

Signaling Behaviors of Intrinsically Disordered Proteins

Lara Clemens

Contents

1	Abstract	5
2	Introduction/Background/Motivation	5
2.1	Intrinsically Disordered Proteins	5
2.2	T Cell Receptor Zeta Chain	5
2.3	Previous Modeling Attempts - FJC as Model for IDPs	6
2.4	The Premise	7
3	Model Development	7
3.1	Code Validation	9
3.1.1	Root-mean-square end-to-end distance	9
3.1.2	R_{ee} Distribution	10
3.1.3	Occlusion Probability of End Site VS Analytical Result	11
3.2	Case Study: T Cell Receptor Zeta Chain	12
4	Local structuring of disordered signaling proteins gives rise to cooperativity and sequential binding	14
4.1	Introduction	14
4.2	Model and Methods	15
4.3	Results	16
4.3.1	Surface presence influences polymer configuration	16
4.3.2	Disorder inhibits ligand binding	16
4.3.3	Cooperativity may arise from local structuring in intrinsically disordered proteins . . .	17
4.3.4	Sequential binding naturally arises from polymer characteristics	18
4.3.5	Negative cooperativity may arise from local disordering of proteins	20
4.3.6	Future Work: Quantifying cooperativity from reversible phosphorylation with Hill coefficients	20
4.3.7	Future Work: Exploring sequence of binding from dephosphorylation and reversible phosphorylation	20
4.4	Discussion	21

5	Electrostatics	21
5.1	Introduction	21
5.2	Model and Methods	22
5.2.1	Electrostatic Potentials	24
5.3	Results	26
5.3.1	Model in which basic residues and tyrosines have similar energetics reveals puzzle in existing data	26
5.3.2	Initial parameter exploration of more general energetic model yields fit to polymer distribution	28
5.3.3	Future Work	30
6	Results: Simultaneous Binding	31
6.1	Introduction	31
6.2	Model and Methods	31
6.3	Results	32
6.3.1	Negative cooperativity arises from simultaneous binding of ligands to an IDP	32
6.3.2	Future Work: Sequence of binding multiple ligands	34
6.3.3	Future Work: Multiple chains towards a full model of a T cell receptor signalosome	35
6.4	Discussion	36
7	Role of the surface on tether-tether reactions within a signaling cluster	36
7.1	Introduction	36
7.1.1	Tethering	36
7.1.2	SHP-1	36
7.2	Model and Methods	36
7.3	Results	40
7.3.1	Accessibility of a reaction site is reduced by the presence of a surface	40
7.3.2	Reactions between tethered reactants are net enhanced by a surface	40
7.3.3	Reach parameter is unaffected by presence of surface	43
7.3.4	Matrix-bound versus surface-bound experiments confirm model predictions	45
7.3.5	Future Work: Catalysis surface factor $\kappa_{cat}^{(teth2teth)}$ versus l_C and/or versus r_{ligand}	45
8	Timeline	46

9	Appendix	46
9.1	Model Modifications	46
9.1.1	Stiffening	46
9.1.2	Electrostatics	46
9.1.3	Simultaneous Binding	47
9.2	Assumptions	47
9.3	Gillespie Algorithm	47
	References	47

1 Abstract

Disordered proteins are present in at least 40% of human proteins, including signaling molecules. Disordered signaling proteins can exhibit nonlinear signaling behavior, but how this behavior develops from disorder is not well understood. Here we explore how disordered proteins impact singular and multi-site ligand binding through a variety of disorder-specific phenomenons. In addition to basic properties of disordered protein interactions, we investigate disorder-to-order transitions, electrostatic membrane association, simultaneous ligand binding and effects of surface presence. We find that disordered proteins may create positive or negative cooperativity and intrinsic sequential binding. These effects are influenced by the length of the disordered protein, size of the binding ligand, location of the binding sites along the polymer and presence or absence of a surface. Intrinsically disordered proteins themselves may therefore act as signaling modules that contribute complex signaling behavior to a network.

2 Introduction/Background/Motivation

Traditionally, studies of protein function have gone hand-in-hand with studies of protein structure. Proteins such as hemoglobin exhibit complicated behavior, such as cooperativity, by modifying their structure. The cooperative transition of hemoglobin from a closed to open state is well studied with the aid of crystallography and other structure elucidation tools. **CITE**

Of more recent interest is the behavior of molecules that cannot be crystallized. These proteins lack a defined structure and instead are capable of assuming many different conformations. Although examples of intrinsically disordered proteins (IDPs) have been reported since the 1970s, it was only in the past two decades that they became a focus of major research. **[Dunker2008]**

As our understanding of disordered proteins develops, so too will our understanding of a variety of cellular behaviors. These studies will elucidate aspects of signaling, cytoskeleton formation, clustered reactions, and **really. I need a fourth function.** Investigations into how disordered proteins mediate each of these processes may lead to new drug targets or introduce new directions for synthetic biology.

2.1 Intrinsically Disordered Proteins

Studies have shown that disordered proteins or disordered domains are present in at least 40% of human proteins, including those involved with signal propagation.**[Tompa2012]** General functions of IDPs include as tethers between two globular domains **cite**, receptor subunits in signaling pathways **cite**, tethers to the membrane **cite**, and facilitators to actin polymerization **cite Bryant2017 or someone else older?**. Given the ubiquitous nature of IDPs, many questions arise: How does their existence influence cellular functions such as biochemical reactions, signaling networks, or cytoskeletal structure? Are there any benefits to being disordered over structured? and Can IDPs exhibit the same complicated behavior as structured proteins?

2.2 T Cell Receptor Zeta Chain

One example of an intrinsically disordered protein is the CD3 ζ chain, one of six disordered chains composing the T Cell Receptor (TCR) intracellular region. This molecule facilitates signal propagation in the TCR network in the immune system. An antigen binding to the extracellular regions of the TCR creates a signal transmitted via a chain of events into the cell to the intracellular components of the TCR including the CD3 ζ chain. CD3 ζ undergoes multiple phosphorylation by kinase Lck before another molecule, ZAP-70 can attach and propagate the signal downstream. This pathway ultimately regulates T-cell cell fate decisions

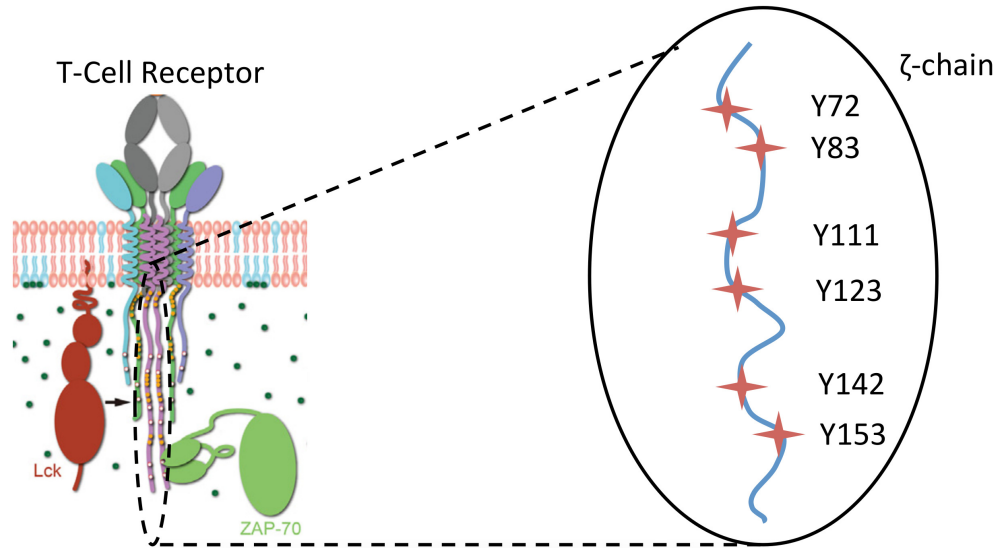


Figure 1: Cartoon of T Cell Receptor network (left) from [Wu2015]. Location of tyrosines in single ζ chain (right).

through cytokines production, (i.e. interleukin-2). Cite - Cell Signal.com?? Presumably I need to cite all of this somewhere?

Experiments with only a reconstituted mouse CD3 ζ dimer, simplified extracellular receptor, kinase, and phosphatase have shown that the number of tyrosines impacts the potency and maximum phosphorylation but not the switch-like response reword [Mukhopadhyay2016]. Initial differential equation models indicate to achieve these characteristics, there would need to be a phosphorylation-dependent enhancement of more than 100-fold. That is, the sixth phosphorylation event would be at least 100-fold faster than the first phosphorylation event. [Mukhopadhyay2016]

Questions:

Their ODE model is single chain, right?

Is a single zeta chain sufficient for signalling?

Mukhop. et al says a single ITAM phos is technically sufficient, but has it been shown that it will?

2.3 Previous Modeling Attempts - FJC as Model for IDPs

Disordered proteins are commonly represented with models from polymer physics.[VanValen2009],[Reeves2011] The distribution of end-to-end distances for disordered domains matches a worm-like chain (WLC) model with persistence length 3.04Å. [Zhou2001] This suggests $l_k = 0.6\text{nm}$...which is not what we use...not sure how to explain that one. Basically we use the lower limit of what is reasonable and could see what difference it makes either by comparing to $N=1/2N_{\text{max}}$ or by resimulating. Find $l_k = 0.5$ in Mukhopadhyay2016 too Models of disordered proteins using freely-jointed chain (FJC) and WLC converge, with the persistence length for the WLC as half the Kuhn length used for the FJC. [VanValen2009]

Alternative models for multisite phosphorylation of IDPs include molecular dynamic, ordinary differential equations, or particle based models. However, FJC or other ‘mesoscale’ approaches reach timescales on the

order of microseconds to seconds, which are computationally out of reach for traditional atomic scale MD. This approach also allows us to capture the steric effects of a disordered chain, which are missed by coarser models.

Representations of disordered proteins as freely-jointed chains have already been used to elucidate properties of IDPs. For instance, the length of tethers between domains controls the effective concentration of one domain seen by the other, such as auto-inhibition of WASP. [VanValen2009] In a second example, the disordered molecule, formin, captures profilin-actin monomers and delivers them to the growing end of actin, increasing the actin polymerization rate. In experimental studies of formin, exerting force on the distal end of formin counterintuitively enhances the polymerization rate two-fold over relaxed formins. [Jegou2013] Models of formin as a freely-jointed chain offer an explanation for this phenomenon. A force exerted on a freely-jointed chain extends the polymer, increasing capture of profilin-actin by increasing the availability of binding sites. This increase of capture rate balances the reduction in delivery rate to have a net positive impact on the actin polymerization rate. [Bryant2017]

2.4 The Premise

Given the above evidence that disordered proteins can give rise to nonlinear signaling behavior (e.g.,), the biological premise of the following projects is that disordered polymer properties alone (e.g.,) can create these effects, for example through steric interactions with rigid ligands. The modeling premise is that since disordered proteins take on a wide ensemble of conformations, modeling approaches which focus on structure alone are insufficient. On the other hand, a ‘mesoscale’ modeling approach able to describe conformational ensembles will be able capture the nonlinear signaling properties arising from disorder.

3 Model Development

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. In our simulation, the FJC is allowed to explore its configuration space through randomized movements. The FJC model tracks where each joint is located, making steric interactions with the ligand easy to simulate compared to using the continuous WLC model. The ligand is simulated as an idealized sphere which may interact with the FJC. We compute quasi-equilibrium statistics of the chain and its bound or unbound ligands using a Monte Carlo (Metropolis) Algorithm.

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 . We can compute the difference in K_D between two states, e.g. fully phosphorylated compared to unphosphorylated. From detailed balance, we have $K_D \equiv k_{\text{off}}/k_{\text{on}}$, which allows us to compute the change in K_D as follows:

$$\frac{k_{on}}{k_{off}} = \exp\left(\frac{-\Delta G}{k_B T}\right) \quad (1)$$

$$\frac{K_{D_1}}{K_{D_2}} = \frac{\frac{1}{\exp\left(\frac{-\Delta G_1}{k_B T}\right)}}{\frac{1}{\exp\left(\frac{-\Delta G_2}{k_B T}\right)}} \quad (2)$$

$$= \frac{\exp\left(\frac{-\Delta G_2}{k_B T}\right)}{\exp\left(\frac{-\Delta G_1}{k_B T}\right)} \quad (3)$$

$$= \exp\left(\frac{\Delta G_1 - \Delta G_2}{k_B T}\right) \quad (4)$$

$$= \exp\left(\frac{(E_1 - TS_1) - (E_2 - TS_2)}{k_B T}\right) \quad (5)$$

$$= \exp\left(\frac{S_2 - S_1}{k_B}\right) \quad (6)$$

$$= \exp\left(\frac{(S_{on}^2 - S_{off}^2) - (S_{on}^1 - S_{off}^1)}{k_B}\right) \quad (7)$$

$$= \exp\left(\frac{(k_B \ln(\frac{\Omega_2 P_2}{\Omega_2})) - (k_B \ln(\frac{\Omega_1 P_1}{\Omega_1}))}{k_B}\right) \quad (8)$$

$$\frac{K_{D_1}}{K_{D_2}} = \frac{P_2}{P_1} \quad (9)$$

$$(10)$$

where $G_j = E_j - TS_j$ is the free energy of binding in the free or rigid 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 be the total number of microstates, with $\Omega_1 * P_1$ being the microstates available when the ligand is bound and $\Omega_2 * P_2$ is the number of microstates available when the polymer is rigid. We also note $E_2 = E_1$ since we are working at equilibrium.

For example, we can consider the change in K_D between a fully rigid polymer, where all binding sites are available and $K_D = K_{D_R}$, compared to a floppy, unstructured polymer, where binding site availability depends on the conformation of the polymer and $K_D = K_{D_F}$. Then we calculate the change in K_D as above, but since the rigid state always allows binding, we may write $P_R = 1$.

$$\frac{K_{D_F}}{K_{D_R}} = \frac{\frac{1}{\exp\left(\frac{-\Delta G_F}{k_B T}\right)}}{\frac{1}{\exp\left(\frac{-\Delta G_R}{k_B T}\right)}} \quad (11)$$

$$= \frac{P_R}{P_F} \quad (12)$$

$$\frac{K_{D_F}}{K_{D_R}} = \frac{1}{P_F} \quad \text{assuming } P_R = 1 \quad (13)$$

$$(14)$$

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

$$\frac{K_{DF}}{K_{DR}} = \frac{1}{(1 - P_{occ})}. \quad (15)$$

This comparison has already been used in Mukhopadhyay et al. as a possibility to account for the change in phosphorylation rates of the TCR ζ chain. [Mukhopadhyay2016] We consider the generalized version for the rest of our work.

Steric occlusion by the polymer will impact 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} .

We calculate the occlusion probability by computing how often 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 surface constraints. To determine if a site is occluded in a given conformation, we check if any of the segment end points are located within the sphere of interest. If a surface is present, we also check if the ligand sphere crosses below the half-space surface designated at $z = 0$. Given that the binding 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.

3.1 Code Validation

There are theoretical solutions for many aspects of the freely-jointed chain. This provides a basis with which to verify our code.

3.1.1 Root-mean-square end-to-end distance

First, we look at the average end-to-end distance of the polymer (RMS). In our simulations, we normalize by the Kuhn length, so all simulations assume $\delta = 1$ and record all other parameters in units of Kuhn lengths. We know from polymer physics that the RMS should increase as $\delta\sqrt{N}$, which given $\delta = 1$, is just \sqrt{N} .

Average end-to-end distance (Root-mean-square end-to-end distance): [Reeves2011]

$$\sqrt{\langle r_{ee}^2 \rangle} = \sqrt{N\delta^2} = \delta\sqrt{N}$$

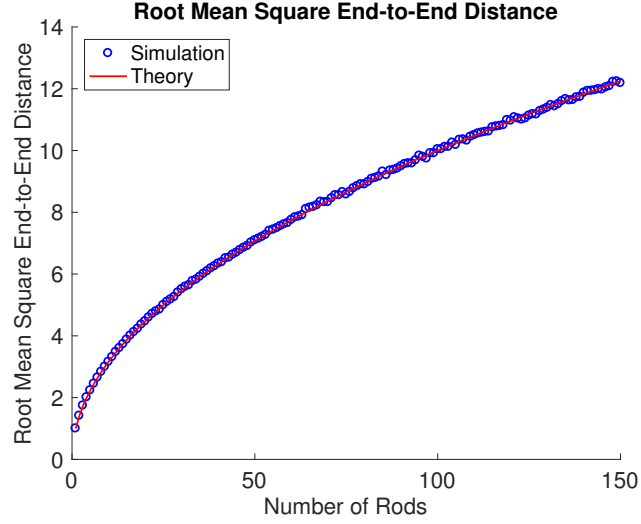


Figure 2: Theoretical root mean square end-to-end distance (red line) against simulated values (blue circles).

3.1.2 R_{ee} Distribution

End-to-end distribution: [VanValen2009], [Reeves2011]

$$P(r_{ee}) = 4\pi r^2 \left(\frac{3}{2\pi N \delta^2} \right)^{\frac{3}{2}} \exp \left(\frac{-3r^2}{2N \delta^2} \right)$$

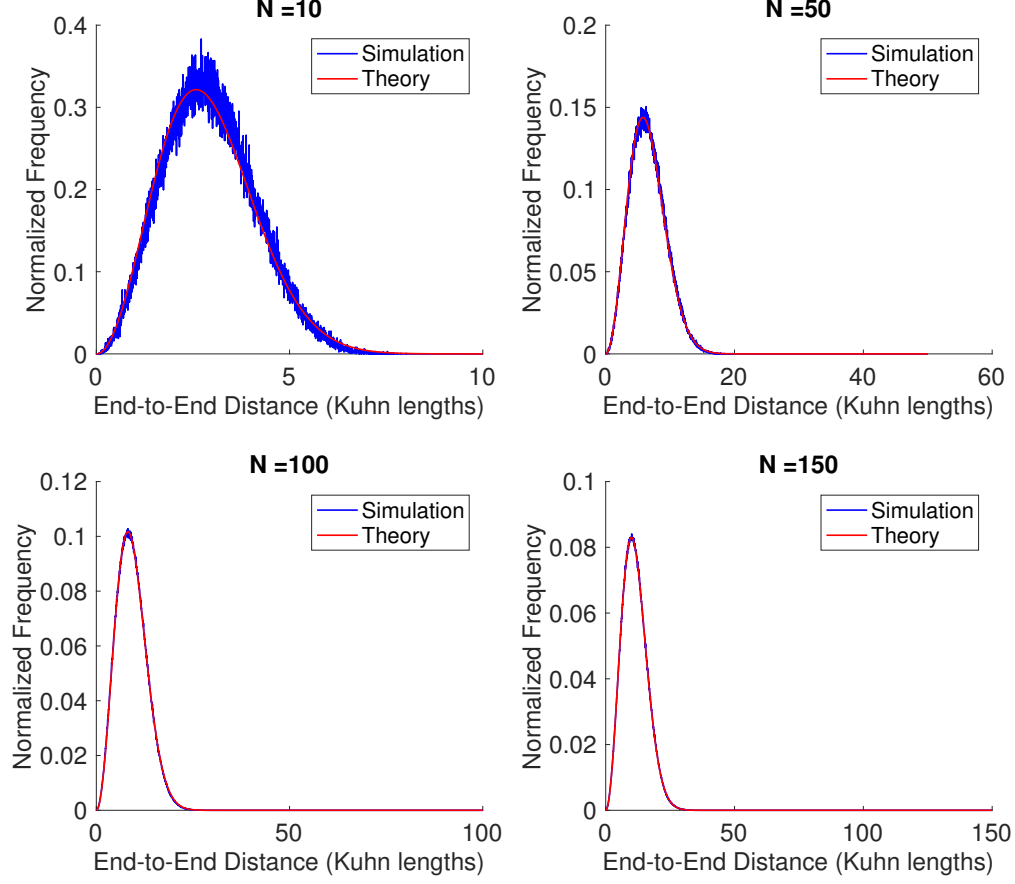


Figure 3: Simulated end-to-end distance distribution (blue) against theoretical distribution (red) for multiple polymer lengths (N).

3.1.3 Occlusion Probability of End Site VS Analytical Result

We consider analytical solutions to the simplest case, where the ligand is attempting to bind at the end of the freely-jointed chain in free-space. The probability that a random walk over time 0 to $\delta * N$, beginning at the edge of a sphere of radius, R , will cross into the sphere at any time is analogous to the probability the ligand is occluded by the freely jointed chain with N segments and Kuhn length δ . In order to solve this analytically, we must assume the random walk begins ϵ away from the sphere or it will always count as a failure. Then we may formulate as follows for the probability of survival, $p(\vec{x}, t)$ when the random walk begins at a starting point, $r = R + \epsilon$:

$$\begin{cases} \frac{\partial p}{\partial t} = D \nabla^2 p, & p = 0 \text{ at } r = R \\ p(\vec{x}, 0) = \delta(\vec{x} - (R + \epsilon)) \end{cases} \quad (16)$$

In order to get a finite estimate to compare, we consider the solution when $N \rightarrow \infty$.

$$\begin{cases} q(r) = \mathbb{P}(\text{not hit sphere} | x(0) = r) \\ 0 = \frac{2}{r}q'(r) + q''(r) \\ q(R) = 0 \\ q(r) \rightarrow 1 \text{ as } r \rightarrow \infty \end{cases} \quad (17)$$

We assume $\epsilon = 1$ Kuhn length and R measured in Kuhn lengths:

$$q(r) = 1 - \frac{R}{r}q(r) = 1 - \frac{R}{R+1} \quad (18)$$

We see that when we compare our simulated binding probability to the analytic solution, we get good agreement. This tells us both our code is working as desired and that $N=100$ is approximately equivalent to $N \rightarrow \infty$.

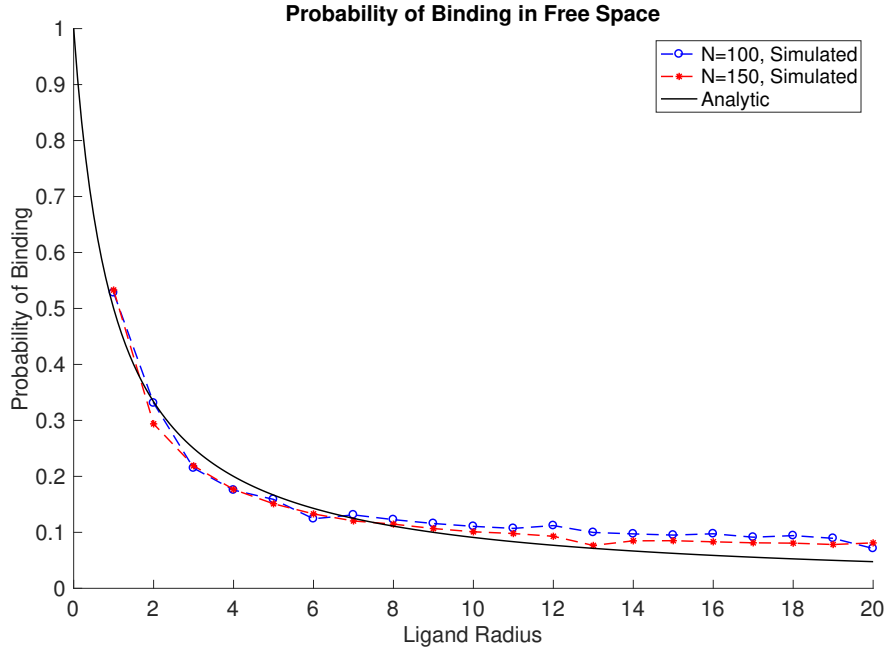


Figure 4: Analytic Solution

3.2 Case Study: T Cell Receptor Zeta Chain

For the following study, we will focus on the mouse TCR CD3 ζ chain. The TCR CD3 ζ chain is a subunit of CD3 consisting of 164 amino acids. Of these, 21 are included in the signal peptide region of the protein. The remaining 143 amino acids make up the extracellular, transmembrane and cytoplasmic regions of the CD3 ζ chain. The cytoplasmic tail is an intrinsically disordered chain of 113 amino acids containing multiple phosphorylation sites, called ITAMs (immunoreceptor tyrosine-based activation motif). There are three ITAMs on the ζ chain, each containing two tyrosines. The tyrosine kinase Lck phosphorylates each tyrosine in the ζ chain and it is thought that phosphorylation of one site influences the binding of the kinase to other sites.

In mouse CD3 ζ , the cytoplasmic tail spans residues 52-164 and the tyrosines are located at residues 72, 83, 111, 123, 142, 153. Therefore, if we were to renumber to begin at the beginning of the cytoplasmic tail, the

region would be $N = 113$ amino acids long, with tyrosines located at $i = 21, 32, 60, 72, 91, 102$. (UniProt, entry P24161) Given an assumption of 0.3 nm per Kuhn length (i.e. one Kuhn length is equivalent to one amino acid), then the tyrosines are similarly located along the 113 segments of the FJC.

Mouse Lck is composed of an SH3, SH2 and protein kinase domain connected by small loops. The domains are 61, 98, and 254 amino acids respectively. (UniProt entry P06240) Using a protein molecular mass calculator, we calculate that the kinase domain is 29.08 kDa. If we assume a protein density of 1.41 g / cm^3 then we can estimate the volume of the kinase domain [Fischer2004]:

$$(29 \times 1000 \text{ Da}) * (1.66 \times 10^{-27} \text{ kg / Da}) * (1000 \text{ g / kg}) / (1.41 \text{ g / cm}^3) = 34 \text{ nm}^3.$$

If we then approximate the kinase domain as a sphere, then we can estimate a radius:

$$\begin{aligned} V &= \frac{4}{3} \pi r^3 \\ 34 \text{ nm}^3 &= \frac{4}{3} \pi r^3 \\ r &\approx 2 \text{ nm} \end{aligned}$$

We measure the maximal length of the kinase domain in PyMol from PDB 3LCK. Rounding up for error, we have a maximal distance of 58 \AA . This gives a maximal spherical estimate with a radius of 2.9 nm, or about ten Kuhn lengths. (Fig. 5a) If we instead measure rough length, width, and height for the kinase domain, we have measurements of 36.6 \AA , 29.4 \AA , and 45.1 \AA respectively. (Fig. 5b) From these we can estimate a sphere with volume corresponding to the volume of the rectangular prism with those dimensions. This estimates a sphere with radius 2.3 nm, or about eight Kuhn lengths. Based on all of these estimates, we choose to represent Lck with a radius of seven Kuhn lengths.

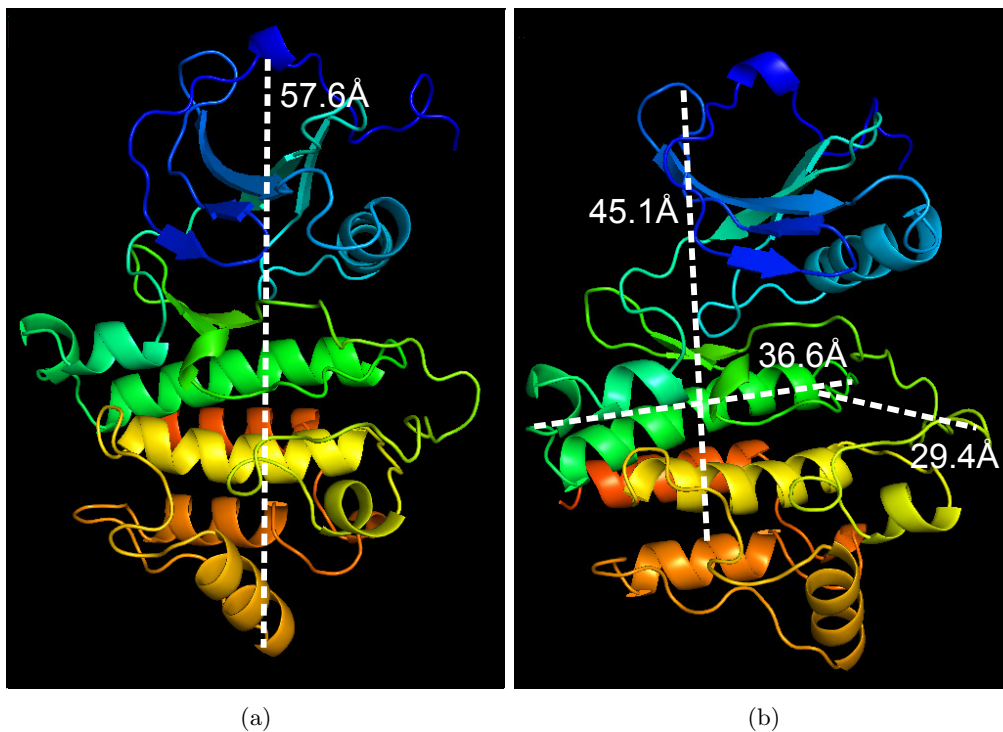


Figure 5: PDB 3LCK - kinase domain of Lck

4 Local structuring of disordered signaling proteins gives rise to cooperativity and sequential binding

4.1 Introduction

There is evidence that indicates some IDPs undergo a disordered-to-ordered transition upon post-translational modification, i.e. phosphorylation. For example, phosphorylation of PAGE4 restricts the number of conformations it may explore, increasing rigidity of the backbone. [He2015] Post-translational modifications may also cause a conformational change, altering the availability of the binding site. The neural protein 4E-BP2 undergoes multisite phosphorylation which leads to a disorder-to-order transition, forming β sheets which hide the eIF4E binding site. [Bah2015]

Initial models of the TCR ζ chain indicate large enhancements of the phosphorylation rate can occur assuming a complete switch from disorder to order upon initial phosphorylation. Binding to a fully structured protein compared to a disordered protein can be 100 fold easier, depending on the persistence length and ligand size. [Mukhopadhyay2016]

We are interested in how cooperative effects may arise in reactions with a disordered protein. One model we explore is disorder-to-order transitions as a result of post-translational modifications. With regards to our case study, CD3 ζ , we assume that each phosphorylation event causes a fraction of the chain to undergo a disorder-to-order transition. Partial structuring will reduce the number of conformations available to the protein. We predict this will increase the number of conformations where a ligand may access its binding site due to a forced linearization of part of the chain. We investigate how this local structuring phenomenon would impact subsequent phosphorylations by Lck along TCR CD3 ζ . Specifically, we want to know if local

structuring of a disordered protein upon multi-site phosphorylation can induce enhancement of subsequent binding events.

4.2 Model and Methods

We model the TCR CD3 ζ chain as a freely jointed chain, where each amino acid is represented as a segment as described in Sec. 3. The unbound ligand is simulated as an idealized ghost sphere located tangential to a single segment. We calculate quasi-equilibrium statistics for the FJC and unbound ligand in both free-space and half-space. For the CD3 ζ chain, we simulate only the cytoplasmic region by an FJC with 113 segments. In our simulation, all lengths are normalized by the Kuhn length, so each rod has length 1 and the radius of the ligand is also measured in Kuhn lengths. We consider a spherical estimate for the kinase Lck, with a radius of 2.1 nm (7 Kuhn lengths) as calculated in 3.2.

We calculate how often a ligand is able to bind to an oriented sphere tangentially attached to the polymer, where ‘able to bind’ refers to the specified sphere being empty of both other polymer segments and the surface, if present.

We include local structuring 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. (Fig. 6) The section of frozen segments acts as a single long segment, still rotating when other neighboring segments rotate. This method of local structuring leaves the original binding site in a primarily available configuration, however, we are interested in how it impacts the neighboring modification targets (i.e. tyrosines to be phosphorylated). We explore how the total number of structured residues per modification impacts the result.

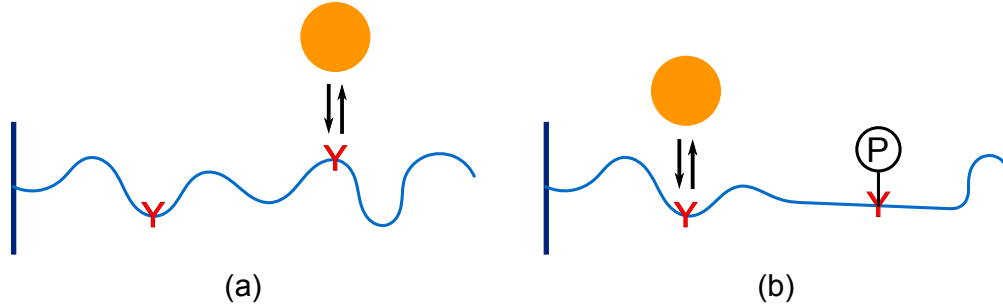


Figure 6: Cartoon of partial stiffening model. Entirely floppy FJC prior to modification (a) transitions to partially rigid chain after modification (b).

With respect to TCR CD3 ζ , phosphorylation events are modeled as creating local structure as described above. We model dephosphorylation of CD3 ζ , using the same parameters for the FJC. We also assume the phosphatase is the same size as the ligand Lck. However, dephosphorylation we model as ‘unfreezing’ nearby segments. Under the premise of eventually modeling reversible phosphorylation, we consider the case where the initial polymer state is only as stiff as it would be from the local stiffening by full phosphorylation. That is, if phosphorylation would locally structure the nearest eleven segments, then the fully phosphorylated form of CD3 ζ would have 66 residues frozen. This is, the state the chain would then begin dephosphorylating from, where each dephosphorylation would unstructure 11 of the initial 66 frozen residues.

4.3 Results

4.3.1 Surface presence influences polymer configuration

We first explore how a freely-jointed chain is affected by the presence of a surface or membrane. Generally, FJCs are most likely to be in a ‘hairball’ conformation due to entropic forces. When we compare the end-to-end distances of a FJC in free space to a FJC in half space, we find a shift in the distribution. The half-space polymer tends to have a larger end-to-end distance than the free-space polymer. (Fig. 7) In half space, the polymer has fewer conformations it can take on since it can only occupy half of the region. This prevents many more ‘hairball’ conformations than it prevents straightened conformations. Overall, the presence of a membrane straightens the polymer.

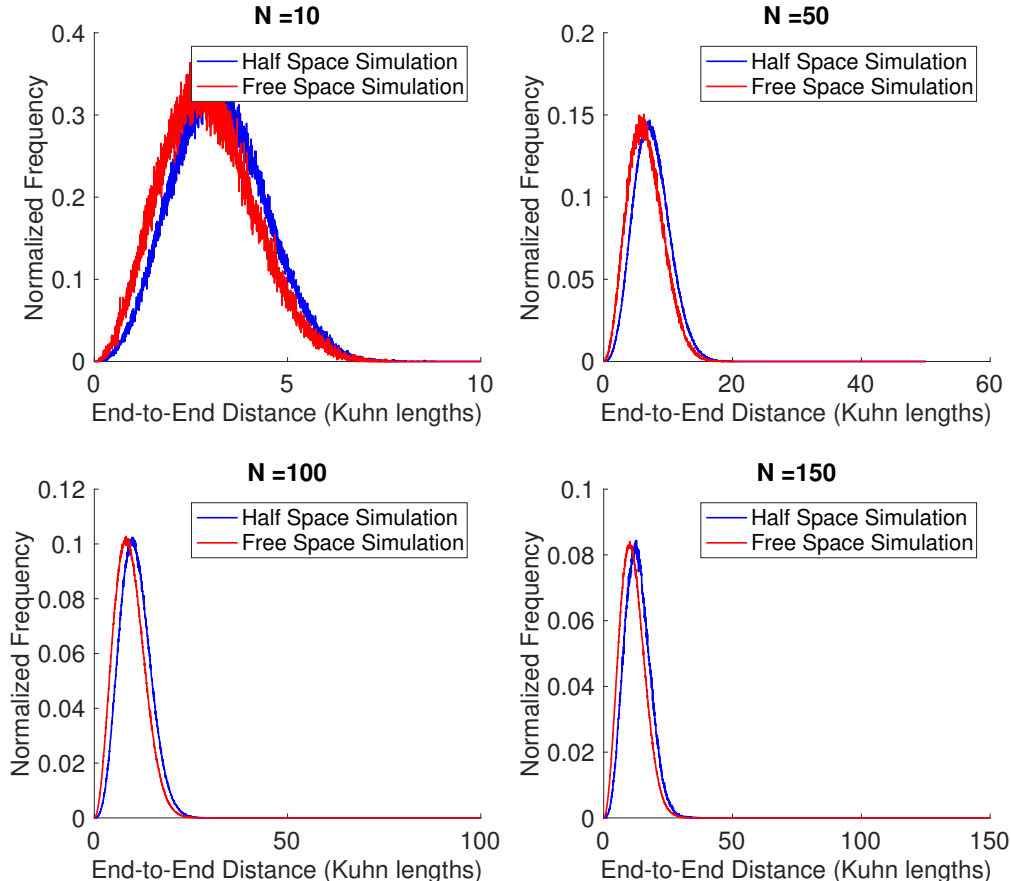


Figure 7: Distributions of polymer end-to-end distance in free space and half space for various polymer lengths (N). For each length, the end-to-end distance of the polymer in half-space is overall straighter than in free space.

4.3.2 Disorder inhibits ligand binding

We calculate how often a ligand may bind to a FJC, under varying characteristics for the FJC. We can see that in free space, the ligand is more occluded when there are more rods and/or the ligand is larger, for a specific binding site. (Fig. 8) These are both intuitive results. A longer chain creates more conformations where any location on the chain will be occluded by other segments. A shorter or less flexible chain will not be able to bend enough to bury a binding site. A larger ligand will require more space near the binding site

to be open, which is naturally less likely. We also find that the binding site location impacts the occlusion probability. A binding site in the middle of the polymer will also be less available to ligand binding since it is more likely sheltered by the rest of the ‘hairball’ polymer.

We compare how FJC parameters influence binding in free-space to binding in half-space. The membrane increases the likelihood of straighter conformations, which would be likely to increase the binding probability. However, the membrane itself may also occlude the ligand since we include an orientation of the binding site in our simulation. Therefore if the binding site is oriented down towards the membrane, then even if the polymer is not occluding the ligand, the membrane still might. When we include a hard plane membrane in our simulation, we see that the presence of a membrane generally decreases the probability of ligand binding. When the ligand is small enough, the difference in binding probabilities between free space and half space binding is minimal. (Fig. 8)

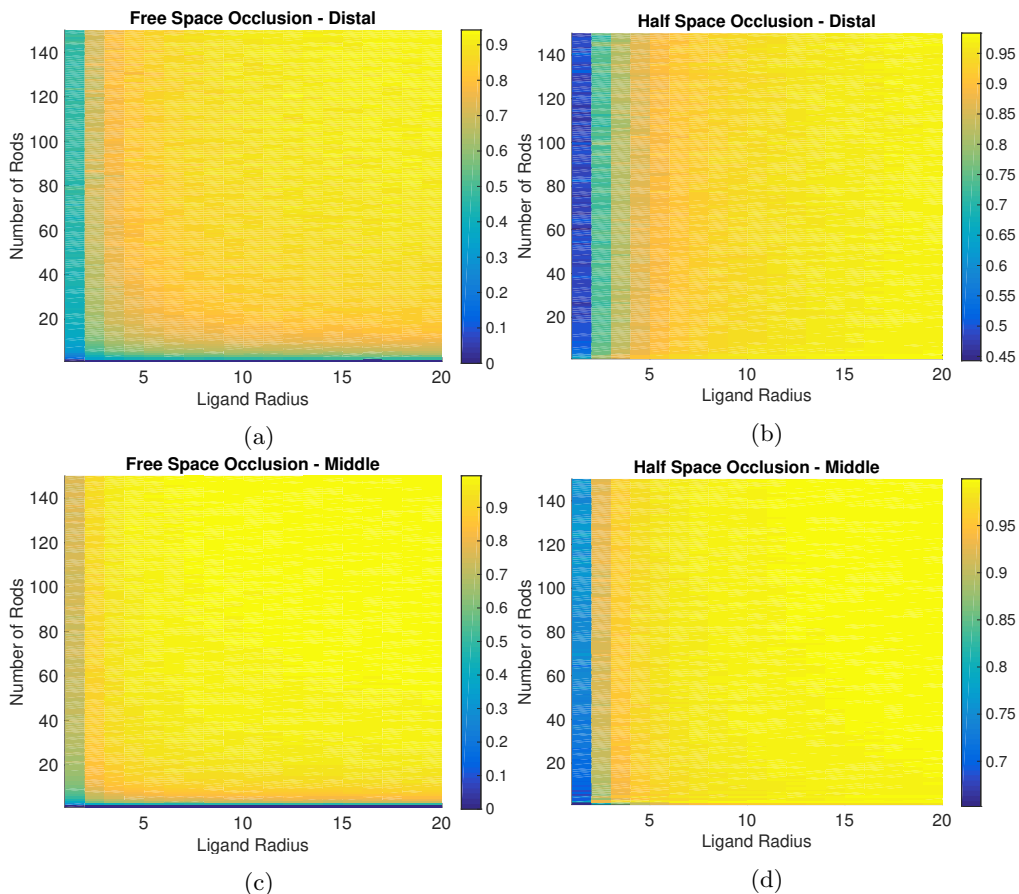


Figure 8: Probability of occluding a ligand of variable radius from a polymer of variable contour length when the binding site is at the distal end or middle of the chain in free space or half space. (a) Free space polymer with binding site at distal end. (b) Half space polymer with binding site at distal end. (c) Free space polymer with binding site in the middle. (d) Half space polymer with binding site in the middle.

4.3.3 Cooperativity may arise from local structuring in intrinsically disordered proteins

We now simulate the kinase Lck binding to the CD3 ζ chain, assuming that each phosphorylation event will create a disorder-to-order transition. If we assume that phosphorylation of tyrosines on the CD3 ζ chain induces local structuring, then neighboring tyrosines will feel a reduction of entropic occlusion. This will

cause them to be more available to binding by a ligand. The magnitude of this effect will vary based on how much structuring occurs, e.g. if it affects the nearest two residues or the nearest ten residues.

We see from these simulations that the average binding rate of a ligand to the unphosphorylated tyrosines increases as the number of phosphorylated tyrosines (and therefore the number of structured residues) increases. From the first to the sixth phosphorylation event, there is a 3-10 fold increase in the average binding rate for moderate ($1/12$ of chain length) to severe ($1/6$ of chain length) local stiffening per phosphorylation. (Fig. 9)

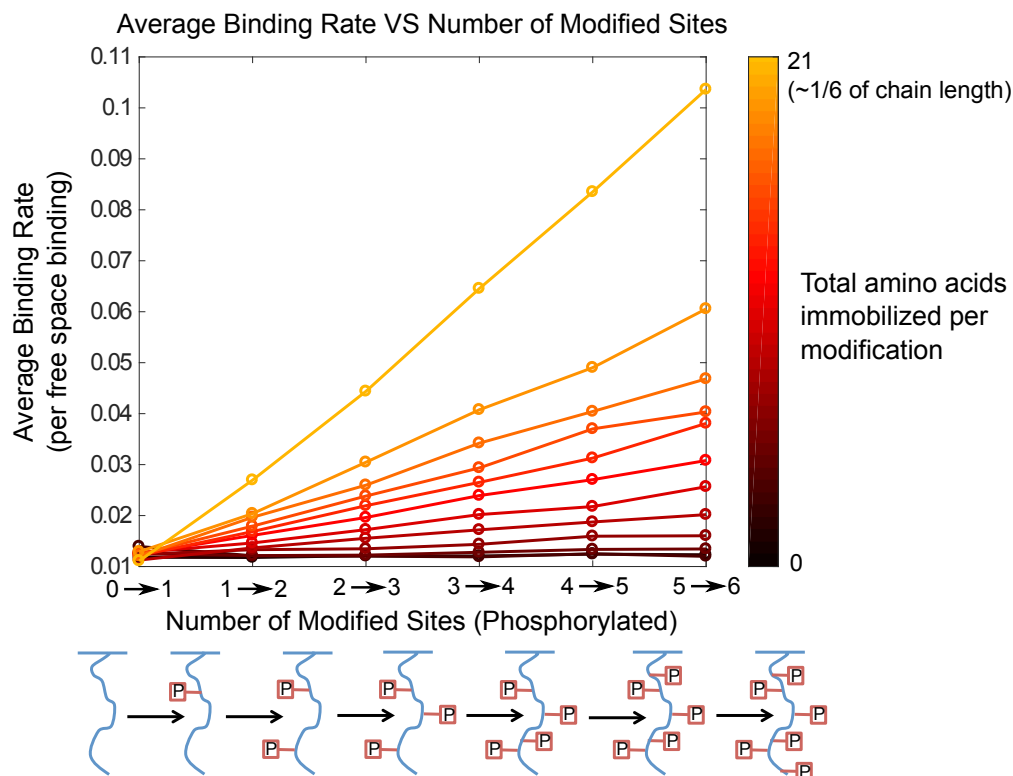


Figure 9: (Above) Simulated average binding rates of Lck to CD3 ζ for different levels of phosphorylation given a specified degree of local structuring (none \rightarrow black, low \rightarrow red, high \rightarrow orange). (Below) Cartoon representing a possible phosphorylation state series of CD3 ζ .

4.3.4 Sequential binding naturally arises from polymer characteristics

We hypothesized that without any modifications due to binding, then with only the presence of a membrane, we would see preferential binding in an outwards to inwards (6-5-4-3-2-1) manner. In regards to our model system, that is to say we expected the tyrosine furthest from the membrane to be phosphorylated first and the tyrosine closest to the membrane to be phosphorylated last. We surmised this from the idea that tyrosine's too close to the membrane would see extra kinase inhibition from membrane occlusion, whereas the tail of the disordered protein would not experience this same effect.

When considering the partial stiffening model, we predict that this effect would be enhanced. In an unphosphorylated state, if the tyrosine furthest from the membrane is most likely to be bound, then partial stiffening would give increased preference to the next furthest from the membrane (here called the fifth tyrosine) as the second to be phosphorylated. This comes from the reduction of total configurations available to the protein. Since the stiffening would occur around the sixth tyrosine, the nearest unphosphorylated tyrosine would be

the most affected. The segment of polymer near the fifth tyrosine would have fewer available conformations and therefore proportionately fewer would be occluding the fifth tyrosine.

We obtained relative probabilities of binding for each tyrosine along the chain at each phosphorylation state. With this information, we are able to explore whether or not there is a dominant sequence of phosphorylation events. We ran a Gillespie algorithm with six events, one for each phosphorylation. From each run, we record which sequence or 'path' of phosphorylation. When we compile the path frequency data, it becomes clear that the first phosphorylation event is dominant. That is to say, any path beginning by phosphorylating the membrane distal tyrosine (6th tyrosine), will occur more often than any path beginning with the membrane proximal tyrosine.

For simplicity, we consider only the paths membrane proximal to membrane distal (123456) and distal to proximal (654321). We see that when a membrane is present, the probability of phosphorylating distal to proximal is much higher than that of phosphorylating proximal to distal. It is also more probable than if all paths were equally likely. This phenomenon is easily explained by the presence of a membrane. The membrane proximal tyrosine has a smaller range of space it may occupy than the distal tyrosine. Therefore, it is more likely to be close to the membrane in a configuration. Since the ligand cannot penetrate the membrane, tyrosines closer to the membrane have a higher probability of being effectively sheltered from ligand binding. This makes the distal to proximal path much more likely, regardless of local stiffening effects. This path preference is enhanced with local structuring since the distal tyrosine is more accessible and local structuring makes neighboring tyrosines even more available.

One would expect then, that a cytoplasmic disordered protein would not experience a path preference between N-terminal to C-terminal or vice versa. When we simulate CD3 ζ without a membrane, there is still a marked preference for C-terminal to N-terminal phosphorylation. This effect arises from the spacing of the tyrosines along the polymer's length. As noted above, the location of a tyrosine impacts how likely a ligand is to bind there. ⁸ Since the tyrosines in CD3 ζ are not equally spaced along the length of the protein, there is a bias for the tyrosine closest to one end, in this case the tyrosine closest to the C-terminal. When we simulate a polymer with the same length as CD3 ζ but equally spaced tyrosines, we see that there is no longer a preference between the two paths.

We are soon going to introduce a ranking system for the paths in order to do a full analysis of all possible paths. This will help analyze the likelihood of any of the non-sequential paths.

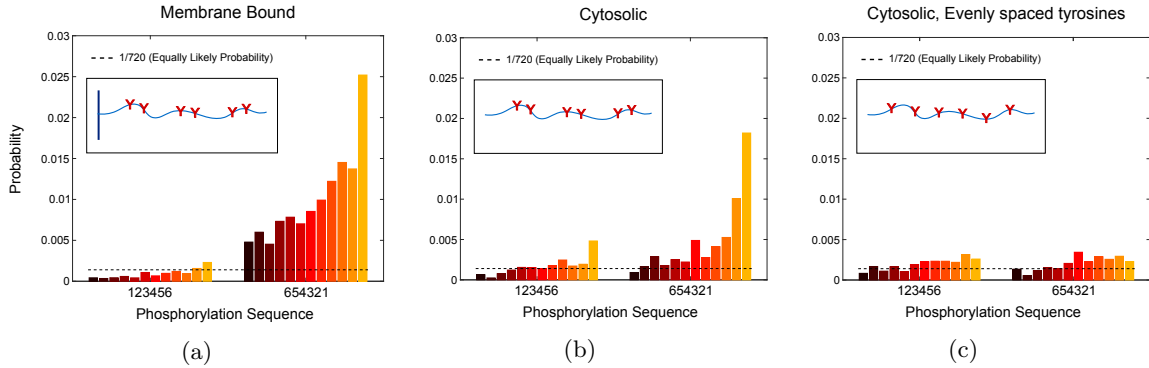


Figure 10: Probability of phosphorylating membrane proximal to distal (123456) compared with distal to proximal (654321). Equally likely probability of all paths indicated with black, dotted line. (a) CD3 ζ parameters, with membrane. (b) CD3 ζ parameters, without membrane. (c) CD3 ζ length, evenly spaced tyrosine locations, without membrane.

4.3.5 Negative cooperativity may arise from local disordering of proteins

We investigate how dephosphorylation of a locally structured protein would impact future dephosphorylation events. Beginning with a model of CD3 ζ where each tyrosine has been locally structured, we investigate how dephosphorylation and consequently, local unstructuring, influence the rate of subsequent dephosphorylations. We find that as the degree of local de/structuring increases, future dephosphorylations become less likely. For moderate de/structuring, this can create up to a 2-fold negative cooperative effect on the dephosphorylation rates from the first to the sixth event. (Fig. 11)

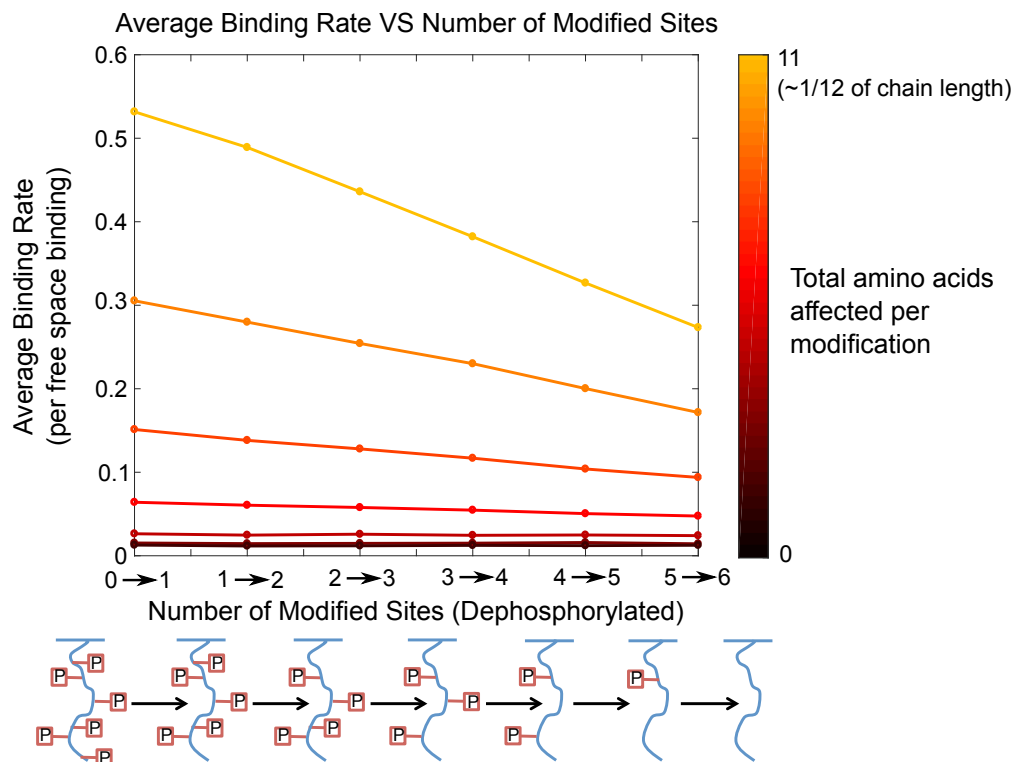


Figure 11: (Above) Simulated average binding rates of Lck to CD3 ζ for different levels of dephosphorylation given a specified degree of local unstructuring (none \rightarrow black, low \rightarrow red, high \rightarrow orange).

4.3.6 Future Work: Quantifying cooperativity from reversible phosphorylation with Hill coefficients

We will create Hill curves of phosphorylation, dephosphorylation, and reversible phosphorylation. This will allow us to better understand and communicate our results with the rest of the biochemistry community. In particular, the Hill curve of reversible phosphorylation will indicate if ultrasensitivity is achievable through this mechanism.

4.3.7 Future Work: Exploring sequence of binding from dephosphorylation and reversible phosphorylation

We will continue to investigate if phosphorylation, dephosphorylation or reversible phosphorylation naturally impose a sequence of binding. This will elucidate whether sequential binding occurs when both phosphorylation and dephosphorylation are impacting the polymer simultaneously.

4.4 Discussion

When binding to disordered proteins, ligands experience significant steric occlusion from their binding sites. We show that this depends on disordered protein length, ligand size, binding site location, and presence of a surface. For disordered proteins with multiple binding sites, disorder leads to global interactions between these sites giving rise to nonlinear effects: positive cooperativity due to disorder-to-order transitions, negative cooperativity from order-to-disorder transitions, and emergent sequential binding. This provides a method of achieving nonlinear signaling behavior from a protein with no native structure.

If we assume the disordered protein undergoes disordered-to-order transitions upon PTM, we find a significant enhancement of the binding rate. We find that for a simulated TCR CD3 ζ chain, there is an enhancement of phosphorylation rate that can help explain our previously published experimental data [Mukhopadhyay2016]. There are several examples of disordered-to-order transitions upon PTM (cite cite cite). It is possible these systems could exhibit this same rate enhancement. Positive cooperativity has been extensively studied both theoretically and experimentally (cite cite cite cite). It gives rise to many signaling behaviors, including sensitivity to ligand concentration.

Conversely, if we assume a protein undergoes an order-to-disorder transition upon PTM (e.g. dephosphorylation), we find a reduction of the binding rate. If the rate reduction upon disorder (e.g. upon dephosphorylation) were to match the rate enhancement from local structuring (e.g. upon phosphorylation), it is in principle possible to negate the nonlinear behavior displayed above. When we simulate dephosphorylation of the TCR CD3 ζ chain, we find only a low reduction of the binding rate compared to the enhancement from phosphorylation. Therefore, we expect to maintain the nonlinear signaling behavior induced by disordered-to-ordered transitions.

Sequential binding arises even when we assume all binding sites are identical. This phenomenon occurs without disordered-to-ordered transitions. Part of this sequential preference is due to the locations of the binding sites along the polymer. Even small differences in binding probabilities can lead to large differences in the probability of a particular sequence. For membrane bound disordered proteins, part of the emergence of sequential binding is due to high preference to bind (membrane-) distal-to-proximal compared to proximal-to-distal. Sequential binding is another phenomenon which has been extensively theoretically studied (cite cite cite) and has been shown to create nonlinear signaling behavior, such as ultrasensitivity. Note that theoretical studies (cite) focus on ‘obligate’ sequential binding, where each event is required for the next to occur. 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.

5 Electrostatics

5.1 Introduction

Subunits of many receptors have been shown to associate with the membrane prior to signaling. [Xu2008],[Shi2013],[Zhang2011] Close association with the membrane inhibits ligands from binding to their target site. Several of these chains have been shown to dissociate from the membrane after the extracellular receptor receives a stimulus. cite cite This mechanism may help prevent unwanted signaling. wording here is mediocre and possibly incorrect?

In particular, (CITE CITE CITE)CD3 ζ associates with the membrane prior to phosphorylation, sequestering the tyrosines in the bilayer. [Aivazian2000],[Zhang2011],[Shi2013] Post-phosphorylation, CD3 ζ is no longer associated with the membrane, but remains anchored by its transmembrane region. [Aivazian2000]

Stimulating the TCR with an antigen will cause the CD3 ζ chain to dissociate from the membrane. However, mutation of the tyrosines to phenylalanine to prevent phosphorylation abrogates this effect. Thus, phosphorylation is required to for ζ chain dissociation from the membrane even with TCR triggering. [Zhang2011]

Basic residue regions in the CD3 ζ and ϵ chains play a major role in maintaining the association between the polymer and the membrane. Studies show that mutation of the basic residue regions is sufficient to cause the protein to dissociate from the membrane.[Zhang2011] Fully phosphorylated tyrosines are also sufficient to pull the protein away from the membrane, even when basic residue regions are present. cite

whoa whoa. Zhang et al claims mutating BRSs actually inhibits phos??? And only associated in 25%?

Calcium influx causes the CD3 ζ chain to dissociate from the membrane prior to phosphorylation. An initial signal causes a calcium dump, creating a positive feedback onto the signaling pathway as more ζ -chains dissociate from the membrane and become phosphorylated, enhancing the signal. [Shi2013] Prior to the calcium influx, the ζ -chain tyrosines remain sequestered in the membrane, preventing signaling. Where the initial signal comes from to instigate the feedback loop remains mysterious.

There are two hypotheses which could explain the initial signaling event. One hypothesis is that a first chain will fall out of the membrane due to steric occlusion created by TCR clustering. When TCR triggering occurs, TCRs cluster. CITE With all six disordered subunits associated with the membrane, when the TCRs cluster, the subunits may become too crowded on the membrane and induce one or more to dissociate and become accessible to phosphorylation.

Independently or in conjunction with the above hypothesis, we formulate a second hypothesis. Although the tyrosines will spend much of their time sequestered in the membrane, there is some probability of entropic forces ehhh transiently pulling them to the cytoplasm. When this occurs, Lck would be able to phosphorylate the tyrosine, creating a large negative charge on the chain. This phosphotyrosine could repel the negative polar heads of the membrane, making it more favorable for that section of the polymer to pull out of the membrane.

We first explore this second hypothesis. We hypothesize that entropic fluctuations of the tyrosines in and out of the membrane are sufficient to lead to an initial phosphorylation event. We also hypothesis that following this initial event, there would be a cooperative enhancement of the accessibility of neighboring tyrosines.

5.2 Model and Methods

A molecular dynamics simulation of CD3 ϵ indicates an approximately Gaussian distribution of the tyrosines, centered at the phospholipid heads of the membrane. [Lopez2015]

We model the polymer-membrane association as a potential acting on each rod in a freely-jointed chain in half space. The potential only acts in the direction of the half space plane (in this case, z-direction). To develop the model, we need to explore parameter space to create potentials which display the same behaviors seen experimentally and through molecular dynamics. In particular, we want to match three conditions:

1. The polymer should display locational distributions consistent with Lopez et al. 2015. [Lopez2015]
2. When basic residues are mutated, the polymer dissociates from the membrane. [Zhang2011]
3. When all tyrosines are phosphorylated, the polymer should dissociate from the membrane.

Based on these goals, we group residues into four distinct categories: tyrosines, phosphotyrosines, basic residues, and remaining residues. Based on these groups, we develop a set of possible relationships to explore (Table).

Table 1: Notation for electrostatic potentials applied to different groups of residues.

Residue Group	Electrostatic Potential Abbreviation
Tyrosines	E_Y
Phosphotyrosines	E_P
Basic Residues	E_B
Remaining Residies	E_R

Table 2: Electrostatic potential relationships to explore for behavior matching experimental results.

Group Number	Electrostatic Potential Relationship
1	$E_Y = E_B < 0$ $E_P = E_R = 0$
2	$E_Y \neq E_B$ $E_Y < 0$ $E_B < 0$ $E_P = E_R = 0$
3	$E_Y \neq E_B$ $E_Y < 0$ $E_B < 0$ $E_P > 0$ $E_R = 0$
4	$E_Y = E_B < 0$ $E_P > 0$ $E_R = 0$
5	$E_B < 0$ $E_Y = E_R = 0$ $E_P > 0$

Table 3: Cytoplasmic amino acid sequence of CD3 ζ and ϵ chains. Basic residues (arginine, lysine, histidine) and tyrosines labeled in sequence (red). Relative numeric location in cytoplasmic sequence and fraction of residues from that group are noted.

Chain	Group	Location in Cytoplasmic Sequence	Numeric Location	# / Total
CD3 ζ	Basic Residues	RAKFSRSAETAANLQDPNQLYNE LNLGRREEYDVLEKKRARDPENMG GKQQRRRNPPQEGVYNALQKDKM AEAYSEIGTKGERRRRGKGHDGLY QGLSTATKDTYDALHMQTLAPR	1,3,5,28,29,37,38, 39,41,48,51,52,53,60,65, 67,72,78,81,82,83,85,91, 99,102,106,113	29/113
	Tyrosines	RAKFSRSAETAANLQDPNQLYNE LNLGRREEYDVLEKKRARDPENMG GKQQRRRNPPQEGVYNALQKDKM AEA YSEIGTKGERRRRGKGHDGLY QGLSTATKDTYDALHMQTLAPR	21,32,60,72,91,102	6/113
CD3 ϵ	Basics Residues	WSKNRKAKAKPVTRGAGAGGRQ RGQNKERPPPVPNPDIPIRKGG RDLYSGLNQRRRI	3,5,6,8,10,14,21,23,27, 29,42,43,46,55,56	15/57
	Tyrosines	WSKNRKAKAKPVTRGAGAGGRQ RGQNKERPPPVPNPDIPIRKGG RDL YSGLNQRRRI	38,49	2/57

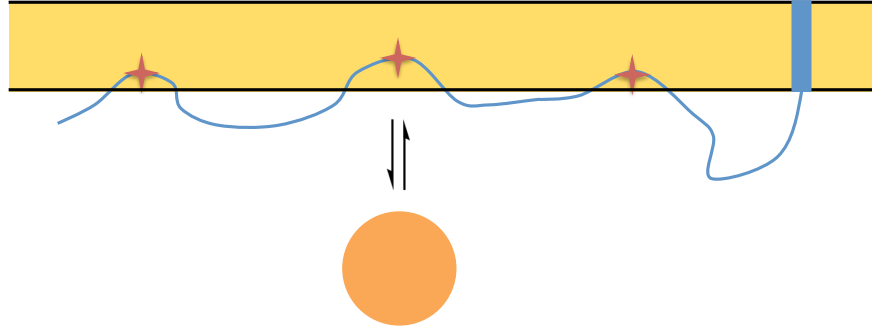


Figure 12: Cartoon of electrostatics model. FJC associates with membrane, burying modification sites within membrane rendering them inaccessible to ligands.

5.2.1 Electrostatic Potentials

The model where basic residues and tyrosines have similar energetics (Table 3, Group 1) is the simplest case to explore. There are many types of electric potential we could explore for the correct behavior. Given that one of the phenomena we wish to match is an approximately Gaussian distribution of both the tyrosines and polymer along the membrane edge, we use a parabolic-constant piecewise potential for our basic residues and tyrosines, instead of a Lennard-Jones potential.

Parabolic-constant piecewise potential:

Piecewise potential (aka, parabola-constant), (*PC*):

$$\begin{cases} width * z^2 - depth & z < \sqrt{\frac{depth}{width}} \\ 0 & z \geq \sqrt{\frac{depth}{width}} \end{cases} \quad (19)$$

We have two possible potentials for the remaining amino acids. In general, the peptide should not be able to penetrate deep into the membrane. Therefore, we can implement a hardwall or softwall constraint. The hardwall will prevent the remaining amino acids from passing the beginning of the membrane. A softwall constraint will allow amino acids to enter the membrane, but will incur an energetic penalty. For this initial model, phosphotyrosines will follow this potential also.

Hardwall:

$$\begin{cases} \infty & z < 0 \\ 0 & z \geq 0 \end{cases} \quad (20)$$

Softwall:

$$\begin{cases} k * z^2 & z < 0 \\ 0 & z \geq 0 \end{cases} \quad (21)$$

We later develop a more complicated but possibly more motivated model with a repulsive force on the phosphotyrosines. There are many more basic residues compared to tyrosines in both the CD3 ζ and ϵ chains. Therefore, in order to account for tyrosine phosphorylation being sufficient to dissociate the polymer from the membrane, we would expect a repulsive force from the phosphorylated tyrosines. We therefore shift future focus to models of Groups 3,4, and 5. We first consider the simple case where tyrosines do not experience their own potential but experience the potential of either the basic residues or experience the same potential as the rest of the amino acids. Since tyrosines are not positively charged, it seems less likely that they would experience the same force as the basic residues. Although their aromatic ring structure may help to anchor the polymer to the membrane once associated with the membrane, it would not drive the polymer to the membrane to begin with. [Lopez2015] Therefore we focus future efforts on Group 5.

In this new model, we introduce a repulsive potential that acts only on the phosphotyrosines. We maintain the same potentials as above for the basic residues and the general amino acids, but now the tyrosines behave as generic amino acids. As our repulsive potential, we will include an exponential distribution where the repulsive effect of the membrane drops off quickly as the phosphotyrosines move further away from the membrane. Eq. 22

Exponential decay:

$$\begin{cases} \infty & z \leq 0 \\ e^{-z} & z > 0 \end{cases} \quad (22)$$

5.3 Results

5.3.1 Model in which basic residues and tyrosines have similar energetics reveals puzzle in existing data

We first explore the simplest model, where basic residues and tyrosines feel the same piecewise parabolic-constant potential and the phosphotyrosines feel the same softwall (or hardwall) potential as the rest of the residues. (Table 2, Group 1) We first explore parameter space to find a set of parameters where the distribution of tyrosines is roughly gaussian centered at the membrane and extending about two Kuhn lengths in each direction, as seen in Lopez et al. 2015. (Condition 1) From this sweep, we see there are multiple parameter sets which achieve this distribution. Fig. ?? For the following exploration of how phosphorylation impacts the distribution, we show the minimal parameters of those tested which give the desired distribution. We establish the following potentials:

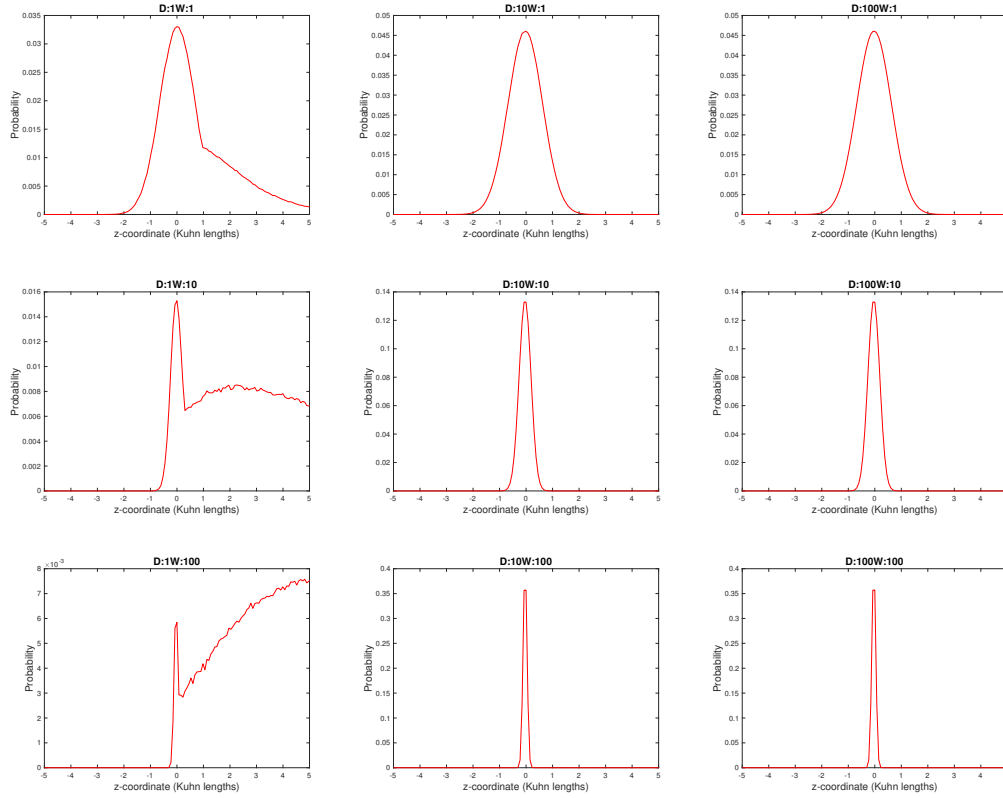


Figure 13: Polymer tip distribution resulting from parameter exploration of electric potentials. For this set, softwall potential width is set at 0.01. Both parabolic-constant potential depth and potential width range from 10^0 to 10^2 .??

Parabolic-constant piecewise potential:

$$\begin{cases} 1 * z^2 - 10 & z < \sqrt{\frac{10}{1}} \\ 0 & z \geq \sqrt{\frac{10}{1}} \end{cases} \quad (23)$$

Softwall:

$$\begin{cases} 0.01 * z^2 & z < 0 \\ 0 & z \geq 0 \end{cases} \quad (24)$$

We explore how phosphorylation, or removing tyrosines from the parabolic-constant potential and adding them to the softwall potential, impacts the distribution of the tyrosines and full chain. We find that under this regime, phosphorylating individual tyrosines does not impact the distribution of neighboring tyrosines. Fig. 14 Additionally, phosphorylating two tyrosines does not influence the distribution of a tyrosines located between the two. Fig. 15 We also find that even when all tyrosines are phosphorylated, the polymer does not fully dissociate from the membrane. This is represented by the distribution of the (basic) polymer tip residue distribution, which remains closely associated with the membrane. A non-basic, non-tyrosine residue located centrally is only minorly affected by the change in phosphorylation state. This result indicates these electric potentials may be insufficient to exhibit the characteristics displayed experimentally.

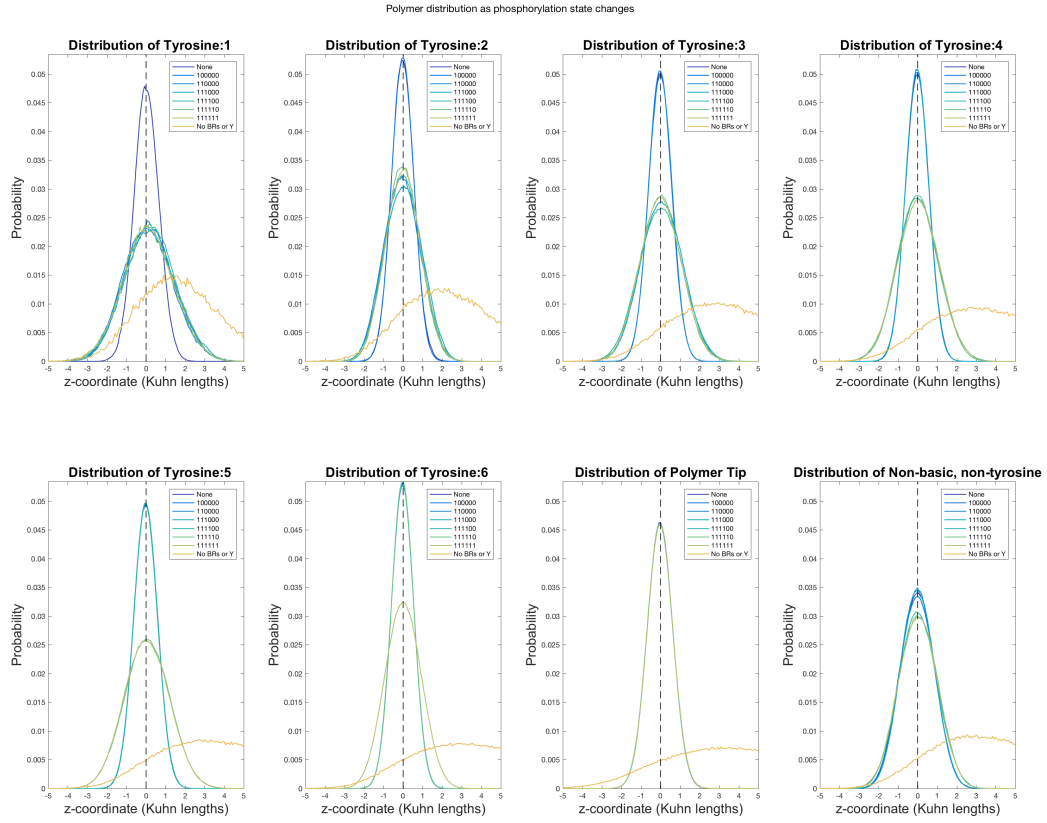


Figure 14: Distribution of tyrosines, tail residue, and single non-basic residue for different phosphorylation states. Each curve has the phosphorylation state denoted as a series of 0's and 1's, with the tyrosines in order membrane-proximal to distal. Tyrosines are phosphorylated sequentially from proximal to distal. Tyrosines are indicated by 0, phosphotyrosines indicated by 1. Each distribution is plotted against the distribution of only the softwall potential. (No BRs, no Y)

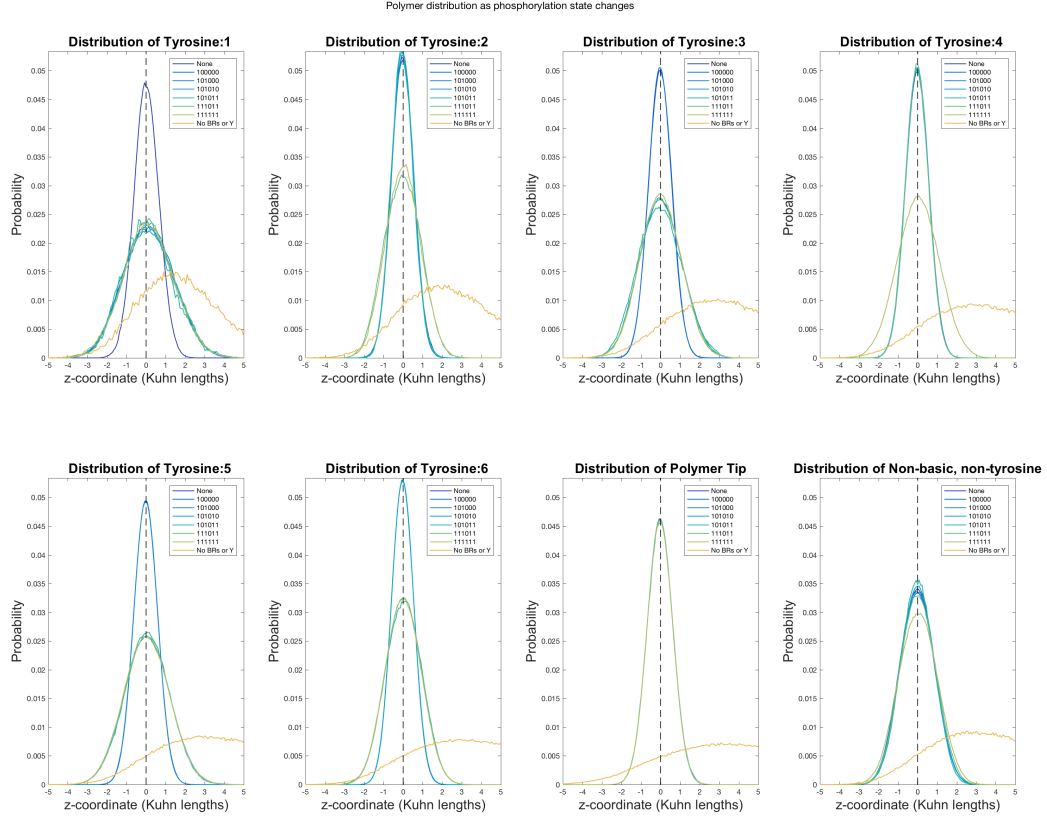


Figure 15: Distribution of tyrosines, tail residue, and single non-basic residue for different phosphorylation states. Each curve has the phosphorylation state denoted as a series of 0's and 1's, with the tyrosines in order membrane-proximal to distal. Tyrosines are indicated by 0, phosphotyrosines indicated by 1. Phosphorylation sequence disordered to examine effect of two phosphorylation events on the distribution of an intermediate residue. Each distribution is plotted against the distribution of only the softwall potential. (No BRs, no Y)

5.3.2 Initial parameter exploration of more general energetic model yields fit to polymer distribution

We first want to match the distribution of the polymer to the molecular dynamics simulations from Lopez et al. We sweep through parameters for the basic residue potential and the softwall potential. The softwall potential permits a Gaussian curve for the tyrosine distributions. The hardwall condition (not shown) prevents the distributions from spreading below the zero axis where we have defined the membrane to be. When we explore these parameters, multiple possibilities for parameter sets emerge which meet the distribution conditions. Below are examples of distributions arising from our parameter exploration for both a tyrosine and a basic residue. Fig. 16, Fig. 17 We note that the tyrosines have a wider distribution than the basic residues since they do not directly feel the attractive potential. From these parameter sets, we may refine our parameter search and begin exploring how the distributions are affected when all tyrosines are phosphorylated.

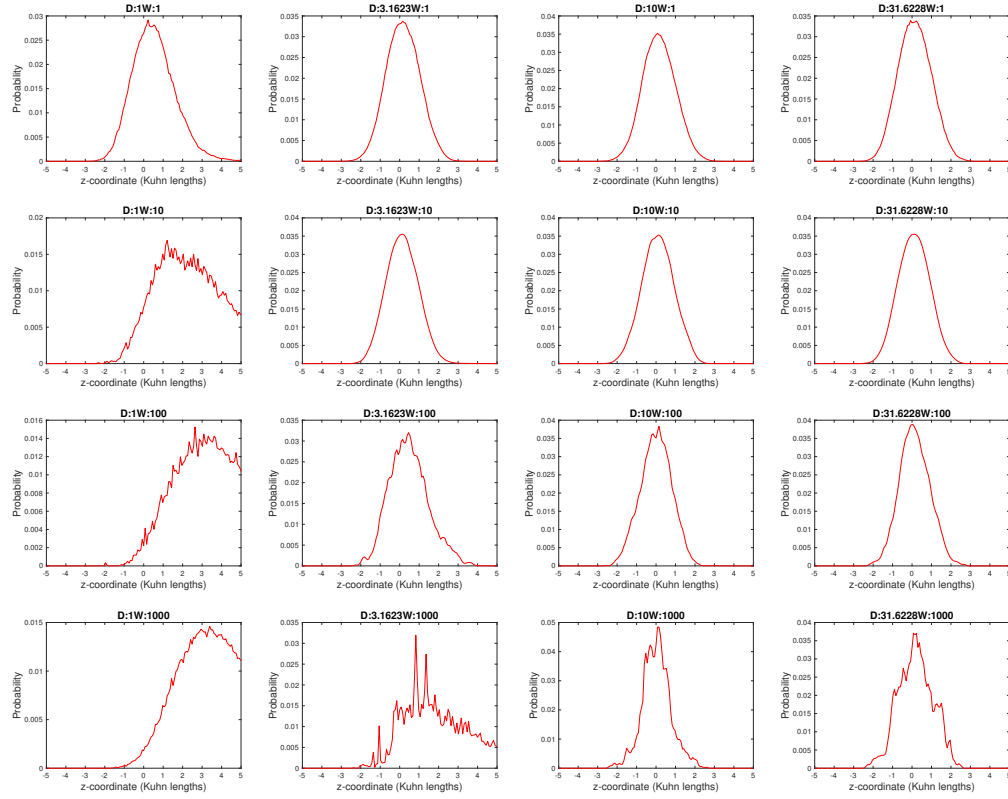


Figure 16: Tyrosine distribution resulting from parameter exploration of electric potentials. For this set, softwall potential is set at 0.1. Parabolic-constant potential depth ranges from 10^0 to $10^{1.5}$, width ranges from 10^0 to 10^3 .

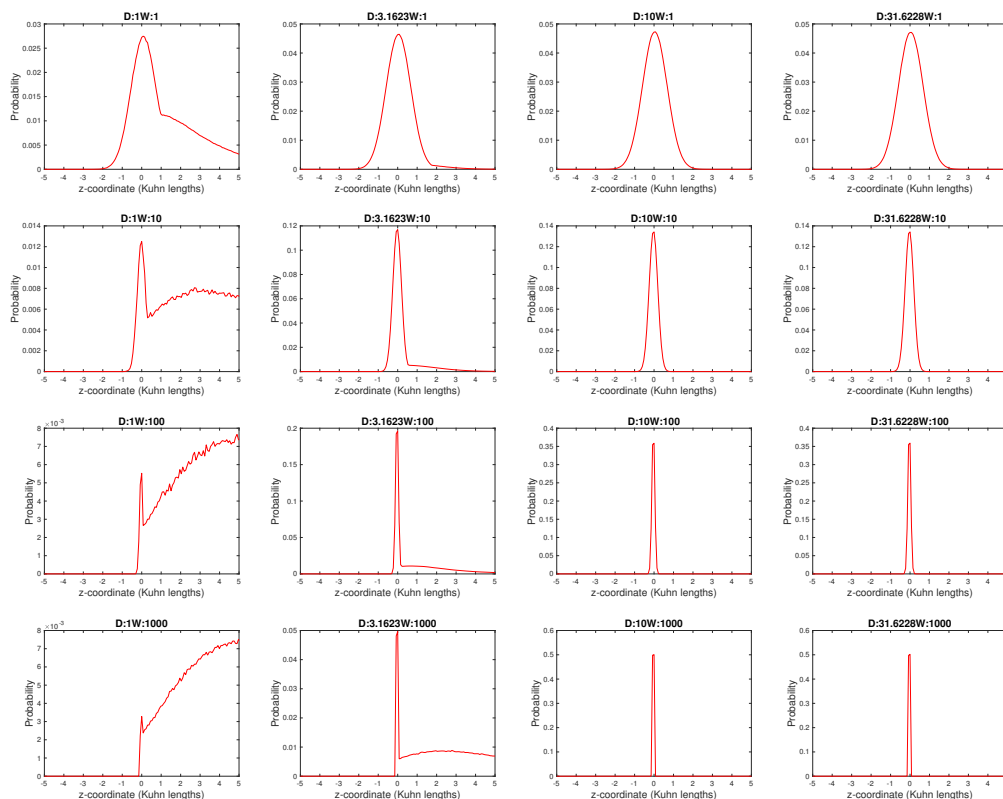


Figure 17: Polymer tip (basic residue) distribution resulting from parameter exploration of electric potentials. For this set, softwall potential is set at 0.1. Parabolic-constant potential depth ranges from 10^0 to $10^{1.5}$, width ranges from 10^0 to 10^3 .

5.3.3 Future Work

First, we want to establish parameters (if they exist) which make our model exhibit the same behaviors as experimental and MD results. In particular, we now need parameters allowing the polymer to be free from the membrane when all tyrosines are phosphorylated. Either there will be only one reasonable parameter set which we can explore or there will be many parameter sets which display the desired properties. If there are multiple parameter sets then we may try to find other experimental results to narrow our parameter regime. Alternatively, we may explore two or three major parameter regimes: weakest, strongest, and median potentials. The results from a wide spread of parameters will give us an indication of how much the parameters matter to the qualitative results.

Second, with our parameter sets, we will now explore the effect of phosphorylation on the accessibility of successive tyrosines to a kinase. We wish to see if there is a cooperative enhancement of the binding rates based on phosphorylation of previous residues. Cooperative phosphorylation could contribute to the first signaling event, explaining how an initial calcium release feeds back on the system, causing more CD3 ζ chains to dissociate from the membrane and create a stronger signal. [Shi2013]

Additionally, we will use this data to explore if there is a natural sequential binding sequence that arises. One would imagine that once a single tyrosine is phosphorylated then its nearest neighbor would be the most likely next target. However it is unclear if there is a first tyrosine that will be most likely to be phosphorylated and whether this will depend dominantly on basic residue distribution or on proximity to the transmembrane region.

We will also explore how clustering of TCRs influences the association of the disordered subunits to the membrane. By modeling multiple disordered chains, we will attempt to estimate how many TCRs must be in a cluster and how tightly clustered they must be for this model to explain the initial triggering event.

6 Results: Simultaneous Binding

6.1 Introduction

Here we consider the case where ligands are able to bind to a polymer and remain bound. For instance, the SH2 domains of ZAP-70 bind to two tyrosines in an ITAM on CD3 ζ and remain bound to phosphorylate downstream.[Wang2010] This leads to the possibility that the polymer conformations are limited by the presence of a bound ligand, or even multiple bound ligands. We expect that as more ligands are added to the polymer, binding site occlusion will increase. The ligand may help to straighten out the polymer by limiting the number of conformations it can take on, but it will also occlude the binding site in many conformations. Depending on the size, number, and relative location of the ligand to the binding site, this effect may be stronger than the decreased occlusion via straightening.

Once TCR CD3 ζ is phosphorylated, the kinase ZAP-70 binds and phosphorylates the transmembrane protein LAT. [Chan1992],[Zhang1998] ZAP-70 is composed of two SH2 domains and a kinase domain.[Chan1992] [Wang2010] These tandem SH2 domains each bind a phosphotyrosine on CD3 ζ . Since minimally two SH2 domains must be able to bind to CD3 ζ , it is natural to wonder how many domains could fit on a disordered chain and how each successive domain impacts the binding of another.

In our simulation, we are able to explore different sizes of bound ligand and different sizes of unbound (incoming) ligand. We explore two scenarios. First, we explore the case where one species binds and remains bound to the disordered protein while a second species tries to transiently bind. For instance, as ZAP-70 binds to the disordered protein, how does this influence further phosphorylation by Lck. Second, we explore the case where a single species binds and remains bound (i.e. bound and unbound species are identical). This will answer how each bound ZAP-70 impacts binding by the next ZAP-70. This is also generalizable to any system where simultaneous binding occurs. For both of these scenarios, we will determine if simultaneous binding of ligands helps or hinders future binding events.

6.2 Model and Methods

We model the TCR CD3 ζ chain as a freely jointed chain, where each amino acid is represented as a segment as described in Sec. 3. The unbound, incoming ligand is simulated as an idealized ghost sphere located tangentially to a single segment. We calculate quasi-equilibrium statistics for the FJC and unbound ligand in both free-space and half-space. For the CD3 ζ chain, we simulate only the cytoplasmic region by an FJC with 113 segments. In our simulation, all lengths are normalized by the Kuhn length, so each rod has length 1 and the radius of the ligand is also measured in Kuhn lengths. We consider a spherical estimate for the kinase Lck, with a radius of 2.1 nm(7 Kuhn lengths) as calculated in 3.2.

To simulate simultaneous binding, each bound ligand is included as an idealized sphere tangentially bound to the polymer. The bound ligands rotate with the polymer, but maintain their tangent orientation to their binding site. The polymer 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 ligand cannot penetrate the membrane. There are separate parameters for the radius of the bound and unbound ligands.

We calculate the probability a ligand is able to bind to the unbound ligand site, where ‘able to bind’ refers to the specified sphere being empty of both other polymer segments, half-space barriers, and bound ligands.

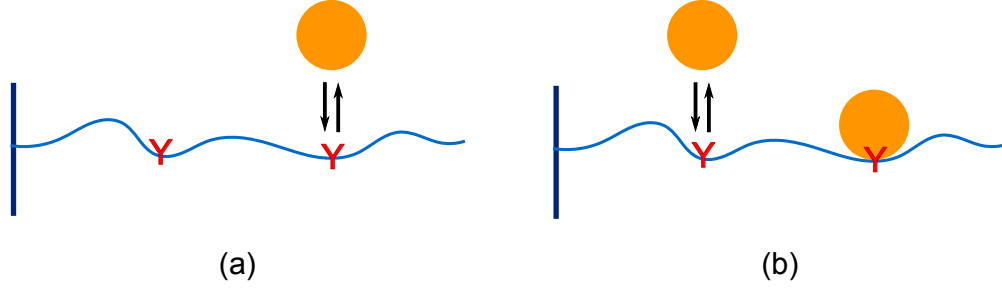


Figure 18: Cartoon of simultaneous binding model. Surface bound FJC prior to ligand binding (a) transitions to FJC with single bound ligand and a second ligand attempting to bind (b).

6.3 Results

6.3.1 Negative cooperativity arises from simultaneous binding of ligands to an IDP

We first explore how two ligand species binding to the same domain may influence each other's binding. We focus on a simplification of ZAP-70 and Lck molecules binding to CD3 ζ . We assume that each of the bound ligands is approximately the size of an SH2 domain (i.e. ZAP-70 is bound to the polymer) and we calculate the probability of a separate species binding to one of the remaining binding sites. We investigate this for many sizes of incoming ligand, including our estimates for Lck and ZAP-70. We find simultaneous binding of multiple SH2 sized domains will cause mild to severe negative cooperativity with the binding of another ligand. Larger ligands have a harder time binding when there are multiple SH2 domains already bound. For a 4.2nm radius incoming ligand, there would be a 30 fold decrease in the probability of binding from no other bound ligands to five SH2 domains bound to the same disordered chain. Only very small ligands would be unhampered by the presence of other bound ligands. An Lck molecule, only slightly larger than an SH2 domain, would experience an approximately eight fold decrease in its ability to bind once five SH2 domains are already bound to the disordered chain. (Fig. 19)

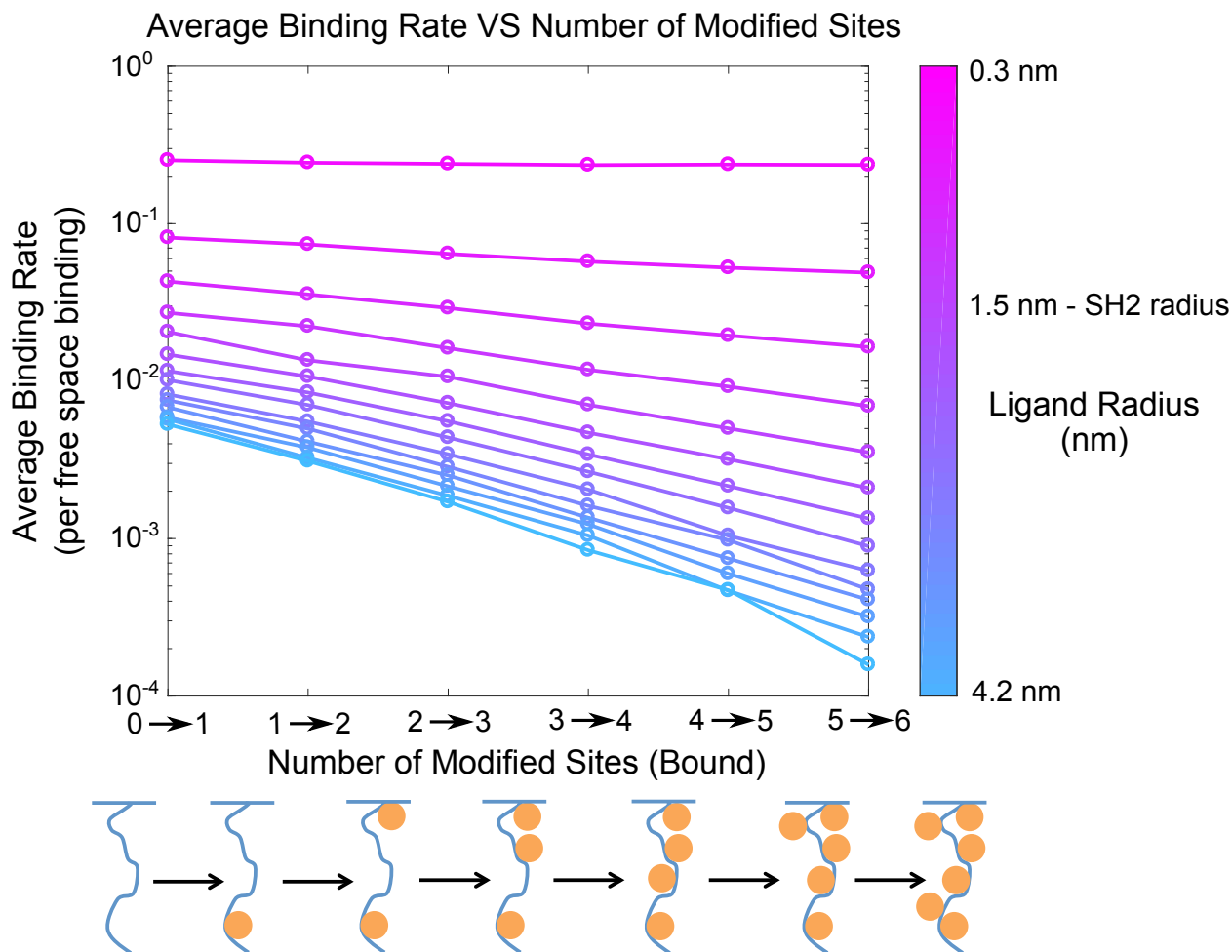


Figure 19: (Above) Simulated average binding rates of Lck to CD3 ζ for specified number of SH2 domain sized (1.5nm radius) ligands previously bound and varying unbound ligand sizes (0.3 nm \rightarrow pink, 4.2 nm \rightarrow blue). (Below) Cartoon representing a possible bound configuration series of CD3 ζ .

The second case of simultaneous binding we explore is when only one ligand species attaches to the disordered chain. We investigate how multiple bound ligands of a given size impacts another ligand of the same size binding. We see the same trend of behavior as above, but more dramatic. Here, if we consider a ligand with radius 2.1nm, we already see a 30 fold decrease between no ligands bound and five bound. This matches intuition, since although a large bound ligand will straighten the polymer more, it will also create a much larger occluded volume making binding site accessibility much less likely. Note that similarly, if the bound ligand size is less than that of an SH2 domain, then we actually see less negative cooperativity than when all bound ligands were SH2 domains. (Fig. 20)

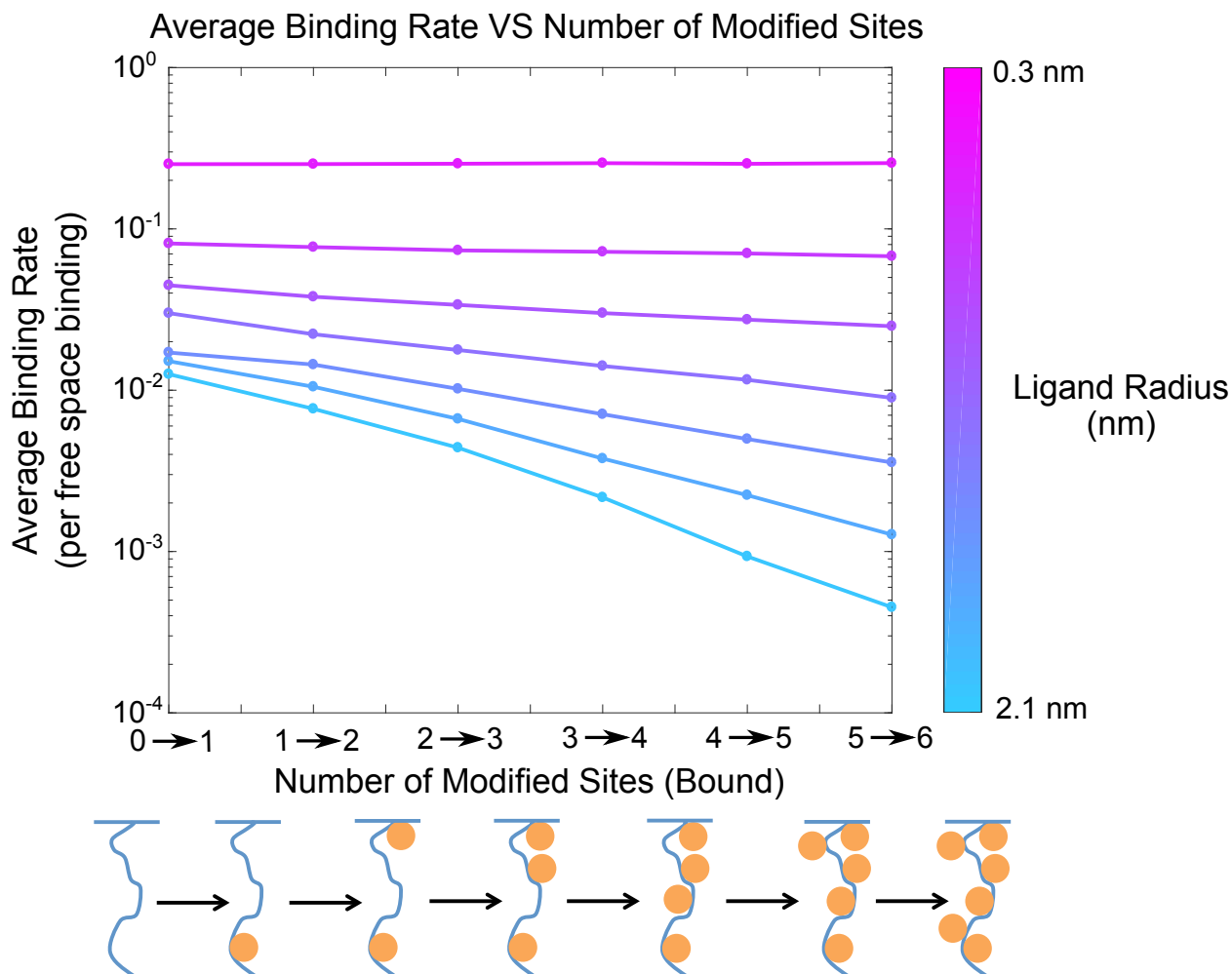


Figure 20: (Above) Simulated average binding rates of Lck to CD3 ζ for specified number of ligands previously bound and varying ligand sizes (0.3 nm \rightarrow pink, 4.2 nm \rightarrow blue). (Below) Cartoon representing a possible bound configuration series of CD3 ζ .

6.3.2 Future Work: Sequence of binding multiple ligands

We obtained relative probabilities of binding for each tyrosine along the chain at each bound state. With this information, we are able to explore whether or not there is a dominant sequence of binding events. We ran a Gillespie algorithm with six events, one for each ligand binding. From each run, we record which sequence or 'path' of binding. When we compile the path frequency data, it becomes clear that the first binding event is dominant. That is to say, any path beginning by binding to the membrane distal tyrosine (6th tyrosine), will occur more often than any path beginning with the membrane proximal tyrosine.

For simplicity, we consider only the paths membrane proximal to membrane distal (123456) and distal to proximal (654321). We see that when a membrane is present, the probability of binding distal to proximal is much higher than that of phosphorylating proximal to distal. This phenomenon is easily explained by the presence of a membrane. The membrane proximal tyrosine has a smaller range of space it may occupy than the distal tyrosine. Therefore, it is more likely to be close to the membrane in a configuration. Since the ligand cannot penetrate the membrane, tyrosines closer to the membrane have a higher probability of being effectively sheltered from ligand binding. This makes the distal to proximal path much more likely.

The preference for distal to proximal binding is enhanced by increasing ligand size. Larger ligands remain able to bind the membrane distal tyrosine with similar probability but will have a lower probability of binding tyrosines close to the membrane. When we simulate CD3 ζ without a membrane, this preference is eliminated. Note that unlike with the local structuring model, a strict comparison of distal to proximal vs proximal to distal binding does not indicate a preference based on uneven spacing of tyrosines and in fact both of these paths are less likely than if all paths were equally probable. Although the most likely tyrosine to be bound is still the distal tyrosine, there is not a strong preference for the distal to proximal sequence because perfectly sequential binding is unlikely. Each bound ligand is more likely to sterically occlude a neighboring tyrosine than one further away. Therefore, if the distal tyrosine is the first bound, the next most likely to be bound is not its neighbor.

We are soon going to introduce a ranking system for the paths in order to do a full analysis of all possible paths. We look to answer if, for simultaneous binding, there a single path or group of paths that stand out as most likely.

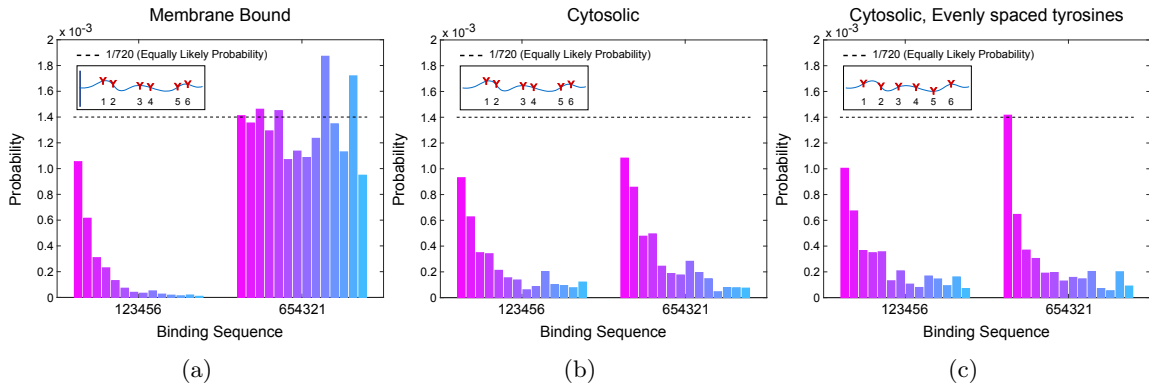


Figure 21: Probability of ligand binding membrane proximal to distal (123456) compared with distal to proximal (654321). Equally likely probability of all paths indicated with black, dotted line. (a) CD3 ζ parameters, with membrane. (b) CD3 ζ parameters, without membrane. (c) CD3 ζ length, evenly spaced tyrosine locations, without membrane.

6.3.3 Future Work: Multiple chains towards a full model of a T cell receptor signalosome

There is work showing that six ZAP-70 molecules bind to the disordered proteins of the T Cell Receptor on average. [ODonoghue2013] We want to eventually report an equilibrium value for the number of molecules able to bind to a CD3 ζ chain. We will also expand our model to include all of the domains of the receptor. This will allow us to test if steric effects are sufficient to explain why only six ZAP-70 molecules would bind to the receptor simultaneously. When all of the chains are included in the simulation, each chain will be able to explicitly occlude binding sites on other domains. Experimental work on CD3 ζ and other receptors is often completed using dimers. [Mukhopadhyay2016] [Dobbins2016] Simulating multiple domains will allow us to better relate our results to the experimental results.

Eventually, we would like to include co-receptors of TCR in our model. This will help expand our understanding of the TCR network, in particular elucidating why co-receptors are necessary for full T cell signaling, e.g. CD4 and CD8. There is evidence that Lck is associated with membrane-bound proteins CD4 and CD8 [Barber1989],[Veillette1988],[Rudd1988]. There is also evidence that the phosphatases involved with TCR triggering, CD45 and CD148, are also membrane bound. [Davis2006] These changes to the model will give finer detail to our understanding of the dynamics of reversible phosphorylation of the CD3 ζ chain.

6.4 Discussion

An Lck molecule, only slightly larger than an SH2 domain, would experience an approximately eight fold decrease in its ability to bind once five SH2 domains are already bound to the disordered chain. This suggests the CD3 ζ chain is mostly phosphorylated before ZAP-70 begins to attach or that phosphorylation by Lck slows down when ZAP-70 attaches. In the second case, this could lead to phosphorylation needing to occur via another mechanism, for instance via the kinase domain of ZAP-70 itself.

Simultaneous binding of ligands leads to negative cooperativity. This endows 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. Similarly, this might reduce the impact of inhibitors, requiring much higher concentrations of inhibitor to completely turn off signaling.

7 Role of the surface on tether-tether reactions within a signaling cluster

7.1 Introduction

7.1.1 Tethering

Disordered proteins aid many reactions by acting as biological tethers. They may be permanent tethers tying two globular domains together or a domain to a surface [example, example, cite cite](#). They may also be transient tethers, catching diffusing particles and tethering them to another molecule, i.e. formin. [cite](#) By tethering molecules or surfaces together, disordered domains effectively increase the local concentration of a particle. This increases reaction rates by improving the probability of two molecules bumping into each other through restriction of their diffusive domain.

Given that a surface reduces the exploration space of a molecule, it will naturally increase the local concentration without a tether. However, it will also reduce the accessibility of a molecule by sometimes occluding, for example, an enzymatic domain from reaching its substrate in the correct orientation. Given that the surface has both positive and negative effects on reaction rates, how does being tethered to a surface impact a reaction compared to being tethered to a smaller domain?

This type of interaction arises frequently with clustered reactions. For instance, the phosphorylation of a membrane tethered disordered domain by a membrane bound kinase. [cite](#) Experimental studies of such reactions are more easily performed on a sparse matrix such as dextran than on a hard surface. Therefore, it is important to understand the effect tethering to a surface has over tethering to a matrix. We explore how the presence of a surface influences the reach of a tethered domain and the overall local concentration.

7.1.2 SHP-1

7.2 Model and Methods

We first investigate the effects of a surface on a single tether without a bound ligand in free-space and half-space. We model the tether as a freely-jointed chain and the unbound ligand as an idealized ghost sphere located tangential to a single segment. To explore the effects of a surface on a single tether, we vary the number of segments and the size of the incoming ligand. We calculate quasi-equilibrium statistics for the FJC and unbound ligand in both free-space and half-space. We calculate how often a ligand is able to bind

Table 4: Structure and number of simulated segments in each explored tether.

Tether Name	Structure	N
PD-1	(37392 B2) Bio-SRAARGTIGARRTGQPLKEDPSAVPVFSV DYGELDFQWREKTPEPPVPSVPEQTEY*ATIVFPSG	64
PEG-3	Bio-(PEG)3-DLQEVTY*IQLDHH	16
PEG-28	Bio-(PEG)28-DLQEVTY*IQLDHH	41

to an oriented sphere tangentially attached to the polymer, where ‘able to bind’ refers to the specified sphere being empty of both other polymer segments and the surface.

We next explore the effects of a surface on three specific tethers: PD-1, PEG-3, and PEG-28. PEG-3 and PEG-28 are tethers created by attaching a single immunoreceptor tyrosine inhibitory motif (ITIM) to a series of PEG linkers. Each tether is modeled as a freely-jointed chain characterized by their lengths as measured in Kuhn lengths. The tethers are made up of PEG linkers and amino acids, each of which contributes approximately 0.3nm to the length. We assume maximal flexibility in our model by assigning a Kuhn length, δ , of 0.3nm. The contour lengths of each tether may be approximated as N Kuhn lengths, where N is the total number of PEG linkers and amino acids. Note that we consider the size of biotin to be negligible for our simulation. (Table 4)

We investigate SHP-1 as a ligand bound to a tether. In order to simulate SHP-1 as a simple sphere, we estimate its size using three different assumptions. We first estimate the volume of SHP-1 based on its molecular weight, 67561Da, as listed in UniProt entry P29350, assuming an average protein density of $1.41\text{g}/\text{cm}^3$. [Fischer2004]

$$(67.5 \times 1000\text{Da}) * (1.66 \times 10^{-27}\text{kg} / \text{Da}) * (1000\text{g} / \text{kg}) / (1.41\text{g} / \text{cm}^3) = 79.468 \text{ nm}^3.$$

Approximating the phosphatase domain as a sphere, we can estimate a radius:

$$\begin{aligned} V &= \frac{4}{3}\pi r^3 \\ 79.468 \text{ nm}^3 &= \frac{4}{3}\pi r^3 \\ r &\approx 2.7 \text{ nm} \end{aligned}$$

For comparison, we also estimate an upper bound on the radius of SHP-1 from PyMol measurements of the SHP-1 structure (PDB 3PS5). Rounding up for measurement error, the longest part of the structure is 86 Å. (Fig. 22) From this we calculate a maximum radius of 4.3nm .

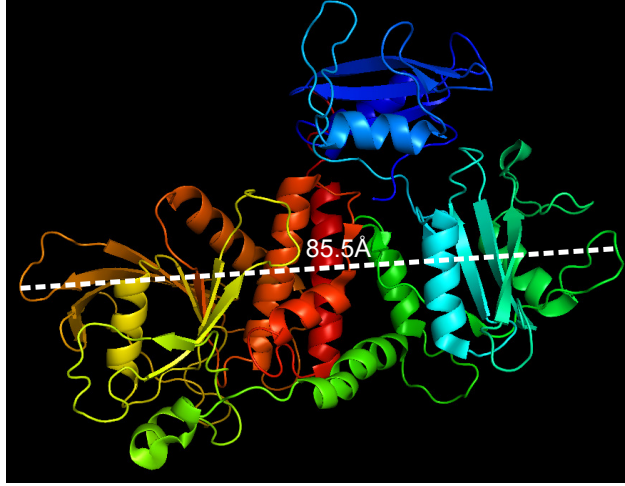


Figure 22: Measurement of maximum length of SHP-1 (PDB 3PS5).

Alternatively, we can estimate the volume of SHP-1 as a rectangular prism and then create a spherical approximation of equal volume. We estimate the length, width and depth of SHP-1 to be 76\AA , 36\AA , and 58\AA respectively, giving a volume of 158688\AA^3 . (Fig. 23) From this, we calculate the radius for a spherical approximation to be 3.4nm .

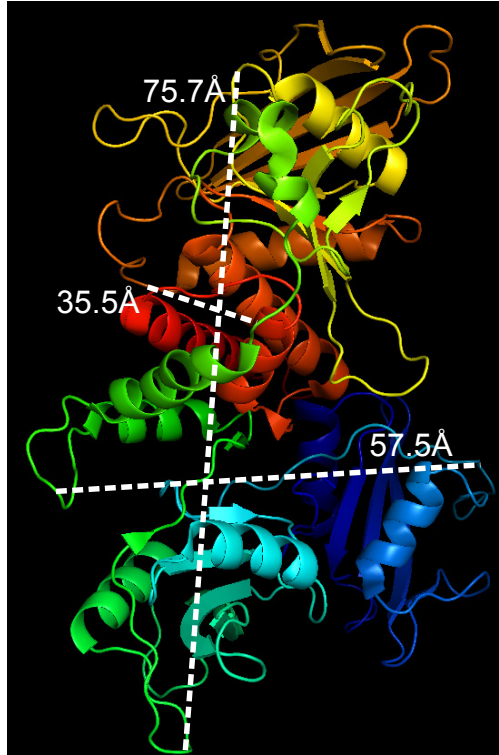


Figure 23: Measurement of rectangular dimensions of SHP-1 (PDB 3PS5).

Using the parameters estimated above, we simulate two equivalent tethers. The idealized SHP1 sphere, or

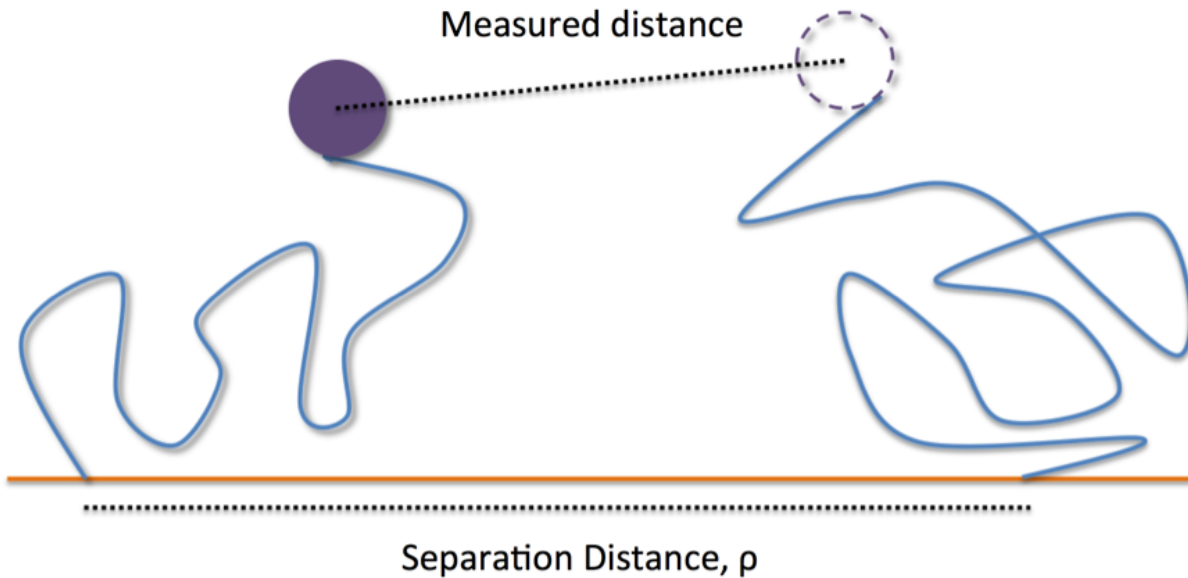


Figure 24: Cartoon of local concentration simulation in half space.

ligand, is bound to the end of one of the tethers. The bound ligand is simulated as an idealized sphere which rotates with the tether. The bound ligand may not occupy the same space as the tether nor may it penetrate the surface if present. The second tether has a ghost ligand attached in the same location. (Fig. 24) We allow each tether to explore conformation space, at each conformation determining if the ghost ligand is occluded by the rest of the tether, or in half-space, by the surface. For the tether with the ghost ligand, we record both the location of the center of the ligand and whether or not it was occluded at each conformation. For the tether with the bound ligand, we only record the location. For these simulations, we assume the two tethers are independent and do not sterically influence each other.

We compute the local concentration as the probability that the bound ligand is near the binding site (where near is defined by within some small radius) and the binding site is not occluded by the rest of the polymer. To calculate this, we separate the tethers from each other by a distance, ρ , measured in Kuhn lengths by adding ρ to the x -coordinate of the ghost ligand. We then measure the distance between the centers of the bound and ghost ligands. If the centers are within a cutoff distance from each other AND the ghost sphere is not occluded, then the bound ligand is able to bind. The local concentration is calculated by dividing the probability the ligand can bind by the volume of the sphere defined by the cutoff radius.

We fit the local concentration curves to a half-gaussian curve, $\sigma_{fit}(\rho) = \sigma_{max} * e^{-\rho^2/l^2}$, as a function of the separation distance between the two tethers. To fit the simulated data, we use Matlab's `fminsearch` function which uses the Nelder-Mead simplex algorithm. The parameters found by the Nelder-Mead algorithm consistently give a smaller sum of least squares than parameters found using Trust-Region-Reflective Least Squares Algorithm or those found by heuristically sweeping parameter space to minimize the sum of least squares.

7.3 Results

7.3.1 Accessibility of a reaction site is reduced by the presence of a surface

We first investigate how the presence of a surface impacts binding of a ligand to a single tether compared to in free-space. We find that a surface consistently net decreases the probability a ligand can bind to a tether, irrespective of tether length or ligand size. For long tethers, the surface reduction is minimal, in some parameters almost eliminating the difference. A larger ligand creates a larger fold decrease in the ability of the kinase to bind in half-space. (Fig. 25)

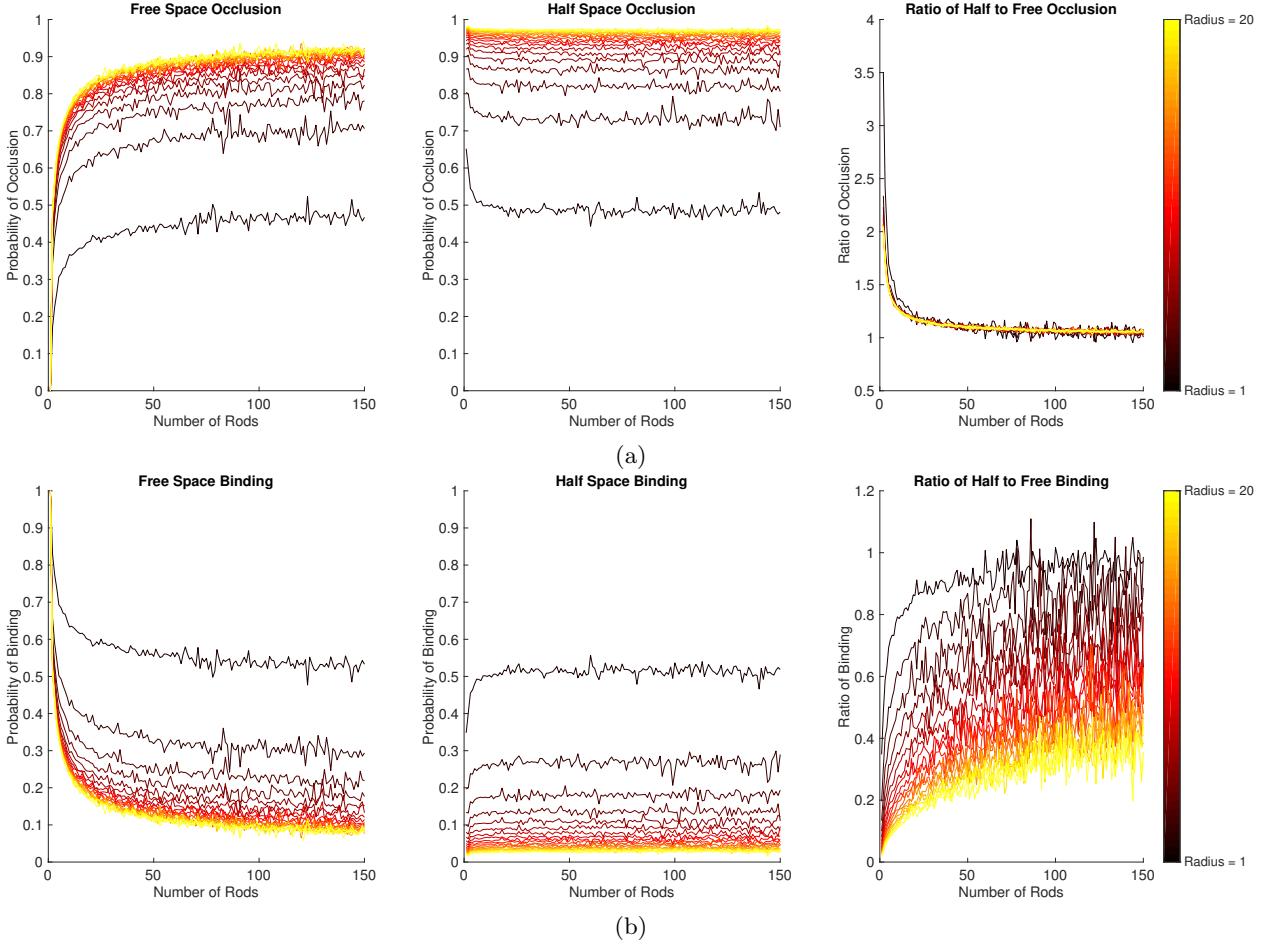


Figure 25: Probability of ligand being occluded, p_{occ} , (top row) and probability of ligand binding, $1 - p_{occ}$, (bottom row) as a function of tether length for various ligand radii. Free-space (left column) half-space (middle), and the ratio of probabilities half-space to free-space (right).

7.3.2 Reactions between tethered reactants are net enhanced by a surface

We explore how effective concentration is impacted by ligand size and polymer length, where effective concentration is defined as how often a ligand bound to the tail of one polymer encounters a binding site on the tail of a second polymer, with the origins of the two polymers separated by a fixed distance. When ligand size is zero, or equivalent to no ligand attached to the polymer, then we see the free space polymers experience an effective concentration predicted by the worm-like chain model. Note that variability from the

theoretical WLC curve is mitigated by decreasing the volume considered a binding encounter. Naturally, when we introduce a surface or ligand of nonzero size, the worm-like chain model no longer describes the effective concentration. We see that the surface enhances the effective concentration over the free space values but that a larger ligand size reduces the effective concentration. (Fig. 26)

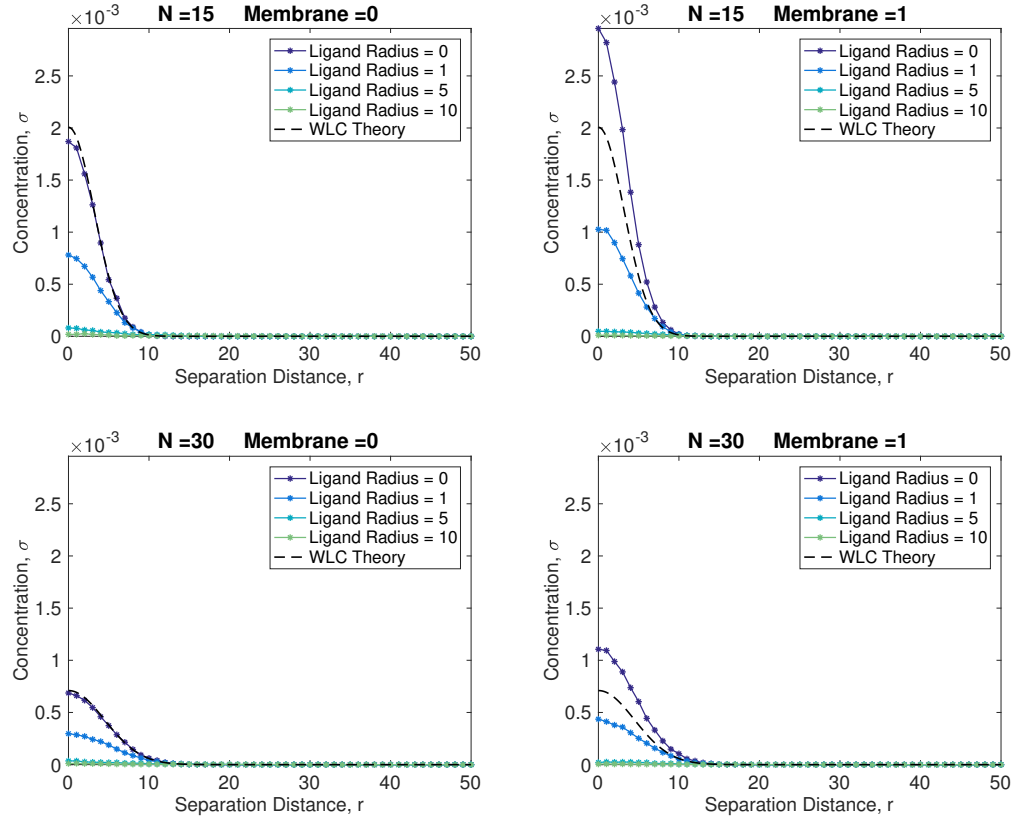


Figure 26: Local concentration as a function of tether separation distance.

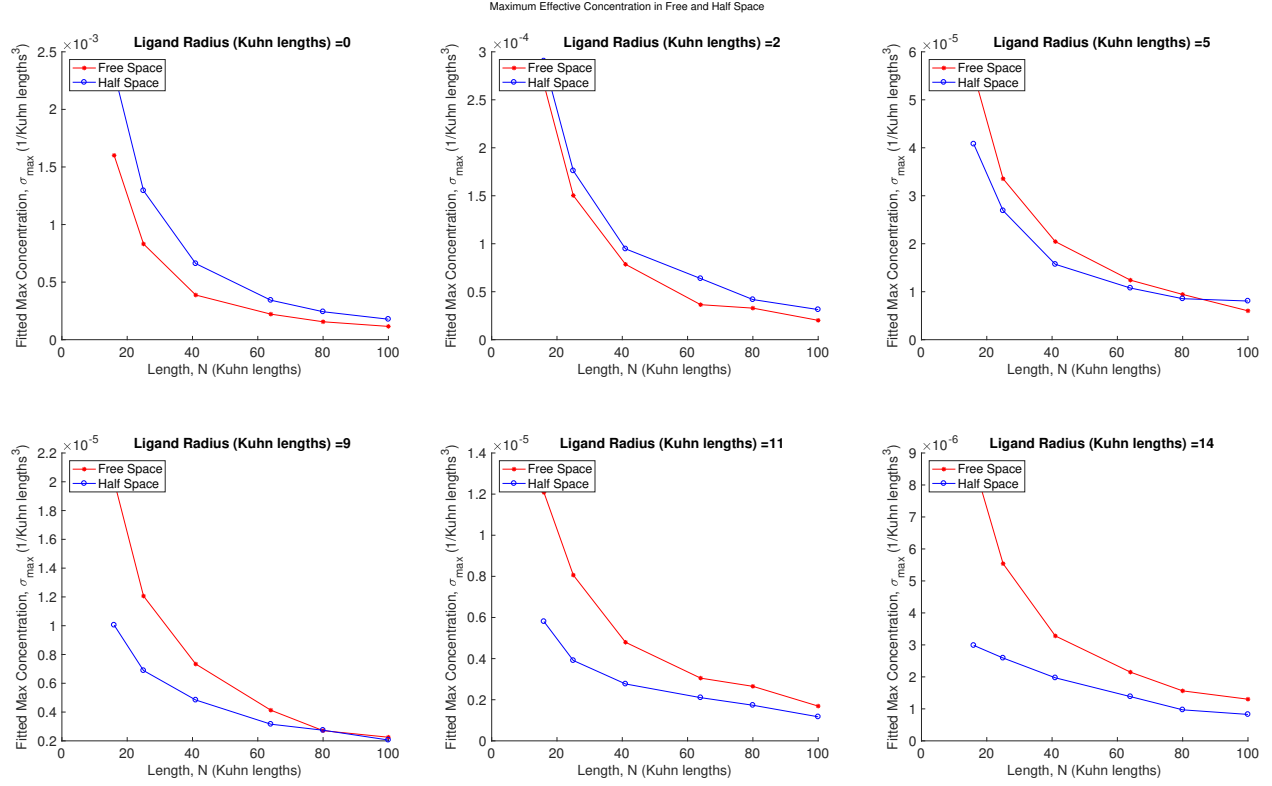


Figure 27: Maximum effective concentration of half-gaussian fit to local concentration data plotted as a function of tether length, N for various ligand radii in free-space (red) and half-space (blue).

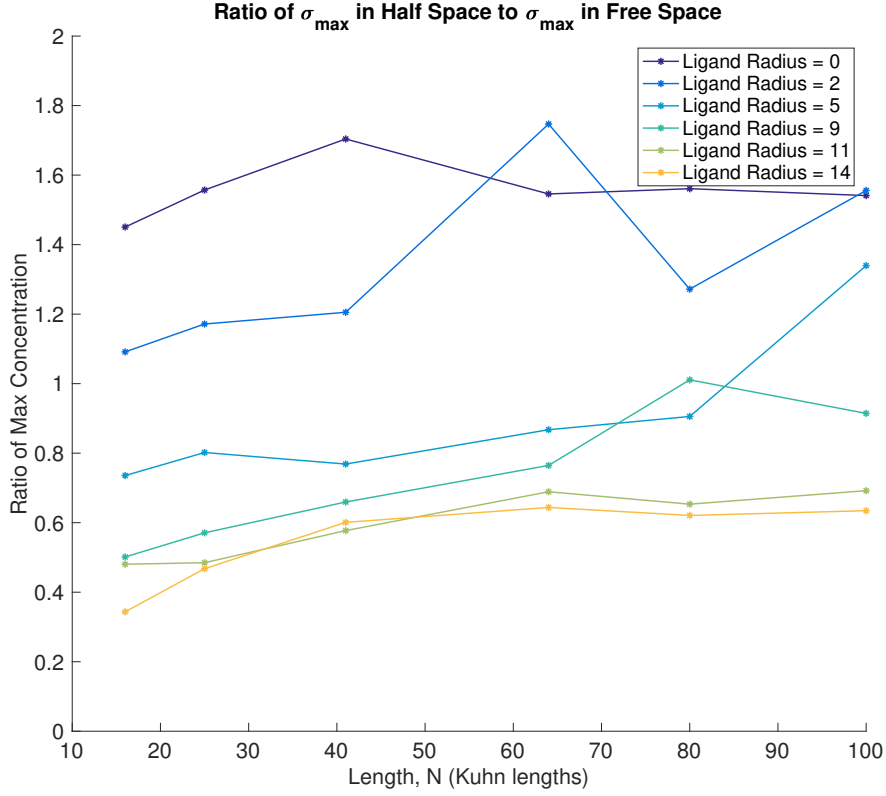


Figure 28: Ratio of half-space fitted maximum effective concentration to free-space fitted maximum effective concentration as a function of tether length, N (Kuhn lengths), for various ligand radii.

7.3.3 Reach parameter is unaffected by presence of surface

We fit a half-gaussian curve to the simulated local concentration data and compare the ‘reach parameter’, l , in free-space to half-space. Our data shown earlier indicates that the presence of a surface will extend a single polymer compared to in free space. We would expect this straightening to also increase the reach parameter for the effective concentration. However, we find that the reach parameter, l , is unaffected by the presence of a surface. This parameter changes based on polymer length and ligand size, but not presence of a surface.

Reach Parameter in Free and Half Space

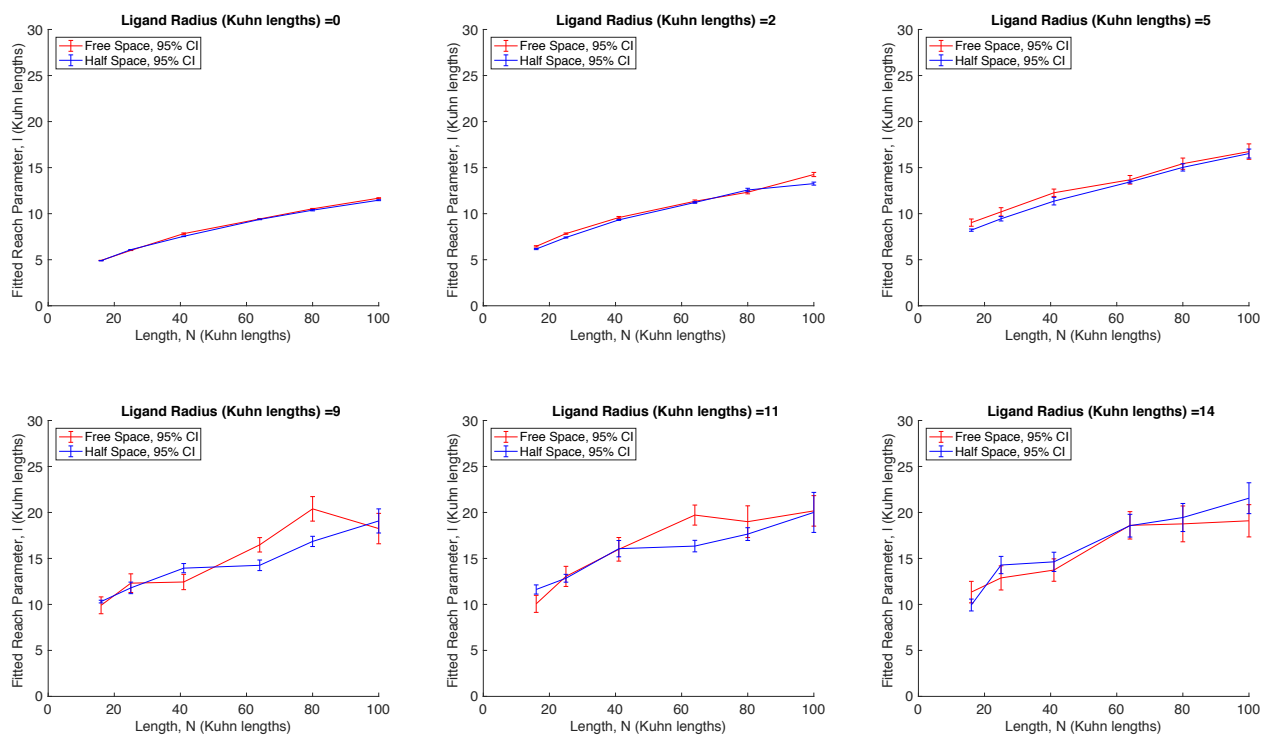


Figure 29: Reach parameter of half-gaussian fit to local concentration data plotted as a function of tether length, N for various ligand radii in free-space (red) and half-space (blue) with 95% confidence intervals for the fit.

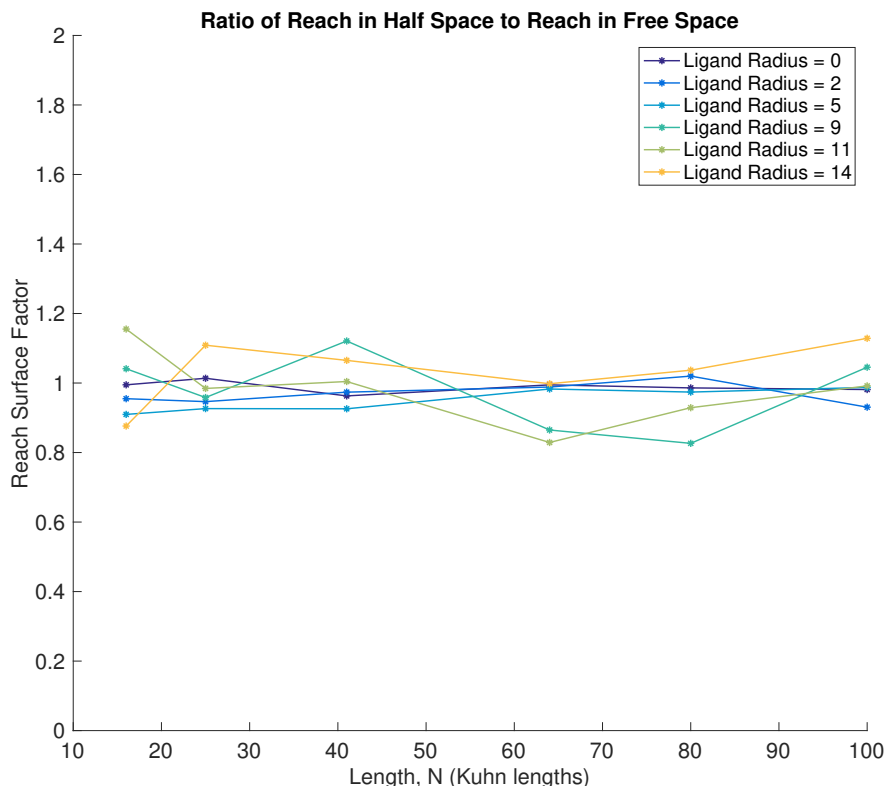


Figure 30: Ratio of half-space fitted reach parameter to free-space fitted reach parameter as a function of tether length, N (Kuhn lengths), for various ligand radii.

7.3.4 Matrix-bound versus surface-bound experiments confirm model predictions

plots/table go here....

7.3.5 Future Work: Catalysis surface factor $\kappa_{cat}^{(teth2teth)}$ versus l_C and/or versus r_{ligand}

Experimental measurements of effective concentration do not distinguish between the effective concentration and the catalytic rate. Since these values are lumped into a single number, it is important to know how the surface impacts the apparent catalytic rate. This might be flat out incorrect. include calculation of how to find catalytic ratio? We will plot the ratio of apparent catalytic rates for various polymer lengths and ligand sizes. This will indicate if there is a relationship between the catalysis factor and the reaction parameters or if there is a single value describing the impact of the surface. The catalysis surface factor may then be used to determine how experiments on a surface differ from on a matrix and possibly give a conversion factor between the two.

8 Timeline

Table 5: Timeline of Coursework. Courses in italics do not count towards degree course requirements.

Fall 2014	Winter 2015	Spring2015	Summer 2015
Dev Bio 212 Math 227A Dev Bio 203A Physics 230A	Dev Bio 212 Math 227B Dev Bio 232	Dev Bio 212 Math 227C Dev Bio 203C M&MG 250	Research
Fall 2015	Winter 2016	Spring2016	Summer 2016
<i>Math 205A</i> <i>Math 220A</i> <i>Math 230A</i> <i>Math 298A</i>	<i>Math 205B</i> <i>Math 220B</i> <i>Math 230B</i> <i>Math 298B</i>	<i>Math 205C</i> <i>Math 220C</i> <i>Math 230C</i> <i>Math 298C</i>	Research
Fall 2016	Winter 2017	Spring2017	Summer 2017
Dev Bio 212 Math 225A MMB 204	Dev Bio 212 Math 290B	Dev Bio 212 CS 284C Physics 230B	Research Preliminary Exam Advancement
Fall 2017	Winter 2018	Spring2018	Summer 2018
Dev Bio 212	Dev Bio 212	Dev Bio 212	Research
Fall 2018	Winter 2019	Spring2019	Summer 2019
Dev Bio 212	Dev Bio 212	Dev Bio 212	Defense

9 Appendix

9.1 Model Modifications

9.1.1 Stiffening

To include this model assumption, we modify the simulation in the following way:

1. Include binary sequence to indicate which sites are “stiff”, sequence specified by user.
2. Create distance metric to indicate how far on either side of the site the effect extends, specified by user.
3. Include function to read binary sequence and create list of “stiff” joints, based on distance metric.
4. Prohibit joints on list from being chosen when simulation chooses joint to move.

9.1.2 Electrostatics

To include this model assumption, we modify the simulation in the following way:

1. Create binary sequence to indicate which sites are “phosphorylated”.
2. Create distance metric to indicate how far on either side of the site the effect extends, specified by user.

3. Create function to read binary sequence and create list of which sites to ignore energy, based on distance metric.
4. Calculate energy of system based on Lennard-Jones or Piecewise potential curve plus hard or soft wall constraints.
 Polymer energy is the sum of the energy for each joint in polymer.
 Energy of sites on phosphorylated list are not included in the sum.
 Depth, width and wall/center location are specified by user.
5. Accept conformations that produce a lower energy than the previous conformation.
6. Reject with normal probability any conformations with higher energy than previous conformation.
7. Rejection keeps polymer in previous conformation, but goes on to next time step. (Metropolis step.)

9.1.3 Simultaneous Binding

To include this model assumption, we modify the simulation in the following way:

1. Create “bound ligands”, with location and radius specified by user. (Location specified by file or binary sequence.)
2. Bound ligands occlude polymer, rejecting conformations that intersect the ligand sphere.
3. Bound ligands occlude each other, rejecting conformations where two ligands intersect each other.
4. If there is a membrane, bound ligands may not fall below membrane, rejecting conformations where this occurs.
5. Rejecting conformations does not step time forward. Continue looking for conformations until one meets constraints.
6. Binding sites are considered ‘occluded’ if a bound ligand intersects the binding site sphere.
7. Binding sites located where bound ligand is located read as always occluded.

9.2 Assumptions

9.3 Gillespie Algorithm