

UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

Scaffolding, Multisite Phosphorylation and Other Aspects of Regulation in Signal Transduction

Permalink

<https://escholarship.org/uc/item/4xz0t9mj>

Author

Lagunes, Leonila

Publication Date

2020

Peer reviewed|Thesis/dissertation

**UNIVERSITY OF CALIFORNIA,
IRVINE**

Scaffolding, Multisite Phosphorylation and Other Aspects of Regulation in Signal Transduction

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Leonila Lagunes

**Dissertation Committee:
Professor Lee Bardwell, Chair
Professor German Enciso
Associate Professor Jun Allard
Assistant Professor Zeba Wunderlich
Assistant Professor Scott Atwood**

2020

Chapter 2 © 2017 American Society for Biochemistry and Molecular Biology
Chapter 3 © 2020 PLOS Computational Biology
All other materials © 2020 Leonila Lagunes

DEDICATION

To

My family

Thank you for the immense support and love!

Para

Mi familia

¡Gracias por el inmenso apoyo y amor!

TABLE OF CONTENTS

	Page
LIST OF FIGURES	iv
LIST OF TABLES	vi
ACKNOWLEDGEMENTS	vii
VITA	viii
ABSTRACT OF THE DISSERTATION	ix
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: The WW domain of the scaffold protein IQGAP1 is neither necessary nor sufficient for binding to the MAPKs ERK1 and ERK2	37
CHAPTER 3: Effect of Magnitude and variability of energy of activation in multisite ultrasensitive biochemical processes	60
CHAPTER 4: The role of the Docking site on phosphorylation of native and novel sites of transcription factor c-Jun by JNK	93
CHAPTER 5: Summary and Conclusions	110

LIST OF FIGURES

	Page
Chapter 1	
Figure 1	Mechanisms of Signal Transduction Regulation 1
Figure 2	Post-Translational Modifications in the Cell 4
Figure 3	Free Energy in the Context of PTMs 10
Figure 4	Dose Response Curves and Multisite Modifications 16
Figure 5	Cooperative and Allosteric Models 23
Figure 6	JNK Signaling Pathway and Transcription Factor c-Jun 25
Chapter 2	
Figure 1	The IQGAP1 scaffold protein 39
Figure 2	The IQ domain of IQGAP1 is necessary for binding to ERK2; the WW domain is not sufficient 44
Figure 3	The IQ domain of IQGAP1 is sufficient for binding to ERK2; the WW domain is not necessary 46
Figure 4	Further characterization of the ERK2-IQGAP1 interaction 47
Figure 5	The IQ domain of IQGAP1 is sufficient for binding to ERK1, MEK1, and MEK2 49
Figure 6	The WW domain does not contribute to the ERK-IQGAP1 interaction 51
Chapter 3	
Figure 1	Generalized MWC Allosteric Model 64
Figure 2	Ultrasensitivity of MWC system 67
Figure 3	Activation Parameters and H in MWC 69
Figure 4	Independent Multisite Modification Model 74
Figure 5	Activation Parameters and H in Independent Model 75
Figure S1	Statistical weights of MWC modification states for n=3 89

Figure S2	Ultrasensitivity of MWC system	90
Figure S3	Parameters and H in MWC	91
Figure S4	Ultrasensitivity and total conformational free energy in MWC	92
Figure S5	Ultrasensitivity and total conformational free energy in MWC with Maintenance costs.	92
Chapter 4		
Figure 1	Transcription factor c-Jun has docking and four phosphorylation sites in residues 1-108	95
Figure 2	c-Jun sites phosphorylated at different rate and under control of D-site	99
Figure 3	Novel sites phosphorylated and under control of the docking site	101
Figure 4	Surrounding sequence in potential novel site	102
Figure 5	Comparison of native c-Jun residue type phosphorylation by JNK2 and its distance to D-site	104
Figure S1	Orientation of D-site determines c-Jun phosphorylation of single sites and Thr sites are under control of D-site.	109

LIST OF TABLES

		Page
Chapter 1		
Table 1	Sign Analysis of ΔG	11
Table 2	Overview of Mechanisms of Modification Factors	21
Table 3	c-Jun Protein Sequence Conservation	27
Chapter 2		
Table 1	Binding assay data for ERK2-IQGAP1 Interaction	41
Table 2	Oligonucleotides used in this study	42
Chapter 3		
Table 1	Ultrasensitivity at knee	71
Table S1	Ultrasensitivity at knee	88
Chapter 4		
Table 1	Oligonucleotides in this study	97

ACKNOWLEDGEMENTS

I would like to thank my family, friends and mentors who have provided endless support and guidance throughout my time here at UC Irvine.

A special thank you to my PhD advisors, Professors Lee Bardwell and German Enciso. I would like to thank Dr. Lee Bardwell for guiding me through an entire PhD program, pushing me to become a fearless scientist, and for showing me how to hold a pipette. Thank you for the millions of hours of mentoring and for showing me that I always have an ally in the room. I would also like to thank Dr. German Enciso for guiding me through the mathematical world, challenging me, and always encouraging me. Thank you for letting me pursue every idea I brought in and writing endless math on the board. Thank you both for being a wonderful source of inspiration and for believing in me.

A thank you to my committee members, Zeba Wunderlich, Scott Atwood, and Jun Allard for all the support from both biology and math worlds. My committee always challenged me to think bigger than myself and provided advice that guided the direction of my thesis over the years.

I would like to thank all the members of the Bardwell lab, past and present. A special thank you to Dr. Jane Bardwell who directly contributed to my scientific growth and for being an amazing science mom. Thank you for sharing your knowledge with me, everything from protocol to being a woman in science to Project Leda.

Words cannot begin to describe my gratitude to my family. My parents for all the support and guidance. My siblings, Keila, Julissa, and Isaac for always being so encouraging and making life amazing! This dissertation could not have been completed without them.

Thank you to Dr. Debra Mauzy-Melitz for being a wonderful GAANN director. My gratitude for believing in me and challenging me to become a better teacher. In addition, I would like to thank all the GAANN fellows for their incredible support in and off campus. A very special thank you to the MCSB community, including the 2014 cohort. Thank you all for the positive encouragement, feedback, and friendship. A special thank you to Lily, Kerrigan, Trini, Neelakshi, Bryan, and Emmanuel for the never-ending support and tea! I would also like to thank the CCBS staff; Karen for always having her door open to me, no matter the question. A thank you to the Developmental and Cell Biology staff who have helped tremendously over the years to complete this thesis smoothly and were always willing and available to assist with anything from putting together writing group to registering for classes my first quarter.

I would like to thank all my CSUF-MARC support, Dr. Amybeth Cohen for always being so supportive and for all the great times! Thank you to Erik, Trini, Rebecca, and Dr. Ruiz for being there for me on countless occasions. Thank you to Dr. Bridgeford as well for teaching to accept my accomplishments.

Last but not least, my never-ending gratitude to my MSC family. Vanessa and John for being there through my entire career. A special thank you for Ariel and Bella for always brightening my days! Thank you to my support systems at MSC: David, Hannah, Saul, Mayra, Andrea, Professor Manafi and everyone else at MSC.

Financial support was granted by various sources over the years: by the Department of Education Graduate Assistance in Areas of National Need grant number P200A120207 & P200A150181, the National Science Foundation Graduate Research Fellowship Program and by the UCI Graduate Division President's Dissertation Year Award.

VITA

Leonila Lagunes

- 2011 A.A. in Mathematics, Santiago Canyon College
- 2014 B.A. in Applied Mathematics, California State University-Fullerton
- 2015 Mathematical, Computational and Systems Biology Gateway Program Graduate, University of California Irvine
- 2017 M.S. in Developmental and Cell Biology , University of California, Irvine
- 2020 Ph.D. in Biological Sciences, University of California, Irvine

FIELD OF STUDY

Mathematical and Computational Systems Cell Biology.

PUBLICATIONS

Lagunes, L., Bardwell, L., and Enciso, G. *Effect of magnitude and variability of conformational free energy in multisite ultrasensitive biochemical processes*. 2020. PloS Comp. Biol. *Accepted*.

Lagunes, L. and Lee, C.H. *Cancer Screening Using Biomimetic Pattern Recognition with Hyper-Dimensional Structures*. 2018. *IEEE 18th IEEE International Conference on Bioinformatics and BioEngineering (BIBE)*. Taichung, 2018, pp. 201-206. Doi: 10.1109/BIBE.208.00046

Bardwell, A.J., **Lagunes, L.**, Zebarjedi, R., and Bardwell, L. *The WW domain of the Scaffolding protein IQGAP1 is neither necessary nor sufficient for binding to the MAPKs ERK1 and ERK2*. 2017. *The Journal of Biological Chemistry*. Vol. 292. Pgs. 8750-8761.

AWARDS

UC President's Dissertation Year Fellowship

National Science Foundation Graduate Research Fellowship Program Recipient

Graduate Assistance in Areas of National Need Fellowship Recipient

UCI Initiative for Maximizing Student Development Program Fellowship

Minority Access to Research Careers (MARC) Program Fellowship Recipient

Ronald E. McNair Post-Baccalaureate Achievement Program Fellowship Recipient

ABSTRACT OF THE DISSERTATION

Scaffolding, Multisite Phosphorylation and Other Aspects of Regulation in Signal Transduction

by

Leonila Lagunes

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2020

Professor Lee Bardwell, Chair

For cells to respond appropriately and timely to internal and external stimulus, they rely heavily on signal transduction cascades to regulate gene and protein activity and response. Protein activity is often regulated by ligand binding, scaffolding, or by post-translational modifications such as multisite phosphorylation. The Mitogen-Activated Protein Kinases (MAPKs) are multisite proteins with critical roles in development and various diseases. The c-Jun N-terminal Kinases (JNK) cascade is one of many MAPK cascades with multisite proteins. Here, we consider mathematical and experimental analyses to understand scaffolding and multisite modifications as protein regulation mechanisms.

Scaffolding is a regulatory mechanism for signal transduction where scaffold proteins bind to many components of a signaling pathway. MAPK scaffold proteins, such as IQ-motif-containing GTPase-activating protein 1 (IQGAP1), are promising targets for novel therapies against cancer and other diseases. Scaffolding is a regulatory mechanism for signal transduction where scaffold proteins bind to many components of a signaling pathway. Thus, it is important to know which domains on IQGAP1 bind to which MAP kinases for developing new therapies. Here, we show with *in vitro* binding assays that the IQ domain of IQGAP1 is both necessary and sufficient for binding to ERK1 and ERK2. Additionally, we show that the WW domain is neither necessary nor sufficient for binding to ERK1 or ERK2. These findings prompt a re-evaluation of how IQGAP1 regulates MAPK cascade proteins.

Protein phosphorylation also regulates a substrate's enzymatic activity, location, stability, and/or interactions with other proteins. Moreover, proteins that are regulated in this way often contain multiple modification sites. In the JNK pathway, transcription factor c-Jun is a multisite substrate shown to regulate cell responses such as cell proliferation, apoptosis, and DNA repair. Understanding kinase-substrate specificity in MAPKs is crucial to advancing medical therapies for diseases. Phosphoproteomic studies have provided insights into the conservation of phosphosites and their evolution across species. However, not much is known about the constraints novel sites experience. Here, we demonstrate with *in vitro* kinase assays that the Docking site (D-site) in c-Jun plays a significant role in the phosphorylation of all native and novel sites of c-Jun. Results indicate that the D-site is necessary for phosphorylation of native and novel sites of c-Jun by JNK2 enzyme.

Mathematical models are useful to study complex biological phenomena such as multisite proteins (i.e., proteins with $n > 1$ modifications sites). Proteins with multiple sites on which they can be modified or bound by ligand have been observed to create an ultrasensitive dose response. Here, we consider the contribution of the individual modification/binding sites to the activation process, and show how their individual values affect the ultrasensitive behavior of the overall system. We use a generalized Monod-Wyman-Changeux (MWC) model that allows for variable free energy contributions at distinct sites, and associate a so-called activation parameter to each site. Our analysis shows that ultrasensitivity generally decreases with increasing activation parameter values and depends on their mean and not on their variability. Additionally, results suggest that a protein can increase its ultrasensitivity by evolving new sites. These results provide insights into the performance objectives of multiple modification/binding sites and thus help gain a greater understanding of signaling and its role in diseases. Together, mathematical and experimental analyses show promising insights into signal transduction regulation through scaffolding, multisite PTMs, and docking.

Chapter 1: Introduction

For cells to successfully respond to internal and external environment stresses, they rely heavily on transduction pathways. Signal transduction provides a way for cells to achieve their performance objectives by responding to stimulus. When a cell receives a signal, typically initiated by the binding of an extracellular ligand to a receptor on the cell membrane, a transduction cascade is activated, **Figure 1a**. The signal is then relayed to reach the appropriate target for response. The response can be anything from altering gene expression, metabolism or cell structure to triggering cell proliferation **Figure 1a**. Between the stimulus and the response, a transduction cascade needs to function properly. Multiple different mechanisms regulate signal transmission, such as scaffolding, ligand or protein binding, and post-translational modifications, leading to the conclusion that protein-protein interactions are key to understanding signal transduction. Unfortunately, many diseases arise when there are defects in signal transduction pathways, such as cancer, diabetes, autoimmune and heart disease. Defects in regulation can lie anywhere in the signal transduction process. Thus, understanding transduction regulation is important for developing new treatments and to further study the process of development. This chapter includes an overview of key aspects of signal transduction regulation like scaffolding and post-translational modifications (PTMs).

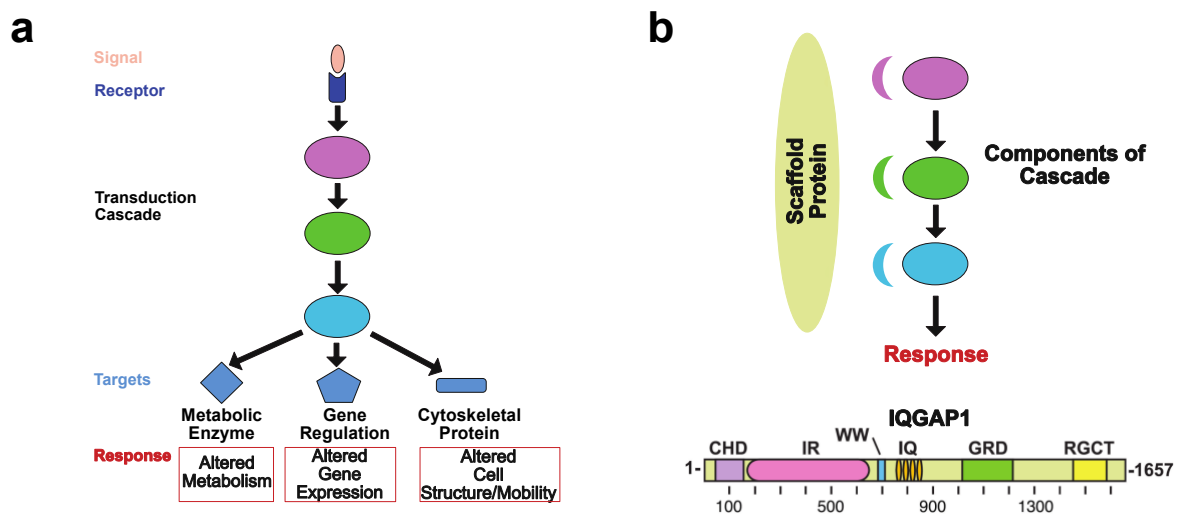


Figure 1: Mechanisms of Signal Transduction Regulation. (a) General schematic of signal transduction pathway. Signal binds to receptor triggering a transduction cascade that ultimately triggers an appropriate response. (b) Scaffold proteins binding to different components of a transduction cascade. Bottom is linear representation of full length human IQGAP1 protein and its known domains. The CHD domain represents the calponin homology domain, the IR is the internal repeated sequence/coiled-coil domain. The WW is the WW domain, IQ domain is made up of

four Isoleucine-Glutamine (IQ) repeats, GRD is the GTPase-activating protein-related domain, and the RGCT is the RasGAP C-terminal domain. Adapted from Reference¹.

Section 1: Scaffold Protein IQGAP1

Scaffolding is a regulatory mechanism for signal transduction where scaffold proteins bind to many components of a signaling pathway. It is theorized that scaffold proteins bind to many components of a cascade (1) to maintain the specificity of the signaling pathway² by localizing components to the same cellular region and/or (2) to serve as catalyzing agents to activate pathways³. However, the theory that scaffold proteins regulate pathway activation is still largely untested.

There are many known scaffolding proteins, (for a review refer to References^{2,4}) For example, in yeast, Ste5 is a scaffold protein involved in the Fus3 Mitogen Activated Protein Kinase (MAPK) pathway that binds to the kinases in the cascade². Ste5 is a scaffold protein that joins the kinases of the mitogen-activated protein kinase (MAPK) cascade for efficient Fus3 activation in yeast mating⁵, making Ste5 a key player in MAPK pathway regulation. There is also scaffold protein c-Jun N-terminal kinase (JNK) interacting protein (JIP1) from the JNK pathway². JIP1 is known to associate with many kinases in the JNK pathway and is theorized to control JNK pathway activation³. JIP1 is also shown to be involved in Alzheimer's disease and differentiation of neurons³, thus being one of many critical scaffolds that regulate signaling pathways involved in development and disease.

Here, we discuss IQ-motif-containing GTPase Activating Protein 1 (IQGAP1) – a scaffold protein that mediates protein-protein interactions of complexes that regulate cytoskeletal formation and intracellular signaling¹. Previous studies show that IQGAP1 interacts with the MAP Kinase pathway⁶, possibly facilitating signal transduction through the MAPK cascade⁷. **Figure 1b** shows a linear schematic of IQGAP1 with key domains highlighted. For many years, it was believed that ERK1 and ERK2 bind to the WW domain of IQGAP1^{8,9}. Other studies showed that IQGAP1 binding to MEK1 and MEK2 (MAPK kinases which phosphorylate and activate ERK1 and ERK2) required the IQ domain¹⁰. Since IQGAP1 binds to so many components of the MAPK pathway, it is crucial to understand how IQGAP1 is interacting and regulating the cascade. Overexpression of IQGAP1 has been found in cancers such as colorectal and ovarian¹¹. Studies also suggested that the WW domain should be targeted to inhibit MAPK signaling¹². Understanding the interactions of IQGAP1 and MAPK pathway kinases can help provide targets for new cancer therapies and other diseases.

Section 2: Multisite Modification as a Regulatory Mechanism

2.1 Post-Translational Modifications in a Cell

Although scaffold proteins provide an effective mechanism of signal transduction regulation, it is reasonable to assume that cascades rely on other, or a combination of, mechanisms to respond to stimuli. Across species, there are conserved PTMs that play key roles in regulating cell functions, at least between mammalian and plant cells¹³. In fact, mammalian cells also use PTMs to regulate transduction cascades; modifications such as protein phosphorylation, methylation, and ubiquitination are a few PTMs. Some consider proteolysis a PTM as well¹⁴. **Figure 2a** includes common PTMs known (Adapted from <http://tcohenlab.web.unc.edu/tau-post-translational-modification-puzzle/>). Interestingly, many proteins are multisite proteins, meaning they can be modified on more than one location. Different PTMs play important roles in cell functions including cell division¹⁵, immune response^{14,16}, and aging¹⁷. Different signals can lead to distinct modifications and thus to the correct response, as observed in Sp-family proteins¹⁸. Understanding PTMs in cell biology is important to further develop existing biological models and build new therapies for diseases.

The modification state of Sic1 in yeast helps regulate cell division. The cyclin kinase inhibitor Sic1 in yeast is targeted for degradation after it is phosphorylated¹⁵. More specifically, kinase Cdc4 phosphorylates Sic1 to trigger cell division¹⁵. Sic1 needs to be phosphorylated on at least six of its nine sites in order for its degradation to be triggered. This degradation allows for proper cell cycle transitions¹⁵.

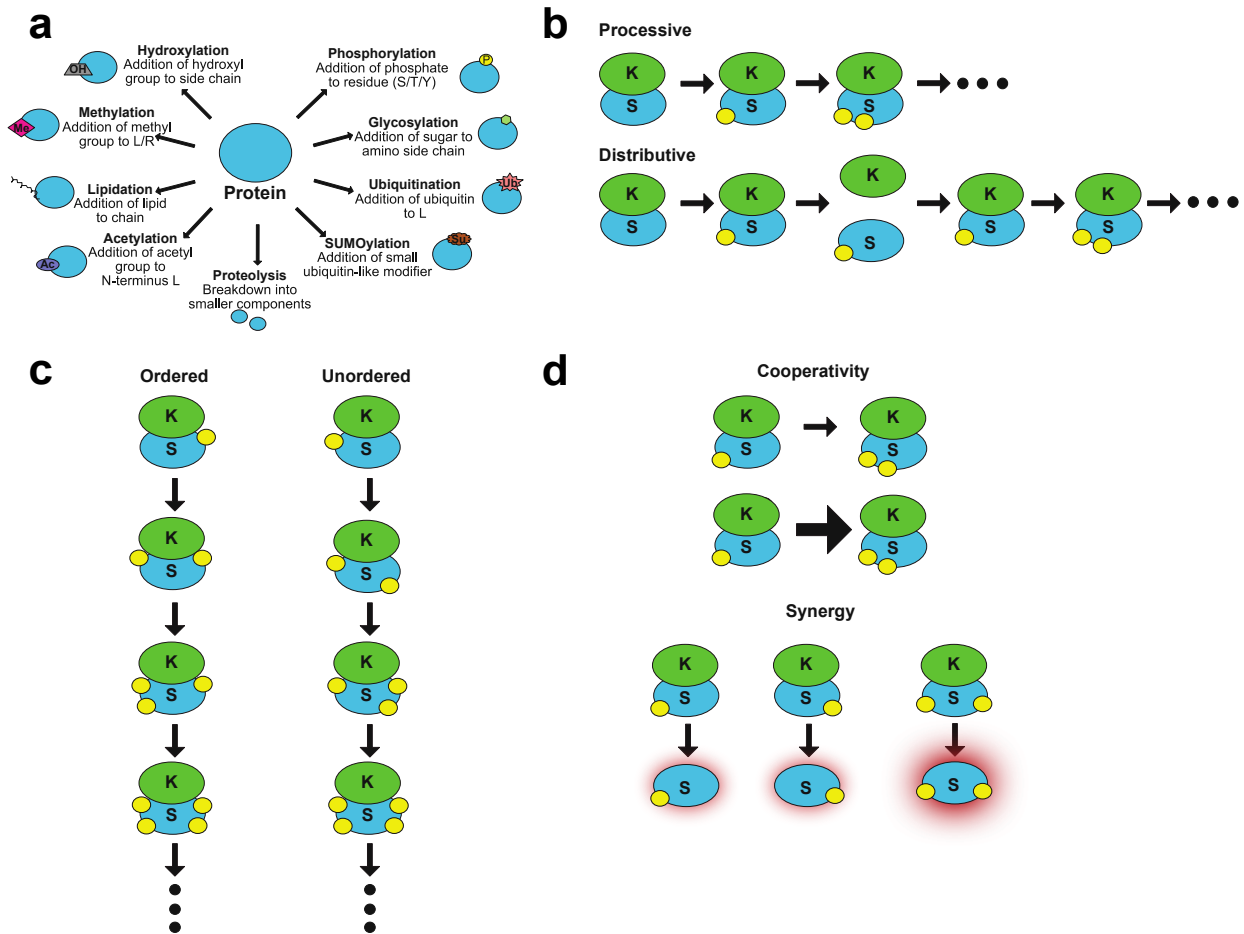


Figure 2. Post-Translational Modifications in the cell. (a) Different types of PTMs observed in cell biology (Adapted from <http://tcohenlab.web.unc.edu/tau-post-translational-modification-puzzle/>). (b) Processivity of modifications. K is the enzyme and S is the substrate. Processive mechanism of modification where one binding event can lead to modification of multiple sites vs Distributive where one modification occurs per binding event. Adapted from Reference¹⁹. The sequential mechanism of modification, where modifications may occur in an ordered fashion or unordered. (d) Cooperativity occurs when the modification of one site enhances (in this example) the modification of a second, denoted by the bigger arrow on the second set as a more favorable reaction. Synergy occurs when the modification of both sites yields a higher response (red glow) than the sum of the yield of the singly modified substrates (lower red glow). Adapted from Reference¹⁹.

In the immune response, PTMs of antigens can influence proper recognition of pathogens, thus showing that PTMs can affect the processing of foreign and native antigens in autoimmune responses¹⁶. More specifically, in a variety autoimmune diseases, immune response is dependent on the modifications of target antigens (Table 2 in Doyle et al.¹⁶). During bacterial infections, bacteria have been shown to interfere with PTMs in the host cells for their own survival and replication by directly affecting activation/inhibition of cell responses¹⁴. Table 1 in Reference¹⁴ shows a collection of bacterial pathogens that interfere with host PTMs for their own survival.

Even in aging, PTMs regulate proteins involved in senescence¹⁷. Some PTMs cause changes/damages to long-lived proteins and thus play a role in the aging process. In fact, there are numerous deleterious non-enzymatic modifications observed in *in vivo* systems¹⁷. For example, proteins in ocular lenses accumulate damage over time, leading to senile cataracts. These proteins are damaged through anodic shifts (greater electronegativity); usually via PTMs such as deamidation, carbamylation, phosphorylation, and oxidation of methionine¹⁷. This serves to show the significance of understanding PTMs as both regulatory and pathological mechanisms.

Aside from the fact that PTMs regulate cell responses, different PTMs yield unique protein activities in proteins like p53²⁰, ion channels²¹, and Sp-family proteins¹⁸. Tumor suppressor protein p53 has more than 30 modification sites²⁰. Having so many different possible combinations of modifications may explain p53's ability to exert different functions²⁰. For example, phosphorylation of Ser15 and Ser20 reduces affinity for negative regulator Hdm2 and promotes recruitment of transcriptional activators. However, phosphorylation of Ser46 induces pro-apoptotic genes, while phosphorylation of Ser392 (after UV exposure) activates DNA binding by stabilizing p53 tetramers²⁰, further suggesting that different modifications of distinct sites regulate different protein activities.

Prabakaran et al. have also shown that different modification-forms elicit different responses within a cell²¹. In membrane signaling proteins, the channel's conductance depends on how many modifications are in place. In microtubules, it has been shown that motility and cytokinesis depend on the location and number of polyglycosylated glutamate residues in beta tubulin²¹. In Way et al., authors show that phosphorylation, acetylation, glycosylation, and SUMOlation of Sp-family proteins in plants control Sp-protein activity¹⁸. For example, phosphorylation of Sp-1 at residues S56 and S101 yields increased binding affinity to DNA, while glycosylation of Sp-1 results in protein translocation¹⁸. This shows that different modifications at distinct sites have a different effect on protein activity. Ultimately, most PTMs help regulate a protein's activity.

Regardless of a protein's function, phosphorylation is theorized to be a strong regulatory mechanism for substrates. However, it is unclear why multisite phosphorylation and modification is so common among proteins. How do enzymes 'know' (1) how to accomplish the modification of a multisite substrate, (2) which sites are regulatory sites, and (3) which proteins are their respective substrates? Here, we explore the current understanding of such mechanisms, among other things.

2.2 Factors of Post-Translational Modifications

Given an un-modified substrate with multiple modification sites available, what are the mechanisms that determine which sites are modified, and with what efficiency? In this section, we will focus on multisite protein phosphorylation as an example. When considering multisite phosphorylation, we can theoretically consider different factors that pertain to modifications such as (1) the rate at which a site is modified, (2) the binding affinity of the kinase to the substrate, (3) the kinase processivity, (4) the sequence of modifications, and (5) interactions between the modification sites.

The *rate* (velocity) at which a single site is modified may differ relative to another site on the same substrate. For instance, Tyrosine Hydroxylase (TH) is phosphorylated *in vitro* by calcium-calmodulin-dependent protein kinase 2 (CaMKII), and the two Serine residues are phosphorylated at different rates based on mass-spectrometry data²². Bevilacqua and colleagues show that the phosphorylation of one of the Serine sites leads to a conformational change of TH and thus makes the second Serine site more accessible to the kinase²².

The phosphorylation rate is not the only aspect of interest, in fact, the *binding affinity* of the kinase to the substrate is also crucial for understanding how phosphorylation occurs. The strength at which the kinase binds to the substrate plays a role on the rate of phosphorylation. In theory, it is possible to assume that if the kinase is binding weakly to the substrate, the rate of phosphorylation is low and vice versa. For example, disassociation constant K_d values have been measured for different kinase-substrate complexes in the MAPK cascade using direct *in vitro* binding assays^{23,24}. To provide an example, Ho and colleagues approximate the K_d of JNK1 binding to MKK7 *in vitro* to be 70-100 μ M, (MKK7 phosphorylates JNK1)^{23,24}. Note that the smaller the K_d values are, the stronger the binding affinity is. Thus, the binding affinity of the kinase-substrate complex can play a significant role on phosphorylation, ultimately regulating signal transduction specificity.

Some substrates contain more than 10 sites phosphorylated by the same kinase²⁵. The question becomes, does how does the enzyme modify all the sites? Theoretically, one possibility is that the enzyme binds to the substrate and modifies all the sites during a single binding event, termed *processive* modification. A second possibility is that the enzyme binds to the target and modifies one site, then unbinds and must re-bind to modify another site, i.e. one modification per binding collision, termed *distributive*. It is also likely that some systems have some combination of the two²⁶, leading to more complexity. For a review on the processivity of phosphorylation refer to¹⁹. An example of processive phosphorylation is in²⁷, where a synthetic peptide of Hck, a kinase from the Src family, phosphorylates the Tyrosine sites on its target (SH2 ligand-containing peptides) in a processive manner. Aubol and colleagues also show that Serine/Arginine

(S/R) protein-specific kinase 1 (SRPK1) phosphorylates ASF/SF2 in a fully processive manner²⁸. As a biological example of distributive phosphorylation, consider transcription factor ATF2 which can be phosphorylated by p38²⁹. Waas and colleagues show that p38 phosphorylates two Threonine residues in a distributive manner²⁹. **Figure 2b** shows a comparison of processive and distributive modification.

Once the enzyme binds to its target, do the combinatorial modifications of the sites matter? More specifically, once modification is initiated, what is the *order* on which the sites are modified? It is possible that sites are modified randomly, in a specific order (**Figure 2c**), or a combination of the two. For example, suppose there are three sites s1, s2, and s3. Will they be phosphorylated in that N-terminal to C-terminal order or in some unordered fashion? In fact, p38 phosphorylates ATF2 in a distributive but unordered manner; Threonine residues can be phosphorylated in any direction²⁹. Similarly, ERK2 phosphorylation by MEKs is shown to be unordered¹⁹. There is also evidence of ordered phosphorylation. Not only does SRPK1 phosphorylate ASF/SF2 in a processive manner, there is an order to the sequence of phosphorylations²⁸. The three sites on Lethal giant larvae (Lgl) in *Drosophila* are phosphorylated by atypical Protein Kinase C (aPKC) in a specific, preferred order³⁰. In rhodopsin phosphorylation, using mass spectrometry, the phosphorylation (and dephosphorylation) occurs in an ordered manner³¹. It is also likely that phosphorylation of a substrate occurs in one sequence and dephosphorylation in another¹⁹. Conceivably, it is possible for substrates with many sites to contain regions where phosphorylation occurs in an ordered fashion while other regions are unordered. When considering the sequential mechanism of modification (and de-modification), substrates can exist in so many states, depending on the mechanism, that obtaining experimental data may prove to be difficult. Mathematical models may be necessary to fully understand the mechanisms of multisite PTMs.

An additional mechanism of PTM regulation lies in possible connections between sites such as (1) cooperativity, (2) synergy, and (3) allostery. *Cooperativity* refers to the scenario where modification of one site enhances (positive cooperativity) or inhibits (negative cooperativity) the modification of a second site. This means that given the first site is modified, the modification of the second site is more favorable (or less) than equally favorable. **Figure 2d** depicts phosphorylation as an example of cooperativity between sites where having the first site phosphorylated, the second phosphorylation event is more likely to occur than if the first phosphorylation was not already present (shown as a denser arrow along the second reaction). In *in vitro* experiments with sodium-hydrogen exchanger regulatory factor 1 (NHERF1) and PKC show there is cooperativity between two sites³². More specifically, the phosphorylation of the Threonine site at residue 95 enhances the phosphorylation of Ser77. Although there are eukaryotic

pathways with substrates containing cooperative sites, proteomic analysis show that not all substrates contain them³³.

Given a substrate with two modification sites, *synergy* is when the interactions of having both sites modified yield a higher effect in protein activity than the sum of the effect of a single site, **Figure 2d** (bottom). For example, consider the three phospho-Serine sites on Lgl that are phosphorylated by aPKC³⁰. The phosphorylation of any one site has been shown to have a much smaller effect on cortical displacement than when all sites are phosphorylated³⁰.

Sometimes confused with cooperativity is allostery. The concept of *allostery* states that ligand binding or modification at one site results in a conformational or activity change at a *distinct* site (Definition from Reference³⁵). The most common example of allostery is hemoglobin. Generally, when one oxygen molecule binds to one of the four sites on hemoglobin, this causes a conformational change in the remaining three sites allowing other oxygen molecules to bind to hemoglobin relatively easier. Since the description of allosteric hemoglobin, there have been many other biological examples of allosteric proteins. For example, phosphor-2-dehydro-3-deoxyheptonate aldolase has been shown to be an allosteric modulator³⁶. In fact, the Allosteric Database provides an excellent resource for allosteric molecules^{37–39} and can be found at (<http://mdl.shsmu.edu.cn/ASD/module/mainpage/mainpage.jsp>) with a large amount of data readily available.

Given the above, enhancing our understanding of the mechanisms of multisite modifications is crucial in moving forward. Although there is extensive biochemical knowledge on the processes of PTMs, there is still a great deal to learn about signal transduction regulation through PTMs. How modification of distinct sites regulate protein activity is still poorly understood but can shed some insights on both development, organism evolution, and medical advances.

2.3 Regulatory Modification Sites

Modifications regulate different responses in a cell. Ubiquitination of p53 results in nuclear transport⁴⁰. In 2013, Nie et al. showed that phosphorylation of Ser418 in FOXP3 regulates transcriptional activity responsible for T cell suppressive functions⁴¹. Additionally, F-box protein S-phase kinase associated protein 2 (Skp2) is acetylated by p300 at Lys68 and Lys71⁴². This acetylation leads to increased Skp2 stability and, as a result, oncogenic function also increases⁴². There are many more examples of different functions PTMs can regulate.

Determining which sites (or combination of sites) have downstream regulatory effects is a challenging task, particularly when considering that proteins containing multiple sites can be modified by different kinases/proteins that give unique responses. For some multisite substrates, different functions are activated when distinct sites are modified²⁵. This requires deeper understanding of how multisite modification regulates protein function. The complexity of multisite proteins is a feature used by experimentalists to determine sites of interest for targeting. Thus, an important step in understanding the effects of multisite phosphorylation, and other PTMs, on signal transduction events is to determine which specific sites (or combination of sites) contribute to protein activation/inhibition.

Experimental data show how some multisite proteins at distinct phosphoforms perform different objectives. For example, p53 is a multisite protein with more than 16 reported sites²⁵. Phosphorylation of Ser20 affects p53's binding to other proteins, whereas phosphorylation of Thr81 affects its stability²⁵. As discussed above, the multisite phosphorylation of p53's Serine residues 18 and 23 play a role in activating p53-dependent apoptosis that other p53 sites do not⁴³.

Phosphorylation has also been shown to regulate transcription factor activity²⁵. Transcription factor Gli1 in the Hedgehog Pathway binds to DNA depending on its phosphorylation state⁴⁴. Hh signaling is initiated when Hh protein binds to receptor Ptch1. This binding activates Smo, which in turn, by an unknown mechanism, promotes the dissociation of Sufu from Gli1, a multisite protein⁴⁵. This dissociation allows Gli1 to be available for further phosphorylation. In fact, Atypical Protein Kinase A (aPKC) has been shown to phosphorylate Gli1 in the DNA binding domain, increasing the binding affinity of Gli1 to DNA⁴⁵. However, when S6K1 phosphorylates Gli1 at Ser84 in the Sufu binding domain, Gli1 is released from its sequesterer Sufu⁴⁶. As a third example, consider transcription factor c-Jun with a total of six phosphorylation sites targeted by JNK⁴⁷. Experimental evidence shows that phosphorylation of the first four sites yields transactivation while phosphorylation of the last two sites are associated with Jun protein degradation⁴⁷.

Since so many proteins contain multiple modifications sites, the combination of modifications a substrate can exist in increases considerably relative a single-site protein. A substrate with one modification site can exist in only one of two states: un-modified or modified. A protein with 2 sites, can be in one of 4 modification-forms. A substrate with n sites can exist in a maximum of 2^n phosphoforms. At a given time, the population of all the substrate molecules in a cell is made up of substrates in different modification-states. With such complexity, mathematical models are necessary for in-depth analysis of the mechanism and function of multisite modifications on protein activity. This brings up an interesting question, if the

modification of some sites contributes more than others to the overall activation of a protein, how does this phenomenon affect signal transduction regulation? Or more specifically, how does this phenomenon affect the energy associated with the modification and protein activation? To address this, we need to understand the free energy of individual sites.

2.4 Gibbs Free Energy of Regulatory Sites

Here, I introduce *Gibbs Free Energy* in the context of PTM's. Gibbs free energy can be thought of as the "available" energy in a system and is mathematically defined as:

$$\Delta G = \Delta H - T \Delta S, \quad (\text{Eq. 1})$$

where ΔH is enthalpy, T is temperature in Kelvin, and ΔS is the entropy of the system. ΔG is usually given in kJ/mol units and is a very useful measurement for evaluating the spontaneity of chemical reactions. To understand ΔG , I believe it is crucial to first explore the individual components of Equation (1) above.

Enthalpy, ΔH , is the amount of heat or energy that is absorbed or released in a system at a constant pressure, also known as the "potential energy." Consider the ball at the top of a hill in **Figure 3**. The potential energy for the ball at the top of the hill is relatively high next to the potential energy at the bottom of the hill since more heat/energy can be released while at the top of the hill. In this example, the system's enthalpy has decreased when the ball rolls down the hill, making $\Delta H < 0$. Holding the other terms in Equation (1) constant will allow ΔG to decrease.

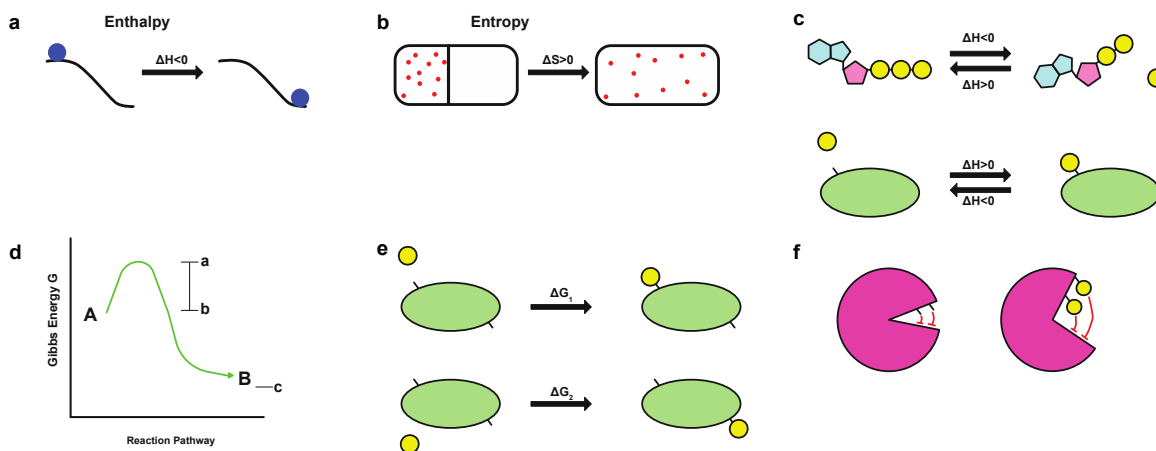


Figure 3: Free Energy in Context of PTMs. Figure depicting key concepts of Gibbs Free Energy ΔG (a) Enthalpy example of ball at top of hill and bottom of hill. Here the forward reaction has a negative ΔH as energy is released. There is more potential energy when ball is at top of the hill than when the ball is at the bottom of the hill. (b) Entropy example where particles are in a contained environment with a barrier vs when the barrier is removed.

Entropy increases, $\Delta S > 0$ since the system is more disordered after the barrier is removed. (c) Cartoon depiction of protein phosphorylation reaction. (Top) An ATP molecule hydrolyzes to ADP and a free phosphate group. In this reaction, energy is released when the bonds between the β and γ phosphates in ATP are broken and vice versa. (Bottom) In phosphorylation, a free phosphate group is covalently bound to a residue on a substrate. The creating of this new bond requires energy and thus $\Delta H > 0$. (d) Example of Gibbs Energy (G) for the reaction pathway that is favorable $A \rightarrow B$. The distance between a and b is the activation energy and c is G at the end of the reaction. In this example $\Delta G < 0$. (e) Example of distinct ΔG values for the phosphorylation event of distinct sites. For example, the phosphorylation of the first site can have an associated ΔG that differs from the phosphorylation of the second site. (f) The ΔG associated with activating a protein and a single site can also vary from site to site. In this example, the phosphorylation of each site contributes to the mechanical conformational change. However, one site might contribute more than another.

Entropy, ΔS , is a measurement for the disorder in a system. For example, let's consider diffusion. Suppose that the molecules in **Figure 3b** are concentrated to the left of a membrane. If the membrane were to be removed, the molecules would diffuse throughout the entire environment, making the system more disordered or more random (to be precise, when the new equilibrium is established, it is much less probable that the molecules will end up clustered together on the left side than that they will end up more spread out). Thus, the system's entropy increases, making $\Delta S > 0$. Holding the other terms in Equation (1) constant will also allow ΔG to decrease.

Now that we understand the components of Equation (1), what can we learn from ΔG ? Let's analyze the sign of ΔG . If $\Delta G < 0$, the system is *spontaneous or exergonic*. Here, free energy is released and will be referred to as "favorable." When $\Delta G > 0$ then the system is *non-spontaneous and endergonic*, where energy is absorbed and referred to as "unfavorable." If $\Delta G = 0$ then the system is in equilibrium. Thus H , T and S can affect a reaction's spontaneity. An interesting question would be under what conditions is a reaction favorable? In **Table 1**, we can see under what conditions a reaction will yield $\Delta G < 0$, where the two columns under ΔG depend on whether the temperature T is low or high.

TABLE 1
Sign Analysis of ΔG

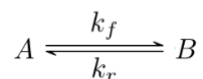
ΔH	ΔS	ΔG	
		T (low)	T (high)
Positive (+)	Positive (+)	Positive (+)	Negative (-)
Positive (+)	Negative (-)	Positive (+)	Positive (+)
Negative (-)	Positive (+)	Negative (-)	Negative (-)
Negative (-)	Negative (-)	Negative (-)	Positive (+)

Consider Gibbs free energy in the context of a PTMs. As an example, let's consider protein phosphorylation. Here, phosphorylation requires the reaction where an ATP molecule hydrolyzes to form an ADP molecule and a phosphate group that is transferred to the substrate (**Figure 3c top**). Here, there

is energy being released from the bonds between the beta and gamma phosphates in ATP. The addition of a phosphate group to a protein's amino acid residue would require energy (**Figure 3c** bottom). Looking at the complete phosphorylation reaction, if the amount of energy released from the conversion of ATP to ADP + a phosphate group is greater than the amount of energy absorbed when adding the phosphate group to an amino acid residue, then $\Delta H < 0$. Although the ΔH of phosphorylation can be negative, this does not necessarily mean that phosphorylation is a spontaneous reaction. From **Table 1**, depending on ΔS and T , the favorability of the reaction can have a positive or negative ΔG . Experimental evidence suggests that the free energy of phosphorylation can be measured⁴⁸⁻⁵².

Additionally, the ΔG of the effect a single phosphorylation has on substrate conformation has been limitedly measured. For example, let's consider scaffold protein Ste5, computational methods have approximated that the ΔG of each site is -8 kcal/mole⁵³. In phosphorylation, the effect of a single phosphate on conformation⁵⁴, protein-protein binding⁵⁵ or protein-membrane binding^{53,56} has been estimated to be about -2 kcal/mol; all favorable reactions.

ΔG is the difference between Gibbs energy at the start and end of a reaction; consider the reactions:



Here, $G = G_{end} - G_{start}$ of the reaction. In **Figure 3d**, we see the value of Gibbs energy as the reaction takes place, termed the reaction pathway. The values of a , b , and c are the Gibbs free energy at key points of interest. At the start of the reaction, there is an activation energy ($b-a$) required and is proportional to k_f . The difference between G at the end of the reaction and the activation energy ($c-b$) is proportional to k_r and is related to the spontaneity of the reaction or equilibrium. We can see that in this example G_{end} (at B) is lower than G_{start} and so $\Delta G < 0$, making the forward reaction favorable. Notice that changing the height of the activation energy barrier does not affect the equilibrium. It can be shown that the equilibrium $K_E = \frac{k_f}{k_r}$. In fact, if a reaction is spontaneous, then the ratio at equilibrium of products to reactants is greater than 1. Since K_E is equal to the ratio at equilibrium of products to reactants then it follows that for spontaneous reactions, $K_E > 1$. Similarly for non-spontaneous reactions. The relationship between ΔG and K_E is given by:

$$K_E = e^{-\frac{\Delta G}{RT}} \quad (\text{Eq. 2})$$

where R is the gas constant and T is temperature in Kelvin. Note that the relationship is exponential. For example, doubling ΔG can increase K_E 10-fold or more. This relationship tells us information about spontaneity given K_E , which is experimentally measurable.

When thinking of PTM's, individual sites can have a unique ΔG associated with them. For example, consider the theoretical substrate in **Figure 3e**. Here, the ΔG of the first modification event could differ from the ΔG of the second. Generally, each site being modified has a unique ΔG of phosphorylation. Even further, the ΔG of *activating* a protein after each modification is theoretically unique at each site. For example, in **Figure 3f**, a substrate can be modified to change conformations, and each site may contribute differently to this conformational change. Although there is significant work done by others to model PTMs and their regulation of conformational changes, not much is known about how the modifications of single sites contribute to substrate activation in multisite systems.

Section 3: Kinase-Substrate Specificity

3.1 The Docking Site: An Overview

Understanding how proteins recognize their substrates is an important aspect of protein modification. Unpacking kinase-substrate specificity and fidelity can also be useful for pharmacology and biochemistry. Without the ability to discriminate substrates, kinases would be able to modify any protein anywhere in the cell, which would certainly lead to negative outcomes. One possible explanation as to how proteins find their respective substrates is that they recognize the modification site and the neighboring residues. For example, MAP kinases are Ser/Thr-Pro driven proteins, meaning they phosphorylate a Ser or Thr residue with a Pro in the +1 position⁵⁷. However, if this were the only aspect of phosphorylation, MAP kinases would phosphorylate just about anything with a Ser/Thr-Pro sequence, which is not the case; suggesting that there must be other factors contributing to specific kinase-substrate phosphorylation⁵⁸. Indeed, extensive experimental data show that some MAP kinases bind tightly to conserved regions that are distinct and separate from the actual phosphosite, termed Docking Sites^{58,59}. Thus, docking sites and docking interactions could be a key to understanding kinase-substrate interactions and the regulation of signaling pathways.

Docking sites are short linear motifs that usually lie 20-100 residues upstream of the modification site and are important for kinase-substrate specificity⁶⁰. These motifs have been shown to be conserved across species and even homologues^{58,61-63}. One possible rationalization as to why docking sites would be conserved is that they allow for higher binding affinity between substrates and their kinases. This would allow docking sites to regulate the substrate activation efficiently⁶⁴. In 1996, Bardwell & Thorner showed

that a docking site found in yeast MEK is conserved in mammalian MEK's⁶⁴. The docking site for MEK-MAPK binding is both structurally and functionally conserved from yeast to mammals⁵⁸. In Jun homologues, the delta domain of Jun is conserved in JunB homologues⁵⁹.

Docking sites have also been found in proteins with other PTMs in addition to phosphorylation. For example, the C-terminus of Hsp70 Interacting Protein (CHIP) is shown to bind to an intrinsically disordered docking site of Interferon Regulatory Factor-1 (IRF-1) which regulates IRF-1 ubiquitination⁶⁵. Additionally, p53 contains multiple sites on which it is modified, and docking sites have been theorized to be key regulators for dynamic and specific binding of p53 and its partners⁴⁰. In dephosphorylation, MAPK phosphatases compete with transcription factor Elk-1 for binding to MAPK via a docking site; ultimately affecting dephosphorylation efficiency⁶⁶. How docking sites regulate PTMs remains an interesting aspect of signal transduction regulation.

3.2 The Docking Site as a Regulatory Mechanism of Modification

Experimental evidence shows that docking sites increase phosphorylation efficiency⁵⁹. There are Ser/Thr family protein kinases that rely on the docking site to mediate interactions as summarized in⁶⁷. It is theorized that the double selection that docking sites provide serve to prevent transcription factors from being activated by promiscuous kinases. However, it is unclear if less discriminating docking sites exist. If they did, they would bind to multiple kinases and allow for response from a variety of stimuli⁶⁰. In this section, we summarize some biological evidence where docking sites serve to regulate PTMs.

The inhibition of docking sites lowers the phosphorylation of multisite proteins in the MAPK cascade⁶⁸. Phosphorylation of c-Jun Serine sites by JNK depends on the presence of the docking site⁶⁹. Kallunki et al. showed that when the docking site is mutated, the overall phosphorylation of c-Jun is decreased relative to wild type (WT)⁶⁹. This would suggest that when the kinase JNK is left to independently select phosphosites, the overall phosphorylation decreases, relative to WT. It is also possible that the docking site helps maintain the kinase and substrate bound longer or with a stronger binding affinity⁶⁴. In fact, when the docking site is inhibited, the Serine sites closest to the docking site are phosphorylated less than when the docking site is left undisturbed⁶⁹. However, JNK it has been reported that JNK phosphorylates the Threonine sites of a docking-site-deficient-mutant more than WT c-Jun; these sites are further downstream to the docking site⁶⁹. Additionally, Bardwell et al. showed that when MKK4 docking peptides compete with c-Jun for JNK binding, c-Jun phosphorylation decreases when more peptide is present⁶⁸. In other words, competition assays also show that when JNK, c-Jun, and Docking Site peptides are together *in vitro*, c-Jun phosphorylation is decreased⁶⁸; this indicates that when the docking site is not available on

JNK, then JNK cannot properly phosphorylate c-Jun. These results suggest that the docking site is part of the regulatory mechanism of phosphorylation of JNK pathway targets.

Additionally, docking sites regulate phosphorylation of yeast High Osmolarity Glycerol (HOG) pathway proteins⁷⁰. Tatebayashi et al. showed that mutations to the docking site of Pbs2 drastically reduced its interaction with Ssk2 and Ssk22, which are downstream targets of the HOG pathway. Interestingly, Tatebayashi et al. showed that the fusion of Pbs2 docking site to other MAP kinases, namely Ste7, allowed for phosphorylation of Ste7⁷⁰, further showing that docking sites are crucial for efficacy and specificity of phosphorylation.

In addition to regulating phosphorylation, docking sites have also been shown to regulate other PTMs. For example, IRF-1 ubiquitination is dependent on CHIP binding to the docking site⁶⁵. IRF-1 ubiquitination is important for delivery to the proteasome and ultimately cell cycle arrest⁶⁵. In bacterial chemotaxis, CheB catalyzes both methylation and demethylation of aspartate receptor of *E. coli*⁷¹. Barnakov et al., showed in 1999 that in vitro CheB has a docking site for the methyltransferase/deamidase of the aspartate receptor⁷¹. In mammalian cells, the STAT transcription factors work to enhance cell growth and the cell cycle⁷². A study showed that Stat3 acetylation by p300 is docking site dependent⁷². Furthermore, Lys acetylation was also shown to be critical for stable Stat3 dimers which are required for DNA binding and transcriptional regulation⁷². These biological examples serve to show that docking motifs may play an important role in regulating protein modifications and ultimately signal transduction.

To summarize, possible hypotheses as to why a docking site is necessary are that (1) it helps the kinase recognize its target protein, (2) it increases the binding affinity of the kinase-substrate complex and (3) may even create an allosteric response for access to phosphosites. Notwithstanding the extensive studies of kinase-substrate specificity, there is still plenty to be understood in the near future with emerging biochemical and theoretical tools.

Section 4: Mathematical models for Multisite Protein Modifications

4.1 Multisite Modifications and Protein Activity

To provide background, here I introduce two key concepts regarding protein response. When asking how one component of a system affects signal transduction, we generally need a measurement of response. Hence, given a substrate and its activator (usually enzymes like kinases and phosphatases), we can measure the amount of active substrate based on factors of modification. The most common tool to measure protein activity is a dose-response curve (**Figure 4a**). These curves help determine a profile for

different aspects of PTMs on the substrate and provide insights on signal transduction regulation. These response functions can give important information regarding protein activity such as (1) *sensitivity* – how much activator is necessary to trigger a response from the substrate, (2) *threshold values* – the concentration of enzyme that maintains a low substrate activation, and (3) *ultrasensitivity* – the switch-like behavior of the response. Sensitivity and threshold are usually measured with Effective Concentration (EC) values that determine the concentration of enzyme that induce a response at a specified percent of the maximal response. For example, EC_{10} , in **Figure 4a**, represents the Effective Concentration of enzyme that activates the protein at 10% of the maximal active protein. Sensitivity and threshold differ in their calculations, sensitivity is usually the EC_{50} , while threshold is normally defined as the EC_{10} . Both of these values provide insights into the regulation of the substrate and may become useful in drug design⁷³.

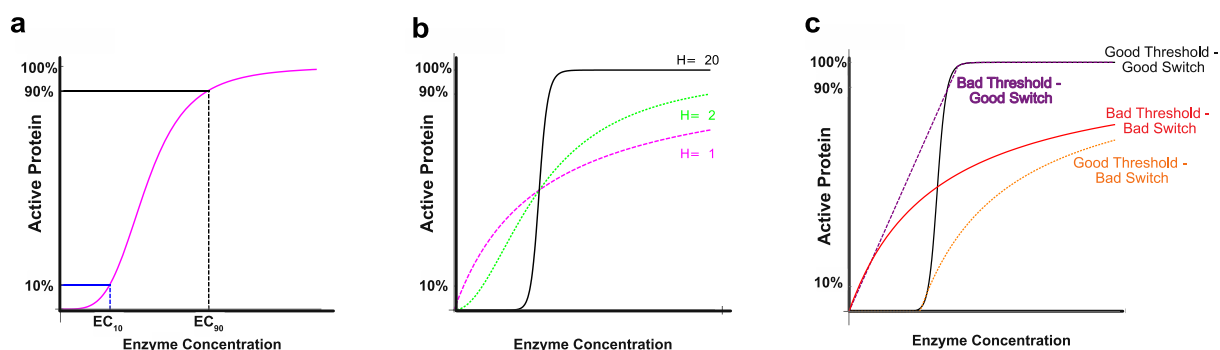


Figure 4: Dose Response Curves and Multisite Modifications. (a) General dose response curve. EC_{10} and EC_{90} are the enzyme concentrations at which there is a 10% and 90% response, respectively and labeled with the dashed lines. (b) Dose response curves with different Hill coefficients. (c) Dose response curves with different possible thresholds and switches: good threshold with a good switch (black), bad threshold with a good switch (purple), good threshold with a bad switch (orange), and bad threshold with a bad switch (red).

A common role for multisite modifications is to create switch-like, or *ultrasensitive*, dose response curves^{74–76}. We can think of ultrasensitivity as a degree of the “switch-ness” of the response, or how sigmoidal the curve is. A quantitative term used is the Hill number/coefficient H from the expression

$$Y = \frac{x^H}{x^H + k^H} \quad (\text{Eq. 3})$$

where x is the input or substrate concentration, Y is the response, and k is a parameter⁷⁴. For examples of ultrasensitivity in signal transduction, refer to⁷⁷. Ultrasensitivity can be measured in different ways⁷⁷. A common measurement of the Hill number H is given by:

$$H = \frac{\ln(81)}{\ln\left(\frac{EC_{90}}{EC_{10}}\right)} \quad (\text{Eq. 4})$$

where EC_{10} and EC_{90} are the Effective Concentration at which there is a 10% of maximal response and 90%, respectively, as seen in **Figure 4a**, based on the Goldbeter-Koshland observations⁷⁸. If there is no ultrasensitivity in the response, the effective Hill number is approximately 1 meaning that the fold change between EC_{10} and EC_{90} is exactly 81, thus for the response to go from 10% active to 90% requires an 81-fold change in the concentration of the activator, as depicted in **Figure 4b** with $H=1$. When the effective Hill number is greater than 1, the response is said to ultrasensitive and the fold change between EC_{10} and EC_{90} is less than 81. However, when the Hill number is less than 1, the response is said to show negative ultrasensitivity. In **Figure 4b**, we see what typical dose response curves with $H>1$ may look like. More specifically, we can see that the “switch-ness” of these curves becomes sharper with higher H and the initial response is lowered, i.e. it takes more input to generate a response. Note that the Goldbeter-Koshland expression is not the only way Hill numbers have been measured, in fact, studies have used max slope - the maximum slope of the curve⁷⁹, the slope at EC_{50} ⁸⁰, and the value of H from fitting data to Equation (1)^{81–85}, and other formulaic values of H ^{86–88}.

One can also think of the Hill number as a measurement of two aspects of a response: (1) the threshold and (2) the switch. The *threshold* is the concentration of activator that maintains a low substrate activation, or at least to some extent. A good threshold is believed to be beneficial in signal transduction, as it would mean that the substrate does not respond to “background” concentrations of its activator until that concentration is sufficiently high. The *switch* is generally the steepness of the response curve or a measurement of how little change in activator concentration yields max response from the substrate; also can be described as a measurement of the sigmodicity of the curve or how close to a step-like function the response is. In **Figure 4c**, examples of what good/bad thresholds and switches might look like. If a response curve has a good threshold and a good switch, we can visualize a curve similar to the black curve. This response can be obtained from the standard Hill function with a large value of H . If the response is too immediate with low activator concentrations and the response is not sharp, it is said to have a bad switch and a bad threshold (red curve). This response is seen in Michaelis-Menten kinetics. If a response has a good threshold but a bad switch (orange dashed curve), we can see that the initial response is minimal until a certain activator/enzyme concentration. After that point, the response is not immediate but graded resembling that of Michaelis-Menten kinetics (bad switch). Interestingly, if a response has a bad threshold and a good switch (purple dashed curve), then the initial response is too quick for low

activator concentrations but still a sharp response. There are not too many mechanisms displaying this behavior to the best of my knowledge thus far but would be interesting to explore.

We do know that multisite modifications are not necessarily going to yield a good switch, but can provide a mechanism for a good threshold⁸⁹. Here, we discuss mathematical models and analyses of multisite protein modification mechanics and their effect on signal transduction regulation. For a recent review of multisite phosphorylation as a mechanism of ultrasensitivity, refer to⁷⁷.

4.2 Mathematical Models of PTMs

Mathematical models are useful to study complex biological phenomena. Consider a protein with n modifications sites. This substrate alone can exist in 2^n states. This large number of combinations leads to a need for computational methods. In fact, mathematical models are excellent tools to examine the mechanisms of the individual working parts and/or their coupling. Both biophysical and non-biophysical models have been used to understand PTMs in the cell. In this section, an overview of the latter type of models in the context of PTMs are discussed. For the purposes of this section, non-biophysical models will be referred to as mathematical simple models. For the most recent review on mathematical simple modeling of multisite protein phosphorylation refer to¹⁹ (Salazar 2009). In recent years, there has been an increase in mathematical modeling to better understand multisite modification using concepts such as (1) rates of modification, (2) binding affinities between enzymes and substrates, (3) processivity of modification, (4) sequence of modification, and (5) binding to scaffolds or docking sites. In **Table 2** and the remainder of this section, we summarize a few mathematical simple models focusing on multisite PTMs. Through mathematical modeling, researchers have been able to gain insights into signal transduction regulation via multisite modifications like phosphorylation.

One key factor of multisite protein modification is the *rate of modification* of individual sites. Using mathematical models that focus on the rates of modifications has been done before. For example, in Reference ⁹⁰, authors investigate how the rates of phosphorylation of individual sites influence protein response. More specifically, it was shown that ultrasensitivity is robust to changes in modification rates. In fact, this paper suggests that in order to reach the maximum ultrasensitivity, sites should have similar rates of modification. For a more comprehensive review on rate of modification mathematical modeling, refer to¹⁹. **Table 2** includes some references that explored parameter space for the rate of modification or ratio of rates of different sites.

It is unlikely that the rates of modification are the only factor regulating protein activity. The *binding affinity between enzyme and substrate* can play a role in the modification of the substrate as well as the specificity. In recent years, researchers have focused on enzyme kinetics since the classic Michaelis-Menten description and there has been a collective effort to better understand how binding affinities regulate proteins using mathematical modeling. In fact, Mukhopadhyay et al. investigated differential binding affinities between ZAP-70 and 3-zeta chain immunoreceptor tyrosine based activation motifs (ITAMs). Mukhopadhyay et al. used a systems approach to show that, in immune response, T-cell antigen receptor (TCR) ultrasensitive response relies on multiple factors namely: having multiple ITAMs, sequential modification, and different binding affinities between ZAP-70 and the ITAMs⁹¹. Additionally, consider an exploration of phosphorylation-dephosphorylation cycles that allows for kinase-phosphatase complexes to form, as seen in phosphoinositides regulation⁸⁴. Szomolay et al. observed ultrasensitivity in these phosphorylation-dephosphorylation cycles that exists outside of a zero-order regime. Interestingly, this system also allowed for non-monotonic bell-shaped response for kinase-phosphatase complex that is robust to varying binding affinities⁸⁴. Unfortunately, the parameter space explored in this study did not span large orders of magnitude. This study shows that even simple deviations of basic biochemical process networks can give rise to “significant consequences on the response of such systems”⁸⁴.

An interesting aspect of PTMs to model is the *processivity* of the events. A huge topic of research lies in the processivity of PTMs. In fact, distributive and processive modifications have been considered extensively. For a review of some of these models and mechanisms, refer to¹⁹. Table 2 includes a brief description of both processive and distributive mechanisms along with references in which each method was explicitly modeled. For example, in⁸¹, a model comparing processive and distributive modifications determined that ERK2 is phosphorylated in a distributive method; also a great example of a mathematical model providing key insights on biological phenomenon. Some studies include mixed mechanisms between processive and distributive modification styles^{26,80,92}. These show that modification processivity plays a role in the regulation of key components of signal transduction.

Similar to processivity, the *order* of modifications can impact protein regulation. Since there are biological systems with sequential and non-sequential phosphorylation, mathematical analysis has also kept up with that fact. In p38 MAPK proteins, a mathematical model of multiple phosphorylations showed that MEK6 phosphorylation by TAO2 kinase processive and the Thr is phosphorylated prior to the Ser⁹³. In a more general multisite protein, Varedi et al. provide a rigorous mathematical analysis to examine how multisite phosphorylation yields ultrasensitivity protein degradation⁹⁴. Here, authors show that the order in which

the phosphorylation events takes place can give rise to a switch-like temporal response of a multisite protein. **Table 2** contains further mathematical models that consider the order of modifications in their analysis.

Since *docking sites and scaffolds* play a crucial role in PTMs, mathematical modeling can provide significant insights. Basu and Liu show that scaffold proteins help create bistability in a generic multisite phosphorylation system^{95,104}. They also show that ultrasensitivity may depend on the total scaffold protein concentration, binding rates, and number of sites. Additionally, mathematical modeling of docking sites have also provided new insights into PTMs in cell biology. A model that considers two-stage binding, docking site binding and target site binding, in phosphorylation-dephosphorylation is discussed in⁸⁰. This model shows that ultrasensitivity can arise when the concentration of the substrate is smaller than that of the enzymes. Authors argue that two-stage binding allows the enzymes to sequester the substrate. This is significant for explaining decision-making in signal transduction when concentrations of substrates are low⁸⁰. Other models that consider scaffolds and docking sites are included in **Table 2**.

A major hole in this type of modeling is that it is unclear how cooperativity and allostery play a role in PTMs and signal transduction regulation. This is why we need allosteric-type models, where we can visualize interactions between sites and the overall effect of PTMs on substrate activity. Allosteric-type models will be discussed in the following section. It is also important to note that there is very little work to understand how single sites contribute to protein activity, a large gap in our understanding of multisite proteins. Lots of room here for further investigation. Although kinetic models as the ones discussed thus far are useful in understanding PTMs, biophysical models have also generated key insights into the mechanisms of signal transduction regulation. For a review on biophysical models, refer to^{96,97}.

TABLE 2
Overview of Mechanisms of Modification Factors

Mechanism of Modification	Description	Examples of Models
Rates of Modification	The rate at which a single site is modified on a substrate	[19, 79, 80, 81, 84, 90, 92, 98-103]
Binding Affinities	The binding strength between substrate and enzyme (binding rate)	[84,90, 91, 99]
Processivity – Distributive	One modification per binding collision between substrate and enzyme	26, 80, 81, 86, 92-94, 100-102, 104-107]
Processivity – Processive	More than one modification per binding collision between enzyme and substrate	[26, 80, 81, 92, 107, 108]
Sequence of Modification	The order in which sites are modified	[82, 85, 86, 91, 93, 94, 100, 101]
Binding to scaffolds or docking sites	Presence and influence of docking site or scaffolds on substrate	[80, 95, 101,10]

4.3. Cooperative and Allosteric Models

Since many proteins exhibit cooperativity and allostery, it is important to also use mathematical models to unpack this behavior. A definitive example is Hemoglobin, with four subunit on which an oxygen molecule can bind. Hemoglobin dose response is switch-like in nature with an approximated Hill number of 2.7. Mathematical models have evolved to try and understand how a four-site protein can exhibit this type of ultrasensitivity. There are prominent models that were specifically designed to incorporate cooperativity and/or allostery such as the Hill model⁷⁴, Adair model¹⁰⁹, Koshland-Nemethy-Filmer (KNF)¹¹⁰, and Monod-Wyman-Changeux (MWC)¹¹¹ model. This section includes a brief description and comparison of these models in the context of PTMs.

The Hill model⁷⁴, has been alluded to in the previous text. Here, we discuss it in the context of multisite cooperativity. The Hill model is one that considers cooperativity between sites; the modification of one site enhances/inhibits the modification of the following modifications. **Figure 5a** includes a substrate that is made up of two subunits with one modification site each. The Hill model assumes that all modifications occurs simultaneous (**Figure 5a**). The Hill Equation for n sites is derived as:

$$y = Y_{MAX} \frac{x^n}{x^n + EC_{50}^n},$$

with similar parameter definitions as in (**Eq. 3**). Unfortunately, it is not physically realistic to assume to all modifications occur simultaneously, as the Hill model does not fit observed data like that for 4-site Hemoglobin which has a Hill number of ~ 2.7 while the Hill model predicts a Hill number of 4.

The Adair model¹⁰⁹ has been extensively studied, expanded and taught throughout the mathematical biology community. The Adair model expands on the Hill model by assuming that modifications do not necessarily take place simultaneously and interactions between sites can occur, making it a more realistic model. In **Figure 5b**, we see how modifications on each subunit can take place one at a time. In the Adair model, modification (or binding) to one site can affect the modification (or binding) of the next. This model still assumes that modification/binding is still identical, meaning that the modification of one site is symmetrical to modification of another. This allows for multiple modification-forms of the substrate. For example, with the two-site substrate in **Figure 5b**, the substrate can exist with no modifications, one, or two. The Adair model then derives the following relation for y , the fractional occupancy of modification sites:

$$y = \frac{K_1x + qK_1x^2}{1 + 2K_1x + qK_1^2x^2}$$

The Adair model has showed that by allowing the modification of the first site to influence the modification of a second, an ultrasensitive dose-response curve can arise. Unfortunately, the Adair model cannot describe how modification (or binding) sites are communicating with each other. Fortunately, other models can shed light on the allosteric ultrasensitive behavior seen in Hemoglobin.

In the KNF model, some of the assumptions in the Hill and Adair models were relaxed. More specifically, the KNF model focused on proteins with more than one subunit by adding a structural component¹¹². In the KNF model, authors assume that each subunit can exist in one of two states: active or inactive¹¹⁰. In addition to subunit states, the KNF model also introduced the concept of “induced fit.” Here, the modification of a site (or ligand binding) on an inactive subunit immediately forces that subunit to change its conformation to an active state. Note that the KNF model can also be used in PTMs in addition to ligand binding. For simplicity, **Figure 5c**, shows a substrate with two subunits and one modification site on each where the circles represent the inactive state and the rectangles the active state. In **Figure 5c**, the modification of the site on the first subunit induces the conformational change from one conformation to another. This possibility of having mixed conformations was a unique perspective on cooperativity between sites. In following with the KNF model, the sequential modification then allows the second subunit to be modified and change its conformation. This model is closer to describing hemoglobin ultrasensitivity than the Adair and Hill models.

The MWC model was also introduced to further describe proteins with multiple subunits. Here, Monod et al. introduced the concept of “concerted binding” where the modification of a subunit forces the entire

protein to shift in its conformation from a “relaxed” to a “tense” state. The MWC model also considers thermodynamic elements to ligand binding. Shibayama showed that hemoglobin follows a the MWC model closer than the KNF model¹¹³. The MWC model has been extended to PTMs in addition to ligand binding. In the context of PTMs, **Figure 5d** shows a schematic of the MWC model where a substrate can be modified while in a relaxed or tense state. The modification of any one site will change the conformation of the entire protein, not just that subunit as in the KNF model. Some models, however, combine the KNF and MWC models such as the Hilser et al. model, where the protein is made up of two key subunits where each subunit can be in a relaxed or tense state (KNF) with MWC type modifications³⁵. The MWC model has been extensively used in understanding areas of biology such as ligand-gated ion channels, bacterial chemotaxis, and genomic accessibility⁸⁷. Here, I cannot review all MWC-type models because there is an extensive amount of work expanding on the work of Monod et al.

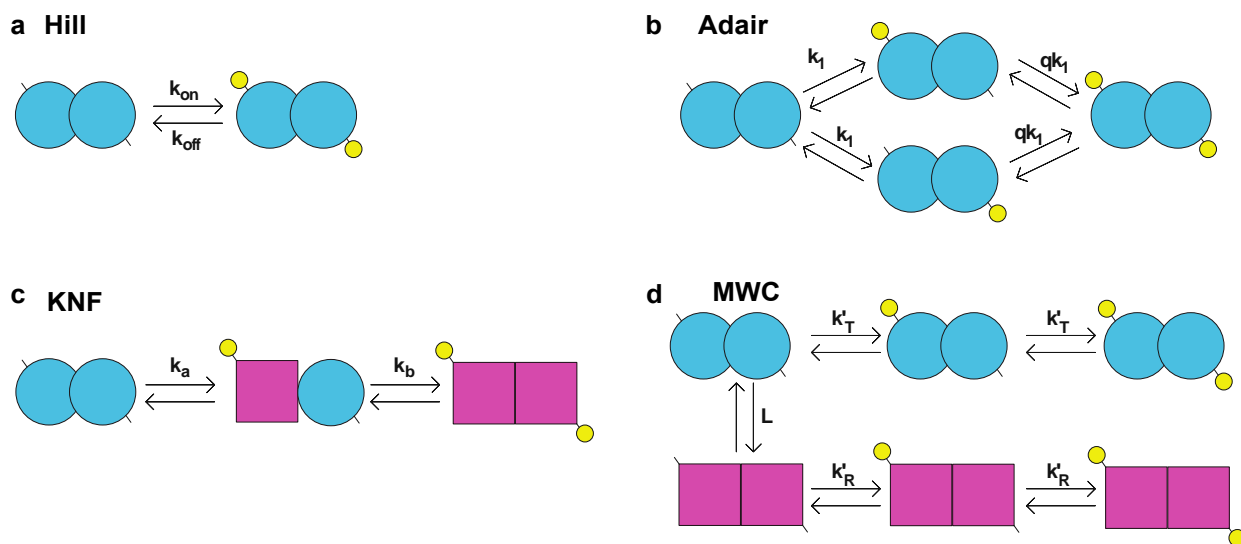


Figure 5: Cooperative and Allosteric Models. Schemes representing models of PTMs each protein here is depicted as one with identical subunits. (a) Hill scheme where modifications occur simultaneously. (b) Adair scheme where modifications can occur in any order and the modification of a second site can occur at a different rate depending on the modification of the first. In (c-d) circles signify an inactive/tense conformations and squares are active/relaxed conformations. (c) Koshland-Nemethy-Filmer (KNF) scheme where there is a conformational change to the subunit modified and follows a sequential modification. (d) Original Monod-Wyman-Changeux (MWC) scheme where modification of any subunit results in conformational change of entire protein.

To summarize, cooperativity and allostery observed in biology can be described using mathematical tools. The leading models include: the Hill, Adair, KNF, and MWC models. All of which have given tremendous insights on protein regulation through PTMs and ligand binding. Thanks to these models we understand

protein regulation more. There is still a lot left to learn about how multisite PTMs regulate proteins, but this is great place to be at right now.

Section 5: JNK pathway as a Biological Model System

5.1. Overview of JNK pathway and Transcription Factor c-Jun

The Mitogen-Activated-Protein Kinases (MAPKs) are critical proteins in various MAPK cascades; here we focus on the c-Jun N-terminal Kinases (JNK) pathway. This pathway regulates many different eukaryotic cell responses including embryonic development and cell death⁴⁷. Similar to other MAPK pathways, the JNK pathway has proved to be a complex system with multiple components.

The JNK pathway is activated by numerous triggers such as developmental cues, inflammation, pathogens, and hormones as seen in **Figure 6a** (adapted from Review⁴⁷). These triggers activate MAPKKs, which then phosphorylate MKK4 and/or MKK7. These MAPKKs then activate JNK by phosphorylation of Thr183 and Tyr185. JNKs (1,2, or 3) will then phosphorylate the target transcription factor. JNKs target a variety of transcription factors such as Elk4, ATF2, Smad2, and c-Jun. Here, we focus on transcription factor c-Jun which is then activated. The c-Jun response based on phosphorylation is associated with cell viability. A simplified version of the JNK pathway is included in **Figure 6a** (adapted from Review⁴⁷). This section provides a short summary of our current understanding of the JNK pathway and multisite phosphorylation of transcription factor c-Jun. More specifically, this section will focus on the functional consequences of phosphorylation of c-Jun sites.

To understand the functional consequences of c-Jun phosphorylation, we must first take a closer look at the main components of c-Jun phosphorylation. There are at least six sites that have been shown to be phosphorylated by JNK⁴⁷ or other kinases. The phosphosites are in residues Ser63, Ser73, Thr91, Thr93, Thr239, and Ser243^{114–116}, all with the S/T-P motif necessary for phosphorylation by proline-directed kinases. **Figure 6b** shows a linear representation of c-Jun and these phosphorylation sites.

Similar to other MAPK pathway proteins, phosphorylation of c-Jun is also docking-site dependent¹¹⁷. The docking site on c-Jun lies between residues 32 and 42 with a sequence⁶⁸ of ³²KILKQSM⁴²LNL, as seen in **Figure 6b**. In ¹¹⁷, inhibition of the docking site showed significantly reduction on c-Jun phosphorylation. In fact, when there exist competing peptides that block access to the docking site, c-Jun phosphorylation efficiency drops as well⁶⁸. Thus, a functional docking sites and the S/T-P motif are both necessary for efficient c-Jun phosphorylation.

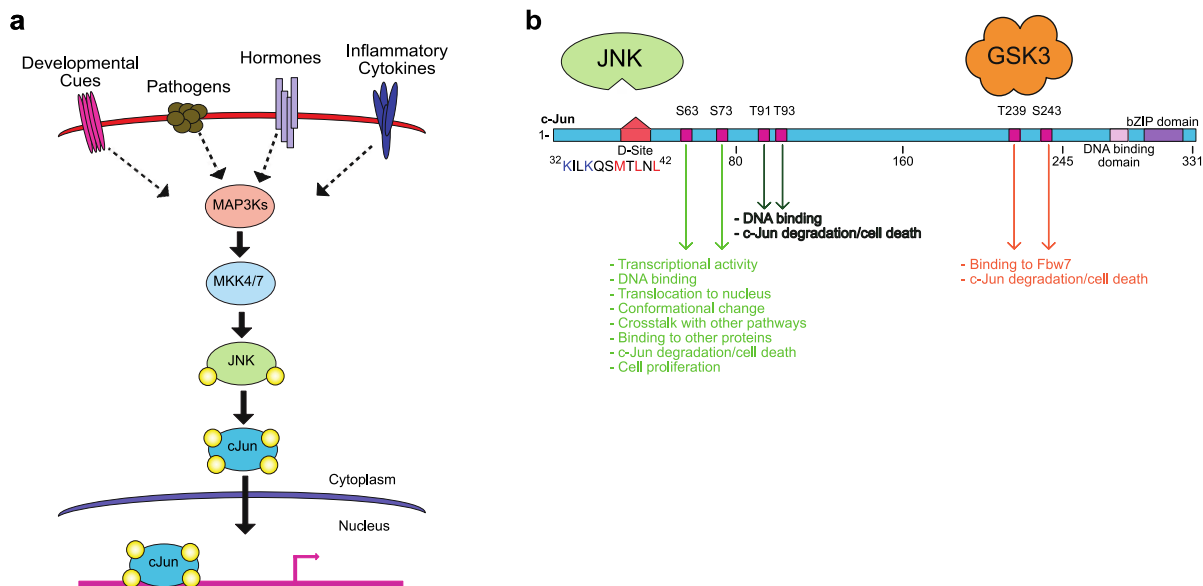


Figure 6: JNK Signaling Pathway and Transcription Factor c-Jun. (a) General schematic of JNK signaling pathway (adapted from Review⁴⁷). Signals that activate JNK pathway include developmental cues, pathogens, hormones and inflammatory cytokines. These signals activate MAP3Ks, which then activate MKK4/7. These kinases then phosphorylate JNK1, JNK2, and/or JNK2 on two regulatory residues. JNKs will then phosphorylate the appropriate target, for simplification, the target shown here is transcription factor c-Jun. c-Jun activity will then take place depending on its phosphorylation form. (b) Linear schematic of c-Jun with two enzymes shown to phosphorylate c-Jun at the residues marked by the magenta squares. The pentagon on c-Jun represents the docking site (D-Site). The rectangles on the C-terminal end of c-Jun represent the DNA binding domain and the basic Leucine Zipper domain. At each pair of sites, the arrows indicate known c-Jun activity upon phosphorylation. More detail regarding each bullet point in text (adapted from Review⁴⁷).

Although solving a crystal structure can prove to be difficult, Jun proteins, including c-Jun, are intrinsically disordered proteins⁴⁷, making it relatively difficult to solve the complete c-Jun crystal structure. However, the DNA binding region of c-Jun has been solved, including the quaternary structures formed for DNA binding¹¹⁸. Fortunately, numerous computational methods are available that can give some information regarding protein structure (RosettaCommons, IUPred2, etc).

5.2. Conservation of c-Jun phosphorylation sites

Understanding the conservation of c-Jun orthologs can provide new insights into the evolution of c-Jun and c-Jun phosphorylation. With emerging technologies, the conservation of phosphosites and their evolution across species is attainable¹³⁵. In this section, I provide a short overview of the conservation of the docking site, the four N-terminal phosphosites, and the 2 C-terminal phosphosites in different species. Sequences were found through UniProt Genetree entries for five different vertebrate classes: *Danio rerio* (zebrafish), *Xenopus tropicalis* (tropical frog), *Chelydra serpentina* (snapping turtle), *Gallus gallus*

(chicken), *Mus musculus* (mouse), and *Homo sapien* (human). I compared the location and sequence alignments for all these sequences.

Table 3 shows a simplified version of the alignments, where each row shows the species name. The center columns show c-Jun protein sequence, and the last column contains the UniProt ID for the complete protein sequence with the exception for *Chelydra serpentina* with transcript ID [ENSCSRT00000030615.1](#). For simplicity, I compared the sequences for the docking site (D-site), the four N-terminal sites, followed by the two C-terminal sites and their surrounding sequences. The protein sequences all start with the start codon Methionine (M) and the hyphen signifies sequences not included in this comparison. The red indicates sequence that is identical to the human c-Jun sequence.

As seen in the protein sequence alignment for c-Jun orthologs, the D-site is completely conserved in human and mouse c-Jun. There is clear conservation of the last five residues of the D-site between all five species. Interestingly the Lysine (K) at the start of the D-site is also conserved. Bardwell et al. 2001 showed that positively-charged Lysine or Arginine residues are conserved in MEK D-sites as well⁵⁸. Hence, even in MAP targets, namely c-Jun, the D-site appears to be highly conserved across orthologs. Not surprisingly, the S/T-P motif is conserved in all five species in **Table 3**, however, the location of the sites is slightly different. For example, Ser63 in humans shows up as Ser61 in zebrafish. Looking at all five classes, the distance from the end of the D-site to the first serine is between 20 and 22 residues; chicken c-Jun has the largest distance between the D-site and the first serine, human c-Jun has 21 residues between the D-site and Ser63. Spacing between the D-site and the other residues is also highly conserved.

Additionally, the sequence surrounding that first Ser phosphosite is slightly different in zebrafish and chickens than in the other classes. It is unclear if this small shift yields different phosphorylation properties or influences c-Jun structure. Perhaps new technologies can help compare the downstream effects of each c-Jun ortholog in more detail. Interestingly, the three N-terminal sites and the two C-terminal sites are perfectly conserved across all these species. It is possible this conservation serves to show that the regulation of c-Jun is robust and relies on multisite phosphorylation. However, that cannot be answered without further exploring the downstream effects of each site's phosphorylation.

TABLE 3
c-Jun Protein Sequence Conservation

Species	Sequence					UniProt
	D-site	Ser	Ser	Thr-Thr	Ser-Thr	
<i>Danio rerio</i>	MST- ²⁸ YNHKALKHMTLNLSDP	⁵⁸ ILTSPDVG	⁶⁹ KLASPELE	⁸⁷ ITTTPTPTQF	²¹³ PGETPPLSPIDM	Q6NZT5
<i>Xenopus tropicalis</i>	MTA- ²⁷ GYNKVLKQSMTLNLSDP	⁵⁷ LLTSPDVG	⁶⁸ KLASPELE	⁸⁶ ITTTPTPTQF	²¹⁷ PGETPPLSPIDM	F6RYL1
<i>Chelydra serpentina</i>	MTA- ²⁹ NNPKVLKQNMTLNLADP	⁵⁹ ILTSPDVG	⁷⁰ KLASPELE	⁸⁸ ITTTPTPTQF	²¹⁹ PGETPPLSPIDM	–
<i>Gallus gallus</i>	MEP- ²⁴ NNAKVLKQSMTLNLSDP	⁵⁶ ILTSPDVG	⁶⁶ KLASPELE	⁸⁴ ITTTPTPTQF	²¹⁵ PGETPPLSPIDM	F1NCN0
<i>Mus musculus</i>	MTA- ²⁹ SNPKILKQSMTLNLADP	⁵⁹ LLTSPDVG	⁷⁰ KLASPELE	⁸⁸ ITTTPTPTQF	²³⁸ PGETPPLSPIDM	P05627
<i>Homo sapien</i>	MTA- ²⁹ SNPKILKQSMTLNLADP	⁶⁰ LLTSPDVG	⁷⁰ KLASPELE	⁸⁸ ITTTPTPTQF	²³⁶ PGETPPLSPIDM	P05412

5.3. Functional consequences of c-Jun phosphorylation

Given that c-Jun is phosphorylation on multiple sites and the docking site mediates these phosphorylation, what are the functional consequences of these phosphorylations? Although the downstream effects of Ser/Thr phosphorylation on c-Jun are complicated, there are studies that provide new information. As reviewed in⁴⁷, generally phosphorylation of Ser63, Ser73, Thr91 and Thr93 is associated with transactivation by recruiting co-activators of c-Jun or histone acetyltransferases. Phosphorylation of Thr293 and Ser243 is implicated in controlling degradation of Jun proteins⁴⁷. Additional studies have been conducted to show the effect phosphorylation of these sites has on c-Jun activity and a general image summary is in **Figure 6b**. The phosphorylation of the four N-terminal phosphorylation sites have been linked to disease and cancers such as lung cancer¹¹⁹, breast cancer¹²⁰, ovarian cancer¹²¹, and Parkinson's disease¹²². For an additional review, see⁴⁷. N-terminal phosphorylation, that of sites Ser63, S73, Thr91, and Thr93, proves to regulate c-Jun DNA binding and cell death. C-terminal phosphorylation, Thr239 and Ser243, appears to have less functional variety as it is associated mostly with c-Jun degradation via protein binding.

Phosphorylation of Ser63 and Ser73 appear to be the most studied. They have been mostly implicated in transcriptional activity; events such as translocation to the cell nucleus^{123,124}, hetero-/homo-dimerization¹²⁵, and DNA binding^{124,126}. It is possible that Ser63 and Ser73 play bigger roles in c-Jun activity, since so many studies focus on these sites. However, it is also possible that functional roles of these sites are more readily available since so many biological methods require antibodies specific for each of the sites, and that is a challenge. Aside from that, it is true that phosphorylation of Ser63 and Ser73 have been connected to crosstalk with other pathways^{120,127}, binding to other proteins, c-Jun degradation^{128,129}, and UV response¹³⁰.

Xie et al. show that a c-Jun knockdown reduces Notch1 gene expression and cell proliferation in HCC70 and MDA-MB-231 cells¹²⁰. In addition, Xie et al. also show that JNK regulates Triple Negative Breast Cancer (TNBC) tumorigenesis by promoting a cancer stem-like cell phenotype through Notch1 signaling via activation of c-Jun when phosphorylated at Ser63 and Ser73. This study indicates that JNK/c-Jun/Notch1

signaling is a potential therapeutic target for TNBC. Phosphorylation of Ser63 and Ser73 also regulate cell survival in neuronal cells¹²². Park et al. show that when the concentrations of L-DOPA, a natural precursor of dopamine, are low, then ERKs will phosphorylate Ser73 and leads to cell survival. However, if L-DOPA concentrations are high, then active JNK or ERK will phosphorylate Ser63 on c-Jun and lead to cell apoptosis. JNK1 and JNK2 phosphorylating c-Jun Ser63 and Ser73 in HeLa cells also show that c-Jun can bind to CBP which is a CREB binding protein that helps in recruitment of transcription machinery¹²⁷. Other proteins that bind to c-Jun that is phosphorylated on residues Ser63 and Ser73 include TCF4 in tumorigenesis¹³¹, ATF-2 in neuronal cell apoptosis¹³², and TPA in DNA binding¹³³ to name a few.

On the C-terminal end of c-Jun, N-terminal to the DNA binding domain, there are two additional phosphorylation sites: Thr239 and Ser243⁴⁷. These sites can be phosphorylated by GSK3¹¹⁶ and ERK2^{126,134} to regulate c-Jun activity. Interestingly, Papavassiliou et al. show that once c-Jun is phosphorylated on the N-terminal sites, c-Jun binds to phosphatase TPA and the C-terminal sites are dephosphorylated. This dephosphorylation leads to efficient c-Jun-DNA binding¹³³. Wei et al. show c-Jun phosphorylation of sites T239 and S243 by GSK3 leads to c-Jun-Fbw7 binding and c-Jun degradation¹¹⁶. Since c-Jun shares sequence and function with c-Myc, authors set out to test c-Jun binding to Fbw7; a protein that regulates substrate degradation. Wei et al. show, *in vivo* and *in vitro*, that T239 and S243 are both necessary for c-Jun binding to Fbw7. This binding is independent of JNK phosphorylation of the N-terminal phosphorylation sites. Once Thr239 and Ser243 are phosphorylated, c-Jun is degraded. Wei et al. show that both phosphorylations are necessary for c-Jun degradation¹¹⁶. In fact, mutation of either site to an Ala allowed c-Jun to prevent its degradation. Furthermore, phosphorylation of the two C-terminal sites by ERK2 leads to c-Jun-Cdc4 binding but phosphorylation of the four N-terminal sites is necessary for binding¹³⁴. Once c-Jun is bound to Fbw7 c-Jun is degraded by the proteasome^{116,134}.

To summarize, **Figure 6b** shows that the phosphorylation of c-Jun can provide different responses. The phosphor-form may play a larger role in determining the actual c-Jun response than previously thought. For example, c-Jun has six sites but the phosphorylation of any two do not yield the same response as the phosphorylation of any other pair. Although it does seem that sites are paired off in terms of the downstream response they provide; phosphorylation of Ser63 is almost always connected to phosphorylation of Ser73, similarly with Thr239 and Ser243. This could be due to their physical distance or accessibility. One thing they all share, however, is that all sites seem to play a role in cell viability either by promoting cell proliferation or apoptosis.

Taken together, it is a combination of factors (binding, PTMs, free energy, etc.) that work in unison for proper cell function. Understanding the roles of these functions on protein activity is key to shed light on development and diseases.

References:

1. Bardwell, A. J., Lagunes, L., Zebarjedi, R. & Bardwell, L. The WW domain of the scaffolding protein IQGAP1 is neither necessary nor sufficient for binding to the MAPKs ERK1 and ERK2. *J. Biol. Chem.* **292**, 8750–8761 (2017).
2. Burack, W. R. & Shaw, A. S. Signal transduction: Hanging on a scaffold. *Curr. Opin. Cell Biol.* **12**, 211–216 (2000).
3. Engström, W., Ward, A. & Moorwood, K. The role of scaffold proteins in JNK signalling. *Cell Prolif.* **43**, 56–66 (2010).
4. Cortese, M. S., Uversky, V. N. & Keith Dunker, A. Intrinsic disorder in scaffold proteins: Getting more from less. *Prog. Biophys. Mol. Biol.* **98**, 85–106 (2008).
5. Choi, K. Y., Satterberg, B., Lyons, D. M. & Elion, E. A. Ste5 tethers multiple protein kinases in the MAP kinase cascade required for mating in *S. cerevisiae*. *Cell* **78**, 499–512 (1994).
6. Hedman, A. C., Smith, J. M. & Sacks, D. B. The biology of IQGAP proteins: beyond the cytoskeleton. *EMBO Rep.* **16**, 427–446 (2015).
7. Dhanasekaran, D. N., Kashef, K., Lee, C. M., Xu, H. & Reddy, E. P. Scaffold proteins of MAP-kinase modules. *Oncogene* **26**, 3185–3202 (2007).
8. Roy, M., Li, Z. & Sacks, D. B. IQGAP1 Binds ERK2 and Modulates Its Activity. *J. Biol. Chem.* **279**, 17329–17337 (2004).
9. Roy, M., Li, Z. & Sacks, D. B. IQGAP1 Is a Scaffold for Mitogen-Activated Protein Kinase Signaling. *Mol. Cell. Biol.* **25**, 7940–7952 (2005).
10. Ren, J. G., Li, Z. & Sacks, D. B. IQGAP1 modulates activation of B-Raf. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 10465–10469 (2007).
11. Briggs, M. W. & Sacks, D. B. IQGAP proteins are integral components of cytoskeletal regulation. *EMBO Rep.* **4**, 571–574 (2003).
12. Jameson, K. L. *et al.* IQGAP1 scaffold-kinase interaction blockade selectively targets RAS-MAP kinase-driven tumors. *Nat. Med.* **19**, 626–630 (2013).
13. Morse, A. M., Whetten, R. W., Dubos, C. & Campbell, M. M. Post-translational modification of an R2R3-MYB transcription factor by a MAP Kinase during xylem development. *New Phytol.* **183**, 1001–1013 (2009).
14. Ribet, D. & Cossart, P. Post-translational modifications in host cells during bacterial infection. *FEBS Lett.* **584**, 2748–2758 (2010).
15. Nash, P. *et al.* Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature* **414**, 514–521 (2001).
16. Doyle, H. A. & Mamula, M. J. Post-translational protein modifications in antigen recognition and autoimmunity. *Trends Immunol.* **22**, 443–449 (2001).
17. Harding, J. J. *et al.* Non-enzymic Post-translational modification of proteins in aging. A review. *Mech. Ageing Dev.* **50**, 7–16 (1989).
18. Waby, J., Bingle, C. & Corfe, B. Post-Translational Control of Sp-Family Transcription Factors. *Curr. Genomics* **9**, 301–311 (2008).
19. Salazar, C. & Höfer, T. Multisite protein phosphorylation - From molecular mechanisms to kinetic models. *FEBS J.* **276**, 3177–3198 (2009).

20. Dai, C. & Gu, W. P53 post-translational modification: Deregulated in tumorigenesis. *Trends Mol. Med.* **16**, 528–536 (2010).
21. Prabakaran, S., Lippens, G., Steen, H. & Gunawardena, J. Post-translational modification: Nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **4**, 565–583 (2012).
22. Bevilaqua, L. R. M., Graham, M. E., Dunkley, P. R., Von Nagy-Felsobuki, E. I. & Dickson, P. W. Phosphorylation of Ser19 Alters the Conformation of Tyrosine Hydroxylase to Increase the Rate of Phosphorylation of Ser40. *J. Biol. Chem.* **276**, 40411–40416 (2001).
23. Ho, D. T., Bardwell, A. J., Abdollahi, M. & Bardwell, L. A docking site in MKK4 mediates high affinity binding to JNK MAPKs and competes with similar docking sites in JNK substrates. *J. Biol. Chem.* **278**, 32662–32672 (2003).
24. Ho, D. T., Bardwell, A. J., Grewal, S., Iverson, C. & Bardwell, L. Interacting JNK-docking sites in MKK7 promote binding and activation of JNK mitogen-activated protein kinases. *J. Biol. Chem.* **281**, 13169–13179 (2006).
25. Holmberg, C. I., Tran, S. E. F., Eriksson, J. E. & Sistonen, L. Multisite phosphorylation provides sophisticated regulation of transcription factors. *Trends Biochem. Sci.* **27**, 619–627 (2002).
26. Suwanmajo, T. & Krishnan, J. Mixed mechanisms of multi-site phosphorylation. *J. R. Soc. Interface* **12**, (2015).
27. Scott, M. P. & Miller, W. T. A peptide model system for processive phosphorylation by Src family kinases. *Biochemistry* **39**, 14531–14537 (2000).
28. Aubol, B. E. *et al.* Processive phosphorylation of alternative splicing factor/splicing factor 2. *PNAS* **100**, 12601–12606 (2003).
29. Waas, W. F., Lo, H. H. & Dalby, K. N. The kinetic mechanism of the dual phosphorylation of the ATF2 transcription factor by p38 mitogen-activated protein (MAP) kinase α . Implications for signal/response profiles of MAP kinase pathways. *J. Biol. Chem.* **276**, 5676–5684 (2001).
30. Graybill, C. & Prehoda, K. E. Ordered multisite phosphorylation of lethal giant larvae by atypical protein kinase C. *Biochemistry* **53**, 4931–4937 (2014).
31. Kennedy, M. J. *et al.* Multiple phosphorylation of rhodopsin and the in vivo chemistry underlying rod photoreceptor dark adaptation. *Neuron* **31**, 87–101 (2001).
32. Weinman, E. J. *et al.* Cooperativity between the phosphorylation of Thr95 and Ser 77 of NHERF-1 in the hormonal regulation of renal phosphate transport. *J. Biol. Chem.* **285**, 25134–25138 (2010).
33. Schweiger, R. & Linial, M. Cooperativity within proximal phosphorylation sites is revealed from large-scale proteomics data. *Biol. Direct* **5**, 1–17 (2010).
34. Chao, C., Herr, D., Chun, J. & Xu, Y. Ser18 and 23 phosphorylation is required for p53-dependent apoptosis and tumor suppression. *EMBO J.* **25**, 2615–2622 (2006).
35. Hilser, V. J., Wrabl, J. O. & Motlagh, H. N. Structural and Energetic Basis of Allostery. *Annu. Rev. Biophys.* **41**, 585–609 (2012).
36. Shumilin, I. A., Zhao, C., Bauerle, R. & Kretsinger, R. H. Allosteric inhibition of 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase alters the coordination of both substrates. *J. Mol. Biol.* **320**, 1147–1156 (2002).
37. Huang, Z. *et al.* ASD v2.0: Updated content and novel features focusing on allosteric regulation. *Nucleic*

- Acids Res.* **42**, 510–516 (2014).
38. Huang, Z. *et al.* ASD: A comprehensive database of allosteric proteins and modulators. *Nucleic Acids Res.* **39**, 663–669 (2011).
 39. Shen, Q. *et al.* ASD v3.0: Unraveling Allosteric regulation with structural mechanisms and biological networks. *Nucleic Acids Res.* **44**, D527–D535 (2016).
 40. Gu, B. & Zhu, W. G. Surf the Post-translational Modification Network of p53 Regulation. *Int. J. Biol. Sci.* **8**, 672–684 (2012).
 41. Nie, H. *et al.* Phosphorylation of FOXP3 controls regulatory T cell function and is inhibited by TNF- α in rheumatoid arthritis. *Nat. Med.* **19**, 322–328 (2013).
 42. Inuzuka, H. *et al.* Acetylation-dependent regulation of Skp2 function. *Cell* **150**, 179–193 (2012).
 43. Xu, Y. Regulation of p53 responses by post-translational modifications. *Cell Death Differ.* **10**, 400–403 (2003).
 44. Atwood, S. X., Chang, A. L. S. & Oro, A. E. Hedgehog pathway inhibition and the race against tumor evolution. *J. Cell Biol.* **199**, 193–197 (2012).
 45. Atwood, S. X., Li, M., Lee, A., Tang, J. Y. & Oro, A. E. GLI activation by atypical protein kinase C λ regulates the growth of basal cell carcinomas. *Nature* **494**, 484–488 (2013).
 46. Wang, Y. *et al.* The Crosstalk of mTOR/S6K1 and Hedgehog Pathways. *Cancer Cell* **21**, 374–387 (2012).
 47. Zeke, A., Misheva, M., Reményi, A. & Bogoyevitch, M. A. JNK Signaling: Regulation and Functions Based on Complex Protein-Protein Partnerships. *Microbiol. Mol. Biol. Rev.* **80**, 793–835 (2016).
 48. Slater, E. C., Rosing, J. & Mol, A. The phosphorylation of potential generated by respiring mitochondria. *Biochim. Biophys. Acta* **292**, 534–553 (1973).
 49. Adams, J. A. & Taylor, S. S. Divalent metal ions influence catalysis and active-site accessibility in the camp-dependent protein kinase. *Protein Sci.* **2**, 2177–2186 (1993).
 50. Mader, A. Glycolysis and oxidative phosphorylation as a function of cytosolic phosphorylation state and power output of the muscle cell. *Eur. J. Appl. Physiol.* **88**, 317–338 (2003).
 51. Cheng, Y., Zhang, Y. & McCammon, J. A. How does the cAMP-dependent protein kinase catalyze the phosphorylation reaction: An ab Initio QM/MM study. *J. Am. Chem. Soc.* **127**, 1553–1562 (2005).
 52. Beard, D. A. A biophysical model of the mitochondrial respiratory system and oxidative phosphorylation. *PLoS Comput. Biol.* **1**, 0252–0264 (2005).
 53. Serber, Z. & Ferrell, J. E. Tuning Bulk Electrostatics to Regulate Protein Function. *Cell* **128**, 441–444 (2007).
 54. Shen, T., Wong, C. F. & McCammon, J. A. Atomistic Brownian dynamics simulation of peptide phosphorylation. *J. Am. Chem. Soc.* **123**, 9107–9111 (2001).
 55. Nishi, H., Hashimoto, K. & Panchenko, A. R. Phosphorylation in protein-protein binding: effect on stability and function. *Structure* **19**, 1807–1815 (2011).
 56. Murray, D. *et al.* Electrostatics and the membrane association of Src: Theory and experiment. *Biochemistry* **37**, 2145–2159 (1998).
 57. Pearson, G. *et al.* Mitogen-activated protein kinase pathways. *Curr. Opin. Cell Biol.* **9**, 180–186 (1997).
 58. Bardwell, A. J., Flatauer, L. J., Matsukuma, K., Thorner, J. & Bardwell, L. A Conserved Docking Site in MEKs Mediates High-affinity Binding to MAP Kinases and Cooperates with a Scaffold Protein to Enhance Signal

- Transmission. *J. Biol. Chem.* **276**, 10374–10386 (2001).
59. Holland, P. M. & Cooper, J. A. Protein modification: Docking sites for kinases. *Curr. Biol.* **9**, 329–331 (1999).
 60. Sharrocks, A. D., Yang, S. H. & Galanis, A. Docking domains and substrate-specificity determination for MAP kinases. *Trends Biochem. Sci.* **25**, 448–453 (2000).
 61. Kusari, A. B., Molina, D. M., Sabbagh, W., Lau, C. S. & Bardwell, L. A conserved protein interaction network involving the yeast MAP kinases Fus3 and Kss1. *J. Cell Biol.* **164**, 267–277 (2004).
 62. Miller, A. S., Kohout, S. C., Gilman, K. A. & Falke, J. J. CheA kinase of bacterial chemotaxis: Chemical mapping of four essential docking sites. *Biochemistry* **45**, 8699–8711 (2006).
 63. Li, W., Young, S. L., King, N. & Miller, W. T. Signaling properties of a non-metazoan Src kinase and the evolutionary history of Src negative regulation. *J. Biol. Chem.* **283**, 15491–15501 (2008).
 64. Bardwell, L. & Thorner, J. A conserved motif at the amino termini of MEKs might mediate high-affinity interaction with the cognate MAPKs. *Protein Seq. Motif* **21**, (1996).
 65. Narayan, V., Pion, E., Landre, V., Müller, P. & Ball, K. L. Docking-dependent ubiquitination of the interferon regulatory factor-1 tumor suppressor protein by the ubiquitin ligase CHIP. *J. Biol. Chem.* **286**, 607–619 (2011).
 66. Bardwell, A. J., Abdollahi, M. & Bardwell, L. Docking sites on mitogen-activated protein kinase (MAPK) kinases, MAPK phosphatases and the Elk-1 transcription factor compete for MAPK binding and are crucial for enzymic activity. *Biochem. J.* **370**, 1077–1085 (2003).
 67. Biondi, R. M. & Nebreda, A. R. Signalling specificity of Ser/Thr protein kinases through docking-site-mediated interactions. *Biochem. J.* **372**, 1–13 (2003).
 68. Bardwell, A. J., Frankson, E. & Bardwell, L. Selectivity of docking sites in MAPK kinases. *J. Biol. Chem.* **284**, 13165–13173 (2009).
 69. Kallunki, T., Deng, T., Hibi, M. & Karin, M. c-Jun can recruit JNK to phosphorylate dimerization partners via specific docking interactions. *Cell* **87**, 929–939 (1996).
 70. Tatebayashi, K., Takekawa, M. & Saito, H. A docking site determining specificity of Pbs2 MAPKK for Ssk2/Ssk22 MAPKKs in the yeast HOG pathway. *EMBO J.* **22**, 3624–3634 (2003).
 71. Barnakov, A. N., Barnakova, L. A. & Hazelbauer, G. L. Efficient adaptational demethylation of chemoreceptors requires the same enzyme-docking site as efficient methylation. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 10667–10672 (1999).
 72. Yuan, Z. L., Guan, Y. J., Chatterjee, D. & Chin, Y. E. Stat3 dimerization regulated by reversible acetylation of a single lysine residue. *Science (80-.)*. **307**, 269–273 (2005).
 73. Steinbach, J. H. & Akk, G. Applying the Monod-Wyman-Changeux allosteric activation model to pseudo-steady-state responses from GABAA receptors. *Mol. Pharmacol.* **95**, 106–119 (2019).
 74. Hill, A. V. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J. Physiol.* 4–6 (1910). doi:10.1093/nq/36.2.211d
 75. Ferrell, J. E. & Ha, S. H. Ultrasensitivity part II: Multisite phosphorylation, stoichiometric inhibitors, and positive feedback. *Trends Biochem. Sci.* **39**, 556–569 (2014).
 76. Enciso, G. Multisite Mechanisms for Ultrasensitivity. *Lect. Notes Math.* **2102**, 3–39 (2013).
 77. Ferrell, J. E. & Ha, S. H. Ultrasensitivity part I: Michaelian responses and zero-order ultrasensitivity. *Trends Biochem. Sci.* **39**, 496–503 (2014).

78. Golbeter, A. & Koshland, D. E. J. An amplified sensitivity arising from covalent modification in biological systems. *PNAS* **78**, 6840–6844 (1981).
79. Amin, M., Porter, S. L. & Soyer, O. S. Split Histidine Kinases Enable Ultrasensitivity and Bistability in Two-Component Signaling Networks. *PLoS Comput. Biol.* **9**, (2013).
80. Martins, B. M. C. & Swain, P. S. Ultrasensitivity in Phosphorylation-Dephosphorylation Cycles with Little Substrate. *PLoS Comput. Biol.* **9**, (2013).
81. Aoki, K., Yamada, M., Kunida, K., Yasuda, S. & Matsuda, M. Processive phosphorylation of ERK MAP kinase in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 12675–12680 (2011).
82. Mukhopadhyay, H. *et al.* Multisite Phosphorylation Modulates the T Cell Receptor ζ -Chain Potency but not the Switchlike Response. *Biophys. J.* **110**, 1896–1906 (2016).
83. Shindo, Y. *et al.* Conversion of graded phosphorylation into switch-like nuclear translocation via autoregulatory mechanisms in ERK signalling. *Nat. Commun.* **7**, 1–10 (2016).
84. Szomolay, B. & Shahrezaei, V. Bell-shaped and ultrasensitive dose-response in phosphorylation-dephosphorylation cycles: the role of kinase-phosphatase complex formation. *BMC Syst. Biol.* **6**, (2012).
85. Trunnell, N. B., Poon, A. C., Kim, S. Y. & Ferrell, J. E. Ultrasensitivity in the Regulation of Cdc25C by Cdk1. *Mol. Cell* **41**, 263–274 (2011).
86. Wang, L., Nie, Q. & Enciso, G. Nonessential sites improve phosphorylation switch. *Biophys. J.* **99**, L41–L43 (2010).
87. Marzen, S., Garcia, H. G. & Phillips, R. Statistical Mechanics of Monod-Wyman-Changeux (MWC) Models. *J. Mol. Biol.* **425**, (2013).
88. Levitzki, A. *Quantitative Aspects of Allosteric Mechanisms*. (Springer-Verlag Berlin Heidelberg, 1978). doi:10.1007/978-3-642-81231-6
89. Gunawardena, J. Multisite protein phosphorylation makes a good threshold but can be a poor switch. *PNAS* **102**, 14617–14622 (2005).
90. Enciso, G. A. & Ryerson, S. The effect of site-to-site variability in ultrasensitive dose responses. *J. Math. Biol.* **74**, 23–41 (2017).
91. Mukhopadhyay, H., Cordoba, S. P., Maini, P. K., van der Merwe, P. A. & Dushek, O. Systems Model of T Cell Receptor Proximal Signaling Reveals Emergent Ultrasensitivity. *PLoS Comput. Biol.* **9**, (2013).
92. Rubinstein, B. Y., Mattingly, H. H., Berezhkovskii, A. M. & Shvartsman, S. Y. Long-Term dynamics of multisite phosphorylation. *Mol. Biol. Cell* **27**, 2331–2340 (2016).
93. Humphreys, J. M., Piala, A. T., Akella, R., He, H. & Goldsmith, E. J. Precisely ordered phosphorylation reactions in the p38 mitogen-activated protein (MAP) kinase cascade. *J. Biol. Chem.* **288**, 23322–23330 (2013).
94. Varedi, M. K. S., Ventura, A. C., Merajver, S. D. & Lin, X. N. Multisite phosphorylation provides an effective and flexible mechanism for switch-like protein degradation. *PLoS One* **5**, (2010).
95. Basu, K. & Liu, X. Mathematical modeling for multisite phosphorylation with scaffold binding in cell signaling. *Math. Methods Appl. Sci.* **38**, 4521–4529 (2015).
96. Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18–29 (2015).
97. Mayer, B. J. & Yu, J. Protein Clusters in Phosphotyrosine Signal Transduction. *J. Mol. Biol.* **430**, 4547–4556

- (2018).
98. Cheng, Z., Liu, F., Zhang, X. P. & Wang, W. Reversible phosphorylation subserves robust circadian rhythms by creating a switch in inactivating the positive element. *Biophys. J.* **97**, 2867–2875 (2009).
 99. Enciso, G., Kellogg, D. R. & Vargas, A. Compact Modeling of Allosteric Multisite Proteins: Application to a Cell Size Checkpoint. *PLoS Comput. Biol.* **10**, (2014).
 100. Kapuy, O., Barik, D., Domingo Sananes, M. R., Tyson, J. J. & Novák, B. Bistability by multiple phosphorylation of regulatory proteins. *Prog. Biophys. Mol. Biol.* **100**, 47–56 (2009).
 101. Liu, X., Bardwell, L. & Nie, Q. A combination of multisite phosphorylation and substrate sequestration produces switchlike responses. *Biophys. J.* **98**, 1396–1407 (2010).
 102. Sadreev, I. I. *et al.* A systems model of phosphorylation for inflammatory signaling events. *PLoS One* **9**, 1–11 (2014).
 103. Srividhya, J., Li, Y. & Pomerening, J. R. Open cascades as simple solutions to providing ultrasensitivity and adaptation in cellular signaling. *Phys. Biol.* **8**, (2011).
 104. Chan, C., Liu, X., Wang, L., Bardwell, L., Nie, Q. & Enciso, G. Protein scaffolds can enhance the bistability of multisite phosphorylation systems. *PLoS Comput. Biol.* **8**, (2012).
 105. Lu, L. X., Domingo-Sananes, M. R., Huzarska, M., Novak, B. & Gould, K. L. Multisite phosphoregulation of Cdc25 activity refines the mitotic entrance and exit switches. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 9899–9904 (2012).
 106. Dushek, O., Van Der Merwe, P. A. & Shahrezaei, V. Ultrasensitivity in multisite phosphorylation of membrane-anchored proteins. *Biophys. J.* **100**, 1189–1197 (2011).
 107. Salazar, C., Brümmer, A., Alberghina, L. & Höfer, T. Timing control in regulatory networks by multisite protein modifications. *Trends Cell Biol.* **20**, 634–641 (2010).
 108. Gonze, D. & Abou-Jaoudé, W. The Goodwin Model: Behind the Hill Function. *PLoS One* **8**, (2013).
 109. Adair, G. S. The Hemoglobin System V1. The oxygen dissociation curve of hemoglobin. *J. Biol. Chem.* (1925).
 110. Koshland, D. E., Nemethy, J. G. & Filmer, D. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry* **5**, 365–385 (1966).
 111. Monod, J., Wyman, J. & Changeux, J. P. On the Nature of Allosteric Transitions: a Plausible Model. *J. Mol. Biol.* **12**, 88–118 (1965).
 112. Yonetani, T. & Laberge, M. Protein dynamics explain the allosteric behaviors of hemoglobin. *Biochim. Biophys. Acta - Proteins Proteomics* **1784**, 1146–1158 (2008).
 113. Shibayama, N. Functional analysis of hemoglobin molecules locked in doubly liganded conformations. *J. Mol. Biol.* **285**, 1383–1388 (1999).
 114. Pulverer, B., Kyriakis, J. M., Avruch, J., Nikolakaki, E. & Woodgett, J. Phosphorylation of c-jun mediated by MAP kinases. *Nature* **354**, 56–58 (1991).
 115. Bogoyevitch, M. A., Ngoei, K. R. W., Zhao, T. T., Yeap, Y. Y. C. & Ng, D. C. H. c-Jun N-terminal kinase (JNK) signaling: Recent advances and challenges. *Biochim. Biophys. Acta - Proteins Proteomics* **1804**, 463–475 (2010).
 116. Wei, W., Jin, J., Schlisio, S., Harper, J. W. & Kaelin, W. G. The v-Jun point mutation allows c-Jun to escape GSK3-dependent recognition and destruction by the Fbw7 ubiquitin ligase. *Cancer Cell* **8**, 25–33 (2005).

117. Bardwell, L. & Shah, K. Analysis of mitogen-activated protein kinase activation and interactions with regulators and substrates. *Methods* **40**, 213–223 (2006).
118. Chen, L. *et al.* Structure of the DNA-binding domains from NFAT, FOS and JUN bound specifically to DNA. *Chemtracts* **12**, 768–773 (1999).
119. Li, Y. *et al.* TOPK promotes lung cancer resistance to EGFR tyrosine kinase inhibitors by phosphorylating and activating c-Jun. *Oncotarget* **7**, 6748–6764 (2016).
120. Xie, X. *et al.* C-Jun N-terminal kinase promotes stem cell phenotype in triple-negative breast cancer through upregulation of Notch1 via activation of c-Jun. *Oncogene* **36**, 2599–2608 (2017).
121. Li, J. *et al.* DGKA provides platinum resistance in ovarian cancer through activation of cJUN-WEE1 signaling. *Am. Assoc. Cancer Res.* (2020). doi:10.1158/1078-0432.CCR-19-2706
122. Park, K. H. *et al.* L-DOPA modulates cell viability through the ERK-c-Jun system in PC12 and dopaminergic neuronal cells. *Neuropharmacology* **101**, 87–97 (2016).
123. Zhou, X. *et al.* Exopolysaccharides from *Lactobacillus plantarum* NCU116 induce c-Jun dependent Fas/FasL-mediated apoptosis via TLR2 in mouse intestinal epithelial cancer cells. *Sci. Rep.* **7**, 1–13 (2017).
124. Szymanska, E., Skowronek, A. & Miaczynska, M. Impaired dynamin 2 function leads to increased AP-1 transcriptional activity through the JNK/c-Jun pathway. *Cell. Signal.* **28**, 160–171 (2016).
125. Vogt, P. K. Jun, the oncoprotein. *Oncogene* **20**, 2365–2377 (2001).
126. Morton, S., Davis, R. J., McLaren, A. & Cohen, P. A reinvestigation of the multisite phosphorylation of the transcription factor c-Jun. *EMBO J.* **22**, 3876–3886 (2003).
127. Arias, J. *et al.* Activation of cAMP and mitogen responsive genes relies on a common nuclear factor. *Nature* **370**, 226–229 (1994).
128. Ip, Y. T. & Davis, R. J. Signal transduction by the c-Jun N-terminal kinase (JNK) - From inflammation to development. *Curr. Opin. Cell Biol.* **10**, 205–219 (1998).
129. Raivich, G. c-Jun Expression, activation and function in neural cell death, inflammation and repair. *J. Neurochem.* **107**, 898–906 (2008).
130. Minden, A. *et al.* c-Jun N-terminal phosphorylation correlates with activation of the JNK subgroup but not the ERK subgroup of mitogen-activated protein kinases. *Mol. Cell. Biol.* **14**, 6683–6688 (1994).
131. Nateri, A. S., Spencer-Dene, B. & Behrens, A. Interaction of phosphorylated c-Jun with TCF4 regulates intestinal cancer development. *Nature* **437**, 281–285 (2005).
132. Watson, A. *et al.* Phosphorylation of c-jun is necessary for apoptosis induced by survival signal withdrawal in cerebellar granule neurons. *J. Neurosci.* **18**, 751–762 (1998).
133. Papavassiliou, A. G., Treier, M. & Bohmann, D. Intramolecular signal transduction in c-Jun. *EMBO J.* **14**, 2014–2019 (1995).
134. Csizmok, V. *et al.* Multivalent Interactions with Fbw7 and Pin1 Facilitate Recognition of c-Jun by the SCFFbw7 Ubiquitin Ligase. *Structure* **26**, 28-39.e2 (2018).
135. Gnad, F., Ren, S., Cox, J., Olsen, J. V., Macek, B., Orosi, M., & Mann, M. PHOSIDA (phosphorylation site database): Management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biology*, **8**(11).(2007)

Chapter 2: The WW domain of the scaffolding protein IQGAP1 is neither necessary nor sufficient for binding to the MAPKs ERK1 and ERK2

This work was published as: Bardwell, A.J., Lagunes, L., Zebarjedi, R., and Bardwell, L. The WW domain of the scaffolding protein IQGAP1 is neither necessary nor sufficient for binding to the MAPKs ERK1 and ERK2. (2017). *J. Biol. Chem.* 292, 8750-8761

My contribution to this paper include the following:

I performed the cloning of mutants IQGAP1 truncations hIQGAP1(1–863), hIQGAP1(1–678), pGEM3Z-IQGAP1(679–863), pGEM3Z-IQGAP1(720–863), and pGEM3Z-IQGAP1 (432–863).

I performed binding assay experiments that were included in Figures 2 and 3.

Chapter 2: The WW domain of the scaffolding protein IQGAP1 is neither necessary nor sufficient for binding to the MAPKs ERK1 and ERK2

Section 1: Abstract

Mitogen-activated protein kinase (MAPK) scaffold proteins, such as IQ motif containing GTPase activating protein 1 (IQGAP1), are promising targets for novel therapies against cancer and other diseases. Such approaches require accurate information about which domains on the scaffold protein bind to the kinases in the MAPK cascade. Results from previous studies have suggested that the WW domain of IQGAP1 binds to the cancer-associated MAPKs ERK1 and ERK2, and that this domain might thus offer a new tool to selectively inhibit MAPK activation in cancer cells. The goal of this work was therefore to critically evaluate which IQGAP1 domains bind to ERK1/2. Here, using quantitative *in vitro* binding assays, we show that the IQ domain of IQGAP1 is both necessary and sufficient for binding to ERK1 and ERK2, as well as to the MAPK kinases MEK1 and MEK2. Furthermore, we show that the WW domain is not required for ERK-IQGAP1 binding, and contributes little or no binding energy to this interaction, challenging previous models of how WW-based peptides might inhibit tumorigenesis. Finally, we show that the ERK2-IQGAP1 interaction does not require ERK2 phosphorylation or catalytic activity and does not involve known docking recruitment sites on ERK2, and we obtain an estimate of the dissociation constant (K_d) for this interaction of 8 μM . These results prompt a re-evaluation of published findings and a refined model of IQGAP scaffolding.

Section 2: Introduction

The RAS-RAF-MEK-ERK signaling pathway, often referred to as the RAS/MAPK cascade, has been a focus of cancer drug development (1, 2). The success of small-molecule inhibitors of RAF and MEK have validated these efforts; however, the emergence of clinical drug resistance remains a major challenge (3, 4).

Signal propagation through the MAPK cascade is facilitated by scaffold proteins such as KSR, Paxillin, and IQGAP1 (5). Scaffold proteins bind to and assemble multiple elements of signaling/regulatory pathways. They are thought to tether their bound components near each other, thereby increasing the rate at which one activates the other. Furthermore, they operate in distinct subcellular locations and in a spatiotemporally regulated manner (5–7). For these reasons, scaffold proteins provide new therapeutic approaches to cancer and other diseases (8).

IQGAP proteins are evolutionarily conserved in eukaryotes (8–18). They function as scaffold proteins that facilitate the formation of complexes that regulate both cytoskeletal dynamics and intracellular signaling.

IQGAP proteins have been highly studied because of their relevance to basic biology and to human disease.

Originally discovered in 1994 (19), IQGAP1 is the founding and best-studied member of a family that includes 3 paralogs in humans (IQGAP1, IQGAP2, and IQGAP3). IQGAP1 overexpression has been implicated in the progression of many cancers (18, 20), and the presence of IQGAP1 has been shown to promote RAS-driven tumorigenesis in mouse models (21). In addition, many bacterial and viral pathogens, including the Ebola and Marburg viruses, have been shown to hijack IQGAP1 during the course of infection (9, 10).

Consistent with their proposed role as scaffold proteins, mammalian IQGAP1 orthologs are over 1600 amino acids long, and contain multiple domains that can mediate protein-protein interactions (**Fig. 1A**). From amino to carboxyl terminus, these domains include a calponin-homology domain, a region containing several internal repeat sequences that have the capacity to form coiled-coils (IR), a WW domain, an IQ domain (consisting of four closely spaced IQ motifs), a GTPase-activating protein-related domain, and a RasGAP C-terminal domain. Multiple binding partners have been identified for most of these domains (13).

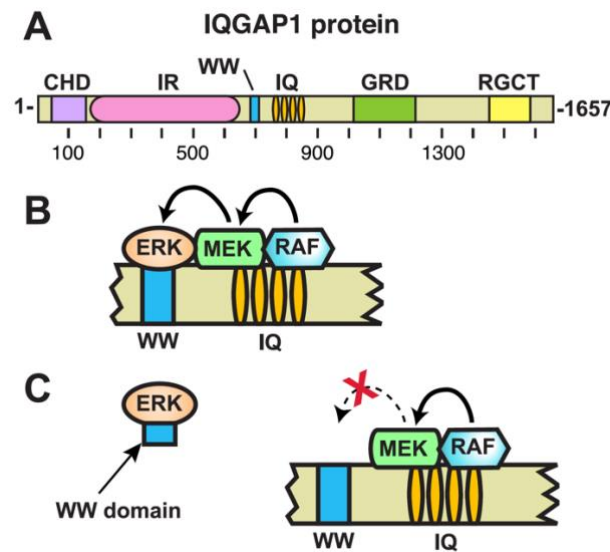


Figure 1. The IQGAP1 scaffold protein. **A**, schematic depicting full-length human IQGAP1 protein, and the domains it contains. CHD, calponin homology domain; IR, internal repeated sequence/coiled-coil domain; WW, WW domain; IQ, IQ domain; GRD, GTPase-activating protein-related domain; RGCT, RasGAP C-terminal domain. **B**, schematic interpretation of proposed model of the function of IQGAP1 as a scaffold protein for the MAPK pathway, based on a similar figure in Ref. 8. According to the model, the IQ domain of IQGAP1 binds to RAF and MEK, and the nearby WW domain binds to ERK. These interactions are thought to facilitate RAF phosphorylation of MEK, and MEK

phosphorylation of ERK. **C**, proposed mechanistic model for the antitumor efficacy of the isolated IQGAP WW domain studied by Jameson et al. (21). In this model, the WW domain binds to ERK and blocks the ability of ERK to productively interact with IQGAP1 (8, 21). The WW domain fragment studied by Jameson et al. (21) consisted of IQGAP1 residues 680–711, plus N-terminal myc and polyarginine tags.

A scaffolding function for IQGAP1 was first proposed when it was observed to link Ca²⁺/calmodulin and Cdc42 signaling (22–24). More recent data also suggest that IQGAP can act as a scaffold in the Wnt pathway (25). However, perhaps the best characterized example of IQGAP1 scaffold function is in its interactions with elements of the RAS/MAPK pathway. As shown in **Fig. 1B**, both ERK1 and ERK2 (MAP kinases that are activated in numerous human cancers) and MEK1 and MEK2 (MAPK kinases that activate ERK1 and ERK2) have been shown to bind to IQGAP1. Upstream components of the MAPK pathway have also been shown to bind to IQGAP1, including the MEK activator BRAF, as well as multiple receptor tyrosine kinases (9).

For many years, the field has believed that ERK1 and ERK2 bind to the WW domain of IQGAP1 (26, 27). WW domains are compact units that fold into a three-stranded- β -sheet structure (28), and have been shown to bind to Pro-rich sequences such as PPXY and PPPR, or to phospho-Ser/Thr-Pro sequences (29–34). ERK1 and ERK2 are the only proteins purported to interact with the WW domain of IQGAP1 (13). In contrast, the binding of BRAF and MEK1/2 to IQGAP1 requires the presence of the IQ domain (26, 35). The IQ domain of IQGAP1 consists of four tandem IQ motifs (**Fig. 1A**). IQ motifs are found in many calcium-regulated proteins (36). They consist of a stretch of about 18–25 amino acid residues, and form amphiphilic α helices that can bind to calmodulin and S100-family proteins, among other ligands.

The assertion that the WW domain of IQGAP1 binds to ERK1 and ERK2 has been widely cited in primary research papers (e.g. see Refs. 21, 26, and 37–50) and reviews (e.g. Refs. 8–18). Indeed, it recently motivated a high-profile translational study in which cancer cells were treated with a cell-permeable version of the WW domain of human IQGAP1 (8, 11, 21). The idea underlying this study was that the WW domain fragment would competitively bind to ERK1/2 and prevent these MAP kinases from interacting with IQGAP1, thus selectively inhibiting MAP kinase activation (**Fig. 1C**). Indeed, the idea seemed to work, in as much as the WW domain fragment inhibited the proliferation, migration, and tumorigenesis of breast, colorectal, and melanoma tumor cells that contained activating mutations in the RAS/MAPK pathway (8, 11, 21).

Herein we re-examined the binding of ERK1 and ERK2 to IQGAP1. In contrast to previous findings, we show that the WW domain of IQGAP1 is neither necessary nor sufficient for binding to ERK1 and ERK2. Rather, the IQ domain of IQGAP1 is both necessary and sufficient for high-affinity ERK binding. Our results

thus prompt a re-evaluation of several highly cited published studies, and suggest a new model for IQGAP scaffolding function.

Section 3: Experimental Procedures

3.1 Genes

The mammalian genes used in this study were human IQGAP1 (NCBI accession number NM_003870), human ERK1 (MAPK3; NM_002746), human ERK2 (MAPK1, NP_620407), rat Erk2 (NM_053842), human MEK1 (MAPK2K1, NM_002755), and human MEK2 (MAP2K2, NM_030662).

3.2 Plasmids for the production of GST fusion proteins

The vector used for generating GST fusion proteins was pGEX-LB, a derivative of pGEX-4T-1 (Amersham Biosciences) (85). In pGEX-LB, an encoded Pro residue is replaced with a Gly-Gly-Gly-Gly-Gly-Ser-Gly coding sequence to promote the independent functioning of the GST and fusion moieties. Plasmid GST-hERK1 encodes a fusion of GST to full-length human ERK1, GST-rERK2 encodes a fusion of GST to full-length rat ERK2 (85). GST-hrERK2 encodes a fusion of GST to a sequence that encodes human ERK2 protein; this was generated for this study from the rat sequence by site-directed mutagenesis (the human and rat protein sequences differ in only 3 positions). The K54A mutant, the T185A/Y187F mutant, and the DGM (D318A/D321A/L115A/Q119A) were generated from this plasmid by site-directed mutagenesis. See **Table 2** for primer sequences. The QuikChange and QuikChange Multi kits (Agilent) were used for all site-directed mutagenesis reactions.

Table 1
Binding assay data for ERK2-IQGAP1 interaction

Experiment ^a	Binding ^b	K_d ^c
	%	μM
A-1	37.5	3.0
A-2	25.1	5.3
A-3	18.8	7.7
A-4	16.3	9.2
A-5	16.7	8.9
A-6	28.9	4.4
A-7	11.9	13.2
A-8	16.4	9.1
A-9	19.2	7.5
Mean	21.2	7.6
Median	18.8	7.7
Standard deviation	7.9	3.1
Standard error	2.6	1.0

^a Binding reactions (200 μl) contained ~ 1 pmol (~ 5 nM) ³⁵S-labeled, *in vitro* translated, full-length IQGAP1 and 25 μg (1.8 μM) of GST-hERK2 fusion protein. Every experiment contained duplicate points (also known as technical replicates); these were averaged to obtain the "% binding" number shown.

^b Percent of the input ³⁵S-labeled protein that bound to the GST fusion protein.

^c Calculated based on the known input concentrations and percent binding, as described elsewhere (65).

Table 2
Oligonucleotides used in this study

Name	Sequence (5' → 3') ^a	Use
LL-hIQG1-P864stop-s	ACAAGACTCTCATCAATGCTGAGGAT TAA CCCTATGGTTGTGGTCC	IQGAP1 (1–863)
JB-hIQG1-L720stop-s	ATTTTGTGCAAAATCTATGCAG TAA TCTCGGGAGGAGATCCAGAGTTC	IQGAP1 (1–719)
LL-hIQG1-G679stop-s	AGCCAAGAAGAAAAAAGTGGCAGT TAA GATAATAACAGCAAG	IQGAP1 (1–678)
hIQGAP1(679-X)	GGAGGCGGTGGATCCACC ATG CGGAGATAATAACAGCAAGTGG	IQGAP1 (679–863)
hIQGAP1(720-X)	GGAGGCGGTGGATCCACC ATG CTTTCTCGGGAGGAGATCCA	IQGAP1 (720–863)
hIQGAP1(X-863)	GCCGCTCGAGTCCGACT TAA TCTCAGCATTGATGAGAG	679 and 720 to 863
LL-hIQGAP1-432-Y	GGTACCCGGGGATCCACC ATG GAGCTGGTTACCCCTGCAGCG	IQGAP1 (432–863)
LL-hIQGAP1-Y-863	TGCCCTGCAGGTCGACT TAA TCTCAGCATTGATGAGAG	IQGAP1 (432–863)
JB-W685A-s	TAGGAGATAATAACAGCAAG GCG GTAAGCACTGGTAAAAG	IQGAP1 (1–863)wmut
JB-P710A-s	CCAGGAAGGAGGATGGGATGA AGCC CCAAATTTTG	IQGAP1 (1–863)wmut
JB-Y696R-Y697A-N699A-s	GTGAAGCACTGGGTAAAAGGTGGATATT TCGT GCCAC GCT CTGGAGACCCAGGAAGG	IQGAP1 (1–863)wmut
JB-rERK2-L44V-s	GGCATGGTTTGTCTGCTTATGATAAT GTCA ACAAAGTTTGA	hERK2
JB-rERK2-ins8AGlong-s	CCACCATGGCGGGCGGGCGGGCGGG GCGGGC CCGGAGATGGTCCGCGGCAGGTGT	hERK2
JB-hrERK2-K54A	AACAAAGTTCGAGTTGCTAT GCG GAAATCAGTCTCTTTTGAGCAC	hERK2 K54A
JB-hrERK2T185A/Y187F	TCATACAGGGTTCTT GCG AGAGTTTGTAGCCACGCGTTGG	hERK2 T185A Y187F
JBhrERK2-D318A/D321A-s	CCTGGAGCAGTATTAT GCC CAAGT GCT GAGCCATTGC	hERK2 DGM
JBhrERK2-L115A/Q119A-s	TGGAGACAGATCTTTACAA GCC TTGAAGAC GCG CACCTCAGCAATGATCATA	hERK2 DGM

^a BamHI and Sall restriction sites used in cloning are underlined. Mutagenized codons are shown in bold. Introduced start and stop codons are also shown in bold.

3.3 Plasmids for the production of in vitro translated IQGAP1 and Derivatives

A human IQGAP1 cDNA in expression vector pCR-Blunt II-TOPO was obtained from Dharmacon/GE Healthcare; this clone is from the mammalian gene collection, accession number BC139731. In vitro transcription and translation of full length IQGAP1 was possible directly from the T7 promoter in this plasmid. The IQGAP truncations IQGAP1(1–863), IQGAP1(1–719), and IQGAP1(1–678) were derived from pCR-Blunt II-TOPO-IQGAP1 by introducing stop codons at codons 864, 720, and 679 of the IQGAP1 coding sequence via site-directed mutagenesis. The coding strand primers for these mutagenesis reactions were LL-hIQG1-P864stop-s, JB-hIQG1-L720stop-s, and LL-hIQG1-G679stop-s, respectively (**Table 2**).

The plasmid encoding IQGAP1(1–863)wmut was derived from IQGAP1(1–863) by site-directed mutagenesis in two stages. First, the W685A and P710A substitutions were introduced with primers JB-W685A-s and JB-P710A-s. This was done in a single step using QuikChange Multi. Next, this intermediate derivative was used as the template in a mutagenesis reaction using coding strand primer JB-Y696R-Y697AN699A-s and the corresponding antisense primer. The final product contains the substitutions W685A, Y696R, Y697A, N699A, P710A.

To construct pGEM3Z-IQGAP1(679–863) (used for the in vitro transcription and translation of IQGAP1(679–863)), a polymerase chain reaction (PCR) was performed with primers hIQGAP1(679-X) and hIQGAP1(X-863) (**Table 2**). The resulting product was then inserted into pGEXLB using recombination-based cloning (Cold Fusion, System Biosciences). Next, the insert was excised from this vector by digestion with restriction enzymes BamHI and Sall, and ligated into plasmid pGEM3Z (Promega), which had been cut with the same enzymes. Plasmid pGEM3Z-IQGAP1(720–863) was constructed using a similar strategy.

Plasmid pGEM3Z-IQGAP1 (432–863) was also constructed using a similar strategy, except that the PCR product was digested directly with BamHI and Sall prior to insertion into the corresponding sites of pGEM3Z. See **Table 2** for primer sequences. All IQGAP1 derivatives were confirmed by DNA sequencing.

3.4 Protein purification

GST fusion proteins were expressed in bacteria, purified by affinity chromatography using glutathione-Sepharose (GE Healthcare), and quantified as described elsewhere (85, 86). In vitro transcription and translation Proteins labeled with [³⁵S]methionine were produced by coupled transcription and translation reactions (T7, Promega). Translation products were partially purified by ammonium sulfate precipitation (65), and resuspended in binding buffer (20 mM Tris-HCl, pH 7.5, 125 mM KOAc, 0.5 mM EDTA, 1 mM DTT, 0.1% (v/v) Tween 20, 12.5% (v/v) glycerol) prior to use in binding assays.

3.5 Protein binding assays

Protein binding assays were performed as described previously (85, 86). Quantification of binding was performed on a Typhoon TRIO Imager using phosphorimaging mode. Percent binding was determined by comparing the input with the amount that was co-sedimented. Each binding assay presented in this paper was repeated at least three separate times (i.e. three independent experiments), with duplicate points (i.e. technical replicates) in each experiment. Technical replicates in a given experiment are averaged together to obtain a single data point. We define “independent experiments” as experiments performed on different days, with fresh batches of GST fusion proteins and in vitro translated proteins.

Statistical analysis

Statistical analysis of binding assay results was performed using Welch’s unequal variance t test with two tails (87). This was accomplished in Microsoft Excel using the T.TEST function, setting the “tail” option to 2, and the “type” option to 3.

Section 4: Results

4.1 IQ domain of IQGAP1 is necessary for binding to ERK2

An initial goal of this study was to verify and more precisely delineate the domain(s) of IQGAP1 that were necessary and sufficient for binding to the MAP kinases ERK1 and ERK2. As a first step in this process, we set out to confirm the finding of Roy et al. (27), who first showed that human IQGAP1 binds to ERK2. In this study, Roy et al. (27) used rat ERK in co-sedimentation assays with in vitro translated human IQGAP1.

We used a very similar approach; rat ERK2 was fused at its N-terminus to *Schistosoma japonicum* glutathione S-transferase (GST), and the resulting fusion protein (GST-rERK2) was expressed in bacteria and purified by adsorption to glutathione-Sepharose beads. GST-rERK2 (or GST alone as a negative control) was then incubated with full-length human IQGAP1 that had been produced in radiolabeled form by in vitro translation (Fig. 2A). Bead-bound complexes were collected by sedimentation, washed extensively, and analyzed by SDS-PAGE and autoradiography.

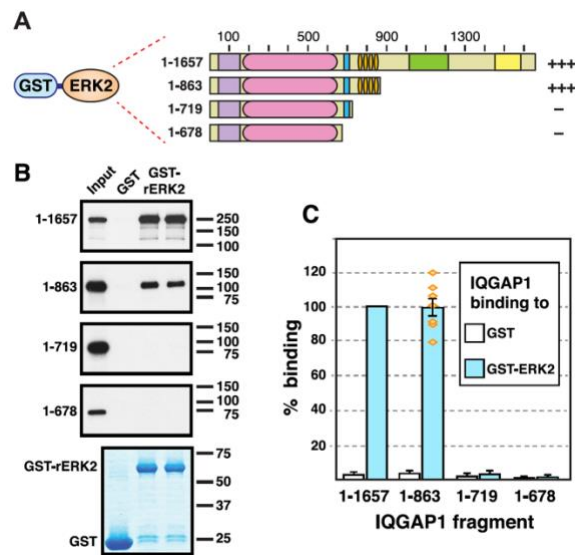


Figure 2. The IQ domain of IQGAP1 is necessary for binding to ERK2; the WW domain is not sufficient. **A**, rat ERK2, fused to GST, was tested for binding to full-length human IQGAP1, or to truncated derivatives of IQGAP1. Qualitative results of these experiments are shown on the right: +++ indicates strong binding; — indicates minimal binding. **B**, as shown in **A**, 35S radiolabeled full-length human IQGAP1 protein and truncated derivatives were prepared by in vitro translation and partially purified by ammonium sulfate precipitation, and portions (10% of the amount added in the binding reactions) were resolved on a 10% SDS-polyacrylamide gel (Input). Samples (μ l pmol) of the same proteins were incubated with 25 μ g of GST or GST-ERK2 bound to glutathione-Sepharose beads, and the resulting bead-bound protein complexes were isolated by sedimentation and resolved by 10% SDS-PAGE on the same gel. The gel was analyzed by staining with GelCode Blue (Thermo Scientific) for visualization of the bound GST fusion protein (a representative example is shown in the lowest panel) and by X-ray film exposure for visualization of the bound radiolabeled protein (upper four panels). **C**, quantification of the binding of IQGAP1 derivatives to GST or GST-ERK2, normalized to the percent binding of full-length IQGAP1 to GST-ERK2. The results shown are the average of at least 5 independent repetitions of the binding assay shown in **A** and **B**, with duplicate points (i.e. technical replicates) in each repetition. S.E. bars are shown ($n = 5$ to 7). The scatter of the individual normalized data points is also shown for the binding of ERK2 to IQGAP1(1– 863). The means for ERK2-IQGAP1 and ERK2-IQGAP1(1– 863) binding were significantly different from all other the means shown ($p = 0.01$), but were not significantly different from each other ($p = 0.98$, thus the null hypothesis that the population means are the same cannot be rejected with confidence). The minimal binding of ERK2 to IQGAP1(1–719) was not significantly different from that of ERK2 to IQGAP1(1– 678) ($p = 0.91$), nor was it significantly different from the minimal binding of GST alone to IQGAP1(1–719) ($p = 0.41$).

Human IQGAP1 is a 1632-residue protein, with a calculated molecular mass of 189 kDa. As shown in **Fig. 2B** (Input lane) IQGAP1 migrated with an apparent molecular mass of 250 kDa on a 10% SDS-PAGE gel. Also as shown in **Fig. 2B**, full-length IQGAP1 bound efficiently to ERK2. Furthermore, this binding was specific, because only trace precipitation of IQGAP1 occurred when GST was used instead of the GST-rERK2 fusion protein.

To delineate the domain(s) of IQGAP1 involved in binding to ERK2, we utilized a series of C-terminal truncation mutants of IQGAP1 (**Fig. 2A**). These mutants were constructed using site-directed mutagenesis to introduce translation termination (“stop”) codons after codons 678, 719, or 863. These derivatives all contain the CHD and IR domains, but differ in the presence of the WW and IQ domains. The IQGAP1(1–678) mutant protein lacks both the IQ and WW domains; IQGAP1(1–719) lacks the IQ domain but contains the WW domain; and IQGAP1(1–863) contains both the IQ and WW domains (**Fig. 2A**).

IQGAP1(1–863) was previously shown to bind rat ERK2 about as well as full-length IQGAP1 did (27). We confirmed this finding (**Fig. 2B**). Indeed, when we quantified the results of 7 independent binding assays, the binding efficiency of full length IQGAP1 was not significantly different from that of IQGAP1(1–863) (**Fig. 2C**).

In stark contrast, both IQGAP1(1–719) and IQGAP1(1–678) exhibited negligible binding to GST-rERK2 (**Fig. 2B**), and this minimal binding was not significantly different from each other, nor was it significantly different to their binding to GST alone (**Fig. 2C**). Thus, these results indicate that the IQ domain is necessary for binding to ERK2 (because 1–863 bound whereas 1–719 did not), and also show that the WW domain is not sufficient for binding (because 1–719 did not bind).

4.2 Human ERK2 binds to human IQGAP1

As noted above, the original discovery of ERK-IQGAP binding was made using rat ERK2 and human IQGAP1 (27); we used this same cross-species configuration in **Fig. 2**. To ascertain if the same pattern of interactions seen in **Fig. 2** would also be observed using human ERK2, we fused human ERK2 to GST and purified this GST-hERK2 protein from bacteria. As shown in **Fig. 3**, full-length human IQGAP1 bound equivalently to both rat ERK (rERK2) and human ERK2 (hERK2). Likewise, IQGAP1(1–863) bound equivalently to both rat and human ERK2. Finally, IQGAP1(1–719) and IQGAP1(1–678) bound to neither ERK2 ortholog. We performed 9 independent, quantitative binding assay experiments between human ERK2 and IQGAP1, with technical replicates in each experiment. From these data we were able to obtain an estimate of 7.6 μ M for the dissociation constant (Kd) of this interaction (**Table 1**).

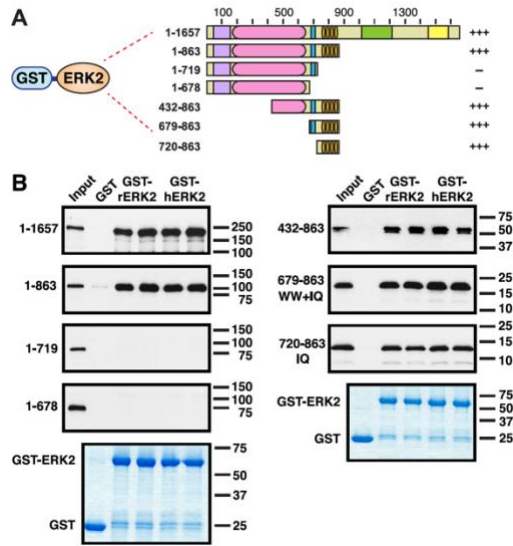


Figure 3. The IQ domain of IQGAP1 is sufficient for binding to ERK2; the WW domain is not necessary. **A**, rat or human ERK2, fused to GST, were tested for binding to full-length human IQGAP1, or to fragments of IQGAP1. Qualitative results of these experiments are shown on the right: +++ indicates strong binding; — indicates minimal binding. **B**, autoradiograms of representative experiments of binding assays are described in **A**. Each binding assay shown was repeated three separate times (i.e. three independent experiments), with duplicate points (i.e. technical replicates) in each experiment. Other details are as described in the legend to **Fig. 2**.

4.3 The IQ domain is sufficient for binding to ERK2

To ask if the IQ domain of IQGAP1 is sufficient for binding to ERK2, we made three additional IQGAP1 fragments. As shown in **Fig. 3A**, IQGAP1(432–863) contains half of the IR domain, the (entire) WW domain, and the IQ domain. IQGAP1(679–863) contains only the WW and IQ domains. Finally, IQGAP1(720–863) contains just the IQ domain. These mutants were tested for binding both to rat ERK2 (rERK2) and human ERK2 (hERK2).

IQGAP1(432–863) was previously shown to bind to rat ERK2 (27). As shown in **Fig. 3B**, we confirmed this finding, and extended it to human ERK2. In vitro-translated IQGAP1(432–863) migrated on SDS-PAGE gels as two forms (**Fig. 3B**): a major, slower migrating form, corresponding to the complete translation product, and a minor, slightly faster migrating form of lower molecular mass. Such faster migrating forms are often seen in cell-free translation reactions, and are typically caused by a low frequency of premature translation termination or internal initiation (51).

Also as shown in **Fig. 3B**, both IQGAP1(679–863) and IQGAP1(720–863) bound to both rat and human ERK2. Importantly, because IQGAP1(720–863), which contains just the IQ domain, bound to ERK2, we conclude that the IQ domain is sufficient for binding to ERK2.

4.4 ERK2 phosphorylation is not required for binding to IQGAP1

ERK2, like most other MAP kinases, is activated by dual phosphorylation at a Thr and Tyr residue in its activation loop. Dual phosphorylation causes remodeling of the activation loop conformation so as to reorganize active site residues, open up substrate specificity determinants, and expose a hydrophobic docking pocket used by some substrates containing FXFP-type docking sites (52–54). Dual phosphorylation of ERK1 and ERK2 is catalyzed by MEK1 and MEK2 during physiological pathway activation. However, ERK2 protein is known to exhibit a low level of autophosphorylation on these residues under various conditions, including bacterial expression (55).

To ascertain whether phosphorylation of ERK2 was necessary for its ability to bind to IQGAP1, we constructed two mutant versions of human ERK2 incapable of undergoing autophosphorylation, and compared their ability to bind IQGAP1 with wild-type ERK2. The first ERK2 mutant, K54A, contains a substitution of a highly conserved catalytic lysine residue; this substitution has been shown to render ERK2 catalytically inactive (56, 57). The second mutant, T185A/Y187F (hereafter designated ERK2-AF), changes the dual phosphorylation sites to non-phosphorylatable residues (54). As shown in **Fig. 4**, both ERK2-K54A and ERK2-AF bound to full-length IQGAP1 and IQGAP1(1–863) comparably to wild-type ERK2. These results indicate that phosphorylation and activation of ERK2 is not required for its ability to bind to IQGAP1.

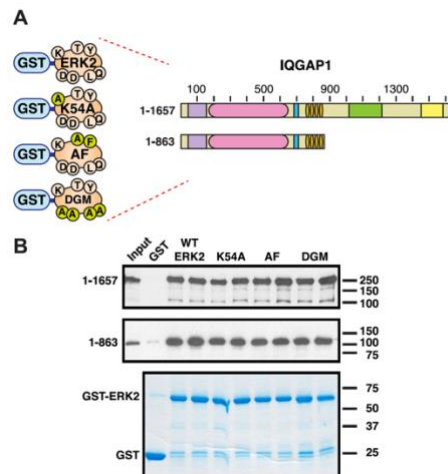


Figure 4. Further characterization of the ERK2-IQGAP1 interaction. **A**, human ERK2, or mutant derivatives thereof, fused to GST, were tested for binding to full-length human IQGAP1, or to IQGAP1(1–863). The ERK2 alleles tested were the wild-type allele (ERK2), catalytically inactive (K54A mutation, K54A), unphosphorylatable and unactivatable (T185A Y185F mutations, AF), and docking groove mutated (L115A, Q119A, D318A, D321A mutations, "DGM"). Small circles on the schematics indicate the wild-type residues and the alterations thereof. **B**, autoradiograms of representative experiments of binding assays described in **A**. Each binding assay shown was repeated three separate

times (i.e. three independent experiments), with duplicate points (i.e. technical replicates) in each experiment. Other details are as described in the legend to **Fig. 2**.

4.5 The docking recruitment site of ERK2 is not involved in IQGAP1 binding

The interaction of MAP kinases with scaffold proteins, other kinases, substrates, and phosphatases often involves the MAPK binding to a short linear motif, a MAPK-docking site, on its binding partner. A well-known class of MAPK-docking sites, designated “D-sites,” has the consensus Lys/Arg₁₋₃-X₁₋₆-Φ-X-Φ, where “X” is any residue and “Φ” is a hydrophobic residue. D-sites were first identified in MAPK kinases (58, 59) and certain transcription factors (60, 61), and were subsequently found in numerous other MAPK partners, including scaffold proteins such as yeast Ste5 and mammalian JIP1 and JIP3 (62–64). A second docking motif (consensus Leu-X₁₋₂-Arg/Lys₂₋₅), related to the D-site, is found in MAPK-activated kinases such as RSK1 and MAPKAP2 (65, 66). Both classes of docking sites are known to bind to a charged surface patch and adjacent hydrophobic cleft on MAPKs referred to as the D-recruitment site or “docking groove” (67, 68).

The IQ domain of IQGAP1 contains two stretches that loosely fit the D-site consensus sequence, including 790KQKKAYQDRLAY801 and 822RKRYRDLQY831. This observation suggested to us the possibility that, as is true of several other MAPK scaffold proteins, ERK2-IQGAP1 binding might be mediated by the docking groove of ERK2. To address this possibility, we constructed a mutant version of human ERK2 that contained 4 amino acid substitutions (L115A, Q119A, D318A, D321A) known to disrupt docking groove-mediated interactions (62, 69–72). As shown in **Fig. 4**, both full-length IQGAP1 and IQGAP1(1–863) bound comparably to wild-type ERK2 and to the ERK2 docking-groove mutant (DGM).³ Hence, the docking groove of ERK2 does not appear to play a significant role in ERK2-IQGAP1 binding.

Another docking motif found in MAPK binding partners has been named the “DEF motif” (consensus FXFP) (73). The IQ domain of IQGAP1 contains no matches to this consensus. Moreover, the complementary binding site on ERK2, designated the F-recruitment site, is only fully formed upon ERK2 phosphorylation and activation (52, 74), and we showed above that ERK2 phosphorylation and activation are not required for IQGAP1 binding.

To summarize, our results strongly suggest that ERK2-IQGAP1 binding is not mediated by any known MAPK-docking sites on IQGAP1 or recruitment sites on ERK2. Further studies will be required to delineate the region of ERK1/2 that mediates binding to IQGAP1.

4.6 The IQ domain is necessary and sufficient for binding to ERK1, MEK1, and MEK2

In prior studies, ERK1 has not been studied as extensively as ERK2 with regard to IQGAP1 binding. ERK1 has been shown to co-immunoprecipitate with full-length IQGAP1 from MCF-7 cells (26), and to increase the binding of MEK1 to IQGAP1 in vitro (26). It is generally assumed that ERK1 binds to the WW domain of IQGAP1, as is purported for ERK2. However, the domain on IQGAP1 to which ERK1 binds has not, to our knowledge, been carefully mapped.

MEK1 and MEK2 have also been shown to bind to IQGAP1 (26). In this case, domain mapping experiments indicated that the IQ domain of IQGAP1 was necessary for MEK binding (26). However, whether or not this domain is sufficient for MEK binding has not been addressed.

To investigate these questions, we expressed and purified full-length human ERK1, MEK1, and MEK2 as GST fusions, and tested them for binding to full-length IQGAP1 and to the panel of IQGAP1 deletion mutants (**Fig. 5A**).

Like GST-ERK2, GST-ERK1 is efficiently expressed and translated in *Escherichia coli*, resulting in abundant production of the expected full-length product (**Fig. 5B**, bottom panels). As shown in Fig. 5B, GST-ERK1 bound efficiently to full-length IQGAP1.

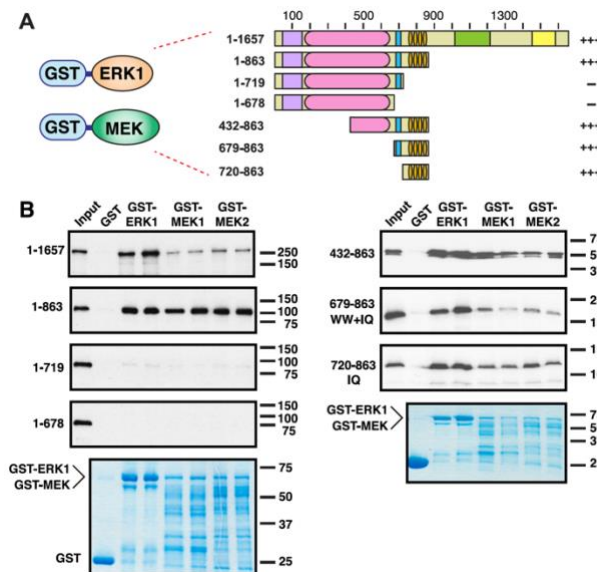


Figure 5. The IQ domain of IQGAP1 is sufficient for binding to ERK1, MEK1, and MEK2. **A**, human ERK1, MEK1, or MEK2, fused to GST, was tested for binding to full-length human IQGAP1, or to fragments of IQGAP1. Qualitative results of these experiments are shown on the right: +++ indicates strong binding; — indicates minimal binding. **B**, autoradiograms of representative experiments of binding assays described in **A**. Each binding assay shown was repeated three separate times (i.e. three independent experiments), with duplicate points (i.e. technical replicates) in each experiment. Other details are as described in the legend to **Fig. 2**.

Bacterial production of full-length GST-MEK1 and GSTMEK2 is less efficient (compared with the GST-ERKs), resulting in both the production of the expected full-length product and of a series of lower molecular weight bands (**Fig. 5B**, lower panels). These bands are presumably attributable to premature transcription and/or translation termination (and possibly also some low level of proteolysis, although the presence/absence of protease inhibitors did not change the pattern appreciably). Despite their less-than-optimal expression, both GST-MEK1 and GST-MEK2 bound to full-length IQGAP1 (**Fig. 5B**), confirming previous observations (26).

When tested with the IQGAP1 domain-deletion mutants, the three proteins tested (human ERK1, human MEK1, and human MEK2) displayed the same pattern of binding interactions as seen in **Fig. 3** for ERK2: they bound to derivatives containing the IQ domain, including the “IQ-domain only” construct 720–863, but did not bind to derivatives lacking the IQ domain (**Fig. 5B**). This pattern indicates that the IQ domain is necessary and sufficient for the interaction of ERK1, MEK1, and MEK2 with IQGAP1.

4.7 The WW domain does not contribute binding energy to the interaction

The results presented above (Figs. 2–5) clearly demonstrate that the IQ domain is necessary and sufficient for high-affinity binding to ERK2, whereas the WW domain is neither necessary nor sufficient. However, these results do not exclude the possibility that the WW domain of IQGAP1 contributes to the binding energy of the ERK-IQGAP interaction.

To further investigate this question, we constructed a mutant version of IQGAP1(1–863) that contained five different substitution mutations, each of which has been shown to be inactivating in other WW domains; this mutant protein is designated “IQGAP1(1–863)wwmut.” The sequence of the core of the WW domain of human IQGAP1 is shown in **Fig. 6A**, where it is aligned with WW domains from human WWOX1 and human PIN1. A Y33R mutation in WWOX1 was previously shown to abolish its interaction with several ligands (75). At the molecular level, this mutation was interpreted as compromising the “aromatic cradle” structure that is essential for the formation of WW domain-ligand complexes. The first mutation in the IQGAP1(1–863)wwmut is an analogous substitution, Y696R. Jager et al. (76) carried out an extensive substitution analysis of the WW domain of the human PIN1 protein, and identified four positions that partially or completely unfolded the protein when substituted with alanine: Trp11, Tyr24, Asn26, and Pro37. These residues are all highly conserved in the WW domain family, and identical residues

are found in homologous positions in human IQGAP1. IQGAP1(1–863)wmut contains analogous substitutions in each of these four residues: W685A, Y697A, N699A, and P710A.

To summarize, the WW domain of IQGAP1(1–863)wmut contains five different amino acid substitutions, any one of which would be expected to compromise its ability to fold and/or bind ligand. Nevertheless, no obvious difference was seen when this mutant was compared with wild-type IQGAP1(1–863) for its ability to bind to rat ERK2, human ERK2, or human ERK1 (**Fig. 6, B–D**). We conclude that the WW domain contributes little or no binding energy to the ERK-IQGAP1 interaction.

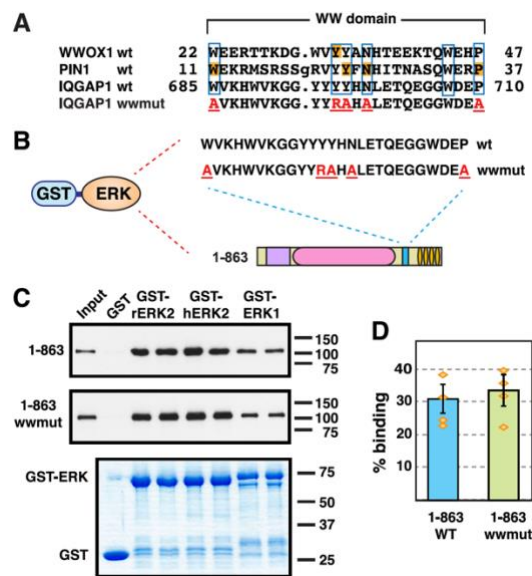


Figure 6. The WW domain does not contribute to the ERK-IQGAP1 interaction. **A**, the top three lines show an alignment of the amino acid sequences of the first WW domain from human WWOX1 (accession number NP_057457, residues shown are 22–47), and the single WW domains in human PIN1 (NP_006212, residues 11–37) and human IQGAP1 (NP_003861, residues 685–710). Residues identical in all three domains are boxed; these include the two tryptophan residues (positions 685 and 707 in IQGAP1) that give the WW domain its name. Residues that were the site of inactivating mutations in other studies (75, 76) are shaded orange. The bottom line shows the sequence of the quintuply-mutated WW domain in the derivative IQGAP1(1–863)wmut; residues mutated to alanine are shown in red and underlined. **B** and **C**, human ERK1, human ERK2, and rat ERK2, fused to GST, were tested for binding to human IQGAP1(1–863) or IQGAP1(1–863)wmut. Other details are as described in the legend to **Fig. 2**. **D**, quantification of the binding of IQGAP1(1–863) or IQGAP1(1–863)wmut to GST-hERK2. The results shown are the average of 4 independent repetitions of the binding assay shown in **B** and **C**, with duplicate points (i.e. technical replicates) in each repetition. S.E. bars are shown (n=4). The scatter of the individual data points is also shown. The ERK2-IQGAP1(1–863) and ERK2-IQGAP1(1–863)wmut interactions were not significantly different from each other (p=0.57, thus the null hypothesis that the population means are the same cannot be rejected with confidence).

Section 5: Discussion

IQGAP proteins are highly studied, evolutionarily conserved scaffold proteins that act as integrators for a number of signaling/regulatory pathways, including the RAS/MAPK pathway. Recently, there has been interest in interdicting the IQGAP-MAPK interaction as a therapeutic strategy in cancer (8, 11, 21). Successful efforts in this direction will be contingent upon accurate information regarding which domains of the IQGAPs bind to the various kinases in the MAPK cascade. For example, the precise identification of the JNK-binding site in the JIP1 scaffold protein led to the development of both peptide and small molecule inhibitors of JNK (77–79).

The core of the RAS/MAPK cascade consists of the MAPK kinases MEK1 and MEK2, which phosphorylate and activate the MAPKs ERK1 and ERK2. Here we investigated the binding of human IQGAP1 to these core kinases, and presented five significant findings.

First and most importantly, we showed that, contrary to what the field has longed believed, the IQ domain of IQGAP1 is both necessary and sufficient for high-affinity binding to the ERK1 and ERK2 MAPKs (**Figs. 2, 3, and 5**). We also showed that the IQ domain of IQGAP1 is necessary and sufficient for binding to the MAPK kinases MEK1 and MEK2 (**Fig. 5**).

In addition, we quantified the strength of the interaction between ERK2 and IQGAP1, determining that the dissociation constant (K_d) for this interaction is about 8 μ M (**Table 1**). Dissociation constants in the low micromolar range have also been observed for other MAPK-scaffold interactions (62, 77, 80).

Further characterizing the ERK2-IQGAP1 interaction, we showed that it was not dependent on the kinase activity of the ERK2, nor on the activation of ERK2 (**Fig. 4**). We also demonstrated that it did not involve known MAPK-docking sites in IQGAP1 or docking-recruitment sites on ERK2 (**Fig. 4**).

Finally, we asked if the WW domain contributed in any significant way to the ERK-IQGAP1 binding interaction. Arguing against this possibility, we found that in constructs lacking the IQ domain, there was only trace binding of WW-containing derivatives to ERK. Furthermore, even this negligible binding was not significantly different from the binding of constructs lacking the WW domain, nor from the binding of either type of construct to GST alone. In other words, in the absence of the IQ domain, there was only minimal background binding in all cases (**Figs. 2, 3, and 5**).

We also sought to address the question of whether or not the WW domain could help the IQ domain to bind to ERK2 via cooperative interactions. To do this, we compared the binding of an IQGAP1 derivative containing an intact WW domain to an otherwise identical construct containing five WW-domain substitution mutations that have been shown (in other WW domains) to be critical for folding and/or

substrate binding. The binding of these two constructs was virtually indistinguishable (**Fig. 6**). Thus, not only is the WW domain neither necessary nor sufficient for the ERK-IQGAP1 (or MEK-IQGAP1) interaction; we could find no evidence that it contributes to these interactions in any way whatsoever.

In this regard, we note that there is nothing in the primary amino acid sequence of ERK1 and ERK2 that would suggest that they are likely to interact canonically with any WW domain. First, neither ERK protein contains the cognate core motif, PPXY, which is dominant among ligands of WW domains. Moreover, neither ERK protein contains a polyproline stretch of any sort; such stretches, with proper flanking residues, can bind to certain types of WW domains (81). Indeed, ERK1 and ERK2 do not even have 2 prolines in a row anywhere in their sequence.

The assertion that the WW domain of IQGAP1 binds to ERK1 and ERK2 originates from Roy et al. (27). Here we used the same proteins as Roy et al. (27) (rat ERK2 and human IQGAP1), and an extremely similar experimental approach (in vitro binding assays with bacterially expressed ERK proteins and in vitro-translated IQGAP1 derivatives), and reached an opposite conclusion.

Although we confirmed the finding of Roy et al. (27) that ERK2 binds to full-length IQGAP1, as well as to IQGAP1(1–863) and IQGAP1(432–863), we cannot readily explain two of their reported results. First, they reported that a mutant designated “ Δ WW” (which is deleted of residues 643–744) did not bind to ERK2. This mutant is missing the entire WW domain (which spans approximately residues 680–710), yet contains essentially all of the IQ domain (which spans approximately residues 744–856). Based on our finding that the IQ domain is necessary and sufficient for ERK binding, we would expect this mutant to bind. However, it is possible that the lack of binding observed by Roy et al. (27) was caused by improper folding of this internally deleted protein.

Roy et al. (27) also reported that a mutant that they designated “ Δ IQ” (which is missing residues 699–905) bound to ERK2 as well as full-length IQGAP1 did (27). Remarkably, this derivative also lacks the C-terminal end of the WW domain, including the second of the two defining tryptophan residues. Thus, the WW domain in this mutant is partially deleted and most likely non-functional. Nevertheless, Roy et al. (27) used the positive binding results obtained with this mutant as part of their argument that the WW domain mediates ERK-IQGAP1 binding. Because this derivative lacks the entire IQ domain, we would expect this mutant not to bind. Possibly, this protein also did not fold properly, and did so in a way that made it nonspecifically sticky. Another possibility is that this particular internal deletion caused a conformational change that unmasked or created a second ERK-binding site elsewhere in the protein.

Jameson et al. (21) showed that an IQGAP-WW domain fragment (consisting residues 680–711 of human IQGAP1, plus short tags) could inhibit RAS- and RAF-driven tumorigenesis, and could bypass acquired resistance to the RAF inhibitor vemurafenib (21). Our results clearly call into question the mechanistic interpretation that these effects were due to the titration of ERK proteins by the WW domain (that is, our results call into question the model shown in Fig. 1C). An obvious alternative hypothesis is that the anti-tumor activity of the WW domain is attributable to its binding to some other ligand(s). However, no other ligands have been identified for the WW domain of IQGAP1, although we note that MAPK-AP2 (a protein kinase regulated by ERK1/2 and p38) was predicted as a WW ligand, as it contains a perfect PPXY motif (82). Given the apparent efficacy of the WW domain of IQGAP1 to inhibit tumor growth and invasiveness, identifying the true ligand(s) of this WW domain should now be prioritized.

Our results suggest a new model of IQGAP scaffolding in which both MEK and ERK bind to the IQ domain, in close proximity for binding sites for RAF (35) and receptor tyrosine kinases (83, 84).

References

1. Dhillon, A. S., Hagan, S., Rath, O., and Kolch, W. (2007) MAP kinase signalling pathways in cancer. *Oncogene* 26, 3279–3290
2. Lawrence, M. C., Jivan, A., Shao, C., Duan, L., Goad, D., Zaganjor, E., Osborne, J., McGlynn, K., Stippec, S., Earnest, S., Chen, W., and Cobb, M. H. (2008) The roles of MAPKs in disease. *Cell Res.* 18, 436–442
3. Caunt, C. J., Sale, M. J., Smith, P. D., and Cook, S. J. (2015) MEK1 and MEK2 inhibitors and cancer therapy: the long and winding road. *Nat. Rev. Cancer* 15, 577–592
4. Lito, P., Rosen, N., and Solit, D. B. (2013) Tumor adaptation and resistance to RAF inhibitors. *Nat. Med.* 19, 1401–1409
5. Dhanasekaran, D. N., Kashef, K., Lee, C. M., Xu, H., and Reddy, E. P. (2007) Scaffold proteins of MAP-kinase modules. *Oncogene* 26, 3185–3202
6. Bhattacharyya, R. P., Reményi, A., Yeh, B. J., and Lim, W. A. (2006) Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu. Rev. Biochem.* 75, 655–680
7. Langeberg, L. K., and Scott, J. D. (2015) Signalling scaffolds and local organization of cellular behaviour. *Nat. Rev. Mol. Cell Biol.* 16, 232–244
8. Stuart, D. D., and Sellers, W. R. (2013) Targeting RAF-MEK-ERK kinase scaffold interactions in cancer. *Nat. Med.* 19, 538–540
9. Hedman, A. C., Smith, J. M., and Sacks, D. B. (2015) The biology of IQGAP proteins: beyond the cytoskeleton. *EMBO Rep.* 16, 427–446
10. Smith, J. M., Hedman, A. C., and Sacks, D. B. (2015) IQGAPs choreograph cellular signaling from the membrane to the nucleus. *Trends Cell Biol.* 25, 171–184
11. Sanchez-Laorden, B., Viros, A., and Marais, R. (2013) Mind the IQGAP. *Cancer Cell* 23, 715–717
12. Brown, M. D., and Sacks, D. B. (2006) IQGAP1 in cellular signaling: bridging the GAP. *Trends Cell Biol.* 16, 242–249
13. White, C. D., Erdemir, H. H., and Sacks, D. B. (2012) IQGAP1 and its binding proteins control diverse biological functions. *Cell Signal.* 24, 826–834
14. Malarkannan, S., Awasthi, A., Rajasekaran, K., Kumar, P., Schuldt, K. M., Bartoszek, A., Manoharan, N., Goldner, N. K., Umhoefer, C. M., and Thakar, M. S. (2012) IQGAP1: a regulator of intracellular spacetime relativity. *J. Immunol.* 188, 2057–2063
15. Abel, A. M., Schuldt, K. M., Rajasekaran, K., Hwang, D., Riese, M. J., Rao, S., Thakar, M. S., and Malarkannan, S. (2015) IQGAP1: insights into the function of a molecular puppeteer. *Mol. Immunol.* 65, 336–349
16. Osman, M. A., Sarkar, F. H., and Rodriguez-Boulan, E. (2013) A molecular rheostat at the interface of cancer and diabetes. *Biochim. Biophys. Acta* 1836, 166–176
17. Wu, Y., and Chen, Y. C. (2014) Structure and function of IQ-domain GTPase-activating protein 1 and its association with tumor progression (review). *Biomed. Rep.* 2, 3–6
18. Johnson, M., Sharma, M., and Henderson, B. R. (2009) IQGAP1 regulation and roles in cancer. *Cell Signal.* 21, 1471–1478
19. Weissbach, L., Settleman, J., Kalady, M. F., Snijders, A. J., Murthy, A. E., Yan, Y. X., and Bernards, A. (1994) Identification of a human rasGAP related protein containing calmodulin-binding motifs. *J. Biol. Chem.* 269, 20517–20521
20. White, C. D., Brown, M. D., and Sacks, D. B. (2009) IQGAPs in cancer: a family of scaffold proteins underlying tumorigenesis. *FEBS Lett.* 583, 1817–1824

21. Jameson, K. L., Mazur, P. K., Zehnder, A. M., Zhang, J., Zarnegar, B., Sage, J., and Khavari, P. A. (2013) IQGAP1 scaffold-kinase interaction blockade selectively targets RAS-MAP kinase-driven tumors. *Nat. Med.* 19, 626–630
22. Ho, Y. D., Joyal, J. L., Li, Z., and Sacks, D. B. (1999) IQGAP1 integrates Ca²/calmodulin and Cdc42 signaling. *J. Biol. Chem.* 274, 464–470
23. Osman, M. A., and Cerione, R. A. (1998) Iqg1p, a yeast homologue of the mammalian IQGAPs, mediates cdc42p effects on the actin cytoskeleton. *J. Cell Biol.* 142, 443–455
24. Briggs, M. W., and Sacks, D. B. (2003) IQGAP proteins are integral components of cytoskeletal regulation. *EMBO Rep.* 4, 571–574
25. Carmon, K. S., Gong, X., Yi, J., Thomas, A., and Liu, Q. (2014) RSPO-LGR4 functions via IQGAP1 to potentiate Wnt signaling. *Proc. Natl. Acad. Sci. U.S.A.* 111, E1221–1229
26. Roy, M., Li, Z., and Sacks, D. B. (2005) IQGAP1 is a scaffold for mitogen activated protein kinase signaling. *Mol. Cell. Biol.* 25, 7940–7952
27. Roy, M., Li, Z., and Sacks, D. B. (2004) IQGAP1 binds ERK2 and modulates its activity. *J. Biol. Chem.* 279, 17329–17337
28. Macias, M. J., Wiesner, S., and Sudol, M. (2002) WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS Lett.* 513, 30–37
29. Lu, P. J., Zhou, X. Z., Shen, M., and Lu, K. P. (1999) Function of WW domains as phosphoserine- or phosphothreonine-binding modules. *Science* 283, 1325–1328
30. Bedford, M. T., Chan, D. C., and Leder, P. (1997) FBP WW domains and the Abl SH3 domain bind to a specific class of proline-rich ligands. *EMBO J.* 16, 2376–2383
31. Chen, H. I., and Sudol, M. (1995) The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules. *Proc. Natl. Acad. Sci. U.S.A.* 92, 7819–7823
32. Hu, H., Columbus, J., Zhang, Y., Wu, D., Lian, L., Yang, S., Goodwin, J., Luczak, C., Carter, M., Chen, L., James, M., Davis, R., Sudol, M., Rodwell, J., and Herrero, J. J. (2004) A map of WW domain family interactions. *Proteomics* 4, 643–655
33. Otte, L., Wiedemann, U., Schlegel, B., Pires, J. R., Beyermann, M., Schmieler, P., Krause, G., Volkmer-Engert, R., Schneider-Mergener, J., and Oschkinat, H. (2003) WW domain sequence activity relationships identified using ligand recognition propensities of 42 WW domains. *Protein Sci.* 12, 491–500
34. Sudol, M., Chen, H. I., Bougeret, C., Einbond, A., and Bork, P. (1995) Characterization of a novel protein-binding module: the WW domain. *FEBS Lett.* 369, 67–71
35. Ren, J. G., Li, Z., and Sacks, D. B. (2007) IQGAP1 modulates activation of B-Raf. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10465–10469
36. Bähler, M., and Rhoads, A. (2002) Calmodulin signaling via the IQ motif. *FEBS Lett.* 513, 107–113
37. Tekletsadik, Y. K., Sonn, R., and Osman, M. A. (2012) A conserved role of IQGAP1 in regulating TOR complex 1. *J. Cell Sci.* 125, 2041–2052
38. Ang, B. K., Lim, C. Y., Koh, S. S., Sivakumar, N., Taib, S., Lim, K. B., Ahmed, S., Rajagopal, G., and Ong, S. H. (2007) ArhGAP9, a novel MAP kinase docking protein, inhibits Erk and p38 activation through WW domain binding. *J. Mol. Signal.* 2, 1
39. Heil, A., Nazmi, A. R., Koltzsch, M., Poeter, M., Austermann, J., Assard, N., Baudier, J., Kaibuchi, K., and Gerke, V. (2011) S100P is a novel interaction partner and regulator of IQGAP1. *J. Biol. Chem.* 286, 7227–7238
40. Gao, C., Frausto, S. F., Guedea, A. L., Tronson, N. C., Jovasevic, V., Leaderbrand, K., Corcoran, K. A., Guzmán, Y. F., Swanson, G. T., and Radulovic, J. (2011) IQGAP1 regulates NR2A signaling, spine density, and cognitive processes. *J. Neurosci.* 31, 8533–8542

41. Shannon, K. B. (2012) IQGAP family members in yeast, Dictyostelium, and mammalian cells. *Int. J. Cell Biol.* 2012, 894817
42. Kim, H., White, C. D., Li, Z., and Sacks, D. B. (2011) Salmonella enterica serotype Typhimurium usurps the scaffold protein IQGAP1 to manipulate Rac1 and MAPK signalling. *Biochem. J.* 440, 309–318
43. Gorman, J. A., Babich, A., Dick, C. J., Schoon, R. A., Koenig, A., Gomez, T. S., Burkhardt, J. K., and Billadeau, D. D. (2012) The cytoskeletal adaptor protein IQGAP1 regulates TCR-mediated signaling and filamentous actin dynamics. *J. Immunol.* 188, 6135–6144
44. Rigotherier, C., Auguste, P., Welsh, G. I., Lepreux, S., Deminière, C., Mathieson, P. W., Saleem, M. A., Ripoche, J., and Combe, C. (2012) IQGAP1 interacts with components of the slit diaphragm complex in podocytes and is involved in podocyte migration and permeability in vitro. *PLoS ONE* 7, e37695
45. Goto, T., Sato, A., Shimizu, M., Adachi, S., Satoh, K., Iemura, S., Natsume, T., and Shibuya, H. (2013) IQGAP1 functions as a modulator of disheveled nuclear localization in Wnt signaling. *PLoS ONE* 8, e60865
46. Cheung, K. L., Lee, J. H., Shu, L., Kim, J. H., Sacks, D. B., and Kong, A. N. (2013) The Ras GTPase-activating-like protein IQGAP1 mediates Nrf2 protein activation via the mitogen-activated protein kinase/extracellular signal-regulated kinase (ERK) kinase (MEK)-ERK pathway. *J. Biol. Chem.* 288, 22378–22386
47. Wallrabe, H., Cai, Y., Sun, Y., Periasamy, A., Luzes, R., Fang, X., Kan, H. M., Cameron, L. C., Schafer, D. A., and Bloom, G. S. (2013) IQGAP1 interactome analysis by in vitro reconstitution and live cell 3-color FRET microscopy. *Cytoskeleton* 70, 819–836
48. Goto, T., Sato, A., Adachi, S., Iemura, S., Natsume, T., and Shibuya, H. (2013) IQGAP1 protein regulates nuclear localization of β -catenin via importin- β 5 protein in Wnt signaling. *J. Biol. Chem.* 288, 36351–36360
49. Liu, J., Guidry, J. J., and Worthylake, D. K. (2014) Conserved sequence repeats of IQGAP1 mediate binding to Ezrin. *J. Proteome Res.* 13, 1156–1166
50. Lu, R., Herrera, B. B., Eshleman, H. D., Fu, Y., Bloom, A., Li, Z., Sacks, D. B., and Goldberg, M. B. (2015) Shigella effector OspB activates mTORC1 in a manner that depends on IQGAP1 and promotes cell proliferation. *PLoS Pathog.* 11, e1005200
51. Struhl, K. (1991) Reverse biochemistry: methods and applications for synthesizing yeast proteins in vitro. *Methods Enzymol.* 194, 520–535
52. Lee, T., Hoofnagle, A. N., Kabuyama, Y., Stroud, J., Min, X., Goldsmith, E. J., Chen, L., Resing, K. A., and Ahn, N. G. (2004) Docking motif interactions in MAP kinases revealed by hydrogen exchange mass spectrometry. *Mol. Cell* 14, 43–55
53. Canagarajah, B. J., Khokhlatchev, A., Cobb, M. H., and Goldsmith, E. J. (1997) Activation mechanism of the MAPkinase ERK2 by dual phosphorylation. *Cell* 90, 859–869
54. Prowse, C. N., Deal, M. S., and Lew, J. (2001) The complete pathway for catalytic activation of the mitogen-activated protein kinase, ERK2. *J. Biol. Chem.* 276, 40817–40823
55. Seger, R., Ahn, N. G., Boulton, T. G., Yancopoulos, G. D., Panayotatos, N., Radziejewska, E., Ericsson, L., Bratlien, R. L., Cobb, M. H., and Krebs, E. G. (1991) Microtubule-associated protein 2 kinases, ERK1 and ERK2, undergo autophosphorylation on both tyrosine and threonine residues: implications for their mechanism of activation. *Proc. Natl. Acad. Sci. U.S.A.* 88, 6142–6146
56. Cohen-Armon, M., Visochek, L., Rozensal, D., Kalal, A., Geistrikh, I., Klein, R., Bendetz-Nezer, S., Yao, Z., and Seger, R. (2007) DNA-independent PARP-1 activation by phosphorylated ERK2 increases Elk1 activity: a link to histone acetylation. *Mol. Cell* 25, 297–308
57. Robinson, M. J., Harkins, P. C., Zhang, J., Baer, R., Haycock, J. W., Cobb, M. H., and Goldsmith, E. J. (1996) Mutation of position 52 in ERK2 creates a nonproductive binding mode for adenosine 5-triphosphate. *Biochemistry* 35, 5641–5646
58. Bardwell, L., and Thorner, J. (1996) A conserved motif at the amino termini of MEKs might mediate high-affinity interaction with the cognate MAPKs. *Trends Biochem. Sci.* 21, 373–374

59. Bardwell, L., Cook, J. G., Chang, E. C., Cairns, B. R., and Thorner, J. (1996) Signaling in the yeast pheromone response pathway: specific and high affinity interaction of the mitogen-activated protein (MAP) kinases Kss1 and Fus3 with the upstream MAP kinase kinase Ste7. *Mol. Cell. Biol.* 16, 3637–3650
60. Kallunki, T., Deng, T., Hibi, M., and Karin, M. (1996) c-Jun can recruit JNK to phosphorylate dimerization partners via specific docking interactions. *Cell* 87, 929–939
61. Yang, S. H., Yates, P. R., Whitmarsh, A. J., Davis, R. J., and Sharrocks, A. D. (1998) The Elk-1 ETS-domain transcription factor contains a mitogen activated protein kinase targeting motif. *Mol. Cell. Biol.* 18, 710–720
62. Kusari, A. B., Molina, D. M., Sabbagh, W., Jr, Lau, C. S., and Bardwell, L. (2004) A conserved protein interaction network involving the yeast MAP kinases Fus3 and Kss1. *J. Cell Biol.* 164, 267–277
63. Kelkar, N., Gupta, S., Dickens, M., and Davis, R. J. (2000) Interaction of a mitogen-activated protein kinase signaling module with the neuronal protein JIP3. *Mol. Cell. Biol.* 20, 1030–1043
64. Dickens, M., Rogers, J. S., Cavanagh, J., Raitano, A., Xia, Z., Halpern, J. R., Greenberg, M. E., Sawyers, C. L., and Davis, R. J. (1997) A cytoplasmic inhibitor of the JNK signal transduction pathway. *Science* 277, 693–696
65. Bardwell, L., and Shah, K. (2006) Analysis of mitogen-activated protein kinase activation and interactions with regulators and substrates. *Methods* 40, 213–223
66. Cargnello, M., and Roux, P. P. (2011) Activation and function of the MAPKs and their substrates, the MAPK-activated protein kinases. *Microbiol. Mol. Biol. Rev.* 75, 50–83
67. Roskoski, R., Jr. (2012) ERK1/2 MAP kinases: structure, function, and regulation. *Pharmacol. Res.* 66, 105–143
68. Peti, W., and Page, R. (2013) Molecular basis of MAP kinase regulation. *Protein Sci.* 22, 1698–1710
69. Fang, L., Lu, W., Choi, H. H., Yeung, S. C., Tung, J. Y., Hsiao, C. D., Fuentes-Mattei, E., Menter, D., Chen, C., Wang, L., Wang, J., and Lee, M. H. (2015) ERK2-dependent phosphorylation of CSN6 is critical in colorectal cancer development. *Cancer Cell* 28, 183–197
70. Chang, C. I., Xu, B. E., Akella, R., Cobb, M. H., and Goldsmith, E. J. (2002) Crystal structures of MAP kinase p38 complexed to the docking sites on its nuclear substrate MEF2A and activator MKK3b. *Mol. Cell* 9, 1241–1249
71. Molina, D. M., Grewal, S., and Bardwell, L. (2005) Characterization of an ERK-binding domain in microphthalmia-associated transcription factor and differential inhibition of ERK2-mediated substrate phosphorylation. *J. Biol. Chem.* 280, 42051–42060
72. Tanoue, T., Adachi, M., Moriguchi, T., and Nishida, E. (2000) A conserved docking motif in MAP kinases common to substrates, activators and regulators. *Nat. Cell. Biol.* 2, 110–116
73. Jacobs, D., Glossip, D., Xing, H., Muslin, A. J., and Kornfeld, K. (1999) Multiple docking sites on substrate proteins form a modular system that mediates recognition by ERK MAP kinase. *Genes Dev.* 13, 163–175
74. Piserchio, A., Ramakrishnan, V., Wang, H., Kaoud, T. S., Arshava, B., Dutta, K., Dalby, K. N., and Ghose, R. (2015) Structural and dynamic features of F-recruitment site driven substrate phosphorylation by ERK2. *Sci. Rep.* 5, 11127
75. Aqeilan, R. I., Donati, V., Palamarchuk, A., Trapasso, F., Kaou, M., Pekarsky, Y., Sudol, M., and Croce, C. M. (2005) WW domain-containing proteins, WWOX and YAP, compete for interaction with ErbB-4 and modulate its transcriptional function. *Cancer Res.* 65, 6764–6772
76. Jäger, M., Dendle, M., and Kelly, J. W. (2009) Sequence determinants of thermodynamic stability in a WW domain: an all- β -sheet protein. *Protein Sci.* 18, 1806–1813
77. Barr, R. K., Boehm, I., Attwood, P. V., Watt, P. M., and Bogoyevitch, M. A. (2004) The critical features and the mechanism of inhibition of a kinase interaction motif-based peptide inhibitor of JNK. *J. Biol. Chem.* 279, 36327–36338

78. Ho, D. T., Bardwell, A. J., Abdollahi, M., and Bardwell, L. (2003) A docking site in MKK4 mediates high affinity binding to JNK MAPKs and competes with similar docking sites in JNK substrates. *J. Biol. Chem.* 278, 32662–32672
79. Stebbins, J. L., De, S. K., Machleidt, T., Becattini, B., Vazquez, J., Kuntzen, C., Chen, L. H., Cellitti, J. F., Riel-Mehan, M., Emdadi, A., Solinas, G., Karin, M., and Pellecchia, M. (2008) Identification of a new JNK inhibitor targeting the JNK-JIP interaction site. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16809–16813
80. Bhattacharyya, R. P., Reményi, A., Good, M. C., Bashor, C. J., Falick, A. M., and Lim, W. A. (2006) The Ste5 scaffold allosterically modulates signaling output of the yeast mating pathway. *Science* 311, 822–826
81. Chen, H. I., Einbond, A., Kwak, S. J., Linn, H., Koepf, E., Peterson, S., Kelly, J. W., and Sudol, M. (1997) Characterization of the WW domain of human yes-associated protein and its polyproline-containing ligands. *J. Biol. Chem.* 272, 17070–17077
82. Einbond, A., and Sudol, M. (1996) Towards prediction of cognate complexes between the WW domain and proline-rich ligands. *FEBS Lett.* 384, 1–8
83. McNulty, D. E., Li, Z., White, C. D., Sacks, D. B., and Annan, R. S. (2011) MAPK scaffold IQGAP1 binds the EGF receptor and modulates its activation. *J. Biol. Chem.* 286, 15010–15021
84. Bañón-Rodríguez, I., Gálvez-Santisteban, M., Vergarajauregui, S., Bosch, M., Borreguero-Pascual, A., and Martín-Belmonte, F. (2014) EGFR controls IQGAP basolateral membrane localization and mitotic spindle orientation during epithelial morphogenesis. *EMBO J.* 33, 129–145
85. Bardwell, A. J., Flatauer, L. J., Matsukuma, K., Thorner, J., and Bardwell, L. (2001) A conserved docking site in MEKs mediates high-affinity binding to MAP kinases and cooperates with a scaffold protein to enhance signal transmission. *J. Biol. Chem.* 276, 10374–10386
86. Gordon, E. A., Whisenant, T. C., Zeller, M., Kaake, R. M., Gordon, W. M., Krotee, P., Patel, V., Huang, L., Baldi, P., and Bardwell, L. (2013) Combining docking site and phosphosite predictions to find new substrates: identification of smoothelin-like-2 (SMTNL2) as a c-Jun N-terminal kinase (JNK) substrate. *Cell Signal.* 25, 2518–2529
87. Ruxton, G. D. (2006) The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test. *Behav. Ecol.* 17, 688–690

Chapter 3: Effect of magnitude and variability of energy of activation in multisite ultrasensitive biochemical processes

Section 1: Abstract

Protein activity is often regulated by ligand binding or by post-translational modifications such as phosphorylation. Moreover, proteins that are regulated in this way often contain multiple ligand binding sites or modification sites, which can operate to create an ultrasensitive dose response. Here, we consider the contribution of the individual modification/binding sites to the activation process, and how their individual values affect the ultrasensitive behavior of the overall system. We use a generalized Monod-Wyman-Changeux (MWC) model that allows for variable conformational free energy contributions from distinct sites, and associate a so-called activation parameter to each site. Our analysis shows that the ultrasensitivity generally increases as the conformational free energy contribution from one or more sites is strengthened. Furthermore, ultrasensitivity depends on the mean of the activation parameters and not on their variability. In some cases, we find that the best way to maximize ultrasensitivity is to make the contribution from all sites as strong as possible. These results provide insights into the performance objectives of multiple modification/binding sites and thus help gain a greater understanding of signaling and its role in diseases.

Section 2: Introduction

Cellular systems rely heavily on signal transduction and environmental sensing pathways to successfully respond to internal and external environmental signals and conditions. To regulate signal transduction cascades, mammalian cells use ligand binding or post-translational modifications (PTMs) such as protein phosphorylation, methylation or ubiquitination. Several forms of disease can arise when there are defects in signal transduction pathways, including cancer, diabetes, and heart disease.

Many proteins regulated by ligands or PTMs are multisite proteins, that is, they have multiple sites on which they can be modified or where a ligand can bind. For example, activation of mitogen activated protein kinases requires phosphorylation on two sites [1], and the hemoglobin tetramer has four sites where oxygen can bind ([2] and references therein). In fact, some proteins have more than 150 modification sites [3]. Ligand binding/PTMs can either promote or inhibit protein activity through conformational changes [4, 5], and can influence the target's enzymatic activity, location, stability, or interactions with other macromolecules [6].

A common role for multisite modifications lies in the creation of switch-like, or ultrasensitive, dose response curves [7–9]. These are positive, monotonically increasing, sigmoidal functions that have two important properties: first, they respond minimally to low levels of input; second, once the input is sufficiently large, they switch from a low output to a near maximal output in response to a relatively small increase in the input. In other words, ultrasensitive systems can both filter out low-level noise and respond with a high gain over an appropriate range of

input. Ultrasensitivity has important roles in signal transduction, and a widely-studied problem is how to implement such dose responses using common biochemical reactions [2].

A classical 1965 model by Monod, Wyman and Changeux (MWC) [10] uses multisite modifications to create ultrasensitive responses. This model remains highly influential today [11–14]. In the MWC model, the target molecule/receptor can be in either an active (relaxed) conformation or an inactive (tense) conformation, and ligand binding/PTMs can influence the probability that the target is in one state or the other. One way to envision this is that ligand binding promotes a conformational change that flips the target from inactive to active (or from active to inactive in the case of an inhibitory ligand). In the MWC model, this is equivalent to the point of view that the ligand binds preferentially to the active conformation, a phenomenon known as conformational selection. In multisite MWC models, there are multiple binding sites for ligand, each of which can be either empty or bound. Such models exhibit cooperativity in ligand binding, as the binding of some ligands to the target will promote flipping to the active state, and in the active state, all binding sites have a higher affinity for ligand. In other words, the presence of ligand increases the probability of the receptor existing in the state with higher ligand affinity, thereby increasing the probability of the next ligand binding. In addition to ‘cooperativity’, the term ‘allostery’ is frequently used in conjunction with MWC models, and refers to the effect that one ligand binding to the target has on additional “distant” binding sites in the same molecule, as well as to the effect that ligand binding has on the conformational change that activates the target. The concepts of ultrasensitivity, allostery and cooperativity are important not only in understanding the logic of cellular regulation, but also with regard to disease pathology and drug discovery [15].

Classical mathematical models of allostery and cooperative ligand binding, such as the MWC model, were based on observations of cooperativity between symmetric subunits of oligomeric proteins, such as hemoglobin (a tetramer), threonine deaminase (also a tetramer) and aspartate transcarbamylase (a hexamer) [16]. Given that the molecules under study consisted of multiple identical or very similar subunits, it made sense to treat all binding sites as identical. More recently, however, the concept of allostery has been expanded to include monomeric proteins, where binding of a ligand at one site can result in modulation of function or binding at a (perhaps) distant site in the same polypeptide chain [17]. For instance, binding or modification events occurring in an intrinsically-disordered segment of a protein can promote its folding, and this can be communicated to an adjacent segment, with the net effect that a coupled folding-and-binding event or PTM in one region of the protein influences subsequent interactions or modifications at a distant site(s) within the same monomer [18]. Yet another example is hetero-oligomers that display cooperativity such as the ATPase rings in the proteasome and CCT chaperonin complex [19–22]. In such cases, there is no reason to expect that binding/modification sites will be identical, or that they will make identical contributions to the underlying conformational change once bound/modified.

In the current paper, we set out to explore multisite systems in which the modification of some sites may have a

stronger effect on the induced conformational change than the modification of other sites. To do this, we generalize the classical MWC system and assign different parameters to different sites. We aim to determine what combinations of parameters lead to a high level of ultrasensitivity. Each site i is assigned an activation parameter c_i , generalizing the parameter c in the original formulation of the MWC model. Small values of the parameter c_i correspond to a strong ability for the i -th site to activate the protein. One can also associate to each site a corresponding conformational free energy contribution, that is, the difference in the Gibbs energy function associated to the site i , $\Delta G_i = rt \ln(c_i)$. In other words, $rt \ln(c_i)$ is the site-specific free energy contribution to tense-to-relaxed flipping from ligand binding at site i . Notice that the conformational free energy contribution ΔG_i is negative number when the activation parameter c_i is less than one, and becomes more negative as c_i approaches 0. Also note that a large negative ΔG_i (and hence a small c_i) corresponds to a strong conformational free energy contribution, which will promote flipping to the active state. In contrast, if $c_i > 1$, then ΔG_i will be positive, meaning that the modification does not promote flipping to the relaxed/active state but instead makes it more likely that the target will stay in the tense state.

Our main results can be summarized as follows. First, making the conformational free energy contribution associated with ligand binding to a single site i more favorable (that is, making this free energy change more negative, which is equivalent to making the activation parameter c_i smaller) has a strong tendency to increase the ultrasensitivity of the system, as measured by its associated Hill coefficient H . This effect is not guaranteed as there are some exceptions, especially for low values of the number of sites n , but it holds in most circumstances and under several orders of magnitude for the parameters in the system.

Second, for a fixed number of sites n , each of which is at least moderately active, one can calculate the average of the activation parameters and get a good approximation of the Hill coefficient of the system by assuming that all sites have this average activation parameter. That is, the Hill coefficient is approximately independent on the variability of the parameters c_i , only on their mean value.

Third, we find that when the cost of site maintenance is taken into account, one can obtain an optimal ultrasensitive behavior by focusing on a subset of the sites. The strategy is to have a subset of the sites be equally active, and all other sites have a low or negligible conformational free energy contribution. This prediction has been indeed observed in a number of experimental systems, where only a subset of the sites have the ability to activate the protein. In addition, we demonstrate that there are diminishing marginal ultrasensitivity increases in response to conformational free energy contribution improvements, which allows us to predict a maximal effective conformational free energy contribution per site, on the order of -2 to -4 kcal/mol. This prediction follows from first principles of the mathematical model, and it is surprisingly consistent with experimental data for a typical protein phosphorylation site [23–26].

For completeness, the last sections contain a study of ultrasensitive behavior in a non-allosteric multisite model

where all sites are independent from each other, applicable in some cases where the MWC allosteric assumptions are not satisfied. It was found that this system has a more complex relation between activation parameters and ultrasensitivity, which was explored both through computations and mathematical analysis.

Section 3: Derivation of the MWC Dose Response

In this section, we carry out a generalization of the MWC model to account for different activation parameters at distinct sites. See Enciso and Ryerson [8], where a similar generalization was carried out for protein modification efficiencies.

Consider a substrate with n sites in modified-form I , in one of two states, relaxed (R_I) or tense (T_I), where $I \in \{0, 1\}^n$ is a binary vector representing the modified-form of the substrate. In the case of protein activation models, relaxed and tense states correspond to different levels of activity. We assume that all modified-forms in the relaxed (R) conformation are active and have the same activity, whereas all modified-forms in the tense (T) conformation are inactive, and have the same (low) activity. Under MWC assumptions, the relaxed state has a higher affinity to the ligand than the tense state, this is assumed here for the states R_I and T_I . The unmodified state, $I = \vec{0} = (0, 0, \dots, 0)$, and the fully-modified state, $I = \vec{1} = (1, 1, \dots, 1)$, are the two extreme modified-forms, and a total of 2^n modified-forms are possible. The modification of site i on the substrate will result in modified-form J , where $J = I \cup \{i\}$. In other words J is the modified-form consisting of adding one more modification at site i to the modified-form I . For example, in Fig. 1a, a two-site substrate can be in the relaxed state with no modifications $R_{(0,0)}$ and be reversibly modified to $R_{(1,0)}$ or $R_{(0,1)}$ and subsequently to $R_{(1,1)}$. A substrate in a relaxed state can also flip to the tense state in that form. For instance, $R_{(1,1)}$ can flip to the $T_{(1,1)}$ state. Similarly, the tense substrate in modified-form $T_{(1,0)}$ can be reversibly modified to $T_{(1,1)}$. We call u the kinase concentration in the case of multisite phosphorylation and u denotes the ligand concentration in the case of ligand binding.

The general system can be described by the chemical reaction network in Fig. 1b. The parameter α_i is the microscopic association constant for ligand binding at site i . Note that α_i is an affinity, i.e. not a dissociation constant but the inverse thereof. If instead of ligand binding the protein is modified by post-translational modification (phosphorylation, acetylation, etc.), α_i represents the modification efficiency of site i , which, for example, will be determined by the relative suitability of the site to be phosphorylated by a given kinase and dephosphorylated by a given phosphatase. The parameter L is the equilibrium constant between $R_{\vec{0}}$ and $T_{\vec{0}}$. This network has the property of detailed balance, as the product of the equilibrium constants around any closed cycle of states is 1; this is the same as saying that the net free energy change around any closed cycle of states is 0. For this reason each forward and reverse reaction pair is in equilibrium.

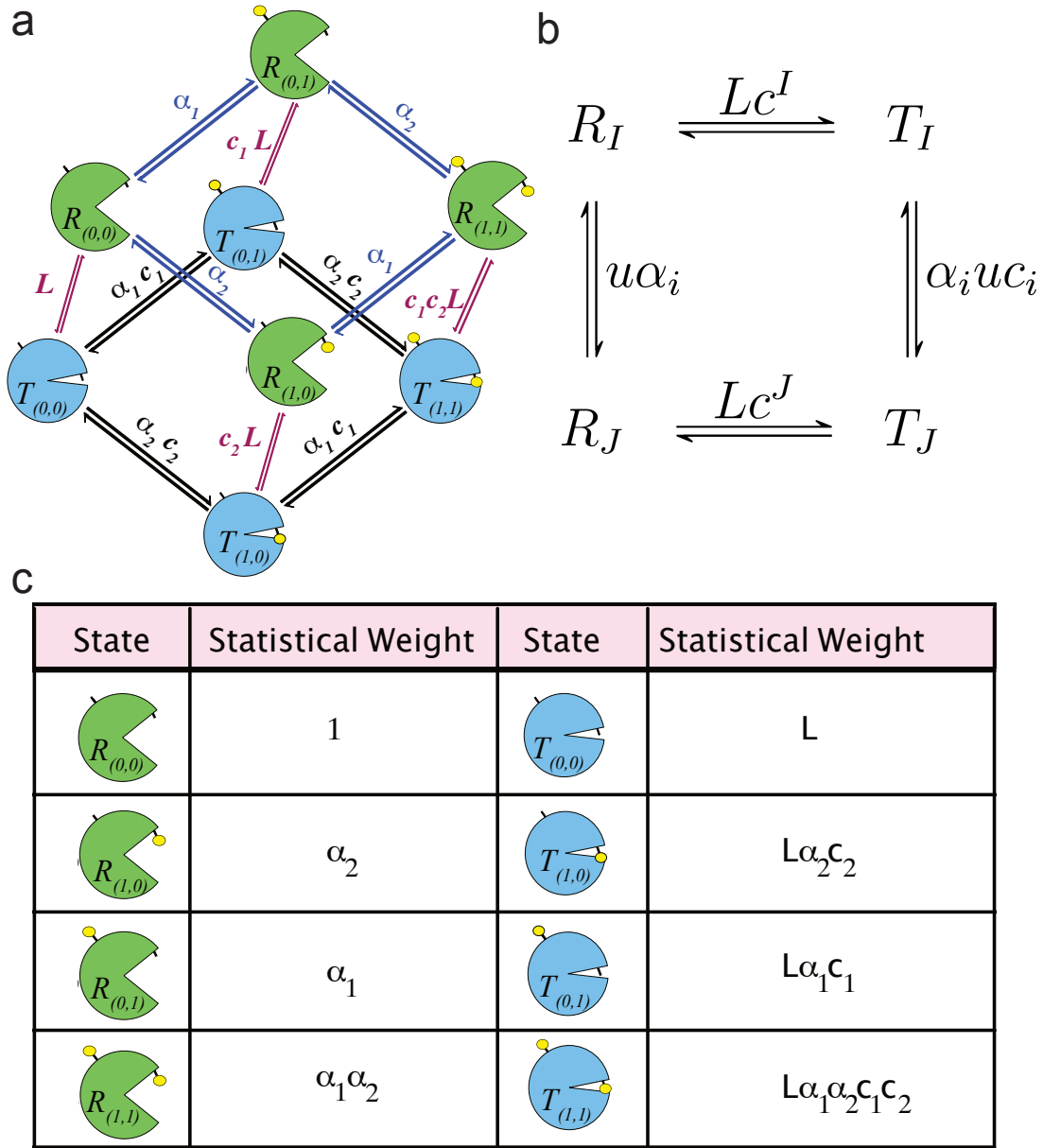


Figure 1: Generalized MWC Allosteric Model. (a) The figure shows the eight possible states of a target molecule/receptor regulated by the MWC mechanism and containing two sites for ligand binding or post translational modification (i.e., $n = 2$). The four states shown in blue (with closed ‘mouth’) are the tense, inactive states, while the four states shown in green (with open ‘mouth’) are the relaxed, active states. Modification/ligand binding is indicated by the presence of absence of a small yellow ball. The L , α and c parameters are explained in the text. (b) Chemical reaction network demonstrating the possible modified-forms of a receptor with n sites, where I is the index vector for the modified-form and J is the index vector after adding one more modification at site i . (c) Table of statistical weights for each state possible with $n = 2$.

We use the notation $c^I = \prod_{i=1} c_i$. Notice that Lc^I is the equilibrium constant between the relaxed (active) state R_I and the tense (inactive) state T_I . In this sense, one can think of c_i as the contribution to this equilibrium

constant made by each individual site i , and ΔG_i as the free energy differential between active and inactive protein contributed by modification at site i .

The statistical weights for each state possible when $n = 2$ is listed in Fig. 1c. The probability of a state, say $R_{(1,0)}$ is defined as the ratio between the statistical weight and the partition function Z . In the $n = 2$ case, $Z = 1 + \alpha_1 + \alpha_2 + \alpha_1\alpha_2 + L + L\alpha_1c_1 + L\alpha_2c_2 + L\alpha_1\alpha_2c_1c_2$. Fig. S1 shows a table of the statistical weights of each modification state when $n = 3$ and the associated Z . For general n , $Z = 1 + L + \sum_{i=1}^n (\alpha_i + L\alpha_i c_i)$.

We can write the dose response for total substrate as a function of enzyme u . Following a similar analysis to that of Enciso & Ryerson [8], since the system is in detailed balance, for every index I ,

$$R_I u \alpha_i = R_J \text{ and } R_I L c^I = T_I$$

Solving for R_I and T_I , we can relate R_I to $R_{\bar{0}}$ (relaxed protein with no modifications) by induction as:

$$R_I = u^{|I|} R_{\bar{0}} \alpha^I \text{ and } T_I = u^{|I|} R_{\bar{0}} \alpha^I L c^I.$$

Note that

$$\begin{aligned} \sum_I u^{|I|} &= \sum_{i=0}^n \binom{n}{i} u^i = (u+1)^n, \\ \sum_I c^I u^{|I|} &= \sum_{i=0}^n \sum_{|I|=i} c^I u^{|I|} = \sum_{i=0}^n u^i \sum_{|I|=i} c^I = \sum_{i=0}^n u^i \rho_i(c) = \sum_{i=0}^n \rho_i(uc) = \prod_{j=1}^n (uc_j + 1), \\ \sum_I u^{|I|} \alpha^I &= \prod_{i=1}^n (u\alpha_i + 1) \text{ and similarly, } \sum_I u^{|I|} \alpha^I c^I = \prod_{i=1}^n (uc_i \alpha_i + 1). \end{aligned}$$

Here, $\rho_i(c) = \sum_{|I|=i} c^I$ is the symmetric polynomial in i with entries c [8]. For example, consider $n = 2$ with $c = (c_1, c_2)$. Here, $\rho_2(c) = c_1 c_2$, $\rho_1(c) = c_1 + c_2$, and $\rho_0(c) = 1$. At various points we are able to rewrite a sum into a product, using the principle that if x_i is a constant for $i = 1, 2, \dots, n$, then $\sum_I x^I = \prod_{j=1}^n (x_j + 1)$. For general n , the above allows us to write

$$\begin{aligned} S_T &= \sum_I R_I + T_I = R_{\bar{0}} \sum_I u^{|I|} \alpha^I + L R_{\bar{0}} \sum_I c^I u^{|I|} \alpha^I = R_{\bar{0}} \prod_{i=1}^n (u\alpha_i + 1) + L R_{\bar{0}} \prod_{j=1}^n (uc_j \alpha_j + 1), \\ R_{\bar{0}} &= S_T \frac{1}{\prod_{i=1}^n (u\alpha_i + 1) + L \prod_{i=1}^n (uc_i \alpha_i + 1)}. \end{aligned}$$

The response of this system is given by the total concentration of relaxed protein, regardless of its level of modifications.

That is,

$$f(u, c, \alpha) = \sum_I R_I = R_0 \sum_I u^{|I|} \alpha^I = R_0 \prod_{i=1}^n (u\alpha_i + 1) = \frac{S_T \prod_{i=1}^n (u\alpha_i + 1)}{\prod_{i=1}^n (u\alpha_i + 1) + L \prod_{i=1}^n (uc_i\alpha_i + 1)}.$$

That is

$$f(u, c, \alpha) = \frac{S_T}{1 + L \prod_{i=1}^n \frac{uc_i\alpha_i + 1}{u\alpha_i + 1}} = S_T \omega(\eta(u, c, \alpha)), \quad (1)$$

where $\eta(u, c, \alpha) = \prod_{i=1}^n \frac{uc_i\alpha_i + 1}{u\alpha_i + 1}$ and $\omega(x) = \frac{1}{1 + Lx}$. Notice from this functional form when any c_i is equal to 1, it simply multiplies the dose response by one and becomes the same as a system with $n - 1$ sites. This is a nontrivial comment which is not obvious from the system otherwise, but it is biologically intuitive. If a substrate has weak sites, they only contribute weakly or not at all to increase the Hill coefficient. For fixed parameter values c and α , we define the maximal response $f^\infty(c) = \lim_{u \rightarrow \infty} f(u, c, \alpha)$. A simple calculation shows that $f^\infty(c) = \frac{1}{1 + Lc_1c_2 \dots c_n}$ and depends only on c_1, c_2, \dots, c_n and is independent of α . This maximal output value will allow us to normalize response curves across different parameter values in the sections below.

Since the effect of varying the modification parameters α_i was extensively described in Enciso and Ryerson [8], in the majority of the discussion here, we will assume that the α_i are equal to each other, and in fact we can set $\alpha_i = 1$. To see this, one can re-scale u by defining $\bar{u} = u\bar{\alpha}$. The new dose response has the same Hill coefficient as the old system, however the new system satisfies $\alpha_i = 1$ for all i . This helps to better understand the effect of individual activation parameters.

Section 4: Computational Results on MWC Ultrasensitivity

Recall from the previous section that $f(u, c, \alpha)$ represents the dose response for the generalized MWC system with parameters c_i and α_i , for $i = 1, 2, \dots, n$, where n is the number of sites, and u is the ligand/enzyme concentration.

In this section we carry out a computational analysis of the dose response and its Hill coefficient, where the parameters c_i are sampled logarithmically. More specifically, $\log(c_i)$ is chosen with uniform distribution between $[10^{-4}, 0.9]$. The parameter $L \geq 1$ was fixed, and the parameters α_i were chosen to be identical to each other, $\alpha_i = \bar{\alpha}$, here the value of $\bar{\alpha}$ does not affect the Hill coefficient.

We calculated the Hill coefficient H by solving for EC_{90} and EC_{10} with a standard numerical solver. Here, we solved for u such that $f(u, c, \alpha) - \beta f^\infty(c) = 0$ for both $\beta = 10\%$ and 90% . With both EC_{10} and EC_{90} , we can

calculate H as

$$H = \frac{\ln(81)}{\ln\left(\frac{EC_{90}}{EC_{10}}\right)} \quad (2)$$

derived in [27]. $H > 1$ implies the dose response curve is ultrasensitive, while $H = 1$ implies there is no ultrasensitivity, and $H < 1$ shows negative ultrasensitivity. One can also think of $H > 1$ showing that the dose response has a good switch [28]; the larger the value of H the more ultrasensitive the dose response curve.

Fig. 2a displays the dose response curves in this system for $n = 2, 4, 8$, $c_i = 0.01$, $L = 1000$, and $\alpha_i = \bar{\alpha} = 1$. These functions show that when all the sites contribute equally, the Hill coefficient tends to increase with the number of sites. In Figs. 2b-d, c_1 and c_2 were increased from 10^{-4} to 0.9 and each $c_i = 0.01$ for $i \geq 3$, $\alpha_i = 1$, and $L = 1000$. In these figures, H decreases for increasing c_1 only, suggesting that H increases with increasing conformational free energy (recall that larger values of c correspond to lower activation contributions). Note that for the $n = 2$ case, there are cases for large values of c_1 where the Hill number is undefined. Fig. S2 shows similar data on a linear scale.

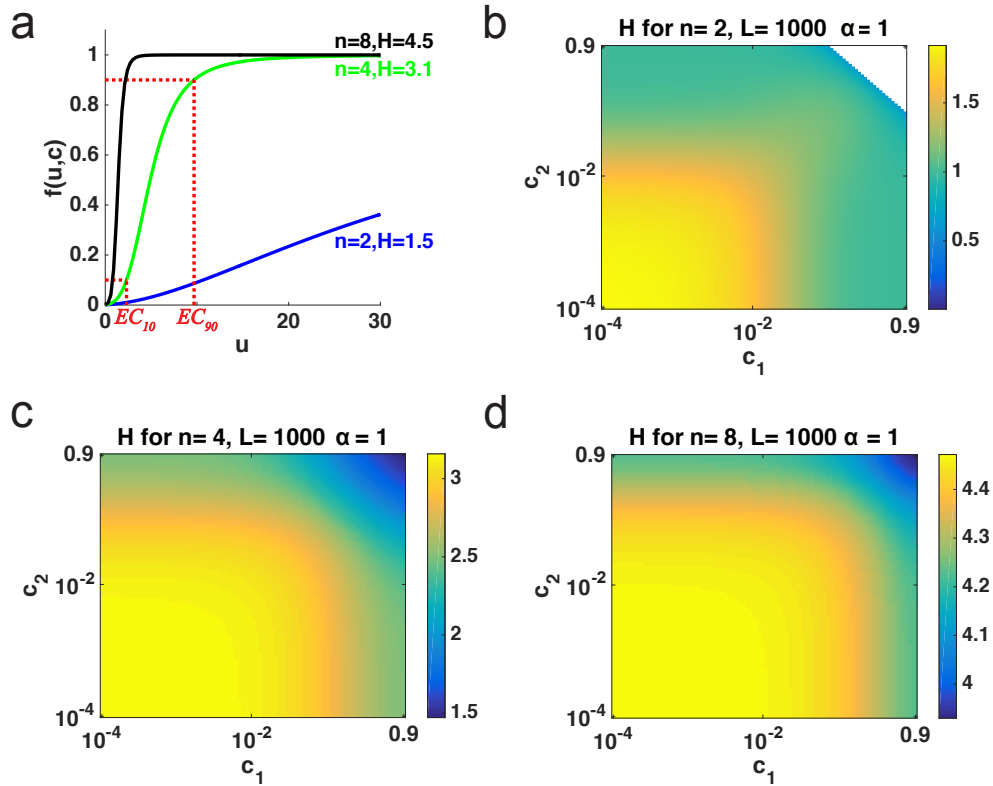


Figure 2: Ultrasensitivity of MWC system.. (a) Dose response curve, $f(u, c, \alpha)$, when $n = 2, 4$, and 8 for increasing u with $c_i = 0.01$, $\alpha_i = \bar{\alpha} = 1$ and $L = 1000$. (b-d) Heat maps for H when $c_1, c_2 \in [10^{-4}, 0.9]$ with $L = 1000$ and $\alpha_i = \bar{\alpha} = 1$ and (d) $n = 2$, (e) $n = 4$, $c_i = 0.01$ for $i \geq 3$, similarly with (f) $n = 8$. White indicates undefined H values.

In Fig. S3e, we show a Monte Carlo approach to study whether H is always a decreasing function of c_i . By

symmetry, we take any individual c_i parameter to be c_1 , without loss of generality. To find the proportion of cases where H decreases on c_1 , we find a numerical approximation to $H_{c_1}(c)$ as follows:

$$H_{c_1}(c) = \frac{\partial H(c)}{\partial c_1} \approx \frac{H(c_1 + \Delta x, c_2, \dots, c_n) - H(c_1, c_2, \dots, c_n)}{\Delta x},$$

for small Δx to determine if $\frac{\partial H(c)}{\partial c_1} < 0$. Here, for different values of L and n with $\alpha_i = \bar{\alpha} = 1$, for 1000 simulations, we sampled $c_i \in [10^{-4}, 0.9]$ logarithmically for $i = 1, 2, \dots, n$. The proportion of simulations where H decreases on c_1 is almost always one for $n > 4$. For $n = 2$ there are many parameter sets where that is not the case.

In Fig. 3 we further analyze the effect of varying the activation parameters on the Hill coefficient. In Fig. 3a, we sampled a vector $c \in \mathbb{R}^n$ with entries in the interval $[10^{-4}, 0.9]$ logarithmically. This vector of activation parameters has arithmetic mean \bar{c} and coefficient of variation $CV(c) = \frac{\sigma(c)}{\bar{c}}$. To each vector c one can assign a second vector $\hat{c} = (\bar{c}, \bar{c}, \dots, \bar{c})$ for which $CV(\hat{c}) = 0$. After calculating H for each case, we can see in Fig. 3a, that when there is no variation between c_i (solid line), with parameter values $\alpha_i = 1$ and $L = 1000$, H decreases with increasing mean of c for $n = 2, 3, 4, 6$ and 8 . Any variation among the c_i does not significantly affect H when $\bar{c} < 10^{-2}$. For larger values of \bar{c} , H depends on the variability among the c_i as well as their mean. When α_i is sampled from the range $[0.1, 10]$ the dependence of H on \bar{c} is less clearly defined, compared to the case $\alpha_i = 1$ (see supplemental Figure S3a-b).

In Fig. 3b, for the same parameter values c_i and $\alpha_i = 1$ we plot H against the total conformational free energy ΔG_{tot} defined as

$$\Delta G_{tot} = rt \ln \left(\prod_{i=1}^n c_i \right) \quad (3)$$

where r is the gas constant and t is the temperature (traditionally R and T but labeled as r and t , respectively, to maintain consistent notation and not be mistaken for the MWC tense and relaxed states). For fixed n , as the total conformational free energy increases, ultrasensitivity generally tends to increase, and after some threshold, it tends to level off. To increase ultrasensitivity at that point, a substrate cannot profitably utilize more conformational free energy, but instead must evolve more sites. As a concrete example, let us consider an MWC molecule with two sites under selective pressure to increase its ultrasensitivity. Changes to the microscopic modification affinities/efficiencies (i.e., the ΔG_i 's) will either decrease ultrasensitivity (if the changes are unbalanced), or at best leave ultrasensitivity unaltered (if the changes are balanced) [8]. Thus, the only viable options to increase ultrasensitivity are to (a) evolve another site, or (b) strengthen the conformational free energy of the existing sites. At first, significant increases to ultrasensitivity can result from the second option. A mutation that strengthens the conformational free energy of one of the sites will move the molecule up and to the right in the cloud of points for $n = 2$ shown in Fig. 3b, with the largest jump coming from strengthening the weakest site. As additional mutations of this type arise and

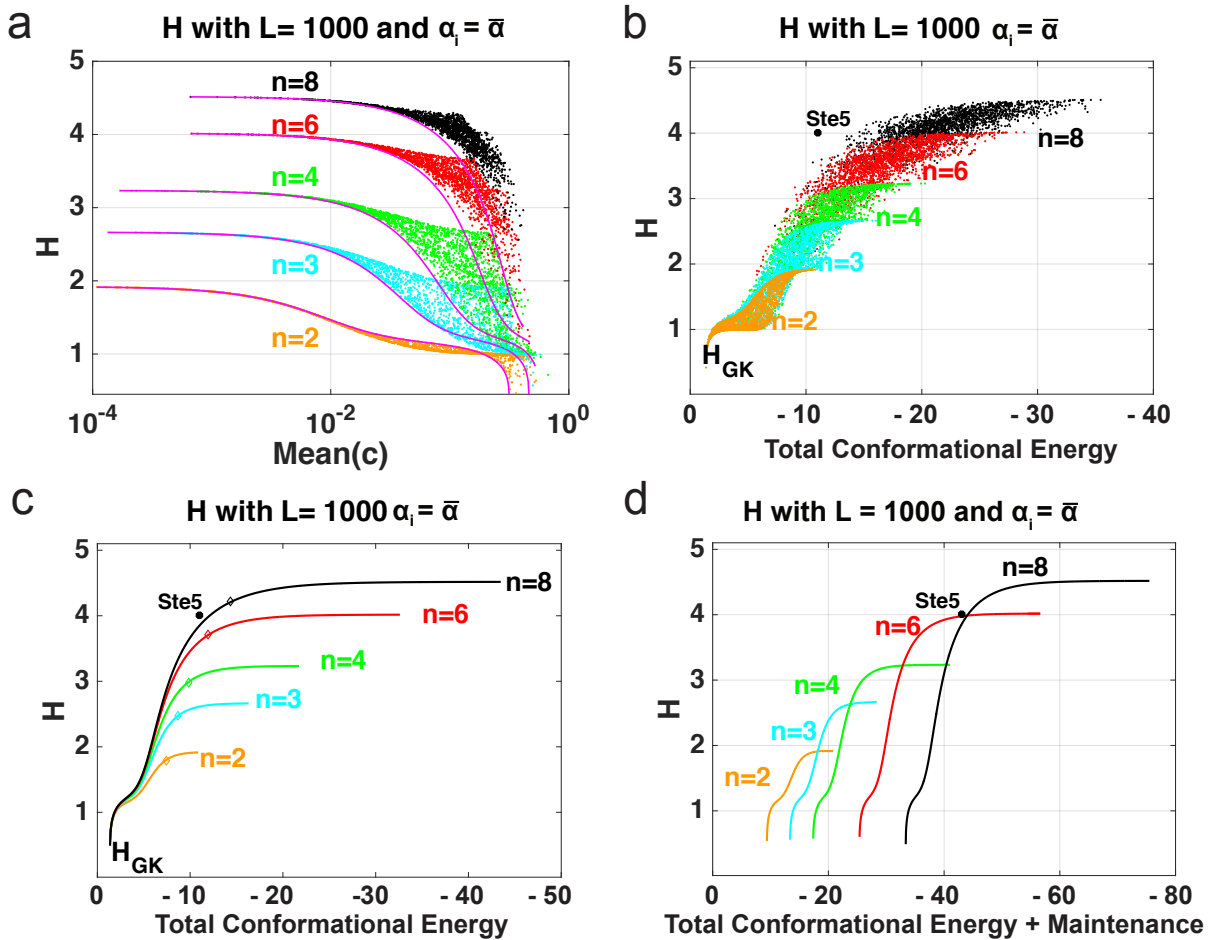


Figure 3: Activation Parameters and H in MWC. (a) Scatter plot for H when c_i are independently and logarithmically chosen from the interval $[10^{-4}, 0.9]$ (dots) and when c_i are all identical (solid line), $L = 1000$, and $\alpha_i = \bar{\alpha}$ for $n = 2, 3, 4, 6, 8$. (b) Scatter plot for H for increasing total conformational free energy (Eq. 3) when $c_i = c \in [10^{-4}, 0.9]$, $\alpha_i = \bar{\alpha}$, and $L = 1000$ for $n = 2, 3, 4, 6, 8$. Asterisk is the approximated H for Ste5 from [24]. (c) Scatter plot for H for increasing total conformational free energy (Eq. 3) when $c_i = \bar{c} \in [10^{-4}, 0.9]$, $\alpha_i = \bar{\alpha}$, and $L = 1000$ for $n = 2, 3, 4, 6, 8$. Diamonds represent the knee of the curve. (d) H for increasing total conformational free energy where $c_i = \bar{c} \in [10^{-4}, 0.9]$, $\alpha_i = \bar{\alpha}$ and $L = 1000$ and with a maintenance cost of 4 kcal/mol per site where $n = 2, 3, 4, 6, 8$. The Ste5 data point is added for illustration purposes with the same maintenance cost for each of the 8 phosphorylation sites.

become fixed by natural selection, the molecule will move to the top right of the cloud; here the conformational free energies will be approximately balanced and have magnitudes of approximately -2 to -4 kcal/mol. At this point, substantial improvement to ultrasensitivity (i.e., an increase of the Hill number by greater than 0.5 units) can only arise if the molecule evolves an additional site.

To view this more clearly, consider Fig. 3c, which shows ultrasensitivity for increasing values of total conforma-

tional free energy where parameter values have been set to $c_i = \bar{c}$ and $\alpha_i = \bar{\alpha} = 1$ and fixed n and L . In other words, when each c_i is the same, meaning the conformational free energy is the same in all sites, we can see that ultrasensitivity generally increases and eventually levels off as the conformational free energy increases. Fig. 3c also helps to make a prediction of the energy that each site optimally contributes, given a total conformational free energy.

Calculating the "knee" of the curve provides a rough estimate of the total conformational free energy required before the ultrasensitivity begins to level off. The knee of a saturating curve is a mathematical definition that captures the point at which the curve is reaching saturation. It is defined in our context as $\max_{a \leq x \leq b} |Y_n(x) - \ell(x)|$, where $Y_n(x)$ is the Hill coefficient curve in Fig. 3c for n sites, a, b are the lowest and highest total energy values among the data points for n sites respectively, and $\ell(x)$ is a secant line joining $(a, Y_n(a))$ and $(b, Y_n(b))$. For a more detailed explanation, refer to Figure 2b in Ref [29].

The approximated knee of each curve in Fig. 3c was found and is depicted with a diamond and listed in Table 1. Notice that the energy for saturation increases roughly linearly with the number of sites. In each case, the amount of energy per site is approximately -2 to -4 kcal/mol.

This analysis is consistent with some previous experimental findings [23, 24]. The approximated H for Ste5 from [24] with $n = 8$ phosphorylation sites is indicated with an asterisk in Fig. 3b and 3c. We derive the value -1.6 kcal/mol per site in this system, which is equivalent to a 10-fold affinity increase per site as approximated in [24]. Not only does this data point approximately lie close to the curve for $n = 8$, but in fact it lies close to the knee of the curve when all sites contribute an equal amount, as predicted in the above analysis. The marginal effect of an additional kcal of free energy is dependent on only one other parameter, namely L (assuming the α_i are roughly equal to each other). If L ranges from 30 to 10,000, the analysis is roughly similar (see Fig. S3c-d), and it leads to an energy range of around -2 to -4 kcal/mol per site (see also Table S1).

There are ways of evaluating ultrasensitivity other than the Goldbeter-Koshland method [27]. In Fig. S4, we measure ultrasensitivity in two additional ways: (1) fitting the dose response curve to the Hill function [9] and (2) Levitzki's n_{50} [30]. We see similar results, thus the qualitative results here do not depend on how ultrasensitivity is measured.

In Fig. 3d, similar to Fig. 3b, we plot H for increasing values of total conformational free energy where now we take into account a maintenance cost for each site, denoted by M_c . Such a maintenance cost may arise, for example, if there is rapid turnover of a post-translational modification, as has been observed for phosphorylation-dephosphorylation of some substrates [31, 32]. This type of rapid dynamics in modification-demodification cycles could constitute a non-negligible expenditure of energy for the cell.

The total activation energy including maintenance can be calculated as $\Delta G_{tot} + M_c * n$, where ΔG_{tot} is given

Table 1: Ultrasensitivity at knee				
L	n	H	total cfe	cfe/site
1000	2	1.80	-7.37	-3.69
	3	2.47	-8.67	-2.89
	4	2.99	-9.80	-2.45
	6	3.71	-11.97	-1.99
	8	4.21	-14.39	-1.80

Table 1: Ultrasensitivity as measured by the Goldbeter-Koshland formula described in Eqn. (2) along with the approximated knee of the curves in Fig. 3c for fixed values of L and n . The knee of curve occurs at a single value of total conformational free energy (cfe) with a Hill number H . Parameters $\alpha_i = \bar{\alpha} = 1$.

by (Eq. 3) and $M_c = 4$ kcal/mol, which was arbitrarily chosen. In this figure, we assume for simplicity that energy is equally distributed among all sites. Once the cost of maintenance is taken into account, one can see more clearly that for each level of total conformational free energy there is an optimal value of n . For instance, if the total energy is -20 kcal/mol, then the optimal number of sites is $n = 3$; any fewer sites will not have as high ultrasensitivity, while any larger number of sites requires an excessive amount of maintenance. If there are more than four sites in this system, it is beneficial to eliminate or silence the remaining sites. Similar qualitative results can be seen in Fig. S5 with different maintenance cost values.

To summarize, in this section we have shown that (1) increasing the conformational free energy at a single site has a strong tendency to increase the ultrasensitivity of the response, with some exceptions, (2) for fixed n , the ultrasensitivity depends on the mean of the free energies of activation and very little on their variance, and (3) we estimate from first principles an effective energy range of -2 to -4 kcal/mol per site, which is consistent with experimental data.

Section 5: Derivation of Independent Dose Response

The assumption of cooperativity between sites plays a role in the ultrasensitive behavior of the dose response curves. However, if we do not assume cooperativity between sites, will we observe the same effect in the previous section on H ? In this section, we use a non-allosteric model and carry out a similar study as for the generalized MWC model. The proposed model has been used elsewhere [8, 33] but here it is generalized for the first time to have different activation parameters at different sites.

Consider a substrate with n modification sites in modified-form $I \in \{0, 1\}^n$, where we no longer assume that there is cooperativity between sites. The substrate can be in one of two states, A_I (active) or B_I (inactive) and thus

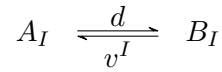
gives 2^n possible modified-forms.

The substrate S in modified-form I can be described by the chemical reaction in Fig. 4a, where $v_i > 1$ represents the conformational free energy of the i -th modification site. Each v_i can also be related with the binding energy of the i -th modification site in the MWC model through the formula

$$\Delta G_i = -rt \ln(v_i).$$

That is, the larger the value of v_i , the larger the free energy. The parameter v_i can also be thought of as the inverse of $v_i = \frac{1}{c_i}$ of the activation parameter in the MWC model. For notation purposes, $v^I = \prod_{I_i=1} v_i$.

To obtain the dose response for total substrate S_T as a function of enzyme concentration u , we use mass action kinetics on the chemical reaction below



As in the MWC model, d is the reaction rate constant and is analogous to L . The associated differential equation for the active substrate is:

$$\frac{dA_I}{dt} = v^I B_I - dA_I, \quad (4)$$

with conservation of mass equation for the substrate in modified-form I :

$$S_I = A_I + B_I.$$

We allow this reaction to reach equilibrium by assuming that this activation/deactivation reaction is much faster than protein modification. This is a reasonable assumption in the case of protein phosphorylation. Solving for steady state of (4),

$$\begin{aligned} 0 &= v^I(S_I - A_I) - dA_I = v^I S_I - A_I(v^I + d), \text{ that is} \\ A_I(v^I + d) &= v^I S_I \text{ and} \\ A_I &= \frac{v^I}{v^I + d} S_I \end{aligned}$$

In order to calculate the activity level of a substrate in modified-form I , we defined the function $Q_I(v)$, which can be considered to be the fraction of time a protein is active, as

$$A_I = \frac{v^I}{d + v^I} S_I, \quad \frac{A_I}{S_I} = \frac{v^I}{d + v^I}.$$

Hence,

$$Q_I(v) = \frac{v^I}{d + v^I}. \quad (5)$$

To further understand $Q_I(v)$, consider the case when $n = 2$. If both sites are modified, then $Q_{(1,1)} = \frac{v_1 v_2}{d + v_1 v_2} \approx 1$ for large values of v_i relative to d . If neither site is modified, then $Q_{(0,0)} = \frac{1}{d + 1}$. In other words, the activity of a protein increases with the amount of modifications. Notice that the activity level will depend on the activation parameters of the specific sites and the overall number of sites modified.

We can also determine the substrate concentration of S_I , as a function of enzyme concentration u . We can accomplish this by first considering the fraction of substrate, p_i , that is modified on the i -th site, at steady state. Then, from [8], given k_i is the disassociation rate constant of the i -th site,

$$p_i(u) = \frac{u}{u + k_i}. \quad (6)$$

We assume that the modification states of the different sites are independent of each other, an assumption that is in a sense the opposite of cooperativity. In other words, the modification of one site does not influence the modification of another. This allows to calculate the proportion of substrate in state I as

$$\frac{S_I}{S_T} = \prod_{I_i=1} p_i \prod_{I_i=0} (1 - p_i) = \prod_{I_i=1} \frac{u}{u + k_i} \prod_{I_i=0} \left(1 - \frac{u}{k_i + u}\right) = \prod_{I_i=1} \frac{u}{u + k_i} \prod_{I_i=0} \frac{k_i}{k_i + u},$$

where S_T is the total substrate. The dose response is calculated as follows:

$$f(u, v) = \sum_I Q_I S_I = \sum_I \frac{v^I}{d + v^I} \prod_{I_i=1} \frac{u}{u + k_i} \prod_{I_i=0} \frac{k_i}{k_i + u}. \quad (7)$$

This function has a maximal output value $f^\infty(v)$, which is found in a similar fashion to the MWC maximal output value by evaluating the limit of $f(u, v)$ as $u \rightarrow \infty$. Note that for any I containing a zero (i.e., any modified-form with at least one site un-modified), $\lim_{u \rightarrow \infty} S_I = 0$. Only $I = \vec{1} = (1, 1, \dots, 1)$ will contain a non-zero limit for S_I , giving $f^\infty(v) = Q_{\vec{1}}$. This maximal output value can be used to normalize the dose response curves across different parameter values, similar to the MWC system.

Section 6: Computational Results on Independent System Ultrasensitivity

For multisite proteins, we can determine the proportion of active substrate by calculating $f(u, v)$ from the independent system above. In Fig. 4b, we plot dose response functions for $n = 2, 4$, and 8 with $v_i = 100$, $k_i = 1$ and $d = 1000$.

Similar to the MWC section above, we show how H is affected by the activation parameters of individual sites. In Fig 4c-e, we measure H in a similar fashion to that in Fig. 3, by solving for EC_{10} and EC_{90} given the dose response $f(u, v)$ in (7). Here, $d = 1000$ and $k_i = 1$ and parameters v_1 and v_2 were sampled with values in $[1, 10^8]$ logarithmically and each $v_i = 100$ for $i \geq 3$ for $n = 2, 4$, and 8 . This implies that H does not increase monotonically with increasing v_i , and there is a local minimum for low values of v_i .

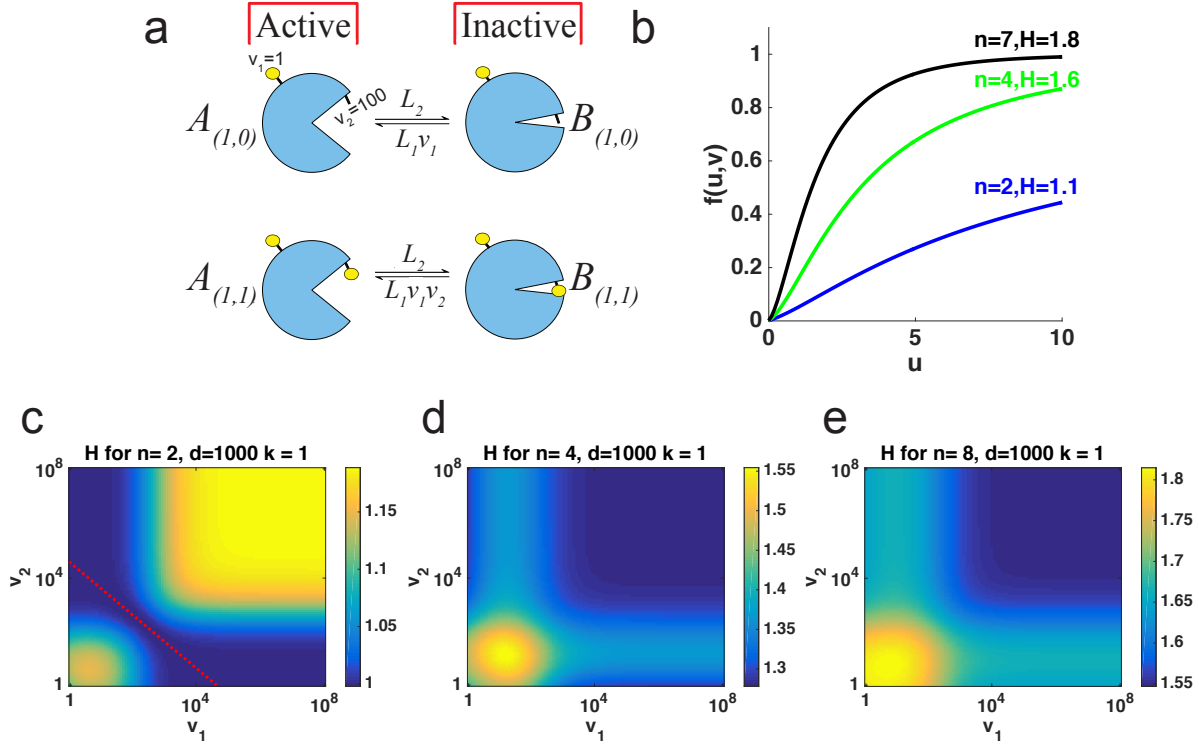


Figure 4: Independent Multisite Modification Model (a) Substrate in modified-form I can be in the inactive state B_I or active A_I state. (b) Dose response curve, $f(u, v)$ when $n = 2, 5, 7$ for increasing kinase concentration u with $v_i = 100$ and $d = 1000$ (c-e) Heat maps for H when $v_1, v_2 \in [10, 10^8]$ with $d = 1000$, and (c) $n = 2$, dashed line used to denote region where $d = \sqrt{v_1 v_2}$. (d) $n = 4$, $v_i = 100$ for all $i \geq 3$, similarly with (e) $n = 7$.

To determine the effect the variability between parameters v_i has on H , we varied parameters v_i , measured H and compared to when all parameters v_i are equal, similar to the MWC system. In Fig. 5a, we sampled a vector with entries from $[1, 10^4]$, logarithmically. This v has an arithmetic mean \bar{v} and coefficient of variation, $CV = \frac{sd(v)}{\bar{v}}$. For each sample, there is a second vector, $\hat{v} = (\bar{v}, \bar{v}, \dots, \bar{v})$ such that $CV(\hat{v}) = 0$. After calculating H for each case, we can see in Fig. 5a, that when there is no variation between v_i (solid line), $k_i = 1$ and $d = 1000$, H can increase or decrease depending on the mean of v for $n = 2, 3, 4$ and 8 . Here, we can see that any variation between the v_i may affect H (asterisks).

In Fig. 5b, we show the same data from Fig. 5a but plotting $CV(v)$ vs H . Here we see that $CV(v)$ has some effect on H , regardless of n . This is particularly interesting since, contrary to MWC, the variation between v_i affects H . We also see that there are values of \bar{v} where H increases and values where it decreases. How often is H increasing with increasing v_1 ?

In Fig. 5c, similar to Supplemental Fig. S3 e, we provide the proportion of simulations where H increases with

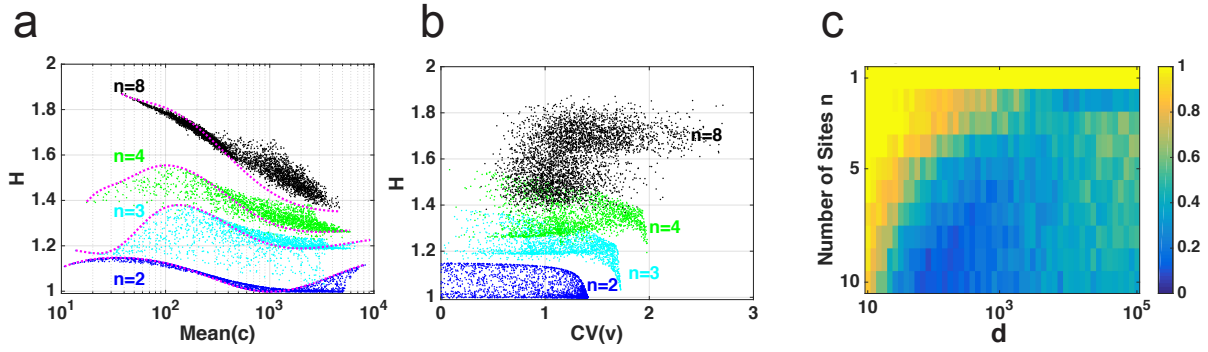


Figure 5: Activation Parameters and H in Independent Model. (a) Scatter plot for H when $v_i \in [10, 10^4]$ (asterisks) and when $v_i = \bar{v}$ for $i = 1, 2, \dots, n$ (solid curve) for different values of n . (b) H vs $\text{CV}(v)$ for different values of n . (c) Heat map showing the proportion of times H increased with increasing v_i as a function of n and d for $v_i \in [10, 10^4]$ and $k_i = 1$ with 1000 simulations of randomly chosen $v_i \in [10, 10^4]$.

increasing v_1 based on n and d . The proportion was found in a similar fashion as in the MWC system. Here, we logarithmically sampled $v_i \in [10, 10^4]$ and $k_i = 1$.

The computational and analytical results described in the section below titled “Independent System Mathematical Analysis” suggest that that $d > \sqrt{v_1 v_2}$ is a biologically reasonable assumption that will give dose response functions where the effect of two modifications is significantly different than the effect of a single modification. Similarly, $d < \sqrt{v_1 v_2}$ gives dose response functions where the effect of a single modification has a similar effect as two modifications, termed “1+” regime. In this “1+” regime we see H increasing on v . When $d = \sqrt{v_1 v_2}$, we have a dose response function where the effect of one modification is approximately 50% of the effect of two modifications, with no ultrasensitivity ($H \approx 1$). We can also see that if d is slightly past the 50% of max activation, H can be maximized by increasing the free activation of energy v .

To summarize, in this section we show that (1) ultrasensitivity increases under specific parameter regimes and (2) may depend on the variability between the activation parameters.

Section 7: MWC System Mathematical Analysis

In this section, we provide a mathematical analysis of the generalized MWC system showing that $H(c, \alpha)$ is roughly independent of the variation of c . We will show that H is essentially a function of \bar{c} and $\bar{\alpha}$. That is, the variability of activation parameters only affects H to the extent that it changes the mean, \bar{c} . Consider $f(u, c, \alpha)$ from Eq. (1) and define

$$\bar{\alpha} = \frac{\alpha_1 + \alpha_2 + \cdots + \alpha_n}{n}, \Delta\alpha = (\alpha_1 - \bar{\alpha}, \alpha_2 - \bar{\alpha}, \cdots, \alpha_n - \bar{\alpha}),$$

$$\bar{c} = \frac{c_1 + c_2 + \cdots + c_n}{n}, \Delta c = (c_1 - \bar{c}, c_2 - \bar{c}, \cdots, c_n - \bar{c}).$$

For notation purposes, let $\hat{c} = (\bar{c}, \bar{c}, \cdots, \bar{c}) \in \mathbb{R}^n$ and $\hat{\alpha} = (\bar{\alpha}, \bar{\alpha}, \cdots, \bar{\alpha}) \in \mathbb{R}^n$. Recall that given a C^2 function f such that $\frac{\partial f}{\partial x}(a, b) = \frac{\partial f}{\partial y}(a, b) = 0$, it holds that

$$f(x, y) = f(a, b) + o(x - a, y - b).$$

We will use this to show that

$$H(c, \alpha) = H(\hat{c}, \hat{\alpha}) + o(\Delta c, \Delta\alpha).$$

This formula demonstrates in particular that H essentially does not vary if the mean of c is preserved, as illustrated in Fig. 3a.

Proposition 1. $H(c, \hat{\alpha}) = H(\hat{c}, \hat{\alpha}) + o(\Delta c)$

Proof. For simplicity, assume $S_T = 1$ and assume for now that u and $\hat{\alpha}$ are fixed. By the approximation of the geometric mean using the arithmetic mean, we have

$$\left(\prod_{i=1}^n uc_i\bar{\alpha} + 1 \right)^{1/n} = \frac{1}{n} \sum_{i=1}^n (uc_i\bar{\alpha} + 1) + o(\Delta c) = (u\bar{c}\bar{\alpha} + 1) + o(\Delta c).$$

Taking the n -th power, $\prod_{i=1}^n (uc_i\bar{\alpha} + 1) = (u\bar{c}\bar{\alpha} + 1)^n + o(\Delta c)$. Let $M > 0$ such that $|\omega(x) - \omega(y)| \leq M|x - y|$ for all $x, y > 0$. Then,

$$\begin{aligned} |f(u, c, \hat{\alpha}) - f(u, \hat{c}, \hat{\alpha})| &\leq M|\eta(u, c, \hat{\alpha}) - \eta(u, \hat{c}, \hat{\alpha})| \\ &= M \left| \prod_{i=1}^n \frac{uc_i\bar{\alpha} + 1}{u\bar{\alpha} + 1} - \prod_{i=1}^n \frac{u\bar{c}\bar{\alpha} + 1}{u\bar{\alpha} + 1} \right| \\ &= \frac{M}{(u\bar{\alpha} + 1)^n} \left| \prod_{i=1}^n (uc_i\bar{\alpha} + 1) - (u\bar{c}\bar{\alpha} + 1)^n \right| = o(\Delta c). \end{aligned}$$

Then $f(u, c, \hat{\alpha}) = f(u, \hat{c}, \hat{\alpha}) + o(\Delta c)$. It follows that

$$EC_{10}(c, \hat{\alpha}) = EC_{10}(\hat{c}, \hat{\alpha}) + o(\Delta c) \text{ and } EC_{90}(c, \hat{\alpha}) = EC_{90}(\hat{c}, \hat{\alpha}) + o(\Delta c).$$

Thus,

$$H(c, \hat{\alpha}) = \frac{\ln(81)}{\ln\left(\frac{EC_{90}(c, \hat{\alpha})}{EC_{10}(c, \hat{\alpha})}\right)} = \frac{\ln(81)}{\ln\left(\frac{EC_{90}(\hat{c}, \hat{\alpha})}{EC_{10}(\hat{c}, \hat{\alpha})}\right)} + o(\Delta c) = H(\hat{c}, \hat{\alpha}) + o(\Delta c).$$

One can assume that $\hat{\alpha}$ doesn't change since H is unaffected by increasing or decreasing its value, as explained in the sections above. Also, the above analysis is carried out for u in a neighborhood of $EC_{10}(\hat{c}, \hat{\alpha})$ and $EC_{90}(\hat{c}, \hat{\alpha})$, hence one can assume that u does not vary significantly. ■

Proposition 2. $H(\hat{c}, \alpha) = H(\hat{c}, \hat{\alpha}) + o(\Delta \alpha)$.

Proof. Similar to Proposition 1,

$$\left(\prod_{i=1}^n (u\bar{c}\alpha_i + 1)\right)^{1/n} = \sum_{i=1}^n \frac{u\bar{c}\alpha_i + 1}{n} + o(\Delta \alpha) = u\hat{c}\hat{\alpha} + o(\Delta \alpha).$$

Taking the n -th power, $\prod_{i=1}^n (u\bar{c}\alpha_i + 1) = (u\bar{c}\hat{\alpha} + 1)^n + o(\Delta \alpha)$. In particular, for $\bar{c} = 1$, $\prod_{i=1}^n (u\alpha_i + 1) = (u\hat{\alpha} + 1)^n + o(\Delta \alpha)$. Therefore, $\eta(u, \hat{c}, \alpha) = \eta(u, \hat{c}, \hat{\alpha}) + o(\Delta \alpha)$. In the same way as in Proposition 1, $H(\hat{c}, \alpha) = H(\hat{c}, \hat{\alpha}) + o(\Delta \alpha)$. ■

Proposition 3. $H(c, \alpha) = H(\hat{c}, \hat{\alpha}) + o(\Delta c, \Delta \alpha)$.

Proof. The first-order Taylor approximation of $H(c, \alpha)$ around $(\hat{c}, \hat{\alpha})$ can be written as

$$\begin{aligned} H(c, \alpha) &= H(\hat{c}, \hat{\alpha}) + \sum_{i=1}^n (c_i - \bar{c}) \frac{\partial}{\partial c_i} H(\hat{c}, \hat{\alpha}) + \sum_{i=1}^n (\alpha_i - \bar{\alpha}) \frac{\partial}{\partial \alpha_i} H(\hat{c}, \hat{\alpha}) + o(\Delta c, \Delta \alpha) \\ &= H(\hat{c}, \hat{\alpha}) + \Delta c \cdot \nabla_c H(\hat{c}, \hat{\alpha}) + \Delta \alpha \cdot \nabla_\alpha H(\hat{c}, \hat{\alpha}) + o(\Delta c, \Delta \alpha). \end{aligned}$$

From Proposition 1, $D_{\Delta c} H(\hat{c}, \hat{\alpha}) = \Delta c \cdot \nabla_c H(\hat{c}, \hat{\alpha}) = 0$. Similarly, from Proposition 2, $D_{\Delta \alpha} H(\hat{c}, \hat{\alpha}) = \Delta \alpha \cdot \nabla_\alpha H(\hat{c}, \hat{\alpha}) = 0$. Therefore, $H(c, \alpha) = H(\hat{c}, \hat{\alpha}) + o(\Delta c, \Delta \alpha)$. ■

Section 8: Independent System Mathematical Analysis

In this section, the following theorem and proposition provide mathematical analysis for the independent system showing that for $n = 2$, H increases when $d < \sqrt{v_1 v_2}$.

Theorem 1. Suppose that $f(u, z) > 0$ is a saturating C^2 function defined for all $u, z > 0$, such that $f_u(u, z) > 0$.

If the function

$$\sigma(u, z) = \frac{f_v(u, z) - f(u, z)f_z^\infty(z)/f^\infty(z)}{uf_u(u, z)} \quad (8)$$

is strictly increasing on u , then $H(z)$ is increasing for every parameter $z > 0$.

Proof. Let $p(z)$ and $q(z)$ represent the EC_{10} and EC_{90} values of the dose response respectively, that is

$$f(p(z), z) = 0.10f^\infty(z) \text{ and } f(q(z), z) = 0.90f^\infty(z),$$

where $f^\infty(z)$ is the maximum value of $f(u, z)$. Differentiating both sides, it follows that

$$f_u(p(z), z)p'(z) + f_v(p(z), z) = 0.10f_z^\infty(z) \text{ and } f_u(q(z), z)q'(z) + f_v(q(z), z) = 0.90f_z^\infty(z).$$

That is

$$p'(z) = \frac{-f_z(p(z), z) + 0.10f_z^\infty(z)}{f_u(p(z), z)} \text{ and}$$

$$q'(z) = \frac{-f_z(q(z), z) + 0.90f_z^\infty(z)}{f_u(q(z), z)}.$$

Recall $H(z) = \frac{\ln(81)}{\ln(q(z)/p(z))}$. Then, $\frac{dH}{dz} = \frac{-\ln(81)(q'(z)p(z) - p'(z)q(z))}{q(z)p(z)\left(\ln\left(\frac{q(z)}{p(z)}\right)\right)^2}$, and $\frac{dH}{dz} > 0$ if and only if

$$\frac{-\ln(81)(q'(z)p(z) - p'(z)q(z))}{q(z)p(z)\left(\ln\left(\frac{q(z)}{p(z)}\right)\right)^2} > 0$$

\iff

$$p'(z)q(z) > q'(z)p(z)$$

\iff

$$\frac{-f_z(p(z), z) + 0.10f_z^\infty(z)}{f_u(p(z), z)}q(z) > \frac{-f_z(q(z), z) + 0.90f_z^\infty(z)}{f_u(q(z), z)}p(z)$$

\iff

$$\frac{f_z(p(z), z) - 0.10f_z^\infty(z)}{p(z)f_u(p(z), z)} < \frac{f_z(q(z), z) - 0.90f_z^\infty(z)}{q(z)f_u(q(z), z)}$$

\iff

$$\frac{f_z(p(z), z) - f(p(z), z)f_z^\infty(z)/f^\infty(z)}{p(z)f_u(p(z), z)} < \frac{f_z(q(z), z) - f(q(z), z)f_z^\infty(z)/f^\infty(z)}{q(z)f_u(q(z), z)}.$$

The last inequality follows since $0.10 = f(p(z), z)/f^\infty(z)$, $0.90 = f(q(z), z)/f^\infty(z)$. Overall, we have that

$$\frac{dH}{dz} > 0 \text{ if and only if } \sigma(p(z), z) < \sigma(q(z), z).$$

Thus, if $\sigma(u, z)$ is an increasing function of u , then $\frac{dH}{dz} > 0$. ■

Given the dose response in (7), does $\sigma(u, v)$ increase on u ? Here, we provide a derivation for the associated σ function for the Independent dose response when $n = 2$. More specifically, we now apply the theorem to the independent model with $z = v_1$.

Proposition 4. For the independent system in (7), for $n = 2$, assuming that $k_i = k$ are equal to each other, $H(v)$ is increasing on v_1 and v_2 if $d < \sqrt{v_1 v_2}$.

Proof. We use the above result and show that when $d < \sqrt{v_1 v_2}$, then $\sigma(u, v)$ is increasing on u .

Consider $f(u, v)$ as given by (7) and $\sigma(u, v)$ from (8). For the case $n = 2$, without loss of generality let $S_T = 1$; a different value of S_T only re-scales the dose response and does not change H . Let $z = v_1$. Notice that setting $z = v_2$ would give the same result by symmetry. Also, let $\phi = \frac{Q_{11, v_1}}{Q_{11}} > 0$, $Q_{I, v_1} = \frac{\partial Q_I}{\partial v_1}$, and $f_u = \frac{\partial f}{\partial u}$. It now follows that

$$\begin{aligned} f(u, v) &= Q_{00}S_{00} + Q_{01}S_{01} + Q_{10}S_{10} + Q_{11}S_{11} \\ f_{v_1}(u, v) &= Q_{00, v_1}S_{00} + Q_{01, v_1}S_{01} + Q_{10, v_1}S_{10} + Q_{11, v_1}S_{11} & f_{v_1}^\infty(v) &= Q_{11, v_1} \\ f_u(u, v) &= Q_{00}S'_{00} + Q_{01}S'_{01} + Q_{10}S'_{10} + Q_{11}S'_{11} & f_u^\infty(v) &= Q_{11}. \end{aligned}$$

Then,

$$\begin{aligned} \sigma(u, v_1) &= \frac{f_{v_1}(u, v_1) - f(u, v_1)f_{v_1}^\infty(v_1)/f^\infty(v_1)}{uf_u(u, v_1)} \\ &= \frac{Q_{00, v_1}S_{00} + Q_{01, v_1}S_{01} + Q_{10, v_1}S_{10} + Q_{11, v_1}S_{11} - (Q_{00}S_{00} + Q_{01}S_{01} + Q_{10}S_{10} + Q_{11}S_{11})\frac{Q_{11, v_1}}{Q_{11}}}{u(Q_{00}S'_{00} + Q_{01}S'_{01} + Q_{10}S'_{10} + Q_{11}S'_{11})} \\ &= \frac{\frac{Q_{10, v_1}ku}{(u+k)^2} + \frac{Q_{01, v_1}ku}{(u+k)^2} + \frac{Q_{11, v_1}u^2}{(u+k)^2} - \left(Q_{00}\frac{k^2}{(u+k)^2} + Q_{10}\frac{ku}{(u+k)^2} + Q_{01}\frac{ku}{(u+k)^2} + Q_{11}\frac{u^2}{(u+k)^2} \right) \phi}{u \left(Q_{00}\frac{-2k^2}{(u+k)^3} + Q_{10}\frac{k(k-u)}{(u+k)^3} + Q_{01}\frac{k(k-u)}{(u+k)^3} + Q_{11}\frac{2uk}{(u+k)^3} \right)} \\ &= \frac{\frac{1}{(u+k)^2} kuQ_{10, v_1} + kuQ_{01, v_1} + u^2Q_{11, v_1} - \phi(k^2Q_{00} + ukQ_{01} + ukQ_{10} + u^2Q_{11})}{\frac{u}{(u+k)^3} - 2k^2Q_{00} + k(k-u)Q_{10} + k(k-u)Q_{01} + 2ukQ_{11}} \\ &= \frac{u+k}{u} \frac{ku(Q_{10, v_1} + Q_{01, v_1}) - \phi k^2Q_{00} - uk\phi(Q_{10} + Q_{01})}{k^2(-2Q_{00} + Q_{10} + Q_{01}) + uk(-Q_{10} - Q_{01} + 2Q_{11})} \\ &= \frac{u+k}{u} \frac{u(Q_{10, v_1} + Q_{01, v_1} - \phi Q_{10} - \phi Q_{01}) - \phi kQ_{00}}{k(Q_{10} + Q_{01} - 2Q_{00}) + u(2Q_{11} - Q_{10} - Q_{01})} \\ &= \frac{u+k}{u} \frac{u(\phi Q_{10} + \phi Q_{01} - Q_{10, v_1} - Q_{01, v_1}) + \phi kQ_{00}}{u(2Q_{11} - Q_{10} - Q_{01}) + k(Q_{10} + Q_{01} - 2Q_{00})} \\ &= \frac{u+k}{C_3u + C_4} \frac{C_1u + C_2}{u}, \end{aligned}$$

where

$$\begin{aligned} C_1 &= \phi Q_{10} + \phi Q_{01} - Q_{10, v_1} - Q_{01, v_1} & C_3 &= 2Q_{11} - Q_{01} - Q_{10} > 0 \\ C_2 &= \phi kQ_{00} > 0 & C_4 &= k(Q_{10} + Q_{01} - 2Q_{00}) > 0. \end{aligned}$$

Let $\tau_1 = \frac{u+k}{C_3u + C_4}$ and $\tau_2 = -\frac{C_1u + C_2}{u}$, so that $\sigma = \tau_1\tau_2$. Note that τ_2 is a strictly increasing function on u since $\tau_2 = -C_1 - \frac{C_2}{u}$.

In the following text, we show that τ_1 is also strictly increasing on u if and only if $d < \sqrt{v_1 v_2}$. To see this, notice that τ_1 is strictly increasing if and only if $\frac{k}{C_4} < \frac{1}{C_3}$, which is equivalent to $C_3k < C_4$. This is equivalent to

$$\begin{aligned}
k(2Q_{11} - Q_{01} - Q_{10}) &< k(Q_{01} + Q_{10} - 2Q_{00}) \\
2Q_{11} - Q_{01} - Q_{10} &< Q_{01} + Q_{10} - 2Q_{00} \\
0 &< Q_{01} + Q_{10} - Q_{00} - Q_{11} \\
0 &< \frac{v_2}{d+v_2} + \frac{v_1}{d+v_1} - \frac{1}{d+1} - \frac{v_1v_2}{d+v_1v_2} \\
0 &< v_2(d+v_1)(d+1)(d+v_1v_2) + v_1(d+v_2)(d+1)(d+v_1v_2) \\
&\quad - (d+v_2)(d+v_1)(d+v_1v_2) - v_1v_2(d+1)(d+v_2)(d+v_1) \\
0 &< v_2(d+1)(d+v_1)[d+v_1v_2 - v_1(d+v_2)] \\
&\quad + (d+v_2)(d+v_1v_2)[v_1(d+1) - (d+v_1)] \\
0 &< v_2d(d+1)(d+v_1)(1-v_1) + d(d+v_2)(d+v_1v_2)(v_1-1) \\
0 &< (v_1-1)[-v_2(d+1)(d+v_1) + (d+v_2)(d+v_1v_2)] \\
0 &< (d+v_2)(d+v_1v_2) - v_2(d+1)(d+v_1) \\
0 &< d^2 + dv_1v_2 + dv_2 + v_1v_2^2 - v_2d^2 - v_1v_2d - v_2d - v_1v_2 \\
0 &< d^2 + v_1v_2^2 - v_2d^2 - v_1v_2 \\
0 &< (v_2-1)(v_1v_2 - d^2) \\
0 &< (v_1v_2 - d^2) \\
d^2 &< v_1v_2 \\
d &< \sqrt{v_1v_2}.
\end{aligned}$$

As long as $d < \sqrt{v_1v_2}$, it follows that τ_1 is an increasing function on u making σ the product of two increasing functions and thus, σ is increasing on u . ■

Theorem 2. Suppose $n = 2$. If $Q_{01} > 1/2$ and $Q_{10} > 1/2$, then H is increasing as a function of v_1 and v_2 .

Proof. If $Q_{01} > 1/2$ and $Q_{10} > 1/2$, it follows that $\frac{v_1}{d+v_1} > \frac{1}{2}$ and $\frac{v_2}{d+v_2} > \frac{1}{2}$. Then,

$$\begin{aligned}
v_1 &> \frac{1}{2}(d+v_1) \text{ and } v_2 > \frac{1}{2}(d+v_2) \\
2v_1 &> d+v_1 \quad \text{and } 2v_2 > d+v_2 \\
v_1 &> d \quad \text{and } v_2 > d.
\end{aligned}$$

Thus, $d^2 < v_1v_2$ and so $d < \sqrt{v_1v_2}$. Thus, by Proposition (4), H is increasing as a function of v_1 and v_2 . ■

The point where $d = \sqrt{v_1 v_2}$ (the dotted line in Fig. 4c) corresponds to the case where the conformational free energy contributed by the singly modified forms is equal to exactly half of that contributed by the doubly modified forms. It can be shown that this situation (which we shall call the ‘linear regime’) results in a dose response curve with a Hill number of 1. If v_1 and/or v_2 are then increased so that $\sqrt{v_1 v_2}$ (the region to the right of the dotted line in Fig. 4c) becomes greater than d , the singly modified forms now have more than half the conformational free energy of the doubly modified forms, and the Hill number increases. The Hill number will continue to increase until the two singly modified forms, collectively, contribute exactly the same conformational free energy as the doubly modified form. At this point, the system is in the ‘1+ regime’, where modification of one, the other, or both sites lead to the same level of activation.

On the other hand, if $d > \sqrt{v_1 v_2}$, then the system is closer to the ‘both or none regime’, where efficient activation only occurs when both sites are modified. Here, increasing v_1 or v_2 reduces the Hill number by pushing the system away from ‘both or none’ and closer to ‘linear’.

Section 9: Discussion

In a protein with multiple ligand binding sites, the individual sites can differ from each other in two ways: in their microscopic ligand binding affinity, and in the energetic contribution they make, once bound or modified, to functional outcomes such as a ligand-induced conformational change in the bound protein. Likewise, for a protein that is post-translationally modified on multiple sites, the individual sites may have different modification efficiencies, and may also, independently, make differential contributions to downstream functional consequences once modified. For example, in the case of phosphorylation, the amino acid sequence around the target phosphoacceptor residue can substantially influence the efficiency of phosphorylation by the relevant kinase, as well as the efficiency of dephosphorylation by cellular phosphatases [34]. Such tuning of the steady-state level of site modification is biochemically distinct and clearly separable from the effects that the phosphorylation of that site will have on the conformation of the substrate, its ability to bind other macromolecules, etc. [35–37].

Previously, Enciso and Ryerson [8] asked the question “how can the microscopic ligand binding affinities (a.k.a. modification efficiencies) be tuned if the goal is to maximize ultrasensitivity?” Interestingly, they found that ultrasensitivity was maximal when the microscopic affinities were balanced. For instance, for a protein with 4 ligand binding sites, ultrasensitivity was maximized when all 4 sites had exactly the same ligand binding affinity. For a protein with 4 phosphorylation sites, ultrasensitivity was maximized when all 4 sites had the same phosphorylation/dephosphorylation efficiency.

Here we examined how differential energetic contributions of the sites might affect the performance objective of ultrasensitivity. We considered a simple model in which binding/modification promotes a conformational change

that flips the modified molecule from an inactive to an active state; this example is readily extended to other known consequences of ligand binding or post-translational modification. We generalized the classic allosteric MWC model to allow for differences in the energetic contributions for any number of distinct sites. We also considered an independent modification model that does not assume allostery or cooperativity among sites. For the generalized MWC system, we found that ultrasensitivity generally increased when the energetic contribution (i.e., the conformational free energy) of one or more of the sites was strengthened. Here, 'strengthened' means that the conformational free energy became more negative; this results in the corresponding activation parameter c becoming smaller. Furthermore, we found that there was no benefit derived from balancing the conformational free energies, nor any penalty for unbalancing them. Our results have implications for understanding the potential trajectories that can be pursued by a protein under selective pressure to increase the ultrasensitivity of its response to modification.

Regarding our finding that decreasing the activation parameter c_i of individual sites has a strong tendency to increase the Hill coefficient, this result is analogous to work by Rubin and Changeux [38]. In Figure 2 of that work the authors illustrate computationally that for fixed parameter values of the MWC model, decreasing c leads to an increase in a different version of the Hill coefficient. In our paper we are able to consider individual sites, rather than all sites together, so our result is in a sense a generalization of that shown in [38].

Despite the fact that there is no penalty associated with the conformational free energies being unbalanced, our model nevertheless suggests a factor that may tend to lead to roughly balanced conformational free energies: diminishing returns. Successive, equal-valued improvements of conformational free energy are diminishing with respect to their effect on ultrasensitivity. That is, changes that are of equal magnitude to previous changes increase ultrasensitivity by a smaller amount than the previous changes did. Furthermore, changes to weaker sites increase ultrasensitivity more dramatically than equivalent changes to stronger sites. Eventually, the marginal increase in ultrasensitivity caused by additional improvements to conformational free energy becomes negligible. At this point it can be argued that a zone of effective neutrality has been reached, where the probability of fixation of a new mutation that incrementally improves ultrasensitivity will be essentially indistinguishable from the probability of fixation of a neutral mutation [39]. At this point, substantial improvement to ultrasensitivity can only arise if the molecule evolves an additional site.

There is an additional factor that may further promote the balancing of conformational free energies. Using both computational and mathematical analysis, we showed that when the sites are at least moderately active, the ultrasensitivity is mostly dependent on the mean of the activation parameters and is largely independent from their variance (Fig. 3a). Since increasing the variance of the activation parameters tends to be associated with weaker (less negative) total conformational free energy, a prediction is that the sites tend to have roughly equal

activation parameters. This can be implemented e.g. by bulk electrostatic mechanisms, which are commonly found experimentally [23, 24, 40].

These considerations lead to a prediction that conformational free energies will be roughly balanced, with a kcal/mol value roughly equal to the point where ultrasensitivity starts to level out substantially. As shown in Figure 3b and 3c and Tables 1 and S1, this “leveling out point” is roughly between -2 to -4 kcal/mol per site, depending on the number of sites n and the level of basal activation (which is determined by the parameter L). This range of -2 to -4 kcal/mol does assume that the efficiency of modification is roughly constant across all sites, but is otherwise surprisingly independent of other parameters. For example, the range found changed very little upon variation of the value of L from 30 to 10,000, therefore covering most biochemically realistic values for this constant. The range of approximately -2 to -4 kcal/mol corresponds to activation coefficient (c) values between approximately 0.05 and 0.001. Such c values are all within the range reported for classic “MWC enzymes” such as threonine deaminase, glucose-6-phosphate deaminase, aspartate transcarbamoylase and glyceraldehyde-3-phosphate dehydrogenase [41–44]. Moreover, with regard to phosphorylation, the effect of a single phosphate on conformation [25], protein-protein binding [26] [Nishi, 2011] or protein-membrane binding [23, 24] has been estimated to be about 2 kcal/mol.

We also showed (Fig 3 d) that when the number of sites is large, and a hypothetical maintenance cost per site is included (such as might arise from rapid phosphorylation-dephosphorylation cycles [31, 32]), an optimal strategy to maximize ultrasensitivity can be to focus on a subset of the sites, and essentially keep the other sites silent. In such cases, evolving another site is not a viable strategy to increase ultrasensitivity, and it can be argued that there is an optimal number of functional sites that will maximize benefit (ultrasensitivity) while containing cost.

This analysis applies for other forms of multisite modification other than phosphorylation such as ligand binding, methylation, acetylation, etc. When the multisite substrate has a symmetric structure (such as hemoglobin which is a tetramer), one can assume that the conformational free energy is similar across all sites. In this sense the current study is most relevant when the substrate structure is more heterogeneous, such as in the case of phosphorylation. Although phosphorylation consumes energy and is not thermodynamically closed, the MWC model is still a popular model to describe it [8, 45, 46]. It is also mathematically more amenable than the non-allosteric, independent model that we also included for completeness.

Work by Kafri et al [19] has previously studied a mathematical model of chaperon-containing TCP-1 protein that has several sites with different ATP binding affinities. This system can provide very interesting parallels with our framework. Their mathematical model shows that when a protein has multiple sites with different ligand affinities, the Hill coefficient can be reduced leading to apparent negative cooperativity. We do observe a similar effect (see eg Fig. 3), although that model has important differences such as variability in modification affinity rather than conformational free energy.

The analysis in this manuscript is limited to systems in equilibrium, i.e. the long term response to a constant input. For non-equilibrium systems, and in situations where energy is used, recent work by Estrada et al. [47] shows that one can obtain a larger Hill coefficient. The authors use techniques similar to kinetic proofreading, which can give rise to large response differences given small differences in ligand affinity. A full discussion of non-equilibrium dynamics is however outside of the scope of our work.

Many dose response curves for allosterically-regulated proteins can be well-modeled by the standard MWC model. Our goal in generalizing the MWC model was to explore the qualitative theoretical consequences of allowing the conformational free energies of different sites to vary, and not to make a tool for empirical fitting to data. On this point, however, it should be noted that Stefan et al. [48] have shown how an extended MWC model such as the one developed here can be used in parameter estimation, and experimental methods to measure MWC parameters are constantly improving [49–51].

Data Accessibility

Source code for figures is available on GitHub at: <https://github.com/lilagunes/uci-dcb-ultCfe>

References

1. Pearson, G., Robinson, F., Gibson, T. B., Xu, B., Karandikar, M., Berman, K., & Cobb, M. H. (2001). Mitogen-Activated Protein (MAP) Kinase Pathways : Regulation and Physiological Functions. *Endocrine Reviews*, 22(2), 153183.
2. Enciso, G. A. (2013). Multisite Mechanisms for Ultrasensitivity in Signal Transduction. In *Nonautonomous Dynamical Systems in the Life Sciences* (pp. 199-224). Springer, Cham. <https://doi.org/10.1007/978-3-319-03080-7>
3. Olsen, J. V., Blagoev, B., Gnäd, F., Macek, B., Kumar, C., Mortensen, P., & Mann, M. (2006). Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. *Cell*, 127(3), 635648. <https://doi.org/10.1016/j.cell.2006.09.026>
4. Audagnotto, M., & Dal Peraro, M. (2017). Protein post-translational modifications: In silico prediction tools and molecular modeling. *Computational and Structural Biotechnology Journal*, 15, 307319. <https://doi.org/10.1016/j.csbj.2017.03.004>
5. Barford, D., Hu, S. H., & Johnson, L. N. (1991). Structural mechanism for glycogen phosphorylase control by phosphorylation and AMP. *J Mol Biol*, 218(1), 233260. [https://doi.org/10.1016/0022-2836\(91\)90887-C](https://doi.org/10.1016/0022-2836(91)90887-C)
6. Liu, X., Bardwell, L., & Nie, Q. (2010). A combination of multisite phosphorylation and substrate sequestration produces switch-like responses. *Biophysical Journal*, 98(8), 13961407. <https://doi.org/10.1016/j.bpj.2009.12.4307>
7. Ferrell, J. E., & Ha, S. H. (2014). Ultrasensitivity part II : multisite phosphorylation , stoichiometric inhibitors , and positive feedback. *Trends in Biochemical Sciences*, 39(11), 556569. <https://doi.org/10.1016/j.tibs.2014.09.003>
8. Enciso, G. A., & Ryerson, S. (2017). The effect of site-to-site variability in ultrasensitive dose responses. *J Math Biol*, 74(12), 2341. <https://doi.org/10.1007/s00285-016-1013-5>

9. Hill, A. V. (1910). The possible effects of the aggregation of the molecule of hemoglobin on its dissociation curves. *J. Physiology*, 40, iv-vii. <https://doi.org/10.1017/CBO9781107415324.004>
10. Monod, J., Wyman, J., & Changeux, J.P. (1965). On the Nature of Allosteric Transitions: A Plausible Model. *J Mol Biol.*, 12, 88-118.
11. Motlagh, H. N., Wrabl, J. O., Li, J., & Hilser, V. J. (2014). The ensemble nature of allostery. *Nature*, 508(7496), 331339. <https://doi.org/10.1038/nature13001>
12. Tsai, C. J., & Nussinov, R. (2014). A unified view of how allostery works. *PLoS Comput Biol*, 10(2), e1003394. <https://doi.org/10.1371/journal.pcbi.1003394>
13. Marzen, S., Garcia, H. G., & Phillips, R. (2013). Statistical mechanics of Monod-Wyman-Changeux (MWC) models. *J Mol Biol*, 425(9), 14331460. <https://doi.org/10.1016/j.jmb.2013.03.013>
14. Martins, B. M., & Swain, P. S. (2011). Trade-offs and constraints in allosteric sensing. *PLoS Comput Biol*, 7(11), e1002261. <https://doi.org/10.1371/journal.pcbi.1002261>
15. Nussinov, R., & Tsai, C. J. (2013). Allostery in disease and in drug discovery. *Cell*, 153(2), 293305. <https://doi.org/10.1016/j.cell.2013.03.034>
16. Changeux, J. P. (2012). Allostery and the Monod-Wyman-Changeux model after 50 years. *Annu Rev Biophys*, 41, 103133. <https://doi.org/10.1146/annurev-biophys-050511-102222>
17. Berlow, R. B., Dyson, H. J., & Wright, P. E. (2018). Expanding the Paradigm: Intrinsically Disordered Proteins and Allosteric Regulation. *J Mol Biol*, 430(16), 23092320. <https://doi.org/10.1016/j.jmb.2018.04.003>
18. Hilser, V. J., & Thompson, E. B. (2007). Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc Natl Acad Sci U S A*, 104(20), 83118315. <https://doi.org/10.1073/pnas.0700329104>
19. Kafri, G., Willison, K. R., & Horovitz, A. (2001). Nested allosteric interactions in the cytoplasmic chaperonin containing TCP-1. *Protein Science*, 10(2), 445-9.
20. Spiess, C., Meyer, A. S., Reissmann, S., & Frydman, J. (2004). Mechanism of the eukaryotic chaperonin: protein folding in the chamber of secrets. *Trends Cell Biol*, 14(11), 598–604. <https://doi.org/10.1016/j.tcb.2004.09.015>
21. Collins, G. A., & Goldberg, A. L. (2017). The Logic of the 26S Proteasome. *Cell*, 169(5), 792–806. <https://doi.org/10.1016/j.cell.2017.04.023>
22. Gruber, R., & Horovitz, A. (2016). Allosteric Mechanisms in Chaperonin Machines. *Chem Rev*, 116(11), 6588–6606. <https://doi.org/10.1021/acs.chemrev.5b00556>
23. Murray, D., Hermida-matsumoto, L., Buser, C. A., Tsang, J., Sigal, C. T., Ben-tal, N., Honig, B., Resh, M., & Mclaughlin, S. (1998). Electrostatics and the Membrane Association of Src : Theory and Experiment. *Biochemistry*, 37(8), 21452159. <https://doi.org/10.1021/bi972012b>
24. Serber, Z., & Ferrell, J. E. (2007). Tuning Bulk Electrostatics to Regulate Protein Function. *Cell*, 128, 441-444. <https://doi.org/10.1016/j.cell.2007.01.018>

25. Shen, T., Wong, C. F. & McCammon, J. A. (2001). Atomistic Brownian dynamics simulation of peptide phosphorylation. *J Am Chem Soc*, 123(37), 9107–11.
26. Nishi, H., Hashimoto, K., & Panchenko, A. R. (2011). Phosphorylation in protein-protein binding: effect on stability and function. *Structure*, 19(12), 1807–1815. <https://doi.org/10.1016/j.str.2011.09.021>
27. Goldbeter, A., & Koshland, D. E. (1981). An amplified sensitivity arising from covalent modification in biological systems. *Proc Natl Acad Sci U S A*, 78(11), 6840-6844. <https://doi.org/10.1073/pnas.78.11.6840>
28. Gunawardena, J. (2005). Multisite protein phosphorylation makes a good threshold but can be a poor switch. *Proc Natl Acad Sci U S A*, 102(41), 14617-14622.
29. Satoppa, V., Albrecht, J., Irwin, D., & Raghavan, B. (2011). Finding a “Kneedle” in a Haystack: Detecting Knee points in System Behavior. 2011 31st International Conference on Distributed Computing Systems Workshop. IEEE. DOI: 10.1109/ICDCSW.2011.20
30. Levitzki, A. (1978). Quantitative aspects of allosteric mechanisms. Springer-Verlag, Berlin.
31. Kleiman, L. B., Maiwald, T., Conzelmann, H., Lauffenburger, D. A., & Sorger, P. K. (2011). Rapid phospho-turnover by receptor tyrosine kinases impacts downstream signaling and drug binding. *Mol Cell*, 43(5), 723–737. <https://doi.org/10.1016/j.molcel.2011.07.014>
32. Gelens, L., & Saurin, A. T. (2018). Exploring the Function of Dynamic Phosphorylation-Dephosphorylation Cycles. *Dev Cell*, 44(6), 659–663. <https://doi.org/10.1016/j.devcel.2018.03.002>
33. Enciso, G., Kellogg, D. R., & Vargas, A. (2014). Compact Modeling of Allosteric Multisite Proteins: Application to a Cell Size Checkpoint. *PLoS Computational Biology*, 10(2). <https://doi.org/10.1371/journal.pcbi.1003443>
34. Miller, C. J., & Turk, B. E. (2018). Homing in: Mechanisms of Substrate Targeting by Protein Kinases. *Trends Biochem Sci*, 43(5), 380–394. <https://doi.org/10.1016/j.tibs.2018.02.009>
35. Narayanan, A., & Jacobson, M. P. (2009). Computational studies of protein regulation by post-translational phosphorylation. *Curr Opin Struct Biol*, 19(2), 156–163. <https://doi.org/10.1016/j.sbi.2009.02.007>
36. Nishi, H., Shaytan, A. & Panchenko, A. R. (2014). Physicochemical mechanisms of protein regulation by phosphorylation. *Front Genet*, 5, 270.
37. Csizmok, V. & Forman-Kay, J. D. (2018). Complex regulatory mechanisms mediated by the interplay of multiple post-translational modifications. *Curr Opin Struct Biol*, 48, 58–67.
38. Rubin, M. M., & Changeux, J. P. (1966). On the nature of allosteric transitions: implications of non-exclusive ligand binding. *J Mol Biol*, 21(2), 265–274. [https://doi.org/10.1016/0022-2836\(66\)90097-0](https://doi.org/10.1016/0022-2836(66)90097-0)
39. Hurst, L. D. (2009). Fundamental concepts in genetics: Genetics and the understanding of selection. *Nature Reviews Genetics*, 10(2), 83–93. <https://doi.org/10.1038/nrg2506>
40. Pufall, M. A., Lee, G. M., Nelson, M. L., Kang, H., Velyvis, A., & Kay, L. E. (2005). Variable Control of Ets-1 DNA Binding by Multiple Phosphates in an Unstructured Region. *Science*, 309, 142-146.

41. Eisenstein, E. (1995). Allosteric regulation of biosynthetic threonine deaminase from *Escherichia coli*: effects of isoleucine and valine on active-site ligand binding and catalysis. *Arch Biochem Biophys*, 316(1), 311–318. <https://doi.org/10.1006/abbi.1995.1042>
42. Zonszein, S., Alvarez-Anorve, L. I., Vazquez-Nunez, R. J., & Calcagno, M. L. (2014). The tertiary origin of the allosteric activation of *E. coli* glucosamine-6-phosphate deaminase studied by sol-gel nanoencapsulation of its T conformer. *PLoS One*, 9(5), e96536. <https://doi.org/10.1371/journal.pone.0096536>
43. Howlett, G. J., Blackburn, M. N., Compton, J. G., & Schachman, H. K. (1977). Allosteric regulation of aspartate transcarbamoylase. Analysis of the structural and functional behavior in terms of a two-state model. *Biochemistry*, 16(23), 5091–5100. <https://doi.org/10.1021/bi00642a023>
44. Kirschner, K., Gallego, E., Schuster, I., & Goodall, D. (1971). Co-operative binding of nicotinamide-adenine dinucleotide to yeast glyceraldehyde-3-phosphate dehydrogenase. I. Equilibrium and temperature-jump studies at pH 8.5 and 40 degrees C. *J Mol Biol*, 58(1), 29–50. [https://doi.org/10.1016/0022-2836\(71\)90230-0](https://doi.org/10.1016/0022-2836(71)90230-0)
45. Trunnell, N. B., Poon, A. C., Kim, S. Y., & Ferrell, J. E. (2011). Ultrasensitivity in the Regulation of Cdc25C by Cdk1. *Molecular Cell*, 41(3), 263–274. <https://doi.org/10.1016/j.molcel.2011.01.012>
46. Van Zon, J. S., Lubensky, D. K., Altena, P. R. H., & Ten Wolde, P. R. (2007). An allosteric model of circadian KaiC phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America*, 104(18), 7420–7425. <https://doi.org/10.1073/pnas.0608665104>
47. Estrada, J., Wong, F., DePace, A., & Gunawardena, J. (2016). Information Integration and Energy Expenditure in Gene Regulation. *Cell*, 166(1), 234–244. <https://doi.org/10.1016/j.cell.2016.06.012>
48. Stefan, M. I., Edelstein, S. J. & Le Novere, N. (2009). Computing phenomenologic Adair-Klotz constants from microscopic MWC parameters. *BMC Syst Biol*, 3, 68.
49. Klein, M. T., Vinson, P. N. & Niswender, C. M. (2013). Approaches for probing allosteric interactions at 7 transmembrane spanning receptors. *Prog Mol Biol Transl Sci*, 115, 1–59.
50. Velyvis, A., Yang, Y. R., Schachman, H. K. & Kay, L. E. (2007). A solution NMR study showing that active site ligands and nucleotides directly perturb the allosteric equilibrium in aspartate transcarbamoylase. *Proc Natl Acad Sci U S A*, 104(21), 8815–20.
51. Sharon, M., & Horovitz, A. (2015). Probing allosteric mechanisms using native mass spectrometry. *Curr Opin Struct Biol*, 34, 7–16. <https://doi.org/10.1016/j.sbi.2015.05.002>

Supplemental Figures

Table S1: Ultrasensitivity at knee									
n	L=30			L=1,000			L=10,000		
	H	total cfe	cfe/site	H	total cfe	cfe/site	H	total cfe	cfe/site
2	1.43	-5.49	-2.74	1.80	-7.37	-3.69	1.83	-8.64	-4.32
3	1.75	-6.86	-2.29	2.47	-8.67	-2.89	2.63	-10.04	-3.35
4	1.95	-8.19	-2.05	2.99	-9.80	-2.45	3.28	-11.19	-2.80
6	2.20	-11.12	-1.85	3.71	-11.97	-1.99	4.25	-13.33	-2.22
8	2.34	-14.23	-1.78	4.21	-14.39	-1.80	4.94	-15.39	-1.92

Table S1: Ultrasensitivity as measured by the Goldbeter-Koshland formula described in Eqn. (2) along with the approximated knee of curves similar to those in Fig. 3c for fixed values of L and n . Parameters $\alpha_i = \bar{\alpha} = 1$.


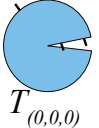

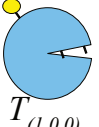

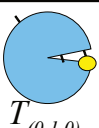
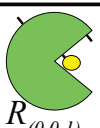


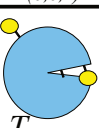
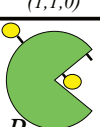
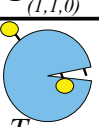
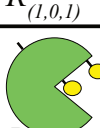
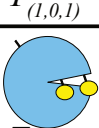
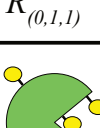
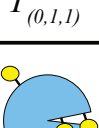
State	Statistical Weight	State	Statistical Weight
 $R_{(0,0,0)}$	1	 $T_{(0,0,0)}$	L
 $R_{(1,0,0)}$	α_1	 $T_{(1,0,0)}$	$L\alpha_1c_1$
 $R_{(0,1,0)}$	α_2	 $T_{(0,1,0)}$	$L\alpha_2c_2$
 $R_{(0,0,1)}$	α_3	 $T_{(0,0,1)}$	$L\alpha_3c_3$
 $R_{(1,1,0)}$	$\alpha_1\alpha_2$	 $T_{(1,1,0)}$	$L\alpha_1\alpha_2c_1c_2$
 $R_{(1,0,1)}$	$\alpha_1\alpha_3$	 $T_{(1,0,1)}$	$L\alpha_1\alpha_3c_1c_3$
 $R_{(0,1,1)}$	$\alpha_2\alpha_3$	 $T_{(0,1,1)}$	$L\alpha_2\alpha_3c_2c_3$
 $R_{(1,1,1)}$	$\alpha_1\alpha_2\alpha_3$	 $T_{(1,1,1)}$	$L\alpha_1\alpha_2\alpha_3c_1c_2c_3$

Figure S1. Statistical weights of MWC modification states for $n=3$. Table of each possible modification state in the generalized MWC system above when $n = 3$ and the corresponding statistical weight of that state. When $n = 3$, the associated partition function $Z = 1 + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_1\alpha_2 + \alpha_1\alpha_3 + \alpha_2\alpha_3 + \alpha_1\alpha_2\alpha_3 + L + \alpha_1c_1L + \alpha_2c_2L + \alpha_3c_3L + \alpha_1\alpha_2c_1c_2L + \alpha_1\alpha_3c_1c_3L + \alpha_2\alpha_3c_2c_3L + \alpha_1\alpha_2\alpha_3c_1c_2c_3L$

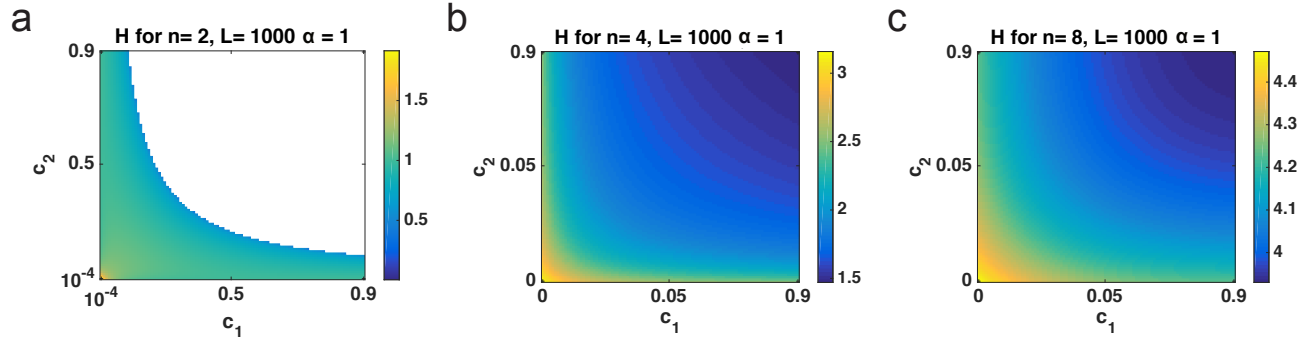


Figure S2: Ultrasensitivity of MWC system. Heat maps for H when $c_1, c_2 \in [10^{-4}, 0.9]$ with $L = 1000$ and $\alpha_i = \bar{\alpha} = 1$ and (a) $n = 2$, (b) $n = 4$, $c_i = 0.01$ for $i \geq 3$, similarly with (c) $n = 8$. These figures are the same data points from Figure 2 in a linear scale.

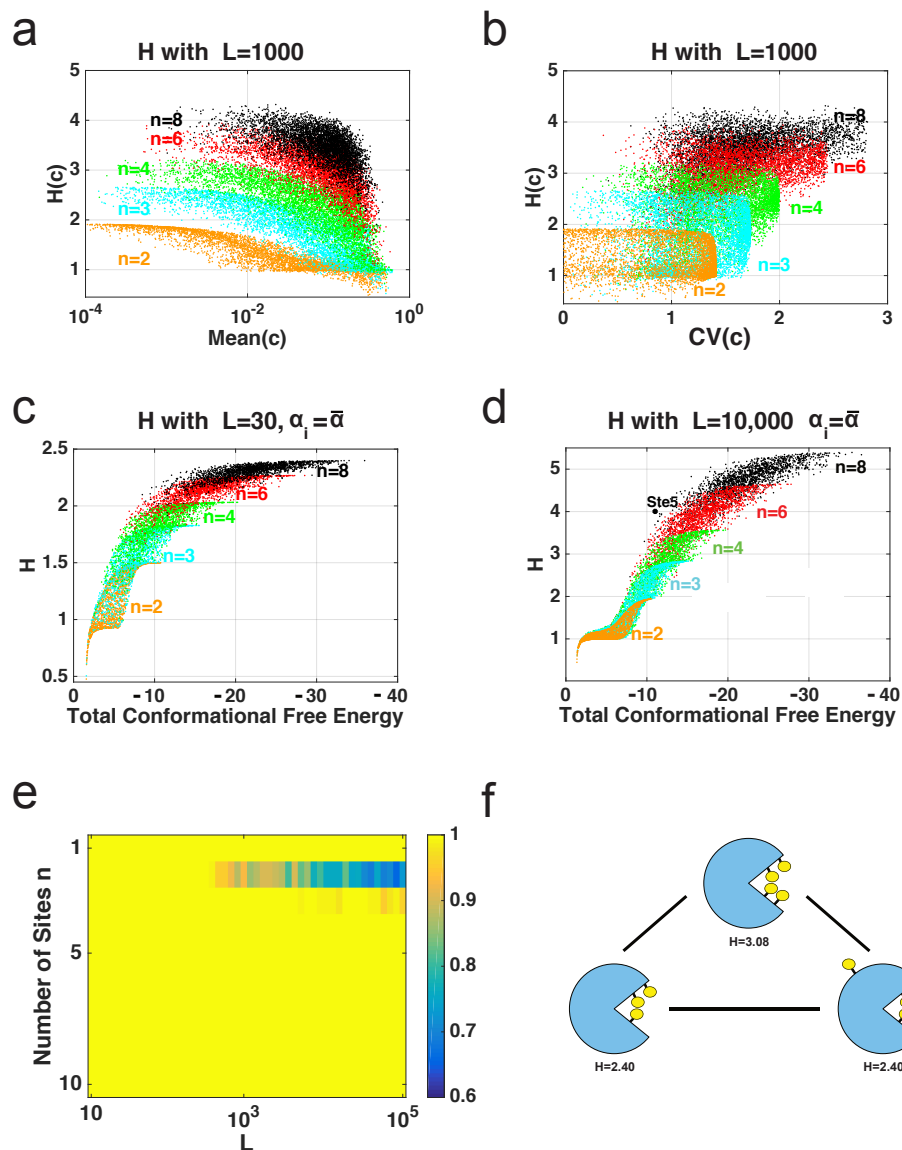


Figure S3. Parameters and H in MWC. (a) Scatter plot for H and the arithmetic mean of c where c_i are independently and logarithmically chosen from $[10^{-4}, 0.9]$, $L = 1000$, and α_i are independently and logarithmically chosen from $[0.1, 10]$ for $n = 2, 3, 4, 8$. (b) H from (a) with the coefficient of variation (CV) along the x-axis. (c) Scatter plot for H when increasing total conformational free energy with $c_i \in [10^{-4}, 0.9]$, $L = 30$ and $\alpha_i = \bar{\alpha}$ for $n = 2, 3, 4, 8$. (d) Scatter plot for H for when $L = 10,000$. (e) Proportion of 10000 parameter sets in which H decreased when a c_i is marginally increased for different values L and n . (f) H values for key scenarios. When a substrate has 3 sites, $H = 2.40$ when $\alpha_i = \bar{\alpha} = 1$, $L = 1000$, and $c_i = 0.01$. Adding a site with $c_4 = 1$ will yield the same H . However, if $c_4 = 0.01$, $H = 3.08$.

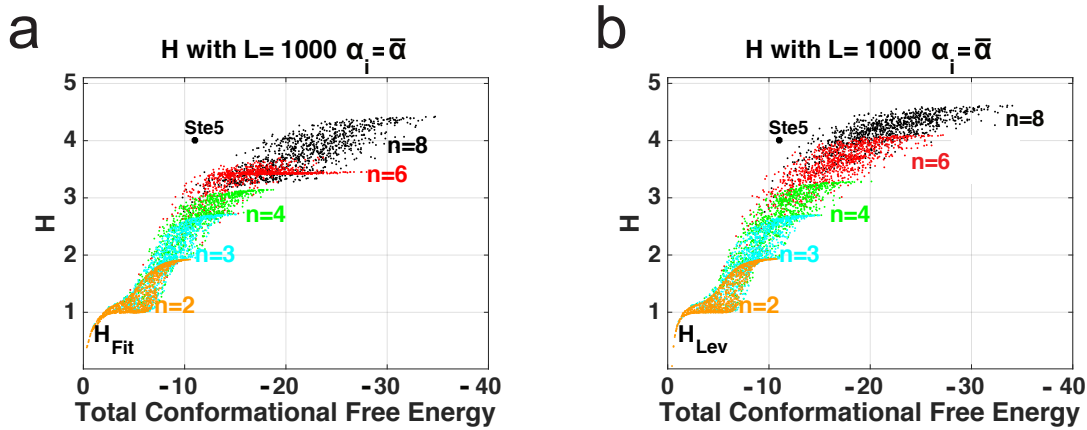


Figure S4. Ultrasensitivity and total conformational free energy in MWC. Scatter plots for ultrasensitivity when increasing total conformational free energy with $c_i \in [10^{-4}, 0.9]$, $L = 1000$ and $\alpha_i = \bar{\alpha} = 1$ for $n = 2, 3, 4, 8$ and 10000 points. Ultrasensitivity is measured with (a) a non-linear regression fit to the Hill function $f = \frac{x^H}{k^H + x^H}$, where H is the Hill number labeled H_{Fit} and (b) a generalized Levitzki derivation for ultrasensitivity [30] as $H_{Lev} = 4 * EC_{50} * f'(EC_{50}, \alpha, c)$ where $f'(EC_{50}, c, \alpha)$ is the derivative of the dose response function evaluated at the EC_{50} , the effective enzyme/ligand concentration at which there is a 50% maximal protein response, label H_{Lev} . We can consider H_{Lev} as the sensitivity at 50% maximal response. EC_{50} was found with the standard MatLab fzero solver and the derivative with diff after normalizing to the $f^\infty(c)$.

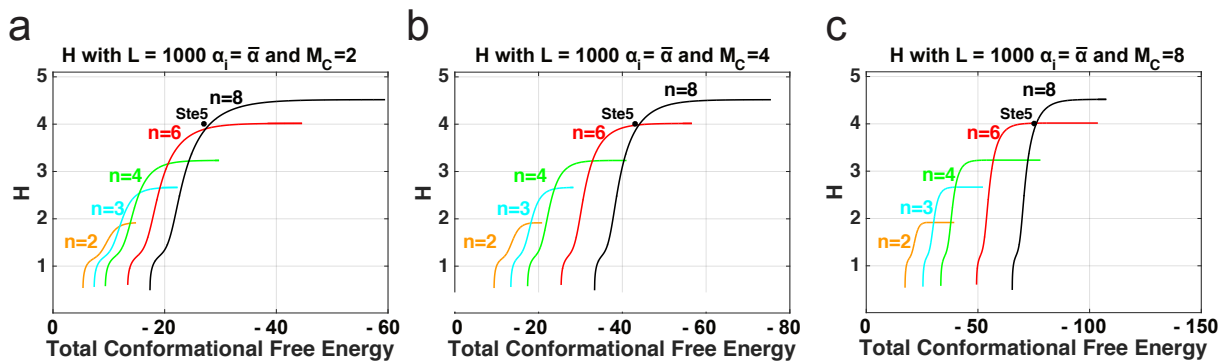


Figure S5. Ultrasensitivity and total conformational free energy in MWC with Maintenance costs. Scatter plots for ultrasensitivity when increasing total conformational free energy with $c_i \in [10^{-4}, 0.9]$, $L = 1000$ and $\alpha_i = \bar{\alpha} = 1$ for $n = 2, 3, 4, 8$ and a maintenance cost of (a) $M_c = 2$, (b) $M_c = 4$ (from Fig. 3) and (c) $M_c = 8$. The Ste5 data point is added for illustration purposes with the same maintenance cost for each of the 8 phosphorylation sites.

Chapter 4: The Role of Docking on phosphorylation of native and novel sites of transcription factor c-Jun by JNK

Section 1: Abstract

One of the most important mechanisms that regulates protein activity is multisite phosphorylation, as it regulates a substrate's enzymatic activity, location, stability, or interactions with other proteins. Understanding kinase-substrate specificity in MAPKs has been crucial to advancing medical therapies for diseases. Here, we demonstrate that the docking site in the c-Jun transcription factor plays an essential role in the phosphorylation of all native and novel sites by the JNK2 MAPK. Here, we characterize the role of the D-site on the initial phosphorylation rates of native and novel sites on c-Jun. We performed *in vitro* kinase assays with JNK2 enzyme, on purified wild-type and mutant derivatives of c-Jun1-108, which contain the native N-terminal phosphorylation sites. Quantitative *in vitro* kinase assays show that c-Jun sites are phosphorylated at different rates and mutation of the D-site decreases the phosphorylation of individual native sites. In fact, removal of the distance between the D-site and a native phosphosite decreases the phosphorylation of the native site. Our results also show that the type of residue, Ser vs Thr, does not influence phosphorylation of native sites. Additionally, we show that introduction of novel sites shows novel phosphorylation that is still under control of the D-site. Our results also show that the surrounding sequence of a novel site does not greatly influence phosphorylation of that novel site. Further showing that the D-site plays a critical role in regulating c-Jun phosphorylation and could provide an additional target for new therapies and provide additional insights into the regulation of multisite MAPK substrates through docking.

Section 2: Introduction

For mammalian cells to respond appropriately and timely to internal and external stimuli, they rely heavily on signal transduction cascades¹. The Mitogen-Activated-Protein Kinases (MAPKs) are critical proteins in various MAPK cascades, including the c-Jun N-terminal Kinases (JNK) pathway². In the JNK pathway, activated MAPKKs phosphorylate MKK4 and MKK7. Once MKK4/7 are phosphorylated, JNK will be activated by phosphorylation of Thr183 and Tyr185, which in turn will phosphorylate target transcription factors such as c-Jun, Atf-2, Elk1, and p53. Transcription factor c-Jun can then move into the cell nucleus and activate its target genes³. The JNK pathway has been shown to regulate critical cell responses such as cell proliferation, apoptosis, inflammation, metabolism and DNA repair⁴.

The JNK pathway has over 40 target genes associated with melanoma and other cancers⁵⁻⁷, nonalcoholic fatty liver disease⁸, Parkinson's Disease⁹, Alzheimer's Disease¹⁰, mitochondrial diseases¹¹, and other diseases³. Making the JNK pathway a target for new therapies^{7,12}. Targets of c-Jun include fosB, wee1, and jun1/2^{3,13,14}. Thus, understanding the mechanisms that regulate JNK target c-Jun can provide new insights for treatments or pharmacology.

In MAP kinases, an active MAPK, activated by multisite phosphorylation, will phosphorylate target substrates. How an activated MAPK phosphorylates its target substrates is not fully understood. MAPKs phosphorylate substrates on a Ser or Thr that are followed immediately by a Pro in the +1 position^{15,16} (Davis1993, Bardwell2006). In fact, removal of the +1 Pro significantly reduces overall substrate-kinase specificity^{16,17}. Even in orthologues that do not contain the S/T-P sequence, phosphorylation is lower when the +1 Proline is not present next to the phosphosite²⁵. However, mutation of neighboring residues they do not make a big difference in the efficiency of the phosphorylation of the target S/T site¹⁶. Thus, in theory, the idea is that the S/T-P consensus sequence is necessary and sufficient for a MAPK to phosphorylate the target. However, in the cell, there are many proteins that contain a S/T-P throughout their sequence (80%) that are not substrates for MAPKs¹⁶. Additionally, studies that test the ability of MAPKs to phosphorylate short peptides containing S/T-P show poor phosphorylation, while the phosphorylation of true MAPK targets is efficient. Suggesting that a second site, the Docking site (D-site) binds to MAPK with a much higher affinity than the actual target phosphosite¹⁸.

The D-site is a region on a substrate that is usually far away from the phosphorylation site. Consensus sequences of common D-sites in MAPKs is made up of the following general motifs: 2-3 basic residues + spacer + hydrophobic-X-hydrophobic motif or K/R₂₋₃-X₁₋₆-ϕ-X-ϕ. On a 3D crystal structure for JNK, the D-site binds to one region of the kinase while the active phosphosite of the target protein binds in a region physically distant to the D-site³.

Inhibition of this D-site shows a significant reduction in substrate phosphorylation¹⁶. Even blocking access to the D-site by a competing D-site peptide significantly reduces phosphorylation efficiency by almost 10-fold in MAP kinases¹⁹. Other studies also show that MAP kinases, MEKs, share the same docking sites downstream of the MAP cascade²⁰. In fact, there are scaffolds and phosphatases share the same D-site consensus sequence, competing for binding¹⁶. Mechanistically, the D-site and the S/T on the target substrate need to bind to the kinase. Further motivation to understand the role of docking in transcription factor regulation.

Here, we ask the question as to what the role of the D-site is on the phosphorylation of novel and native sites. We explore factors that may regulate the phosphorylation of native and novel sites in addition to the D-site such as: the distance between the D-site and phosphosite, a Ser vs Thr in the same position, and the sequence surrounding the target site.

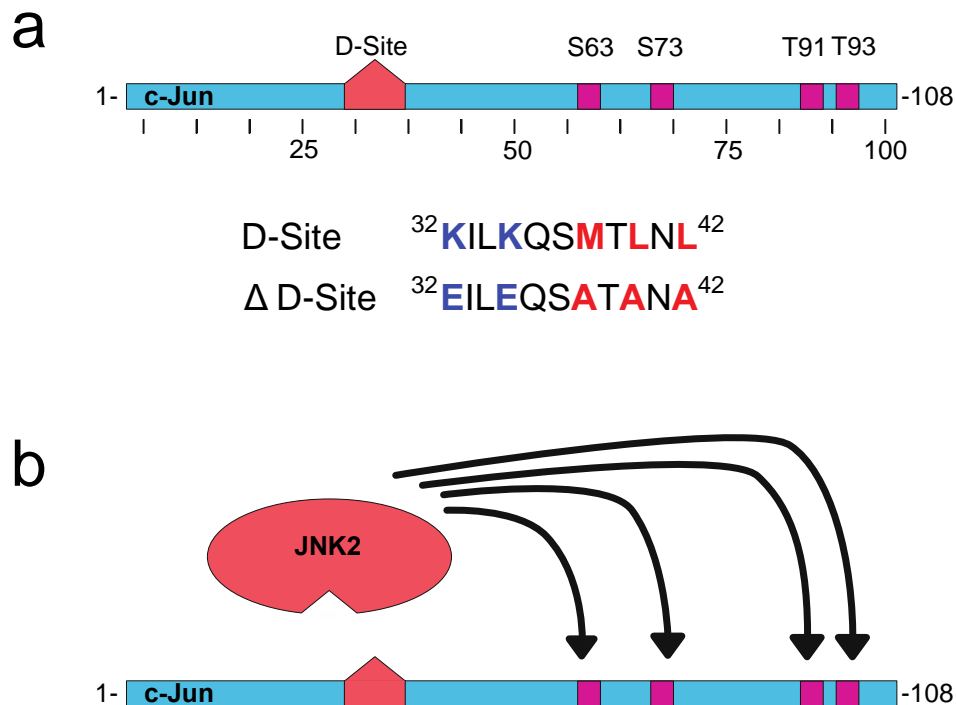


Figure 1: Transcription factor c-Jun has docking and four phosphorylation sites in residues 1-108. (a) Schematic of c-Jun₁₋₁₀₈ with docking site (D-site) labeled as pentagon and phosphorylation sites S63, S73, T91, and T93 in magenta boxes. (b) Simple schematic to demonstrate that JNK2 enzyme has D-site recruitment region that binds to c-Jun and phosphorylates the four target sites.

Our results can be summarized as follows. Phosphorylation rates of single sites can be different than the phosphorylation rate of all four sites. Not only are the phosphorylation rates different, N-terminal docking site is necessary for the phosphorylation of each site. Introduction of novel SP residues in c-Jun shows interesting phosphorylation; some, but not all novel sites can be phosphorylated by JNK2. Additionally, phosphorylation of novel sites requires presence of a functional D-site. To that end, we explored factors that may influence phosphorylation of native and novel sites. Sequence surrounding novel sites did not significantly increase phosphorylation of a potential novel site. Our results also indicate that the type of residue does not significantly influence phosphorylation. In other words, Ser vs Thr phosphosites at the same residue have similar phosphorylation. Lastly, the amino acid distance between the D-site and a

native site also influences phosphorylation, our results indicate that if the distance between the D-site and a native site is too small, there is a loss in phosphorylation of the native site.

Section 3: Experimental Methods

3.1 Genes

The mammalian genes used in this study were human c-Jun and human JNK2.

3.2 Plasmids for the production of GST fusion proteins

The vector used for generating GST-fusion proteins was pGEX-LB, a derivative of pGEX-4T-1 (Amersham-Pharmacia Biotech). In pGEX-LB, an encoded Pro residue is replaced with a Gly-Gly-Gly-Gly-Gly-Ser-Gly coding sequence to promote the independent functioning of the GST and fusion moieties. Plasmid GST-c-Jun₁₋₁₀₈ encodes a fusion of GST to human c-Jun with residues 1-108 generated by introducing a stop codon after the 108th residue in c-Jun with Site-Directed Mutagenesis (SDM). A Null mutant GST-c-Jun₁₋₁₀₈ AAAA was made as a negative control in which all native phosphorylation sites were mutated into non-functional residues; each of the Ser and Thr were mutated, by SDM, into Ala. The single site c-Jun mutants were generated from the null plasmid with SDM by re-introducing the native Ser or Thr, labeled GST-c-Jun₁₋₁₀₈ S63, S73, T91, and T93. See Table I for primer sequences.

The GST-Δ-c-Jun₁₋₁₀₈ mutants were generated with the native Docking-Site (D-Site) disrupted from the GST-c-Jun₁₋₁₀₈ AAAA plasmid with SDM. See Table I for primer sequences. Single phosphorylation sites were then re-introduced, similar to the single site GST-c-Jun₁₋₁₀₈ mutants described above.

Novel sites were introduced to GST-c-Jun₁₋₁₀₈ AAAA by mutating two residues into a Ser-Pro sequence. Using SDM, GST-c-Jun₁₋₁₀₈ F9S mutant was generated by mutating the Phe-Tyr 9th and 10th residues into a Ser-Pro sequence. Similarly with mutants GST-c-Jun₁₋₁₀₈ L53S and GST-c-Jun₁₋₁₀₈ I81S. See Table I for primer sequences. The docking site for mutants with a potential novel site was also mutated in the same fashion as the GST-Δ-c-Jun₁₋₁₀₈ mutants described above.

Additional mutants in this study include mutant GST-c-Jun₁₋₁₀₈ H52G+L53S mutant, GST-c-Jun₁₋₁₀₈ del-63S, and GST-c-Jun₁₋₁₀₈ S63T. The GST-c-Jun₁₋₁₀₈ H52G+L53S mutant was created with SDM by mutating the His in residue 52 to a Gly of the GST-c-Jun₁₋₁₀₈ L53S mutant. For mutant GST-c-Jun₁₋₁₀₈ del47-56+63S, the residues between residue 47 and 56 were deleted using SDM. Finally, the GST-c-Jun₁₋₁₀₈ S63T mutant was

generated by mutating the native Ser at residue 63 to a Thr in the single site GST-c-Jun₁₋₁₀₈ S63 mutant. See Table I for primer sequences.

The GST- γ -c-Jun₁₋₁₀₈ mutants were generated with the Native D-site introduced at the C-terminal end of the GST-c-Jun₁₋₁₀₈ AAAA plasmid. GST-c-Jun₁₋₁₀₈ AAAA was ligated with BamH1 and Sal1 along with D-site oligos. Individual native phosphorylation sites were then re-introduced with SDM. See Table I for primer and oligo sequences. All c-Jun derivatives were confirmed by standard DNA sequencing.

TABLE 1
Oligonucleotides in this study

Name	Sequence (5' → 3')	Use
LL.cJun1-X LL.cJunX-108	GGAGGCGGTGGATCCACCATGACTGCAAAGATGGAAACGAC GCCGCTCGAGTCGACTTACTCCTGCTCATCTGTACGT	pGEX-c-Jun(1-108)
LL-Jun-S63A LL-Jun-S63A anti LL-Jun-S73A LL-Jun-S73A anti LL-Jun-T91A T93A LL-Jun-T91A T93A anti	GACCTCCTCACCGCGCCCGACGTGG CCACGTCGGGCGCGGTGAGGAGGTC CTCAAGCTGGCGGCGCCGAGCTGG CCAGCTCGGGCGCGCCAGCTTGAG ACATCACCACCGCGCCGCCCCACCCAG CTGGGTGGGGGCCGGCGCGGTGGTATGT	pGEX-c-Jun(1-108) AAAA
LL.Jun-A91T LL.Jun-A91T antisense LL.Jun-A93T sense LL.Jun-A93T antisense	ACATCACCACCACGCCGCCGCC GGGGCGGGCGTGGTGGTATGT CCACCGCGCCGACCCACCCAG CTGGGTGGGGGTGCGCGCGGTGG	pGEX-c-Jun(1-108) T91 pGEX-c-Jun(1-108) T93
LL.cJunDS1 LL.cJunDS_RC	CCCCTGGGTGCGGCCGCTTCGCGGTGCGCG- -TCTGCTCCAGGATCTCGGGTFACTGTAGCC GGCTACAGTAACCCCGAGATCCTGGAGCAGA- -GCGCGACCCGGAACGCGGCCGACCCAGTGGG	pGEX- Δ -c-Jun(1-108)
LL.9S_C LL.9S_RC LL.53.c LL.53_rc LL.81S_c LL.81S_rc	GGCGTTGAGGGCATCGCCAGGGCTGGTCGTTTCCATCTTTGCA TGCAAAGATGGAAACGACCAGCCCTGGCGATGCCCTCAACGCC CCGAGTTCTTGCGGGGCTGTGCGGCTTCAGGC GCCTGAAGCCGCACAGCCCGCCAAGAAGCTCGG CGTTGCTGGACGGGCTTATCAGGCGCTCCAGCTCG CGAGCTGGAGCGCCTGATAAGCCCGTCCAGCAACG	pGEX-c-Jun(1-108) F9S pGEX-c-Jun(1-108) L53S pGEX-c-Jun(1-108) I81S
LL.H52G_S53 LL.H52G_S53_antisense	CTTGGCGGGGCTGCCGGCTTCAGGCTC GAGCCTGAAGCCGGGACGCCCGCCAAG	pGEX-c-Jun(1-108) H52G + L53S
LL.del47-56 LL.del47-56-antisense	GGCCGACCCAGTGAAGTTCGGACCTCC GGAGGTCCGAGTTCAGTGGGTGCGCC	pGEX-c-Jun(1-108) del-S63
LL.cJun1-108S63T LL.cJun1-108S63T_rc	CCACGTCGGGCGTGGTGGAGGAGGTC GACCTCCTCACCGCCCGACGTGG	pGEX-c-Jun(1-108) S63T

3.3 Protein purification

GST fusion proteins were expressed in bacteria, purified by affinity chromatography using glutathione-Sepharose (GE Healthcare) and quantified as described elsewhere.

3.4 Protein kinase assays

Protein kinase assays were performed as described previously²⁰. Protein kinase assay reactions (20 μ L) contained Kinase Assay Buffer (New England BioLabs, 0071610), 10 mg/mL BSA, 0.5 μ M substrate, active enzyme JNK2 (10 mU/ μ L), 2 nM ATP, and 1 μ Ci of [γ -³²P] ATP. Reactions took place for 5 and 10 minutes at 30°C. Substrate phosphorylation was then determined by SDS-PAGE (12% gels). Quantification of phosphorylation was performed on a Typhoon TRIO+ Imager using phosphorimaging mode. Percent phosphorylation was determined by comparing the input with the amount incorporated and normalization to GelCode Blue Stain substrate bands. Statistical analysis of kinase assay results was performed using Welch's unequal variance t-test with two tails. This was accomplished in Microsoft Excel using the T.TEST function, setting the 'tail' option to 2, and the 'type' option to 3. All figures shown are arithmetic means of 3-6 experiments with 2-3 technical replicates.

Section 4: Results

4.1 c-Jun sites are phosphorylated at different rates

Mechanistically, the S/T on the target protein needs to bind to the activating site on the kinase before the transfer of a phosphate can take place. However, if the rate of phosphorylation is slow, then theoretically, phosphorylation would take a long time to occur, which not realistic. Here, we set out to determine the phosphorylation rates of each of the four sites on c-Jun in the region upstream of the DNA binding domain. To determine the rate of phosphorylation of single sites, GST-c-Jun₁₋₁₀₈ mutant substrates were mixed in vitro with JNK2 enzymes and radiolabeled (³²P) ATP in a kinase assay (**Fig. 2a-b**). This assay provides a readout on the ability of JNK2 to phosphorylate c-Jun mutants. At each time point (0, 10 and 20 minutes), we measure the initial phosphorylation and the rate of phosphorylation is the slope of % incorporation of (³²P) ATP for phosphorylation across on time. **Fig. 2a** includes a linear representation of each of the c-Jun mutants on which only one phospho-site is available for phosphorylation and the remaining three are mutated to phosphor-deficient sites. Schematics include WT c-Jun. As a negative control for c-Jun phosphorylation, a GST-c-Jun₁₋₁₀₈ mutant was generated with four phosphor-dead sites (not included in **Fig. 2**). When compared to the relative phosphorylation rate of WT GST-c-Jun₁₋₁₀₈, the relative phosphorylation rate of each site is significantly different from WT (**Fig. 2b-c**). The relative phosphorylation rate of the S63 and S73 mutants (those with only one Ser residue intact), appears to be higher than the relative rate of phosphorylation of the T91 and T93 mutants (**Fig. 2c**). Hence, the relative phosphorylation rates of single sites can differ.

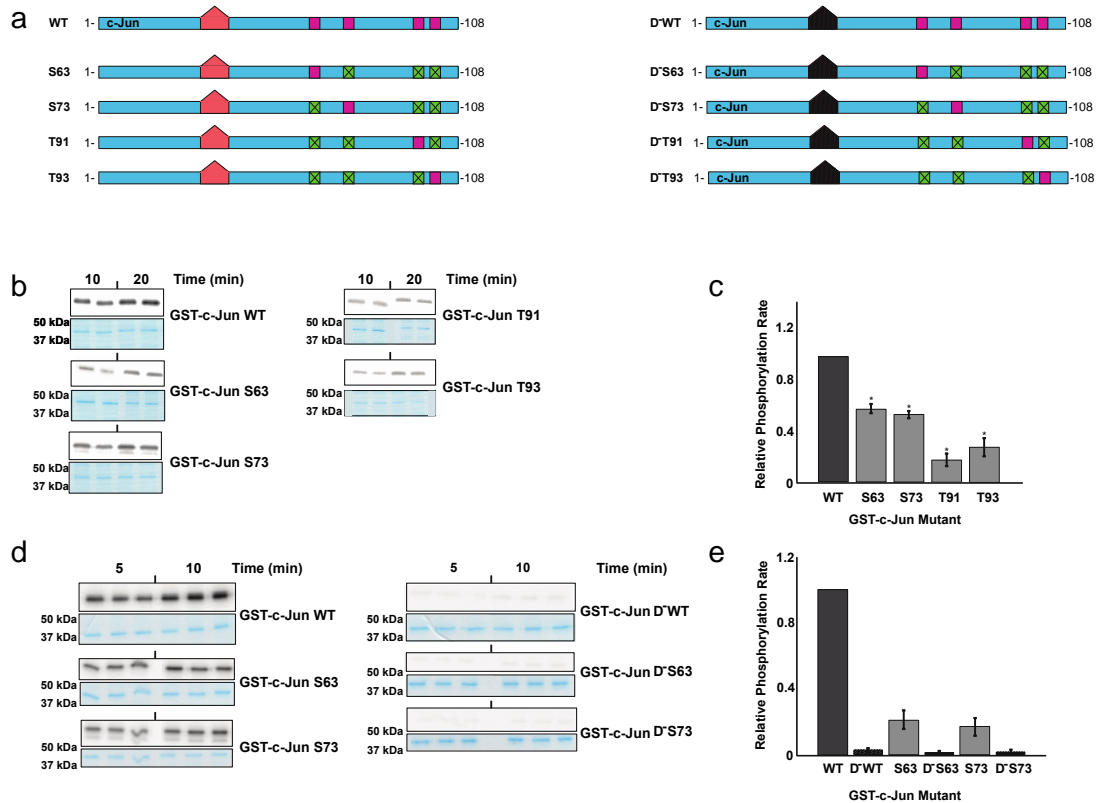


Figure 2: c-Jun sites phosphorylated at different rate and under control of D-site. (a Left) Linear representation of c-Jun₁₋₁₀₈ single site mutants. D-site is represented as pentagon, Ser and Thr sites are bright magenta boxes, and the phosphosite mutated to Ala are the green magenta boxes. (a Right) Linear representation of Δ -c-Jun mutants. The hashed pentagon represents the mutated D-site and phosphosites have same color scheme to the right panel. (b left) Each GST-c-Jun₁₋₁₀₈ mutants (1 μ M) were tested for phosphorylation by JNK2 enzyme at different two time points 10 minutes (left lanes 1-2) and 20 minutes (right lanes 3-4). Coomassie stain are beneath the radioblots showing GST-c-Jun substrates with labels on the right hand side. (c) Quantification of c-Jun phosphorylation by JNK2. Error bars represent standard error of mean (n=6). (d-e) Similar to (b-c) with the inclusion of GST- Δ -c-Jun substrates (n=3).

4.2 Mutation of D-Site decreases phosphorylation of individual native sites

Since c-Jun has a D-site on which JNK binds, does the D-site provide additional regulation of the phosphorylation of individual sites? Theoretically, if the kinase binds to the docking site during a phosphorylation event, then the substrate is tethered to the kinase, making binding to the phosphosite easier. To assess the ability for the D-site to control the phosphorylation of single sites, and to measure the relative phosphorylation rates of single sites in the absence of a functional D-site, GST- Δ -c-Jun₁₋₁₀₈ constructs were evaluated on for their phosphorylation rates with kinase assays including JNK2, GST- Δ -

c-Jun₁₋₁₀₈ mutants, and radiolabeled (³²P) ATP for different time points (0, 5, and 10 minutes). Linear representations of Δ-c-Jun₁₋₁₀₈ constructs are included in **Fig. 2a** (right).

As shown in **Fig. 2c**, c-Jun mutant with a mutated D-site and all four phosphosite intact, GST-Δ-c-Jun₁₋₁₀₈ WT, shows a decrease in phosphorylation at two different time points, consistent with previous work²⁰. The phosphorylation of GST-Δ-c-Jun₁₋₁₀₈ S63 and S73 mutants (those with only one phosphor-Ser available for phosphorylation and mutated D-site) show a decrease in phosphorylation than when the D-site is not mutated (**Fig. 2d-e**). Showing that a functional D-site is necessary for evident phosphorylation of single sites. Thus, the phosphorylation rate of Ser63 and S73 sites decreases when the D-site is not available. Similarly for GST-Δ-c-Jun₁₋₁₀₈ T91 and T93 mutants, which showed decreased phosphorylation when the D-site is mutated (Supplemental **Fig. S1**). Therefore, the phosphorylation of single sites is regulated by a functional D-site.

4.3 Addition of D-Site to C-terminal end of c-Jun yield no phosphorylation of individual native sites

Since we see that the D-site is necessary, not only for WT c-Jun phosphorylation, but for the phosphorylation of single sites, it is not clear what factors influence D-site control of phosphorylation. Does the location of the docking site relative to the phosphosites control phosphorylation? In other words, is it necessary that the D-site be N-terminal to all the phosphosites? To address this question, GST-γ-c-Jun₁₋₁₀₈ mutants were generated which have a mutated native D-site and the domain of a functional D-site is included on the C-terminal end of c-Jun (Supplemental **Fig. S1**). As seen in Supplemental **Fig. S1**, addition of the D-site to the C-terminal end of c-Jun and a mutated native D-site results in little to no phosphorylation of WT and single phospho-site c-Jun mutants. Thus, phosphorylation of single sites requires an N-terminal D-site.

4.4 Introduction of novel sites shows novel phosphorylation under control of the docking site

The D-site is necessary for the phosphorylation of all c-Jun native sites, however, not a lot is known about the evolution of new phospho-sites. More specifically, how easy is it to evolve a new site? Does a novel site exhibit similar traits as the native sites? To explore the potential phosphorylation of novel sites, c-Jun constructs were generated containing potential new phospho-sites (**Fig. 3a**). Residues FY at position 9 and 10 were mutated to SP to generate GST-c-Jun₁₋₁₀₈ F9S mutant while the native phosphosites are mutated to phosphor-dead residues (see Methods section above). Residues were mutated to SP sequences in order to maintain the S/T-P consensus target sequence required for JNK phosphorylation. Similarly with residues LR at position 53 and 54 for mutant GST-Δ-c-Jun₁₋₁₀₈ L53S and residues IQ at position 81 and 82 for mutant

GST- Δ -c-Jun₁₋₁₀₈ I81S; local sequences for each mutant is included in Fig. 3a (bottom). Similar to the assays above, phosphorylation was measured for each of these mutants that contain a new potential site with the noted difference that only one time point is measured (at 5 minutes). As seen in Fig. 3b, phosphorylation of residues F9S and L53S is weak. However, phosphorylation of I81S is apparent (Fig. 3c). This suggests that some, but not all novel sites can be phosphorylated on c-Jun. Interestingly, this brings up many questions. Why only I81S? What factors are influencing the phosphorylation of I81S? In the following section, the role of the D-site on novel site phosphorylation is explored in more detail.

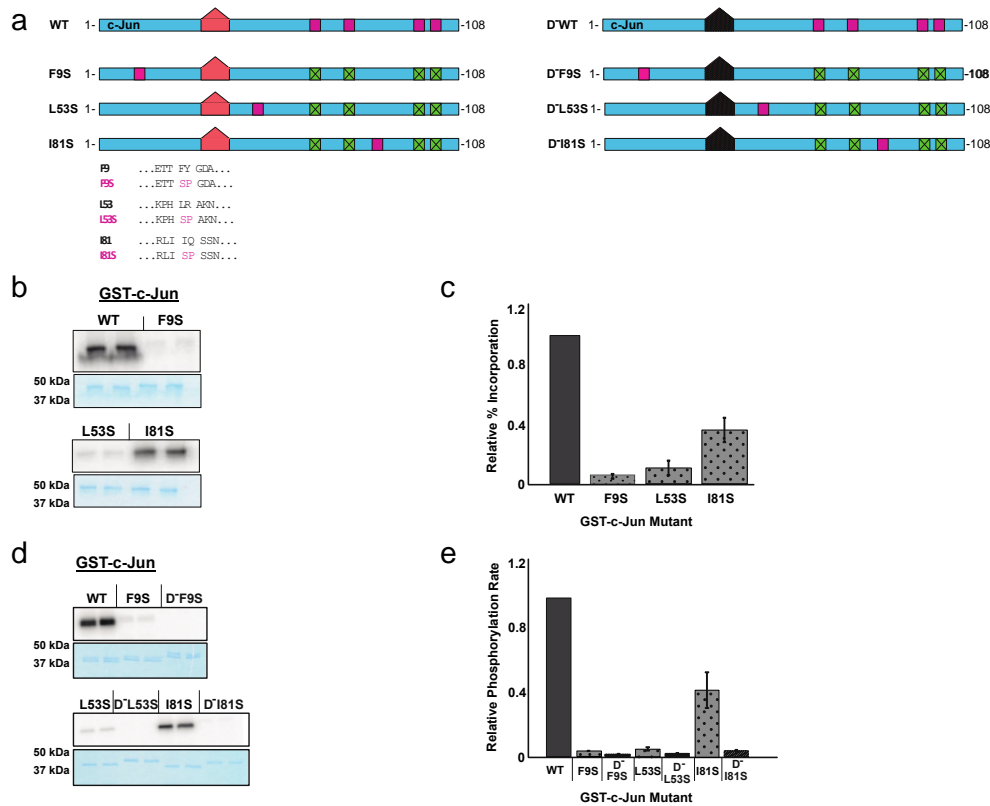


Figure 3: Novel sites phosphorylated and under control of the docking site. (a) Linear representation of c-Jun₁₋₁₀₈ with introduction of novel sites with docking site intact (left) and docking site mutated (right). Sequence surrounding the residues of interest in lower part of panel comparing WT sequence to mutated SP sequence. (b) Radioblot and Coomassie stain of kinase assays with GST-c-Jun₁₋₁₀₈ novel site mutants and JNK2 at one time point (10 minutes) in duplicates. (c) Quantification of c-Jun phosphorylation by JNK2. Error bars represent standard error of mean (n=3). (d-e) Similar to (b-c) with the inclusion of GST- Δ -c-Jun₁₋₁₀₈ novel site mutants, those with a mutated D-site. Similar to (c), (e) includes standard error of mean (n=3).

To explore the role of the D-site on the phosphorylation of novel sites on c-Jun, mutants containing novel sites and a mutated D-site were generated and tested for phosphorylation using similar kinase assays as

above. Each mutant, GST- Δ -c-Jun₁₋₁₀₈ F9S, L53S, and I81S displayed little to no phosphorylation in addition to the single novel site mutants in the previous section (Fig. 3d-e). Since GST-c-Jun₁₋₁₀₈ I81S is phosphorylated but the GST- Δ -c-Jun₁₋₁₀₈ I81S is not, the D-site influences the phosphorylation a novel site as well the native sites. Hence, the D-site is necessary for the phosphorylation of novel phospho-sites.

4.5 Surrounding sequence of novel site is not sufficient for phosphorylation

The surrounding residues to a phosphosite have been shown to play little role in phosphorylation of native sites¹⁶ (Bardwell2006 and references therein). To explore the possibility that the sequence surrounding a potential novel site could direct phosphorylation, mutant GST-c-Jun₁₋₁₀₈ H52G+L53S was tested for phosphorylation. This mutant contains and L53S residue (along with the Pro in the +1 residue) and the His52 was mutated to a Gly (see Methods section). H52 was mutated to a Gly since the logo plot for JNK2 demonstrates a preference for Gly in the -1 position (UniProtKB Reference # [P45984](#)). Local sequences for c-Jun L53, L53S, H52G+L53S, and S63 are included in Fig. 4a. In Fig. 4b-c, kinase assays reveal that addition of a replacement of His52 to a Gly are not sufficient for increasing the phosphorylation of L53S. Furthermore, a Gly in the -1 position does not necessarily control the phosphorylation of novel sites.

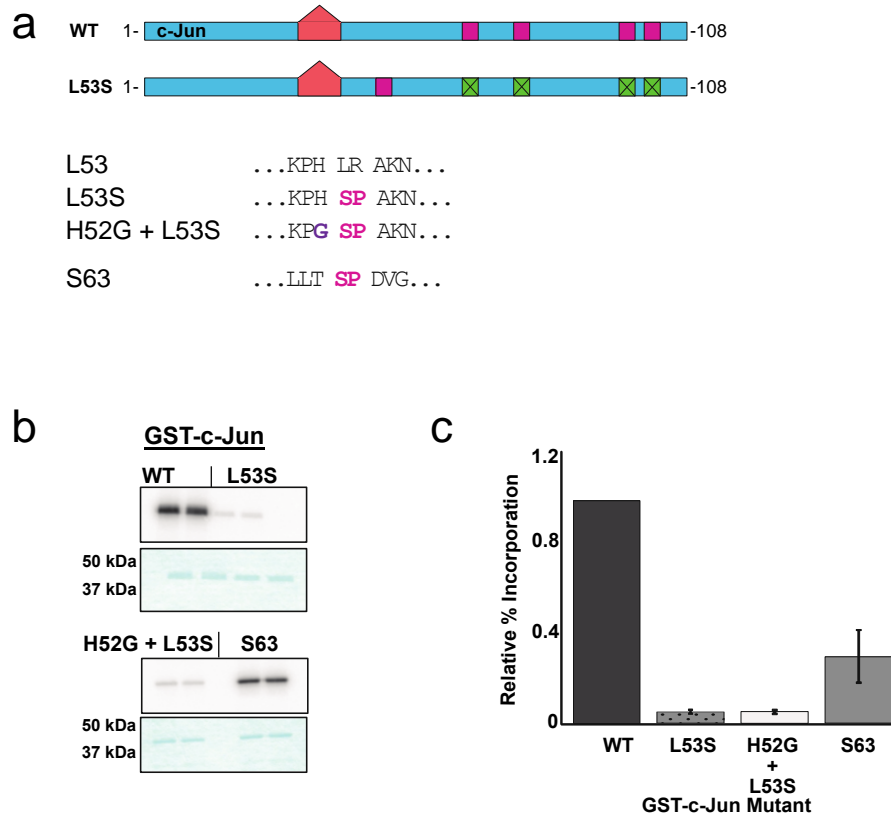


Figure 4: Surrounding sequence in potential novel site. (a) Linear representation of c-Jun₁₋₁₀₈ WT and novel site L53S with same color scheme as above. Local sequence surrounding S53 for WT L53, L53S, H52G+L53S, and S63 mutants are provided for comparison. (b) Radioblot and Coomassie stain of kinase assays including JNK2 and GST-c-Jun₁₋₁₀₈ WT (lanes 1-2), L53S (lanes 3-4), H52G+L53S (bottom lanes 1-2) and S63 (bottom lanes 3-4). (c) Quantification of c-Jun phosphorylation by JNK2. Error bars represent standard error of mean (n=3).

4.6 Distance between D-site and phosphosite influences phosphorylation of native Ser

In addition to direction and local sequences, the distance between the target site and the D-site could control the phosphorylation of novel sites. To that end, here mutant GST-c-Jun₁₋₁₀₈ del-S63 was generated such that the residues 47-56 were deleted and only S63 is available for phosphorylation (see Methods section). Linear representations of GST-c-Jun₁₋₁₀₈ WT, S63, and del-S63 are included in **Fig. 5a** along with local sequences for comparison. As seen in **Fig. 5b-c**, deletion of the region between the D-site and S63 displays a loss in phosphorylation. Suggesting that the distance between the phosphosite and the D-site play a role in phosphorylation.

4.7 Type of residue does not influence phosphorylation of native sites

To test the hypothesis that the type of residue (Ser vs Thr) does not influence phosphorylation of the target site, we mutated S63 to a Thr, mutant GST-c-Jun₁₋₁₀₈ S63T. We compare the initial phosphorylation rate of GST-c-Jun₁₋₁₀₈ WT, S63T, and S63. Since the linear schematic is similar to S63, the local sequences are instead included in **Fig. 5a**. GST-c-Jun₁₋₁₀₈ S63T mutant was tested for phosphorylation in similar *in vitro* kinase assays. Output from the assay shows that the phosphorylation of S63 is similar to that of S63T (**Fig. 5d**). Indeed, there is no discernable difference in the initial phosphorylation rate between S63 and S63T (**Fig. 5e**). Thus, if Ser63 had been a Thr, there is little to no preference for faster phosphorylation than as a Ser.

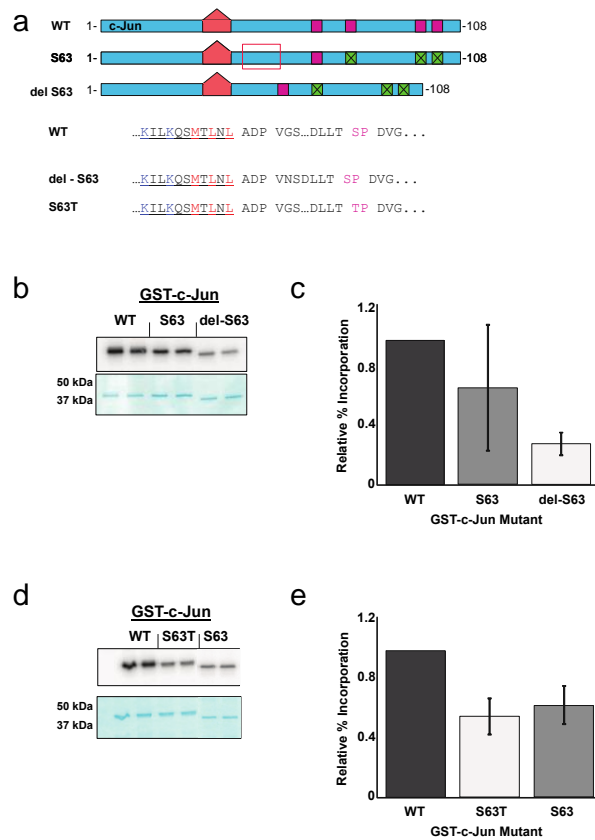


Figure 5: Comparison of native c-Jun residue type phosphorylation by JNK2 and its distance to D-site. (a) Linear representation of c-Jun₁₋₁₀₈ WT, native S63 single mutant, and mutant with residues 46-57 deleted; cartoon contains same color scheme as above. Local sequences surrounding S63 for WT, del-S63, and S63T mutants are provided for comparison. (b) GST-c-Jun₁₋₁₀₈ mutant with residues 46-57 deleted and only native S63 available was tested for phosphorylation. Radioblot and Coomassie stain show WT (lanes 1-2), S63 (lanes 3-4), and del-S63 (lanes 5-6) show phosphorylation of each mutant after 10 minute incubation time, Coomassie stain below radioblot. (c) Quantification of c-Jun phosphorylation by JNK2. Error bars represent standard error of mean (n=3). (d) GST-c-Jun₁₋₁₀₈ mutant with S63 mutated to a Thr was also tested for phosphorylation by JNK2. Radioblot shows phosphorylation of WT (lanes 1-2), S63T (lanes 3-4), and S63 (lanes 5-6) after 10 minute incubation time, Coomassie stains below radioblot showing substrate in assay. (e) Quantification of c-Jun phosphorylation by JNK2. Error bars represent standard error of mean (n=3).

Section 5: Discussion

This study examined the role of the docking site on the phosphorylation of native and novel sites on transcription factor c-Jun. We focused on the phosphorylation rates of native Ser/Thr sites and novel Ser sites. We find that the D-site plays a significant role on the phosphorylation of all sites, native and novel. Our findings highlight the importance of the D-site on kinase-substrate specificity and the efficiency of phosphorylation.

Docking-driven phosphorylation of individual sites – We find that single sites can be phosphorylated at different rates (**Fig. 1**). Closer inspection of the phosphorylation rates of the single phosphosite mutants shows the possibility that the sites display negative cooperativity. More specifically, we see that single Ser mutants are phosphorylated at high percentages (~60%) relative to WT c-Jun with all four sites available for phosphorylation. If the sites displayed positive cooperativity, one would expect to see a lower percent of phosphorylation such that all sites need to be available in order to phosphorylate any one site. However, our observations indicate that phosphorylation of one site can weakly inhibit the phosphorylation of nearby sites, i.e. negative cooperativity. This negative cooperativity can be attributed to electrostatic interference, conformational changes of the intrinsically disordered substrate that limit access to surrounding the target sites, or other aspects of phosphorylation regulation.

In general, the mutation of the D-site on c-Jun decreased the overall phosphorylation rate of each of the native sites. In other words, the D-site directly influences the phosphorylation of each native Ser/Thr site (**Fig. 2**). Suggesting that the D-site controls the phosphorylation of every site. Thus, JNK2 prefers the ability to bind to c-Jun D-site to phosphorylate Ser/Thr sites. Similar results were observed in previous work²⁰.

Evolution of novel sites and the D-site – Over the past years, there has been an increase in phosphoproteomic studies²¹. New technologies have also provided insights into the conservation of phosphosites and their evolution across species²¹. However, not much is known about the constraints of novel sites. Here, In addition to finding the relative phosphorylation rates for single phosphosites, we find the phosphorylation rates of novel sites that can be phosphorylated by JNK2. The introduced novel sites contained the S-P consensus sequence required for MAPK phosphorylation. We find that some, but not all, novel sites can be phosphorylated by JNK2 (**Fig. 3**). More specifically, the site on residue I81, was clearly phosphorylated. Interestingly, I81S was the novel in between two native sites. The novel sites at residues F9 and L53 were not discernably phosphorylated. F9S was a native site N-terminal to the D-site, which might explain the low phosphorylation. L53S, however, was a native site that was C-terminal to the D-site but N-terminal to any native phosphosite.

We set out further investigate possible factors that allow I81S to be phosphorylated by JNK2 and not L53S. Namely, we altered the +1 residue from a His to a Gly. We find that this sole mutation does not increase the efficiency of phosphorylation of this novel site (**Fig. 4**). Possible explanations for this phenomenon may revolve around the location of L53S. For example, perhaps L53S is too close to the D-site. All the sites that has been phosphorylated are at least 20 residues away from the D-site.

Additionally, perhaps a Gly in the +1 position is not sufficient for phosphorylation. Although studies have suggested that a S/T-P proline is sufficient for MAPK phosphorylation¹⁶ (Bardwell2006), this region of c-Jun perhaps requires more consensus in addition to other factors. For example, Jacobs et al. suggests that stronger binding affinity can occur with a motif closer to P-X-S/T-P²².

The mechanical or physical properties of c-Jun can also play a role in the phosphorylation of c-Jun native sites. The crystal structure for the DNA binding domain is c-Jun is available²³, however, the regulatory region has yet to be solved. This solution could shed light on the structure of c-Jun outside of the DNA binding domain. Computational software can be used to predict the structure of c-Jun such as the one described by Meszaros et al.; which predict a mostly intrinsically disordered region around all the native and novel sites²⁴. Thus, it is possible that L53S is not as accessible as the other sites for JNK2 phosphorylation.

Lastly, the type of residue may direct phosphorylation rates as well. More specifically, does a Thr in the same location as a Ser have a different rate? We find that mutating a native Ser to a Thr does not affect the overall phosphorylation of that site (**Fig. 5**). Further support that distance from the D-site or local phosphosite sequence may direct phosphorylation in addition to a function D-site.

Conclusions – In conclusion, we have investigated the role of the docking site on the phosphorylation of native and novel sites. Our findings provide insights into mechanisms that help regulate signal transduction in pathways involved in development and diseases.

References:

- [1]. Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V. & Reece, J. B. *Biology*. (Pearson, 2016).
- [2]. Chang, L., & Karin, M. (2001). Mammalian MAP kinase signaling cascades. *Nature*, *410*(6824), 37–40. <https://doi.org/10.1038/35065000>
- [3]. Zeke, A., Misheva, M., Reményi, A., & Bogoyevitch, M. A. (2016). JNK Signaling: Regulation and Functions Based on Complex Protein-Protein Partnerships. *Microbiology and Molecular Biology Reviews*, *80*(3), 793–835. <https://doi.org/10.1128/mubr.00043-14>
- [4]. Weston, C. R., & Davis, R. J. (2007). The JNK signal transduction pathway. *Current Opinion in Cell Biology*, *19*(2), 142–149. <https://doi.org/10.1016/j.ceb.2007.02.001>
- [5]. Schummer, P., Kuphal, S., Vardimon, L., Bosserhoff, A. K., & Kappelmann, M. (2016). Specific c-Jun target genes in malignant melanoma. *Cancer Biology and Therapy*, *17*(5), 486–497. <https://doi.org/10.1080/15384047.2016.1156264>
- [6]. Karin, M. (1992). Signal transduction from the BCR. *The FASEB Journal*, *6*, 2581–2590.
- [7]. Kappelmann, M., Bosserhoff, A., & Kuphal, S. (2014). AP-1/c-Jun transcription factors: Regulation and function in malignant melanoma. *European Journal of Cell Biology*, *93*(1–2), 76–81. <https://doi.org/10.1016/j.ejcb.2013.10.00>
- [8]. Dorn, C., Engelmann, J. C., Saugspier, M., Koch, A., Hartmann, A., Müller, M., ... Hellerbrand, C. (2014). Increased expression of c-Jun in nonalcoholic fatty liver disease. *Laboratory Investigation*, *94*(4), 394–408. <https://doi.org/10.1038/labinvest.2014.3>
- [9]. Silva, R. M., Kuan, C. Y., Rakic, P., & Burke, R. E. (2005). Mixed lineage kinase-c-jun N-terminal kinase signaling pathway: A new therapeutic target in Parkinson's disease. *Movement Disorders*, *20*(6), 653–664. <https://doi.org/10.1002/mds.20390>
- [10]. Ferrer, I., Seguí, J., & Planas, A. M. (1996). Amyloid deposition is associated with c-Jun expression in Alzheimer's disease and amyloid angiopathy. *Neuropathology and Applied Neurobiology*, *22*(6), 521–526. <https://doi.org/10.1111/j.1365.2990.1996.tb01130.x>
- [11]. Filosto, M., Tonin, P., Vattemi, G., Savio, C., Rizzuto, N., & Tomelleri, G. (2003). Transcription factors c-Jun/activator protein-1 and nuclear factor-kappa B in oxidative stress response in mitochondrial diseases. *Neuropathology and Applied Neurobiology*, *29*(1), 52–59. <https://doi.org/10.1046/j.1365.2990.2003.00411.x>
- [12]. Darnell, J. E. (2002). Transcription factors as targets for cancer therapy. *Nature Reviews Cancer*, *2*(10), 740–749. <https://doi.org/10.1038/nrc906>
- [13]. Herdegen, T., Skene, P., & Bähr, M. (1997). The c-Jun transcription factor - Bipotential mediator of neuronal death, survival and regeneration. *Trends in Neurosciences*, *20*(5), 227–231. [https://doi.org/10.1016/S0166-2236\(96\)01000-4](https://doi.org/10.1016/S0166-2236(96)01000-4)
- [14]. Nateri, A. S., Spencer-Dene, B., & Behrens, A. (2005). Interaction of phosphorylated c-Jun with TCF4 regulates intestinal cancer development. *Nature*, *437*(7056), 281–285. <https://doi.org/10.1038/nature03914>
- [15]. Davis, R. J. (1993). The mitogen-activated protein kinase signal transduction pathway. *Journal of Biological Chemistry*, *268*(20), 14553–14556.
- [16]. Bardwell, L. (2006). Mechanisms of MAPK signaling specificity. *Biochemical Society Transactions*, *34*(5), 837–841. <https://doi.org/10.1042/BST0340837>
- [17]. Songyang, Z., Lu, K. P., Kwon, Y. T., Tsai, L. H., Filhol, O., Cochet, C., Cantley, L. C. (1996). A structural basis for substrate specificities of protein Ser/Thr kinases: primary sequence preference of casein kinases I and

- II, NIMA, phosphorylase kinase, calmodulin-dependent kinase II, CDK5, and Erk1. *Molecular and Cellular Biology*, 16(11), 6486–6493. <https://doi.org/10.1128/mcb.16.11.6486>
- [18]. Sharrocks, A. D., Yang, S. H., & Galanis, A. (2000). Docking domains and substrate-specificity determination for MAP kinases. *Trends in Biochemical Sciences*, 25(9), 448–453. [https://doi.org/10.1016/S0968-0004\(00\)01627-3](https://doi.org/10.1016/S0968-0004(00)01627-3)
- [19]. Bardwell, A. J., Frankson, E. & Bardwell, L. (2009). Selectivity of docking sites in MAPK kinases. *J. Biol. Chem.* 284, 13165–13173
- [20]. Bardwell, A. J. & Bardwell, L. (2015). Two hydrophobic residues can determine the specificity of Mitogen-activated protein Kinase docking interactions. *J. Biol. Chem*, 26661–26674
- [21]. Gnad, F., Ren, S., Cox, J., Olsen, J. V., Macek, B., Oroshi, M., & Mann, M. (2007). PHOSIDA (phosphorylation site database): Management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biology*, 8(11). <https://doi.org/10.1186/gb-2007-8-11-r250>
- [22]. Jacobs, D., Glossip, D., Xing, H., Muslin, A. J. & Kornfeld, K. (1999). Multiple docking sites on substrate proteins form a modular system that mediates recognition by ERK MAP kinase. *Genes Dev.* 13, 163–175
- [23]. Chen, L., Glover, M., Hogan, P., Rao, A., Harrison, S. C., Szilak, L., ... Vinson, C. (1999). Structure of the DNA-binding domains from NFAT, FOS and JUN bound specifically to DNA. *Chemtracts*, 12(10), 768–773.
- [24]. Mészáros, B., Erdős, G., & Dosztányi, Z. (2018). IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Research*, 46(W1), W329–W337. <https://doi.org/10.1093/nar/gky384>
- [25]. Kallunki, T., Deng, T., Hibi, M., & Karin, M. (1996). c-Jun can recruit JNK to phosphorylate dimerization partners via specific docking interactions. *Cell*, 87(5), 929–939. [https://doi.org/10.1016/S0092-8674\(00\)81999-6](https://doi.org/10.1016/S0092-8674(00)81999-6)

SUPPLEMENTAL FIGURE

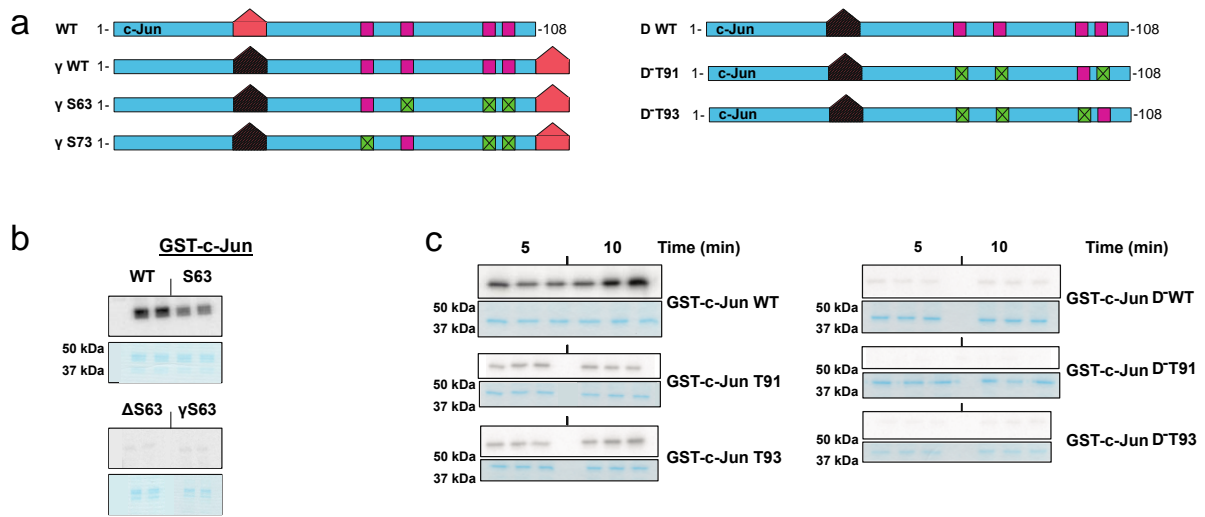


Figure S1: Orientation of D-site determines c-Jun phosphorylation of single sites and Thr sites are under control of D-site. (a Left) Linear representation of the γ -c-Jun₁₋₁₀₈ single site mutants which have an added D-site on the C-terminal end of c-Jun. The hashed pentagon represents the mutated D-site and the solid pentagon is the native D-site. Ser and Thr sites are bright magenta boxes, and the phosphosite mutated to Ala are the green magenta boxes. (a Right) Linear representation of Δ -c-Jun mutants which have only either Thr91 or Thr93 available for phosphorylation. (b) Each GST- γ -c-Jun₁₋₁₀₈ mutant (1 μ M) was tested for phosphorylation by JNK2 enzyme. Coomassie stain are beneath the radioblots showing GST-c-Jun substrates with labels on the right hand side. (c) Similar kinase assays to (b) with the inclusion of GST- Δ -c-Jun substrates for Thr91 and Thr93 mutants (n=1).

Chapter 5: Conclusion and Discussion

In summary, I have presented in this dissertation different mechanisms for signal transduction regulation in cells. Proper function of signal transduction is crucial for proper development or to prevent diseases. Mechanisms that regulate transduction cascades include scaffolding, binding, and Post-Translational Modifications (PTMs). Therefore, targeting these mechanisms can further expand our understanding of signal transduction. The Mitogen-Activated Protein Kinase (MAPK) cascade is one of the most involved cascades. The MAPK cascade regulates cell responses such as cell proliferation, division, survival and death¹. Since the MAPK cascade is so prominent, understanding different aspects of its regulation is important for our advancement in medicine and development. There are many different mechanism that regulate the MAPK cascade such as PTMs, ligand or protein binding, and scaffolding.

Section 1: Scaffold Protein IQGAP1 and cell signaling

This work includes a study on scaffolding as a mechanism of signal transduction regulation. IQGAP1 is a scaffold protein that interacts number of signaling/regulatory pathways, including the MAPK cascade proteins². IQGAP1 has also been a recent therapeutic target in cancer^{3,4}. As one can imagine, it is important that to generate successful therapies we need to understand which domains on IQGAP1 bind to the proteins in the MAPK cascade.

IQGAP1 has different domains, seen in Figure 1 of Chapter 1. Two key domains of IQGAP1 are the WW domain and the IQ domain, which consists of four closely spaced IQ motifs. The work included in Chapter 2 of this dissertation shows that the IQ domain of IQGAP1 is both necessary and sufficient for binding to ERK2, a MAP kinase. However, the WW domain was not necessary nor sufficient for ERK2 binding. This is interesting when considering the fact that previous work suggested otherwise. It fact, it has been believed that that ERK1 and ERK2 bind to the WW domain of IQGAP1^{5,6}. Some possible explanations for this contradiction are included in the Discussion section of Chapter 2.

The observation that IQGAP-ERK binding depends on the presence of the IQ domain and not the WW domain asks for a re-evaluation of IQGAP1 binding to MAP kinases. This study provides further understanding on how scaffolds work to regulate signal transduction. In other words, the IQ domain binds to ERK1 and ERK2, perhaps the fact that IQGAP1 has different domains that bind to different MAP kinases helps these kinases find each other. In theory, IQGAP1 brings these kinases together for efficient and timely cell response. Thus, knowing which domains of IQGAP1 bind to which components of the MAPK cascade is crucial for successful development of new therapies. Additionally, this knowledge can provide

new insights into synthetic biology as a mechanism for generating scaffold proteins to target specific kinases. Further discussion is also included in the Discussion section of Chapter 2.

As new technologies are developed, the idea is that one day we should be able to easily determine if IQGAP1 binds to different MAPK-pathway kinases simultaneously or in a more sequential manner. This would help determine which domains need to be directly inhibited in treatment development. Furthermore, one speculation of potential future work includes the usability of CRISPR tools to measure the effect each IQGAP1 domain has on tumorigenesis. In 2013, Jameson et al. used a knockout mouse model to show that the WW domain was sufficient for inhibiting tumorigenesis and bypasses drug resistance⁷. However, our results suggest that this needs to be re-evaluated or further explored. With a CRISPR system, it would be more feasible to see that the IQ domain is playing a bigger role in tumorigenesis. In theory, editing mouse genomes to supply lines with and without certain IQGAP1 domains can prove to be a powerful tool, not only for more rigorous mouse studies, but as a tool to prevent tumorigenesis.

Section 2: Aspects of PTMs in Protein Regulation Through Mathematical Modeling

In addition to scaffolding, cells also rely on different mechanisms to relay a signal for cell function. To regulate signal transduction, different proteins use a wide variety of PTMs such as phosphorylation, methylation, and ubiquitination⁸. A short list is also included in Figure 1 of Chapter 1. Proteomic studies also show that a large number of proteins are multisite⁹, meaning they can be modified on more than one location. How multisite PTMs regulate protein activity remains a question in the field.

Different factors that influence PTMs include rates of modification, binding affinities, scaffolding, processivity, and order of modifications, as discussed in Chapter 1. Since so many proteins are multisite, unpacking everything there is to know about PTMs can be challenging. However, mathematical models are useful tools to theoretically test hypotheses involving these factors and protein activity. For example, numerous mathematical models have been brought forth describing phosphorylation-dephosphorylation cycles¹⁰. Interactions between sites also plays a role in protein activity. Site cooperativity is a well-studied phenomenon, first described in hemoglobin. Cooperativity is the behavior when the modification of one site, greatly enhances (positive) or inhibits (negative) the modification of a second site. In the past, mathematical models have described cooperative or allosteric protein behavior such as the Hill model¹¹, the Adair model¹², Koshland-Nemethy-Filmer (KNF)¹³, and the Monod-Wyman-Changeux (MWC)¹⁴ model. The MWC allosteric mathematical model is a well-known model for cooperative proteins like

hemoglobin¹⁴. A general schematic for this model is included in Chapter 1 Section 4, where a substrate can exist in one of two conformations with different modification forms.

This dissertation includes a mathematical model of PTMs and protein activity in the context of cooperativity. A generalized MWC model was built that allows for variable conformational free energy at distinct sites and associate a so-called activation parameter to each site. The main question pertains to how this activation parameters affects protein response, mainly the ultrasensitivity of the response. This model shows that that the ultrasensitivity, measured with a Hill number¹⁵, generally decreases with increasing activation parameter values and depends on their mean and not on their variability. In other words, variability in conformational free energies does not significantly change ultrasensitivity compared to balanced conformational free energies between sites. This was not something that could be explored with the classic MWC model, which assumes that all sites contribute equally to the activation of the substrate.

Additionally, our model predicts that there is an optimal Hill number that will be reached for a fixed number of sites. Since ultrasensitivity levels off after some total conformational free energy, our model suggests that in order for a substrate to increase its degree of ultrasensitivity, it can evolve a new site to do so. Again, the original MWC model could not provide any insight on that. Mainly because the original MWC assumes all sites contribute the same conformational free energy; thus, if a new site were to evolve, it must also contribute the same as the native sites. However, it is not clear if in cells that is the case; does a newly evolve site contribute the same to the substrate activation and does it have the same modification properties? Ideally, this model is generalized to address those questions.

This model also brings forth a need for an area that is not closely studied; determining the conformational free energy of distinct sites can give new information regarding regulatory sites and protein response. There are tools available today that help measure free energy of a state, measuring the conformational free energy in phosphorylation¹⁶, protein-protein binding¹⁷, and protein-membrane binding^{18,19}. However, as new technologies emerge, measuring the conformational free energy will become more feasible. Theoretically, this model can be expanded to fit the needs of the field.

For completeness, a mathematical model in which there is no cooperativity between sites is also introduced. This model proves to be more complex than the MWC model but provides different insights. In fact, model shows that ultrasensitivity can vary if the conformational free energies between sites also varies. This is contrary to the generalized MWC model. One explanation could be that if the modification

of one site is independent of whether or not a previous site is modified, this allows the system to be less specific. In other words, independence between sites allows the substrate to respond quickly at low kinase concentrations, in the context of multisite phosphorylation. Independence between sites could reduce the system to behave similar to a Michaelis-Menten system or a one-site system. One-site systems are known to produce graded response curves which are not ultrasensitive²⁰. There are multisite systems that have also displayed graded responses. For example, Lee et al. show that multisite phosphorylation of p53 provides a mechanism for a non-ultrasensitive response when binding to CBP/p300 domains²¹. Although it is unclear if the phosphorylation sites on p53 discussed in Reference²¹ act independently, our non-allosteric model can be used to recapitulate the behavior seen. In fact, both models discussed here can be expanded to multisite proteins whether they are phosphorylated or obtain another PTM.

These results provide insights into the performance objectives of multiple modification/binding sites and thus help gain a greater understanding of signaling and its role in diseases. As new mathematical tools and computational methods arise, simulations will perhaps one day become standard in biological studies. Furthermore, a combination of high-throughput techniques and mathematical modeling hold a promising approach for unpacking PTMs as cell signaling mechanisms.

Section 3: Transcription Factor c-Jun Regulation

The c-Jun N-terminal Kinase (JNK) cascade is one pathway that contains critical proteins for cell function²². The JNK pathway contains kinases that are multiply phosphorylated during transduction²³. A general schematic of the JNK pathway is included in Figure 5 of Chapter 1, where transcription factor c-Jun is one of the targets of the pathway. Transcription factor c-Jun is a substrate with multiple phosphorylations²³.

The response depends on which set of sites are phosphorylated. Put simply, N-terminal phosphorylation is associated with transcription²⁴⁻²⁷ while C-terminal phosphorylation is associated with c-Jun degradation^{28,29}. Knowing some of the downstream consequences of c-Jun phosphorylation provides important information on JNK pathway regulation in cell death and proliferation. However, there are key aspects of phosphorylation that remain unveiled: the role of the docking site on the phosphorylation rates of single sites and the effect of novel sites on c-Jun activation.

How substrates find their respective targets remains an interesting question in the field. Docking sites have been theorized to play a role in that specificity. The docking site (D-site) is an important domain on MAP kinases for kinase-substrate specificity³⁰. Studies have shown that c-Jun D-site accessibility is necessary for c-Jun phosphorylation by JNK2 enzyme³¹. Since c-Jun requires the D-site for JNK2

phosphorylation, and JNK2 phosphorylates the four N-terminal c-Jun sites, determining the role of the D-site on the phosphorylation of single c-Jun sites can provide new information regarding c-Jun activity regulation.

The native sites on c-Jun are: Ser63, Ser73, Thr91, Thr93, Thr239 and Ser243. In this dissertation, I discuss a study on the phosphorylation of the four N-terminal sites of transcription factor c-Jun. Using *in vitro* kinase assays and JNK2 enzyme, we show that the phosphorylation rates of single native c-Jun sites can differ from one to the other. Additionally, results suggest that the D-site controls the phosphorylation of every single site. Suggesting, that the D-site is necessary for phosphorylation of any site. Previous work showed that the D-site is necessary for c-Jun phosphorylation^{23,30,32} but it was unclear if each site was under the control of the D-site. Here we show that they are. In fact, even with the introduction of novel sites that contains the same S-P motif, the D-site controls the phosphorylation of any potential new site.

As mentioned previously, the effect of novel sites on c-Jun activation is poorly understood. The MAPK cascade is conserved across species, from yeast to humans and MAP kinases are expressed in all eukaryotic cells³³. Additionally, D-sites in MAP kinases are conserved across species³⁴ as well as the S/T-P motif required for phosphorylation³³. Further, the JNK pathway is also conserved between species, at least between mice and humans as of Reference³³. However, the D-site on c-Jun is poorly conserved between c-Jun and its family proteins JunB, JunD³⁵ and in oncogenic v-Jun²⁸. v-Jun also contains two missense mutations in Ser243 (a known c-Jun phosphorylation site) and Cys269²⁸. Other JNK targets, namely Bc12 family members, also do not have conserved phosphorylation sites, while in c-Jun, the four N-terminal phosphorylation sites are conserved²³. In general, proteomic sequence alignment suggests that many phosphorylation sites evolve rapidly³⁶.

My prediction regarding what possible functions the evolution of a novel site on c-Jun could have are as follows. All of the introduced novel sites are N-terminal to the native Thr239 and Ser243 and are relatively close to the native Ser63, Ser73, Thr91, and Thr93. My prediction is that if a new site evolved in this region between the D-site and Thr93, that site would also contribute to cell survival, much like the native N-terminal sites. If the conformational free energy of that novel site were favorable, I suspect that c-Jun would have a more ultrasensitive response than without the new site, even if the ultrasensitivity increased a tiny amount, every little bit counts. This could be tested in cell lines expressing c-Jun with all native sites and with a c-Jun mutant containing a novel site.

Our results suggest that the distance between the D-site and a phosphosite also determines phosphorylation. More specifically, if the distance between the D-site and a native site is too small, there is a loss in phosphorylation of the native site. In theory, there could be an optimal distance between the D-site and phosphosites that depends on the overall structure of c-Jun. If the D-site is too close to the phosphosite, c-Jun could be folding in such a way that prevents access of the site to JNK. Determining the structure of c-Jun and c-Jun mutants could help address this as new biochemical tools become available.

Lastly, the D-site and the phosphosites, and the distance between them, are highly conserved in c-Jun orthologs, specifically in five vertebrate classes, as seen in **Table 3** of Chapter 1. Since the main difference between orthologs lies in the location of key residues, such as human Ser63 vs Ser59 in chicken, it is possible that the structure is different in these species. In theory, different structure could yield different function for c-Jun orthologs. The surrounding sequence of all the phosphosites is generally very well conserved in orthologs, suggesting that perhaps function is very similar. I believe that since c-Jun regulates such crucial cell responses, like proliferation and cell death, c-Jun regulation needs to be robust. That robustness is carried through evolution in different species by conserving the phosphosites and the docking site. This theory needs to be tested, of course, and with emerging technologies, the hope is that unlocking c-Jun regulation will provide new insights into evolution and pathology.

Determining the structure of c-Jun can provide aid in developing new drugs. For example, the JNK kinase inhibitors were designed using high-throughput screening and X-ray crystal structure information to target the JNK1-ATP binding site³⁷ or the docking site³⁸. These inhibitors can be used to develop therapies in different diseases^{37,38}. Additionally, the crystal structure of c-Jun homodimer³⁹ and c-Fos-c-Jun heterodimer⁴⁰ has been solved, these structures only include the DNA binding domain. The DNA binding domain of c-Jun is arguably the active region of the transcription factor. However, the regulatory region contains the phosphorylation sites and determining the structure of this region will provide new insights into fundamental cell functions c-Jun regulates such as cell death and survival. To the best of my ability, complete structures of c-Jun, for any biological model, are yet to be determined. As new technologies arise such as NMR, determining protein structures will one day become feasible and cost-effective. Allowing the scientific community to ask questions we cannot comprehend today.

Section 4: Conclusions

A possible experiment that can be designed to complement the MWC model generalized in this thesis would be to measure c-Jun response as the concentration of JNK enzyme increases to obtain a dose

response. Defining c-Jun response can pose a challenge since the phosphorylation of the two C-terminal sites regulate c-Jun degradation²⁸ and the phosphorylation of the N-terminal sites regulate transcriptional activity²³. However, for the experimental data shown here, a possible experiment to determine the ultrasensitivity of c-Jun would be to measure the effective binding of c-Jun mutants to DNA with increasing JNK concentration. This would help determine how single sites influence ultrasensitivity. Thus allowing a fit of this data to the generalized MWC model to predict possible activation parameters and ultimately the conformational free energy contributions of authentic sites. Further, we can compare c-Jun activity when a novel site is present and measure the conformational free energy contribution of a novel site as well. This could provide new insights into c-Jun regulation and possibly identify specific sites that contribute to disease or developmental irregularities for targeting.

Potential future directions the generalized MWC model can take include addition of a docking site. Binding to the D-site is not necessarily a PTM but clearly plays a big role in the regulation of multisite proteins like the MAP kinases. In theory, binding to the D-site can be considered an allosteric reaction, where the binding to the D-site could create a conformational change in the protein that allows the target sites to be accessible for modification. This would suggest that D-site also contributes to the activation of the protein although it is not a PTM. Additionally, the generalized MWC expanded further and in combination with a KNF model, where the modification of a protein subunit determines if that subunit alone is relaxed or tense. More specifically, a combination of the generalized MWC and KNF models would include a multi-subunit protein where each site could have a unique modification efficiency and conformational free energy contribution that determine the state (active/relaxed or inactive/tense). This would be more like generalizing the KNF model with an activation parameter included for each site. This would give a broader picture of how proteins and protein complexes, such as chaperone proteins, are regulated through multiple modifications or subunits and their thermodynamic properties.

A combination of theoretical and experimental methods can provide new insights on signal transduction regulation. Collectively, this work provides some new insights to the field of signal transduction regulation. First, scaffold protein IQGAP1 specific domains bind to MAP kinases, ERK1 and ERK2. Second, the conformational free energies of distinct sites play a role in creating ultrasensitive dose response curves. Third, transcription factor c-Jun phosphorylation sites can be phosphorylated at different rates. Furthermore, the D-site controls the phosphorylation of both native and novel c-Jun sites. Moreover, the contributions of this dissertation allow for better understanding of signal transduction regulation.

References:

1. Pearson, G. *et al.* Mitogen-activated protein (MAP) kinase pathways: Regulation and physiological functions. *Endocr. Rev.* **22**, 153–183 (2001).
2. Hedman, A. C., Smith, J. M. & Sacks, D. B. The biology of IQGAP proteins: beyond the cytoskeleton. *EMBO Rep.* **16**, 427–446 (2015).
3. Stuart, D. D. & Sellers, W. R. Targeting RAF-MEK-ERK kinase-scaffold interactions in cancer. *Nat. Med.* **19**, 538–540 (2013).
4. Sanchez-Laorden, B., Viros, A. & Marais, R. Mind the IQGAP. *Cancer Cell* **23**, 715–717 (2013).
5. Roy, M., Li, Z. & Sacks, D. B. IQGAP1 Binds ERK2 and Modulates Its Activity. *J. Biol. Chem.* **279**, 17329–17337 (2004).
6. Roy, M., Li, Z. & Sacks, D. B. IQGAP1 Is a Scaffold for Mitogen-Activated Protein Kinase Signaling. *Mol. Cell. Biol.* **25**, 7940–7952 (2005).
7. Jameson, K. L. *et al.* IQGAP1 scaffold-kinase interaction blockade selectively targets RAS-MAP kinase-driven tumors. *Nat. Med.* **19**, 626–630 (2013).
8. Morse, A. M., Whetten, R. W., Dubos, C. & Campbell, M. M. Post-translational modification of an R2R3-MYB transcription factor by a MAP Kinase during xylem development. *New Phytol.* **183**, 1001–1013 (2009).
9. Li, H. *et al.* SysPTM: A systematic resource for proteomic research on post-translational modifications. *Mol. Cell. Proteomics* **8**, 1839–1849 (2009).
10. Salazar, C. & Höfer, T. Multisite protein phosphorylation - From molecular mechanisms to kinetic models. *FEBS J.* **276**, 3177–3198 (2009).
11. Hill, A. V. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J. Physiol.* 4–6 (1910). doi:10.1093/nq/36.2.211d
12. Adair, G. S. The Hemoglobin System V1. The oxygen dissociation curve of hemoglobin. *J. Biol. Chem.* (1925).
13. Koshland, D. E., Nemethy, J. G. & Filmer, D. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry* **5**, 365–385 (1966).
14. Monod, J., Wyman, J. & Changeux, J. P. On the Nature of Allosteric Transitions: a Plausible Model. *J. Mol. Biol.* **12**, 88–118 (1965).
15. Golbeter, A. & Koshland, D. E. J. An amplified sensitivity arising from covalent modification in biological systems. *PNAS* **78**, 6840–6844 (1981).
16. Shen, T., Wong, C. F. & McCammon, J. A. Atomistic Brownian dynamics simulation of peptide phosphorylation. *J. Am. Chem. Soc.* **123**, 9107–9111 (2001).
17. Nishi, H., Hashimoto, K. & Panchenko, A. R. Phosphorylation in protein-protein binding: effect on stability and function. *Structure* **19**, 1807–1815 (2011).
18. Murray, D. *et al.* Electrostatics and the membrane association of Src: Theory and experiment. *Biochemistry*

- 37, 2145–2159 (1998).
19. Serber, Z. & Ferrell, J. E. Tuning Bulk Electrostatics to Regulate Protein Function. *Cell* **128**, 441–444 (2007).
 20. Prabakaran, S., Lippens, G., Steen, H. & Gunawardena, J. Post-translational modification: Nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **4**, 565–583 (2012).
 21. Lee, C. W., Ferreon, J. C., Ferreon, A. C. M., Arai, M. & Wright, P. E. Graded enhancement of p53 binding to CREB-binding protein (CBP) by multisite phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 19290–19295 (2010).
 22. Chang, L. & Karin, M. Mammalian MAP kinase signalling cascades. *Nature* **410**, 37–40 (2001).
 23. Zeke, A., Misheva, M., Reményi, A. & Bogoyevitch, M. A. JNK Signaling: Regulation and Functions Based on Complex Protein-Protein Partnerships. *Microbiol. Mol. Biol. Rev.* **80**, 793–835 (2016).
 24. Vogt, P. K. Jun, the oncoprotein. *Oncogene* **20**, 2365–2377 (2001).
 25. Morton, S., Davis, R. J., McLaren, A. & Cohen, P. A reinvestigation of the multisite phosphorylation of the transcription factor c-Jun. *EMBO J.* **22**, 3876–3886 (2003).
 26. Zhou, X. *et al.* Exopolysaccharides from *Lactobacillus plantarum* NCU116 induce c-Jun dependent Fas/FasL-mediated apoptosis via TLR2 in mouse intestinal epithelial cancer cells. *Sci. Rep.* **7**, 1–13 (2017).
 27. Szymanska, E., Skowronek, A. & Miaczynska, M. Impaired dynamin 2 function leads to increased AP-1 transcriptional activity through the JNK/c-Jun pathway. *Cell. Signal.* **28**, 160–171 (2016).
 28. Wei, W., Jin, J., Schlisio, S., Harper, J. W. & Kaelin, W. G. The v-Jun point mutation allows c-Jun to escape GSK3-dependent recognition and destruction by the Fbw7 ubiquitin ligase. *Cancer Cell* **8**, 25–33 (2005).
 29. Papavassiliou, A. G., Treier, M. & Bohmann, D. Intramolecular signal transduction in c-Jun. *EMBO J.* **14**, 2014–2019 (1995).
 30. Sharrocks, A. D., Yang, S. H. & Galanis, A. Docking domains and substrate-specificity determination for MAP kinases. *Trends Biochem. Sci.* **25**, 448–453 (2000).
 31. Bardwell, A. J., Frankson, E. & Bardwell, L. Selectivity of docking sites in MAPK kinases. *J. Biol. Chem.* **284**, 13165–13173 (2009).
 32. Bardwell, L. & Shah, K. Analysis of mitogen-activated protein kinase activation and interactions with regulators and substrates. *Methods* **40**, 213–223 (2006).
 33. Widmann, C., Gibson, S., Jarpe, M. B. & Johnson, G. L. Mitogen-activated protein kinase: Conservation of a three-kinase module from yeast to human. *Physiol. Rev.* **79**, 143–180 (1999).
 34. Bardwell, A. J., Flatauer, L. J., Matsukuma, K., Thorner, J. & Bardwell, L. A Conserved Docking Site in MEKs Mediates High-affinity Binding to MAP Kinases and Cooperates with a Scaffold Protein to Enhance Signal Transmission. *J. Biol. Chem.* **276**, 10374–10386 (2001).
 35. Gupta, S. *et al.* Selective interaction of JNK protein kinase isoforms with transcription factors. *EMBO J.* **15**, 2760–2770 (1996).

36. Tan, C. S. H. *et al.* Comparative analysis reveals conserved protein phosphorylation networks implicated in multiple diseases. *Sci. Signal.* **2**, (2009).
37. Stocks, M. J. *et al.* Structure-driven HtL: Design and synthesis of novel aminoindazole inhibitors of c-Jun N-terminal kinase activity. *Bioorganic Med. Chem. Lett.* **15**, 3459–3462 (2005).
38. De, S. K. *et al.* Design, synthesis, and structure-activity relationship of substrate competitive, selective, and in vivo active triazole and thiadiazole inhibitors of the c-jun n-terminal kinase. *J. Med. Chem.* **52**, 1943–1952 (2009).
39. Junius, F. K., O'Donoghue, S. I., Nilges, M., Weiss, A. S. & King, G. F. High resolution NMR solution structure of the leucine zipper domain of the c-Jun homodimer. *J. Biol. Chem.* **271**, 13663–13667 (1996).
40. Glover, J. N. M. & Harrison, S. C. Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature* **373**, 257–261 (1994).