Mechanistic modeling identifies emergent properties of tandem SH2 domain interactions

Reagan Portelance[1], **Anqi Wu**[1], **Alekhya Kandoor**[1], **and Kristen M. Naegle**[1,✉]

[1]Department of Biomedical Engineering and the Department for Genome Sciences, University of Virginia, Charlottesville, Virginia, United States of America.

Tandem SH2 domains occur in key protein mediators of phosphotyrosine signaling and have the capacity to drive high affinity interactions through the avidity that results with bisphosphorylated protein partners. However, challenges have prevented the broad exploration of tandem SH2 domain avidity and here we utilize advances in both computational modeling and experimental approaches to predict and test tandem SH2 domain recruitment. Theoretical model behavior suggests that maximum avidity occurs with closely spaced or flexibly linked phosphotyrosine sites, combined with moderate monovalent affinities – exactly around the affinities of SH2 domains with individual phosphotyrosine sites. Surprisingly, despite sequence diversity, structure-based analysis showed remarkably conserved three-dimensional spacing between SH2 domains across all tandem SH2 families, which we interrogate experimentally, suggesting evolutionary optimization for avidity interactions. The combination of structure-based analysis of domain spacing with available monovalent experimental data appears to be sufficiently accurate to rank order predict high affinity interactions of tandem SH2 domain recruitment to the EGFR C-terminal tail. These approaches lay the groundwork for larger utility in multivalent prediction and testing to help better understand protein interactions that drive cell signaling.

SH2 domains | protein interactions | computational biology | phosphotyrosine | avidity | bivalent | biolayer interferometry | BLI

Correspondence: ***kmn4mj@virginia.edu***

Introduction

Avidity is the property that defines the effective binding strength of interactions that involve multiple monovalent interactions. Frequently, avidity is greater than the sum of the individual monovalent affinities, which occurs due to biophysical constraints where the initial interaction locks the unbound interface into a region of space constrained by their shared linker, thus increasing its "effective concentration" and, as a result, its binding driving force (1). It is a well-documented phenomenon in antibody binding, where multiple antibody epitopes boost affinity (2), which scientists have exploited for antibody-drug conjugates to increase therapeutic impact (3, 4). Cell signaling also relies on multivalent interactions to drastically increase sensitivity and specificity in network connections, and one subset of multivalent proteins – termed tandem SH2 proteins due to possessing two SH2 domains – is a strong example of this (5, 6). Recruitment of the tyrosine kinase ZAP70, a key regulator of T cell receptor activation, requires both phosphotyrosine (pTyr) sites in the immunoreceptor tyrosine-based activation motif (ITAM) sequences in the CD3ζ chain of the T-cell receptor. Not only are the two SH2 domains of ZAP70 recruited with 100-fold higher affinity than either SH2 domain individually, but deletion of one of ZAP70's SH2 domains prevents localization of the protein to the T-cell receptor (7, 8). This phenomenon has been well documented in other immune receptors that regulate the T-cell synapse. For example, SYK – the homolog of ZAP70 – gets recruited to Fc receptors for immunoglobin E (FcϵRI) with three times greater affinity than either individual domain, and phosphorylation within the SH2 linker has been shown to alter the the binding affinity, suggesting that the interaction occurs through avidity which is dependent on the biophysical constraints of the linker (9). Furthermore, the programmed cell death protein 1 (PD-1) and B- and T- lymphocyte attenuator (BTLA) both contain an immunoreceptor tyrosine-based inhibitory motif (ITIM) and immunoreceptor tyrosine-based switch motif (ITSM), providing two pTyr sites in close proximity, which recruit tandem SH2 domain containing phosphatases (PTPN6 and PTPN11) to the complex. Avidity is necessary for BTLA recruitment – like the TCR:ZAP70 interaction, PTPN11 recruitment to BTLA is lost if one of the pTyr sites is removed (10). PD-1 can recruit PTPN11 as long as the ITSM motif is intact – albeit with lower affinity – and there is evidence that PTPN11's preferred interaction crosslinks two PD-1 proteins in cis by binding to both ITSM motifs (11), an unexpected but potent example of avidity that provides additional context to the importance of receptor dimerization and clustering.

In addition to PTPN6/PTPN11 and ZAP70/SYK, three additional families (PI3K, PLCγ, and RASA1) covering ten total human proteins also contain tandem SH2 domains. Although this accounts for only 10% of the human SH2 domain-containing proteins, they are key proteins that transition signaling from receptor-proximal phosphotyrosine signaling to downstream lipid signaling (PI3K and PLCγ) and serine/threonine signaling (RASA1). Multisite phosphorylation is also more broadly observed beyond immune receptor systems. It occurs on other receptor tyrosine kinases (RTKs) and on downstream targets. For example, GAB1 gets robustly phosphorylated to act as a docking protein for multiple tandem SH2 proteins such as PTPN11, PI3K, and RASA1, allowing for simultaneous regulation of different signaling networks (12). The binding of PTPN11 to Y627/Y659 in GAB1 has been shown to occur at a stronger rate when both pTyr sites are available, which impacts activation of PTPN11 and has downstream consequences for the EGFR-MAPK pathway (13).

Despite the clear importance that bivalent interactions of

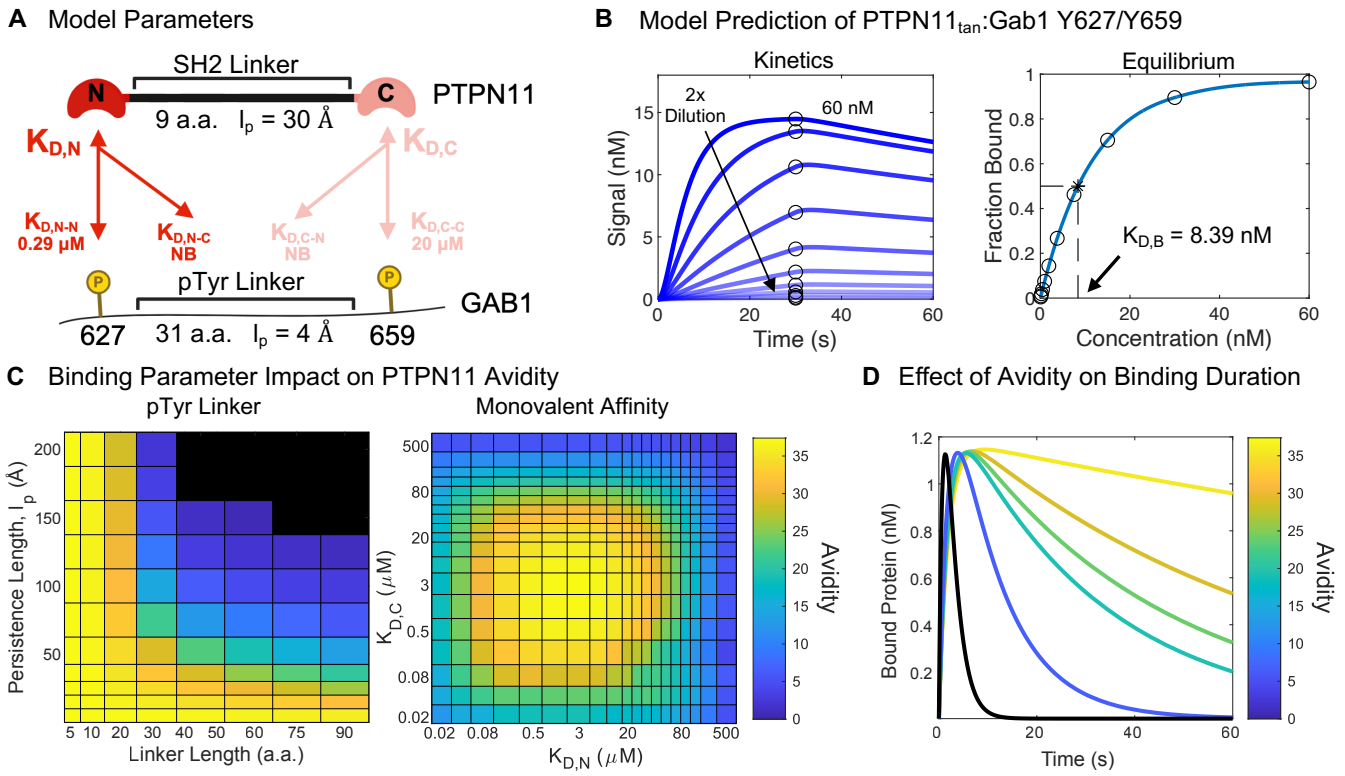

Fig. 1. Computational simulations to explore how parameters impact avidity. **A)** The generalized ODE model considers linker parameters for both partners and the monovalent affinities of all pairs. An interaction that is not expected to occur is listed as a non-binder (NB). This toy model is annotated with parameters specifically used to model tandem SH2 domain binding of PTPN11 to doubly phosphorylated Y627/Y659 of GAB1 based on literature and sequence values. **B)** A serial dilution simulation of surface plasmon resonance (SPR) experiments of $\text{PTPN11}_{\text{tan}}$ and GAB1 was performed using the ODE model. Equilibrium concentrations of bound protein were used to generate a concentration effect curve and estimate the $K_{D,B}$. **C)** We used the model to independently explore the impact of ligand linker biophysics and affinity of the monovalent species. Here, the PTPN11 tandem SH2 domain parameters were used, and we either fixed the monovalent affinities as noted in Fig. 1A and changed the linker parameters, or we fixed both linkers and changed the monovalent affinities. **D)** The kinetics of transient binding (interaction does not reach equilibrium) of the SH2 domain are plotted here under different model parameters that yield varying levels of avidity.

tandem SH2 proteins play in immune signaling, we generally lack a strong, comprehensive understanding of tandem SH2 domain recruitment to multi-phosphorylated proteins across the broader human proteome. A challenge to such an understanding is the complication of modeling multivalent interactions in a framework that captures the biophysics of the constrained interaction and the many species that are formed (such as an N-terminal SH2 domain bound to both pTyr sites and bivalent species bound in different orientations). Errington et al. developed an ordinary differential equation (ODE) model that accounts for the biophysical parameters of the linker constraint, using this to scale the driving force of the second interaction in the complex once the first has formed (14). An additional challenge involves the experimental validation of multivalent pTyr interactions, especially across larger regions of proteins like growth factor RTK tails. Peptide synthesis of phosphorylated proteins becomes both costly and complex, with limitations on total size and number of pTyr residues that can be incorporated (15). We developed a synthetic toolkit that can produce large amounts of multiply phosphorylated protein, including sites across the entirety of the EGFR C-terminal tail in a single protein construct (16). Using both of these advances, computational and experimental, we set out to more broadly understand the rules that guide tandem SH2 domain interactions in phosphotyrosine signaling networks.

In this work, we first establish an integrated computational and experimental approach and validation its ability to recapitulate the behavior of a known bivalent interaction (PTPN11:GAB1). Next, we use the model to explore the theoretical regimes where high avidity interactions occur in tandem SH2 domain recruitment, defining the role of linker biophysics and monvalent affinities in shaping overall avidity. In evaluating the ability to generalize this approach across the tandem SH2 domain family, we found that despite high sequence variability between SH2 domains, structure-based analysis suggests that there is a much more constrained 3-dimensional spacing that exists between SH2 domains, making them much more similar to each other regarding the partners they bind with high affinity. Finally, we ask if this model can be used to identify and rank tandem SH2 domain recruitment to the EGFR C-tail sites, demonstrating that even EGFR has the capacity to recruit partners with sub-micromolar affinities, which has implications in more broadly understanding the dynamics of RTK signaling.

Results

Establishing a computational model of tandem SH2 binding. Errington et al.'s ordinary differential equation (ODE) model simulates an SPR experiment between a multi-

valent receptor-ligand pair, tracking all possible binding configurations over time – 15 of which exist in a bivalent system (14). In our adaptation of the model, the tandem SH2 protein plays the role of the ligand, and the bisphosphorylated protein plays the role of the receptor. The model uses monovalent k_{on} and k_{off} rate constants to quantify the affinity of an individual interaction and uses these in combination with the SH2 concentration to determine the driving force behind any binding event. Linker parameters including the sequence length and protein flexibility – as determined by the persistence length (l_p) – are used to calculate the "effective concentration" of an unbound species following initial binding of the first domain, thus altering the driving force of the subsequent interaction and producing avidity. We adapted the model to allow monovalent affinity parameterization of all possible SH2-pTyr interactions and to improve control over the persistence length parameter. We modeled dynamic behavior under different SH2 domain concentrations, extracting the fraction bound to both SH2 domains (in either configuration) at equilibrium to estimate the bivalent dissociation constant ($K_{D,B}$). We additionally simulated the estimate of the monovalent domain interactions with the ligand (by setting the other domain affinity to 1M and effectively removing that interaction). We refer to these as $K_{D,N}$ and $K_{D,C}$ and calculate avidity as the inverse of the bivalent dissociation constant ($K_{A,B}$) divided by the sum of the inverse of the monovalent affinities ($K_{A,N}$ and $K_{A,C}$) (Equation 1):

$$Avidity = \frac{K_{A,B}}{K_{A,N} + K_{A,C}} \quad \textbf{(1)}$$

Simulations recapitulate avidity in PTPN11:GAB1 interaction. To test the ability of this modeling approach to be suitably parameterized from available data and to predict avidity, we selected an interesting cytosolic interaction that is known to occur bivalently – PTPN11 binding to Y627/Y659 bisphosphorylated GAB1 (13). The linker separating the PTPN11 SH2 domains is 9 amino acids long, and the GAB1 pTyr sites are separated by 31 amino acids. The persistence length of the SH2 linker was set to 30Å as has been measured in interdomain spanning protein linker. The pTyr linker was set to 4Å to match previously measured persistence lengths of intrinsically disordered proteins (17–19) (Fig. 1A). Estimates of monovalent K_D values were taken from a deep reanalysis of fluorescence polarization experiments by Ronan et al. (20). The N-SH2 domain of PTPN11 is estimated to bind to site Y627 with a monovalent K_D of 287 nM whereas the C-SH2 domain was not found to bind to either site. Since there was a reasonable detection limit in the experiments of 20 μM (21, 22), we selected to include the C-SH2 domain binding to Y659 with 20 μM, a value representing the highest affinity possible. For simplicity, we assumed no binding of the C-SH2 domain to site Y627 to prevent interference with the N-terminal interaction (using 1M constant for this). Rate constants were set with a constant k_{off} of 1 s^{-1}, adjusting k_{on} for the desired K_D. Although specific rates would be important to dynamic models, here we are specifically interested in using equilibrium behavior to understand overall binding affinity between a tandem SH2 domain and a bisphosphorylated partner.

Simulations were performed with a series of eleven PTPN11 concentrations, and the resulting concentration effect curve provided a predicted $K_{D,B}$ of 8.39 nM and avidity of 34, meaning the presence of both SH2 domains makes the binding affinity of the interaction 34-fold stronger than the individual monomers (Fig. 1B). In short, the capacity for tandem SH2 domain interaction with GAB1 converts a moderate affinity interaction that could at best be 287 nM to a strong interaction of 8.27 nM, which has significant implications for the degree and duration of the PTPN11:GAB1 interaction.

The role of linkers on shaping avidity. Since the linker region separating the SH2 domains and pTyr sites defines the effective concentration of the unbound domain upon initial binding of the protein, changing the linker length and persistence length impacts avidity. Having demonstrated that PTPN11 is capable of experiencing avidity, we used this interaction to explore the theoretical effects of avidity with a binding partner of different linker biophysics. Using the same linker parameters for PTPN11 and setting both $K_{D,N}$ and $K_{D,C}$ at a constant 4 μM, we simulated a series of different receptor linker and persistence lengths and calculated avidity (Fig. 1C). We observed maximum avidity occurs when the pTyr linker is relatively short or when it is highly flexible. As the distance between the pTyr sites increases and the linker becomes more rigid, avidity decreases until bivalent binding is not possible. Across the theoretical range of linkers, it happens that the predicted interaction for GAB1 is nearly maximum, suggesting that it is well-designed for avidity.

The role of monovalent affinities on shaping avidity. The other key parameter that determines the driving force behind a tandem SH2 interaction is the monovalent affinities of each SH2 domain. The linker helps constrain the biophysical space the protein resides in, which can increase the effective concentration of the unbound domain, but the intrinsic affinity still plays a role in determining binding. To test the impact of monovalent K_D on avidity we fixed the linker parameters to that of PTPN11 and GAB1 and then independently tested monovalent affinities (1C). The strongest avidities occurred when each monovalent affinity was moderately high. As either $K_{D,N}$ or $K_{D,C}$ became too large or too small, avidity began to decrease. The behavior in these two regimes is quite different, however, from a protein-protein interaction standpoint. Under high monovalent affinities, interactions are still high, and there is diminishing benefit or need for a second recruitment site. On the opposite end of the avidity drop-off, under low enough monovalent affinities, the partners are not interacting and there is no initial binding event that seeds the biophysical constraint that leads to a secondary binding event. Theoretically, the range where avidity is strongest occurs in the typical range of human SH2 domain binding (0.1 to 10 μM) (23).

The primary model of avidity is each domain binding with near exclusivity to one cognate partner on the other protein with little to know cross-reactivity, such as in the

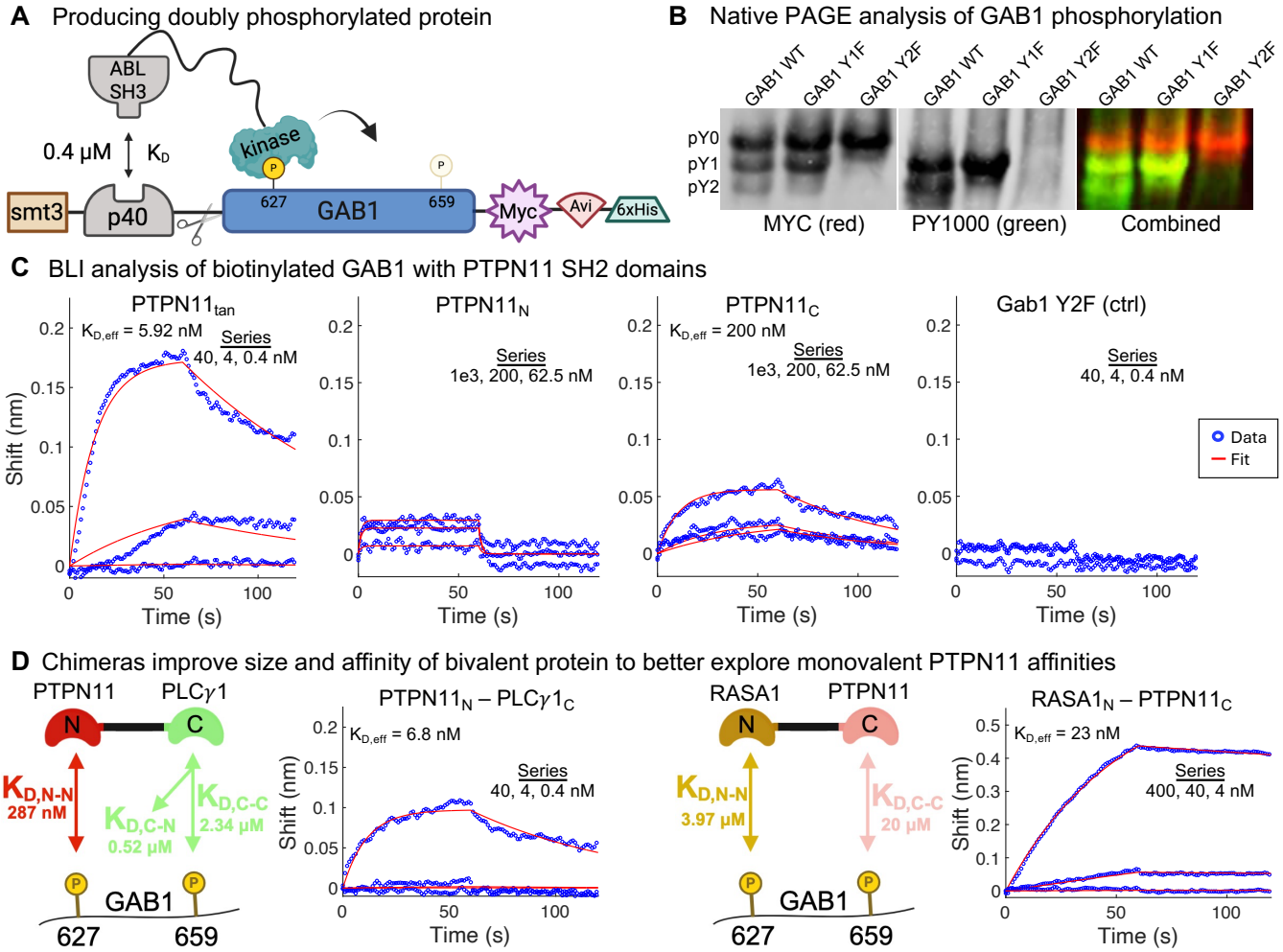

Fig. 2. Experimental testing of the PTPN11:GAB1 interaction. **A)** SISA-KiT was used to phosphorylate both pTyr sites within the GAB1 region expressed using SRC kinase catalytic domain. The secondary interaction used in SISA-KiT to enhance phosphorylation of the substrate is the ABL SH3 domain with the p40 polyproline sequence (APTYSPPPPP). **B)** The degree of phosphorylation and confirmation of multisite phosphorylation on purified GAB1 was evaluated by running native PAGE and immunoblotting with MYC and confirming by pan-specific phosphoantibody. GAB1 WT refers to the GAB1 region with both tyrosines and the Y1F and Y2F mutate one and both tyrosines to phenylalnine, respectively, as a control for where on the gel the species run. **C)** Purified, phosphorylated GAB1 was biotinylated using BirA ligase and streptavidin BLI tips were used to measure the kinetics of ligand with SH2 domain. Shown here are the background corrected on- and off-kinetics and the fit of the BLI data with a 1:1 kinetic model. Depending on the degree of effective binding affinity, we adjusted the serial dilution series to avoid saturating the system and the specific concentrations of SH2 domain in the experiment are indicated. Where fit was sufficiently reliable in the experiment, we report an effective affinity ($K_{D,eff}$). We used the double Y2F mutant as a negative control. **D)** Chimeras were produced by swapping out the $PTPN11_N$ and $PTPN11_C$ SH2 domains with $RASA1_N$ and $PLC\gamma1_C$, respectively. Simulations of these chimeras interacting with GAB1 were performed in the model, and BLI experiments were performed to calculate $K_{D,eff}$ for each new species.

PTPN11-GAB1 model (Fig. 1A). To explore how cross-reactivity between partners impacts avidity, we altered the relative contributions of each potential domain/pTyr interaction towards the overall $K_{D,N}$ and $K_{D,C}$ by increasing the degree of cross-reactivity for both domains. Monovalent affinity remained constant, with $K_{D,N-N}$ and $K_{D,C-C}$ decreasing to compensate for the increases in $K_{D,N-C}$ and $K_{D,C-N}$ (Fig. S1). The maximum avidity decreases, though remains in the greater than 10-fold range, when both domains have reasonably strong affinity (2.5μM) for the ligand, except in conditions where the cross-reactivity is much higher than the on-target reactivity. However, moving to a model where there is one strong recruitment domain (2.5μM) and a weaker domain (25μM) significantly diminishes the avidity possible. Hence, the characteristics of high avidity include strong preferences for separate domain specificity with minimal cross-reactivity.

The effects of avidity on binding duration. Signaling proteins bind in a transient manner, meaning they bind, unbind, and typically re-bind a number of times before fully dissociating and moving away within the cell (24). Since we have established that avidity within tandem SH2 proteins likely occurs due to an increased effective concentration of unbound domain upon initial binding of the first SH2 domain, it stands to reason that a multivalent protein in the process of unbinding from a target where one domain is still bound would experience the same increased driving force. In this manner, we hypothesize that protein interactions that experience stronger avidity will re-bind with a stronger affinity, thus extending the amount of time two proteins are engaged. To test this hypothesis, the PTPN11:GAB1 interaction was used as a baseline, and linker and monovalent K_D parameters were altered to achieve different avidities, including a simulated monovalent condition where no avidity occurs. The different condi-

tions were then used to simulate a transient binding interaction where there is no longer a constant influx of PTPN11, to maintain a steady concentration and drive the system to equilibrium, and instead the proteins are allowed to bind and fully unbind from one another (Fig. 1D). As hypothesized, the conditions that represented a larger avidity also resulted in the proteins remaining bound for a longer period of time.

Establishing an experimental model of tandem SH2 domain binding. Although the adapted ODE modeling approach appears to describe a known bivalent interaction relatively well, experimental validation of model outputs is necessary, especially since there may often be uncertainty in the parameterization of monovalent affinities and linker biophysics. Therefore, we set out to establish a complementary experimental validation. We selected biolayer interferometry (BLI) for measuring binding kinetics between an SH2 domain and a phosphorylated peptide partner immobilized to a BLI probe (i.e. in the loading step we bind a biotinylated phosphopeptide to a streptavidin-coated BLI tip and in the association phase we present an SH2 domain, followed by the dissociation step in fresh buffer). However, synthesizing multiply phosphorylated peptides is challenging and expensive, especially at the length necessary to cover physiologically relevant partners. For example, sufficient flanking sequence coverage of the GAB1 Y627/Y659 partners is 40 amino acids. This limitation was overcome by utilizing our lab's SISA-KiT system which involves co-expressing a protein fused to a p40 polyproline sequence with a constitutively active tyrosine kinase fused to an ABL SH3 domain (Fig. 2A). The affinity between the ABL SH3 domain and the p40 sequence localizes the kinase to our protein of interest, driving phosphorylation and helping us achieve multi-site phosphorylation. This approach has the benefit of cheaply and easily producing large batches of phosphorylated proteins, although it produces a mixture of peptide species, including nonphosphorylated, singly phosphorylated, and doubly phosphorylated partners. In some ways, this experimental system is a strength – it captures the complexity of ligand presentation in cells, where it is unlikely that tandem SH2 domains interact with a pure population of doubly phosphorylated partners. Hence, here we set out to establish the relevancy of this experimental system to test model predictions of tandem SH2 domain binding.

As part of this experimental approach, we also tested the monovalent interactions of the N-terminal and C-terminal PTPN11 domains with the phosphorylated GAB1, allowing us to evaluate both the tandem affinity and the constitutive components that make up the interaction. Initially, we focused on creating SH2 domain binding dead point mutants (mutating the invariant arginine in the SH2 domain that coordinates pTyr binding), in order to retain the mass of the association step, which BLI is sensitive to. However, a recent evaluation of available structures highlighted that in tandem SH2 and high local concentration interactions, the SH2 domain can engage in a non-canonical manner (i.e. engagement occurs even in the absence of the pTyr-arginine interaction) (25). Hence, to avoid complications from that possibility, we measured isolated domain interactions of monovalent SH2 domains. By presenting the same phospholigand species mixture to individual N- or C-terminal domains, the measurement reflects the combined $K_{D,N}$ and $K_{D,C}$ interaction with both phosphorylation sites and can be compared then directly to the tandem SH2 domain interaction. Like other experimental kinetic or thermodynamic analyses, BLI signal represents the mixture of all configurations of the SH2 domain protein bound to the ligand and hence we refer to the measured affinity from BLI as an "effective" affinity ($K_{D,eff}$).

In order to extract an effective affinity from BLI, we performed experimental serial dilutions of SH2 domain. Due to the expense of reagents, we limited experiments to three serial dilutions of SH2 domain and simultaneously fit on- and off-rates for the interaction across the series (calculating $K_{D,eff}$ as the off-rate over the on-rate). We also tested the domain species against a fully mutated ligand (Y2F control) made in the same manner, to ensure binding was phosphosite specific. The SH2 domains were produced in a pGEX backbone (a GST fusion), but we Prescission-cleaved the GST tag to ensure the SH2 domains were monomeric (26). Finally, following advice that was accumulated from a thorough analysis of fluorescence polarization data (20), we report the lowest $K_{D,eff}$ of duplicate experiments, except in cases of poor fit. Using PTPN11:GAB1, we explore the ability of the model to match meaningful interaction data of tandem SH2 domains with phosphorylated partners.

Experimentally testing PTPN11:GAB1 interaction. We cloned the human GAB1 sequence from amino acids 590 to 694 (the entire C-terminus before a new tyrosine was introduced) into the SISA-KiT substrate vector and incorporated an Avi-tag for biotinylation of the substrate, using Nickel-based purification of the ligand. We additionally performed mutagenesis, mutating both tyrosine sites to a phenylalanine (GAB1 Y2F) to prevent phosphorylation and act as a negative control. Phosphorylation of these proteins and an additional version of GAB1 with a mutation only at site Y659 (Gab1 Y1F) was evaluated using native PAGE, which allows for separation of phosphospecies due to the strong negative charge provided by the phosphate ion (Fig. 2B). By native PAGE, we estimate that 20% of our GAB1 protein was doubly phosphorylated and 30% of the protein was singly phosphorylated.

We hypothesized that if the degree of doubly phosphorylated protein was sufficient to induce bivalent binding, then the tandem SH2 domain binding would be significantly stronger than the monovalent SH2 domain binding. During BLI testing, we saw significantly higher $PTPN11_{tan}$ binding, compared to the monovalent species (Fig. S2). In fact, we found that it was necessary to customize the serial dilution range as presenting high concentrations of the tandem species led to nonlinearities in binding and the inability to accurately extract on and off rates (Fig. S3). Therefore we maximize concentrations in the series for the monovalent species, while dropping the range for the bivalent species (Fig. 2C-D), matching the serial dilution range to the affinity of the interactions.

Consistent with the literature values, individual PTPN11

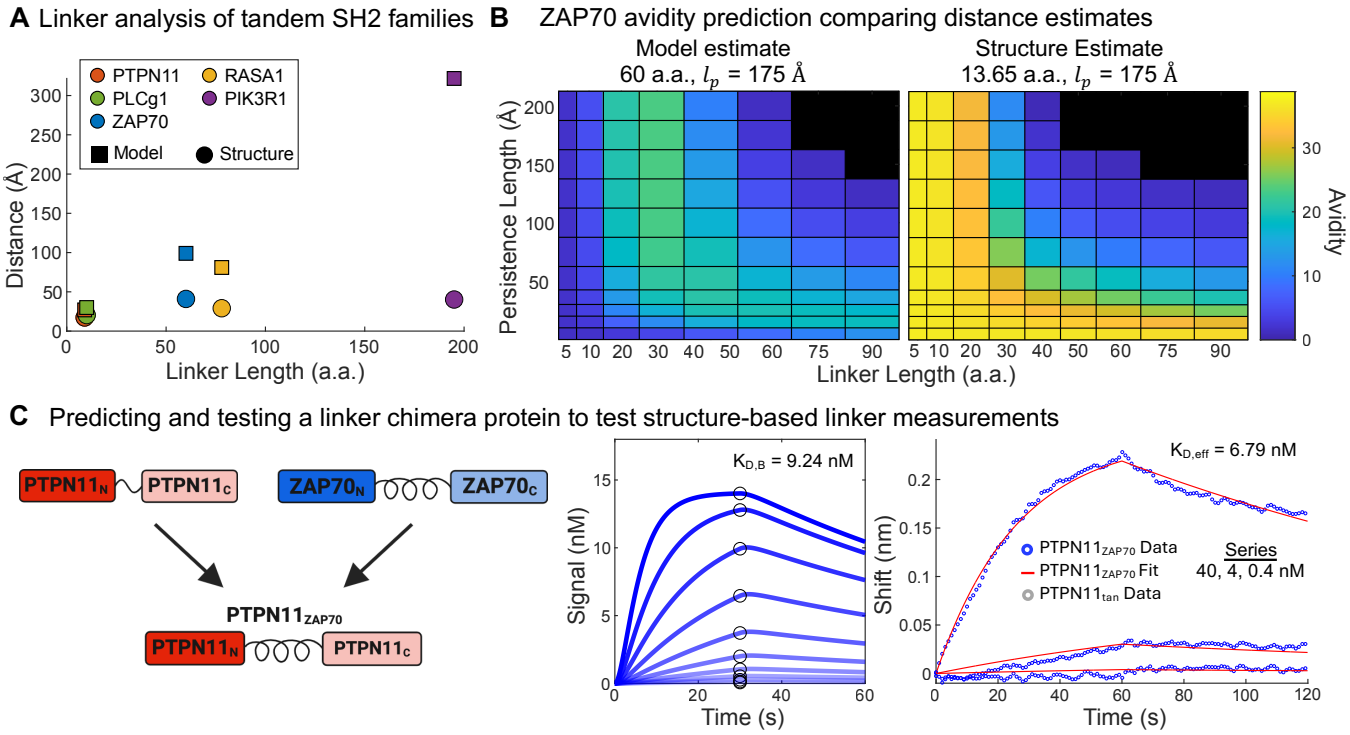

Fig. 3. Evaluating linkers between all tandem SH2 domain families in the human proteome. **A)** The end-to-end distance versus the primary sequence length for each tandem SH2 protein family is shown. Distance was estimated based on the number of amino acids and was measured from available PDB structures (here the average distance is shown where multiple structures were available). **B)** We used the model parameterized by either the primary sequence-based estimate or the structure-extracted distance to characterize avidity space. **C)** To test if the structure extracted estimate of SH2 distances was accurate, we created a chimera protein, cloning the ZAP70 linker between the PTPN11 SH2 domains. We modeled and experimentally measured the effective affinity of the $\text{PTPN11}_{\text{ZAP70}}$ chimera with doubly phosphorylated GAB1 Y627/659.

domains bound weakly to GAB1, whereas the tandem PTPN11 domains bound strongly with an estimated $K_{D,\text{eff}}$ of 5.92 nM, 34-fold stronger than the highest monovalent affinity, matching the avidity predicted by the model. In contrast with the literature affinities, it was PTPN11_{N} that appeared to be below the signal of our experimental range and instead PTPN11_{C} which bound with a $K_{D,\text{eff}}$ of 200 nM. The signal range of BLI is limited most by the highest concentration of SH2 domain we are able to produce, though we cannot rule out the possibility that for some reason the isolated N-terminal SH2 domain of PTPN11 undergoes some degree of misfolding and activity loss. However, the bivalent affinity demonstrates strongly that both domains are functionally intact in the tandem domain species. The model parameters of a monovalent species of 287 nM replicates well the 200 nM monovalent interaction and the prediction of the bivalent interaction was 8.39 nM, which is very similar to the experimentally measured 5.92 nM interaction. Together, these suggest that, as a set of interactions, the model affinities and the linker parameters were well calibrated to predict the overall increase in binding affinity in tandem interactions, which is two orders of magnitude better than the best monovalent affinity.

SH2 chimeras to interrogate model parameters. To overcome the limitations of testing SH2 monomers with BLI, we developed a new approach for evaluating the discrepancy in monovalent affinities of each PTPN11 domain. We selected two alternate tandem SH2 domain proteins, RASA1 and PLCγ1, which also had previously characterized binding affinities with GAB1 Y627 and Y659 (20). We made two domain "swaps" that would allow us to interrogate model performance under new affinities, and confirm the potentially hidden or misannotated PTPN11 monovalent affinities (Fig. 2D). The PTPN11_{N} - $\text{PLC}\gamma\text{1}_{\text{C}}$ chimera replaces what we experimentally found to be the important PTPN11:GAB1 interaction with a new C-terminal domain with previously measurable affinity to both GAB1 sites. The RASA1_{N} - PTPN11_{C} chimera replaces the immeasurable PTPN11 N-terminal domain with an interaction previously characterized with GAB1 Y627 as 4μM. We simulated and experimentally tested these chimeras, keeping the originally proposed PTPN11 domain interaction parameters (Fig. 1A). For the PTPN11_{N} - $\text{PLC}\gamma\text{1}_{\text{C}}$ chimera, the model and BLI experiments were in strong agreement with a $K_{D,B}$ of 6.98 nM and a $K_{D,\text{eff}}$ of 6.8 nM, suggesting that our parameterization of the PTPN11 N-SH2 domain is fairly accurate, and the monomer is binding with an affinity of roughly 300 nM. However, for the RASA1_{N} - PTPN11_{C} chimera, the model and experimental results differed by a wide margin. The model provided a $K_{D,B}$ prediction of 94.3 nM whereas BLI yielded a $K_{D,\text{eff}}$ of 23 nM. This disagreement suggests that the parameterization of the PTPN11 C-SH2 affinity of 20 μM (the affinity used to represent the original signal range of the literature value) was inaccurate, and the monomer is capable of binding with a stronger affinity as supported by the

$PTPN11_C$ BLI results.

These results highlight the importance of having accurate experimental data to parameterize the ODE model. However, the relatively strong agreement between the model and experimental results of the $PTPN11_{tan}$:GAB1 interaction, despite the uncertainty surrounding $PTPN11_C$ demonstrates the ability of the model to handle errors in one domain if the remaining monovalent interaction is relatively strong and can drive avidity. Furthermore, these results introduce the idea that certain pTyr sites within the human proteome that are not expected to interact with any SH2 domain-containing protein, based on available monovalent binding experiments, may actually be a critical binding site to enhance bivalent binding of a tandem SH2 domain partner, changing both amount and duration of the interaction.

Exploring generalizability of tandem SH2 domain binding. Having established that the computational approach can successfully predict high affinity bivalent interactions when parameterized by previously defined monovalent affinities and estimates of linker parameters and the ability to confirm experimentally, we set out to explore the generalizability of this approach to more broadly predict tandem SH2 domain recruitment. First, we asked if we could develop predictions for each family member of the tandem SH2 domains, which vary in the SH2 domain linker region. Secondly, we ask if we can scan a set of different pTyr pairs along a receptor tyrosine kinase C-terminal tail and identify the recruitment of a tandem SH2 domain partner to the receptor.

Expanding to all tandem SH2 domain family members. Having established that the linker between the SH2 domains determines the idealized partner linker parameters, we initially hypothesized that each SH2 domain family would require family-specific models (i.e. linker parameters). Across the five structurally homologous groups of tandem SH2 proteins, the SH2 linker is highly variable. Based on sequence alone, PTPN11 and PLCγ1 are similar in length being 9 and 10 amino acids long, respectively. They are also both flexible with an estimated persistence length of 30Å. On the other hand, the ZAP70 linker is 60 amino acids long and rich in alpha helices, making it significantly more rigid. Similarly, the PIK3R1 linker is rigid and 195 amino acids long. The persistence length of alpha helices is typically between 150 – 200Å, so the persistence length for ZAP70 and PIK3R1 was estimated at 175Å (27, 28). RASA1's SH2 domains are split by an SH3 domain, providing moderate rigidity and a length of 78 amino acids (29). The ODE model would predict that each of these linkers constrain the SH2 domains to a wide range of separation distances, and to confirm this, we examined the PDB files of each tandem SH2 protein, measuring the physical distance from one end of the linker to the other, and calculated the average distance (Fig. 3A). Surprisingly, while the model-estimated distances between SH2 domains varied widely, the structure-estimated distances were much more consistent across all families – there was only a 23Å difference across all tandem SH2 domains based on structure (17.3Å for PTPN11 versus 40.9Å for ZAP70), versus a model predicted 290Å difference (27Å for PTPN11 versus 321.75Å for PIK3R1). This suggests that despite exhibiting high variability within the sequence of their SH2 linkers, tandem SH2 proteins evolved to maintain a strongly conserved physical distance between domains.

To evaluate the effect of the modeled linker versus the structure-extracted linker distances on avidity, we performed a computational experiment. We set $K_{D,N}$ and $K_{D,C}$ constant at 4 μM and performed a pTyr linker sweep with ZAP70 using both the model-estimated and structure-estimated linker lengths (Fig 3B). There is a striking difference, not just in the idealized pTyr partner linker parameters that create optimal avidity, but also in a lower overall maximal avidity that can be achieved by ZAP70 if its SH2 domains were separated by the larger distance expected by the model. These data suggest that the highly conserved physical separation of all tandem SH2 domains likely plays a role in enabling them to bind with high avidity.

To confirm that the tandem SH2 domain separation is parameterized best by the structure-extracted distance, we created a linker "swap". Choosing the helical ZAP70 linker, which has the largest structure-predicted difference in separation distance compared to PTPN11, we replaced the PTPN11 SH2 domain linker with the ZAP70 linker (Fig 3C). The ($PTPN11_{ZAP70}$) chimera is predicted (parameterized by the structure-based distance) to bind with high avidity on a similar scale to the original PTPN11 wild type protein (9.24 nM compared to 8.39 nM of PTPN11). Experimentally, the $PTPN11_{ZAP70}$ bound to GAB1 with an effective affinity of 6.79nM, closely matching the model estimate and, as expected, being only slightly higher than the original $PTPN11_{tan}$ $K_{D,eff}$ of 5.92 nM (Fig. S4). In comparison, when parameterized to match the longer model-predicted linker length of ZAP70, the interaction was predicted to achieve an effective affinity of 369.6 nM, clearly demonstrating that the structure-based estimates of distance are better representations of the biophysical structure of tandem SH2 proteins. To confirm accuracy, new simulations of PTPN11:GAB1 dynamics were performed using the structure-based PTPN11 linker parameters where the linker length is 5.75 amino acids (SFig. S5). Since there is a minimal difference between the model- and structure-based estimates of the PTPN11 linker, the impact of this change was negligible, and our conclusions from the PTPN11:GAB1 studies remain sound. The results also suggest that not only are the tandem SH2 domain family members constrained to similar physical separation, but that it means they are all "tuned" relatively similarly to bind similarly spaced pTyr residues.

Ranking EGFR phosphotyrosine pair recruitment of PLCγ1. Given the emergent property of avidity and the high degree of multiply phosphorylated proteins in the human proteome, it seems likely that protein interactions with bivalent protein partners should consider more than just the monovalent affinities. Here, we wished to ask if the ODE model parameterized by available data, such as prior monovalent interaction data, the physical separation of SH2 domains,

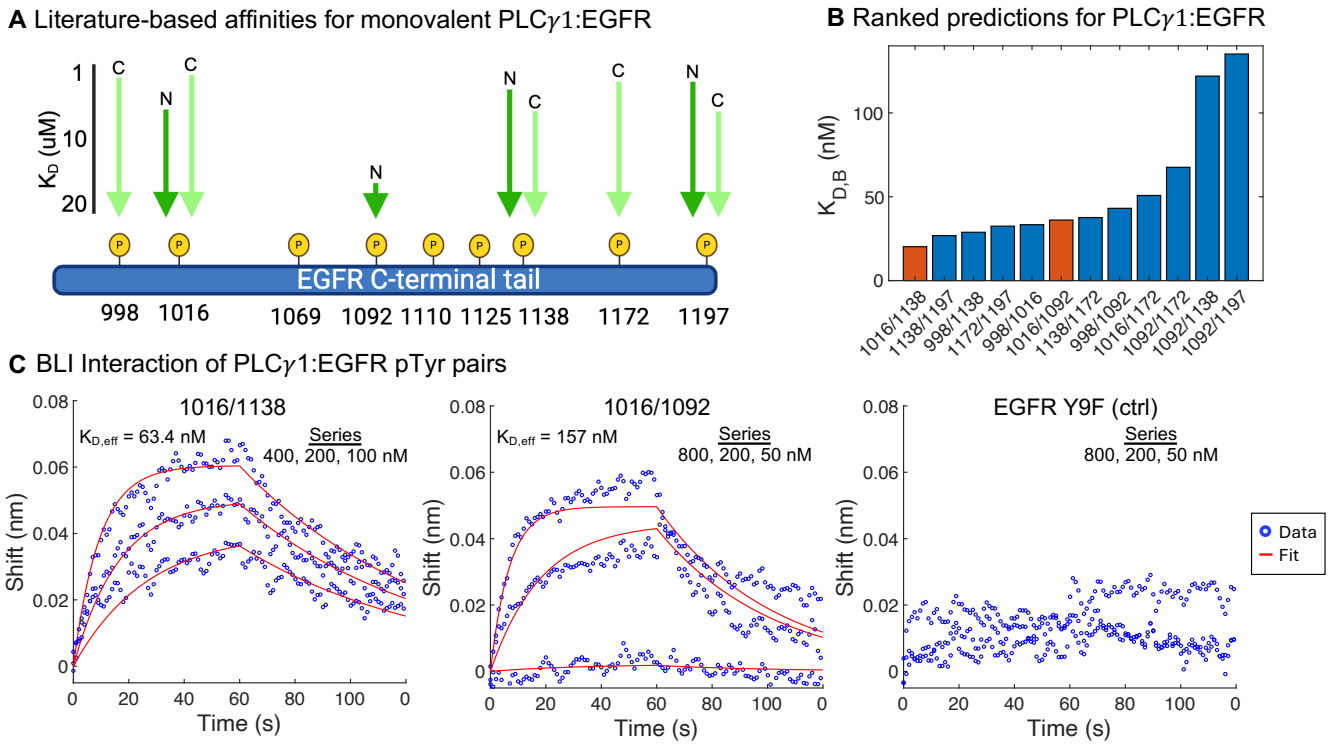

Fig. 4. Predicting and testing bivalent recruitment of PLCγ1 to EGFR C-terminal tail phosphotyrosine pairs. **A)** Literature based monovalent affinities of PLCγ1 with each pTyr site on the EGFR C-terminal tail (20). **B)** Using literature affinity values, structure-based SH2 domain distance, and estimated biophysical distances for the disordered tail, we predicted the bivalent binding $K_{D,B}$ of PLCγ1 to EGFR pTyr pairs. **C)** We selected three pTyr pairs predicted to have different effective affinities and measured these experimentally using BLI and a Y7F form of the EGFR C-terminal tail (ensuring only the two of the nine possible sites were phosphorylated).

and the pTyr linker biophysics could be broadly generalizable to identifying high affinity interactions and their relative strength. We elected to study recruitment of PLCγ1 to different pairs along the EGFR C-terminal tail. The EGFR C-tail contains nine pTyr sites, and each SH2 domain of PLCγ1 has an expected monovalent affinity towards at least four of these sites (Fig. 4A). We used the ODE model to predict the $K_{D,B}$ of each potential pTyr pair. The biophysical estimation approach was limited by memory when predicting two of the longest span pairs on the tail (998/1197 and 1016/1197) and were not considered. All other remaining pairs were predicted and ranked based on bivalent avidity (Fig. 4B), suggesting a broad range of effective affinities and some with very high avidity.

To experimentally validate these predictions, we selected two pTyr pairs predicted to exhibit strong bivalent affinity and tested them using BLI (Fig. 4C, Fig. S6). Using the framework of a fully tyrosine mutated EGFR Y9F construct (30), we introduced the pairs of tyrosines and phosphorylated these using the SISA-KiT approach. Unfortunately, the size of this construct made separation of phosphospecies by native PAGE impossible for quantification and we also failed to resolve the species in Phos-tag based approaches (31). However, we used phosphospecific antibodies to the EGFR sites to confirm that we had successfully produced phosphorylation of both sites in the kinase conditions selected (SFig. S7), with additional confirmation provided by high avidity interactions measured in BLI experiments using the phospholigand.

As predicted, both pTyr pairs bound with strong bivalent affinity, and their predicted rank order was properly reflected with 1016/1138 exhibiting a stronger $K_{D,\mathrm{eff}}$ of 63.4 nM compared to 1016/1092's $K_{D,\mathrm{eff}}$ of 157 nM. While the relative affinity matches the predicted order, it should be noted that the exact $K_{D,\mathrm{eff}}$ value for each EGFR pair differs from its predicted $K_{D,B}$ by a wider margin than observed in the PTPN11:GAB1 interaction with systematically stronger predicted affinity values by the model. Our model assumption was similar flexibility of GAB1 and EGFR C-tail (a persistence length of 4Å), but the model versus experimental discrepancy points to the Ctail as having physically higher rigidity. However, the C-tail clearly maintains a relatively high degree of flexibility as evidenced by the ability for PLCγ1 – with its short 10 amino acid SH2 linker – incurring strong bivalent binding with sites separated by 121 amino acids in the 1016/1138 pair. EGFR Y1016 (also known as Y992 in another isoform) has been shown to be a key interaction requirement for PLCγ1 (32), consistent with the high monovalent affinities on the order of a μM. However, the model and experiments support that PLCγ1 interaction is further strengthened, to the sub-μM range by avidity and secondary recruitment, even with the low monovalent interaction of Y1092, consistent with theoretical regimes identified by the mechanistic model.

Discussion

A guiding principle in the field of phosphotyrosine signaling is that the relatively low affinity of SH2 domains (on the order

of μM) is important to the reversibility of signaling – allowing the dissociation of signaling complexes and phosphatase access (26, 33, 34). This study, along with existing knowledge in the immune receptor system, indicates that tandem SH2 domain containing proteins are recruited with a strikingly higher affinity (on the order of nM). Notably, we found that the normal range of affinity for an SH2 domain (0.1 - 10 μM) matches well to the range of monovalent K_D values that maximize avidity, providing a new perspective on SH2 affinity. Perhaps the relatively low affinity of SH2 domains evolved because it optimizes the advantage of multivalency within the tandem SH2 proteins to allow for these dramatic improvements in affinity. We found it interesting that the tandem SH2 domain families appear to have evolved relatively similar geometries in 3D structure that set them up exceedingly well for high avidity interactions, which also helps explain why post-translational modification of the SH2 domain linker impacts binding partners (9). Given the breadth of signaling proteins covered by the tandem SH2 domain family, there is a possibility that almost all phosphotyrosine signaling is governed to some degree by tandem SH2 domain interaction avidity. Additionally, SH2 domains rarely exist as single domain proteins, and it is likely that recruitment to phosphotyrosine sites are coupled with other multivalent interactions (such as SH3-polyproline or PTB-pTyr interactions). These nM-order interactions have the capacity to strengthen and extend signaling duration, which ultimately shapes cell response (35, 36).

With the already difficult task of measuring monovalent interactions between domains and all of their possible partner ligands, tandem domain interactions creates an additional layer of combinatorially complexity. We propose that this modeling framework is a useful step at identifying possible high affinity interactions within the combinatorially complex space of multiply phosphorylated protein partners. As a framework, using structure-based extraction for defining the domain linker conditions was successful. Cognate targets of high affinity will entail those that have high monovalent interaction strength alone, or, for avidity the existence of at least one site of sufficient recruitment potential and a second site that has the capacity to stabilize this interaction. Despite some differences in available monovalent interaction data, we were able to predict, with a surprising degree of accuracy, high affinity interactions. We found that for flexible, disordered ligand partners the effective physical distance between phosphotyrosines that were capable of binding was much larger than we anticipated (for example 121 amino acids apart in the EGFR C-terminal tail), which means that when considering bisphosphorylated partners, one must scan beyond immediately adjacent pairs when considering possible high affinity interactions, though the intervening structural linker must be evaluated for the possible fit to the partner biophysics.

The conformation of bivalent interaction affinity, especially for something as large as the EGFR C-terminal tail, would not be possible via peptide synthesis. Hence, the synthetic approach to phosphoprotein production was invaluable for these assays and can likely extend to testing other bivalent interactions where one or more site is a phosphorylated tyrosine residue. However, these in vitro interaction data do not yet account for complexities in real protein interactions within the cell. For example, it is possible that multiply phosphorylated sites on the disordered RTK tail changes the flexibility of the tail and therefore avidity partnerships (here, we eliminated other tyrosine sites to isolate specific interactions). Additionally, those other phosphotyrosines, along with the isolated sites of study in an experiment and model, have the capacity to interact with other domains also changing the physical dynamics of the tail and creating competition. On the other hand, avidity has the potential to deeply shape competition, preventing interactions due to the frequent rebinding of a low affinity partner due to high local concentration effects. Therefore, these experiments and models are really the first step towards asking what could happen in a cell, but may not reflect necessarily what does happen in the complex cell environment. However, we propose that the avidity of tandem SH2 proteins, and other multivalent interactions, should be predicted and assessed in order to then further guide cell-based analysis.

Materials and Methods

Plasmid cloning and mutagenesis. We used plasmid backbones as described in Ryan et al. (16) to clone GAB1 and EGFR into the toolkit substrate backbones (sVd with Avi-tag) for co-expression and phosphorylation with a kVh-kinase (also described in Ryan et al.). SH2 domains were cloned into a pGEX backbone, with an N-terminal GST fusion. We used either restriction/ligation cloning or In-Fusion cloning for engineering of backbone vectors, insertion of tags and targeting, and all kinase and substrate insertions. We used Snapgene to design primers, and sequenced verified by Sanger sequencing. Sequencing was performed to get full coverage of insert. Some constructs were whole plasma sequenced to verify the entire plasmid. For cloning, we used NEB Phusion polymerase (part no M0530S) and included extensions for cloning into the target backbones using NEB enzymes. We used Agilent's QuickChange mutagenesis kit (part no 210518) to perform point mutations. We used carbenicillin at 100μg per mL for selection of ampicillin plasmids (substrate and GST plasmids) and kanamycin at 50μg per mL for selection of kanamycin plasmids (kinase plasmids). Subcloning Efficiency DH5α Competent Cells: ThermoFisher Scientific Cat. 18265017 were used for DNA propagation. The sources for SH2 domains were: PTPN11 – from pBABE-PTPN11 was a gift from Dr. Matt Lazzara (37); RASA1 – pDONR223_RASA1_WT was a gift from Jesse Boehm & William Hahn & David Root (Addgene plasmid # 81778 ; http://n2t.net/addgene:81778 ; RRID:Addgene_81778); ZAP70 – pDONR223-ZAP70 was a gift from William Hahn & David Root (Addgene plasmid # 23887 ; http://n2t.net/addgene:23887 ; RRID:Addgene_23887); PLCγ1 – pGEX PLCg1(NC)-SH2 was a gift from Bruce Mayer (Addgene plasmid # 46471 ; http://n2t.net/addgene:46471 ; RRID:Addgene_46471).

GAB1 – from pCDNA3.1-HA-GAB1 a gift from Dr. Matt Lazzara (38). EGFR C-terminal tail was gifted to us by Linda Pike (39) in both Y9F and Y8F formats. We selected the appropriate Y8F plasmids to mutate a second tyrosine to phenylalanine to create the Y7F plasmids.

Protein expression and purification.

Protein induction and bacterial lysis.. **SH2 Domains** Plasmids were transformed into BL21 gold DE3 (Agilent part no. 230132) and an isolated colony was used to inoculate a 5 mL LB, 5 uL carbenicillin culture which grew overnight at 37°C. 2.5 mLs of this culture was used to inoculate a 100 mL LB, 100 uL carbenicillin culture which shook at 37C until it reached the desired OD600 (0.7-1.0). The culture was then induced using 0.5 mM IPTG and grown overnight in an 18C shaker. The culture was spun down at 8500 rpm for 10 minutes and the pellet was frozen at -20°C until lysis.

Phosphorylated Proteins Plasmids were co-transformed along with a kinase plasmid into BL21 gold DE3 (Agilent part no. 230132) and an isolated colony was used to inoculate a 5 mL LB, 5μL carbenicillin/kanamycin culture which grew overnight at 37°C. 2.5 mLs of this culture was used to inoculate a 100 mL LB, 100μL Carbenicillin/Kanamycin culture which shook at 37C until it reached the desired OD600 (0.7-1.0). The culture was then induced using 0.5 mM IPTG and grown overnight in an 18C shaker. The culture was spun down at 8500 rpm for 10 minutes and resuspended in 100 mL LB, 100 uL Carbenicillin/Kanamycin, and 0.2% L-arabinose. The culture was grown for another 4 hrs in a 37°C shaker and then spun down at 8500 rpm for 10 minutes and the pellet was frozen at -20°C until lysis.

Lysis We resuspended bacterial cell pellets in a 10:1 ratio of final growth volume to lysis buffer. Lysis buffer was 50mM Tris-HCl at pH7.0 and 150mM NaCl, supplemented with protease inhibitors (EMD Millipore 539137), phosphatase inhibitors (Millipore Sigma P5726), and PMSF (1:1000). For small volume lysis of 1mL to 5mL, we used bead beating, adding 400μL of beads to each 1mL of resuspended cell pellet and bead beat for 2 minutes. For larger lysis volumes, we sonicated resuspended pellets for 5 minutes with 40% pulse sequence. Following lysis, we frequently sampled and evaluated the raw lysate, prior to clarification by centrifugation at 13,000 rpm for 10 minutes at 4°C.

Protein Purification. **Nickel/6xHis Purification**. We used Genscript Ni-NTA MagBeads, using a magnetic tube stand for isolating beads during decanting. Prior to protein binding, we equilibrated beads in lysis buffer (50mM Tris-HCl, 300mM NaCl (pH 8.0), 0.5% TX-100, 80 mM imidazole). Beads were isolated by magnet and decanted. We repeated equilibration two total times, then added 1mL of clarified lysate to beads and incubated at 4°C for 1hr up to overnight, while rotating. Isolated beads were washed in a 20-fold excess (relative to bead volume) of wash buffer (50mM Tris-HCl, 300mM NaCl, 40mM Imidazole, 0.5% Triton-X (pH 8.0)) a total of three times. Protein was eluted using 2 times the resin bed volume with elution buffer (50mM Tris-HCl, 300mM NaCl, 500mM Imidazole pH 8.0). We repeated elution two total times, typically finding the highest yield in the first two elutions. We dialyzed protein following manufacturer directions for the Pierce Slide-A-Lyzer mini dialysis with a 3.5kDa cutoff into 50mM Tris-HCl, 150mM NaCl at pH. 7.8. Proteins were aliquoted and frozen at -80°C.

GST Purification We used Pierce Glutathione Agarose (part no. 16100) and followed the manufacturer's base protocol for GST-based purification. Beads were equilibrated in buffer (50mM Tris, 150mM NaCl) and protein capture was performed for one hour to overnight at 4°C. Beads were washed (three times) with a 10-fold excess volume of equilibration buffer, with rotation. We used PreScission protease-based elution according to manufacturers directions (GenScript GST-PreScission protease part no. Z02799), incubating protease with beads resuspended in 50 mM Tris-HCl, 150 mM NaCl overnight at 4°C GST-fused PreScission.

Biotinylation. Following nickel- purification and dialysis, phosphorylated substrates were biotinylated using the BirA500 kit (Avidity EC 6.3.4.15). We reacted 10nmol of substrate with 2.5μg of BirA ligase for between 30 minutes and 5 hours at 30°C depending on the protein concentration. Biotinylation was confirmed through 680IRdye-conjugated streptavidin, incubated for one hour at 4°C (at 1:1500), prior to final washes and LICOR-based imaging.

Western-based analysis:. Samples were reduced in Laemmli loading buffer (Boston BioProducts) and boiled for 10 minutes at 95C, unless they contained imidazole, where we boiled at 70°C for 5 to 10 minutes to avoid breaking protein bonds. We performed standard SDS-PAGE analysis using purchased precast NuPAGE gels from Invitrogen, typically using 4-12% gradients or single 10% gels, running in a MOPS SDS running buffer (Novex Invitrogen). Transfer to nitrocellulose membranes occurred in a buffer (Novex Invitrogen) with 20% methanol (membranes were pre-wetted in transfer buffer). Pre-stained ladder from LI-COR (Chameleon Duo) was used along with IRDye-680 and -800 secondary antibodies for infrared scanning. We used Intercept Blocking Buffer (LI-COR), diluted 1:1 in Tris Buffered Saline (TBS) for blocking membranes (one hour at room temperature or overnight at 4°C) and preparing primary and secondary antibodies. Following primary incubation (1 hour at room temperature or overnight at 4°C) and secondary incubation (1 hour at room temperature at 1:10,000, we washed membranes in a TBS-T (1% tween solution) or TBS solution. Antibody stripping was done with 0.2N NaOH for 30 minutes, as needed for efficient stripping. Membranes were scanned on a LI-COR Odyssey system.

Antibodies: We used the following antibodies for verification of expression and phosphorylation: Myc antibody Mouse Bio X Cell BE0238 at 1:3000 dilution; Rabbit PY1000 antibody CST 8954S at 1:1500 dilution; EGFR pY992/EGFR pY1016 Rabbit CST 2235 at 1:1000 dilution; ERBB2 pY1196 (with cross-reactivity of EGFR pY1138 and pY1016) Rabbit CST 6942 at 1:1000 dilution; EGFR pY1068/EGFR pY1092 Rabbit CST 3777 at 1:1000 dilution.

We used the following secondary antibodies from LI-COR Biosciences (at 1:10,000 dilution): Donkey Anti-Mouse IgG 680RD, Donkey Anti-Rabbit IgG 800CW; Donkey Anti-Rabbit igG 680RD; Donkey Anti-Mouse IgG 800CW.

Native PAGE: Phosphorylation efficiency of phosphoproteins was evaluated by running protein samples by native PAGE to allow for separation of protein species with varying levels of phosphorylation. A 9% acrylamide separating gel was poured with 1.8 mLs 30% acrylamide, 1.5 mLs 1.5 M Tris-HCl, 600 μLs 10% APS, and 6 μL TEMED. A stacking gel was poured using 533 μL 30% acrylamide, 1 mL 1.5 M Tris-HCl, 40 μL 10% APS, and 8 μL TEMED. Gels were cast using the Bio-Rad Mini-PROTEAN® Tetra Handcast System. 1L of 1x Running Buffer (250mM Tris-HCl and 1.92M Glycine) was used, and gels were run at 4°C at 90V until separation was achieved. Samples were prepared in loading dye: 40% Glycerol, 248 mM Tris-HCl, 0.02% Bromophenol Blue.

Protein concentration estimation: We used PAGE-based separation and Coomassie staining (ThermoScientific Pierce Coomassie Brilliant Blue G250 part no 20279), followed by 680-infrared detection by LICOR for estimating protein concentration compared to a BSA standard curve.

Biolayer Interferometry (BLI). Experiments were run on the Gator® Pilot BLI instrument using Streptavidin (SA) probes (Gator SKU 160002), MAX plates (Gator SKU 130062) for probe loading, and black flat plates (Gator SKU 130150) for sample loading. A sampling rate of 10 Hz was used, and a temperature of 30°C was maintained within the instrument. 250 uL of buffer (50 mM Tris-HCl, 150 mM NaCl, 1% Bovine Serum Albumin) was loaded into the MAX plate to allow for equilibration of the probes, and a 10-minute, 400 rpm equilibration step was performed at the beginning of each experiment. 180 uL of sample was loaded into the black flat plate and held at a tilt throughout the experiment. Experiments were run with a 30 second baseline in buffer, 120 second load of biotinylated, phosphorylated protein, 30 second baseline in buffer, 60 second association of SH2 domain, and 60 second dissociation in buffer. The load wells contain equal concentrations of phosphorylated protein, whereas the association wells contain a serial dilution of SH2 domain within the first three wells of the column. The final well contains buffer to act as a reference. Plates were shaken throughout the experiment with PTPN11:GAB1 experiments being shaken at 1000 rpm and PLCγ1:EGFR experiments being shaken at 400 rpm.

BLI Analysis. Kinetic analysis was performed within the Gator One® software (version is 2.16.6.0130). Data was aligned on the Y-axis at the association step, and an inter-step correction was applied along with Savitzky-Golay filtering to reduce noise. The reference well is subtracted from the other three wells to account for non-binding artifacts within the signal. A 1:1 global binding model is fit to both the association and dissociation curves of each set of three serial dilutions with an unlinked Rmax to quantify the dissociation constant, K_D. BLI figures contained within this paper uses data processed in the manner described above, and both the data and binding fit curves have been re-plotted in MATLAB®. The data is subsampled to include every 10th data point. Data provided in the supplement was not subsampled but did undergo the other pre-processing steps.

ODE Modeling of tandem SH2 domain interactions. The original code was written by Dr. Wesley Errington in the Sarkar lab at the University of Minnesota (14) and shared by compressed folder by Dr. Sarkar. The code was adapted to include greater flexibility within the parameters. Most notably, additional k_{on} and k_{off} parameters were incorporated into the differential equations to allow for different affinities by each SH2 domain for each pTyr site, and persistence length was made universally adjustable instead of binning all protein interactions into two possible groups, flexible and rigid. The output of the original code is a series of plots mapping the binding behavior of the proteins under different initial concentrations of SH2 domain. Fifteen binding configurations are possible within a bivalent system, but for the purposes of this paper, only subsets were summed together and plotted as a single metric of binding. When analyzing tandem SH2 domains, the two bivalently bound configurations (inline and twisted) were isolated, whereas configurations detailing single domain binding were utilized when simulating monovalent binding. Additional code was written to extract the equilibrium concentration of each of these plots and feed that into a concentration effect curve (CEC). The CEC operates by plotting the fraction of bound species at equilibrium – calculated by dividing the equilibrium concentration by the starting concentration of phosphorylated protein (R0) – against the starting concentration of SH2 domain (L0). Within the model, the concentration of SH2 domain is assumed to remain constant throughout the simulation. The value of L0 that corresponds to 50% of the phosphorylated protein being bound at equilibrium is the dissociation constant (K_D). All results provided were generated using Matlab R2024b. All parameters have been provided in Data S1. We used the University of Virginia's High Performance Computing environment Rivanna for model predictions. All Matlab code is provided in our repository at https://github.com/NaegleLab/BivalentModeling_tanSH2.

Estimation of diameters from structure. For simulations provided in this paper, estimates of the molecular weight of SH2 domains and pTyr sites were calculated using the following website https://web.expasy.org/compute_pi/. For pTyr sites, the sequence spanning two sites N-terminal of the pTyr site and four sites C-terminal was used to estimate molecular weight. Sites Y627 and Y659 within Gab1 were calculated and used as an estimate for all pTyr pairs.

SH2 diameters We identified PDB entries containing the domain of interest (SH2) and their corresponding domain boundaries using CoDIAC (25). We included wild-type and mutant structures only if the mutations resided well outside the domain regions. To estimate domain diameters, we assumed each domain adopts an approximately spherical shape. Based on this assumption, the diame-

ter was defined as the longest distance between any two atoms within the domain. To compute this, we measured all pairwise distances between backbone atoms (N and C) within the defined domain boundaries and recorded the maximum distance as the domain's diameter for multiple PDB structures available for the same domain. Since the ODE code assumed the same size for both entities, and since we saw relatively small differences between the N and C terminal diameters, we averaged all diameters for both domains across all available structures. All diameter calculations are provided in Data S2. Code for extraction of domain diameters is provided at https://github.com/NaegleLab/BivalentModeling_tanSH2.

Estimation of linkers from structure. **pTyr linkers** Linker length was estimated as the number of amino acids between the pTyr sites or SH2 domains, unless a PDB file was available for the protein in which case linker length was estimated using the same PDB-extraction based approach as SH2 domains.

SH2 domain linkers Structures exist for all tandem SH2 domain family members and so we used those to estimate the linker lengths. We used UniProt defined SH2 domain boundaries to define the linker (as the first amino acid after the N-terminal SH2 domain and up to the amino acid before the start of the C-terminal SH2 domain). Then for all available structures, we calculated the average distance between all atoms of the two amino acids farthest apart from each other on the linker region (i.e. the residues immediately adjacent to the domains). All linker calculations are provided in Data S2. Code for extraction of linker distances is provided at https://github.com/NaegleLab/BivalentModeling_tanSH2.

Estimation of monovalent KD values. Estimates of monovalent K_D were taken from Ronan et al.s meta-analysis of fluorescence polarization experiments (20). All k_{off} values were set to 1 s^{-1}, and k_{on} values were adjusted to match the estimated K_D.

ACKNOWLEDGEMENTS

Research reported in this publication was supported by the National Institute Of General Medical Sciences of the National Institutes of Health under Award Numbers R35GM138127 and T32GM008715 (to Reagan Portelance). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We would like to thank Dr. Casim Sarkar and Dr. Wesley Errington for sharing the Matlab code and giving guidance on model adaptations. We would additionally like to thank Devon Semoy and Dr. Neel Shah for their helpful guidance in native PAGE for the separation of phosphospecies.

Bibliography

1. Ravi S. Kane. Thermodynamics of multivalent interactions: influence of the linker. *Langmuir : the ACS journal of surfaces and colloids*, 26(11):8636–8640, June 2010. ISSN 1520-5827 0743-7463. doi: 10.1021/la9047193. Place: United States.
2. Georges Vauquelin and Steven J. Charlton. Exploring avidity: understanding the potential gains in functional affinity and target residence time of bivalent and heterobivalent ligands. *British journal of pharmacology*, 168(8):1771–1785, April 2013. ISSN 1476-5381 0007-1188. doi: 10.1111/bph.12106. Place: England.
3. Natalia Nuñez-Prado, Marta Compte, Seandean Harwood, Ana Álvarez Méndez, Simon Lykkemark, Laura Sanz, and Luis Álvarez Vallina. The coming of age of engineered multivalent antibodies. *Drug discovery today*, 20(5):588–594, May 2015. ISSN 1878-5832 1359-6446. doi: 10.1016/j.drudis.2015.02.013. Place: England.
4. Angel M. Cuesta, Noelia Sainz-Pastor, Jaume Bonet, Baldomero Oliva, and Luis Alvarez-Vallina. Multivalent antibodies: when design surpasses evolution. *Trends in biotechnology*, 28(7):355–362, July 2010. ISSN 1879-3096 0167-7799. doi: 10.1016/j.tibtech.2010.03.007. Place: England.
5. Elizabeth A. Ottinger, Martyn C. Botfield, and Steven E. Shoelson. Tandem SH2 Domains Confer High Specificity in Tyrosine Kinase Signaling*. *Journal of Biological Chemistry*, 273 (2):729–735, January 1998. ISSN 0021-9258. doi: 10.1074/jbc.273.2.729.
6. Yunxin Deng, Artem K. Efremov, and Jie Yan. Modulating binding affinity, specificity, and configurations by multivalent interactions. *Biophysical journal*, 121(10):1868–1880, May 2022. ISSN 1542-0086 0006-3495. doi: 10.1016/j.bpj.2022.04.017. Place: United States.
7. N. Isakov, R. L. Wange, W. H. Burgess, J. D. Watts, R. Aebersold, and L. E. Samelson. ZAP-70 binding specificity to T cell receptor tyrosine-based activation motifs: the tandem SH2 domains of ZAP-70 bind distinct tyrosine-based activation motifs with varying affinity. *The Journal of experimental medicine*, 181(1):375–380, January 1995. ISSN 0022-1007 1540-9538. doi: 10.1084/jem.181.1.375. Place: United States.
8. J. Sloan-Lancaster, J. Presley, J. Ellenberg, T. Yamazaki, J. Lippincott-Schwartz, and L. E. Samelson. ZAP-70 association with T cell receptor zeta (TCRzeta): fluorescence imaging of dynamic changes upon cellular stimulation. *The Journal of cell biology*, 143(3):613–624, November 1998. ISSN 0021-9525 1540-8140. doi: 10.1083/jcb.143.3.613. Place: United States.
9. Timothy Travers, William K. Kanagy, Rachael A. Mansbach, Elton Jhamba, Cedric Cleyrat, Byron Goldstein, Diane S. Lidke, Bridget S. Wilson, and S. Gnanakaran. Combinatorial diversity of Syk recruitment driven by its multivalent engagement with FcεRIγ. *Molecular Biology of the Cell*, 30(17):2331–2347, August 2019. ISSN 1059-1524. doi: 10.1091/mbc.E18-11-0722. Publisher: American Society for Cell Biology (mboc).
10. Xiaozheng Xu, Takeya Masubuchi, Qixu Cai, Yunlong Zhao, and Enfu Hui. Molecular features underlying differential SHP1/SHP2 binding of immune checkpoint receptors. *eLife*, 10:e74276, November 2021. ISSN 2050-084X. doi: 10.7554/eLife.74276. Publisher: eLife Sciences Publications, Ltd.
11. Nikolaos Patsoukis, Jonathan S. Duke-Cohan, Apoorvi Chaudhri, Halil-Ibrahim Aksoylar, Qi Wang, Asia Council, Anders Berg, Gordon J. Freeman, and Vassiliki A. Boussiotis. Interaction of SHP-2 SH2 domains with PD-1 ITSM induces PD-1 dimerization and SHP-2 activation. *Communications Biology*, 3(1):128, March 2020. ISSN 2399-3642. doi: 10.1038/s42003-020-0845-0.
12. Anatoly Kiyatkin, Edita Aksamitiene, Nick I. Markevich, Nikolay M. Borisov, Jan B. Hoek, and Boris N. Kholodenko. Scaffolding protein Grb2-associated binder 1 sustains epidermal growth factor-induced mitogenic and survival signaling by multiple positive feedback loops. *The Journal of biological chemistry*, 281(29):19925–19938, July 2006. ISSN 0021-9258 1083-351X. doi: 10.1074/jbc.M600482200. Place: United States.
13. Jess M. Cunnick, Lin Mei, Craig A. Doupnik, and Jie Wu. Phosphotyrosines 627 and 659 of Gab1 Constitute a Bisphosphoryl Tyrosine-based Activation Motif (BTAM) Conferring Binding and Activation of SHP2 *. *Journal of Biological Chemistry*, 276(26):24380–24387, June 2001. ISSN 0021-9258. doi: 10.1074/jbc.M010275200. Publisher: Elsevier.
14. Wesley J. Errington, Bence Bruncsics, and Casim A. Sarkar. Mechanisms of noncanonical binding dynamics in multivalent protein–protein interactions. *Proceedings of the National Academy of Sciences*, 116(51):25659–25667, December 2019. doi: 10.1073/pnas.1902909116. Publisher: Proceedings of the National Academy of Sciences.
15. Tim Bilbrough, Emanuele Piemontese, and Oliver Seitz. Dissecting the role of protein phosphorylation: A chemical biology toolbox. *Chemical Society Reviews*, 51(13):5691–5730, July 2022. ISSN 1460-4744. doi: 10.1039/D1CS00991E.
16. Margaret M. Ryan, Reagan Portelance, Graham F. Newman, Gabrielle Martinez, Swathi Shekharan, Anqi Wu, Savannah Angel, Katherine E. Schaberg, Petra Gilmore, Robert Sprung, Reid Townsend, and Kristen M. Naegle. A signaling inspired synthetic toolkit for efficient production of tyrosine phosphorylated proteins. *bioRxiv*, page 2024.12.22.629992, January 2025. doi: 10.1101/2024.12.22.629992.
17. Huan-Xiang Zhou. Quantitative Relation between Intermolecular and Intramolecular Binding of Pro-Rich Peptides to SH3 Domains. *Biophysical Journal*, 91(9):3170–3181, November 2006. ISSN 0006-3495. doi: 10.1529/biophysj.106.090258.
18. Alexander F. Chin, Dmitri Toptygin, W. Austin Elam, Travis P. Schrank, and Vincent J. Hilser. Phosphorylation Increases Persistence Length and End-to-End Distance of a Segment of Tau Protein. *Biophysical Journal*, 110(2):362–371, January 2016. ISSN 0006-3495. doi: 10.1016/j.bpj.2015.12.013.
19. Guillaume Stirnemann, David Giganti, Julio M. Fernandez, and B. J. Berne. Elasticity, structure, and relaxation of extended proteins under force. *Proceedings of the National Academy of Sciences*, 110(10):3847–3852, March 2013. doi: 10.1073/pnas.1300596110. Publisher: Proceedings of the National Academy of Sciences.
20. Tom Ronan, Roman Garnett, and Kristen M. Naegle. New analysis pipeline for high-throughput domain–peptide affinity experiments improves SH2 interaction data. *Journal of Biological Chemistry*, 295(32):11346–11363, August 2020. ISSN 0021-9258. doi: 10.1074/jbc.RA120.012503.
21. Ronald J. Hause, Kin K. Leung, John L. Barkinge, Mark F. Ciaccio, Chih-pin Chuu, and Richard B. Jones. Comprehensive Binary Interaction Mapping of SH2 Domains via Fluorescence Polarization Reveals Novel Functional Diversification of ErbB Receptors. *PLoS ONE*, 7(9):e44471, September 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0044471.
22. Kin K Leung, Ronald J Hause, John L Barkinge, Mark F Ciaccio, Chih-Pin Chuu, and Richard B Jones. Enhanced prediction of Src homology 2 (SH2) domain binding potentials using a fluorescence polarization-derived c-Met, c-Kit, ErbB, and androgen receptor interactome. *Molecular & cellular proteomics : MCP*, 13(7):1705–23, July 2014. ISSN 1535-9484. doi: 10.1074/mcp.M113.034876.
23. Tomonori Kaneko, Rakesh Joshi, Stephan M. Feller, and Shawn SC Li. Phosphotyrosine recognition domains: the typical, the atypical and the versatile. *Cell Communication and Signaling*, 10(1):32, November 2012. ISSN 1478-811X. doi: 10.1186/1478-811X-10-32.
24. Georges Vauquelin and Steven J. Charlton. Long-lasting target binding and rebinding as mechanisms to prolong in vivo drug action. *British journal of pharmacology*, 161(3):488–508, October 2010. ISSN 1476-5381 0007-1188. doi: 10.1111/j.1476-5381.2010.00936.x. Place: England.
25. Alekhya Kandoor, Gabrielle Martinez, Julianna M. Hitchcock, Savannah Angel, Logan Campbell, Saqib Rizvi, and Kristen M. Naegle. CoDIAC: A comprehensive approach for

interaction analysis reveals novel insights into SH2 domain function and regulation. *bioRxiv*, page 2024.07.18.604100, January 2025. doi: 10.1101/2024.07.18.604100.

26. J E Ladbury, M a Lemmon, M Zhou, J Green, M C Botfield, and J Schlessinger. Measurement of the binding of tyrosyl phosphopeptides to SH2 domains: A reappraisal. *Proceedings of the National Academy of Sciences of the United States of America*, 92(8):3199–203, May 1995. ISSN 0027-8424.
27. Sivaraj Sivaramakrishnan and James A. Spudich. Systematic control of protein interaction using a modular ER/K α-helix linker. *Proceedings of the National Academy of Sciences*, 108 (51):20467–20472, December 2011. doi: 10.1073/pnas.1116066108. Publisher: Proceedings of the National Academy of Sciences.
28. Kathryn G. Wilcox, Marlee E. Dingle, Ankit Saha, Michael J. A. Hore, and Svetlana Morozova. Persistence length of α-helical poly-l-lysine. *Soft Matter*, 18(35):6550–6560, 2022. ISSN 1744-683X. doi: 10.1039/D2SM00921H. Publisher: The Royal Society of Chemistry.
29. Magnus Kjaergaard. Estimation of Effective Concentrations Enforced by Complex Linker Architectures from Conformational Ensembles. *Biochemistry*, 61(3):171–182, February 2022. ISSN 0006-2960. doi: 10.1021/acs.biochem.1c00737. Publisher: American Chemical Society.
30. Kamaldeep Gill, Jennifer L. Macdonald-Obermann, and Linda J. Pike. Epidermal growth factor receptors containing a single tyrosine in their C-terminal tail bind different effector molecules and are signaling-competent. *Journal of Biological Chemistry*, 292(50):20744–20755, December 2017. ISSN 0021-9258. doi: 10.1074/jbc.M117.802553.
31. Eiji Kinoshita, Emiko Kinoshita-kikuta, and Tohru Koike. Phos-tag SDS-PAGE systems for phosphorylation profiling of proteins with a wide range of molecular masses under neutral pH conditions. *Proteomics*, 12:192–202, 2012. doi: 10.1002/pmic.201100524.
32. D Rotin, B Margolis, M Mohammadi, R J Daly, G Daum, N Li, E H Fischer, W H Burgess, A Ullrich, and J Schlessinger. SH2 domains prevent tyrosine dephosphorylation of the EGF receptor: Identification of Tyr992 as the high-affinity binding site for SH2 domains of phospholipase C gamma. *The EMBO Journal*, 11(2):559–567, February 1992. ISSN 0261-4189. doi: 10.1002/j.1460-2075.1992.tb05087.x.
33. Tomonori Kaneko, Haiming Huang, Xuan Cao, Xing Li, Chengjun Li, Courtney Voss, Sachdev S. Sidhu, and Shawn S. C. Li. Superbinder SH2 Domains Act as Antagonists of Cell Signaling. *Science Signaling*, 5(243):ra68–ra68, September 2012. doi: 10.1126/scisignal.2003021. Publisher: American Association for the Advancement of Science.
34. John E. Ladbury and Stefan T. Arold. *Energetics of Src Homology Domain Interactions in Receptor Tyrosine Kinase-Mediated Signaling*, volume 488. Elsevier Inc., 2011. ISBN 978-0-12-381268-1. doi: 10.1016/B978-0-12-381268-1.00007-0.
35. T. C. Grammer and J. Blenis. Evidence for MEK-independent pathways regulating the prolonged activation of the ERK-MAP kinases. *Oncogene*, 14(14):1635–1642, April 1997. ISSN 0950-9232. doi: 10.1038/sj.onc.1201000. Place: England.
36. Silvia D. M. Santos, Peter J. Verveer, and Philippe I. H. Bastiaens. Growth factor-induced MAPK network topology shapes Erk response determining PC-12 cell fate. *Nature Cell Biology*, 9(3):324–330, March 2007. ISSN 1476-4679. doi: 10.1038/ncb1543.
37. Janine M. Buonato, Ingrid S. Lan, and Matthew J. Lazzara. EGF augments TGFβ-induced epithelial-mesenchymal transition by promoting SHP2 binding to GAB1. *Journal of Cell Science*, 128(21):3898–3909, January 2015. ISSN 1477-9137. doi: 10.1242/jcs.169599.
38. Christopher M. Furcht, Janine M. Buonato, and Matthew J. Lazzara. EGFR-activated Src family kinases maintain GAB1-SHP2 complexes distal from EGFR. *Science Signaling*, 8 (376):1–13, 2015. ISSN 19379145. doi: 10.1126/scisignal.2005697.
39. Kamaldeep Gill, Jennifer L Macdonald-obermann, and Linda J Pike. Epidermal growth factor receptors containing a single tyrosine in their C-terminal tail bind different effector molecules and are signaling-competent. *JBC*, 292(50):20744–20755, 2017. doi: 10.1074/jbc.M117.802553.