

Biochemistry on a leash: The roles of tether length and geometry in signal integration proteins

David Van Valen

Department of Applied Physics,
California Institute of Technology, Pasadena, CA

Mikko Haataja

Department of Mechanical and Aerospace Engineering
Princeton Institute for the Science and Technology of Materials (PRISM),
Princeton University, Princeton, NJ

Rob Phillips¹

Department of Applied Physics,
California Institute of Technology, Pasadena, CA

¹Corresponding author. Address: Department of Applied Physics, California Institute of Technology, MC 114-96 1200 East California Blvd, Pasadena, CA 91125, U.S.A., Tel.: (626)395-3374, Fax: (626)395-5867

Abstract

We use statistical mechanics and simple ideas from polymer physics to develop a quantitative model of proteins with flexibly tethered ligands and receptors that control the activity of these proteins. We predict how the properties of tethers influence the function of these proteins and demonstrate how their tether length dependence can be exploited to construct proteins whose integration of multiple signals can be tuned. One case study to which we apply these ideas is that of the Wiskott Aldrich syndrome proteins (WASPs) as activators of actin polymerization. More generally, tethered ligands competing with those free in solution are a common phenomena in biology, making this an important specific example of a widespread biological idea.

Key words: Tether; WASP; Actin; Biochemistry on a Leash

1 Introduction

Flexible chains of amino acids, referred to here as tethers, are important for the function of natural and synthetic proteins. Four examples of natural proteins in which such tether motifs are thought to play an active role are actin polymerization by formin proteins, the autoregulation of Src family kinases, the methylation of bacterial chemotaxis receptors, and the inactivation of potassium channels. Several of these examples are highlighted schematically in figure 1.

Actin is an important biopolymer; it is important for cell division, motility, and other tasks like phagocytosis. Given its vital role in the life of a cell, it is no surprise that the formation of actin filaments is tightly controlled by the cell. Formins are one class of proteins that participate in this process by promoting the growth of actin filaments. Formins have been shown to increase the rate of polymerization by as much as 5 or 15 fold; how they achieve this speed up remains an open question [1, 2]. Recently, a physical model for this speed up has been proposed that treats the activation process like a ball in a cup [3]. In this model, depicted in figure 1a, flexible tethers connect binding sites for profilin-bound actin to an FH2 domain, the portion of formins known to bind to the barbed end of actin filaments. The flexible tethers serve to deliver actin monomers to the barbed end, enabling the polymerization process to proceed more rapidly than permitted by the diffusion limit. While this model has yet to be validated, it is an intriguing hypothesis that such tethers play an active role in formin-mediated actin polymerization

Src kinases are another family of proteins in which tethers are thought to play a prominent role [4, 5]. Src kinases are signaling proteins that control a number of cellular processes including adhesion and cell division. In particular, Src kinases are thought to posses both an inactive and active conformation. They also possess SH2 and SH3 domains, protein domains that recognize and bind to small peptide sequences. In Src kinases, these domains are connected to their respective ligands by flexible tethers. In the current model of Src regulation, the SH2 and SH3 domains bind to their tethered ligands to stabilize the inactive conformation; this model is shown in figure 1b. The receptor-ligand interaction is suboptimal, enabling ligands free in solution to compete for binding with the tethered domain and activate the Src kinase [4, 5]. Experimental and theoretical studies have shown that the regulation is dependent on the composition of the tethers connecting the multiple subunits [6, 7]. In this system and others, the role tethers play can be thought of as biochemistry on a leash. In these cases, there is a competition between the tethered ligands and those that are free in solution. Simple physical models and clever experiments are starting to reveal how this competition plays out [8]. A quantitative picture will deepen our understanding of how signaling proteins work.

Tethers are also believed to play an important role in the methylation of the chemotaxis receptors Tar and Tsr. The flexible C-terminus of these receptors is thought to bind to the methyl-transferases CheR and CheB, linking them to the receptor with a flexible tether. While tethered, these proteins can methylate the chemotaxis receptor, an event crucial for adaptation. The flexible tethers linking the transmembrane receptor with the methyl-transferase has been the focus of recent theoretical work [9]. Through use of a dynamic description of the tethers, Windisch *et. al.* have demonstrated how tethering can lead to a more efficient search for substrates as well as control the orientation of the tethered CheR and CheB [9].

The inactivation of the voltage gated *Shaker* potassium channel is another instance in which tethers are thought to play an important role in the biological function of a protein. The *Shaker* potassium channel adopts an open conformation when a large transmembrane voltage is applied, leading to a flux of potassium ions across the membrane. Once open, the channel can be spontaneously inactivated in a voltage independent fashion; mutational studies by Hoshi *et. al.* identified an N-terminal cytoplasmic portion of the ion channel as the portion responsible for this rapid inactivation [10–13]. The current model for inactivation posits that the cytoplasmic domain consists of a ball connected to a leash [12,13]. The ball can bind to the pore and block the flux of potassium ions. In this model, the tether serves to effectively increase the concentration of balls seen by the open pore. Interestingly, the degree of inactivation has been shown to be dependent on the length of the tether. Mutants with longer tethers have less inactivation while mutants with shorter tethers exhibited more inactivation. Work by Timpe *et. al.* demonstrated that this behavior was consistent with a model in which the tethers were modeled as a random walk [14]. As exhibited by these examples, simple polymer models have much to tell us about how tethers can influence the behavior of natural proteins.

Tethers have also been used to design new proteins with improved or novel functions. Applications vary from increasing folding rates of homodimers to designing multivalent ligands [8,15,16]. Recent work by Krishnamurthy *et. al.* has demonstrated how tethers can be used to construct ligands that bind to multiple sites on a protein [8]. In their work, they connected human carbonic anhydrase and its ligand using a flexible polyethylene glycol tether. In this system, they were able to quantify the competition between free and tethered ligands and measure how this competition varied when the length of the tether was tuned. While this work has important consequences for the design of multi-valent ligands, the results also speak to how biochemistry on a leash plays out in natural biological systems. Indeed, we argue later that tether length is an important dial that is key for modulating the behavior of synthetic proteins and a useful probe to understand how natural proteins work.

Another instance in which tethers have been used to design new proteins is the rewiring of the WASP family of signaling proteins [17]. Wiskott Aldrich Syndrome proteins, or WASPs, are a family of proteins that regulate the creation of new actin filaments by the Arp2/3 complex [18–23]. One common feature of these proteins is that activation of Arp2/3 is carried out by a C-terminal VCA domain. The VCA domain binds both the Arp2/3 complex and actin, and has been shown to deliver the first monomer in a new actin filament [20,24]. As with Src kinases, regulation occurs through auto-inhibitory interactions that stabilize an inactive conformation [18,23,25]. Chemical signals can bind and competitively relieve this auto-inhibition; the binding of Cdc42 and PIP₂ to the family member N-WASP is a classic example [22,25].

As with regulatory networks, reprogramming is often a metric for understanding [26–31]. Recent work in modular recombination has produced synthetic N-WASPs by tethering a VCA domain to a receptor and its associated ligand [17]. The interaction between the receptor and its tethered ligand serve to mimic the auto-inhibitory interactions in native N-WASP. Some of the richest results of this recombination technique involve constructs comprised of multiple receptor-ligand pairs on the same tether (see figure 6). In these cases, a binding partnership between one receptor-ligand pair will physically shorten the tether connecting the remaining pair, resulting in a different free energy

of closure and novel signal integration behavior.

In conjunction with reprogramming, WASPs have also been the focus of quantitative modeling [17, 22, 32, 33]. Engineering and theoretical approaches have proven useful, as they serve to challenge and sharpen our understanding. In particular, one role of theory is to test models by having dials that can be tuned to produce falsifiable predictions. The cartoons that depict these proteins suggest the geometry of the tethers can be tuned to elicit different biological functions. Experiments have shown that changing tether lengths can affect the behavior of synthetic proteins; theoretical studies have reported similar findings [7, 15–17, 34]. The goal of this paper is to examine how the general features of tethers influence the behavior of natural and synthetic proteins. In particular, we develop a quantitative model for tethered ligand-receptor pairs and show through a case study how the features of tethers can control the behavior of signaling proteins.

The outline of this paper is as follows. In section 2 we present our model for the interplay between free and tethered ligands. In section 3 we outline a simple random walk model used to describe flexible tethers. In section 4 we apply the model to predict the tether length dependence of a one-input synthetic N-WASP construct. We also extend our statistical mechanical model to study the input/output properties of multiple input constructs and illustrate how tethers can influence the signal integration properties of these proteins. We finish our discussion in section 5 by discussing the role tether geometry plays in protein-mediated signal integration and discussing how this model can be used to examine other cases of biochemistry on a leash.

2 Statistical Mechanical Model of Tethered Receptor-Ligand Pairs

In this work, we restrict ourselves to the case where tethered receptor-ligand binding inhibits protein activity. Our motivating example for selecting this motif are the synthetic WASP proteins that regulate actin polymerization by Arp2/3. A schematic for one of these constructs, which we call the simple switch, is outlined in figure 2. The receptor and tethered ligand, also called *cis*-ligand, are each connected to a tether. The two tethers originate from distinct, fixed locations in space. This is likely the case if there are some intervening secondary structural elements between the origins of the two tethers. In this motif, the switch is considered to be off when the receptor and *cis*-ligand are bound and on otherwise.

The starting point for our model is the assumption that the switches are effectively in thermal equilibrium. Because life is an inherently non-equilibrium process, care must be exercised when attempting to apply an equilibrium model to a biological system. Our model of the simple switch is one such case. We imagine that the active states of the switch are coupled to some productive reaction, as envisioned in figure 11. The intuition behind our equilibrium assumption is a separation of time scales between ligand binding and this productive reaction. Binding of ligands and the spatial fluctuations of unstructured amino acids are fast, occurring on a time scale of milliseconds or less [35–37]. For our motivating example, this process is Arp2/3 mediated actin polymerization, a process that occurs on the timescale of seconds or longer. Because of this separation of timescales, we expect the productive reaction to only care about the equilibrium behavior of the switches. We give this argument a more concrete mathematical treatment in the appendix.

We are interested in finding the probability that a simple switch occupies an active state, a

quantity we label p_{on} . In our model, this quantity is our metric for the activity of the simple switch. To proceed with the calculation, we first invoke the standard statistical mechanics toolkit. Statistical mechanics tells us that the probability of the system being in a given collection of microscopic states is given by

$$p_{state} = \frac{ge^{-\beta\epsilon}}{Z}, \quad (1)$$

where g is the degeneracy of the collection, $\beta \equiv \frac{1}{k_B T}$, ϵ is the energy of the collection, and Z is a normalization factor called the partition function. For our problem, there are three factors that determine the statistical weight of a collection - the strength of the interaction between the receptor and its ligands, the configurations of the ligands in solution, and the configurations of the tethers. Simple arguments lead to expressions for all three factors.

We treat the interaction between the receptor and its ligands by letting free ligand bind to the receptor with a binding energy of ϵ_f and cis-ligand bind with energy ϵ_t . As shown in figure 2, this identifies three different energy levels; one in which no ligand is bound, one in which free ligand is bound, and one in which cis-ligand is bound. Applying equation 1, we can write the probability the switch is active as

$$p_{on} = \frac{\sum_{i \in \text{active states}} g_i e^{-\beta\epsilon_i}}{\sum_{i \in \text{all states}} g_i e^{-\beta\epsilon_i}}. \quad (2)$$

The configurations of the ligands in solution and the configurations of the tethers contribute to the degeneracy of each energy level. Assuming independence of the contributions, we can write the degeneracy of the i^{th} energy level as $g_i = \Omega_i^{sol}\Omega_i^{tet}$, where Ω_i^{sol} and Ω_i^{tet} are the number of configurations available to the ligands in solution and tethers respectively. In what follows, we will assume that the ligands are in excess of the switches; that is, the switches do not compete with each other for ligands in the solution. If there is one ligand in solution, then the number of configurations for that ligand scales with the volume of the solution (here we invoke ideas appropriate to a dilute solution). With L indistinguishable ligands in a solution of volume V , the number of configurations available to them scales as $\Omega_L^{sol} \propto V^L/L!$. When a ligand in solution binds to the receptor, its available volume is restricted to a small volume v_s . The number of configurations then scales as $\Omega_{L-1}^{sol} \propto V^{L-1}v_s/(L-1)!$. We can account for the configurations of the tethers by appealing to basic polymer physics. The physical intuition is that the tethers can explore more configurations when the receptor is free than when it is bound to its cis-ligand, making it entropically favorable for the switch to be on. Models from polymer physics allow us to quantify this effect. What we wish to know are Ω_{on}^{tet} and Ω_{off}^{tet} , the number of configurations available to the tether in the on and off states, respectively. In the active state, we assume the tethers are unconstrained. In the inactive state, the configurations available to the tethers are restricted because their binding of the receptor with its cis-ligand forces the tether ends to lie near each other. The ratio between Ω_{off}^{tet} and the total number of possible tether configurations is the probability that the tether ends are close enough for binding to occur. In other words, if the probability of the two tether ends being

close is p_{loop} , then

$$\frac{\Omega_{off}^{tet}}{\Omega_{on}^{tet}} = p_{loop}. \quad (3)$$

With these expressions in hand we compute the statistical weight for each state. Further, as shown in figure 2, we can write the probability that the switch is active as

$$p_{on} = \frac{\Omega_{on}^{tet}\Omega_L^{sol} + \Omega_{on}^{tet}\Omega_{L-1}^{sol}e^{-\beta\epsilon_f}}{\Omega_{on}^{tet}\Omega_L^{sol} + \Omega_{on}^{tet}\Omega_{L-1}^{sol}e^{-\beta\epsilon_f} + \Omega_{off}^{tet}\Omega_L^{sol}e^{-\beta\epsilon_t}}. \quad (4)$$

By dividing the top and bottom of this expression by $\Omega_{on}^{tet}\Omega_L^{sol}$, this expression reduces to

$$p_{on} = \frac{1 + \frac{\Omega_{L-1}^{sol}}{\Omega_L^{sol}}e^{-\beta\epsilon_f}}{1 + \frac{\Omega_{L-1}^{sol}}{\Omega_L^{sol}}e^{-\beta\epsilon_f} + \frac{\Omega_{off}^{tet}}{\Omega_{on}^{tet}}e^{-\beta\epsilon_t}}. \quad (5)$$

By definition, $\Omega_{L-1}^{sol}/\Omega_L^{sol} = v_s[L]$, where $[L]$ denotes the concentration of ligand. Thus, the probability p_{on} can be expressed as

$$p_{on} = \frac{1 + \frac{[L]}{\frac{1}{v_s}e^{\beta\epsilon_f}}}{1 + \frac{[L]}{\frac{1}{v_s}e^{\beta\epsilon_f}} + \frac{\frac{1}{v_s}p_{loop}}{\frac{1}{v_s}e^{\beta\epsilon_t}}}. \quad (6)$$

At this point, it is convenient to define a probability density

$$J = \frac{p_{loop}}{v_s}, \quad (7)$$

characterizing the concentration of the tethered ligand in the vicinity of the tethered receptor. In this simple model, the receptor and tethered ligand are treated as point objects and are free to bind when the tether ends are within a small volume of each other. Naturally, the looping probability will depend on what this volume is; however, the probability density will not. Furthermore, we can formulate our entire problem in terms of probability densities and calculate them directly from different polymer models. Readers familiar with DNA looping will find this probability density very similar to the J factor first used by Jacobson and Stockmayer to describe DNA cyclization [38]. Using the probability density, we can rewrite p_{on} as

$$p_{on} = \frac{1 + \frac{[L]}{\frac{1}{v_s}e^{\beta\epsilon_f}}}{1 + \frac{[L]}{\frac{1}{v_s}e^{\beta\epsilon_f}} + \frac{J}{\frac{1}{v_s}e^{\beta\epsilon_t}}} = \frac{1 + \frac{[L]}{K_d}}{1 + \frac{[L]}{K_d} + \frac{J}{K'_d}}, \quad (8)$$

where $K_d = \frac{1}{v_s}e^{\beta\epsilon_f}$ and $K'_d = \frac{1}{v_s}e^{\beta\epsilon_t}$ denote dissociation constants. Grouping terms into dissociation constants is useful because they provide a scale for concentrations and because they can be measured experimentally.

The physical interpretation of Eq.(8) is straightforward. The denominator reflects a competition between the ligands in solution and ligands that are tethered. Binding of free ligand to the receptor

has a favorable energy but reduces the entropy of the ligands in solution. Binding of cis-ligand to the receptor is also energetically favorable but reduces the configurations available to the tethers, thus reducing their entropy. When $\frac{[L]}{K_d} \gg \frac{J}{K'_d}$, the free ligand wins the competition and $p_{on} \approx 1$.

When $\frac{J}{K'_d} \gg \max\left[\frac{[L]}{K_d}, 1\right]$, on the other hand, the cis-ligand wins and $p_{on} \approx 0$. Also note that the probability density J has the units of concentration, and hence can be thought of as an effective concentration. In other words, it is the concentration of cis-ligand seen by the receptor. To quantify the interplay between the entropy of the solution and the entropy of the tethers, we need a physical model of the tethers. One such model, the random walk model, is outlined in the following section.

3 Polymer Physics of Protein Tethers

In order to have a predictive framework for thinking about the tether-length dependence of signaling protein function, we must have a quantitative description of the competition between the tethered ligand and its untethered counterparts in solution. Specifically, we wish to employ a polymer model to describe the tethers and calculate the probability density of the tether ends. Our ultimate goal is to compute the effective concentration of cis-ligand seen by the receptor. We note that the meeting of two tethers with distinct origins is a problem that has been considered previously in the literature [39]. There are a variety of possible approaches to modeling tethers and in this paper we employ the simplest of them, a random walk model. The limitations of this model are briefly discussed at the end of the section.

In the random walk model, the tether is coarse-grained into a collection of independent segments, as depicted in figure 3. Each segment is allowed to rotate freely in space. The tether is composed of N statistically independent Kuhn segments of length b . The i^{th} segment has orientation \vec{r}_i and the end-to-end-distance vector is given by $\vec{R} = \sum_{i=1}^N \vec{r}_i$. It follows from the central limit theorem that the probability distribution for the end-to-end distance vector, $G(\vec{R})$, takes the form of a Gaussian. The spatial components of \vec{R} are independent for long tethers, so we can write $G(\vec{R}) = G_x(x)G_y(y)G_z(z)$. The central limit theorem demands that G_x , G_y , and G_z be Gaussian as well. Gaussian distributions are characterized by their mean and variance. Symmetry demands the means be zero; the variance is given by $\langle \vec{R}^2 \rangle = Nb^2$. Independence implies the variances are given by $\langle x^2 \rangle = \langle y^2 \rangle = \langle z^2 \rangle = \frac{Nb^2}{3}$. Using these relations, we see that the probability distribution of \vec{R} is given by

$$G(\vec{R}; \xi, L) = \left(\frac{3}{4\pi\xi L} \right)^{\frac{3}{2}} \exp\left[-\frac{3\vec{R} \cdot \vec{R}}{4\xi L}\right], \quad (9)$$

where $L = Nb$ is the total length of the random walk and $\xi = \frac{b}{2}$ is its persistence length.

Because the receptor and cis-ligand are each connected to a tether, this simple model allows us to find their probability densities at each point in space and compute the effective concentration of cis-ligand seen by the receptor. We start by assuming the receptor's tether has length L_1 and begins at some point \vec{r}_1' while the cis-ligand's tether has length L_2 and begins at \vec{r}_2' . If we denote the end of the receptor's tether as \vec{r}_1 and the end of the cis-ligand's tether as \vec{r}_2 , then the probability distribution for the tether ends are given by $G(\vec{r}_1 - \vec{r}_1'; \xi_T, L_1)$ and $G(\vec{r}_2 - \vec{r}_2'; \xi_T, L_2)$ respectively. To find the probability density of the cis-ligand as seen by the receptor, we must

count the tether conformations where the receptor and cis-ligand are in close proximity to each other. This is accomplished by integrating the joint probability distribution of the two tether ends, $G(\vec{r}_1 - \vec{r}'_1; \xi_T, L_1) G(\vec{r}_2 - \vec{r}'_2; \xi_T, L_2)$, over all possible positions of the two tether ends while employing a delta function to ensure that we only count the conformations where the tether ends are in the same location. The effective concentration is then

$$\begin{aligned} J &= \int d^3 r_1 d^3 r_2 \delta^{(3)}((\vec{r}_1 - \vec{r}'_1) - (\vec{r}_2 - \vec{r}'_2)) G(\vec{r}_1 - \vec{r}'_1; \xi_T, L_1) G(\vec{r}_2 - \vec{r}'_2; \xi_T, L_2) \\ &= G(\vec{D}; \xi_T, L_T), \end{aligned} \quad (10)$$

where $L_T = L_1 + L_2$ is the combined length of both tethers, ξ_T is their persistence length, and $\vec{D} = \vec{r}'_1 - \vec{r}'_2$ is the vector between the origin of the two tethers. Note that in the above calculation, we have made use of the fact that the convolution of two Gaussians is itself a Gaussian.

The random walk model provides a simple way to compute the effective concentration of cis-ligand. However, this model is not without its flaws. For example, the wormlike chain (WLC) model, which attributes a bending energy for different tether conformations, has been shown to fit force-extension measurements and distributions of length sizes extracted from protein databases [40–43]. Other models, such as the rotational isomeric state model or the three-bead rotating chain are justified on experimental and structural grounds [44, 45]. More complicated models like the WLC are necessary to examine short tethers. On the other hand, the WLC and RW models provide the same quantitative description for long tethers (unpublished data). Further, none of the above models address the influence of self avoidance and nearby macromolecules on tether conformation. These excluded volume effects, particularly the role of nearby macromolecules, have been the subject of theoretical modeling [9, 46]. Recent experimental work suggests that they play an important role in tethered receptor-ligand systems [8]. While these concerns are important, we have chosen this model because our goal was to develop a simple model to explore tethering in proteins and to illustrate the kinds of effects such tethering can give rise to. Furthermore, we expect that our *qualitative* conclusions will remain valid even when self avoidance becomes important. Further consideration of excluded volume effects is beyond the scope of this work and will be the focus of future studies.

4 Signaling Proteins: Roles of Tether Length and Geometry

At this point, we are in a position to apply the tools from Sections 2 and 3 to the investigate the roles of tether length and geometry on the behavior of signaling proteins. First, we will discuss the case of simple switches, which contain a single ligand-receptor pair. Then, we will describe the cooperative integration of signals in more complex switches which contain multiple ligand-receptor pairs.

4.1 Simple Switches

Our motivating example for the statistical mechanical model are the reprogrammed WASPs that link Arp2/3 dependent actin polymerization to chemical signals. In particular, we are interested

in predicting how the properties of the tethers that connect the different domains influence the function of these proteins. To obtain such a prediction requires a number of model parameters. Fortunately, a number of studies exist on the structural biology of WASPs and the physical nature of amino acid chains that provide us with physical parameters that can be used to render the model predictive. Using the rule of thumb that one amino acid contributes ~ 0.4 nm to a protein's contour length, we estimate the flexible portions of the VCA domain to be ~ 21.6 nm in length. These portions are the black segments shown in figure 4a. The receptor, which in the case of the simple switch is a PDZ domain, is also thought to have unstructured residues at the C-terminus; we estimate these amino acids to be ~ 2.8 nm in length. For the simple switch studied by Dueber *et al.*, we assume these are the only contributions to the tether [17]. Experimental and theoretical evidence suggests unstructured amino acids have a persistence length of ~ 0.4 nm [40–43, 47].

The remaining unknown parameter is the size of the rigid portion of the VCA domain, D as indicated schematically in fig. 4. Recall that D arises in the model because the tethers for the receptor and ligand originate from different points. There are a number of reasons to expect this to be the case for the VCA domain. First, experiments have shown that N-WASP will bind Arp2/3 whether or not it is in an active conformation [22]. This data leads to the working assumption that the simple switch is always bound to Arp2/3. While this may not always be the case, only the subpopulation of Arp2/3's in solution that are bound to an N-WASP will contribute significantly to the creation of new actin filaments. Hence, we are justified in limiting the discussion to this case. Because the VCA domain of N-WASP is mostly unstructured, it is reasonable to ask where on Arp2/3 the VCA domain binds. Yeast two-hybrid experiments have shown that the acidic residues (i.e. the A domain in VCA) bind the protein p21, a member of the Arp2/3 complex [19]. The verprolin domain, or V domain, is known to bind actin monomers. Because the conserved domain, or C domain, has homology to the V domain, it has been proposed that it binds to the actin homolog Arp2 [47]. The five-fold decrease in the dissociation constant for VCA binding to Arp2/3 when the C domain is present supports this hypothesis [20]. If these two interactions exist, then they likely serve to hold the two tether origins fixed in space at separate points. This is illustrated in figure 4. Hence, the justification for applying this simple model to WASPs rests on the existence of these two interactions.

To find D , we turn to the titration experiments of the simple switch by Dueber *et al.* [17]. In this experiment, the relative activity, a measure of the speed up in actin polymerization caused by activated WASPs, was measured as a function of ligand concentration. We fit this experiment with fixed tether lengths by assuming that p_{on} is equal to the relative activity and using least-squares to find the effective concentration that best matched their data. With this fit, we estimate D to be ~ 7.94 nm. This value is close to the distance between Arp2 and p21 in the Arp2/3 complex, the two members of the Arp2/3 complex that have been proposed as binding sites for the C and A domains respectively. Finally, the dissociation constants for the receptor's interaction with both free and cis-ligand is taken to be $\sim 8 \mu\text{M}$ [17]. With these numbers, we can predict how the effective concentration (and hence the activity) changes with tether length.

The prediction for the tether length dependence of the simple switch is shown in figure 5. Lengthening short tethers increases effective concentrations while the opposite effect is seen for very long tethers. The crossover between the two regimes occurs at $L_T = \frac{D^2}{2\xi_T}$. This scaling

behavior arises because a portion of the output domain is treated as a rigid rod. To enter a looped state, the tethers must cross this rod, a task that is difficult when the tether length is comparable to the rod size D . In this regime, extending the tethers across the rod becomes easier with longer tethers. When the tethers are very long, the opposite is seen. Because the rod is short compared to tethers, it has a small effect on how often the two tether ends meet. If the rod is ignored, then our simple polymer model treats the whole construct as a random walk. Longer walks decrease the chance the walk ends at the beginning, implying that longer tethers will decrease the likelihood of PDZ seeing its cis-ligand. In particular, in this regime the effective concentration obeys the scaling law $J \propto L_T^{-3/2}$.

There are several points about this model that are worth noting. First, we are assuming that p_{on} is equivalent to the relative activity that is measured in experiments. While the relative activity is likely a monotonically increasing function of p_{on} , one cannot say the two are equivalent until this has been proven experimentally. The use of this assumption may undermine a quantitative interpretation of our results, though there are other cases such as transcriptional regulation where this kind of thinking is appropriate [48]. This assumption is a reasonable first step and allows us to use our model to interpret experimental results and generate predictions. Second, the short tether regime may be more relevant for experiments. Looping likely renders the construct inactive through a conformational change in the VCA domain. Very long tethers may decouple the connection between looping and switch inactivation. Indeed, some constructs with long tethers are constitutively active [17]. Short tethers avoid this issue and hence might be more useful experimentally. Third, our results depend on our assumptions about the physical properties of the VCA domain while bound to Arp2/3. If one of the putative interactions does not exist, then the VCA domain would remain unstructured while bound to Arp2/3, and the effective concentration would decrease monotonically with increasing tether length with a $-\frac{3}{2}$ scaling exponent. Finally, we note that the tether length prediction here will have interesting consequences for how constructs with multiple tethered receptor-ligand pairs behave in our model. Indeed, the rich behavior afforded by the case of multiple receptor-ligand pairs associated with the same overall tether will serve as the basis of our discussion below.

4.2 Complex Switches

In this section, we extend the statistical mechanical model described above to examine constructs with multiple receptor-ligand pairs, or complex switches. Such motifs might be one way in which evolution has expanded the signaling inventory by permitting multiple inputs to a given protein that activates other downstream components. The topologies of the complex switches under discussion are shown in figure 6. Complex switches accept multiple chemical inputs and exhibit interesting signal integration behavior. Some of the most interesting cases are reprogrammed N-WASP constructs that display antagonistic integration; one input is stimulatory while the other is inhibitory. Our goal is to demonstrate how tethers can play an active role in this decision making and propose physical mechanisms for how signaling proteins integrate multiple signals.

As with the simple switch, we can use statistical mechanics to construct a states and weights diagram for the two topologies in consideration; this is shown in figure 7. We restrict ourselves to the case where receptors only interact with ligands of the same color and where looping is the

mechanism of inhibition. We point the reader's attention to two important features of this extension. First, the topology of the complex switch, that is the arrangement of receptors and ligands on the tether, is crucial for determining what role tethers play in the switch's signal integration properties. In class I switches, the black receptor-ligand pair controls the length of the white receptor's tether. Because the white receptor-ligand pair controls the switch's activation, the behavior with respect to tether length is crucial for determining how the switch responds to white and black inputs. Conversely, class II switches do not share this topology. Interactions between the black receptor-ligand pair merely bring the white receptor-ligand pair into closer proximity (assuming the black receptor is smaller than the size of the cylinder). Hence, in the model, the switch's behavior is determined more by the origins of the tethers, a parameter governed by the secondary structure of the receptors, not the tether's physical properties.

Second, the model now contains three effective concentrations that must be calculated. These are

- The effective concentration of cis-ligand the black receptor sees when the white receptor/cis-ligand pair is unbound.
- The effective concentration of cis-ligand the white receptor sees when the black receptor/cis-ligand pair is unbound
- The effective concentration of cis-ligand the white receptor sees when the black receptor/cis-ligand pair is bound.

Calculating these effective concentration is not straightforward, as we can use structures from the PDB in some instances to abandon our point particle description of the receptors and ligands. Because of the physical size of the receptor, the tether ends for the receptor and cis-ligand are held at a fixed distance when the two species are bound; physical structures allow us to estimate this distance. In what follows, we return to our case study of reprogrammed N-WASP to demonstrate how this might be done.

4.2.1 Class I Switches

As with the simple switch, we can use statistical mechanics to find the statistical weight of each state. The states and weights are shown in Figure 7. Using these results, we can rewrite the pictorial representation of p_{on} algebraically as

$$p_{on} = \frac{1 + \frac{[L_B]}{K_B} + \frac{[L_W]}{K_W} + \frac{[L_B]}{K_B} \frac{[L_W]}{K_W} + \frac{J_1}{K'_B} + \frac{J_1}{K'_B} \frac{[L_W]}{K_W}}{1 + \frac{[L_B]}{K_B} + \frac{[L_W]}{K_W} + \frac{[L_B]}{K_B} \frac{[L_W]}{K_W} + \frac{J_1}{K'_B} + \frac{J_1}{K'_B} \frac{[L_W]}{K_W} + \frac{J_1}{K'_B} \frac{J_3}{K_W} + \frac{J_2}{K'_W} + \frac{[L_B]}{K_B} \frac{J_2}{K'_W}}. \quad (11)$$

Here J_1 is the effective concentration of cis-ligand seen by the black receptor when the white pair is unbound, J_2 is the effective concentration of cis-ligand seen by the white receptor when the black pair is unbound, and J_3 is the effective concentration of cis-ligand seen by the white receptor when the black pair is bound. Further, K_B is the dissociation constant for free black ligand, K'_B is the

dissociation constant for black cis-ligand, K_W is the dissociation constant for free white ligand, and K'_W is the dissociation constant for white cis-ligand.

In the reprogrammed N-WASP constructs designed by Dueber *et. al.*, an SH3 domain was used as the black receptor and a PDZ domain was used as the white receptor [17]. Fortunately, a crystal structure of SH3 allows us to relax our point particle approximation for the black receptor during our computation of J_1 . The crystal structure of an SH3 domain bound to its ligand is shown in figure 8.

As seen in the figure, a tether connecting the C-terminus of an SH3 domain to the N-terminus of a ligand would have its beginning and end separated by a distance $D_{SH3} \sim 2.14$ nm in the bound state. Hence, J_1 is given by the expression

$$J_1 = \left(\frac{3}{4\pi\xi_T L_2} \right)^{\frac{3}{2}} \exp \left[-\frac{3D_{SH3}^2}{4\xi_T L_2} \right]. \quad (12)$$

The calculation of J_2 and J_3 is more straightforward. When the black receptor/cis-ligand pair is unbound, the tether connecting the two contributes to the length of the white receptor's tether. The total length of the white receptor's tether is then $L_1 + L_2 + L_3$, and J_2 is given by

$$J_2 = \left(\frac{3}{4\pi\xi_T (L_1 + L_2 + L_3)} \right)^{\frac{3}{2}} \exp \left[-\frac{3D^2}{4\xi_T (L_1 + L_2 + L_3)} \right]. \quad (13)$$

When the black receptor/cis-ligand pair is bound, the tether connecting the two is looped out and does not contribute to the length of the white receptor's tether. The total length of the white receptor's tether is then $L_1 + L_3$, and J_3 is given by

$$J_3 = \left(\frac{3}{4\pi\xi_T (L_1 + L_3)} \right)^{\frac{3}{2}} \exp \left[-\frac{3D^2}{4\xi_T (L_1 + L_3)} \right]. \quad (14)$$

4.2.2 Class II Switches

As with class one switches, we can use the pictorial representation of class two switches to write p_{on} algebraically as

$$p_{on} = \frac{1 + \frac{[L_B]}{K_B} + \frac{[L_W]}{K_W} + \frac{[L_B][L_W]}{K_B K_W}}{1 + \frac{[L_B]}{K_B} + \frac{[L_W]}{K_W} + \frac{[L_B][L_W]}{K_B K_W} + \frac{[L_B]}{K_B} \frac{J_2}{K'_W} + \frac{J_2}{K'_W} + \frac{J_1}{K'_B} \frac{[L_W]}{K_W} + \frac{J_1}{K'_B} + \frac{J_1}{K'_B} \frac{J_3}{K'_W}}. \quad (15)$$

The effective concentrations and dissociation constants have the same meaning for class II switches as they had for class I switches.

The calculation of J_1 and J_2 is straightforward, as we can use the results from the simple switch with tether lengths L_1 and $L_1 + L_2$ respectively. These effective concentrations are given by

$$J_1 = \left(\frac{3}{4\pi\xi_T L_1} \right)^{\frac{3}{2}} \exp \left[-\frac{3D^2}{4\xi_T L_1} \right], \text{ and} \quad (16)$$

$$J_2 = \left(\frac{3}{4\pi\xi_T (L_1 + L_2)} \right)^{\frac{3}{2}} \exp \left[-\frac{3D^2}{4\xi_T (L_1 + L_2)} \right]. \quad (17)$$

To find J_3 , we can again appeal to a crystal structure of an SH3 domain bound to its ligand. When SH3 is bound to its cis-ligand, the tether for the white receptor, in this case PDZ, originates from the N-terminus of the SH3 domain and the tether for the white ligand, in this case PDZ ligand, originates from the C terminus of the SH3 domain's cis-ligand. The tether origins in this case are separated by a distance D_{PDZ} , and J_3 is consequently given by the expression

$$J_3 = \left(\frac{3}{4\pi\xi_T L_2} \right)^{\frac{3}{2}} \exp \left[-\frac{3D_{PDZ}^2}{4\xi_T L_2} \right]. \quad (18)$$

The model parameters used for the case study on reprogrammed N-WASPs are in table 1.

4.2.3 Applications to reprogrammed N-WASP

To illustrate how tethers can play an active role in signal integration, we used our model to analyze two classes of reprogrammed N-WASP constructs engineered by Dueber *et al.* [17]. Curiously, the constructs corresponding to class I switches display antagonistic signal integration in experiments; the results for the application of our model to these switches are shown in figure 9. As seen in the figure, the simple switch's scaling behavior with tether length provides a physical mechanism for the antagonistic gating seen in these constructs. Because we assume looping is the mechanism for inhibition, only binding between the white receptor/cis-ligand pair will be inhibitory. The role of the interaction between the black receptor/cis-ligand pair is to tune the length of the tether for the white receptor. As shown in the case study for the simple switch, lengthening the white receptor's tether will increase its local concentration of cis-ligand. In the absence of free ligand, the black receptor is bound to its cis-ligand with high probability. The addition of black ligand to the solution competitively removes this interaction, allowing the tether connecting the black receptor/cis-ligand pair to be added to the length of the white receptor's tether. As a result, changes in tether length are already a part of the natural repertoire of these signaling motifs. The effect of this change in tether length on the binding of the black receptor/cis-ligand pair is shown in figure 9b. When the interaction between the black receptor and its cis-ligand is removed, the effective concentration the white receptor sees of its cis-ligand slides up the curve to a higher point for the parameters corresponding to the constructs engineered by Dueber *et al.* [17]. The white receptor's cis-ligand functions as an inhibitor, so removing the black receptor/cis-ligand interaction represses the switch. Because the role of free black ligand is to competitively remove that interaction, inputs through the black receptor are antagonistic.

The topology of class I switches afford tethers the opportunity to play an active role in how the two signals are processed. As seen in Figure 5, there is a hump in the curve of effective concentration versus tether length. Where the system sits on the curve determines how class I switches will integrate the white and black signals. For the constructs engineered by Dueber *et al.*, the parameters placed the system on the increasing part of the curve, where longer tethers lead to higher effective concentrations. This model predicts that for these constructs, lengthening the tether connecting the black receptor and its cis-ligand (L_2) will lead to stronger antagonistic integration. Finally, we note that if the system were on the decreasing part of the curve (e.g. by employing a large L_2), then the model predicts cooperative integration. In other words, both black and white ligands activate the switch.

The results for the application of our model to reprogrammed N-WASP constructs corresponding to class II switches is shown in figure 10. These constructs have been shown to exhibit cooperative integration in experiments, and can behave as either AND or OR gates. As seen in figure 10, the model displays the cooperative integration seen in experiments. Both black and white ligands activate the switch, with maximum activation achieved when both ligands are present. Mechanistically, the tether length dependence plays a less prominent role in the behavior of switches with this topology. Recall that for this topology, both binding of either black or white cis-ligand to the appropriate receptor is inhibitory. The cooperativity in the model arises because the black receptor/cis-ligand interaction controls the location of the tether origins for the white receptor and its cis-ligand. Binding of the black receptor/cis-ligand pair brings the tether origins for the white receptor/cis-ligand pair closer together, causing the white receptor to see a higher effective concentration of its cis-ligand. Binding of free black ligand not only removes an inhibitor but reduces the effective concentration the white receptor sees of its cis-ligand, making it easier for free white ligand to bind and activate the switch.

4.2.4 SH3/cis-ligand interactions

The case study on reprogrammed N-WASP is useful for demonstrating how tethers can be involved in signal integration. However, during our analysis, we noted a discrepancy between theory and experiment for the interaction strength between SH3 and its cis-ligand. Experimentally, class I switches with $K'_B = 10 \mu\text{M}$ display antagonistic integration and switches with $K'_B = 660 \mu\text{M}$ only respond to inputs through the white receptor; the reverse is seen in our model. Two possible sources for this discrepancy are the SH3/cis-ligand interaction strength or the effective concentration SH3 sees of its cis-ligand. If either are reduced by two orders of magnitude, then our model qualitatively agrees with the experimental data.

Deciphering which term is the root of the discrepancy is non trivial. Concentrations are scaled by dissociation constants in the theory, making the two cases indistinguishable with the current data. Further, we found the same issue in each construct we analyzed that contained an SH3 domain. A simple switch made from an output domain and SH3 recently constructed by Dueber *et al.* is best described by an effective concentration of $\sim 0.2 \mu\text{M}$ for SH3's cis-ligand [49]. This is to be compared to $\sim 50 \mu\text{M}$ for the simple PDZ switch. It is difficult to explain two order of magnitude difference by a larger distance between the two tether origins, as such a distance is larger than the physical dimensions of the Arp2/3 complex. The behavior of the SH3 domain remains a mystery; one possible explanation is that the SH3/cis-ligand interaction strength is context dependent. The dissociation constants for SH3's interaction with free ligand has been measured; the strength of this interaction may differ if the ligand is part of a larger construct [50–52].

Alternatively, the excluded volume effects alluded to in section 3 may play a significant role. The presence of large macromolecules like Arp2/3 may influence the conformations available to the tether, leading to a decrease in the effective concentration of cis-ligand. Recent work by Krishnamurthy *et. al.* has suggested that such an effect is responsible for an order of magnitude decrease in the effective concentration of cis-ligand seen by human carbonic anhydrase in a tethered receptor-ligand system [8]. Whether this phenomenon could cause a two order of magnitude effect is unclear. A new round of experiments, similar to the work of Krishnamurthy *et. al.*, that explores

the tether length dependence of SH3/cis-ligand binding would be useful in exploring this issue [8]. While much stands to be learned when theory and experiment disagree, the resolution of this discrepancy is beyond the scope of our work. The key qualitative point to emerge from these calculations which transcends the details of our choice of parameters is the synergy between the different inputs that can exist for a single tether.

5 Discussion

We have used statistical mechanics and a random walk model to quantitatively explore the general theme of proteins with tethered receptor-ligand pairs. Through a case study on synthetic WASPs, we illustrated how tethers can influence protein-mediated decision making. We applied our model to constructs where multiple receptor-ligand pairs are connected by the same tether. In this case, we saw that the expected tether length dependence when the tether origins of the receptor and cis-ligand are separated provides a physical mechanism for antagonistic integration. These results demonstrate how tethers can play an active role in signal integration

There are a number of directions for future work. With regards to the case study, our assumptions about the binding of the VCA domain to Arp2/3 should be tested. The flexibility of regions assumed to be unstructured in our work can be ascertained with force-extension experiments. We have generated several predictions relevant to synthetic WASPs, namely that

- **Simple Switch:** For short tethers, an increase in tether length leads to higher effective concentrations.
- **Antagonistic Switch:** Lengthening the tether between SH3 and its cis-ligand will lead to stronger antagonistic integration.

Both of these predictions can be tested experimentally. We suggest using FRET measurements or fluorescence quenching as a more direct assay of probabilities [36, 37, 53]. Third, the role excluded volume plays in this system, particularly with regards to the SH3 domain, should be measured with tether lengthening experiments akin to the work by Krishnamurthy *et al* [8].

We end by noting that biochemistry on a leash is a thread that occurs throughout biology. Flexible tethers have been proposed to play a role in numerous processes, as mentioned above. A combination of simple physical models with experiments that vary tether lengths may lead to a clearer picture of the role tethers play in these systems.

6 Appendix: A Kinetic Model for the Simple Switch

In Section 2 we developed a statistical mechanical model to compute the probability that a switch is active by invoking the assumption of thermodynamic equilibrium. Before applying this model to a real biological system, it is important to understand the limits of this assumption. The argument was built around the separation between the time scales of ligand binding and a productive subsequent activation reaction. Here, we give this argument a proper quantitative underpinning by examining a kinetic model of the simple switch.

Our kinetic model is summarized pictorially in figure 11. Ligands free in solution bind and unbind to the receptor with rate constants k_+ and k_- . The cis-ligand binds and unbinds to the receptor with rate constants k_c and k_- . In this model, we assume that ligand binding is diffusion limited. That is, the on rate is determined by how long it takes a ligand to come into contact with the receptor; this time is dependent on whether the ligand is free or if it is connected to a tether. The off rate is determined by the strength of the receptor-ligand interaction. Note that we treat the free ligand and cis-ligand as if they were identical and assign them the same off rate. As argued earlier, the switches are coupled to a slow productive reaction; in the case of WASPs, this is actin polymerization. This coupling is modeled by the final reaction; switches in the active state are depleted at a rate R to generate product.

It is straightforward to use chemical kinetics to convert our pictorial representation in figure 11 into a set of rate equations. The full linear set of kinetic equations that governs the system is

$$\frac{dS}{dt} = k_- S_L + k_- S_T - k_c S - k_+ S L - R S, \quad (19)$$

$$\frac{dS_L}{dt} = k_+ S L - k_- S_L - R S_L, \quad (20)$$

$$\frac{dS_T}{dt} = k_c S - k_- S_T, \quad (21)$$

$$\frac{dP}{dt} = R S + R S_L, \quad (22)$$

where S is the concentration of free switches, S_L is the concentration of switches bound to free ligand, S_T is the concentration of switches bound to tethered ligand, L is the concentration of free ligand, and P is the concentration of product. Note that we assume the concentration of free ligand is high and is essentially constant. Also note that in the above equations, we have neglected the spatial variations of the concentrations. This is a reasonable assumption if the system is initially well-mixed.

The goal of our analysis is to compare a full treatment of the dynamics of the kinetic equations and determine when, if at all, an effective equilibrium approximation might hold. To this end, consider first the case $R = 0$. Starting from an initial condition of S_0 free switches, the characteristic relaxation time $\tau_{rel} \sim 1/\min[k_-, k_+ L, k_c]$ governs the relaxation rate of the system towards thermal equilibrium. Once in equilibrium, the concentrations satisfy

$$S = p_S S_{tot}, \quad (23)$$

$$S_L = p_{S_L} S_{tot}, \quad (24)$$

$$S_T = p_{S_T} S_{tot}, \quad (25)$$

where $S_{tot} = S + S_L + S_T$ is the concentration of all switch species present in the system, and the

probabilities p_S , p_{S_L} , and p_{S_T} are given by

$$p_S = \frac{1}{1 + \frac{L}{K_d} + \frac{\frac{k_c}{k_+}}{K_d}}, \quad (26)$$

$$p_{S_L} = \frac{\frac{L}{K_d}}{1 + \frac{L}{K_d} + \frac{\frac{k_c}{k_+}}{K_d}}, \text{ and} \quad (27)$$

$$p_{S_T} = \frac{\frac{\frac{k_c}{k_+}}{K_d}}{1 + \frac{L}{K_d} + \frac{\frac{k_c}{k_+}}{K_d}}. \quad (28)$$

Now, when $R > 0$, we would like to know under which condition(s) we may regard the switches as being effectively in equilibrium, that is, $S_i(t) \approx p_i S_{tot}(t)$ where $i = [S, S_L, S_T]$. The approximate kinetic equations are given by

$$\frac{dS}{dt} = p_S \frac{dS_{tot}}{dt}, \quad (29)$$

$$\frac{dS_L}{dt} = p_{S_L} \frac{dS_{tot}}{dt}, \quad (30)$$

$$\frac{dS_T}{dt} = p_{S_T} \frac{dS_{tot}}{dt}, \quad (31)$$

$$\frac{dS_{tot}}{dt} = -R(p_S + p_{S_L}) S_{tot}, \text{ and} \quad (32)$$

$$\frac{dP}{dt} = R(p_S + p_{S_L}) S_{tot}. \quad (33)$$

That is, the relative ratios of S , S_L , and S_T remain fixed during the time evolution of the system. For brevity, we refer to the full kinetic model as FKM and the kinetic model with effective equilibrium as EKM.

Our strategy to find the limits of the equilibrium assumption is to compare the exact description with our expectation from equilibrium statistical mechanics. To do so, we numerically integrate the differential equations subject to a particular choice of initial conditions and compare the resulting concentrations of the various species. This integration requires values for the k_+ , k_- , k_c , and R . To find k_+ and k_- , we turned to the experiments of Gianni *et al.* and Dueber *et al.* [17, 35]. For PDZ domains, Gianni *et al.* have measured k_+ and find $k_+ \approx 8 \mu\text{M}^{-1}\text{s}^{-1}$ [35]. Their work also suggests that differences in K_d 's among different PDZ's and ligands can be attributed primarily to the off rate, k_- . As a result, we can use the measured value of k_+ in conjunction with the measured K_d to back out the value of k_- . Using the equation

$$k_- = K_d k_+, \quad (34)$$

we find that $k_- = 64 \text{ s}^{-1}$ to match $K_d = 8 \mu\text{M}$ measured by Dueber *et al.* [17].

Determining the value of k_c is more involved. Indeed, the details of the loop formation kinetics are described in the next section and here we spell out the conceptual basis for determining this

rate constant. This parameter has two components. The first component is a collision rate, k_{col} , that describes how often the cis-ligand and receptor collide. The second component is a reaction probability, ρ , that describes how often binding occurs given a collision. This component is likely independent of the properties of the tether. If the two components are independent, then we can approximate the reaction rate as

$$k_c = \rho k_{col}. \quad (35)$$

k_{col} is calculated in the following section; for a tether length of 24.4 nm, we find that $k_{col} = 1.79 \times 10^6 \text{ s}^{-1}$. To estimate ρ , we again turn to the experiments of Gianni *et al.* [35]. Consider one receptor in a solution of ligands. The idea is to use the tether-free experiment to estimate parameters that we will later use to describe the kinetics of the tethered ligand. In particular, we need to determine the unknown constant ρ . Their data shows the reaction rate, k_+ , is $8 \mu\text{M}^{-1}\text{s}^{-1}$. Again, we can separate this into two components, implying

$$k_+ = \rho k_{col}^+. \quad (36)$$

We can compute the collision rate, k_{col}^+ for this reaction using the equation

$$k_{col}^+ = 4\pi D_L b, \quad (37)$$

where D_L is PDZ-ligand's diffusion coefficient in water and b is its approximate radius. This is a well known result from Smoluchowski [54]. We set $b = 1 \text{ nm}$ to estimate the size of a PDZ domain. We can estimate D_L using the Stokes-Einstein relation,

$$D_L = \frac{k_B T}{6\pi\eta r_L}, \quad (38)$$

where η is the viscosity of water and r_L is the radius of the ligand. We take $\eta = 10^{-3} \text{ kg m}^{-1}\text{s}^{-1}$ and $r_L = .565 \text{ nm}$ to obtain $D_L = 3.86 \times 10^{-10} \text{ m}^2\text{s}^{-1}$. These values give $k_{col}^+ = 2.92 \times 10^3 \mu\text{M}^{-1}\text{s}^{-1}$, implying that $\rho = 2.74 \times 10^{-3}$. Using this value for ρ , we estimate $k_c = 4900 \text{ s}^{-1}$.

The parameter R is the nucleation rate for new actin filaments. To estimate R , we turned to experiments performed by Zalvsky *et. al.* [55]. In their work, they measured the creation rate for new filaments and fit their work to a four step kinetic model. In this model Arp2/3, N-WASP, and actin come together to form a complex, which is then activated at a given rate to produce new actin filaments. This activation rate is identical to our definition of R ; for N-WASP, they found $R = 0.034 \text{ s}^{-1}$ [55]. Alternatively, the ATP hydrolysis rate on Arp2 can serve as a surrogate for the nucleation rate. Work by Dayel *et. al.* has shown that the VCA domain feeds Arp2/3 its first actin monomer and triggers ATP hydrolysis on Arp2 [24]. They have also demonstrated that the creation of new filaments proceeds at a rate close to the rate of ATP hydrolysis. Assuming that ATP hydrolysis and the creation of a new filament go hand-in-hand, the ATP hydrolysis rate can be a surrogate for the nucleation rate. Dayel *et. al.* estimate that a lower bound for the ATP hydrolysis rate is $R \sim 0.1 \text{ s}^{-1}$. While the use of either value is reasonable, precision is not important. As long as R captures the appropriate time scale, then it will suffice to determine whether we can treat the system with an equilibrium model. The parameters used in the simulation are summarized in table 2.

Computing the collision rate — To compute k_{col} , we turn to analytic methods that have been previously developed [56–58]. The outline of this calculation is as follows. First, we find the probability distribution of the distance between the receptor and its cis-ligand. This distribution is then used as an effective entropic potential that governs the dynamics of the cis-ligand. Next, we treat the cis-ligand as a point particle that undergoes one-dimensional diffusion inside this potential. Its behavior is described by a one-dimensional Fokker-Planck equation. We then find the mean first passage time, which is defined as the average time for the ligand to first come within a short distance of the receptor provided it starts at the bottom of the potential well. The inverse of the mean first passage time is our estimate for k_{col} .

To compute the probability distribution for the end-to-end distance between the receptor and its cis-ligand, we return to the calculation of the effective concentration presented in the main text. The effective concentration is the probability density of the cis-ligand in the vicinity of the receptor. To compute the new quantity of interest, we want the probability density of the cis-ligand at some vector \vec{R} away from the receptor. Let receptor's tether begin at \vec{r}_1' and end at \vec{r}_1 and let the cis-ligand's tether begin at \vec{r}_2' and end at \vec{r}_2 . Let us continue to describe the tethers with the random walk model. The probability density of interest is given by

$$G(\vec{R}) = \int d^3 r_1 d^3 r_2 G(\vec{r}_1 - \vec{r}_1'; \xi_T, L_1) G(\vec{r}_2 - \vec{r}_2'; \xi_T, L_2) \delta^3(\vec{R} - (\vec{r}_2 - \vec{r}_1)), \quad (39)$$

$$= \int d^3 x_1 G(\vec{r}_1 - \vec{r}_1'; \xi_T, L_1) G(\vec{R} + \vec{r}_1 - \vec{r}_2'; \xi_T, L_2), \quad (40)$$

$$= \left(\frac{3}{4\pi L_T \xi_T} \right)^{\frac{3}{2}} \exp \left[-\frac{3(\vec{R} - \vec{D})^2}{4L_T \xi_T} \right], \quad (41)$$

where $\vec{D} = \vec{r}_2' - \vec{r}_1'$ and $L_T = L_1 + L_2$. The probability distribution of the end-to-end distance is then

$$G(R) = R^2 \int d\Omega G(\vec{R}), \quad (42)$$

$$= R^2 \int_0^{2\pi} d\phi \int_{-1}^1 d(\cos \theta) \left(\frac{3}{4\pi L_T \xi_T} \right)^{\frac{3}{2}} \exp \left[-\frac{3}{4L_T \xi_T} (R^2 + D^2 - 2RD \cos \theta) \right], \quad (43)$$

$$= \left(\frac{3}{4\pi L_T \xi_T} \right)^{\frac{1}{2}} \frac{2R}{D} \sinh \left(\frac{3RD}{2L_T \xi_T} \right) \exp \left[-\frac{3}{4L_T \xi_T} (R^2 + D^2) \right]. \quad (44)$$

This distribution can be converted into an effective entropic potential using the equation

$$U_{eff}(R) = -k_B T \log G(R) + U_0, \quad (45)$$

where U_0 is a constant defined such that $\min U_{eff}(R) = 0$. The shape of this potential using $D = 7.94$ nm, $\xi_T = 0.4$ nm, and $L_T = 24.4$ nm is shown in figure 12. It is important to note that the tethered ligand will behave like it moves in this entropic potential only if we look at time scales that are long compared to the time scale of the tether's mechanical relaxation. A comparison of the

contact rate for flexible amino acid chains obtained through fluorescence quenching measurements and the theory described by Szabo *et. al.* and Hyeon *et. al.* suggest that this is indeed the case and this approach is valid [36, 37, 56, 58].

Because we assume the ligand undergoes one-dimensional diffusion in a potential, the probability distribution of its location is governed by the Fokker-Planck equation,

$$\frac{\partial p(R, t)}{\partial t} = D_0 \frac{\partial}{\partial R} \left(\frac{\partial}{\partial R} p + \beta \frac{\partial U_{eff}}{\partial R} p \right), \quad (46)$$

where p is the probability distribution of the ligand and D_0 is the cis-ligand's apparent diffusion constant. Because we assume the receptor's location is fixed at the origin, D_0 is actually the sum of both the cis-ligand's and the receptor's diffusion constant. The mean first passage time is given by the equation

$$\tau = \frac{1}{D_0} \int_a^r dR' \int_{R'}^{L_T} dR e^{\frac{U_{eff}(R') - U_{eff}(R)}{k_B T}}, \quad (47)$$

where a is the end-to-end distance of the tether at which the receptor and its cis-ligand are considered bound and r is the location of the potential's minimum [58]. We take a to be 2 nm, the approximate diameter of a PDZ domain. We take D_0 to be $6.04 \times 10^8 \text{ nm}^2\text{s}^{-1}$, the sum of D_L and the diffusion constant of a sphere of radius 1 nm in water. We computed this integral numerically using MATLAB to find $k_{col} = 1.79 \times 10^6 \text{ s}^{-1}$.

We numerically integrated the differential equations for the FKM and EKM using the MATLAB function `ode15s`. For the full model (FKM), we chose initial conditions that matched the concentrations used in the work of Dueber *et al.* [17]. These were $S = 50 \text{ nM}$, $S_L = 0 \text{ nM}$, $S_T = 0 \text{ nM}$, $L = 200 \mu\text{M}$, and $P = 0 \text{ nM}$. The results are shown in figure 13a, where we plot the probability for the switch to be in either of its three states as a function of time. As shown in figure 13a, using these initial conditions results in an initial transient during which the concentrations of S_L and S_T rapidly increase. Following the initial transient, the probabilities remain constant for the duration of the simulation and are equal to the equilibrium probabilities set by equilibrium statistical mechanics. This behavior is independent of the initial conditions, as the system will display the same behavior provided the total concentration of all switch species, concentration of ligand, and rate constants remain the same.

We conclude that after the initial transients, the system enters a dynamic regime that is well characterized by the model invoking effective equilibrium. Hence, for our case study, the separation of timescales allows us to apply an equilibrium statistical mechanical model to a non-equilibrium system. The kinetic model also provides a vehicle to examine when this is not the case. During the initial transients, the three states are not in effective equilibrium. If actin polymerization occurred on a time scale comparable to or faster than the other rate constants, as shown in figure 13b, then the behavior of the system differs from the behavior based on equilibrium statistical mechanics considerations.

7 Acknowledgements

We thank Lacramioara Bintu, John Dueber, Hernan Garcia, Paul Grayson, Alexander Grosberg, Jané Kondev, Wendell Lim, Tharathorn Rimchala, Boo Tseng, Paul Wiggins, and Brian Yeh for helpful discussions. David Van Valen and Rob Phillips acknowledge the support of a Fannie and John Hertz Foundation Yaser Abu-Mustafa Fellowship, a UCLA/Caltech MSTP Fellowship, and an NIH Pioneer Award. Mikko Haataja acknowledges support from the National Science Foundation through NSF-DMR Grant No. 0449184.

References

- [1] Romero, S., C. Le Clainche, D. Didry, C. Egile, D. Pantaloni, and M.F. Carlier (2004) Formin is a processive motor that requires profilin to accelerate actin assembly and associated ATP hydrolysis. *Cell* **119**, 419–429.
- [2] Kovar, D.R., E.S. Harris, R. Mahaffy, H. Higgs, and T. Pollard (2006) Control of the assembly of ATP- and ADP-actin by formins and profilin. *Cell* **124**, 423–435.
- [3] Vavylonis, D., D. Kovar, B. O'Shaughnessy, and T. Pollard (2006) Model of formin-associated actin filament elongation. *Mol Cell* **21**, 455–466.
- [4] Nguyen, J. and W. Lim (1997) How Src exercises self-restraint. *Nat Struct Biol* **4**, 256–260.
- [5] Hubbard, S. (1999) Src autoinhibition: let us count the ways. *Nat Struct Biol* **6**, 711–714.
- [6] Gonfloni, S., J. Williams, K. Hattula, A. Weijland, R.K. Wierenga , and G. Superti-Furga (1997) The role of the linker between the SH2 domain and catalytic domain in the regulation and function of Src. *The EMBO Journal* **16**, 7261–7271.
- [7] Young, M., S. Gonfloni, G. Superti-Furga, B. Roux, and J. Kuriyan (2001) Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell* **105**, 115–126.
- [8] Krishnamurthy, V., V. Semetey, P. Bracher, N. Shen, and G. Whitesides (2007) Dependence of effective molarity on linker length for an intramolecular protein-ligand system. *J Am Chem Soc* **129**, 1312–1320.
- [9] Windisch, B., D. Bray, and T. Duke (2006) Balls and Chains - A Mesoscopic Approach to Tethered Protein Domains. *Biophys J* **91**, 2383–2392.
- [10] Zagotta, W., T. Hoshi, W. Aldrich (1989) Gating of single Shaker potassium channels in *Drosophila* muscle and in *Xenopus* oocytes injected with Shaker mRNA. *Proc Natl Acad Sci* **86**, 7243–7247.
- [11] Zagotta, W. and R. Aldrich (1990) Voltage-dependent gating of Shaker A-type potassium channels in Drosophila muscle. *J Gen Physiol* **95**, 29–60.

- [12] Hoshi, T., W. Zagotta, and W. Aldrich (1990) Biophysical and Molecular Mechanisms of *Shaker* Potassium Channel Inactivation. *Science* **250**, 533–538.
- [13] Zagotta, W., T. Hoshi, W. Aldrich (1990) Restoration of Inactivation in Mutants of *Shaker* Potassium Channels by a Peptide Derived from ShB. *Science* **250**, 568–571.
- [14] Timpe, L., and L. Peller (1995) A Random Flight Chain Model for the Tether of the *Shaker* K⁺ Channel Inactivation Domain. *Biophys J* **69**, 2415–2418.
- [15] Robinson, C., and R. Sauer (1998) Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. *Proc Natl Acad Sci USA* **95**, 5929–5934.
- [16] Nagi, A., and L. Regan (1997) An inverse correlation between loop length and stability in a four-helix-bundle protein. *Fold Des* **2**, 67–75.
- [17] Dueber, J., B. Yeh, K. Chak, and W. Lim (2003) Reprogramming control of an allosteric signaling switch through modular recombination. *Science* **30**, 1904–1908.
- [18] Higgs, H., L. Blanchoin, and T. Pollard (1999) Influence of the C-terminus of Wiskott-Aldrich Syndrome Protein (WASP) and the Arp2/3 complex on actin polymerization. *Biochemistry* **38**(46), 15212–15222.
- [19] Machesky, L. and R. Insall (1998) Scar1 and the related Wiskott-Aldrich syndrome protein, WASP, regulate the actin cytoskeleton through the Arp2/3 complex. *Curr Biol* **8**, 1347–1356.
- [20] Marchand, J.B., D. Kaiser, T. Pollard, and H. Higgs (2001) Interaction of WASP/Scar proteins with actin and vertebrate Arp2/3 complex. *Nat Cell Biol* **3**, 76–82.
- [21] Mullins, R.D. and L. Machesky (2000) Actin assembly mediated by Arp2/3 complex and WASP family proteins. *Methods Enzymol* **325**, 214–237.
- [22] Prehoda, K., J. Scott, R.D. Mullins, and W. Lim (2000) Integration of multiple signals through cooperative regulation of the N-WASP-Arp2/3 complex. *Science* **290**, 801–806.
- [23] Rohatgi, R., L. Ma, H. Miki, M. Lopez, T. Kirchhausen, T. Takenawa, and M. Kirschner (1999) The interaction between N-WASP and the Arp2/3 complex links Cdc42-dependent signals to actin assembly. *Cell* **97**, 221–231.
- [24] Dayel, M. and R.D. Mullins (2004) Activation of Arp2/3 complex: addition of the first subunit of the new filament by a WASP protein triggers rapid ATP hydrolysis on Arp2. *PLoS Biol* **2**, 476–485.
- [25] Kim, A., L. Kakalis, N. Abdul-Manan, G. Liu, and M. Rosen (2000) Autoinhibition and activation mechanisms of the Wiskott-Aldrich syndrome protein. *Nature* **404**, 151–158.
- [26] Ptashne, M. and A. Gann (2004) *Genes & Signals* Cold Spring Harbor Laboratory Press (Cold Spring Harbor, New York).

- [27] Hochschild, A., and M. Ptashne (1988) Interaction at a distance between λ repressors disrupts gene activation. *Nature* **336**, 353–357.
- [28] Joung, J.K., L.U. Le, and A. Hochschild (1993) Synergistic activation of transcription by Escherichia coli cAMP receptor protein. *Proc Natl Acad Sci USA* **90**, 3083–3087.
- [29] Joung, J.K., D.M. Koepp, and A. Hochschild (1994) Synergistic activation of transcription by bacteriophage λ cI protein and E. coli cAMP receptor protein. *Science* **265**, 1863–1866.
- [30] Dodd, I.B., K.E. Shearwin, A.J. Perkins, T. Burr, A. Hochschild, and J.B. Egan (2004) Co-operativity in the long-range gene regulation by the λ cI repressor. *Genes Dev* **18**, 344–354.
- [31] Ackers, G., A. Johnson, and M. Shea (1982) Quantitative model for gene regulation by λ phage repressor. *Proc Natl Acad Sci USA* **79**, 1129–1133.
- [32] Graham, I. and T. Duke (2005) The logical repertoire of ligand-binding proteins. *Phys Biol* **2**, 159–165.
- [33] Buck, M., W. Xu, and M. Rosen (2004) A two-state allosteric model for autoinhibition rationalizes WASP signal integration and targeting. *J Mol Biol* **338**, 271–285.
- [34] Zhou, H.X. (2006) Quantitative relation between intermolecular and intramolecular binding of pro-rich peptides to SH3 domains. *Biophys J* **91**, 3170–3181.
- [35] Gianni, S., A. Engstrom, M. Larsson, N. Calosci, F. Malatesta, L. Eklund, C. Ngang, C. Travagliini-Alcocatelli, and P. Jemth (2005) The kinetics of PDZ domain-ligand interactions and implications for the binding mechanism. *J Biol Chem* **280**, 34805–34812.
- [36] Hudgins, R., F. Huang, G. Gramlich, and W. Nau (2002) A fluorescence-based method for direct measurement of submicrosecond intramolecular contact formation in biopolymers: an exploratory study with polypeptides. *J Am Chem Soc* **124**, 556–564.
- [37] Wang, X., E. Bodunov, and W. Nau (2003) Fluorescence quenching kinetics in short polymer chains: Dependence on chain length. *Opt Spectrosc* **95**, 560–570.
- [38] Jacobson, H. and W. Stockmayer (1950) Intramolecular Reaction in Polycondensations. I. The Theory of Linear Systems. *J Chem Phys* **18**, 1600–1606.
- [39] Ito, K., J. Chuang, C. Alvarez-Lorenzo, T. Watanabe, N. Ando, and A. Grosberg (2003). Multiple point adsorption in a heteropolymer gel and the Tanaka approach to imprinting: experiment and theory. *Prog. Polym. Sci.* **28**, 1490–1515.
- [40] Rief, M., M. Gautel, F. Oesterhelt, J. Fernandez, and H. Gaug (1997) Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **276**, 1109–1112.
- [41] Rief, M., J. Fernandez, and H. Gaub (1998) Elastically coupled two-level systems as a model for biopolymer extensibility. *Phys Rev Lett* **81**, 4764–4767.

- [42] Lim, R., N.P. Huang, J. Koser, J. Deng, K.H.A. Lau, K. Schwarz-Herion, B. Fahrenkrog, and U. Aebi (2006) Flexible phenylalanine-glycine nucleoporins as entropic barriers to nucleocytoplasmic transport. *Proc Natl Acad Sci USA* **103**, 9512–9517.
- [43] Zhou, H.X. (2001) Loops in proteins can be modeled as worm-like chains. *J Phys Chem B* **105**(29), 6763–6766.
- [44] P. Flory (1969) Statistical Mechanics of Chain Molecules. John Wiley and Sons, Inc, New York, NY.
- [45] H. Seyed-allaei (2005) Three-bead rotating chain model shows universality in the stretching of proteins. *Phys Rev E* **72** 041908.
- [46] Segall, D., P. Nelson, and R. Phillips (2006) Excluded-Volume Effects in Tethered Particle Experiments: Bead Size Matters. *Phys Rev Lett* **96**, 088306.
- [47] Chereau, D., F. Kerff, P. Graceffa, Z. Grabarek, K. Langsetmo, and R. Dominguez (2005) Actin-bound structures of Wiskott-Aldrich syndrome protein (WASP)-homology domain 2 and the implications for filament assembly. *Proc Natl Acad Sci USA* **102**(46), 16644–16649.
- [48] Bintu, L., N. Buchler, H. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips (2005) Transcriptional regulation by the numbers: models. *Curr Opin Genet Dev* **15**, 116–124.
- [49] Dueber, J., E. Mirsky, and W. Lim (2007) Engineering synthetic signaling proteins with ultrasensitive input/output control. *Nat Biotechnol* **25**, 660–662.
- [50] Posern, G. *et. al.* (1998) Development of highly selective SH3 binding peptides for Crk and CRKL which disrupt Crk-complexes with DOCK180, SoS and C3G. *Oncogene* **16**, 1903–1912.
- [51] Knudsen, B., J. Zheng, S. Feller, J. Mayer, S. Burrel, D. Cowburn, and H. Hanafusa (1995) Affinity and specificity requirements for the first Src homology 3 domain of the Crk proteins. *EMBO J* **14**, 2191–2198.
- [52] Nguyen, J., C. Turck, F. Cohen, R. Zuckerman, and W. Lim (1998) Exploiting the basis of proline recognition by SH3 and WW domains: Design of N-substituted inhibitors. *Nature* **282** 2088–2092.
- [53] Ward, M., J. Wu, and Y. Rao (2004) Visualization of spatially and temporally regulated N-WASP activity during cytoskeletal reorganization in living cells. *Proc Natl Acad Sci USA* **101**, 970–974.
- [54] Smoluchowski, M. (1916) *Phys Z* **17** 557.
- [55] Zalvsky, J., L. Lempert, H. Kranitz, and R.D. Mullins (2001) Different WASP family proteins stimulate different Arp2/3 complex-dependent actin-nucleating activities. *Curr Biol* **11**, 1903–1913.

- [56] Szabo, A., K. Schulten, and Z. Schulten (1980) First passage time approach to diffusion controlled reactions. *J Chem Phys* **72**, 4350–4357.
- [57] Jun, S., J. Bechhoefer, and B. Ha (2003) Diffusion-limited loop formation of semiflexible polymers: Kramers theory and the intertwined time scales of chain relaxation and closing. *Europhys Lett* **64**, 420–426.
- [58] Hyeon C. and D. Thirumalai (2006) Kinetics of interior loop formation in semiflexible chains. *J Chem Phys* **124**, 104905–104918.
- [59] Robinson, R., K. Turbedsky, D. Kaiser, J.B. Marchand, H. Higgs, S. Choe, and T. Pollard (2001) Crystal strucure of Arp2/3 complex. *Science* **294**, 1679–1684.
- [60] Samuel, J. and S. Sinha (2002) Elasticity of semiflexible polymers. *Phys Rev E* **66**, 050801.
- [61] Wu, X., B. Knudsen, S. Feller, J. Zheng, A. Sali, D. Cowburn, H. Hanafusa, and J. Kuriyan (1995) Structural basis for the specific interaction of lysine-containing proline-rich peptides with the N-terminal SH3 domain of c-Crk. *Structure* **3**, 215–226.

Tables

Class 1		Class 2	
Parameter	Value	Parameter	Value
D	7.94 nm	D	7.94 nm
D_{SH3}	2.14 nm	D_{PDZ}	2.35 nm
L_1	21.6 nm	L_1	21.6 nm
L_2	7.2 nm	L_2	6.8 nm
L_3	6.8 nm		
K_B	0.1 μM	K_B	0.1 μM
K_W	8 μM	K_W	8 μM

Table 1: Model parameters for class I and class II complex switches. These parameters were taken from experiments performed by Dueber *et. al.* [17] and correspond to constructs H1-H4 for class one switches and H14-H15 for class two switches.

Parameter	Value
k_+	$8 \mu\text{M}^{-1} \text{s}^{-1}$
k_-	64s^{-1}
k_c	$4.9 \times 10^3 \text{s}^{-1}$
R	.034 s^{-1}

Table 2: Numerical values for the rate constants used in the kinetic model.

Figure Legends

- **Figure 1:** Biological examples of biochemistry on a leash. Receptors and ligands connected by a flexible tether are a common motif in biological systems. (a) Formin mediated actin polymerization [3]. In one model, formin proteins can use flexible tethers to grab actin monomers in solution and deliver them to the end of a growing actin filament. (b) Inactivation of Src tyrosine kinase (Courtesy: David Goodsell, Scripps Research Institute, La Jolla, CA). A flexible linker connects the SH2 domain and the tail portion of Src tyrosine kinase. It is thought that the binding of the phosphorylated tail to SH2 locks the protein in an inactive conformation. Theoretical studies suggest that the properties of the tether can influence the protein’s function. (c) Activation of Arp2/3 by synthetic WASPs [59]. WASPs link Arp2/3 mediated actin polymerization to chemical signals by controlling the delivery of the first actin monomer to the growing actin filament. Synthetic WASPs have an interesting structural motif where a receptor and ligand are connected by a flexible tether. The competition between the tethered ligand and ligands free in solution gives rise to varying levels of actin polymerization.

- **Figure 2:** States and weights for the simple switch. The tethered ligand functions as an inhibitor and locks the protein into an inactive state when bound to the receptor. Ligands in solution can compete with cis-ligand to activate the switch. Statistical mechanics assigns statistical weights to each state and allows us to compute the probability of the protein to be in an active state, as shown in the ratio at the bottom.
- **Figure 3:** A random walk model for flexible polypeptide chains. (a) In this model, the polypeptide chain is separated into statistically independent segments, called Kuhn segments, of length b . The orientation of two monomers is perfectly correlated if they are in the same segment and completely uncorrelated if they are in different segments. (b) The probability distribution for the end-to-end distance vector for a chain of length $L = 20\xi_p$ computed with both the wormlike chain and random walk models. The computation for the wormlike chain was implemented using the method of Samuel *et. al.* [60]. A comparison between the two polymer models shows that the random walk model is an acceptable approximation of the wormlike chain model for long chains.
- **Figure 4:** A physical model for the output domain of N-WASP. The VCA domain is thought to be a mixture of alpha helices and unstructured amino acids. (a) Separation of structure into residues assumed to have secondary structure and residues assumed to be unstructured. Here, cylinders represent stable alpha helices. (b) Model for binding of the VCA domain to Arp2/3 [59]. In our model, binding of VCA to Arp2/3 fixes the location of the C and A domains. This allows us to treat the C and A domains as a rigid cylinder and the rest of the VCA domain as a flexible tether.
- **Figure 5:** The tether length dependence of the simple switch. (a) The probability of the switch being active as a function of ligand concentration and tether length. The parameters used for the case study on reprogrammed N-WASP are described in the main text. Zero additional amino acids corresponds to the experimental construct with a tether length of ~ 24.4 nm. The additional amino acids used by Dueber *et al* consist of serine-glycine repeats. Assuming the relative activity is equivalent to p_{on} , we can use experimental data from Dueber *et al* to fit for unknown parameters in our model [17]. (b) The tether length dependence of the effective concentration of cis-ligand. We find that for short tethers, an increase in tether length leads to higher effective concentrations and hence a lower probability of occupying an active state. The reverse effect is seen in the model for very long tethers.
- **Figure 6:** The topologies of complex switches. This figure outlines the topologies of the complex switches discussed in this work and highlights the relevant tether lengths. The statistical mechanical model used to examine the simple switch can be extended to examine constructs with multiple tethered receptor-ligand pairs. Note that there are two distinct mechanisms for signal integration. For class one switches, the interaction between the black receptor-ligand pair serves to control the length of the white receptor's tether. The tether length dependence is then crucial for this construct's signal integration behavior. For class two switches, the interaction between the black receptor-ligand pair serves to bring the white

receptor-ligand pair into closer proximity. This likely increases the effective concentration of white cis-ligand and leads to cooperativity for the binding of the two inputs.

- **Figure 7:** The states and weights for class I and class II complex switches. Statistical mechanics can be used to compute the thermodynamic weight of each state and the probability the switch occupies an active state. A complete loop must be formed to inactivate the switch. (a) Class I switches have been observed to display antagonistic gating in experiments. (b) Class II switches display cooperative integration in experiments.
- **Figure 8:** NMR structure of the SH3 domain. We can estimate the effective concentration of cis-ligand seen by an SH3 domain in class one switches by appealing to a structure of the bound complex, identification number 1PRM in the Protein Databank [61]. This structure can also be used to compute the effective concentration of cis-ligand seen by a PDZ domain in class two switches.
- **Figure 9:** An activation profile for an antagonistic class I switch. For this profile, $K'_B = 660 \mu\text{M}$ and $K'_W = 8 \mu\text{M}$. The color represents p_{on} , with low values being blue and high values being black. When [white Ligand] is increased the switch is activated, reflected by a color shift of blue to yellow or orange. When [black Ligand] is increased, the switch is deactivated. This effect is best seen at high values of [white Ligand]. Because the role of the black receptor/cis-ligand pair is to determine the length of white receptor's tether, the simple scaling properties of the simple switch provide a physical mechanism for antagonistic integration. The removal of SH3's interaction with its cis-ligand lengthens the white receptor's tether. The longer tether length for L_2 increases the effective concentration of cis-ligand seen by the white receptor and represses the switch.
- **Figure 10:** An activation profile for class II switches. The color represents p_{on} , with low values being blue and high values being black. (a) This construct exhibits the same behavior as an AND gate; both inputs are necessary to achieve maximal activation. Dissociation constants used were $K'_B = 1000 \mu\text{M}$ and $K'_W = 8 \mu\text{M}$. (b) This construct has behavior consistent with an OR gate; either input is sufficient to achieve maximum activation. Dissociation constants used were $K'_B = 1000 \mu\text{M}$ and $K'_W = 100 \mu\text{M}$.
- **Figure 11:** A kinetic model of the simple switch. This toy model assumes simple chemical kinetics for all reactions. Ligands free in solution bind and unbind to the receptor with rate constants k_+ and k_- while the tethered ligand binds and unbinds with rate constants k_c and k_- . Switches in an active state are consumed at a rate R to generate product.
- **Figure 12:** The potential landscape for the cis-ligand. The probability distribution of the distance between the receptor and its tethered ligand can be used to define an entropic potential. The cis-ligand is then assumed to undergo one-dimensional diffusion inside this potential well. We then use the Fokker-Planck equation to find the average time needed for the cis-ligand to encounter the absorbing boundary assuming it starts at the bottom of the potential well. The inverse of this mean first passage time is an estimate for the collision rate.

- **Figure 13:** A simulation of the full kinetic model. (a) The probabilities and concentrations for S , S_L , and S_T as a function of time for $R = .034 \text{ s}^{-1}$. In the full kinetic model, the switches achieve a rapid pre-equilibria between the three states S , S_L , and S_T . After the initial transient, the probabilities for each of the three states is constant and equal to the probability predicted by equilibrium statistical mechanics. (b) The probabilities and concentrations for S , S_L , and S_T as a function of time for $R = 1000 \text{ s}^{-1}$. Because the timescale of the productive reaction is the same as that for ligand unbinding, the system never reaches an effective between the states of the simple switch. For this choice of parameters, equilibrium statistical mechanics considerations provide a poor description for the behavior of this system.

Figures

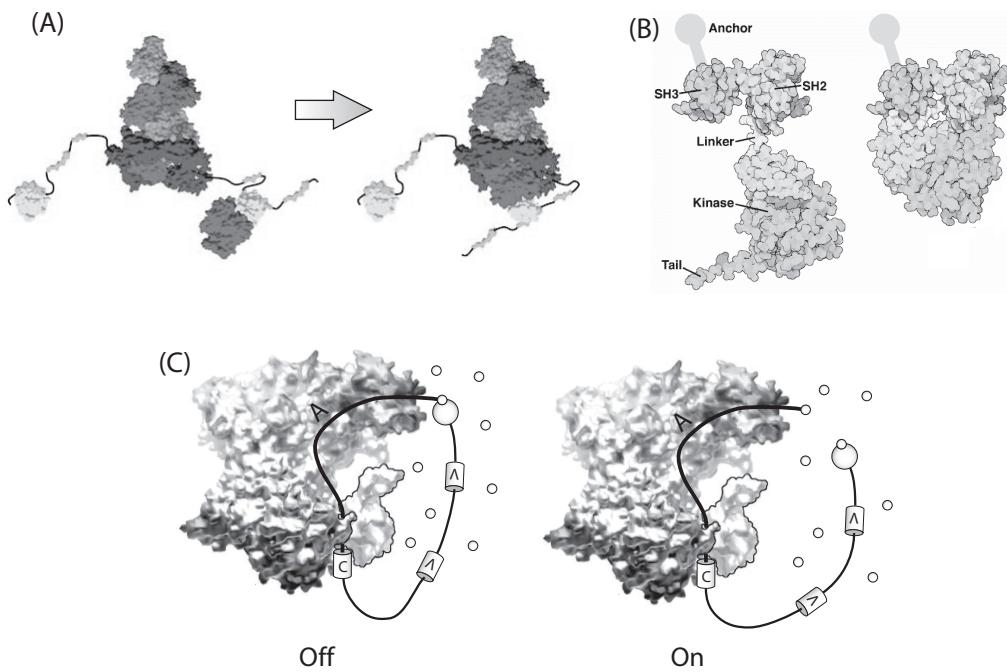


Figure 1

State	Output	Boltzmann Factors	Thermodynamic Weight
	On	$\Omega_{on}^{tet}\Omega_L^{sol}$	1
	On	$\Omega_{on}^{tet}\Omega_{L-1}^{sol}e^{-\beta\epsilon_f}$	$\frac{[L]}{K_d}$
	Off	$\Omega_{off}^{tet}\Omega_L^{sol}e^{-\beta\epsilon_t}$	$\frac{J}{K'_d}$
$p_{on} = \frac{\text{On states}}{\text{On states} + \text{Off state} + \text{Other states}}$			

Figure 2

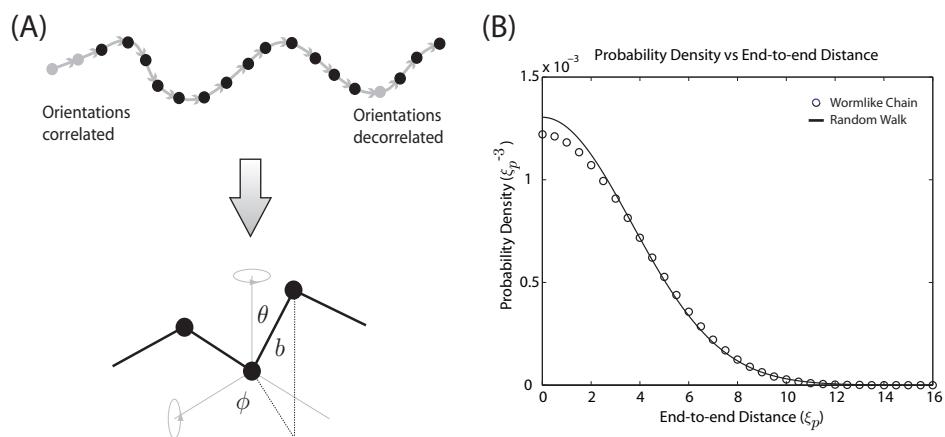


Figure 3

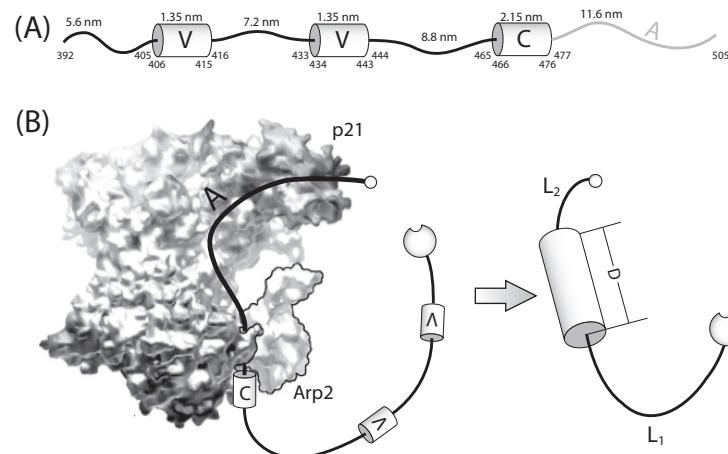


Figure 4

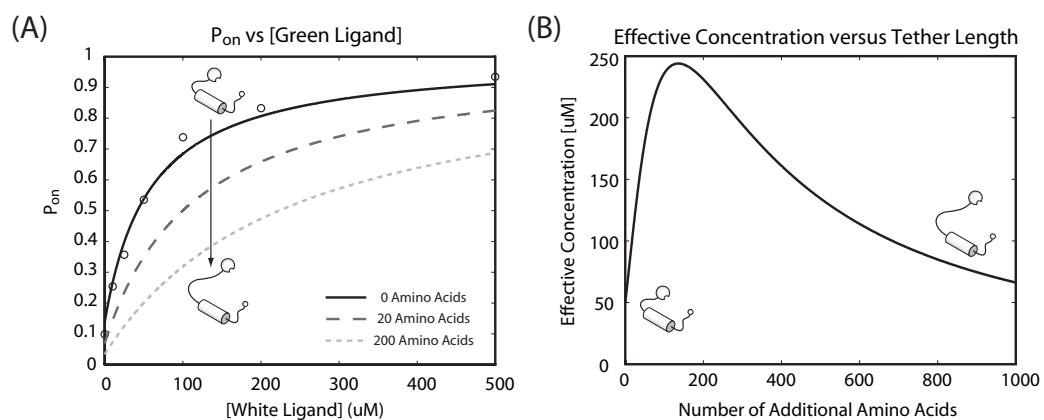


Figure 5

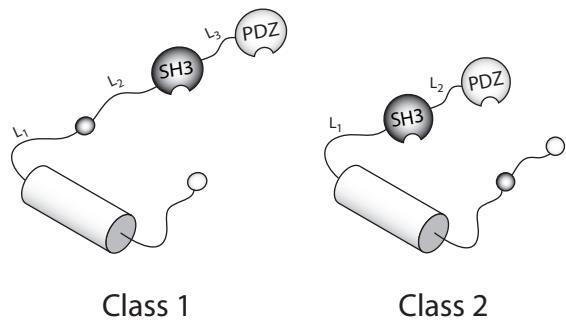


Figure 6

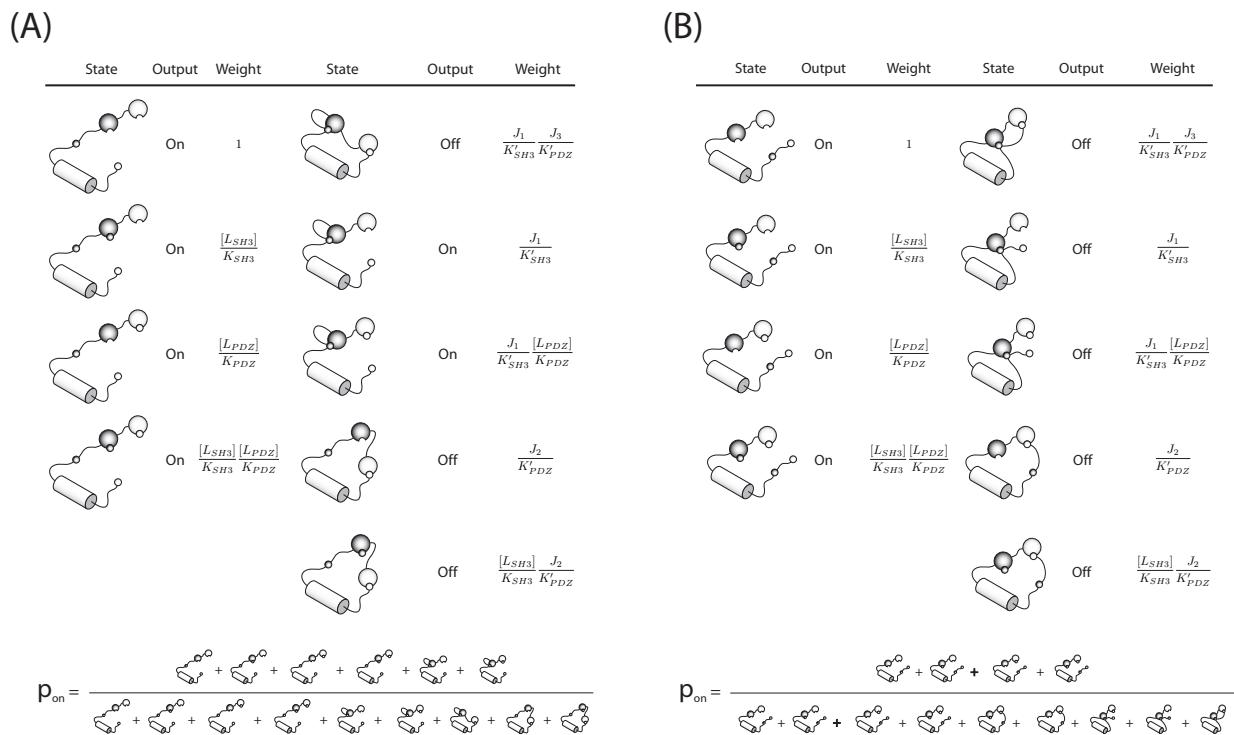


Figure 7

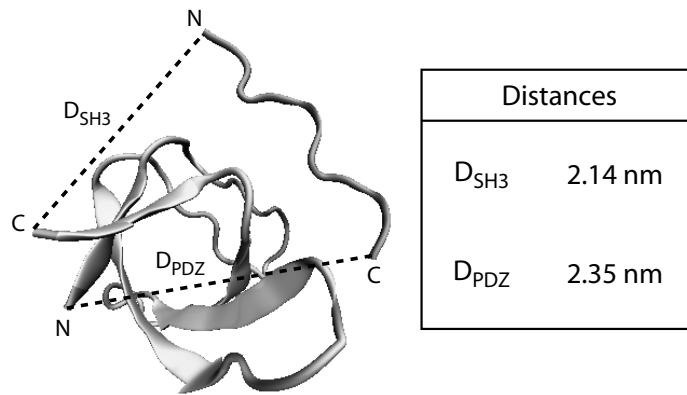


Figure 8

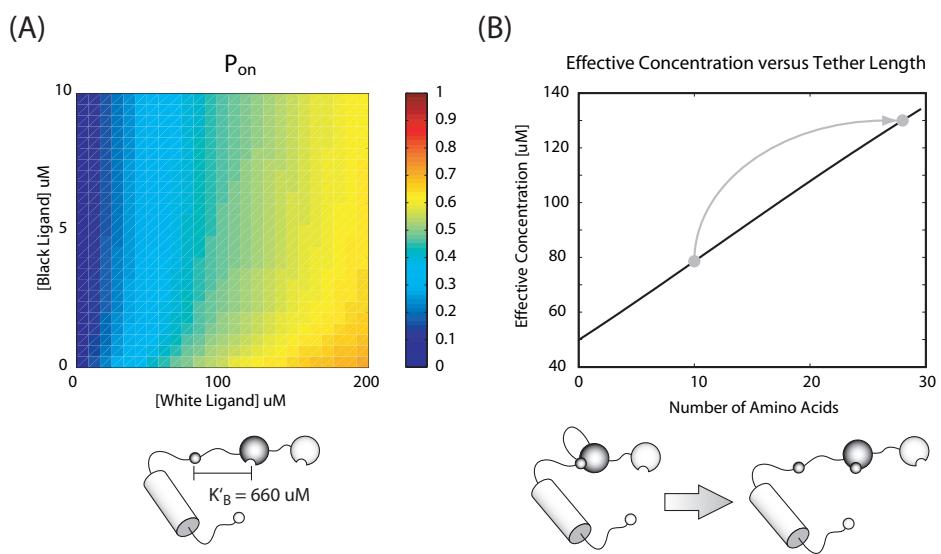


Figure 9

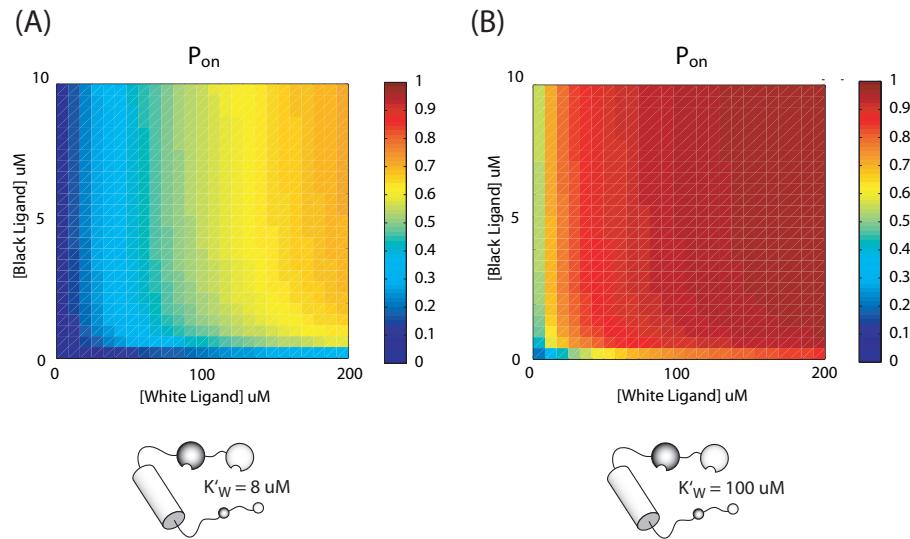


Figure 10

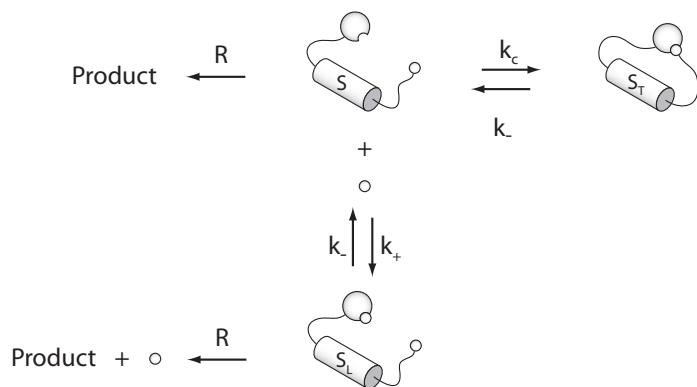


Figure 11

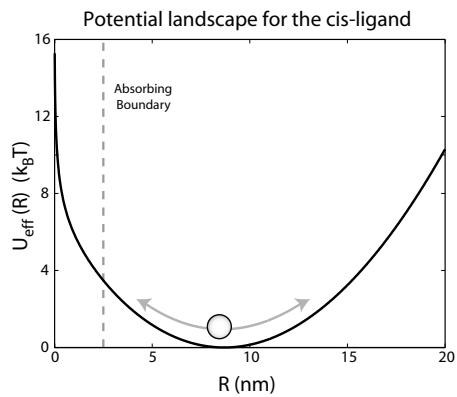


Figure 12

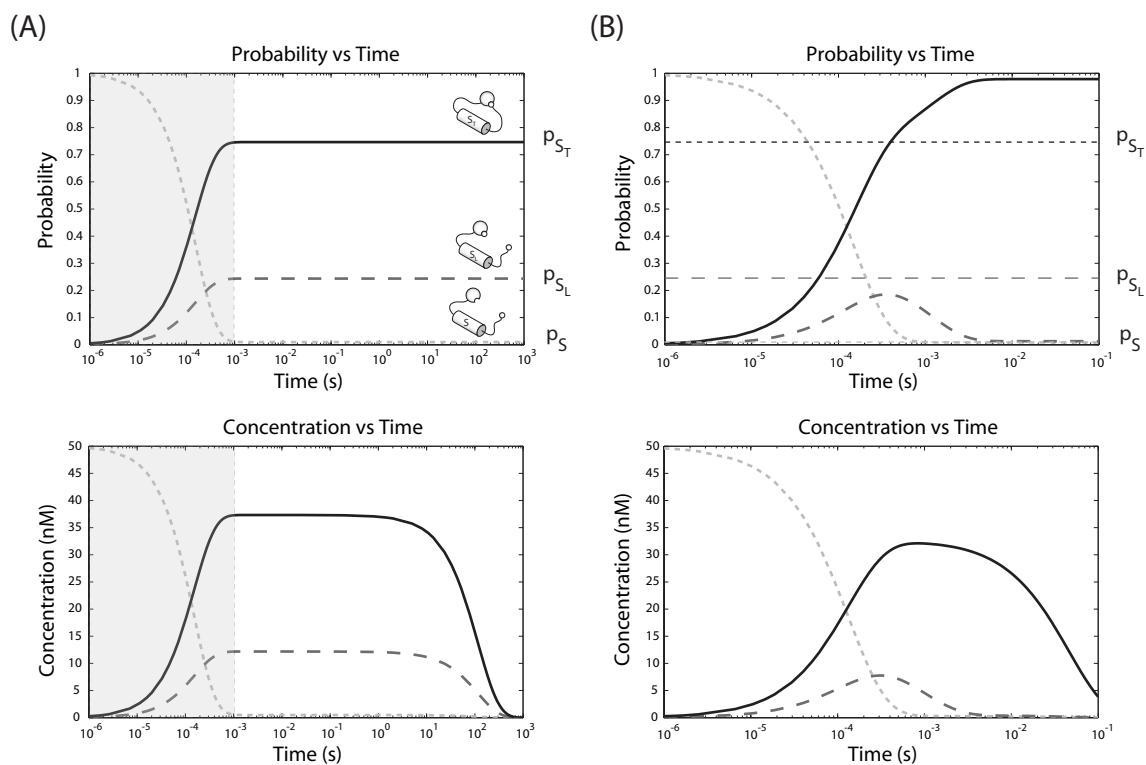


Figure 13