the subsequent (de)phosphorylations. However, when E_{P0} is stronger ($E_{P0} \geq 2 k_B T$), the average kinase binding rate increases for future phosphorylation events (Fig. 2.12A.i, 2.14A.i). For CD3ϵ, phosphorylation can introduce an almost 2-fold increase in the binding rate for the second phosphorylation (Fig. 2.12A.i) For CD3ζ, phosphorylation can create a 2-4 fold increase in the binding rates from the first to the sixth phosphorylation event, depending on the strength of the phosphate-membrane interaction (Fig. 2.14A.i). These results indicate phosphoryla-

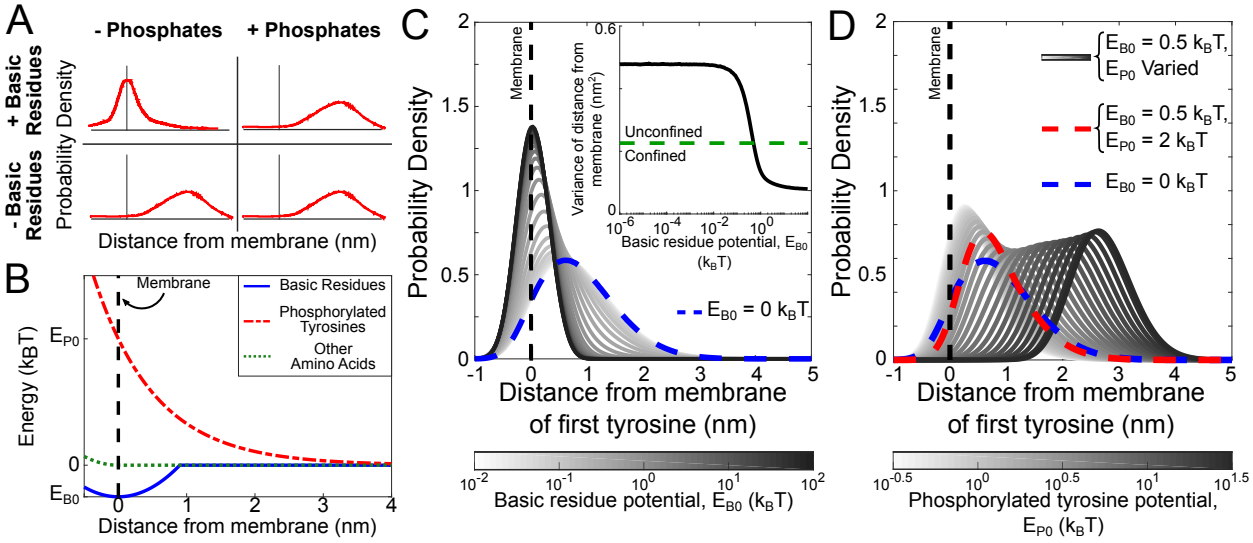


Figure 2.10: **Simplified membrane interaction model sets constraints on the strength of basic residue attraction and phosphorylation-driven repulsion from membrane.** (A) Schematic diagrams reflecting probability density of tyrosines under different conditions, based on previous experimental studies. Probability density of tyrosines is narrow and close to the membrane for wild type CD3ϵ, but widens and shifts away from the membrane when it is phosphorylated or the basic residues are mutated. (B) Interaction potentials used in simplified model. Basic residues experience an attractive potential of depth E_{B0} near the membrane and experience zero potential one Debye length ($\sim$1nm) away from the membrane (blue solid line). Phosphorylated tyrosines experience a repulsive potential with strength E_{P0} (red dashed line). Unphosphorylated tyrosines and all other amino acids are hindered from entering the membrane but otherwise experience no potential (green dotted line). (C) Probability density of the distance from the membrane of the 1st tyrosine of CD3ϵ, assuming no phosphorylated tyrosines, for varying strengths of the basic residue potential, E_{B0} (white: low; black: high). The tyrosine moves close to the membrane as E_{B0} is increased. Probability density when there is no basic residue potential (E_{B0} = 0 k_BT) is shown as blue dotted line. (Inset) Variance of probability density of 1st tyrosine over range of basic residue potential strengths. Green dashed line shows the characteristic $E_{B0} = 0.5 k_BT$ required to confine the tyrosine to the membrane, defined in the text. (D) Probability density of the location of 1st tyrosine assuming all tyrosines are phosphorylated, for E_{B0} = 0.5 k_BT (the value required to confine the tyrosine to the membrane assuming no tyrosines are phosphorylated), and varying phosphorylated tyrosine potential strength, E_{P0} (low - white, high - black). Probability density when E_{P0} = 2 k_BT is shown as red dashed line, reflecting the E_{P0} value needed to approximately return to the distribution when E_{B0} = 0 k_BT (blue dashed line).

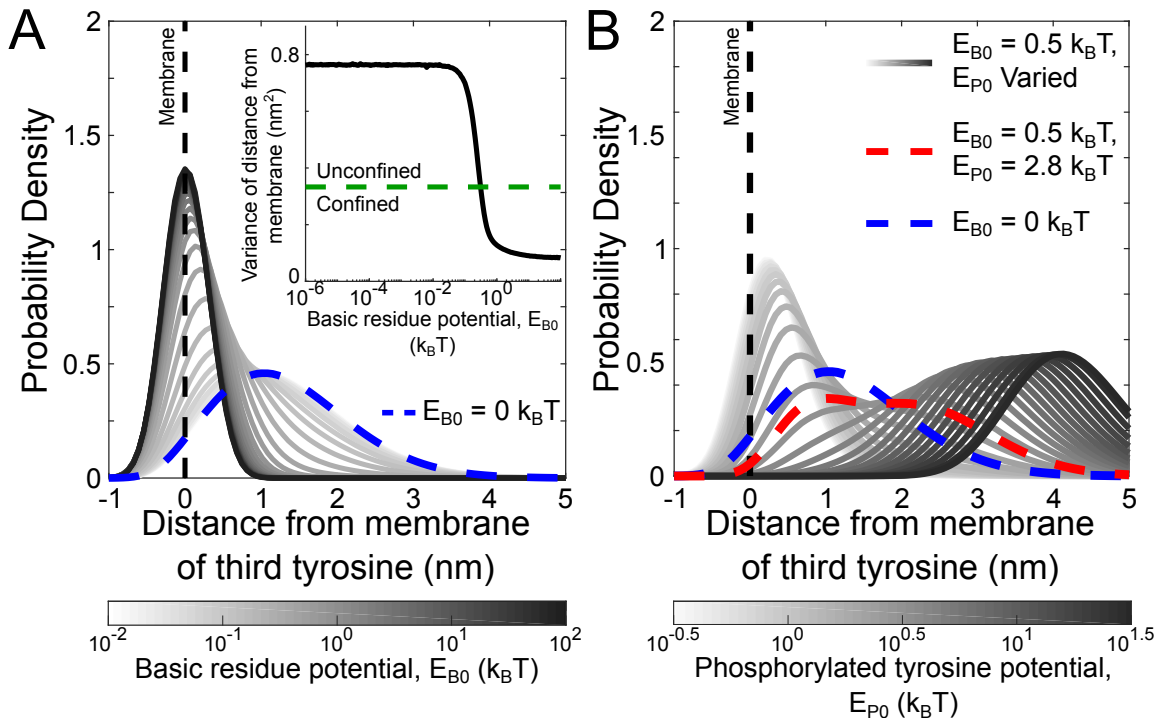


Figure 2.11: **Simplified membrane interaction model sets constraints on the strength of basic residue attraction and phosphorylation-driven repulsion from membrane for CD3ζ.** (A) Probability density of the distance from the membrane of the 3rd tyrosine of CD3ζ, assuming no phosphorylated tyrosines, for varying strengths of the basic residue potential, E_{B0} (white: low; black: high). The tyrosine moves close to the membrane as E_{B0} is increased. Probability density when there is no basic residue potential (E_{B0} = 0 k_BT) is shown as blue dotted line. (Inset) Variance of probability density of 3rd tyrosine over range of basic residue potential strengths. Green dashed line shows the characteristic $E_{B0} = 0.5 k_BT$ required to confine the tyrosine to the membrane, defined in the text. (B) Probability density of the location of 3rd tyrosine assuming all tyrosines are phosphorylated, for E_{B0} = 0.5 k_BT (the value required to confine the tyrosine to the membrane assuming no tyrosines are phosphorylated), and varying phosphorylated tyrosine potential strength, E_{P0} (low - white, high - black). Probability density when E_{P0} = 2.8 k_BT is shown as red dashed line, reflecting the E_{P0} value needed to approximately return to the distribution when E_{B0} = 0 k_BT (blue dashed line).

tion of membrane-associated disordered domains is able to modulate binding rates, creating cooperativity between phosphorylation events.

Interestingly, the average phosphatase binding rate also increases as more sites are dephosphorylated (Fig. 2.12A.ii, 2.14A.ii). For phosphorylation, each phosphorylation event helps dissociate more of the chain from the membrane, reducing the steric influence of the membrane on the kinase. However, for dephosphorylation, the initial steric influence is the entropic penalty from the disordered domain. As more residues become dephosphorylated, more sections of the chain become membrane associated, restricting their freedom to primarily two dimensions. The remaining phosphorylated tyrosines remain dissociated from the membrane by the repulsive phosphate-membrane interaction and therefore more accessible to the phosphatase. The cooperative effect of dephosphorylation only produces a small increase (CD3ϵ: 1.15-fold; CD3ζ: 1.5-fold) in binding rate since the membrane still creates a large steric barrier to phosphatase binding (Fig. 2.12A.ii, 2.14A.ii).

In a reversible system, weak phosphate-membrane interactions behave with approximately Michaelis-Menten kinetics (Hill coefficient = 1). Stronger phosphate-membrane interactions ($E_{P0} > 2$) cause the dose-response curve to diverge from Michaelis-Menten kinetics under both constant and sterically-influenced dephosphorylation (Fig. 2.12B, 2.14B). The effective Hill coefficients, defined as the maximum logarithmic slope, for this family of curves increases over the range of the phosphate-membrane interaction strength. Under both constant and steric dephosphorylation, the Hill coefficient increases to 1.2 and 1.3 for CD3ϵ and CD3ζ respectively at $E_{P0} = 10 k_B T$ (Fig. 2.12C, 2.14C). This suggests reversible phosphorylation of membrane associated domains is able to create weak ultrasensitivity, dependent on the strength of the interaction between the phosphorylated tyrosines and the membrane.

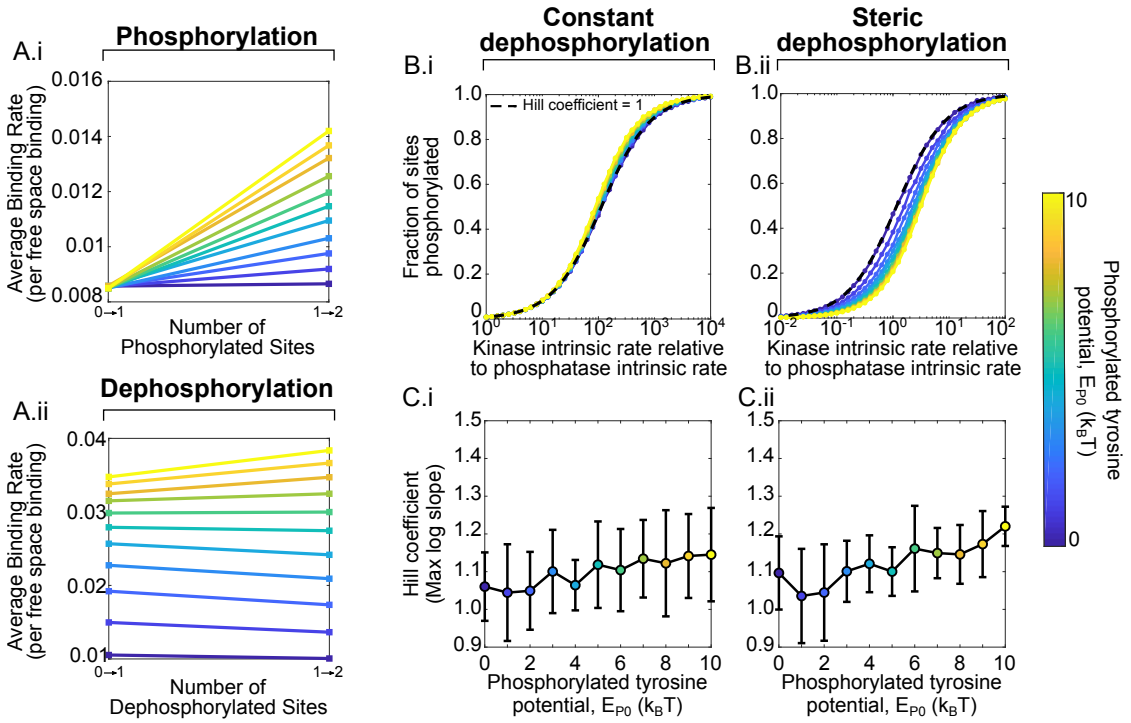


Figure 2.12: **Phosphorylation-driven modulation of membrane association leads to cooperativity.** (A) Average binding rates of (i) kinase and (ii) phosphatase binding to CD3ϵ at different (de)phosphorylation states for varying strengths of phosphorylated tyrosine potential (E_{P0}) (blue: weak; yellow: strong). Both kinase and phosphatase binding rates increase with each phosphorylation and also with increasing E_{P0}. (B) Fraction of sites phosphorylated over kinase intrinsic rate compared to phosphatase intrinsic rate for varying strengths of phosphorylated tyrosine potential (E_{P0}), assuming a phosphatase with (i) negligible size (constant dephosphorylation) or (ii) 2.1 nm radius, equivalent to the kinase (steric dephosphorylation). Black dashed line indicates linear dose response, i.e., Hill coefficient 1. (C) Hill coefficients calculated from maximum logarithmic slope of the Hill curves for varying strengths of phosphorylated tyrosine potential (E_{P0}), assuming a phosphatase of (i) negligible size or (ii) radius of 2.1 nm, equal to the kinase. Error bars indicate root-mean-square error from a cubic polynomial fit to slope. For (A)-(C), both kinase and phosphatase are assumed to have radius 2.1 nm. Rates are normalized to the free-space binding rate.

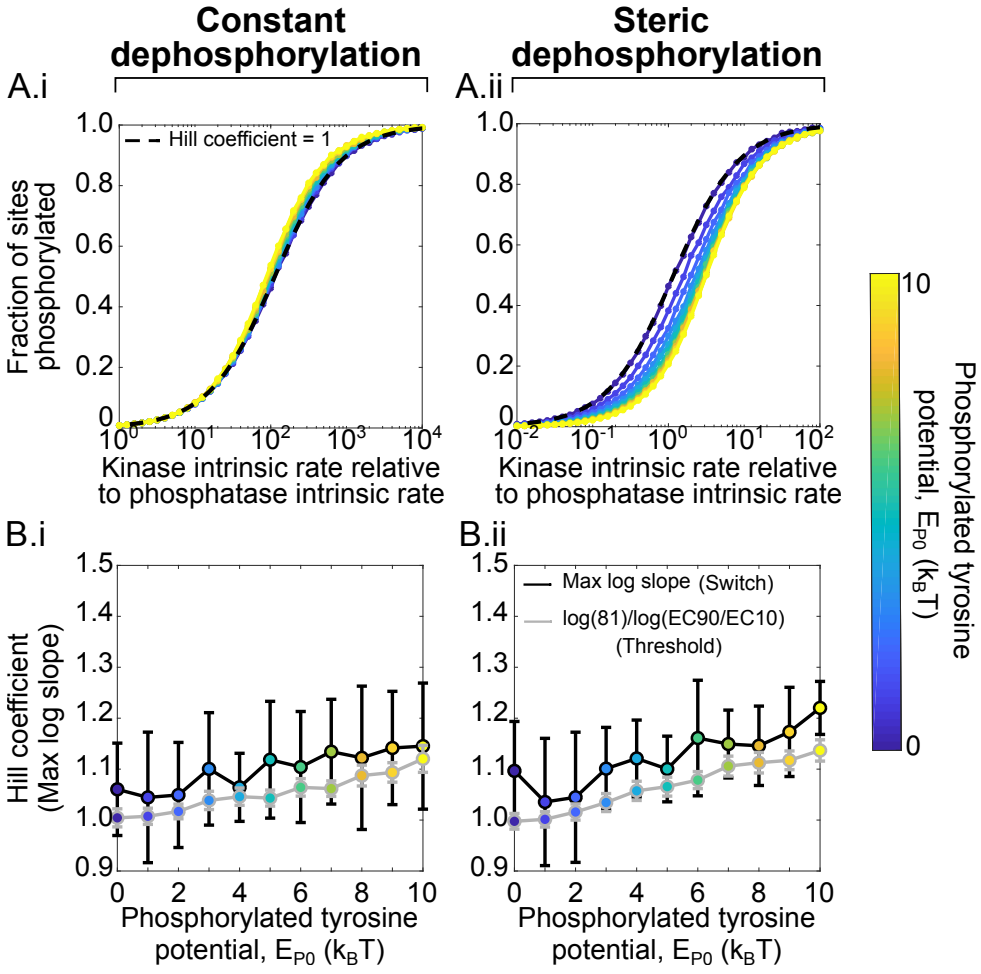


Figure 2.13: **Emergent cooperativity from binding rate enhancement to CD3ϵ makes a weak switch.** (A) Fraction of sites phosphorylated over kinase intrinsic rate compared to phosphatase intrinsic rate for varying strengths of phosphorylated tyrosine potential (E_{P0}), assuming a phosphatase with (i) negligible size (constant dephosphorylation) or (ii) 2.1 nm radius, equivalent to the kinase (steric dephosphorylation). Black dashed line indicates linear dose response, i.e., Hill coefficient 1. (B) Hill coefficients for varying strengths of phosphorylated tyrosine potential (E_{P0}), assuming a phosphatase of (i) negligible size or (ii) radius of 2.1 nm, equal to the kinase. Hill coefficients calculated from maximum log-log slope (black line) or log(81)/log(EC90/EC10) (gray line) of the dose response curves. Error bars for max log slope indicate root-mean-square error from a cubic polynomial fit to slope. Error bars for log(81)/log(EC90/EC10) indicate standard deviation of hill coefficients from bootstrap sampling from dose-response curve. In both cases, the Hill coefficient increases with the strength of phosphorylated tyrosine potential, even when dephosphorylation is assumed to be sterically hindered (B). Rates are normalized to the free-space binding rate.

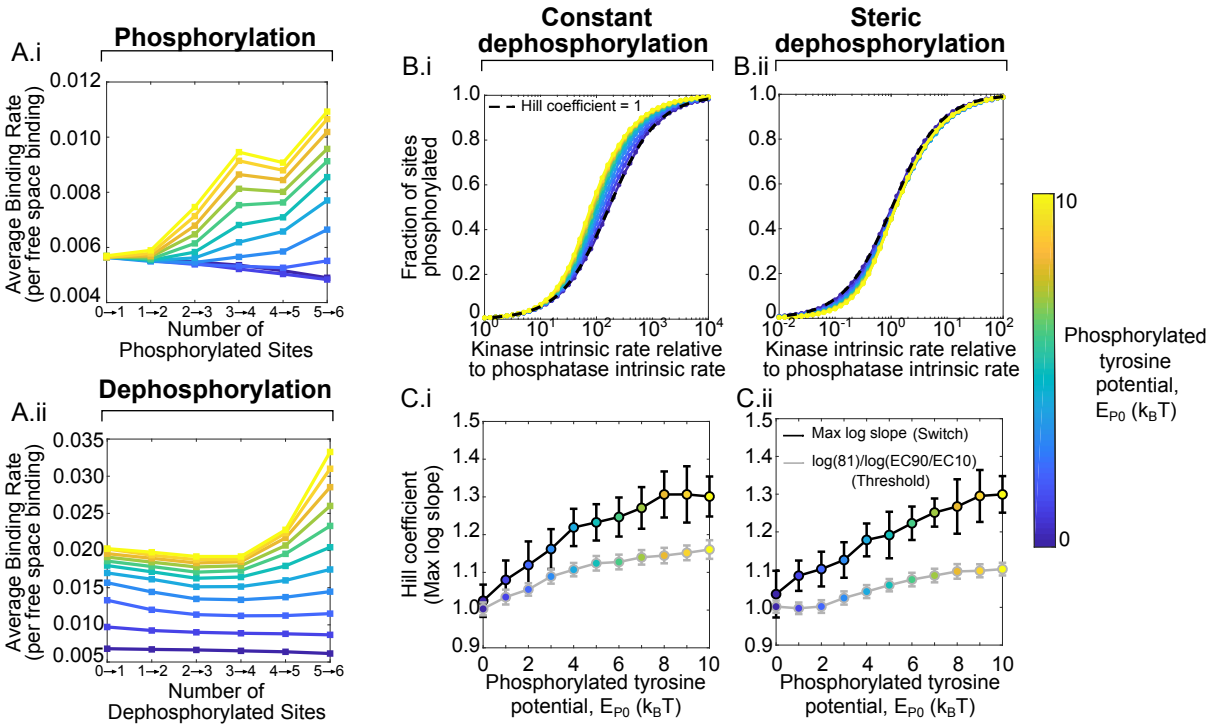


Figure 2.14: **Phosphorylation-driven modulation of membrane association of CD3ζ leads to cooperativity and sequential binding.** (A) Sequence-dependent average binding rates of (i) kinase and (ii) phosphatase binding to CD3ζ at different (de)phosphorylation states for varying strengths of phosphorylated tyrosine potential (E_{P0}) (blue: weak; yellow: strong). Both kinase and phosphatase binding rates increase with each phosphorylation and also with increasing E_{P0}. (B) Fraction of sites phosphorylated over kinase intrinsic rate compared to phosphatase intrinsic rate for varying strengths of phosphorylated tyrosine potential (E_{P0}), assuming a phosphatase with (i) negligible size (constant dephosphorylation) or (ii) 2.1 nm radius, equivalent to the kinase (steric dephosphorylation). Black dashed line indicates linear dose response, i.e., Hill coefficient 1. (C) Hill coefficients for varying strengths of phosphorylated tyrosine potential (E_{P0}), assuming a phosphatase of (i) negligible size or (ii) radius of 2.1 nm, equal to the kinase. Hill coefficients calculated from maximum log-log slope (black line) or log(81)/log(EC90/EC10) (gray line) of the dose response curves. Error bars for max log slope indicate root-mean-square error from a cubic polynomial fit to slope. Error bars for log(81)/log(EC90/EC10) indicate standard deviation of hill coefficients from bootstrap sampling from dose-response curve. In both cases, the Hill coefficient increases with the strength of phosphorylated tyrosine potential, even when dephosphorylation is assumed to be sterically hindered. For (A)-(C), both kinase and phosphatase are assumed to have radius 2.1 nm. Rates are normalized to the free-space binding rate.

2.3.4 III. Multiple simultaneous binding

One of the first molecular participants in T cell activation downstream of TCR is Zap-70, a molecule with two SH2 domains that bind phosphorylated tyrosines [5, 6, 40–42]. A recent *in vivo* experimental study found that approximately six Zap-70 molecules are bound per TCR [42], despite having ten possible simultaneous binding sites. This raises two questions: Do entropic constraints arising from polymer flexibility prevent more Zap-70 ligands from binding? And, if so, from a design viewpoint, are there advantages to the cell in not using the full array of possible Zap-70 binding sites? To address this question, we simulate binding of a ligand, Zap-70, to the full T Cell Receptor, when other ligands are already bound, shown schematically in Fig. 2.1D, and ask what impact the previously-bound ligands have on kinetics.

Simultaneous binding of full TCR is not sterically prohibited at physiological ligand radii.

We first examine the feasibility of fitting ten Zap-70 on a single T Cell Receptor. For large enough ligand radii, configuring ten molecules on a single receptor is not possible due to the restriction created by the binding sites and membrane. To determine whether this is a concern at physiological ligand radii, we simulate ten ligands tangentially attached to implicit TCR domains. The ten ligands move, energetically restricted by their interactions with each other and the contour limits of their attached polymers. They are also limited by an impassable membrane. Simulations end when a configuration is attained where the ten ligands do not conflict with each other or the membrane, and do not pass the contour limit of the polymers. We find that for physiological ligand radii (i.e. at least up to 6.9 nm radius), we can find a configuration where all ten ligands are able to bind to the T Cell Receptor without conflict (Fig. 2.15).

Multiple binding to multi-site disordered domain gives rise to anti-cooperativity

We calculate the average binding rates of ligands to all six membrane-bound chains of the T Cell Receptor, assuming a given number of previously bound ligands. For small ligands (e.g., 0.3 nm radius), the average binding rate remains approximately constant regardless of how many ligands are already bound (Fig. 2.16C). Since each ligand is very small, binding another ligand has little impact on the entropy of the chains, explaining why each addition does not influence the binding rate of the next ligand.

For larger ligands, the first binding event is significantly less likely than for small ligands, since a larger volume needs to be sterically accessible to accomodate the large ligand. As more ligands are simultaneously bound to the TCR, the average binding rate decreases. For a large ligand like Zap-70, with approximately 2.7 nm radius, the binding rate for the tenth ligand is approximately 20,000 times lower than the binding rate of the first ligand (Fig. 2.16C). Therefore, for large ligands, the extra steric barrier introduced by the previously-bound ligands significantly decreases the binding rate of future binding partners. This steric barrier influences the binding rate, but does not prohibit simultaneous binding to all ten sites (Fig. 2.15).

Sterically-induced negative cooperativity is one way of regulating the number of Zap-70 that associate with a given TCR. The first couple Zap-70 are able to bind to the domain relatively easily, but it quickly becomes prohibitive to bind more. The extra binding sites also allow flexibility of where Zap-70 bind and allow more to simultaneously bind to the domain before steric influences dominate.

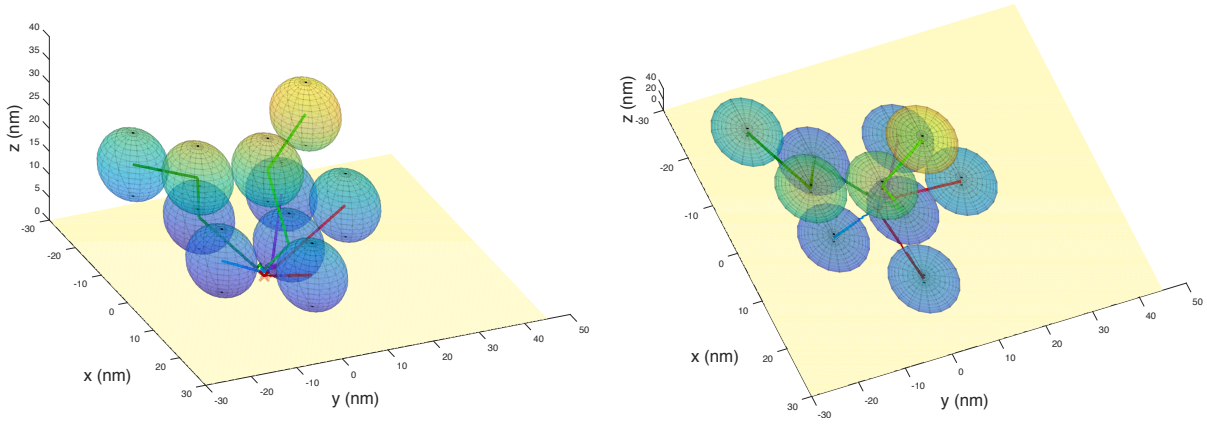


Figure 2.15: **Simultaneous binding of full TCR is not sterically prohibited at physiological ligand radii.** Two views of single TCR configuration with ten bound ligands where each ligand has radius of 6.9 nm. In this configuration, ligands do not interact with membrane (yellow) or other ligands. Colored lines represent TCR cytoplasmic domains.

2.4 Discussion

Disordered proteins play a critical role in many biological systems, including the first steps of T cell triggering. Previous work has demonstrated that despite their lack of structure, disordered domains can confer nonlinear behavior to signaling networks [6]. In this work, we developed a coarse grain model of a generalized disordered domain interacting with a spherical ligand. We adapted this model to specifically explore how the disordered domains of the T Cell Receptor behave in several contexts.

Relation to other work

Preferential Sequential Binding

Binding to multi-site proteins is typically classified into one of two groups: (1) sequential, or ordered, binding where binding occurs in a specific sequence, and (2) random binding, where no such sequence is required and binding can occur at any site. For CD3ζ, previous work

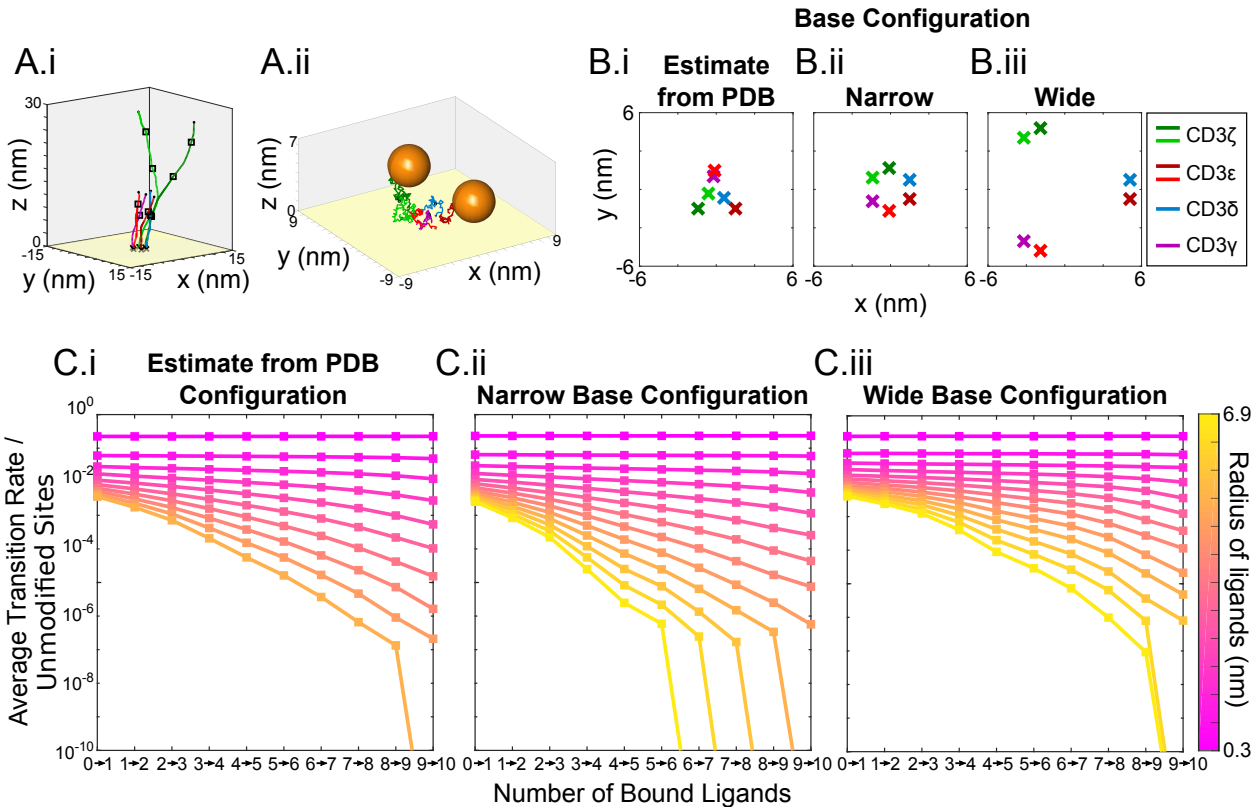


Figure 2.16: **Simultaneous binding reduces binding rates and implies sequential binding to multi-site disordered domain.** (A) Snapshots of simulated TCR subunits and binding sites (black squares) (Ai) in extended conformation and (Aii) at equilibrium with two bound ligands. (B) Three TCR subunit configurations are explored: (Bi) Estimated from PDB structure of TCR, (Bii) 1.5nm between dimer centers, and (Biii) 5nm between dimer centers. (C) Average binding rates to TCR against number of ligands bound for varying ligand size (color bar; pink: small; blue: large) and TCR subunit configuration (Ci) Estimated from PDB structure of TCR, (Cii) 1.5nm, and (Ciii) 5nm. The average binding rate increases with distance between TCR subunits. The average binding rate decreases as more ligands are bound. This effect increases with increasing ligand radius. Rates are normalized to the free-space binding rate.

offers a contradiction: some evidence suggests TCR ζ is sequentially phosphorylated, while others found no evidence for this [43–45]. Out work offers a possible resolution. We find that multi-site disordered domains can display 'preferential' (i.e. non-obligate) sequential binding through entropically derived binding rate differences. The presence of a membrane introduces additional steric limitations, enhancing the differences in binding rate and therefore exacerbating the observed preference.

Sequential binding has been extensively theoretically studied and has been shown to create nonlinear signaling behavior, such as ultrasensitivity. However, theoretical studies focus on 'obligate' sequential binding, where each event is required for the next to occur [45–48]. Our work reveals emergent 'preferential' sequential binding, where the sequence is likely to form but not required. How these differ is an interesting question for future study.

Disordered-to-Ordered Transitions

We find that for a multisite disordered domain, modifications that lead to changes in the entropy of the domain can lead to cooperativity between the binding sites. Specifically, local disordered-to-ordered transitions are sufficient to create positive cooperativity of PTMs, and conversely, ordered-to-disordered transitions can create negative cooperativity. Positive cooperativity can increase the sensitivity of a response to changes in ligand concentration and increase the potency of signaling responses. We find that for CD3ζ, enhancement of binding rates through disordered-to-ordered transitions can lead to positive cooperativity. This offers one explanation to explain previous experimental evidence showing cooperative effects between phosphorylations of CD3ζ [6]. There are several recorded instances of proteins undergoing disordered-to-ordered transition upon post-translational modification [20, 49, 50]. It is possible these systems could also exhibit rate enhancements as well.

Phosphorylation-mediated Membrane Association

Prior to T cell triggering, the CD3ζ, ϵ domains of the TCR associate with the cell membrane, with the tyrosines primarily embedded in the bilayer [25–27]. Post-triggering, they dissociate from the membrane, revealing the tyrosines for phosphorylation [25]. There is evidence that fully phosphorylated domains do not associate with the membrane [25]. However, given that tyrosine phosphorylation is one of the first events in TCR triggering, it is unclear how the domains either dissociate from the membrane to become phosphorylated, or, become phosphorylated to dissociate. We capture these behaviors in a simple model, including interactions between the membrane, basic residues, and phosphorylated tyrosines. Although tyrosine accessibility is low while the domains are membrane-associated, we find that a single phosphorylation can increase the binding rate of Lck to the rest of the tyrosines. Cooperativity of phosphorylations during membrane-association would offer one way for the CD3ζ,ϵ to both become phosphorylated and dissociate from the membrane. Prior to TCR triggering, phosphorylation of the first tyrosine is rare enough that phosphatases could dephosphorylate, preventing extraneous signaling. Upon TCR ligation, the phosphatase CD45 is thought to locally evacuate, which could increase the likelihood of phosphorylation [51].

TCR signaling occurs within seconds of the TCR binding an antigen [52–54]. It is unclear if the degree of ultrasensitivity created by phosphorylation of membrane-associated CD3ζ would be sufficient to achieve this fast response. Additional factors may be needed to induce efficient phosphorylation of the domain, such as TCR clustering or local changes in lipid composition. TCR clustering could also lead to more steric influences between the cytoplasmic membrane-associated domains, leading to an increased likelihood of dissociation [55]. Changes in lipid composition near the TCR have been shown to influence the membrane-association of the TCR domains [3, 28, 56]. Any of these could lead to increased tyrosine

accessibility, which when combined with the cooperative effect of phosphorylation, could lead to the rapid phosphorylation of the cytoplasmic domains.

Multiple Simultaneous Binding

Recent experimental evidence suggests only six ZAP70 bind to a single TCR on average. Why only six, when there are ten ITAMs on a TCR? The tight placement of the disordered domains of the TCR limits ZAP70 binding by decreasing individual binding rates. Binding sites near the membrane experience more occlusion from other domains in this configuration than expected from looser configurations. This provides a natural regulation to the number of ZAP70 able to bind to the TCR. This offers both a way to regulate the signal produced from a single T cell receptor while maintaining the ultrasensitive effects derived from having more binding sites than necessary for signaling [48].

We find that multiple simultaneous binding of ligands to the full T Cell Receptor leads to negative cooperativity. Negative cooperativity can endow signaling systems with two features: high turnover of ligands even in high ligand concentration (which might be advantageous if ligands are involved in other reactions), and constant signaling activity in low ligand concentration [31]. Similarly, this might reduce the impact of inhibitors, requiring much higher concentrations of inhibitor to completely turn off signaling [31].

Limitations of Model

Many previous studies have used differential equations or Markov chains to model binding to multisite molecules, including CD3ζ [45–48]. These models have offered insights into how multisite phosphorylation creates ultrasensitivity in a system. However, these methods make assumptions about the phosphorylation rates and do not include physical influences, such as

the chain configurations or interactions with the membrane. Our model begins from basic principles and calculates binding rates to each site, which can then be fed into other models.

Our model makes several assumptions, including flexibility of the domain and the spherical nature of the ligands. Given the distribution of amino acids along the domain, it is likely the persistence length is not perfectly uniform along the domain and may be stiffer in some areas. For example, proline-enriched regions tend to have a higher persistence length than other disordered segments [57]. However, by assuming the shortest persistence length reasonable for a protein (one segment represents one amino acid), we are able to consider the upper estimate for how entropy influences binding. Previous studies indicate the distribution of end-to-end distances of disordered domains match those of a worm-like chain with persistence length 3.04Å [32]. Additionally, this estimate is supported by recent studies quantifying the persistence length for CDϵ [27].

Other possible extensions to the freely-jointed chain model include using segments with non-zero width, and restricting angular freedom based on amino acid preferences.

For this work, ligand sizes are estimated as spheres. A more detailed representation could include several more parameters to describe the size (e.g. oval, rectangular prism), and/or represent each ligand with multiple domains. For instance, ZAP70 would have three separate but connected domains to represent the two SH2 and one catalytic domains. This would significantly complicate the simulation, since all tests for occlusion would need involve more complicated geometries, and we do not expect it to influence the qualitative results of our study.

This method also does not include atomistic details of the disordered domain or ligand. As computational power increases, alternate methods of studying disordered domains such as molecular dynamics, could give a detailed view of CD3ζ phosphorylations and how they affect interactions between the disordered domain and the membrane or itself.

Chapter 3

Formin

3.1 Background

Actin is a major part of the cytoskeletal network influencing cell structure, division, and movement [58, 59]. Understanding the process of actin polymerization, where actin monomers (Fig. 3.1 (orange circles)) are added to a growing actin polymer (Fig. 3.1 (orange chain)), is therefore critical to understanding these cellular characteristics.

Actin polymerization is often facilitated by members of a protein family called formins. Formins contain a structured Formin Homology 2 (FH2) domain and two disordered FH1 domains [60]. The FH2 domain is bound to the growing (barbed) end of actin, processing as monomers attach [61, 62]. The two identical FH1 domains behave as flexible 'arms' (Fig. 3.1 (blue chains)), which are able to explore the nearby space and bind profilin-actin molecules. FH1 domains contain polyproline tracks — proline-enriched regions of the amino acid sequence — which act as binding sites for profilin-actin molecules [61, 62]. To facilitate actin assembly, formins perform a 'capture-and-deliver', capturing profilin-actin molecules in solution and then delivering them to the growing end of the polymer [61, 63–65]. Upon

delivery, the actin monomer of the profilin-actin binds to the barbed end, dissociating from the profilin, and the profilin dissociates from the FH1 domain [63].

In humans alone, there are fifteen distinct formins [66]. Features such as length, number of binding sites, location of binding sites, and dimerization state distinguish these formins from each other. The behaviors and mechanisms of different formins are not well understood and are of growing interest in the field. What is it about these flexible arms that allow them to enhance actin polymerization? How do characteristics of specific formins influence this effect? Here we explore how formin features affect actin polymerization rates. Specifically, we investigate the effects of the following: (1) having two disordered domains compared to a single domain, (2) dimerization at the N-terminal ends of the domains, and (3) length and binding site locations as derived from true formin sequences.

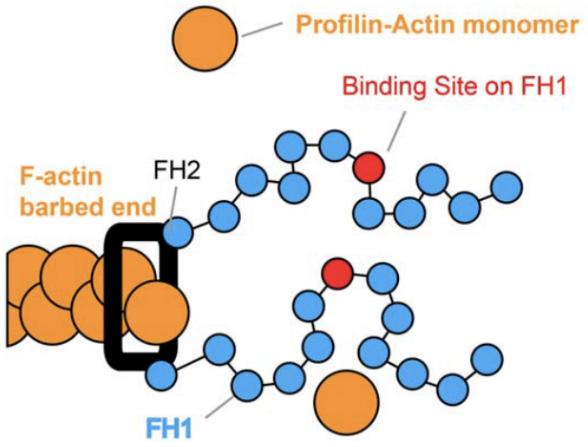


Figure 3.1: **Cartoon of formin structure.** Formin FH2 domain (black rectangle) attached to F-actin (orange chain). Profilin-actin monomers (orange circle) can bind to binding site (red circles) of formin FH1 domain (blue chain). Figure adapted from [11].

3.2 Methods

We model the FH1 domains of formins as freely jointed chains, interacting with spherical profilin-actin molecules. The basis of this model is further described in Sections 2.2.1 and 2.2.2. We simulate two types of formins: (1) theoretical formins with fixed length and evenly spaced polyproline tracks, and (2) biologically motivated formins with length and binding sites designed to represent specific formin family members (e.g. Fhod1, Capu).

3.2.1 Parameters for generic formin simulations

Formin amino acid sequences include stretches of proline-rich regions. Polyproline regions often have a higher propensity to form polyproline II helices, which increases the persistence length [67]. In our model, we choose to model formin with 60 segments of length 1.2 nm, twice the length of an amino acid. Five polyproline tracks are located at 1/10, 3/10, 5/10, 7/10, and 9/10ths of the length of the filament. Additional parameters and methods match Bryant et al. 2017 [11].

To investigate how formin dimerization, and in particular, N-terminal dimerization, influences the rate of actin polymerization, we simulated three cases: monomeric formin, formin dimerized at the C-terminus (C-term dimer), and formin dimerized at both the N and C terminals simultaneously (NC-term dimer). The dimerized formin is separated at the C-terminals by 5 nm in both dimerized forms. For the NC-terminal dimer, the N-terminals are separated by approximately 1nm. Outside of these restrictions, each domain is allowed to move freely.

3.2.2 Parameters for biological formins

We consider two specific formins: Fhod and Capu. For simplicity we only consider polyproline tracks with consecutive prolines.

Parameters for Fhod were determined from UniProt entry Q9Y613. The Fhod FH1 domain has 129 amino acids, with four polyproline tracks. We simulate formin with Kuhn length 1.2 nm. Therefore, we use a FJC with 41 Kuhn segments, with polyproline tracks centered at 8, 13, 16, and 21 segments away from the C-terminus of FH1 (actin barbed end).

Parameters for Capu were determined from UniProt entry Q24120. The Capu FH1 domain has 81 amino acids, with five polyproline tracks. We simulate formin with Kuhn length 1.2 nm, with polyproline tracks centered at 5, 13, 19, 28, and 36 segments away from the C-terminus of FH1.

3.3 Results

3.3.1 Multiple filaments reduce site specific capture rates, but increase overall polymerization rate

Overall, dimerization appears to reduce the rate of polymerization (Fig. 3.2B). More interestingly, dimerization and the specific type of dimerization influence the relative polymerization rates of each polyproline track. When the capture and delivery steps are compared, the changes in capture rate seem to contribute most of the interesting behavior. In general, dimerized formin experiences more steric hinderance or occlusion of the polyproline tracks since there are two domains able to occlude monomers. This creates a lower capture rate compared to monomeric formin (Fig. 3.2A). However, the NC-terminal dimer decreases

the capture rate even further since both ends of the formin domains are constrained to be close together and therefore incur more steric hindrance. Experiments on a Fhod variant *in vitro* also show N-terminal dimerization reduces actin polymerization rate compared to Fhod dimerized only at the C-terminal (personal communication, Quinlan Lab, UCLA).

3.3.2 N-terminal dimerization alters relative contributions of binding sites

In the monomeric form, the first (here defined as closest to the FH2 domain) and last (furthest from the FH2 domain) polyproline tracks have higher capture rates than the middle three tracks. Since the first and last site are equidistant from their respective ends, they experience similar occlusion and therefore have similar capture rates. In the C-terminal dimer, the last polyproline track shows a higher capture rate than the first track and the capture rate of the first track more similarly matches that of the middle sites (Fig 3.2A.ii). Since the C-terminal of each domain is fixed 5nm apart, the first polyproline track on each domain sees much more of the other domain compared to the last track, which is able to explore a larger region of space. Therefore, in the C-terminal dimer, the first site experiences more occlusion than the last track, giving it a smaller capture rate. In the NC-terminal dimer, both ends are forced into close proximity, but now the N-terminal ends are only 1 nm apart, creating even more steric occlusion than the 5nm C-terminal separation. This causes the first polyproline track to now have a higher capture rate than the last polyproline track (Fig. 3.2Aiii).

Since the capture rate is the rate limiting step in polymerization, these trends carry over to the polymerization rates for each dimer (Fig. 3.2B). With C-terminal dimerization, polyproline tracks closest to the N-terminus will contribute the most to polymerization whereas with NC-terminal dimerization, tracks closest to the C-terminus will contribute more.

There is a small increase in the delivery rate (not shown) of the NC-terminal dimer compared to the C-terminal dimer, which can be accounted for by the change in end-to-end distance distribution of the two dimers. The NC-terminal dimer on average has a smaller end-to-end distance than the C-terminal dimer, allowing it to deliver profilin-actin monomers more efficiently (Fig. 3.2C).

To more closely investigate how differences in polyproline track locations influence the capture rates, we also looked at the probability a polyproline track is available for binding as a function of where that track is located as a on the domain (Fig. 3.2D). For monomeric formin, sites near each end are significantly more available than sites in the middle, but this difference only matters for sites less than 2/10ths from either end. All sites further than 2/10ths from either end experience similar levels of occlusion. In the C-terminal dimer, sites close to the C-terminus are less often available compared to the monomeric form whereas sites at the N-terminal behave similar to if the formin were monomeric. In the NC-terminal form, the C-terminal sites behave as in the C-terminal dimer but N-terminal sites are now more occluded.

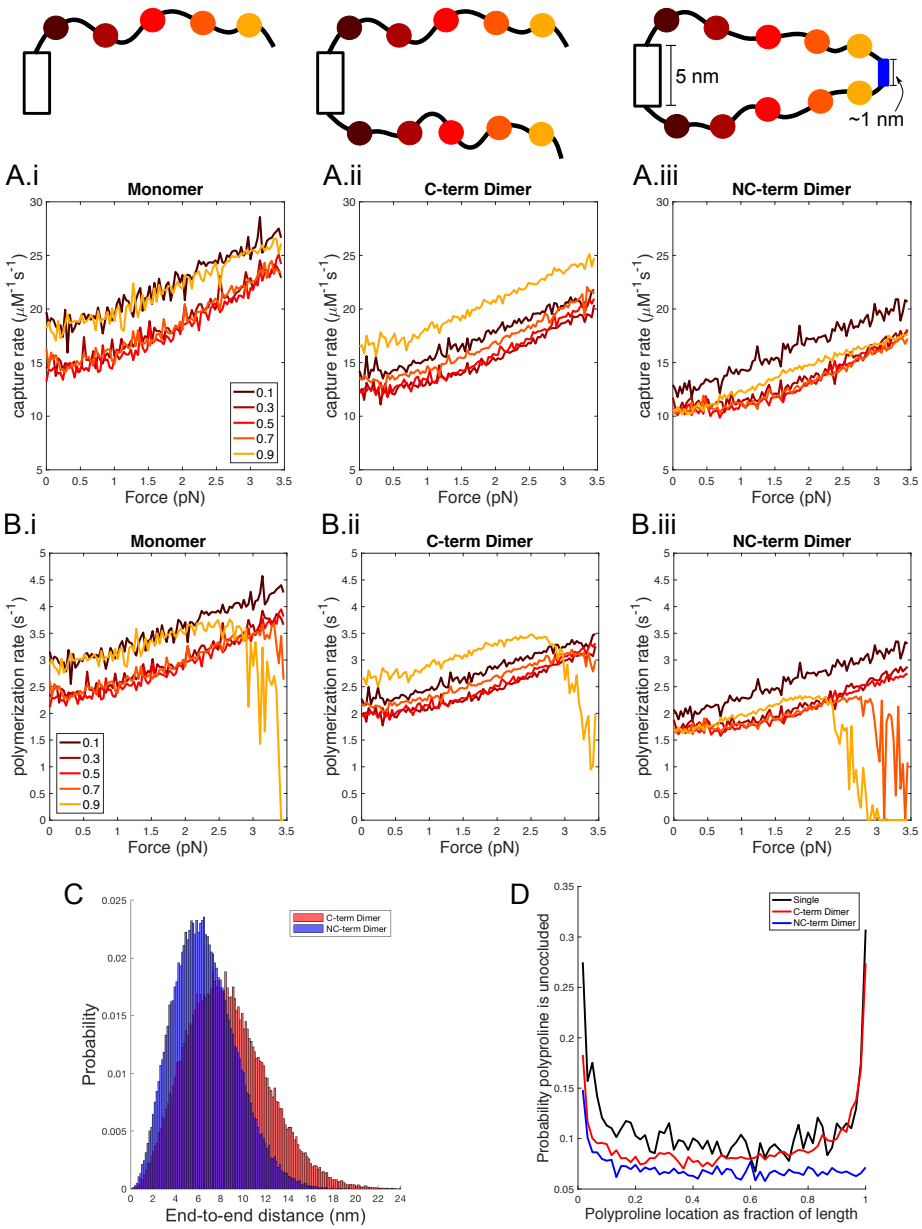


Figure 3.2: **Formin dimerization decreases actin polymerization rate and influences relative polyproline track contribution.** (above) Cartoons of each simulated formin domain. The FH1 domains are shown as black curves with the polyproline tracks shown as colored circles. The FH2 domain is represented as a black rectangle. (A) Capture rate ($\mu M^{-1}s^{-1}$) and (B) polymerization rate (s^{-1}) versus force (pN) for each polyproline track. Polyproline tracks are located at 0.1 (dark red), 0.3, 0.5, 0.7, and 0.9 (yellow) fractions of the domain length. Rates shown for formin in (i) monomeric, (ii) C-terminal dimeric, and (iii) NC-terminal dimeric forms. (C) End-to-end distributions of formin in C-terminal (red) and NC-terminal (blue) forms. (D) Probability polyproline track is unoccluded as a function of polyproline track location as a fraction of domain length. Single (monomeric) formin is in black, C-terminal dimer in red, and NC-terminal dimer in blue.

3.3.3 Development of analysis pipeline for formin homologues

Given that there are many formin homologues, it would be beneficial to have a pipeline to analyze each formin and determine the rate they polymerize actin. We develop a pipeline to do the following: (1) find proteins with FH1 domains from Uniprot, (2) determine the length, binding site locations, and number of prolines in each binding site for the FH1 domains, and (3) determine the polymerization rates for the three dimer forms from a lookup table of rates developed for the pipeline. We create a Python script to access the Uniprot database to find proteins with FH1 domains and determine their length and binding sites. The lookup tables are created by simulating a range of lengths and binding site locations for each dimer form and interpolating the intermediate values. Results from our pipeline indicate differences between actin polymerization rates due to variation in formin FH1 domains.

3.3.4 Differences in actin polymerization rates can emerge from biological variation in FH1 domains

Here, we use our pipeline to analyze two specific formins, Fhod and Capu, to explore how differences in their FH1 domains impact their polymerization rates. The FH1 domain of Fhod is ~ 1.5 times as long as Capu and has four polyproline tracks grouped more tightly near the C-terminus. In comparison, Capu has a shorter FH1 domain containing five polyproline tracks spaced roughly uniform along its length. Additionally, each of Capu's polyproline regions contains more consecutive prolines than any of Fhod's polyproline tracks. We compute the total polymerization rates for Fhod and Capu, as a function of their length, binding site location, and size of polyproline track (how many prolines in the region). We find that in all dimerization states, Capu demonstrates a higher polymerization rate than Fhod (Fig. 3.3, 3.4). This indicates that biological variation in FH1 domains can produce significant differences in actin polymerization rates.

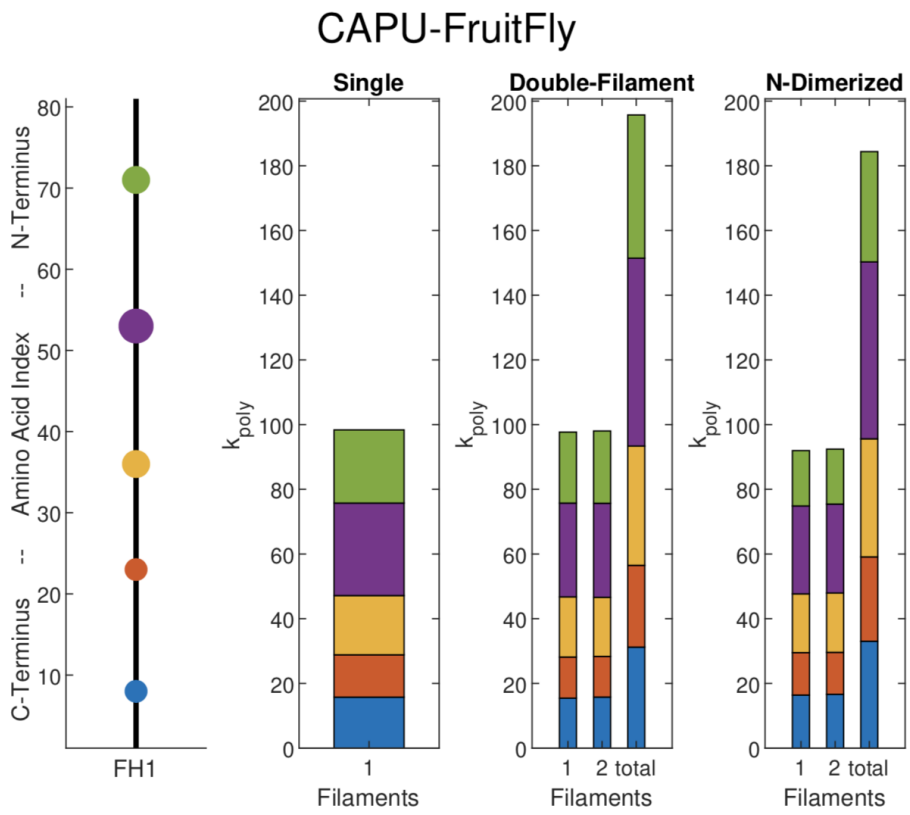


Figure 3.3: **Polymerization rates for the formin, Capu** Schematic of Capu including length and binding sites. Sizes of binding sites represent relative number of prolines in binding site (left). Polymerization rates for a single filament, C-terminal dimer, and NC-terminal dimerized Capu. Contributions from individual binding sites indicated with colored bars. For dimers, the polymerization rates of individual filaments are shown (1,2) and the rates for the whole dimer is shown by 'total'.

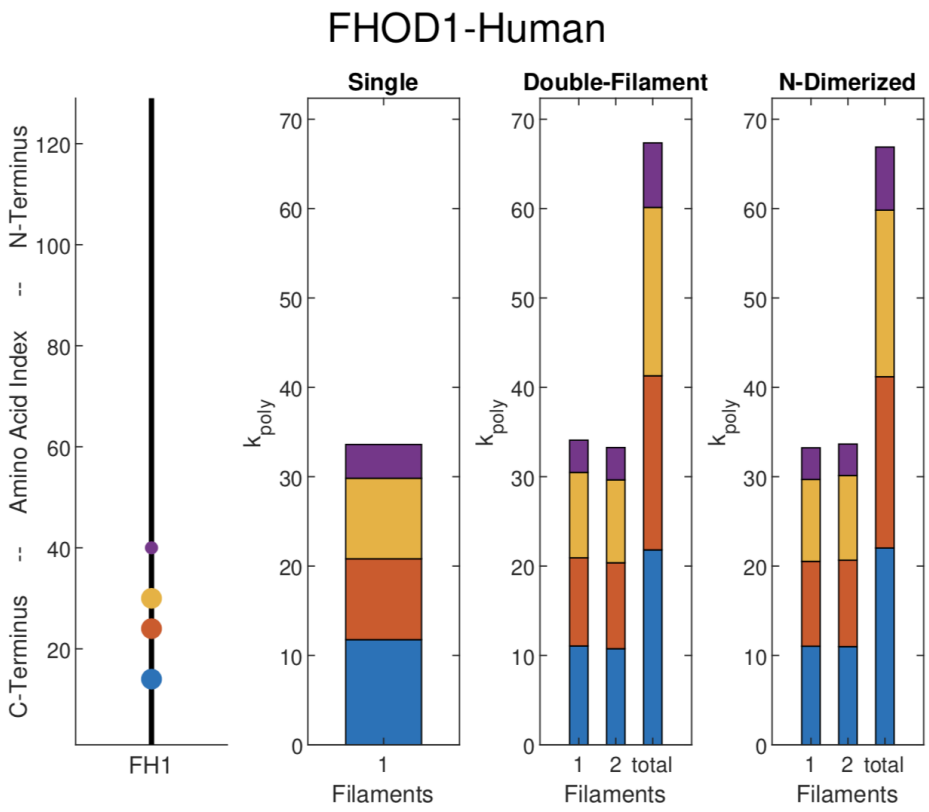


Figure 3.4: **Polymerization rates for the formin, Fhod1** Schematic of Fhod-1 including length and binding sites. Sizes of binding sites represent relative number of prolines in binding site (left). Polymerization rates for a single filament, C-terminal dimer, and NC-terminal dimerized Fhod-1. Contributions from individual binding sites indicated with colored bars. For dimers, the polymerization rates of individual filaments are shown (1,2) and the rates for the whole dimer is shown by 'total'.

3.4 Discussion

In this chapter, we simulated explored how differences in disordered domain characteristics and multiplicity can influence formin polymerization rates. We first simulate a generic formin, with evenly spaced polyproline tracks, to explore how single, C-terminal dimerized, and NC-terminal dimerized FH1 domains influences actin polymerization rates. We find that dimerization of FH1 at the C-terminus reduces the polymerization rates contributed by each polyproline track due to decreased probability of capturing profilin-actin monomers.

Despite this, the total polymerization rate of dimerized formin is increased as the benefit of double the possible binding sites outweighs the increased steric occlusion. Formins where FH1 is dimerized at both the N and C termini experience yet more steric occlusion, but with an overall polymerization rate still above that of a single FH1 domain. Dimerization states therefore offer one method for controlling actin polymerization rates

We also compared the polymerization rates resulting from two specific formins, Fhod and Capu. We find that Capu has a higher polymerization rate than Fhod. This is intuitive since Capu has binding sites distributed closer to the N-terminus of FH1. These experience less occlusion by the rest of the formin domain than the C-terminal binding sites. Additionally, the polyproline tracks of Capu are much longer than the polyproline tracks of Fhod. The longest track of consecutive prolines on Fhod FH1 domains is five prolines long, whereas Capu polyproline tracks range from 6-14 prolines in length. These features mean Capu is more easily able to capture profilin-actin monomers, and therefore has a higher polymerization rate than Fhod.

Variations in sequence such as this could point to formins adapted to different environments in the cell. For instance, Capu would increase actin polymerization effectively, providing critical control for systems requiring quick actin growth. However, a formin like Fhod, where most binding sites are clustered near the actin barbed end, could provide fast polymerization in situations where formins undergo high force. A force extending the formin would increase the capture rate, and decrease the delivery rate [11]. Binding sites primarily near the barbed end could mitigate the reduction in delivery rate by increasing the likelihood of encountering the barbed end.

Chapter 4

Molecular Reach of Immune Receptor Tethers

4.1 Background

T cells play a critical role in the adaptive immune system: recognizing and responding to foreign artifacts, called antigens, in the body. Antigens are presented on the outside of Antigen-Presenting Cells, where T Cells can form a contact with them and determine if the antigen presents a threat. The T Cell response needs to be both specific (only responding to true threats) and sensitive (responding even to small quantities of the threat) [68].

The simplified story of T Cells interacting with antigens revolves around the T Cell Receptor. The T Cell Receptor (TCR) is a complex protein expressed by T Cells with both extracellular and intracellular regions. When a T cell encounters an antigen, the TCR binds to the antigen and makes a threat assessment. T cell triggering occurs when a TCR encounters an antigen that warrants a response. Triggering initiates a signaling cascade, leading to T Cell activation, characterized by cytokine production, proliferation, and target killing. The first

events of T cell triggering include phosphorylation of the intracellular, disordered domains of the TCR by kinase Lck. This is followed by tandem binding to the same domains by the kinase ZAP70, and a phosphorylation cascade that eventually leads to T cell activation [22, 69].

The more complicated story of T cell activation actually involves many more players. Specifically, there are many other proteins called co-receptors which are involved with enhancing or inhibiting T cell activation [22, 70, 71]. For example, engagement of co-receptor CD28 is critical to achieving full T cell activation. CD28 knockdowns result in a reduction of both proliferation and cytokine production [72].

Of significant interest is the co-receptor Programmed Cell Death - 1 (PD-1). Unlike CD28, PD-1 is a co-inhibitory receptor, responsible for reducing T cell activation when engaged [70, 73]. Ligands of PD-1 (e.g. PD-L1, PD-L2) are often expressed by cancer cells, thwarting an immunological response [19, 74]. Understanding PD-1 interactions with T Cell signaling is therefore critical for improving potential immune therapies. Upon ligation, the tyrosine motif of PD-1 is phosphorylated by the kinase Lck. These tandem phospho-tyrosines act as a docking site for the phosphatases SHP-1 and SHP-2. The tethered phosphatases dephosphorylate nearby phosphorylated receptor domains, such as CD28 or CD3ζ. Dephosphorylation of these T Cell stimulating receptors leads to decreased T Cell activation [19, 70, 73, 75].

One notable feature of these co-receptors is their disordered cytosolic domain. Previous studies have shown that characteristics of these disordered domains can influence their function. For example, the length of the tether has been shown to influence signaling reactions by affecting both binding and catalytic rates [12]. Simulations of membrane-anchored tethers have also indicated reaction rates can increase or decrease with tether length, depending on the rate of diffusion in the membrane [7]. Additionally, a bioinformatic analysis indicates binding site motifs associated with activatory functions tend to be located closer to

the membrane-anchored end of the cytoplasmic tail than motifs associated with inhibitory function [12].

There are many reactions where a enzyme bound to a cytosolic domain will interact with another cytosolic domain. One such case is SHP-1/2 bound to PD-1 dephosphorylating CD28 [75]. For reactions like this, the length of the two tethers and the size of the enzyme can determine how far away a reaction can occur. We refer to the distance that describes the probability cloud of interaction as the 'molecular reach' of the reaction (Fig. 4.1).

Given that the length of tethers — and therefore the molecular reach — has been shown to influence reaction rates, some questions arise: Can we quantitatively measure the 'molecular reach' of a reaction? How much of a reaction's molecular reach is contributed by the enzyme versus the tether? Can we measure the 'reach' of different immune receptors?

To address these questions, we consider a simple two-species *in vitro* system involving a disordered, singly phosphorylated peptide and a phosphatase, SHP-1. The phosphatase can bind and unbind to the phosphorylated tyrosine on the peptide. When bound, the phosphatase can dephosphorylate the nearby peptides, effectively removing possible binding sites for other phosphatases. The phosphatase can also dephosphorylate from solution.

Here we use Surface Plasmon Resonance to obtain a time series describing this tethered dephosphorylation reaction. We fit a differential equation model to the time series data, extracting five biophysical parameters: binding and unbinding rates, two distinct catalytic rates, and the molecular reach. We analyze data for four constructed peptides of different lengths and use the molecular reach of their reactions to estimate the reach of the phosphatase alone. Using the molecular reach of the phosphatase, we find the reach of different immune receptor cytoplasmic domains.

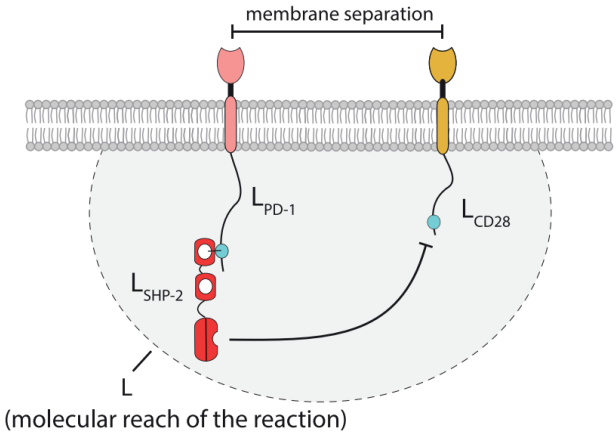


Figure 4.1: **Cartoon illustrating molecular reach.** Cartoon illustrating SHP-2 tethered to PD-1 and dephosphorylating CD28. The reach of PD-1, CD28, and SHP-2 combine to give the overall reach of the reaction, L. Figure adapted from [7], which is licensed under CC BY-SA 4.0.

4.2 Methods

4.2.1 Surface Plasmon Resonance

Tethered reaction *in vitro* experiments are conducted using Surface Plasmon Resonance (SPR). In SPR, light is shined at a surface. At a certain angle, plasmons on the surface will resonate with the light, resulting in an absorption of light. The reflected beam of light will therefore have a reduced magnitude at the resonance angle. If the surface changes, the resonance angle changes. SPR detects this change in angle and reports the time series of these changes converted to arbitrary 'Resonance Units'. This response is related to the amount of mass on the surface at a given time. SPR is a highly sensitive technique able to detect even fast, minute changes in the surface. The following experiments are conducted on a BiaCore T200, with detection events every 0.1 seconds.

For our experimental setup, peptides are irreversibly bound to a surface — in this case a 200 nm layer of dextran — in a channel, or flow cell. The response to the peptide alone determines the 'baseline' for the experiment. A phosphatase is then flowed through the channel and can bind, unbind, and/or dephosphorylate the peptides. SPR detects the changes in the surface and returns a time series of the response. We subtract the 'peptide only' baseline from the resulting curve to establish what changes are only due to interactions with the phosphatase. We also subtract drift from control flow cell to remove artifacts from the buffer.

In the following work, we consider reactions of the phosphatase SHP-1 interacting with many.different immune receptor tails. We use peptides constructed with Polyethylene Glycol (PEG) linkers with the ITIM motif of PD-1. Five such peptides are constructed, each with a different number of PEG linkers (0, 3, 6, 12, and 28) in order to establish different contour lengths. We also use several wild type immune receptor tails, including PD-1, CD28, SLAM, and CD3ζ. We use PD-1 phosphorylated at the second tyrosine, which has 55 amino acids between the membrane and the second tyrosine. We also consider a truncated version of the PD-1 with only 25 amino acids to the second tyrosine, which we call PD-1 25aa. All experiments are conducted at 37° C with a buffer flow rate of $10\mu L/min$.

4.2.2 Differential Equation Model of SPR Time Series

To fit the SPR tethered reaction data, we modify a multicenter particle density partial differential equation model developed by Goyette et al [12]. The original model includes parameters corresponding to the binding and unbinding rates (k_{on},k_{off}) of an enzyme to the receptor, the catalytic rate of the enzyme in solution ($k_{cat}Sol$), the catalytic rate of the enzyme when tethered to the receptor (k_{cat}Teth), and a molecular reach parameter (L). Our data exhibits nonspecific binding of the enzyme to the surface, sometimes reducing the re-

sponse below the baseline. This type of data cannot be well fit by the original version of the model. To account for this, we introduce a nonspecific binding term as a fit parameter. We demonstrate that the magnitude of response attributed to nonspecific binding is approximately linear with time. We also incorporate two parameters to fit the start and stop time of injections. Given that the response is noisy near injection start and stop times, this can more accurately find the start/stop time than by eye. Parameter fits are generated using MATLAB's least-squares curve fitting (lsqcurvefit) function. Model is implemented in MATLAB 2017b.

4.2.3 Estimating molecular reach of enzyme

Combining molecular reaches

The molecular reach of the reaction is influenced by the reach for the tether, L_{tether}, and the reach of the enzyme, $L_{\text{SHP-1}}$. Each of these reaches depend on the length and flexibility of the molecule. Using a worm-like chain model, we can write the probability density function as seen in Eq. 4.1. For a single tether or enzyme, the reach, L, is given by $L = \sqrt{2l_c l_p}$ where l_c is the contour length and l_p is the persistence length.

$$P(r, L) = \left(\frac{3}{2\pi L^2}\right)^{3/2} \exp\left(-\frac{3r^2}{2L^2}\right) \tag{4.1}$$

We can combine the tethers and enzyme to establish the probability density for the reaction, which is also described by Eq. 4.1 [10, 12]. In this case, L describes the molecular reach of the reaction and is a combination of the reach of the tethers and enyzme as seen in Eq. 4.3.

$$L = L_{\text{reaction}} = \sqrt{L_{\text{tether}}^2 + L_{\text{tether}}^2 + L_{\text{SHP-1}}^2} \tag{4.2}$$

$$= \sqrt{2 * L_{\text{tether}}^2 + L_{\text{SHP-1}}^2} \tag{4.3}$$

Estimate of SHP-1 reach

For a disordered domain, the reach can be estimated from the contour length (l_{c}) and the persistence length (l_{p}), or flexibility, of the domain. For the constructed PEG-PD-1 peptides, we estimate the contour length (from membrane to binding site of SHP-1) as being the number of PEG-linkers (N_{PEG}) times the length of a single PEG, ($l_{\mathrm{PEG}} \sim 0.4\,\mathrm{nm}$) [76]. From this, we can derive an approximation for the reach of SHP-1 where the reach is given by the intercept of the line L^2 versus N_{PEG} (Eq. 4.7).

$$L = \sqrt{2 \cdot L_{\text{tether}}^2 + L_{\text{SHP-1}}^2} \tag{4.4}$$

$$L^2 = 2 \cdot L_{\text{tether}}^2 + L_{\text{SHP-1}}^2 \tag{4.5}$$

$$= 2 \cdot (2 l_c l_p) + L_{\text{SHP-1}}^2 \tag{4.6}$$

$$= 4 \cdot N_{\text{PEG}} \cdot l_{\text{PEG}} \cdot l_p + L_{\text{SHP-1}}^2 \tag{4.7}$$

Calculation of L^2

For each reaction, L^2 (as seen in Fig. 4.9) is calculated by averaging the fitted parameter σ^* for all replicates and transforming the average to a single L^2 value for the peptide. Error propagation is used to convert the standard deviation of σ^* to an error for L^2. The parameter L relates to σ^* as seen in Eq. 4.8.

$$L = (\sigma^* \cdot \omega)^{-1/3} \qquad \text{where } \omega = (10^{-6} \cdot 6.023 \cdot 10^{23} \cdot 10^3 \cdot (10^{-9})^3) \tag{4.8}$$

Estimates from PyMol of SHP-1

From the PDB file for an open conformation SHP-1 (PDB 3PS5) and sequence data from UniProt (P29350), we can estimate a probable range of reach values for SHP-1. Direct measurement from PDB structure of N-SH2 binding site to catalytic site (straight line) gives

a short reach estimate of 5.3 nm. Combining the size of the individual structured domains with fully extended disordered regions gives the longest reach estimate of 16.5 nm. Instead of using fully extended disordered regions, we also consider a more probable end-to-end distance for the linkers. Assuming the linkers behave as worm-like chains with persistence length of 0.3 nm, we can calculate an intermediate expected reach of 7.4 nm. All measurements of structured domains were calculated using the measurement tool in PyMol. We use this range of reach estimates to validate the experimentally determined SHP-1 reach.

4.3 Results

4.3.1 Nonspecific binding can be represented as linear drift

One of the challenges of Surface Plasmon Resonance is in detecting signal only from specific binding/unbinding events, not interactions with other molecules (e.g. the dextran matrix) termed non-specific binding. Inspection of the control flow cells indicates that SHP-1 exhibits non-specific binding with a component of the flow cell and that the magnitude of non-specific binding is not consistent between control and experimental flow cells (Fig. 4.3). Therefore, when we subtract the control flow cell response from the experimental flow cell response to remove artifacts from the buffer, the difference between non-specific binding magnitudes results in a shift away from the baseline (Fig. 4.3). The difference in non-specific binding between control and experimental flow cells does not correlate with phosphatase concentrations, or peptide concentrations. The time series of nonspecific binding to the control flow cell indicates an approximately linear increase in magnitude over the course of the phosphatase injection (Fig. 4.3B). Therefore, we account for nonspecific binding in our model as a linear term with a fitted prefactor, nsb.

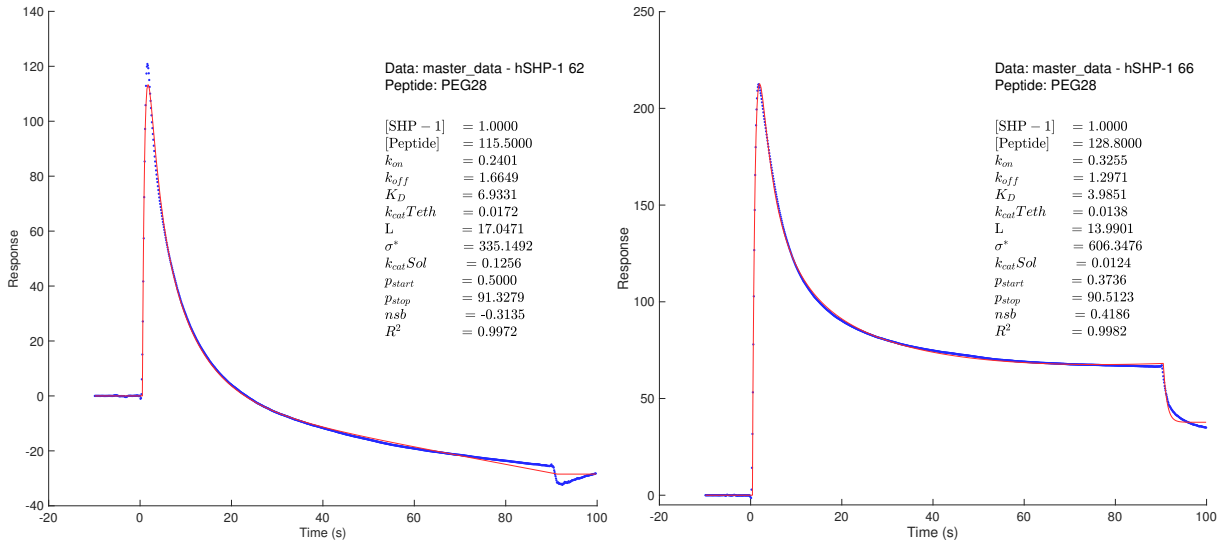


Figure 4.2: **Nonspecific binding relative magnitude depends on control flow cell.** Example SPR traces and fit parameters. (left) Experimental injection in flow cell 4, control in flow cell 3. Curve drops below baseline. (right) Experimental injection in flow cell 3, control in flow cell 4. Curve remains far above baseline.

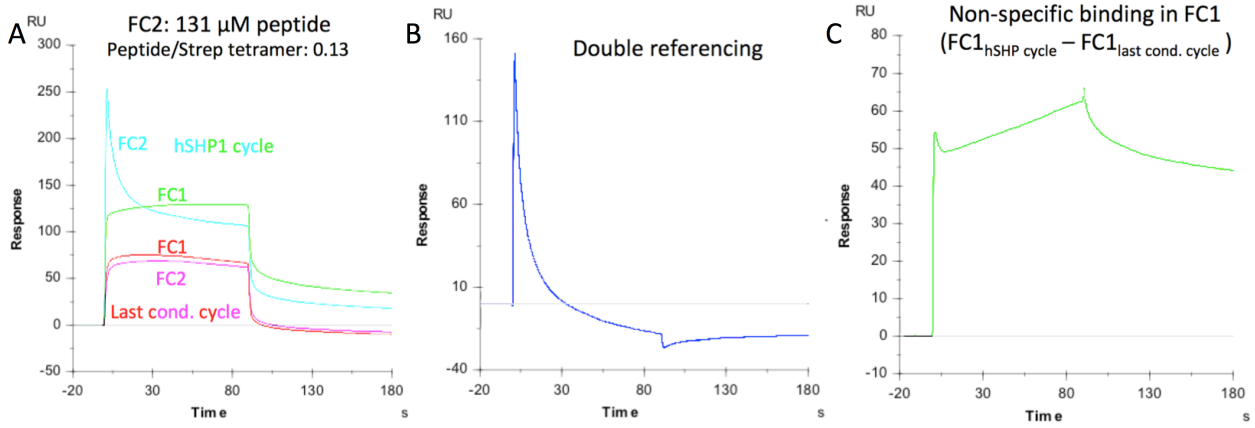


Figure 4.3: **Nonspecific binding can be represented as linear drift.** (A) Raw traces of experiment, normalized to peptide response as baseline. Last buffer cycles plotted for flow cell 1 (red) and 2 (pink). Experimental injection of SHP-1 in flow cell 2 (blue). Control injection of SHP-1 in flow cell 1 (green). Control flow cell response increases more with SHP-1 injection than buffer only. (B) Double referenced response: experimental response minus both control flow cell and buffer cycle. Response decreases below baseline. (C) Nonspecific binding response: Control flow cell with SHP-1 cycle minus control flow cell buffer cycle. Non specific response increases approximately linearly with time.

4.3.2 Parameter identification with Markov Chain Monte Carlo

We use a Markov Chain Monte Carlo to determine the posterior distributions for the parameters of the model. Preliminary results from a single dataset indicate there a no strong correlations between parameters, suggesting each parameter can be independently determined from the data (Fig. 4.4). Small positive correlations exist between k_{on} and k_{off}, indicating a dependence on K_D. However, the small range of k_{on} and k_{off} suggest they can still be determined from the data. Intuitively, there are correlations between $k_{\text{cat}}(\text{Tethered}) \cdot [\text{Peptide}]$ and $\sigma^* \cdot k_{\text{cat}}(\text{Tethered})$ since they both depend on the tethered catalytic rates. Interestingly, the start time, p_{start} shows a sharp cutoff at 0.4 seconds. An important next step is to develop confidence intervals for each parameter.

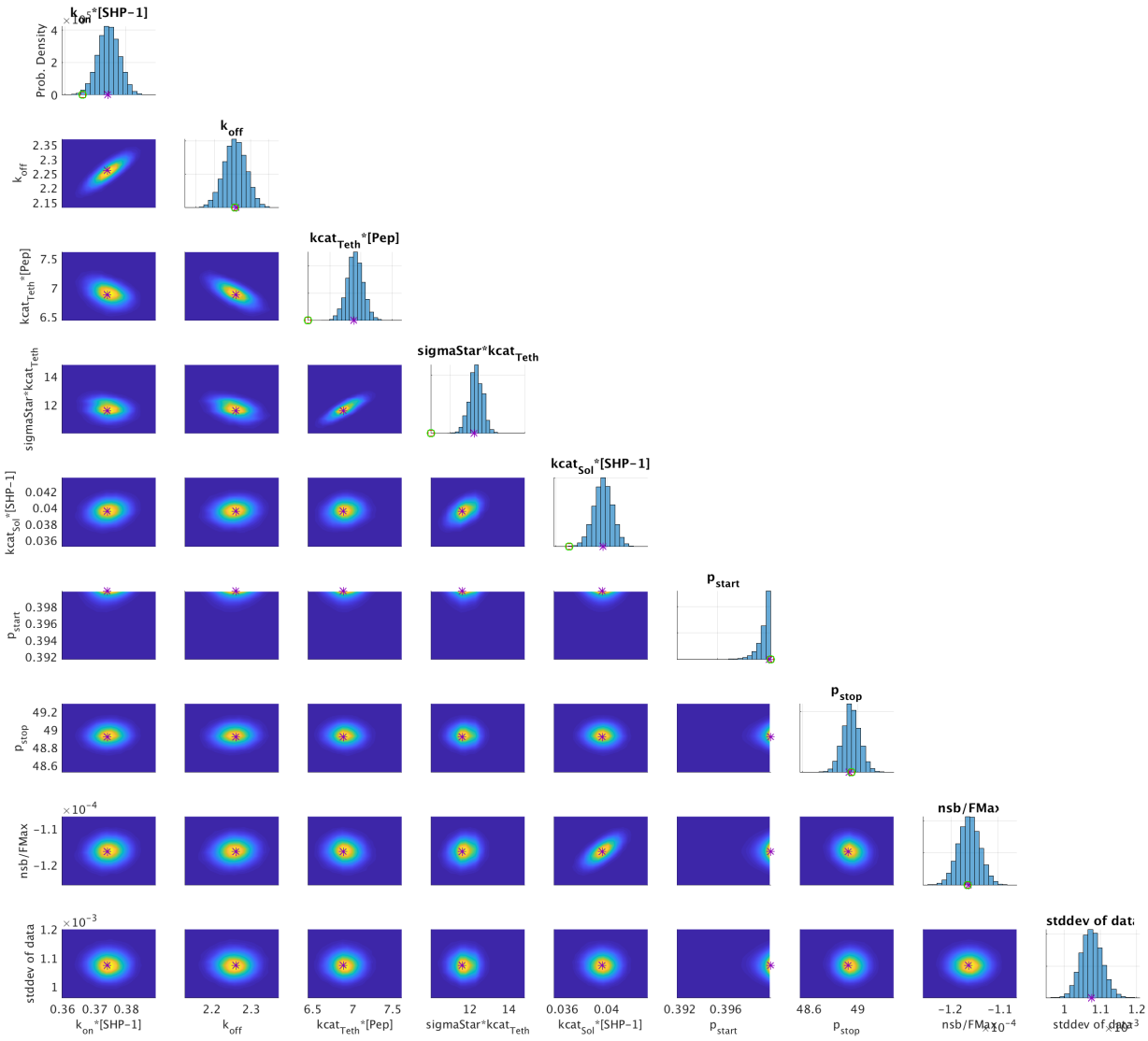


Figure 4.4: **Parameters correlations from Markov Chain Monte Carlo** Histograms (diagonal) and correlations (lower triangle) for model parameters for a single dataset. Green circles indicate starting point of MCMC. Purple asterisks indicate mode of parameters.

4.3.3 Assessment of correlations between biophysical parameters and enzyme, substrate concentrations

We first verify that the biophysical parameters can be identified independently of phosphatase or peptide concentration, which are chosen as part of the experimental set up. We

collect data for SHP-1 interacting with PEG28-PD1 for different phosphatase and peptide concentrations. When we consider each fitted biophysical parameter as a function of phosphatase concentration, we see that for most parameters, there is no correlation between the parameter and SHP-1 concentration (Fig. 4.5). The exception to this is the binding rate, k_{on}, which negatively correlates with SHP-1 concentration (Fig. 4.5). One explanation for this could be that increasing SHP-1 concentration creates crowding of the system. If there is an increased steric hindrance of the binding site due to crowding, this could appear as a decrease in binding rate. There are no strong correlations between the fitted parameters and peptide concentration (Fig. 4.6). This model is therefore capable of determining the biophysical parameters for a reaction without dependence on phosphatase and peptide concentrations.

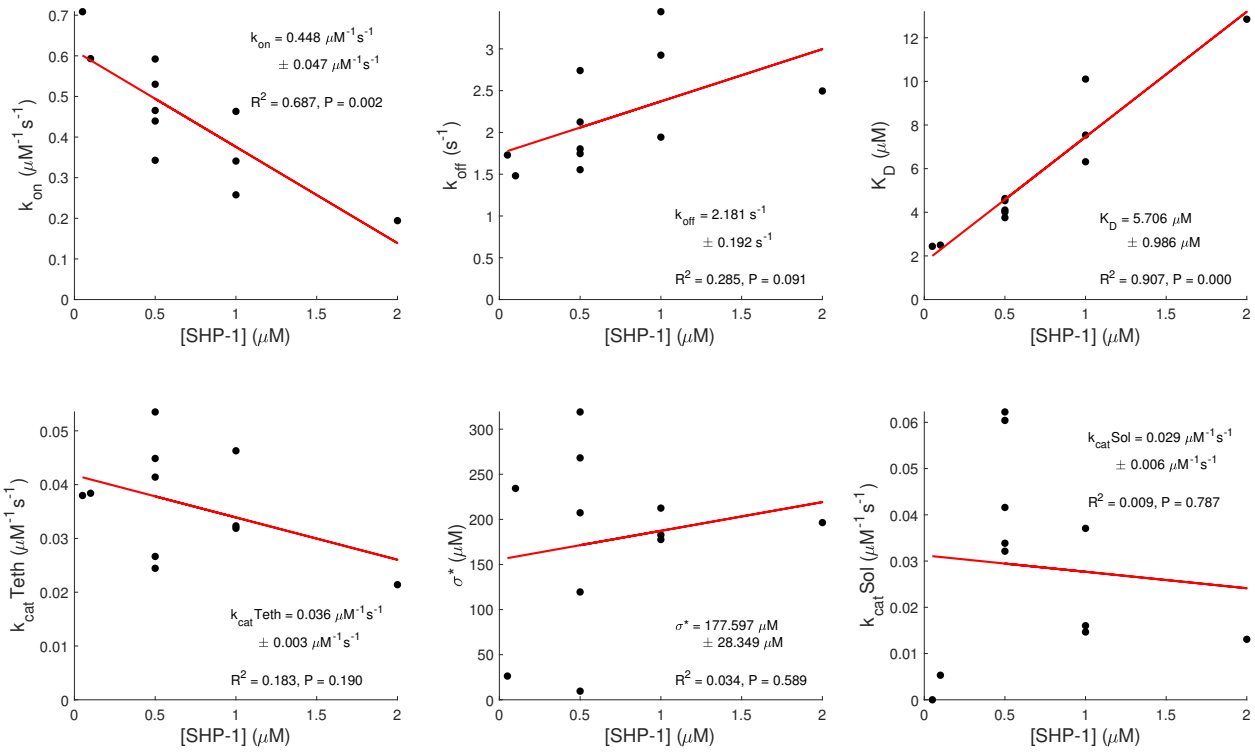


Figure 4.5: **PEG28 biophysical parameter correlations with SHP-1 concentration.** Biophysical parameters plotted against SHP-1 concentration of experiment (black dots). Mean and standard deviations of parameters displayed. Linear regression for each parameter indicated with red line. Coefficient of determination (R^2) and p-value for linear regression displayed. Correlation at 5% significance found for binding rate (k_{on}) and dissociation constant (K_D).

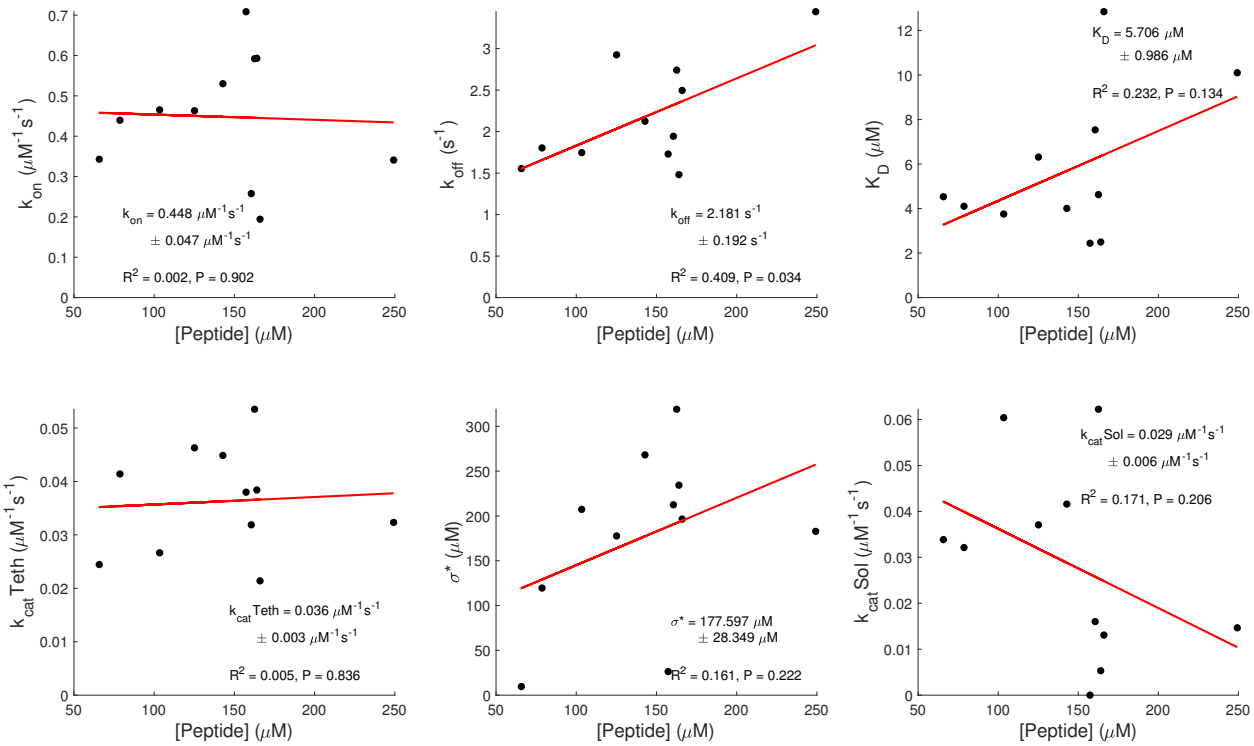


Figure 4.6: **PEG28 biophysical parameter correlations with peptide concentration.** Biophysical parameters plotted against peptide (PEG28-PD1) concentration of experiment (black dots). Mean and standard deviations of parameters displayed. Linear regression for each parameter indicated with red line. Coefficient of determination (R^2) and p-value for linear regression displayed. Correlation at 5% significance found for unbinding rate (k_{off}).

4.3.4 Tether length influences molecular reach of reaction

To examine how tether length impacts the molecular reach of a reaction, we collect data for peptides of five different lengths. These peptides are constructed to have the same binding motif — an ITIM taken from PD-1 — and varying numbers of PEG linkers. Fitting the data each dataset individually, we average the biophysical parameters within each peptide. The parameter, $\sigma*$, is the effective concentration. As the number of PEG linkers in the peptide increases, $\sigma*$ decreases, marking an increase in the molecular reach (Fig. 4.7). This indicates that the length of the tether does influence the molecular reach of the reaction. The longer peptide length expands the region a tethered enzyme is able to access, allowing it

to dephosphorylate peptides in a larger radius. Additionally, since the peptides themselves are longer, they explore a larger region where they may encounter a phosphatase. However, disordered tethers may take on many conformations, most of which will not be extended, but instead be compact (think a hairball). This means that even large changes in the contour length of the peptide may only result in small changes in the domain most explored by the peptide. This is consistent with the current evidence that a increase of $\sim$ 11 nm in the contour length from a peptide with no PEG (PEG0) to PEG28 only changes the molecular reach by $\sim$ 5 nm (Fig. 4.8).

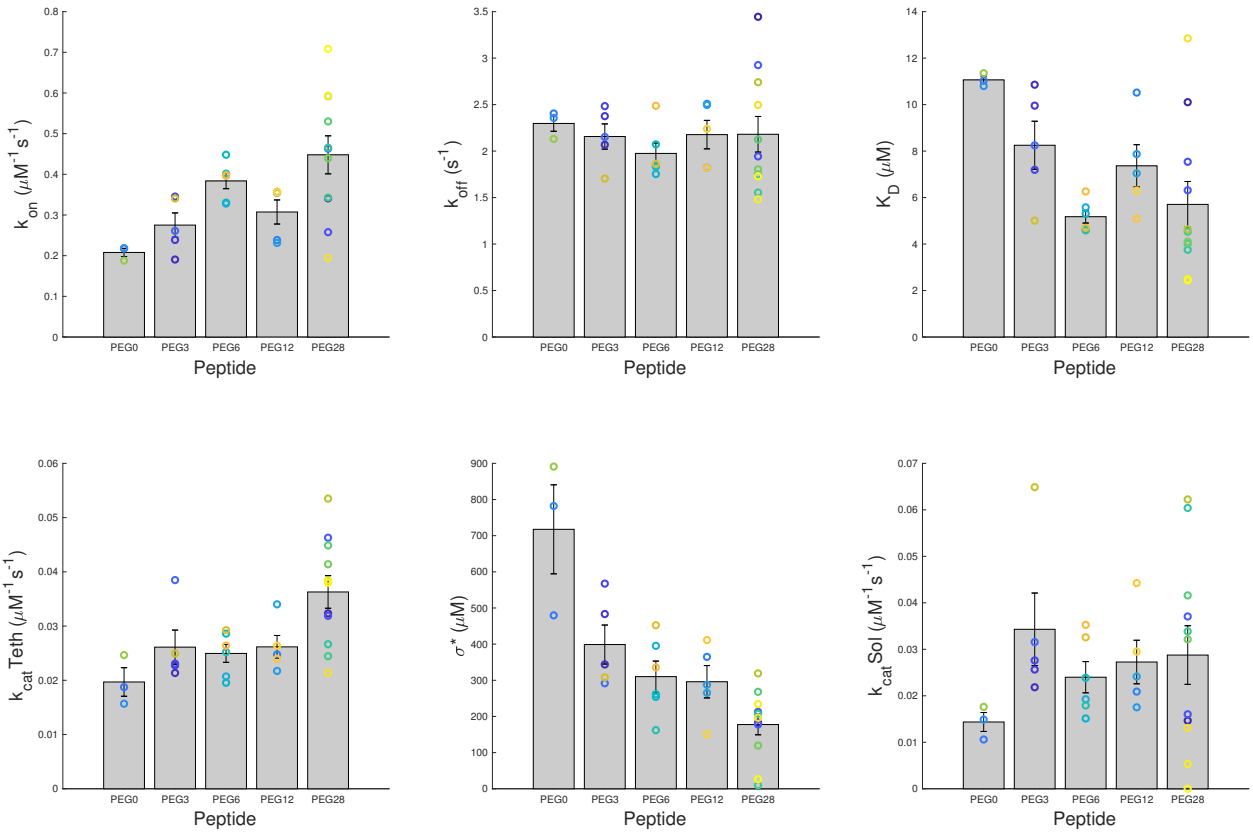


Figure 4.7: **Biophysical parameters for peptides of various lengths.** Biophysical parameter fits (colored dots), mean (gray bar), and standard error (black error bars) plotted for each PEG peptide: PEG0, PEG3, PEG6, PEG12, and PEG28. The binding rate (k_{on}) and tethered catalytic rate ($k_{cat}Teth$) each show increases with tether length (increasing number of PEGs). The effective concentration (σ^*) shows a decrease with increasing tether length.

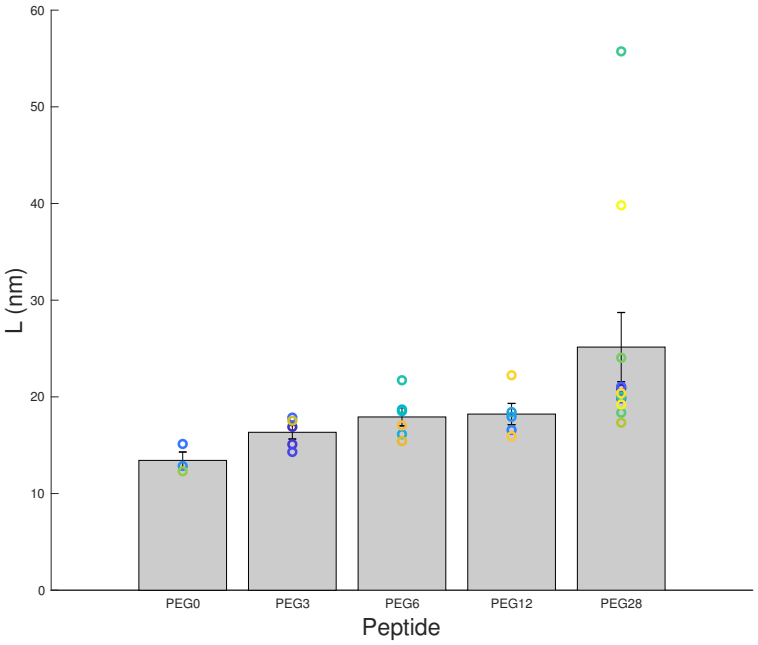


Figure 4.8: **Molecular reach increases with tether length.** Molecular reach (colored dots), mean (gray bar), and standard error (black error bars) plotted for each PEG peptide: PEG0, PEG3, PEG6, PEG12, and PEG28. Molecular reach increases with increasing tether length.

4.3.5 SHP-1 reach is primary contributor to molecular reach of reaction

The molecular reach for these simplified reactions involves two components: the reach of the tether and the reach of the enzyme. Assuming the reach of these molecules can be modeled as worm-like chains, the individual reaches combine as seen in Eq. 4.3. We can calculate the tether reach independently using the formula $L = \sqrt{2l_c l_p}$, where l_c is the contour length and l_p, is the persistence length. We therefore modify Eq. 4.3 to find the reach of the enzyme alone, as seen in Eq. 4.7. This equation shows we can estimate the reach of the enzyme as the intercept of a line plotting the squared reach of reaction versus the number of PEG linkers in the peptide. We find that when we consider the peptides with PEG linkers (PEG3, PEG6, PEG12, and PEG28), the squared reach of reaction is linear with the number of PEG

linkers, as expected from Eq. 4.7 (Fig. 4.9). From this, we estimate the molecular reach of SHP-1 to be 15.9 nm (Fig. 4.9). This value is within extrema estimates of SHP-1 reach we measured using the PDB 3PS5 and Uniprot P29350. The molecular reach of reactions using PEG peptides ranged from 13 nm to 20 nm (Fig. 4.8). This suggests most of the reach of a reaction is determined by the reach of the enzyme and not the reach of the tether.

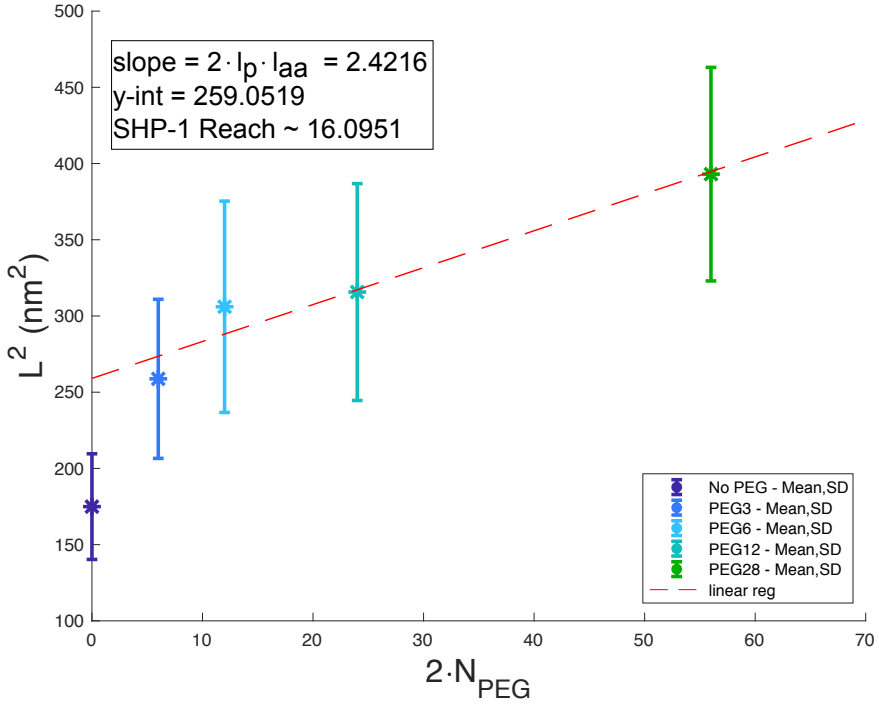


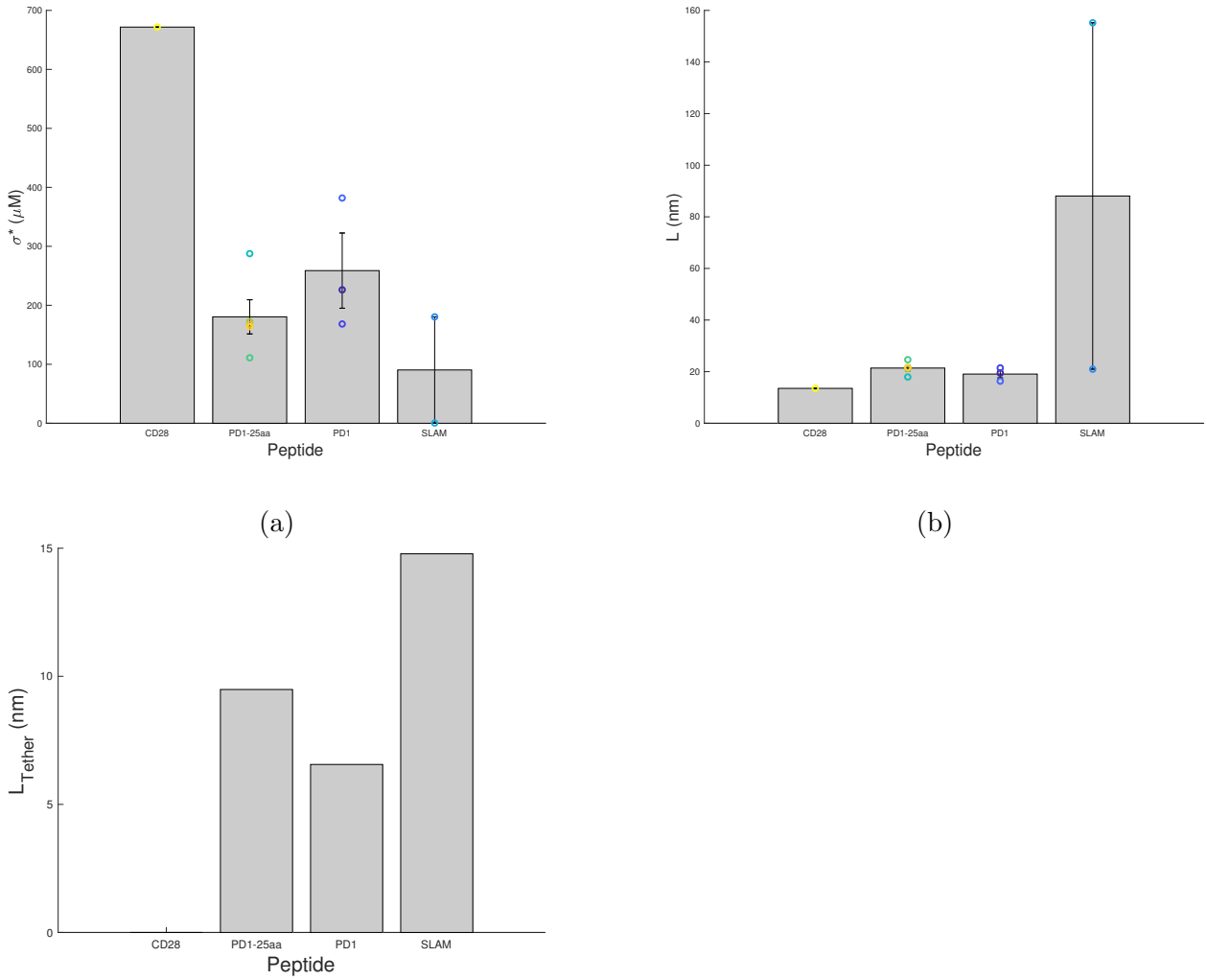
Figure 4.9: **Molecular reach of SHP-1 can be estimated from reach of reaction for different length tethers.** Computed mean square molecular reach (L^2) and standard deviation (colored points) plotted against two times the number of PEG linkers ($2 \cdot N_{PEG}$) for the peptide. Linear regression for PEG3,6,12, and 28 (red line) gives a y-intercept of 259.05 nm^2. The estimate for SHP-1 reach alone (i.e. zero tether length) is 16.1 nm.

Table 4.1: Immune receptors analyzed and number of amino acids before the phosphorylated peptide.

Peptide Name	**Number of amino acids to phosphorylated tyrosine**
CD28	11
PD1-25aa	25
PD1	55
SLAM	68

4.3.6 Different immune receptor tails create small changes in molecular reach

Given the estimate of SHP-1, we can calculate the molecular reach of specific immune receptors tethers. We collected data for a variety of immune receptors including PD-1, CD28, SLAM, and CD3ζ. Using the same fitting procedure as above, we fit the molecular reach of the reactions for each peptide (Fig. 4.10A). Using the molecular reach of reaction and the SHP-1 reach, we calculate the tether reach for each peptide (Fig: 4.10C). Consistent with their contour length from the membrane anchor to the binding site, CD28 shows the smallest molecular reach and SLAM shows the longest (Fig: 4.10C). It is notable that the reach for CD28 is approximately zero. The binding site of CD28 is only 11 amino acids away from the membrane (as measured to the phosphorylated tyrosine) (Table 4.1). The theoretical reach of CD28 by itself is therefore only $\sim 1\,\text{nm}$. This suggests almost the entire reach of SHP-1 dephosphorylating CD28 is made up from SHP-1, with almost no contribution from CD28. For CD3ζ, SHP-1 demonstrated dephosphorylation, but no detectable binding (data not shown).



(a) (b)

(c)

Figure 4.10: **Different immune receptor tails create small changes in molecular reach.** (A) Effective concentration, σ^*, for each dataset by peptide. (B) Reach parameter, L for each dataset by peptide. (A,B) Individual data given by colored points. Mean and standard error given in gray bars. (C) Mean reach of tethers estimated with reach of reaction and reach of SHP-1.

4.3.7 Tethering of SHP-1 increases effective concentration, reaction rates

In this system, dephosphorylation occurs at two rates depending on the mechanism. Dephosphorylation from solution occurs at $r_{\text{dephos}}(\text{Solution}) = k_{\text{cat}}(\text{Solution}) \cdot [\text{SHP-1}]$. When SHP-1 is tethered to a peptide, dephosphorylation occurs at $r_{\text{dephos}}(\text{Tethered}) = k_{\text{cat}}(\text{Tethered}) \cdot \sigma^*$. There is some evidence that the catalytic rate of SHP-1 depends on whether it is in solution or tethered to a peptide. In solution, the N-SH2 domain of SHP-1 is thought to inhibit the catalytic PTP domain [77, 78]. When the C-SH2 domain of SHP-1 binds a phosphotyrosine, the conformation change releases this auto-inhibition [77, 78]. This phenomenon could produces changes in the catalytic rate between SHP-1 in solution compared to tethered SHP-1. Regardless of changes in the catalytic rate, changes in the concentration of SHP-1 could account for changes in the reaction rate between solution and tethered dephosphorylation. For example, in one of the experiments using PEG28, we use a SHP-1 concentration of $0.5\mu M$. In this experiment, the fitted effective concentration (σ^*) is $119.5\mu M$. If the catalytic rates from solution and when tethered are equivalent, then these changes in effective concentration will be reflected in the dephosphorylation rate. This indicates that tethering of SHP-1 can enhance the dephosphorylation rate more than 200-fold compared to dephosphorylation from solution, independent of auto-inhibitory effects.

4.4 Discussion

In this chapter, we quantified the molecular reach of a phosphatase, SHP-1, and many immune receptor cytoplasmic domains including PD-1, CD28, and SLAM. We showed that the reach of the enzyme contributes the majority of the reaction reach, but that changes in tether length do result in small differences in reach. We also demonstrated that enzyme

tethering can increase reaction rates, even without changes in the catalytic rate. This supports the growing body of work suggesting tether lengths can influence reactions [7, 10, 12]. Tether length therefore offers an interesting point of control for evolutionary, or synthetic, regulation of reactions. This work further confirms that disordered domains can behave as functional units, beyond as passive linkers.

Chapter 5

Discussion

In this dissertation, we discussed the importance of disordered proteins in biological systems. Specifically, we investigated how characteristics of disordered domains — including length, binding site location, post-translational modifications, long-lived interactions (e.g. dimerization), and short-lived interactions (e.g. phosphorylation-dependent membrane association) — can alter kinetic rates and interactions with downstream enzymes (Table 5.1). We consider three main systems: (1) T Cell receptor cytoplasmic domains and their interactions with the cell membrane and kinases, (2) formin-family proteins altering actin polymerization rates, and (3) immune co-receptor cytoplasmic domains interacting with the phosphatase SHP-1. We will discuss the contributions of each of these projects separately below.

5.1 Multisite Entropic Binding

T Cell activation is the hallmark of an adaptive immune response. Synthetic regulation of the adaptive immune response is of recent interest, particularly to develop cancer therapeutics (e.g. CAR T Cell Therapy). It is therefore critical to understand one of the earliest steps

Table 5.1: **Table of disordered protein characteristics and emergent behaviors.** Characteristics are features of disordered proteins included as inputs for some of the analyzed models. Emergent behaviors indicate the resulting behaviors a disordered protein can exhibit in a system.

Characteristics	Emergent Behaviors
• length	• concentration magnification
• binding site location	• preferential sequential binding
• local flexibility changes	• positive cooperativity
• changes in electrostatic interactions	• ultrasensitivity
• multiple simultaneous binding	• negative cooperativity
• multi-site dimerization	• rate changes

of T cell activation: T Cell Receptor signaling through phosphorylation of its disordered cytoplasmic domains. Experimental and computational research demonstrate that the disordered domains of the TCR are capable of positive cooperativity between phosphorylations [6]. In Chapter 2, we developed a freely-jointed chain model of the TCR disordered domains to explore how disorder can contribute nonlinear behaviors — such as cooperativity — to a signaling network. Using the freely-jointed chain as our base, we implement variations of the model to describe the multisite disordered domains in three cases: (1) disordered-to-ordered transitions induced by phosphorylation by the kinase Lck; (2) phosphorylation-dependent membrane association; and (3) multiple simultaneous binding of the kinase Zap70. In each model, we compute how each scenario influences the kinetics of the binding reaction. Using these distinct binding rates, we compute how changes in the entropy of the domain due to early binding events impact future binding events.

We find that, independent of post-translational modifications, the entropic fluctuations of the disordered domain naturally lead to binding rate differences between binding sites. Interestingly, we find that these differences lead to emergent preferential sequential binding (a preferred, but not required order of binding) to the domain. Obligate sequential binding (required order of binding) has previously been shown to lead to nonlinear behaviors

such as ultrasensitivity. Exploring the difference in consequences of obligate and preferential sequential binding is an interesting avenue of future work.

When we consider the three different contexts of post-translational modifications, we find that the corresponding rate differences lead to nonlinear behaviors in the system. Specifically, under disordered-to-ordered transitions, positive cooperativity and ultrasensitivity can be attained. We also find that phosphorylation-dependent membrane association could lead to a weak switch-like behavior. Finally, if multiple molecules bind to the TCR Complex at the same time, the added steric interference will create negative cooperativity among the binding sites.

This project offers many insights into how disorder could behave as more than a simple tether, and in fact contribute nonlinear signaling to the T cell activation pathway. These concepts are not only applicable to the T Cell Receptor. Many immune receptors tails, such as PD-1 and CD28, have at least two tyrosines that are phosphorylated as part of the immune response. Our results could easily extend to these molecules, and to other multisite disordered proteins as well.

5.2 Formin

Formins are a family of disordered proteins which modulate actin polymerization rates. Distinct formins can be found among different species and different contexts of actin polymerization (e.g. contractile rings, actin-driven cell motility). Some differences between the formin family members include changes in the number of polyproline regions and length of the formin. Typically, the disordered arms of formins form a dimer at the C-terminus, connected by the structured domain of formin, and are free to explore space at the N-terminus. However, some species of formins exhibit dimerization at both the C-terminus and the N-

terminus. This naturally leads to the following questions: What affect does dimerization at both ends of the disordered domain have on formin-mediated actin polymerization? Do other biological variations in formin sequence impact the polymerization rates?

In Chapter 3, we extend a freely-jointed chain model of the formin FH1 domains to examine how dimerization and polyproline track locations influence actin polymerization rates.

We find that formins dimerized at both the N- and C-terminus reduces the polymerization rate to actin compared to formins dimerized only at the C-terminus. Our collaborators, the Quinlan Lab at UCLA, have demonstrated that the formin, Fhod, dimerized at both the N- and C-terminals has a slower polymerization rate *in vitro* than C-terminal only dimerized Fhod, consistent with our predictions [personal communication]. Our simulations also find that different species of formins, with different length and spacing of binding sites, exhibit different polymerization rates. Specifically, the formin Capu — which has long, evenly spaced polyproline tracks — creates a higher actin polymerization rate compared to Fhod.

These results suggest the specific formin species could determine the actin polymerization rate, thereby regulating processes dependent on actin polymerization.

5.3 Molecular Reach

Disordered domains often appear as tethers anchoring structured molecules to another surface or molecule. These tethers can facilitate reactions, increasing the effective concentration of an enzyme to its substrate compared to the untethered reaction. Previous work has demonstrated that the length of the tether influences the rate of reaction [7, 10, 12]. For reactions where a tethered enzyme modifies another tether, we call the length scale over which this reaction can occur the 'molecular reach' of the reaction. The molecular reach is determined by the reach of both the tethers and the enzyme in the reaction.

In Chapter 4, we use experimental and computational methods to quantify the molecular reach of the reaction and extract the reach of the enzyme and tethers alone. We use surface plasmon resonance to measure time series of the phosphatase SHP-1 interacting with the disordered immune receptor PD-1. We then extract biophysical rate constants and the molecular reach from the data using ordinary differential equations model of the binding, unbinding, and catalysis of SHP-1 to a tether. We model each component of the molecular reach as a worm-like chain in order to extract the reach contributed by SHP-1 compared to the tether. We find that the molecular reach of the reaction is predominantly determined by the reach of SHP-1. However, we also find that changing the tether length will change the molecular reach of the reaction. This work quantifies the contributions of individual molecules to the molecular reach of a reaction. Although the majority of the molecular reach is contributed by the structured enzyme, changing the tether lengths can still influence the reaction. This work therefore suggests that tether length can act as a point of control for the regulation of reactions.

5.4 Conclusion

Each of the above chapters considers how disordered proteins can act as functional units in their own right, contributing important points of regulation to a system. In Chapters 2 and 3, we explore theoretical models of disordered proteins. This method of modeling offers a computationally efficient way to explore characteristics of disordered proteins and how post-translational modifications could influence their kinetics. For immune receptor tails, our models offer predictions for how disorder can confer nonlinear behaviors to signaling networks. In the case of formins, our models predicts force-dependent polymerization rates that change with the biological variation of the sequence. For formin, some of these predictions have been confirmed experimentally (unpublished data). In Chapter 4, we take a different approach to modeling behaviors of disordered proteins. We use *in vitro* experimental data,

differential equations, and a worm-like chain model to extract values for the molecular reach of the phosphatase SHP-1 and immune receptors tails, including PD-1, CD28, and SLAM. This data-driven approach to elucidating disordered protein function verifies the importance of disordered domain lengths in determining reaction rates. This work also highlights the relative contributions of the enzyme and tethers in the magnitude of the molecular reach of reaction. Through each of these models, we discover new facets of the behaviors of disordered proteins and their roles in various biological processes. Together with the growing body of disordered protein research, this work points to disordered proteins playing a vital role in the cell.

Bibliography

[1] B Alberts, K Hopkin, A. D. Johnson, and M Raff, *Essential Cell Biology* (W W NORTON & Company, 2019).

[2] M. Sickmeier, J. A. Hamilton, et al., "DisProt: The database of disordered proteins", Nucleic Acids Research **35**, D786–D793 (2007).

[3] A. B. Sigalov, D. A. Aivazian, V. N. Uversky, and L. J. Stern, "Lipid-binding activity of intrinsically unstructured cytoplasmic domains of multichain immune recognition receptor signaling subunits", Biochemistry **45**, 15731–15739 (2006).

[4] M. Guharoy, B. Szabo, S. C. Martos, S. Kosol, and P. Tompa, "Intrinsic structural disorder in cytoskeletal proteins", Cytoskeleton **70**, 550–571 (2013).

[5] H. Wang, T. A. Kadlecek, et al., "ZAP-70: an essential kinase in T-cell signaling.", Cold Spring Harbor Perspectives in Biology **2**, 1–17 (2010).

[6] H. Mukhopadhyay, B. De Wet, et al., "Multisite phosphorylation modulates the T Cell Receptor ζ-Chain potency but not the switchlike response", Biophysical Journal **110**, 1896–1906 (2016).

[7] Y. Zhang, L. Clemens, et al., "The Influence of Molecular Reach and Diffusivity on the Efficacy of Membrane-Confined Reactions", Biophysical Journal **117**, 1189–1201 (2019).

[8] C. A. Lopez, A. Sethi, et al., "Membrane-Mediated Regulation of the Intrinsically Disordered CD3ϵ Cytoplasmic Tail of the TCR", Biophysical Journal **108**, 2481–2491 (2015).

[9] K. Kasahara, H. Terazawa, T. Takahashi, and J. Higo, "Studies on Molecular Dynamics of Intrinsically Disordered Proteins and Their Fuzzy Complexes: A Mini-Review", Computational and Structural Biotechnology Journal **17**, 712–720 (2019).

[10] D. Van Valen, M. Haataja, and R. Phillips, "Biochemistry on a leash: The roles of tether length and geometry in signal integration proteins", Biophysical Journal **96**, 1275–1292 (2009).

[11] D. Bryant, L. Clemens, and J. Allard, "Computational simulation of formin-mediated actin polymerization predicts homologue-dependent mechanosensitivity", Cytoskeleton **74**, 29–39 (2017).

[12] J. Goyette, C. S. Salas, et al., "Biophysical assay for tethered signaling reactions reveals tether-controlled activity for the phosphatase SHP-1", Science Advances **3**, e1601692 (2017).

[13] R. Van Der Lee, M. Buljan, et al., "Classification of Intrinsically Disordered Regions and Proteins", Chemical Reviews **114**, 6589–6631 (2014).

[14] P Tompa and K.-H. Han, "Intrinsically disordered proteins", Physics Today **65**, 64–65 (2012).

[15] S. Gonfloni, J. C. Williams, et al., "The role of the linker between the SH2 domain and catalytic domain in the regulation and function of Src", EMBO Journal **16**, 7261–7271 (1997).

[16] D. R. Kovar and T. D. Pollard, "Insertional assembly of actin filament barbed ends in association with formins produces piconewton forces", Proceedings of the National Academy of Sciences **101**, 14725–14730 (2004).

[17] S. Romero, C. Le Clainche, et al., "Formin is a processive motor that requires profilin to accelerate actin assembly and associated ATP hydrolysis", Cell **119**, 419–429 (2004).

[18] E. Duchardt, A. B. Sigalov, D. Aivazian, L. J. Stern, and H. Schwalbe, "Structure induction of the T-cell receptor ζ-chain upon lipid binding investigated by NMR spectroscopy", ChemBioChem **8**, 820–827 (2007).

[19] M. E. Keir, M. J. Butte, G. J. Freeman, and A. H. Sharpe, "PD-1 and Its Ligands in Tolerance and Immunity", Annual Review of Immunology **26**, 677–704 (2008).

[20] A. Bah, R. M. Vernon, et al., "Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch.", Nature **519**, 106–9 (2015).

[21] A. Bah and J. D. Forman-Kay, "Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications", Journal of Biological Chemistry **291**, 6696–6705 (2016).

[22] J. E. Smith-Garvin, G. A. Koretzky, and M. S. Jordan, "T Cell Activation", Annual Review of Immunology **27**, 591–619 (2009).

[23] M. Lever, H.-s. Lim, et al., "Architecture of a minimal signaling pathway explains the T-cell response to a 1 million-fold variation in antigen affinity and dose", Proceedings of the National Academy of Sciences **113**, E6630–E6638 (2016).

[24] P. E. Love and S. M. Hayes, "ITAM-mediated signaling by the T-cell antigen receptor.", Cold Spring Harbor Perspectives in Biology **2** (2010).

[25] H. Zhang, S.-P. Cordoba, O. Dushek, and P Anton van der Merwe, "Basic residues in the T-cell receptor ζ cytoplasmic domain mediate membrane association and modulate signaling.", Proceedings of the National Academy of Sciences **108**, 19323–8 (2011).

[26] C. Xu, E. Gagnon, et al., "Regulation of T cell Receptor Activation by Dynamic Membrane Binding of the CD3ϵ Cytoplasmic Tyrosine-Based Motif", Cell **135**, 702–713 (2008).

[27] X. Guo, C. Yan, et al., "Lipid-dependent conformational dynamics underlie the functional versatility of T-cell receptor", Cell Research **27**, 505–525 (2017).

[28] E. Gagnon, D. a. Schubert, S. Gordo, H. H. Chu, and K. W. Wucherpfennig, "Local changes in lipid environment of TCR microclusters regulate membrane binding by the CD3ϵ cytoplasmic domain.", The Journal of Experimental Medicine **209**, 2423–39 (2012).

[29] E. K. Barber, J. D. Dasgupta, S. F. Schlossman, J. M. Trevillyan, and C. E. Rudd, "The CD4 and CD8 antigens are coupled to a protein-tyrosine kinase (p56 lck) that phosphorylates the CD3 complex", Proceedings of the National Academy of Sciences **86**, 3277–3281 (1989).

[30] E. H. Palacios and A. Weiss, "Function of the Src-family kinases, Lck and Fyn, in T-cell development and activation", Oncogene **23**, 7990–8000 (2004).

[31] D. E. Koshland and K. Hamadani, "Proteomics and models for enzyme cooperativity", Journal of Biological Chemistry **277**, 46841–46844 (2002).

[32] H. X. Zhou, "Loops in proteins can be modeled as worm-like chains", Journal of Physical Chemistry B **105**, 6763–6766 (2001).

[33] D. Reeves, K. Cheveralls, and J. Kondev, "Regulation of biochemical reaction rates by flexible tethers", Physical Review E - Statistical, Nonlinear, and Soft Matter Physics **84**, 1–12 (2011).

[34] M. L. Kutys, J. Fricks, and W. O. Hancock, "Monte carlo analysis of neck linker extension in kinesin molecular motors", PLoS Computational Biology **6** (2010).

[35] R. Phillips, J. Theriot, J. Kondev, and H. Garcia, *Physical biology of the cell* (Garland Science, 2012).

[36] W. Wimley and S. White, "Experimentally determined hydrophobicity scale for proteins at membrane interfaces", Nature Structural Biology **3** (1996).

[37] P. Lenz and P. S. Swain, "An Entropic Mechanism to Generate Highly Cooperative and Specific Binding from Protein Phosphorylations", Current Biology **16**, 2150–2155 (2006).

[38] X Shi, Y Bi, et al., "Ca2+ regulates T-cell receptor activation by modulating the charge property of lipids", Nature **493**, 111–115 (2013).

[39] O. Dushek, "Elementary steps int cell receptor triggering", Frontiers in Immunology **2** (2012).

[40] J. R. James, "Tuning ITAM multiplicity on T cell receptors can control potency and selectivity to ligand density", Science Signaling **11** (2018).

[41] C. Klammt, L. Novotná, et al., "T cell receptor dwell times control the kinase activity of Zap70.", Nature Immunology **16**, 961–9 (2015).

[42] G. P. O'Donoghue, R. M. Pielak, A. A. Smoligovets, J. J. Lin, and J. T. Groves, "Direct single molecule measurement of TCR triggering by agonist pMHC in living primary T cells.", eLife **2**, e00778 (2013).

[43] E. Neumeister Kersh, A. S. Shaw, and P. M. Allen, "Fidelity of T cell activation through multistep T cell receptor ζ phosphorylation", Science **281**, 572–578 (1998).

[44] N. S. Van Oers, B. Tohlen, et al., "The 21- and 23-kD forms of TCRζ are generated by specific ITAM phosphorylations", Nature Immunology **1**, 322–328 (2000).

[45] H. Mukhopadhyay, S. P. Cordoba, P. K. Maini, P. A. van der Merwe, and O. Dushek, "Systems Model of T Cell Receptor Proximal Signaling Reveals Emergent Ultrasensitivity", PLoS Computational Biology **9** (2013).

[46] C. Salazar and T. Höfer, "Multisite protein phosphorylation - From molecular mechanisms to kinetic models", FEBS Journal **276**, 3177–3198 (2009).

[47] T. Suwanmajo and J Krishnan, "Mixed mechanisms of multi-site phosphorylation", Journal of the Royal Society Interface **12** (2015).

[48] L. Wang, Q. Nie, and G. Enciso, "Nonessential sites improve phosphorylation switch", Biophysical Journal **99**, L41–L43 (2010).

[49] Y. He, Y. Chen, et al., "Phosphorylation-induced conformational ensemble switching in an intrinsically disordered cancer/testis antigen", Journal of Biological Chemistry **290**, 25090–25102 (2015).

[50] A. F. Chin, D. Toptygin, W. A. Elam, T. P. Schrank, and V. J. Hilser, "Phosphorylation Increases Persistence Length and End-to-End Distance of a Segment of Tau Protein", Biophysical Journal **110**, 362–371 (2016).

[51] O. Leupin, R. Zaru, T. Laroche, S. Müller, and S. Valitutti, "Exclusion of CD45 from the T-cell receptor signaling area in antigen-stimulated T lymphocytes", Current Biology **10**, 277–280 (2000).

[52] X. Wei, B. J. Tromberg, and M. D. Cahalan, "Mapping the sensitivity of T cells with an optical trap: Polarity and minimal number of receptors for Ca2+ signaling", Proceedings of the National Academy of Sciences **96**, 8471–8476 (1999).

[53] S. M. Patrick, S. Kim, N. S. Braunstein, J. L. Thomas, and E. F. Leonard, "Dependence of T cell activation on area of contact and density of a ligand-coated surface", Journal of Immunological Methods **241**, 97–108 (2000).

[54] J. C. Houtman, H. Yamaguchi, et al., "Oligomerization of signaling complexes by the multipoint binding of GRB2 to both LAT and SOS1", Nature Structural and Molecular Biology **13**, 798–805 (2006).

[55] R. N. Germain, "T-cell signaling: The importance of receptor clustering", Current Biology **7** (1997).

[56] W. Wu, C. Yan, et al., "Lipid in T-cell receptor transmembrane signaling", Progress in Biophysics and Molecular Biology **118**, 130–138 (2015).

[57] A. L. Cortajarena, G. Lois, et al., "Non-random-coil Behavior as a Consequence of Extensive PPII Structure in the Denatured State", Journal of Molecular Biology **382**, 203–212 (2008).

[58] R. Dominguez and K. C. Holmes, "Actin Structure and Function", Annual Review of Biophysics **40**, 169–186 (2011).

[59] A. J. Davidson and W. Wood, "Unravelling the Actin Cytoskeleton: A New Competitive Edge?", Trends in Cell Biology **26**, 569–576 (2016).

[60] H. N. Higgs, "Formin proteins: A domain-based approach", Trends in Biochemical Sciences **30**, 342–353 (2005).

[61] A. S. Paul and T. D. Pollard, "Review of the mechanism of processive actin filament elongation by formins", Cell Motility and the Cytoskeleton **66**, 606–617 (2009).

[62] D. Breitsprecher and B. L. Goode, "Formins at a glance", Journal of Cell Science **126**, 1–7 (2013).

[63] D. Vavylonis, D. R. Kovar, B. O'Shaughnessy, and T. D. Pollard, "Model of formin-associated actin filament elongation", Molecular Cell **21**, 455–466 (2006).

[64] A. Paul and T. Pollard, "The Role of the FH1 Domain and Profilin in Formin-Mediated Actin-Filament Elongation and Nucleation", Current Biology **18**, 9–19 (2008).

[65] N. Courtemanche and T. D. Pollard, "Determinants of formin homology 1 (FH1) domain function in actin filament elongation by formins", Journal of Biological Chemistry **287**, 7812–7820 (2012).

[66] E. C. Krainer, J. L. Ouderkirk, et al., "The multiplicity of human formins: Expression patterns in cells and tissues", Cytoskeleton **70**, 424–438 (2013).

[67] A. M. Rosales, H. K. Murnen, S. R. Kline, R. N. Zuckermann, and R. A. Segalman, "Determination of the persistence length of helical and non-helical polypeptoids in solution", Soft Matter **8**, 3673–3680 (2012).

[68] P. A. van der Merwe and O. Dushek, "Mechanisms for T cell receptor triggering.", Nature Reviews Immunology **11**, 47–55 (2011).

[69] R. J. Salmond, A. Filby, I. Qureshi, S. Caserta, and R. Zamoyska, "T-cell receptor proximal signaling via the Src-family kinases, Lck and Fyn, influences T-cell activation, differentiation, and tolerance", Immunological Reviews **228**, 9–22 (2009).

[70] J. L. Riley, "PD-1 signaling in primary T cells", Immunological Reviews **229**, 114–125 (2009).

[71] L. Chen and D. B. Flies, "Molecular mechanisms of T cell co-stimulation and co-inhibition", Nature Reviews Immunology **13**, 227–242 (2013).

[72] A Shahinian, K Pfeffer, et al., "Differential T cell costimulatory requirements in CD28-deficient mice.", Science **261**, 609–12 (1993).

[73] T. Yokosuka, M. Takamatsu, et al., "Programmed cell death 1 forms negative costimulatory microclusters that directly inhibit T cell receptor signaling by recruiting phosphatase SHP2", The Journal of Experimental Medicine **209**, 1201–1217 (2012).

[74] C. Blank and A. Mackensen, "Contribution of the PD-L1/PD-1 pathway to T-cell exhaustion: An update on implications for chronic infections and tumor evasion", Cancer Immunology, Immunotherapy **56**, 739–745 (2007).

[75] E. Hui, J. Cheung, et al., "T cell costimulatory receptor CD28 is a primary target for PD-1mediated inhibition", Science **4** (2017).

[76] F. Kienberger, V. P. Pastushenko, et al., "Static and Dynamical Properties of Single Poly(Ethylene Glycol) Molecules Investigated by Force Spectroscopy", Single Molecules **1**, 123–128 (2000).

[77] J. Yang, L. Liu, et al., "Crystal structure of human protein-tyrosine phosphatase SHP-1", Journal of Biological Chemistry **278**, 6516–6520 (2003).

[78] H. A. Watson, S. Wehenkel, J. Matthews, and A. Ager, "SHP-1: The next checkpoint target for cancer immunotherapy?", Biochemical Society Transactions **44**, 356–362 (2016).