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

1.1 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 [?]. Similarly, a segment of the protein Tau, has been shown to have increased persistence length upon multiple phosphorylation [?]. 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 [?].

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 [?].

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, CD3ζ, 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 CD3ζ. Specifically, we want to know if local structuring of a disordered protein upon multi-site phosphorylation can induce enhancement of subsequent binding events.

1.2 Model and Methods

We model the TCR CD3ζ chain as a freely jointed chain, where each amino acid is represented as a segment as described in Sec. **??**. 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 CD3ζ 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 $2.1\,\mathrm{nm}$(7 Kuhn lengths) as calculated in Sec. **??**.

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. 1). 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.
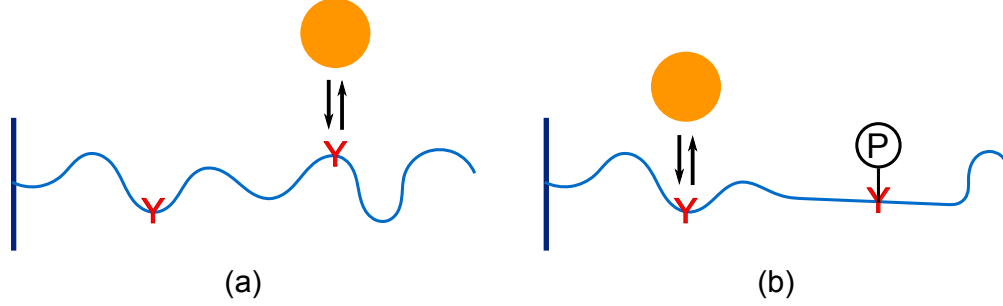
Figure 1: Cartoon of partial stiffening model. Entirely floppy FJC prior to modification (a) transitions to partially rigid chain after modification (b).

For TCR CD3ζ, phosphorylation events are modeled as creating local structure as described above. Similarly, we model dephosphorylation as 'unfreezing' nearby segments. Since we will eventually model 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. For example, if phosphorylation would locally structure the nearest eleven segments, then the fully phosphorylated form of CD3ζ would have 66 residues frozen. We use this as the initial state for dephosphorylation. Each dephosphorylation would unstructure 11 of the initial 66 frozen residues. For simplicity, we assume the phosphatase is the same size as the ligand Lck.

1.3   Results

1.3.1   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. 2). 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.
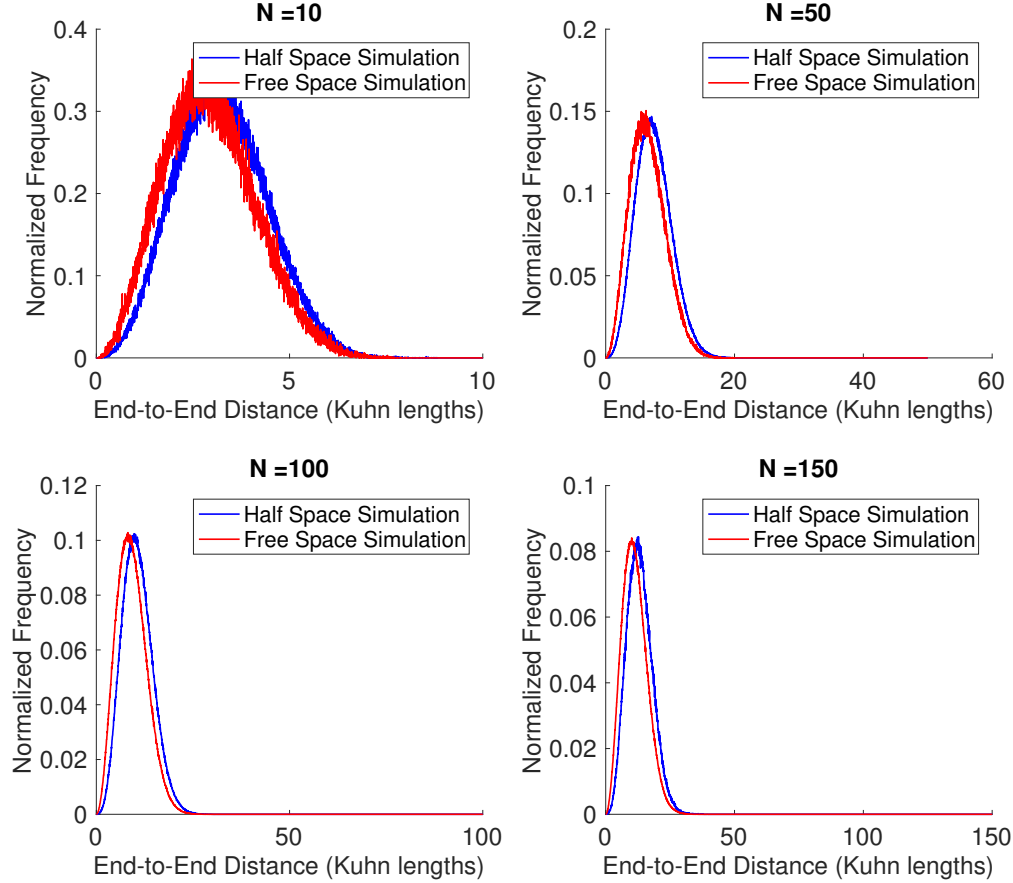
Figure 2: Distributions of polymer end-to-end distance in free space and half space for various polymer lengths (N). For each length, the end-to-end distance of the polymer in half-space is overall straighter than in free space.

1.3.2 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. 3). 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. 3).
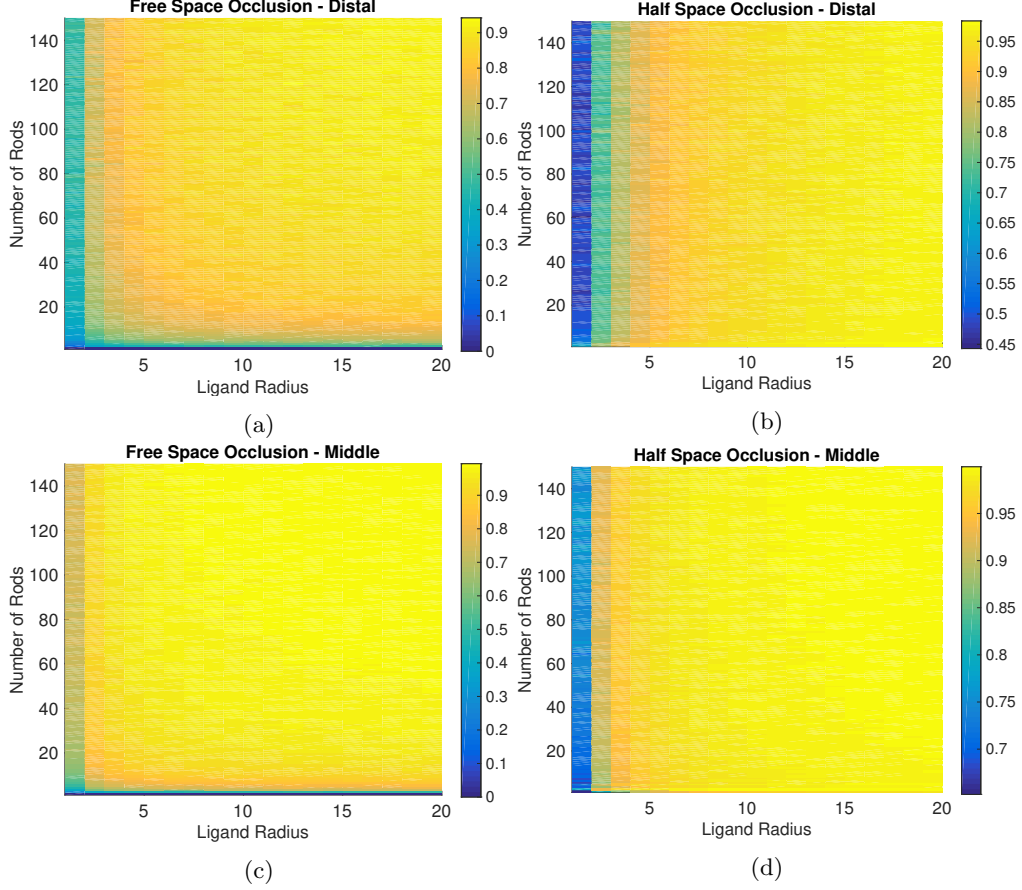
3

Figure 3: 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.

1.3.3 Analytical solution of occlusion probability matches simulated result for binding site at polymer tip

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 δN, 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 occluded. Then we may formulate as follows for the probability of survival, $p(\vec{x}, t)$ when the random walk begins at a starting point, $r = R + \epsilon$:

$$\begin{cases} \frac{\partial p}{\partial t} = D\nabla^2 p, & p = 0 \quad \text{at} \quad r = R \\ p(\vec{x}, 0) = \delta(\vec{x} - (R + \epsilon)) \end{cases} \tag{1}$$

For an approximation, we consider the solution when $N \to \infty$.

4

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

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

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

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

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 $N \to \infty$.
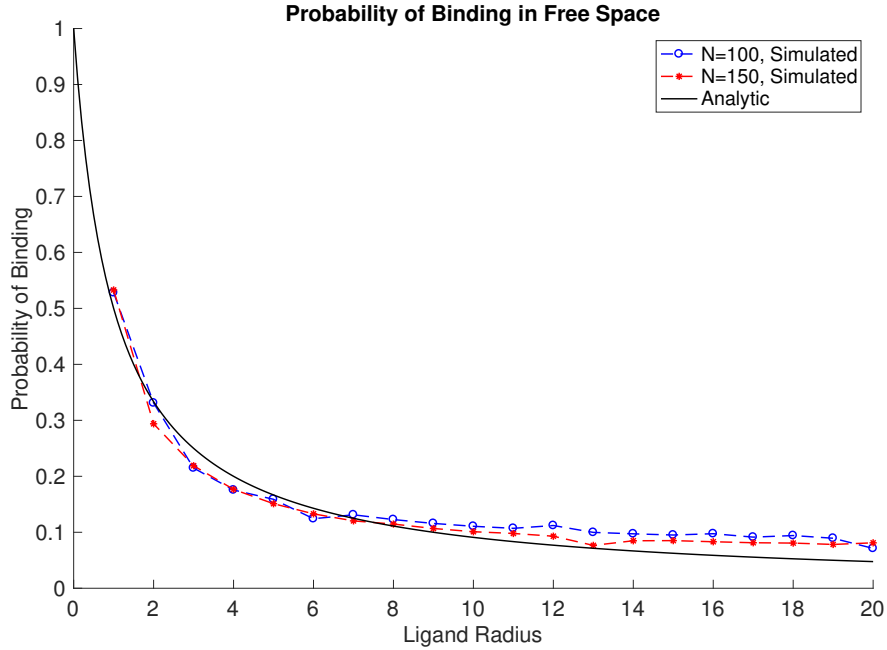

Figure 4: Analytic solution (black, solid) compared to occlusion for free-space polymer of 100 Kuhn lengths (blue) and 150 Kuhn lengths (red) for ligands of radius, R Kuhn lengths.

1.3.4 Cooperativity may arise from local structuring in intrinsically disordered proteins

We now simulate the kinase Lck binding to the CD3ζ chain, assuming that each phosphorylation event will create a disorder-to-order transition. If we assume that phosphorylation of tyrosines on the CD3ζ 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. 5).
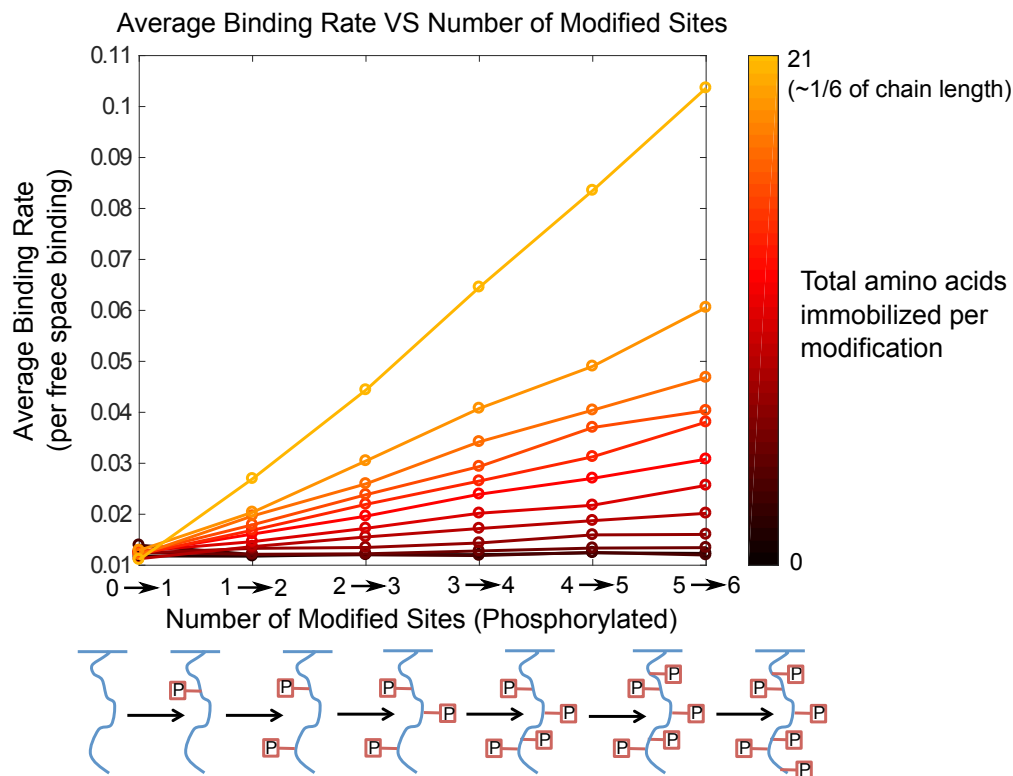

Figure 5: (Above) Simulated average binding rates of Lck to CD3ζ for different levels of phosphorylation given a specified degree of local structuring (none $\rightarrow$ black, low $\rightarrow$ red, high $\rightarrow$ orange). (Below) Cartoon representing a possible phosphorylation state series of CD3ζ.

1.3.5   Sequential binding naturally arises from polymer characteristics

We hypothesized that without any modifications due to binding, with only the presence of a membrane, we would see preferential binding in an membrane-distal to membrane-proximal manner. In 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 expected that tyrosines closest to the membrane would see extra kinase inhibition from membrane occlusion, whereas the tail of the disordered protein would not experience this same effect.

For the local structuring model, we predict that this effect would be enhanced. In an unphosphorylated state, if the tyrosine furthest from the membrane (here referred to as the sixth tyrosine) is most likely to be bound, then local structuring 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 stucturing would occur around the sixth tyrosine, the nearest 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 occurs. When we compile the path frequency data, we find 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 show only the paths membrane proximal to membrane distal (123456) and distal to proximal (654321) (Fig. 6). 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 (Fig. 6). 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. 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 stucturing 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 CD3ζ 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 (Fig. 3). Since the tyrosines in CD3ζ are not equally spaced along the length of the protein, there is a bias for binding 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 CD3ζ but equally spaced tyrosines, we see that there is no longer a preference between the two paths.
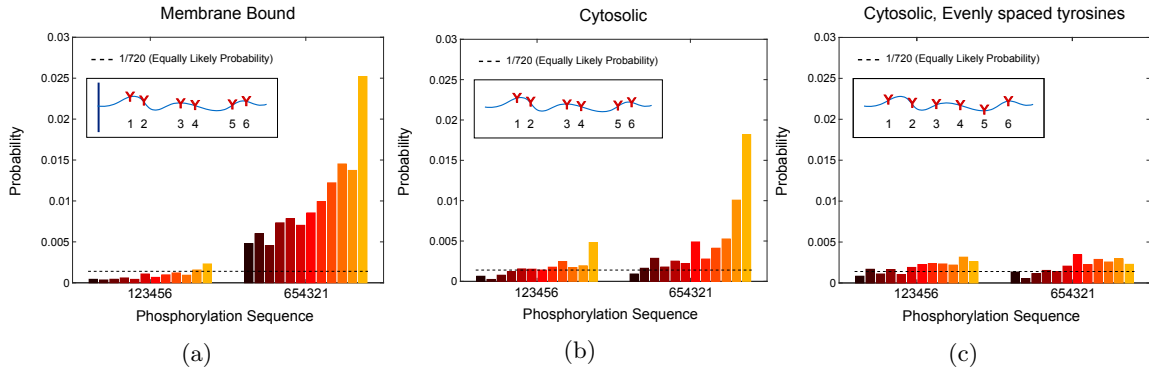

Figure 6: Probability of phosphorylating membrane proximal to distal (123456) compared with distal to proximal (654321). Equally likely probability of all paths indicated with black, dotted line. (a) CD3ζ parameters, with membrane. (b) CD3ζ parameters, without membrane. (c) CD3ζ length, evenly spaced tyrosine locations, without membrane.

1.3.6 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 CD3ζ 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. 7).
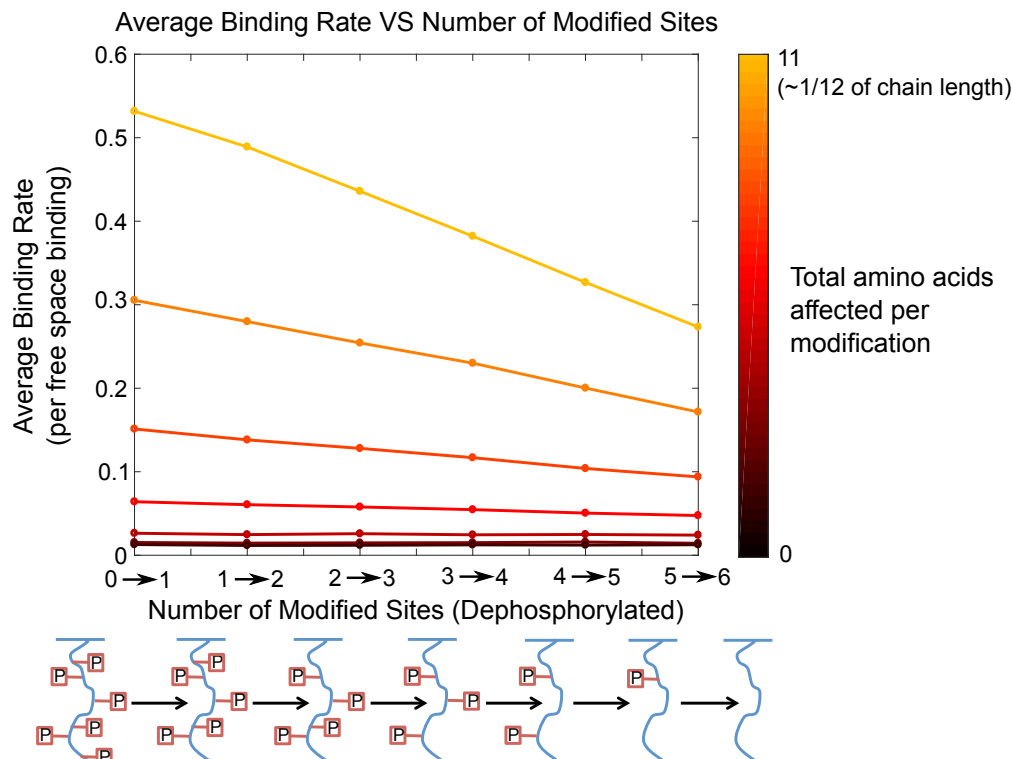

Figure 7: (Above) Simulated average binding rates of Lck to CD3ζ for different levels of dephosphorylation given a specified degree of local unstructuring (none $\rightarrow$ black, low $\rightarrow$ red, high $\rightarrow$ orange).

1.3.7 Future Work: Quantifying cooperativity from reversible phosphorylation with Hill co-efficients

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

1.3.8 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. We will 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. Exploring all sequences. will elucidate whether sequential binding occurs when both phosphorylation and dephosphorylation are impacting the polymer simultaneously.

1.4 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 CD3ζ chain, there is an enhancement of phosphorylation rate that can help explain our previously published experimental data [?]. There are several examples of disordered-to-order transitions upon PTM [? ? ?]. It is possible these systems could exhibit this same rate enhancement. Positive cooperativity has been extensively studied both theoretically and experimentally [? ? ? ?] . 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 CD3ζ 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 [? ? ?] and has been shown to create nonlinear signaling behavior, such as ultrasensitivity. Note that theoretical studies [?] 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.