Title

Construction of Synthetic Signaling Circuits by Modular Recombination

Permalink

https://escholarship.org/uc/item/7bj9v0b6

Author

Yeh, Brian Jen-Chang

Publication Date

2007-06-15

Peer reviewed|Thesis/dissertation

Construction of Synthetic Signaling Circuits by Modular Recombination

by

Brian Jen-Chang Yeh

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry and Chemical Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

To Mom, Dad, and Michelle

Acknowledgements

Current Opinion in Structural Biology 14, 690-699 (2004). (2) Reproduced in part with permission from the *Annual Review of Biochemistry*, Volume 75, ©2006 by Annual Reviews, http://www.annualreviews.org: Bhattacharyya, R.P., Reményi, A., Yeh, B.J., and Lim, W.A. Domains, Motifs, and Scaffolds: The Role of Modular Interactions in the Evolution and Wiring of Cell Signaling Circuits. *Annual Review of Biochemistry* 75, 655-680 (2006).

Chapter 2 is reproduced with permission from AAAS (http://dx.doi.org/10.1126/science.1085945): Dueber, J.E., Yeh, B.J., Chak, K., and Lim, W.A. Reprogramming Control of an Allosteric Signaling Switch Through Modular Recombination. *Science* 301, 1904-1908 (2003).

Chapter 3 is reproduced with permission from Nature (http://dx.doi.org/10.1038/nature05851): Yeh, B.J., Rutigliano, R.J., Deb, A., Bar-Sagi, D., and Lim, W.A. Rewiring cellular morphology pathways with synthetic guanine nucleotide exchange factors. *Nature* 447, 596-600 (2007).

Chapter 4 is reproduced from a Commentary submitted to *Nature Chemical Biology*: Yeh, B.J. and Lim, W.A. Synthetic Biology: Lessons from the History of Synthetic Organic Chemistry.

Construction of Synthetic Signaling Circuits by Modular Recombination

Brian Jen-Chang Yeh

(signature)

Wendell A. Lim, Ph.D.

Abstract

Living cells integrate information from their external environments and demonstrate a wide range of sophisticated behaviors. Most of the rapid responses exhibited by cells are mediated by circuits composed of interconnected signal transduction proteins. What mechanisms allow these proteins and circuits to respond precisely in space and time? Moreover, how do signaling networks evolve, producing new relationships between signals and responses? We chose to address these questions using a synthetic biology approach by engineering signaling proteins with novel input-output relationships.

Many signaling proteins are composed of both catalytic domains and interaction domains, which are physically and functionally modular: the domains can be separated and function in different contexts. This modular structure has led to the hypothesis that new input-output relationships could be generated by recombining catalytic domains with alternative interaction domains. We first tested this hypothesis by engineering variants of

the actin regulatory protein N-WASP (neuronal Wiskott-Aldrich syndrome proteins). These variants demonstrated a diverse array of gating behaviors in response to non-physiological inputs.

We then tested this approach in a cellular context by engineering synthetic Dbl family guanine nucleotide exchange factors (GEFs), which activate Rho family GTPases, the master regulators of the actin cytoskeleton. Microinjection of these GEFs linked specific morphological responses to normally unrelated signaling pathways. In addition, two synthetic GEFs could be linked in series to form a linear cascade, which demonstrated amplification and increased ultrasensitivity when compared to the direct single-GEF circuits.

These results demonstrate the evolutionary plasticity of modular signaling proteins, and suggest that it may be possible to manipulate cellular responses by engineering synthetic signaling networks. This ability will be critical for engineering cells with diverse therapeutic and biotechnological applications.

Table of Contents

List of Tables

List of Figures

Chapter 1

Modularity in Cellular Signal Transduction

Introduction

Living cells must constantly monitor and respond to their environment and internal conditions. In metazoans, individual cells must communicate and respond to other cells in the organism. Thus, cells display a remarkable array of sophisticated signal processing behaviors that rivals or surpasses that of modern computers. Many of these responses are processed by networks of cytoplasmic signaling proteins. Here we review recent advances in our understanding of the fundamental design principles underlying the structure and mechanism of eukaryotic signaling proteins, focusing particularly on how they are functionally linked to one another to form complex circuits capable of information processing. We discuss how the modular organization of signaling proteins may help facilitate the evolution of innovative protein circuits and phenotypes, providing increased fitness in a competitive and changing environment.

Evolvability of Cell Circuitry: Making New Connections

How have the incredibly diverse and complex phenotypes observed in modern eukaryotic organisms evolved? A growing body of work suggests that new phenotypes rarely arise through the evolution of radically new proteins[1]. Rather, innovation is thought to occur through the establishment of novel connectivities between existing or duplicated proteins to generate new regulatory circuits and thereby new regulatory behaviors (**Figure 1-1a**). This model is consistent with the surprisingly small genomes of even very complex

organisms, and the limited number of protein or domain types observed[2-4]. Phenotypic

diversity and complexity appear to arise from new combinations of proteins or protein

domains working as a network, not from the generation of completely new protein

functions. This strategy is similar to that of electronic circuits — a huge variety of

circuits can be built from a finite set of electronic components by wiring them together in

different ways. Thus, a critical question is how new input/output connections can be

established between biological components.

Connecting Transcriptional Nodes: Structural and Functional Modularity

Although this review focuses on protein-based signaling circuits, it is instructive to

consider briefly how new connectivities are generated in transcriptional networks, a

different class of biological regulatory networks (**Figure 1-1b**). Transcriptional control is

mediated by promoters that respond to signals provided by upstream transcription factors

and convert this input into gene expression. Transcriptional nodes are highly modular[1, 5,
6]. First, they display structural modularity: the output region — the coding sequence to

be transcribed — is physically separable from the input regions — the cis-acting

elements that regulate expression. Second, and perhaps more importantly, they display

functional modularity: input and output components still function when separated and can

be recombined to yield new input/output connectivities. For example, insertion of a new

cis-acting element into a promoter can place a gene under the control of a new input

pathway[7]. Alternatively, insertion of a new gene behind a promoter can result in a

radically new output in response to the same input signal. Even linking input and output

elements that have had no previous physiological relationship will often work, in large

part because gene expression is controlled by standardized general transcription machinery. Thus, the highly modular structure of promoters allows the input and output elements to be easily transferred to yield novel node connectivities. Transcriptional nodes are therefore thought to be a highly evolvable system[8]. Recombination of transcriptional input and output components is thought to be a major source of phenotypic variation during evolution[1].

Classical Regulatory Proteins

Historically, the best-studied regulatory proteins are enzymes involved in metabolic pathways, which lack the modularity of transcriptional nodes and therefore present several fundamental problems with respect to generating new input-output connectivities. The output of an enzyme — the reaction it catalyzes and the products generated — is dependent on precise stereochemical requirements; thus, enzymes cannot easily undergo radical changes in output without compromising catalytic activity. Input control of enzymes can be mediated by allosteric effectors; binding of these effectors at allosteric sites is coupled to specific conformational changes at the active site[9, 10]. The intimate and subtle coupling between allosteric and active sites limits the possibility of radically modifying allosteric input control without concomitantly compromising function or stability of the catalytic site. In summary, such metabolic enzymes show little structural or functional modularity; the elements that mediate input and output are found within a single cooperatively-folding unit, and therefore cannot easily be independently modified. Such systems, which we refer to as being tightly integrated, have less readily transferable elements and are therefore not as evolvable as modular systems.

4

Modularity of Eukaryotic Signaling Proteins

Signaling pathways involve enzymes that catalyze reactions such as phosphorylation, dephosphorylation, and nucleotide exchange. The input control of such enzymes determines when, where, and by what they are activated. The output control determines what downstream partners these enzymes act upon once activated.

Signal transduction enzymes differ radically from classical metabolic enzymes: they appear to utilize far more modular mechanisms to determine their input/output connectivities[11]. Over the last decade, our understanding of the design principles of signaling enzymes has increased dramatically due to mechanistic and structural studies as well as the sequencing of multiple eukaryotic genomes. Signaling proteins often contain, in addition to a core catalytic function, multiple independently folding domains or motifs that mediate connectivity by interacting with other signaling elements. These modules are found in different combinations with diverse catalytic functions, suggesting that insertion and recombination of modules may be a common mechanism of evolution of new proteins and connections[2-4].

Eukaryotic signaling proteins appear to have developed a range of modular strategies for controlling their input and output connectivities, all of which involve increased functional separation between core catalytic elements and connectivity elements (**Figure 1-1c**). Three basic mechanisms by which the catalytic activity of kinases and other signaling functions are directed and regulated in a modular manner include the following: the use

of peripheral docking sites, modular interaction domains, and scaffolding and adapter proteins. Each of these mechanisms can be used to select functional upstream and downstream partners as well as, in many cases, to allosterically regulate catalytic activity. These mechanisms represent a continuum of increasing structural modularity in which catalytic function is separated from the elements that determine its wiring (e.g., scaffolds or adapter proteins represent a separation of catalysis and input control into separate gene products).

We will explore the hypothesis that the increasing modularity observed in signaling proteins correlates with higher evolvability: this framework may promote the formation of diverse linkages between catalytic functions via generic, standardized connecting elements. These modular connecting elements may facilitate the evolution of more complex phenotypes, much as standardized components facilitate the design of diverse and complex devices in engineering.

Modular Recognition Domains: Structural Separation of Connectivity and Catalysis

The evolution of metazoans appears to have coincided with an explosion in the use of modular protein domains, including many recognition domains that play a major role in diverse cell signaling processes[2-4] (**Table 1-1**). These include, for example, domains that recognize peptides (e.g., SH3 domains), phosphopeptides (e.g., SH2 domains), and

6

phospholipids (e.g., PH domains). The detailed functions of these diverse domains are reviewed elsewhere[12-20]. Compared with the more specialized Ser/Thr kinase docking sites, such domains represent an even more complete physical separation between elements that mediate connectivity from those that mediate catalytic functions.

Increased Recombinational Possibilities

From a genetic perspective, modular interactions offer more flexibility than docking interactions: both the peptide motifs and their cognate domains can be transferred through simple genetic changes such as recombination and insertion. Thus, both an enzyme and its substrate can make new connections by incorporating a relevant recognition domain or motif (**Figure 1-2a**). Circumstantial evidence for this higher degree of transferability can be found by comparing metazoan genomes. Increasing phenotypic complexity appears to correlate not with the development of new domains (only 7% of human protein families are vertebrate-specific), but rather with an increase in the type and number of new domain combinations: humans have 1.8-fold more distinct protein architectures (arrangements of domains in primary sequence) than do worms and flies[2]. An example of domain mixing and matching is shown in **Figure 1-2b**, illustrating how specific regulatory and catalytic domains can be found in many combinations to yield proteins, and therefore pathways, with highly diverse input/output relationships.

Regulation by Modular Domains

Modular domains can be used not only to physically link partner proteins but also as regulatory elements (**Figure 1-2c**). First, several classes of interaction domains display

conditional recognition. These include phosphopeptide recognition domains such as SH2 domains, for which the linkage of a catalytic domain to its partners depends on a prior phosphorylation event[21]. Similarly, regulated membrane localization can be achieved with lipid recognition modules that bind to rare phosphoinositide species such as PIP3 that are only produced upon activation of PI3 kinase[16].

Modular recognition domains can also play more sophisticated roles in achieving allosteric regulation, most commonly though autoinhibitory mechanisms. Domains can interact in an intramolecular fashion with catalytic domains, either acting as pseudo-substrates or sterically occluding accessibility of the active site[22]. The catalytic function can be specifically switched on by the binding of competitive ligands or by covalent modification events that disrupt the autoinhibitory interaction. In other cases, domains can interact with cognate motifs in a manner that conformationally disrupts catalytic function. In some cases, such as the Src family kinases or the actin regulatory protein N-WASP, multiple domains function together to stabilize an inactive state of their respective catalytic output domains[23-28]. In these cases, the proteins can act as sophisticated switches that respond in complex ways to multiple inputs. For example, a protein might approximate an AND gate if two intramolecular interactions must both be disrupted to release the autoinhibited catalytic function. Interestingly, these modular allosteric switches show behavior very similar to more conventional allosteric proteins: switching involves preferential stabilization of a high activity state by a ligand. However, in the case of modular switches, there is a clear physical separation between the regions of the protein that mediate input regulation and those that mediate output catalytic

activity. Not only does this architecture lend itself to increased transferability of function, but modularity may also allow the incremental construction of switch proteins with multiple layers of input control.

New Examples of Modular Allosteric Signaling Proteins

In recent years the number of signaling proteins that appear to be regulated by modular allosteric mechanisms has exploded. An extensive but not exhaustive list is given in **Table 1-2**. The mechanisms of several examples are shown in **Figure 1-3**.

It has become clear that signaling proteins that utilize modular autoinhibition can display remarkably complex gating behaviors. For example, the Abl kinase appears to be capable of integrating information from three distinct inputs (**Figure 1-3a**). Like the related Src kinase, Abl contains an SH2 domain and an SH3 domain that participate in autoinhibitory interactions. However, Abl contains a third interaction required for autoinhibition: an N-terminal myristoyl group binds in a pocket in the kinase[29]. Myristoyl binding appears to contribute to the nearby SH2 docking site[30]. These findings are consistent with a model in which Abl acts as a three-input gate: it can be activated by exogenous SH2 or SH3 ligands as well as insertion of the myristoyl group into the membrane (alternatively there may exist an unknown hydrophobic ligand that displaces the myristoyl group from its binding pocket). These three inputs would likely function cooperatively, activating the kinase with high specificity.

The WASP family of actin-regulatory proteins also appear to be able to coordinate at least three distinct inputs (**Figure 1-3b**). Previous work has shown that cooperative autoinhibitory interactions allow N-WASP to respond synergistically to a specific combination of inputs: the GTPase Cdc42 and the phosphoinositide PIP2[24, 31]. More recent work has revealed that phosphorylation is a third input that can function intimately with Cdc42 activation[32-34]. The Cdc42-binding module that participates in autoinhibition is referred to as the GTPase binding domain (GBD). A residue in the GBD (Tyr 256) can be specifically phosphorylated in a manner that disrupts its autoinhibitory interaction without perturbing its binding to Cdc42. Thus, both Cdc42 and phosphorylation can function cooperatively to disrupt the same autoinhibitory interaction. This dual activation may provide a type of memory: binding of Cdc42 may transiently disrupt the GBD autoinhibitory interaction, thereby facilitating phosphorylation, which may provide a more long-lived state of activation[32].

The Problem of Domain Discrimination

Although the use of modular domains may allow the rapid generation of new signaling input/output relationships, the expansion of domain families presents a new problem: how can repeated domains in a proteome encode specific information in the context of many related family members (**Table 1-1**)? For example, consider the SH3 domain family, which binds to proline-rich peptides containing the core motif PxxP: the SMART database predicts that there are 31 SH3 domains in the yeast, 132 in C. elegans, 273 in Drosophila, and 894 in humans[35, 36]. How can an ordered array of component connectivities be maintained by such a large set of related domains?

Recent studies suggest that several strategies have evolved for maintaining domain discrimination. First, domains can diverge so far from other family members that they display distinct, non-canonical recognition profiles (**Figure 1-4a**). For example, some SH3 domains have diverged to no longer recognize PxxP motifs: the C-terminal SH3 domain of the T-cell adapter protein Gads, instead recognizes RxxK motifs[37]. This recognition event occurs on a surface distinct from the canonical proline binding pocket[38, 39]. One Gads ligand, hematopoietic progenitor kinase-1 (HPK1), binds primarily through an RxxK motif, but its binding is augmented by a weak secondary PxxP motif, thus illustrating the versatility of this divergent domain[40]. A pair of SH3 domains in p47phox has been found to act as a single unit, using the surface between the two domains to recognize a novel motif[41]. Similarly, non-canonical domains have been found in many other domain families, including the SH2 domain from the protein SAP (also called SH2D1A) that binds unphosphorylated motifs[42] and the C2 domain from PKCδ, of a class of domains that normally binds phospholipids or unphosphorylated peptides, that recognizes phosphotyrosine motifs[19, 43].

A second mechanism for increasing domain-mediated specificity is to use multiple domains to recognize dual ligands in a cooperative manner[44-46] (**Figure 1-4b**). A third mechanism is to use system-wide optimization of the domain interaction network (**Figure 1-4c**). Recent studies in yeast have shown that although many of the ~30 SH3 domains have overlapping specificity as determined by peptide libraries, there appears to be some level of negative selection against sequences that interact in a highly promiscuous

manner[47]. Many physiological partner peptides are optimized for specificity not only by positive selection for binding to the proper domain, but also by negative selection against interaction with competing domains in that genome. Thus, in many cases, individual SH3 domains are only observed to interact with a handful of more than 1500 potential PxxP partners within the genome[48]. Finally, a fourth way to achieve specificity is to segregate domains either through subcellular compartmentalization, differential temporal expression, or tissue-specific expression (**Figure 1-4d**)[47]. Thus, domains with highly overlapping recognition properties might never compete for the same targets.

Nonetheless, even with these mechanisms, there likely comes a point at which the information-encoding capability of a family of domains is saturated, and increasing signaling complexity may require the development of orthogonal domains. For example, SH2 domains are generally only found in metazoans, and it is possible that the development of SH2 domains and tyrosine phosphorylation-based signaling in general may have been a pre-requisite for the evolution of multicellularity, given its need for increased signaling bandwidth (cell-cell signaling in addition to cell-environment signaling). Interestingly, SH2 domains and receptor tyrosine kinases have recently been identified in choanoflagellates, the closest single-celled eukaryotes to the evolutionary branch point of multicelluarity[49].

Conclusions: Modularity and Evolvability of Biological Regulatory Systems

Highly modular architectures are not only found in eukaryotic signaling systems but also in many other systems, including transcription, proteolysis, and cellular trafficking[1, 50, 51]. These systems are characterized by the use of increasingly general, portable elements that can be genetically interchanged to mediate new regulatory connections. The exact domains and motifs that implement these connections vary to some extent, but it does appear that there may be some pressure to maintain a degree of evolvability in such systems. The reuse of similar modular domains in different contexts represents standardization of the means of communication between protein nodes. Standardization is a central feature of highly complex and evolvable systems[52].

Why might there be selective pressure to maintain modularity and evolvability, given that cellular systems cannot actually foresee the need to change their response behaviors? Presumably, in a constantly changing and competitive environment, the lack of an ability to rapidly evolve novel responses might prove to be a disadvantage. In many cases, highly integrated, non-modular systems perform in a more efficient, optimal manner, but such performance would only be optimal for a specific and unchanging environment. During the course of evolution, as environmental pressures shift, there is likely a constant push and pull between the efficiency of integration on one hand and the flexibility and adaptability of modularity on the other. Even in engineered systems like electronic circuits, where modular components provide an advantage in circuit development, there

is often pressure to minimize and integrate circuits once they are well developed. This optimization and integration can lead to a loss of the modularity that was critical during development (e.g., components in integrated circuits do not have transferable functions). Similarly, one might expect that modularity could easily be lost in fundamental housekeeping biological processes, which do not change significantly over evolution. In support of this model, recent bioinformatics studies indicate that tissue-specific proteins, especially those associated with evolutionarily newer functions, tend to have a more modular composition than those proteins that are globally expressed and have a housekeeping function[53]. Hopefully as more families of closely related genomes are sequenced we will be able to gain insight into the actual paths by which new signaling pathways have arisen over the course of evolution.

Table 1-1 Abundance of selected modular domains (and proteins containing them) in commonly studied eukaryotes

	Homo sapiens	*Mus musculus*	*Drosophila melanogaster*	*Caenorhabditis elegans*	*Saccharomyces cerevisiae*
SH3[a]	223 (180)[b]	124 (92)	113 (76)	83 (68)	26 (22)
WW	91 (49)	27 (117)	21 (14)	40 (22)	9 (6)
PDZ	234 (126)	119 (78)	98 (71)	106 (79)	3 (2)
SH2	112 (98)	73 (67)	33 (30)	67 (66)	1 (1)
PTB	34 (30)	14 (12)	7 (7)	23 (20)	0 (0)
14-3-3	8 (8)	4 (4)	4 (4)	2 (2)	1 (1)
BRCT	39 (20)	23 (12)	28 (16)	2 (2)	1 (1)
FHA	16 (16)	9 (9)	17 (17)	12 (12)	13 (12)
C2	149 (99)	94 (63)	51 (36)	93 (64)	22 (11)
Total genes[c]	30,000	30,000	14,000	19,000	6,300

[a] Abbreviations and descriptions of domains in table: SH3 = Src homology 3 domain, binds PxxP peptide ligands[20]; WW = PxxP binding domain named after two conserved Trp residues[20]; PDZ = domain from PSD-95, Dlg, ZO-1, binds C-terminal peptide ligands[15]; SH2 = Src homology 2 domain, binds phospho-Tyr peptide ligands[18]; PTB = phospho-Tyr binding domain[18]; 14-3-3 = phospho-Ser/Thr binding domain[12]; BRCT = breast cancer susceptibility gene, C-terminal domain, binds phospho-Ser/Thr peptide ligands[14]; FHA = forkhead-associated domain, binds phospho-Ser/Thr peptide ligands[13]; C2 = domain from protein kinase C, binds phospholipids and occasionally phospho-Tyr peptide ligands[19]

[b] Source: Simple Modular Architecture Research Tool (SMART) database (http://smart.embl-heidelberg.de/)[35, 36].

[c] Source: Human Genome Project Information, Functional and Comparative Genomics Fact Sheet (www.ornl.gov/sci/techresources/Human_Genome/faq/compgen.shtml)

Table 1-2 Examples of signaling proteins gated by autoinhibition

	Input(s)	Output	Mechanism of Autoinhibition	Ref
			Steric	
Epidermal Growth Factor Receptor (EGFR)	Epidermal Growth Factor	Receptor dimerization	cysteine-rich domain occludes receptor dimerization surface (another cysteine-rich domain)	54, 55
SH2-containing phosphatase 2 (SHP2)	SH2-binding motifs (p-Tyr)	phosphatase	N-terminal SH2 domain sterically blocks phosphatase catalytic site	56
p21-activated kinase (PAK1)	Rac or Cdc42	Ser/Thr kinase	GTPase binding domain (GBD) blocks catalytic site preventing auto-phosphorylation	57
Twitchin	Ca^{+2}/S100 complex	Ser/Thr kinase	pseudo-substrate motif occupies kinase active site; locked into position by adjacent IgG domain.	58
p47phox	phosphorylation by PKC	NADPH oxidase	intramolecular peptide blocks tandem SH3 domains from interacting with membrane associated partner, thereby blocking formation of functional oxidase complex	41
Vav	phosphorylation by Src family kinases	Rho, Rac, Cdc42 GEF (DH-PH module)	amino-terminal extension blocks GTPase interaction site	59
			Conformational	
Src kinases	SH2 and SH3 binding motifs	Tyr kinase	Binding of the SH2 and SH3 domains to intramolecular ligands locks kinase in inactive conformation.	60
c-Abl	SH2 and SH3 binding motifs; possibly membrane targeting of myristoyl group	Tyr kinase	Binding of N-terminal myristoyl group and SH2 and SH3 domains to sites on or adjacent to kinase domain locks kinase in inactive conformation remarkably similar to the autoinhibited structure of Src	29, 30
Neuronal Wiskott Aldrich Syndrome protein (N-WASP)	Cdc42 and PIP2	Arp2/3 stimulation (actin polym.)	The GTPase binding domain (GBD) and a polybasic motif (B) form cooperative intracomplex interactions that conformationally inactivate the N-WASP output domain, blocking its ability to activate the Arp2/3 actin filament nucleating complex.	24, 31

	Input(s)	Output	Mechanism of Autoinhibition	Ref
			Unknown (evidence for autoinhibition given)	
Polo-like kinase (PLK)	phosphorylated Cdc25	Ser/Thr kinase	Polo-box domain reduces activity of kinase domain. This repressive interaction can be dissociated by phosphorylation of the kinase domain at Tyr210.	61
Dbl	PIP2 and PIP3	Rho, Cdc42 GEF (DH-PH module)	N-terminal region binds PH domain and is required for autoinhibition of exchange activity of adjacent DH domain	62
Intersectin	proline-rich region from N-WASP	Cdc42 GEF (DH-PH module)	SH3 domains inhibit the catalytic DH domain. Binding proline-rich region of N-WASP to SH3 domains stimulates DH exchange activity.	63, 64
Cdc24	Rsr1p/Bud1p and Bem1p binding	Cdc42 GEF (DH-PH module)	Phox-Bem1 homolgy (PB1) domain involved in autoinhibition. Rsr1p/Bud1p binds calponin homology (CH) domain which is thought to cause a conformational change that dissociates PB1 autoinhibitory interaction. Bem1 binds the PB1 domain and is thought to trap this active state.	65
chimaerin	phosphatidylserine and phosphatidic acid	Rac GAP	C1 domain required for regulation.	66
P-Rex1	PIP3 and $G_{\beta\gamma}$	Rac GEF (DH-PH module)	Coincidence detector for PIP3 and $G_{\beta\gamma}$ but mechanism of autoinhibition is unknown.	67
Rho-associated kinase (ROCK)	Rho and arachidonic acid	Ser/Thr kinase	Overexpression of a fragment containing the rho-binding domain (RB) and pleckstrin homology domain (PH) inhibits activity of the kinase. Point mutations disrupting binding of RB to Rho have similar effects. RhoA binding to RB is thought to activate kinase	68

a

- **new nodes**
- **new connectivities**

b

INPUT: cis acting factors

OUTPUT: coding seq.

CIS ELEMENT change input control

HIGH MODULARITY

HIGH EVOLVABILITY

CODING SEQ. change output

c

INPUT: upstream regulators

OUTPUT: substrate/ effector

ACTIVE SITE

DOCKING INTERACTIONS

MODULAR DOMAINS

SCAFFOLDS/ADAPTERS

Highly integrated catalytic & regulatory functions

HIGH   LOW

INTEGRATION

MODULARITY, EVOLVABILITY

LOW   HIGH

Separable, standardized components mediate connectivity & regulation of diverse catalytic units

Figure 1-1 Modularity and evolvability of cellular regulatory circuits and nodes. (legend next page)

18

Figure 1-1 Modularity and evolvability of cellular regulatory circuits and nodes. (previous page)

a, Evolution of new regulatory pathways and responses. A simple linear pathway (shown in black) can be converted to a more complex one through the addition of novel nodes that introduce branchpoints, or by the generation of novel functional linkages between existing components, such as the feedback or feedforward circuits depicted. New components and connections are shown in red. **b,** New connectivity with transcriptional nodes. Transcriptional circuits exemplify a highly modular network, as simple recombination events can alter input-output relationships. Introduction of new cis-acting elements such as promoters and enhancers can alter input control, and insertion of a new coding sequence downstream of an existing set of cis-acting elements can impose an existing mode of regulation upon expression of a different gene. **c,** New connectivity with protein/enzyme nodes. Four means of mediating connections between protein nodes are depicted: active site recognition, docking interactions, recognition through modular domain/ligand pairs, and interactions mediated by organizing factors such as scaffolds or adapters. These connection strategies fall on a continuum of modularity vs. integration; greater separation between catalytic functions and interactions that mediate connections lends itself to greater evolvability of the signaling network.
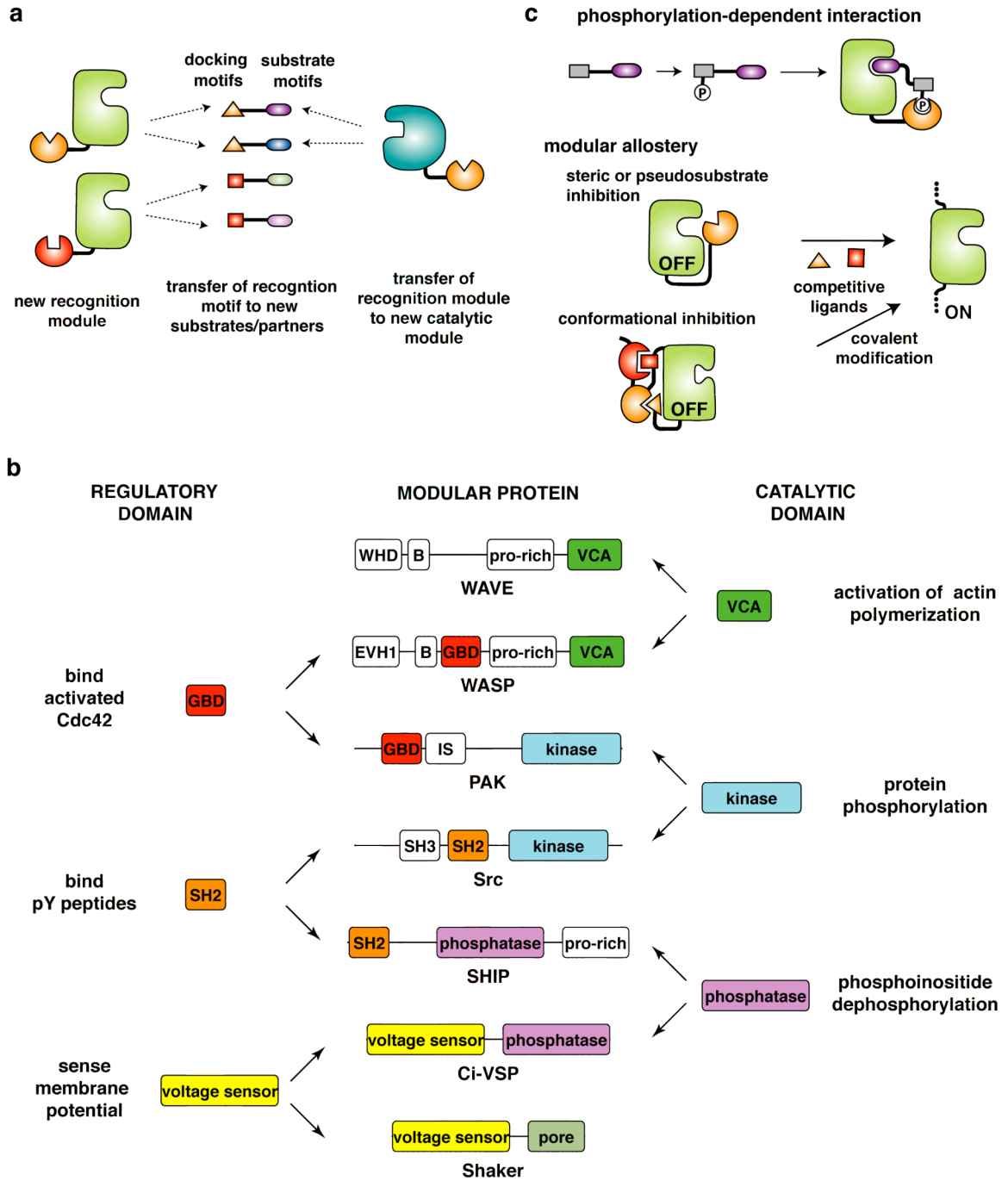
Figure 1-2 Modular interaction domains can mediate new connectivity
(legend next page)

Figure 1-2 Modular interaction domains can mediate new connectivity
(previous page)
a, Transferability of modular recognition and catalytic functions. Modular domains facilitate the formation of new connections between proteins, as standardized recognition domains or their ligands can be swapped onto catalytic modules or substrates via recombination events, opening up a new set of possible enzyme-substrate interactions. **b,** Evidence of evolutionary input/output transfer. Naturally occurring examples are depicted in which domains are reused in various combinations to mediate distinct connections between catalytic activities and target molecules. The VCA (*v*erprolin homology, *c*ofilin homology, *a*cidic) domain, which activates actin polymerization, is common to the actin regulatory proteins WAVE and WASP[69], but it is covalently linked to a different set of interaction domains in each case, contributing to distinct modes of deployment of this output activity. Of these interaction domains, the GTPase binding domain (GBD) of WASP is also found in p21-activated kinase (PAK)[57] and is used to direct its binding partner, activated Cdc42, to each of these two diverse proteins. The kinase domain from PAK is reused in many different contexts. The classical example of Src is depicted, in which the kinase domain is joined with several protein interaction domains, including the Src homology 2 (SH2) domain, which regulates the activation state of the kinase and mediates its interaction with phosphotyrosine-containing peptides[23, 26-28]. This SH2 domain is, likewise, reused in many signaling components, such as the SHIP phosphatase[70]. The phosphatase domain found in SHIP is reused in multitudes of signaling proteins as well, including the recently described voltage-sensing phosphatase from *Ciona intestinalis*, Ci-VSP. Due to a fusion of the phosphatase domain with a voltage-sensing domain more traditionally found in voltage-gated channels such as Shaker, Ci-VSP exhibits regulation of its phosphatase activity by membrane potential[71, 72]. Thus, many complex signaling proteins are built from a relatively small toolkit of standardized components that are combinatorially connected. **c,** Enzyme regulation by modular domains. Some modular domains only recognize their ligands after covalent modification resulting from other cellular signaling processes, thus linking the connectivity of proteins containing these domains to the regulation of these other pathways. In addition, modular domains are often used to regulate enzyme activity more directly. These domains can participate in interactions that inhibit catalysis, either by sterically blocking access to the catalytic site or by preferentially stabilizing an inactive conformation of the catalytic domain. These inactive states can then be reversed upon exposure to competing ligands that bind to the domains, or by covalent modification of the domains or ligands.
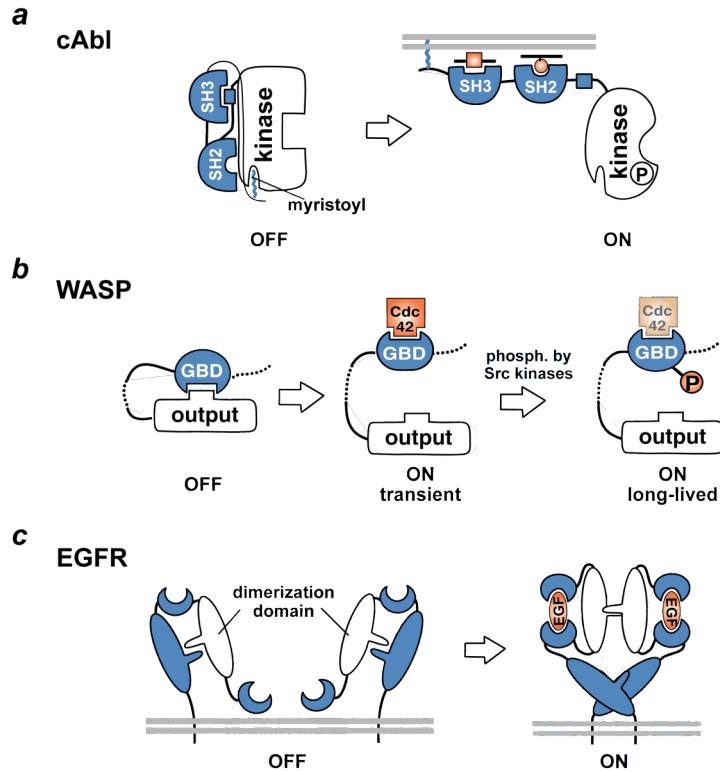
Figure 1-3 New examples of modular allosteric signaling proteins
a, cAbl is a modular allosteric switch that may respond to three or more inputs. The kinase domain is conformationally autoinhibited by three coordinated intramolecular interactions: an SH3-peptide interaction, docking of the SH2 domain on the kinase domain, and docking of an N-terminal myristoyl group in a hydrophobic pocket on the kinase domain[30]. The kinase can be activated by ligands that compete with these autoinhibitory interactions, including exogenous SH3 and SH2 ligands as well as interactions that might displace the buried myristoyl group (e.g., membrane targeting). Relief of autoinhibition by a combination of these mechanisms allows autophophorylation and full activation of the kinase. **b,** WASP family proteins can interact with and activate the Arp2/3 actin-nucleating complex. The output domain, constitutively active in isolation, is autoinhibited by several interactions, including an intramolecular interaction with the GTPase binding domain (GBD). This autoinhibitory interaction can be relieved by binding of the GTPase Cdc42. In addition, recent studies have shown that Src family kinases can phosphorylate tyrosine 256 on the GBD, destabilizing its ability to participate in the autoinihibitory interaction[32]. Phosphorylation is only observed when the protein has been activated by Cdc42. Thus, it has been proposed that phosphorylation may provide "memory" by locking the protein in a longer-lived activated state, even after removal of active Cdc42 as a stimulus. **c,** EGF receptor (EGFR) is activated by ligand-mediated dimerization. However, unlike similar receptors, dimerization does not involve any ligand-mediated bridging interactions. Instead, the receptor has a dimerization domain that in the inactive state is occluded by autoinhibitory interactions. EGF ligand relieves this autoinhibition, indirectly promoting dimerization[55].
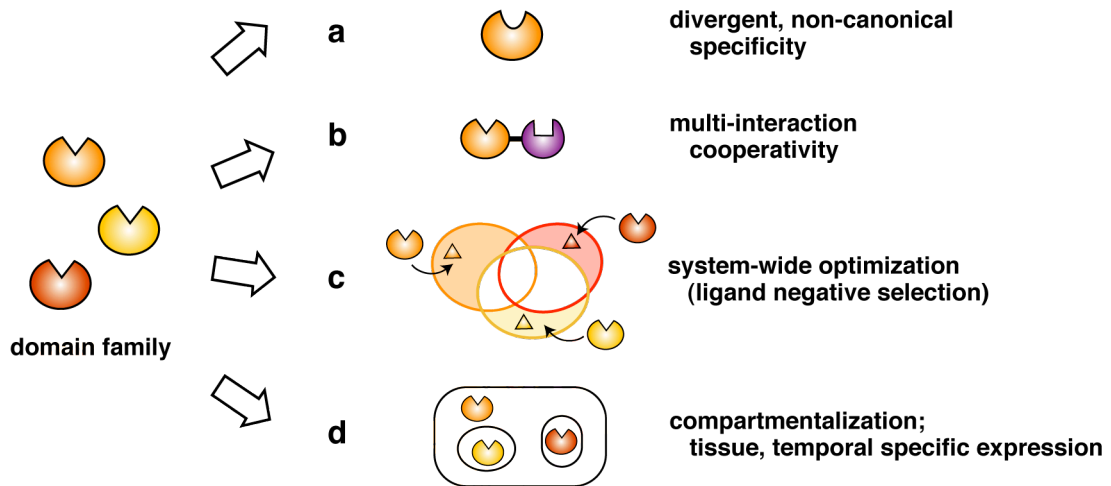
Figure 1-4 Mechanisms of domain discrimination
a, Domains can evolve divergent ligand-binding pockets that recognize sequences that stray from the consensus for the domain family. **b,** Multiple domains can be used in combination to generate a combinatorial increase in selectivity compared with the individual recognition events alone. **c,** Domains and ligands within an organism can coevolve to occupy regions of recognition space with an acceptably low level of cross-recognition. **d,** Domains and ligands can be segregated in space and time so that they are more likely to be co-expressed with genuine interacting partners than with spurious cross-reactive partners.

References

1. Carroll, S.B. Evolution at two levels: on genes and form. *PLoS Biol* **3**, e245 (2005).

2. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).

3. Rubin, G.M. The draft sequences. Comparing species. *Nature* **409**, 820-821 (2001).

4. Venter, J.C. *et al.* The sequence of the human genome. *Science* **291**, 1304-1351 (2001).

5. Hartwell, L.H., Hopfield, J.J., Leibler, S. & Murray, A.W. From molecular to modular cell biology. *Nature* **402**, C47-52 (1999).

6. Remenyi, A., Scholer, H.R. & Wilmanns, M. Combinatorial control of gene expression. *Nat Struct Mol Biol* **11**, 812-815 (2004).

7. Ihmels, J. *et al.* Rewiring of the yeast transcriptional network through the evolution of motif usage. *Science* **309**, 938-940 (2005).

8. Kirschner, M. & Gerhart, J. Evolvability. *Proc Natl Acad Sci U S A* **95**, 8420-8427 (1998).

9. Monod, J., Changeux, J.P. & Jacob, F. Allosteric proteins and cellular control systems. *J Mol Biol* **6**, 306-329 (1963).

10. Perutz, M.F. Hemoglobin structure and respiratory transport. *Sci Am* **239**, 92-125 (1978).

11. Pawson, T. & Nash, P. Assembly of cell regulatory systems through protein interaction domains. *Science* **300**, 445-452 (2003).

12. Bridges, D. & Moorhead, G.B. 14-3-3 proteins: a number of functions for a numbered protein. *Sci STKE* **2004**, re10 (2004).

13. Durocher, D. & Jackson, S.P. The FHA domain. *FEBS Lett* **513**, 58-66 (2002).

14. Glover, J.N., Williams, R.S. & Lee, M.S. Interactions between BRCT repeats and phosphoproteins: tangled up in two. *Trends Biochem Sci* **29**, 579-585 (2004).

15. Harris, B.Z. & Lim, W.A. Mechanism and role of PDZ domains in signaling complex assembly. *J Cell Sci* **114**, 3219-3231 (2001).

16. Lemmon, M.A. & Ferguson, K.M. Signal-dependent membrane targeting by pleckstrin homology (PH) domains. *Biochem J* **350 Pt 1**, 1-18 (2000).

17. Pawson, T., Gish, G.D. & Nash, P. SH2 domains, interaction modules and cellular wiring. *Trends Cell Biol* **11**, 504-511 (2001).

18. Schlessinger, J. & Lemmon, M.A. SH2 and PTB domains in tyrosine kinase signaling. *Sci STKE* **2003**, RE12 (2003).

19. Sondermann, H. & Kuriyan, J. C2 can do it, too. *Cell* **121**, 158-160 (2005).

20. Zarrinpar, A., Bhattacharyya, R.P. & Lim, W.A. The structure and function of proline recognition domains. *Sci STKE* **2003**, RE8 (2003).

21. Marengere, L.E. & Pawson, T. Structure and function of SH2 domains. *J Cell Sci Suppl* **18**, 97-104 (1994).

22. Dueber, J.E., Yeh, B.J., Bhattacharyya, R.P. & Lim, W.A. Rewiring cell signaling: the logic and plasticity of eukaryotic protein circuitry. *Curr Opin Struct Biol* **14**, 690-699 (2004).

23.   Lim, W.A. The modular logic of signaling proteins: building allosteric switches from simple binding domains. *Curr Opin Struct Biol* **12**, 61-68 (2002).

24.   Prehoda, K.E., Scott, J.A., Mullins, R.D. & Lim, W.A. Integration of multiple signals through cooperative regulation of the N-WASP-Arp2/3 complex. *Science* **290**, 801-806 (2000).

25.   Rohatgi, R., Ho, H.Y. & Kirschner, M.W. Mechanism of N-WASP activation by CDC42 and phosphatidylinositol 4, 5-bisphosphate. *J Cell Biol* **150**, 1299-1310 (2000).

26.   Sicheri, F., Moarefi, I. & Kuriyan, J. Crystal structure of the Src family tyrosine kinase Hck. *Nature* **385**, 602-609 (1997).

27.   Williams, J.C. *et al.* The 2.35 A crystal structure of the inactivated form of chicken Src: a dynamic molecule with multiple regulatory interactions. *J Mol Biol* **274**, 757-775 (1997).

28.   Xu, W., Harrison, S.C. & Eck, M.J. Three-dimensional structure of the tyrosine kinase c-Src. *Nature* **385**, 595-602 (1997).

29.   Hantschel, O. *et al.* A myristoyl/phosphotyrosine switch regulates c-Abl. *Cell* **112**, 845-857 (2003).

30.   Nagar, B. *et al.* Structural basis for the autoinhibition of c-Abl tyrosine kinase. *Cell* **112**, 859-871 (2003).

31.   Rohatgi, R. *et al.* The interaction between N-WASP and the Arp2/3 complex links Cdc42-dependent signals to actin assembly. *Cell* **97**, 221-231 (1999).

32. Torres, E. & Rosen, M.K. Contingent phosphorylation/dephosphorylation provides a mechanism of molecular memory in WASP. *Mol Cell* **11**, 1215-1227 (2003).

33. Cory, G.O., Garg, R., Cramer, R. & Ridley, A.J. Phosphorylation of tyrosine 291 enhances the ability of WASp to stimulate actin polymerization and filopodium formation. Wiskott-Aldrich Syndrome protein. *J Biol Chem* **277**, 45115-45121 (2002).

34. Badour, K. *et al.* Fyn and PTP-PEST-mediated regulation of Wiskott-Aldrich syndrome protein (WASp) tyrosine phosphorylation is required for coupling T cell antigen receptor engagement to WASp effector function and T cell activation. *J Exp Med* **199**, 99-112 (2004).

35. Letunic, I. *et al.* SMART 4.0: towards genomic data integration. *Nucleic Acids Res* **32**, D142-144 (2004).

36. Schultz, J., Milpetz, F., Bork, P. & Ponting, C.P. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* **95**, 5857-5864 (1998).

37. Berry, D.M., Nash, P., Liu, S.K., Pawson, T. & McGlade, C.J. A high-affinity Arg-X-X-Lys SH3 binding motif confers specificity for the interaction between Gads and SLP-76 in T cell signaling. *Curr Biol* **12**, 1336-1341 (2002).

38. Harkiolaki, M. *et al.* Structural basis for SH3 domain-mediated high-affinity binding between Mona/Gads and SLP-76. *Embo J* **22**, 2571-2582 (2003).

39. Liu, Q. *et al.* Structural basis for specific binding of the Gads SH3 domain to an RxxK motif-containing SLP-76 peptide: a novel mode of peptide recognition. *Mol Cell* **11**, 471-481 (2003).

40. Lewitzky, M., Harkiolaki, M., Domart, M.C., Jones, E.Y. & Feller, S.M. Mona/Gads SH3C binding to hematopoietic progenitor kinase 1 (HPK1) combines an atypical SH3 binding motif, R/KXXK, with a classical PXXP motif embedded in a polyproline type II (PPII) helix. *J Biol Chem* **279**, 28724-28732 (2004).

41. Groemping, Y., Lapouge, K., Smerdon, S.J. & Rittinger, K. Molecular basis of phosphorylation-induced activation of the NADPH oxidase. *Cell* **113**, 343-355 (2003).

42. Li, S.C. *et al.* Novel mode of ligand binding by the SH2 domain of the human XLP disease gene product SAP/SH2D1A. *Curr Biol* **9**, 1355-1362 (1999).

43. Benes, C.H. *et al.* The C2 domain of PKCdelta is a phosphotyrosine binding domain. *Cell* **121**, 271-280 (2005).

44. Bu, J.Y., Shaw, A.S. & Chan, A.C. Analysis of the interaction of ZAP-70 and syk protein-tyrosine kinases with the T-cell antigen receptor by plasmon resonance. *Proc Natl Acad Sci U S A* **92**, 5106-5110 (1995).

45. Iwashima, M., Irving, B.A., van Oers, N.S., Chan, A.C. & Weiss, A. Sequential interactions of the TCR with two distinct cytoplasmic tyrosine kinases. *Science* **263**, 1136-1139 (1994).

46. Pluskey, S., Wandless, T.J., Walsh, C.T. & Shoelson, S.E. Potent stimulation of SH-PTP2 phosphatase activity by simultaneous occupancy of both SH2 domains. *J Biol Chem* **270**, 2897-2900 (1995).

47. Zarrinpar, A., Park, S.H. & Lim, W.A. Optimization of specificity in a cellular protein interaction network by negative selection. *Nature* **426**, 676-680 (2003).

48. Landgraf, C. *et al.* Protein interaction networks by proteome peptide scanning. *PLoS Biol* **2**, E14 (2004).

49. King, N., Hittinger, C.T. & Carroll, S.B. Evolution of key cell signaling and adhesion protein families predates animal origins. *Science* **301**, 361-363 (2003).

50. Jackson, P.K. *et al.* The lore of the RINGs: substrate recognition and catalysis by ubiquitin ligases. *Trends Cell Biol* **10**, 429-439 (2000).

51. Vale, R.D. & Milligan, R.A. The way things move: looking under the hood of molecular motor proteins. *Science* **288**, 88-95 (2000).

52. Csete, M. & Doyle, J. Bow ties, metabolism and disease. *Trends Biotechnol* **22**, 446-450 (2004).

53. Cohen-Gihon, I., Lancet, D. & Yanai, I. Modular genes with metazoan-specific domains have increased tissue specificity. *Trends Genet* **21**, 210-213 (2005).

54. Burgess, A.W. *et al.* An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. *Mol Cell* **12**, 541-552 (2003).

55. Ferguson, K.M. *et al.* EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization. *Mol Cell* **11**, 507-517 (2003).

56. Hof, P., Pluskey, S., Dhe-Paganon, S., Eck, M.J. & Shoelson, S.E. Crystal structure of the tyrosine phosphatase SHP-2. *Cell* **92**, 441-450 (1998).

57.   Lei, M. *et al.* Structure of PAK1 in an autoinhibited conformation reveals a multistage activation switch. *Cell* **102**, 387-397 (2000).

58.   Kobe, B. *et al.* Giant protein kinases: domain interactions and structural basis of autoregulation. *Embo J* **15**, 6810-6821 (1996).

59.   Aghazadeh, B., Lowry, W.E., Huang, X.Y. & Rosen, M.K. Structural basis for relief of autoinhibition of the Dbl homology domain of proto-oncogene Vav by tyrosine phosphorylation. *Cell* **102**, 625-633 (2000).

60.   Sicheri, F. & Kuriyan, J. Structures of Src-family tyrosine kinases. *Curr Opin Struct Biol* **7**, 777-785 (1997).

61.   Jang, Y.J., Lin, C.Y., Ma, S. & Erikson, R.L. Functional studies on the role of the C-terminal domain of mammalian polo-like kinase. *Proc Natl Acad Sci U S A* **99**, 1984-1989 (2002).

62.   Bi, F. *et al.* Autoinhibition mechanism of proto-Dbl. *Mol Cell Biol* **21**, 1463-1474 (2001).

63.   Hussain, N.K. *et al.* Endocytic protein intersectin-l regulates actin assembly via Cdc42 and N-WASP. *Nat Cell Biol* **3**, 927-932 (2001).

64.   Zamanian, J.L. & Kelly, R.B. Intersectin 1L guanine nucleotide exchange activity is regulated by adjacent src homology 3 domains that are also involved in endocytosis. *Mol Biol Cell* **14**, 1624-1637 (2003).

65.   Shimada, Y., Wiget, P., Gulli, M.P., Bi, E. & Peter, M. The nucleotide exchange factor Cdc24p may be regulated by auto-inhibition. *Embo J* **23**, 1051-1062 (2004).

66. Ahmed, S. *et al.* A novel functional target for tumor-promoting phorbol esters and lysophosphatidic acid. The p21rac-GTPase activating protein n-chimaerin. *J Biol Chem* **268**, 10709-10712 (1993).

67. Welch, H.C. *et al.* P-Rex1, a PtdIns(3,4,5)P3- and Gbetagamma-regulated guanine-nucleotide exchange factor for Rac. *Cell* **108**, 809-821 (2002).

68. Amano, M. *et al.* The COOH terminus of Rho-kinase negatively regulates rho-kinase activity. *J Biol Chem* **274**, 32418-32424 (1999).

69. Stradal, T.E. *et al.* Regulation of actin dynamics by WASP and WAVE family proteins. *Trends Cell Biol* **14**, 303-311 (2004).

70. Rohrschneider, L.R., Fuller, J.F., Wolf, I., Liu, Y. & Lucas, D.M. Structure, function, and biology of SHIP proteins. *Genes Dev* **14**, 505-520 (2000).

71. Murata, Y., Iwasaki, H., Sasaki, M., Inaba, K. & Okamura, Y. Phosphoinositide phosphatase activity coupled to an intrinsic voltage sensor. *Nature* **435**, 1239-1243 (2005).

72. Sands, Z., Grottesi, A. & Sansom, M.S. Voltage-gated ion channels. *Curr Biol* **15**, R44-47 (2005).

Chapter 2

Reprogramming Control of an Allosteric Signaling Switch Through Modular Recombination

Abstract

Many eukaryotic signaling proteins are composed of simple modular binding domains, yet they can display sophisticated behaviors such as allosteric gating and multi-input signal integration, properties essential for complex cellular circuits. To understand how such behavior can emerge from combinations of simple domains, we engineered variants of the actin regulatory protein N-WASP (neuronal Wiskott-Aldrich Syndrome Protein) in which the "output" domain of N-WASP was recombined with heterologous autoinhibitory "input" domains. Synthetic switch proteins were created with diverse gating behaviors in response to non-physiological inputs. Thus, this type of modular framework can facilitate the evolution or engineering of cellular signaling circuits.

Introduction

Cellular behavior is mediated by circuits of interconnected signal transduction proteins. Many of these proteins are allosteric — their catalytic output activity is gated by specific upstream inputs such as ligand binding or covalent modification. Most eukaryotic signaling proteins are composed of modular domains with binding or catalytic functions[1, 2]. It has been proposed that domain recombination could facilitate the evolution of proteins with novel signaling functions[1-4].

Consistent with such a model, complex allosteric gating in some signaling switches is mediated by modular, autoinhibitory interactions[4, 5]. For example, the actin regulatory switch N-WASP[6, 7] (**Figure 2-1A**), which displays sophisticated signal integration, contains an output region ("VCA" domain) that in isolation is constitutively active — it stimulates actin polymerization by binding and activating the actin-related protein (Arp) 2/3 complex. However, two modular domains, a highly basic (B) motif and a GTPase binding domain (GBD) repress activity through autoinhibitory interactions[8, 9]. Two activating stimuli, the phosphoinositide PIP_2 and the activated GTPase Cdc42, bind the B and GBD motifs, respectively, and disrupt autoinhibition[9, 10]. Because the two inputs act cooperatively, N-WASP approximates an AND-gate in which strong activation is only observed in the presence of both inputs[9, 11]. Such multi-input regulation is thought to yield precise spatial and temporal control over actin polymerization.

Results and Discussion

We explored the flexibility of such modular regulation by attempting to use domain recombination to reprogram input control of N-WASP. As a simple test of whether modular autoinhibition is interchangable, we engineered a synthetic signaling switch gated by a single heterologous ligand (**Figure 2-1B**). The design involved tethering an unrelated modular domain-ligand pair — in this case a PDZ domain and its cognate C-terminal peptide ligand — to the termini of the N-WASP output domain. This design would create a potential autoinhibitory interaction that could be relieved by competitive binding of an external PDZ ligand.

Under basal conditions, this synthetic switch was repressed in an *in vitro* actin polymerization assay[12] (Methods, **Figure 2-S1**). Repression required an intact, intramolecular autoinhibitory interaction: constructs containing only one interaction partner were not repressed, and addition of saturating free PDZ domain (~10-fold > K_d) *in trans* to a construct bearing only the PDZ ligand did not yield repression (**Figure 2-S2**). The intramolecular PDZ interaction likely locks the output domain in an inactive conformation or restricts dynamic properties required for activity.

The switch was activated by increasing concentrations of free PDZ ligand (**Figure 2-1C**), with maximal activity close to that of the isolated output domain. Half-maximal activation (K_{act}) required 50 μM input. Precise gating behavior was dependent on the affinity of the autoinhibitory interaction (**Figure 2-S3**); reducing affinity of the internal

ligand resulted in lower basal repression but increased input sensitivity (reduced K_{act}), as would be expected if the intramolecular PDZ interaction was required for repression.

As in electronic circuits, complex cellular regulation often requires multi-input integrating gates (AND, OR, XOR, etc.) used in combinatorial control or feedback and feedforward loops[4]. We attempted the design of synthetic AND-gate switches by covalently tethering two modular domain-ligand pairs to N-WASP's output domain such that the intramolecular interactions might cooperatively repress activity. Such a switch would respond cooperatively to the combination of both competing external ligands (**Figure 2-2A**). Because of increased complexity of two-input switches, we created a combinatorial library in which switch design parameters including domain type, domain-ligand affinity, linker length, and domain architecture were varied (**Figure 2-2B**). To further increase variability two forms of the N-WASP output domain were used (long and short — both display constitutive activity)[13].

Two classes of switches were designed. For the first class — "chimeric" switches — target behavior was dual regulation by PDZ ligand and Cdc42, a non-native and native N-WASP regulator, respectively. These switches were constructed using a PDZ domain and the native N-WASP GBD as regulatory modules. The GBD binds a peptide within the N-WASP output region (residues 461-479), an interaction that is competitively disrupted by activated Cdc42[8]. Although the intramolecular GBD interaction is required for autoinhibition in native N-WASP, it is not sufficient: the interaction does not repress N-WASP activity unless combined with the autoinhibitory interaction of the B module (the

PIP$_2$ responsive element)[9]. For the second class — "heterologous" switches — target

behavior was dual regulation by PDZ and SH3 domain ligands, two non-native inputs.

These switches were constructed using the PDZ domain from α-syntrophin and the SH3

domain from Crk. SH3 domains recognize short proline-rich motifs[14, 15].

A library of 34 such switches (**Figure 2-2B**) was tested for gating by the appropriate high

affinity intermolecular ligands. Activity was tested in the presence of no inputs, each

individual input, and both inputs together. Like most signaling proteins, these modular

allosteric switches did not give simple binary responses; the precise response observed

depended on the input concentrations used. We therefore performed activation screens

under a standard set of input concentrations: 10 μM Cdc42-GTPγS, 200 μM PDZ ligand,

10 μM SH3 ligand. Each of these concentrations is 20 to 100-fold greater than the K$_d$

observed for input ligand binding to its isolated recognition domain.

Switches could be divided into diverse behavioral classes (**Figure 2-2C**). At the

extremes, 5 switches showed little or no basal repression, while 9 were extremely well-

repressed, but could not be activated under any of the tested conditions. Most constructs,

however, showed some type of gating behavior. Of the remaining 20 switches, 16 showed

positive gating (both inputs activate). Two of the proteins displayed antagonistic gating:

one input activates while the other represses (mechanism discussed later). The positively

gated dual input switches could be further subdivided. Two proteins showed OR-gate-like

behavior (roughly equivalent activation in the presence of either individual input or both

together), five proteins showed clear AND-gate-like behavior, while the remaining

constructs showed intermediate behavior. Thus, this relatively small library yielded a diversity of switch behaviors, including several with the targeted AND-gate behavior.

Several design principles were revealed by examining how switch parameters alter behavior. Basal repression and input sensitivity were directly linked to the affinity of autoinhibitory interactions. For example, the chimeric switch C11, which has an intramolecular PDZ ligand with $K_d = 8$ μM, was well-repressed under basal conditions but insensitive: it could not be activated by the standard concentrations of PDZ ligand or Cdc42, even in combination (**Figure 2-3A**). However, if the intramolecular PDZ ligand affinity was reduced ($K_{PDZ} = 100$ μM), the protein then resembled an AND-gate (switch C12).

Heterologous switch behavior was also dependent on affinity of the autoinhibitory interactions. For example, switch H15, which has internal SH3 and PDZ ligands with $K_{SH3} = 10$ μM and $K_{PDZ} = 100$ μM, resembled an OR-gate (**Figure 2-3B**). However, increasing the affinity of the internal PDZ ligand by ~10-fold ($K_{PDZ} = 8$ μM) within the same architecture yielded a well-behaved AND-gate (switch H14). Interestingly, in one architectural context, the 8 μM PDZ affinity was too high to yield AND-gate behavior (switch C11), whereas in another context this affinity was ideal (switch H14). This difference may be due to differences in the affinity of the partner domain; in C11 the partner domain is the GBD, which binds its internal ligand with $K_d = 1$ μM[9, 16], whereas in H14 the partner domain is an SH3 domain with $K_{SH3} = $ ~10 μM. Maintaining a balance

between switch repression and sensitivity may require balancing the affinities of the highly coupled autoinhibitory interactions.

Linker length also affected switch behavior. For example, if the linker length between the PDZ and SH3 domains in H14 was increased from 5 to 20 residues, the switch became more sensitive to the isolated inputs (switch H16), indicative of reduced domain coupling. This finding is consistent with observations that coupling between regulatory domains of Src family kinases depends strongly on conformational and energetic features of the interdomain linker[17]. Within this library, however, increasing interdomain linker length did not uniformly reduce coupling, suggesting these effects are context dependent.

Synthetic AND-gate switches were tested for targeted activation of actin polymerization in *Xenopus* oocyte extracts (**Figure 2-3C**). Carboxylated polystyrene beads were coated with Glutathione-S-transferase (GST) fusions to the relevant input ligands: no ligand (GST alone), SH3 ligand, PDZ ligand, or SH3 and PDZ ligands connected in tandem. When beads were incubated with soluble H14 switch and oocyte extract, actin filament nucleation was only observed on beads coated with the tandem SH3-PDZ ligand, consistent with multi-input targeting.

The combinatorial switch library also yielded switches with the unexpected behavior of antagonistic or negative input control (H1, H2) in which PDZ ligand acted as an activator, but SH3 ligand acted as a repressor (**Figure 2-4A**). Detailed examination of the gating properties of switch H2 in various input concentration regimes revealed that PDZ

ligand always acts as an activator; SH3 ligand, however, increased the basal level of repression (**Figure 2-4B**). Antagonistic regulation is consistent with a model in which the intramolecular PDZ interaction is solely responsible for autoinhibition, and the intramolecular SH3 interaction destabilizes the intramolecular PDZ interaction, but, by itself, has no direct effect on output activity (**Figure 2-4C**). We modeled this scheme by assuming that the state in which both intramolecular interactions take place is unfavorable and unpopulated (**Figure 2-S5**). Such a scheme predicted an activation surface (**Figure 2-4C**) resembling the observed behavior of switch H2 (**Figure 2-4B**). For related switches (H1-H3), the maximum level of repression observed (in the presence of SH3 ligand), directly correlated with PDZ affinity, a trend consistent with repression driven solely by the intramolecular PDZ interaction.

In this type of antagonistic switch, the two domains appear to act in a nested manner: the SH3 intramolecular interaction negatively regulates the PDZ intramolecular interaction, which in turn negatively regulates the output activity (**Figure 2-4C**). Addition of exogenous SH3 ligand, therefore, stabilizes the autoinhibitory PDZ interaction, leading to the observed inhibitory effect. In contrast, in positive integrating switches that resemble AND-gates, the two domains work in concert to negatively regulate output function (**Figure 2-4D**). Consequently, disruption of both intramolecular interactions yields activation.

This unanticipated class of switches highlights a striking feature of the library: subtle changes in switch parameters can lead to dramatic changes in gating behavior. The

architecture of antagonistic switches (H1, H2) is identical to a set of positive switches (H7-H12) except for the size of the output domain (long output in the antagonistic switches; short in the positive switches). The geometry of the output domain must have significant impact on the coupling between regulatory domains, presumably by altering stability of the various conformational states of the switch.

These results demonstrate that multi-domain signaling switches like N-WASP are functionally modular — diverse and complex gating behaviors can be generated through relatively simple recombination events between input and output domains, even among domains with no known evolutionary relationship. By allowing the establishment of novel regulatory relationships between molecules with no previous physiological relationship, such recombination events would be a powerful force driving evolution of novel cellular circuitry[18]. This interchangability exists because in modular allosteric switches, regions that mediate input control are physically separable from output regions. Facile interchange of gating properties is unlikely to occur in conventional allosteric proteins in which input and output activities are centralized in a single folded structure, and gating is mediated by subtle conformational shifts.

Domain recombination space sampled in these experiments proved functionally rich: although constructs showed a range of different gating behaviors (negative/positive; integrating/non-integrating; etc.), nearly all of them show some form of gating. Gating as an emergent property, therefore, does not appear to be extremely rare, as might be expected if only very precise domain arrangements yielded regulation. This modular

framework, in addition to promoting switch protein evolution, could be utilized to engineer proteins with novel regulatory control, and, in principle, novel cellular circuits.

Acknowledgements

Materials and Methods

Protein construction and purification

Switch proteins. Component domains and binding motifs used for switch protein construction are listed in **Table 2-S1**. Plasmids bearing these component DNA sequences were amplified by multi-step polymerase chain reaction (PCR), using primers that encoded for the appropriate intramolecular ligand motifs and linkers. All linkers were poly –Ser-Gly- repeats. Proteins were expressed as fusions to either a cleavable hexa-histidine tag (pET19-derived vector)[19] or glutathione S-transferase (GST) (pGEX4T, Pharmacia) in *Escherichia coli* (BL21-DE3) as previously described[9]. Desired protein

was purified by chromatography on Ni-NTA resin (Qiagen) or glutathione-agarose resin (Sigma). In the case of polyhistidine tagged proteins, the affinity tag was removed by incubation with tobacco etch virus (TEV) protease at 25°C for 2 hrs, after which the reaction mixture was passed over a second Nickel-NTA column. Proteins were further purified using Source S or Q columns (Pharmacia).

Cdc42. A soluble fragment of human Cdc42 (residues 1-179) for use as an input molecule was expressed in *Escherichia coli* (BL21-DE3) as a polyhistidine fusion, and purified as described above for switch proteins. Purified Cdc42 was activated by incubating with 10 fold excess of GTPγS for 15 min at 30°C, followed by addition of 50 fold excess $MgCl_2$ to quench the reaction. Charged Cdc42 was dialyzed in 50 mM KCl, 1 mM $MgSO_4$, 1 mM EGTA, and 10 mM imidazole (pH 7.0).

Peptide synthesis. Peptides were synthesized using conventional solid-phase Fmoc-amino acid chemistry. SH3 and PDZ domain ligand peptides were synthesized on Rink-Amide resin (Novabiochem) and Wang resin (Novabiochem), respectively. All peptides were N-terminally acetylated, cleaved, and purified as described[20]. SH3-peptide affinities were previously measured[21-23], while PDZ-binding affinities were measured by competition with dansylated-peptide as described[20].

Actin polymerization assays

Reagent preparation. Actin was purified from rabbit muscle, as described[24]. Rabbit skeletal muscle actin was pyrene-labeled as described[25]. Arp2/3 was purified from bovine

brain by a two-step purification scheme adapted from Ref 26. Briefly, brains were homogenized in an equal volume of buffer (50 mM PIPES (pH 6.8), 5 mM EGTA, 2 mM $MgCl_2$, 1 mM DTT, 0.1 PMSF, and 0.2 mM ATP) using a blender (Waring). Insoluble materials were separated by centrifugation and the supernatant was incubated with SP-sepharose (Amersham). Fractions eluted with the above buffer + 100 mM KCl were applied to an affinity column composed of GST-tagged human WASP residues 418-502. Arp2/3 was eluted with 200 mM $MgCl_2$, 150 mM NaCl, 0.2 mM ATP, 1 mM DTT, 10 mM imidazole (pH 7.5). Fractions containing protein were pooled and the concentration of $MgCl_2$ reduced to 0.2 mM by dialysis.

Pyrene actin polymerization assay. Actin polymerization assays were performed essentially as previously described[27] with modifications for use of a SpectraMax Gemini XS (Molecular Devices) fluorescent plate reader (excitation: 365 nm; emission: 407 nm). Briefly, a solution consisting of 10% pyrene-labeled actin was converted from a Ca-ATP to a Mg-ATP form by addition of $MgCl_2$ and EGTA to final concentrations of 50 μM and 200 μM, respectively. This solution was incubated at 25°C for 10 min in a 96 half area well plate (Corning) (10 μL per well). To initiate polymerization, the solution containing Arp2/3, switch construct, and appropriate ligand preincubated for 10 min at 25°C was added to the actin mix. Final assay conditions were 1.3 μM actin, 5 nM Arp2/3, 50 nM switch, 50 mM KCl, 1 mM $MgSO_4$, 1 mM EGTA, 0.2 mM ATP, 1 mM DTT, 3 μM $MgCl_2$, and 11.5 mM imidazole (pH 7.0) in a volume of 150 μL.

Raw fluorescence values from pyrene actin polymerization assays were normalized

relative to lower and upper baselines using the equation $(F_{data}-F_{low})/(F_{high}-F_{low})$, where

F_{data} is the fluorescence for each time point and F_{low} and F_{high} are average fluorescence

obtained at the lower and upper baselines, respectively. Reaction half-time ($t_{1/2}$) is

defined as the time required to reach 50% polymerization. A simple metric for relative

activity was based on the experimentally measured halftime (see **Figure 2-S1**): relative

activity = $(t_{1/2}^{max} - t_{1/2})/(t_{1/2}^{max} - t_{1/2}^{min})$ where $t_{1/2}^{max}$ is the halftime observed with no

activator (roughly equivalent to spontaneous actin polymerization; approx. 2000 sec.) and

$t_{1/2}^{min}$ is the halftime observed with the constitutively active output domain (approx. 400

sec. for output A and 600 sec. for output B). Relative activity was always calculated

using values for $t_{1/2}$, $t_{1/2}^{min}$, and $t_{1/2}^{max}$ measured simultaneously (same 96-well plate and

reagents). An alternative metric for activity was the rate of actin polymerization

(fluorescence units/sec) at half-polymerization.

Classification of switch behavior. Proteins from the 2-input switch library were

screened for relative activity under a set of standard input concentrations: chimeric

switches — no activators, 10 μM Cdc42-GTPγS, 200 μM PDZ lig, or both; heterologous

switches — no activators, 200 μM PDZ ligand, 10 μM SH3 ligand, or both. These

concentrations were arbitrarily chosen to be 20-100-fold higher than the K_d for the

interaction of the input molecule with the isolated recognition domain. Switches were

divided into behavioral classes shown in **Figure 2-2C** based on relative activation under

these conditions. Definitions for the classes are given in **Figure 2-S4**.

Modeling of switch behavior. See **Figure 2-S5**.

Bead actin polymerization assay. Spatial control of actin polymerization was assayed as follows. Carboxylated 2 μm polystyrene beads (Polysciences) were coated with GST-input fusions as described[28]. Briefly, 2 μm carboxylated polystyrene beads (2 μL of 2.5% slurry) were incubated in solution containing saturating amounts of GST-input fusion, mixed at the desired percentage with GST alone (no fusion), at a total protein concentration of 2 mg/mL in a volume of 160.5 μL. Beads were incubated with either 69.5 μM of monovalent ligand (GST-SH3 or GST-PDZ) or 34.8 μM of bivalent ligand (GST-SH3-PDZ). After incubation at room temperature for 1 hr, beads were pelleted and washed in XB buffer (100 mM KCl, 0.1 mM CaCl$_2$, 2 mM MgCl$_2$, 5 mM EGTA, 10 mM K•Hepes, pH 7.7). Washed beads were resuspended in 30 μL of XB. 0.5 μL of this bead solution was incubated with 1.18 μL of a 525 μM stock of soluble H14 protein for 15 min at 25°C, then supplemented with 4.5 μL of ~30 mg/mL *Xenopus* extract (generously donated by C. Co and J. Taunton) and 0.25 μL rhodamine-labeled actin prepared as previously described[29]. After incubation for 5 minutes at 25°C, a 1 μL aliquot was removed and sealed between a microscope slide and a 18 mm square coverslip with vaseline:lanolin:paraffin (at 1:1:1). After a further 5 min incubation, beads were observing with an Olympus 1X70 microscope equipped with brightfield and epifluorescent optics. Images were acquired with a Photometrics Cool Snap HQ camera (standard rhodamine filter set). Rhodamine and brightfield images were analyzed identically for each sample using Adobe Photoshop.

Table 2-S1 Components used in switch construction and their properties

Output	Source	Residues	Ref.
Output A	rat N-WASP	392-501	11
Output B	rat N-WASP	429-501	13

Regulatory Domains	Source	Residues	Ref.
PDZ	Mouse α-syntrophin (syn)	77-171	30
GBD	rat N-WASP	196-274	8
SH3	mouse Crk	134-191	31

Intramol. ligands	Sequence	Partner	K_d (μM)	Ref.
PDZ lig. H	GVKESLV	Syn PDZ	8[a]	20, This work
PDZ lig. M	GVKQSLL	Syn PDZ	100[a]	This work
PDZ lig. L	GVKESGA	Syn PDZ	1000[a]	This work
SH3 lig. H	PPPVPPRR	Crk SH3	10	32
SH3 lig. M	PPAIPPRQPT	Crk SH3	100	23
SH3 lig. L	GPPVPPRQST	Crk SH3	700	23
Output (C helix)	Rat N-WASP residues 461-479 (within Output A and B)	GBD	1	9

Input Ligands	Sequence	Partner	Kd (uM)	Ref.
PDZ lig.	Ac-YVKESLV-COOH	Syn PDZ	8	20, This work
SH3 lig.	Ac-PPPALPPKRRR-CONH2	Crk SH3	0.1	21
Cdc42-GTPγS	Residues 1-179	GBD	0.1	9

Intramolecular ligand affinities for partner domains are reported as measured *in trans*.

[a] Affinities measured with a peptides with an N-terminal tyrosine replacing the glycine to accurately measure peptide concentration. PDZ peptides were synthesized with a N-terminal acetyl group and a C-terminal carboxyl group.
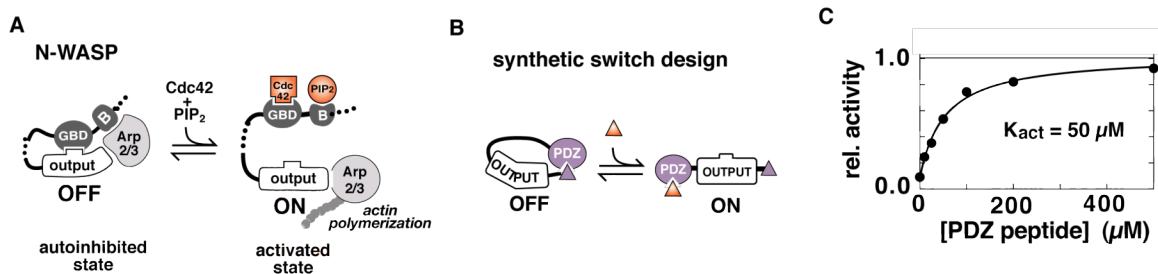
Figure 2-1 Design of synthetic switch gated by heterologous ligand
A, N-WASP is a modular allosteric switch: its output domain constitutively stimulates Arp2/3-mediated actin polymerization, but is repressed by autoinhibitory interactions involving two domains, the GTPase binding domain (GBD) and a Basic (B) motif. Input ligands activate by disrupting autoinhibitory interactions: GTP-loaded Cdc42 binds GBD; PIP_2 binds B motif. These two inputs act synergistically[9, 11]; thus, N-WASP resembles an AND-gate. **B,** Design strategy for a synthetic single input switch using N-WASP's output domain and a PDZ domain-ligand pair as heterologous autoinhibitory module (α-syntrophin PDZ; ligand — NH$_2$-GVKESLV-COOH; $K_d = 8$ μM). **C,** Synthetic switch protein is basally repressed but can be activated by addition of exogenous PDZ ligand. Switches were tested using an *in vitro* pyrene actin polymerization assay (**Figure 2-S1**), using time required to reach 50% polymerization ($t_{1/2}$) as activity metric. Basal repression is observed only in constructs containing the intramolecular PDZ domain-ligand pair (**Figure 2-S2**). Peptide concentration required for half maximal activation (K_{act}) is 50 μM. Studies with variant switches show that degree of repression is correlated with affinity of the intramolecular ligand (K_{PDZ}), while sensitivity to external PDZ ligand shows an inverse correlation (**Figure 2-S3**). Assays in this and all other figures were performed with 50 nM switch protein, 5 nM Arp2/3, and 1.3 μM actin (10% pyrene-actin).

B

ARCHITECTURE	PARAMETERS					INPUT/OUTPUT BEHAVIOR				BEHAVIOR CLASS
	construct code	linker length (no. residues) n	m	output domain	intramol. ligand affinity K_{PDZ} K_{GBD}	Cdc42: − + − + / PDZ lig.: − − + +				
Control Constructs										
OUTPUT A (res. 392-501)		−	−	A	− −					
OUTPUT B (res. 429-501)		−	−	B	− −					
No Switch		−	−	−	− −					
Chimeric (GBD/PDZ) switches										
	C1.	0	9	A	100 1					AND
	C2.	1	9	A	100 1					AND
	C3.	1	9	A	1000 1					Sing
	C4.	5	9	A	8 1					Ins
	C5.	5	9	A	100 1					Ins
	C6.	9	9	A	100 1					Ins
	C7.	20	9	A	100 1					Ins
	C8.	20	9	A	1000 1					OR
	C9.	0	9	A	8 1					Ins
	C10.	1	9	A	100 1					Int
	C11.	5	9	A	8 1					Ins
	C12.	5	9	A	100 1					AND
	C13.	5	9	A	1000 1					Int
	C14.	9	9	A	8 1					Ins
	C15.	20	9	A	8 1					Ins
	C16.	20	9	A	100 1					Int

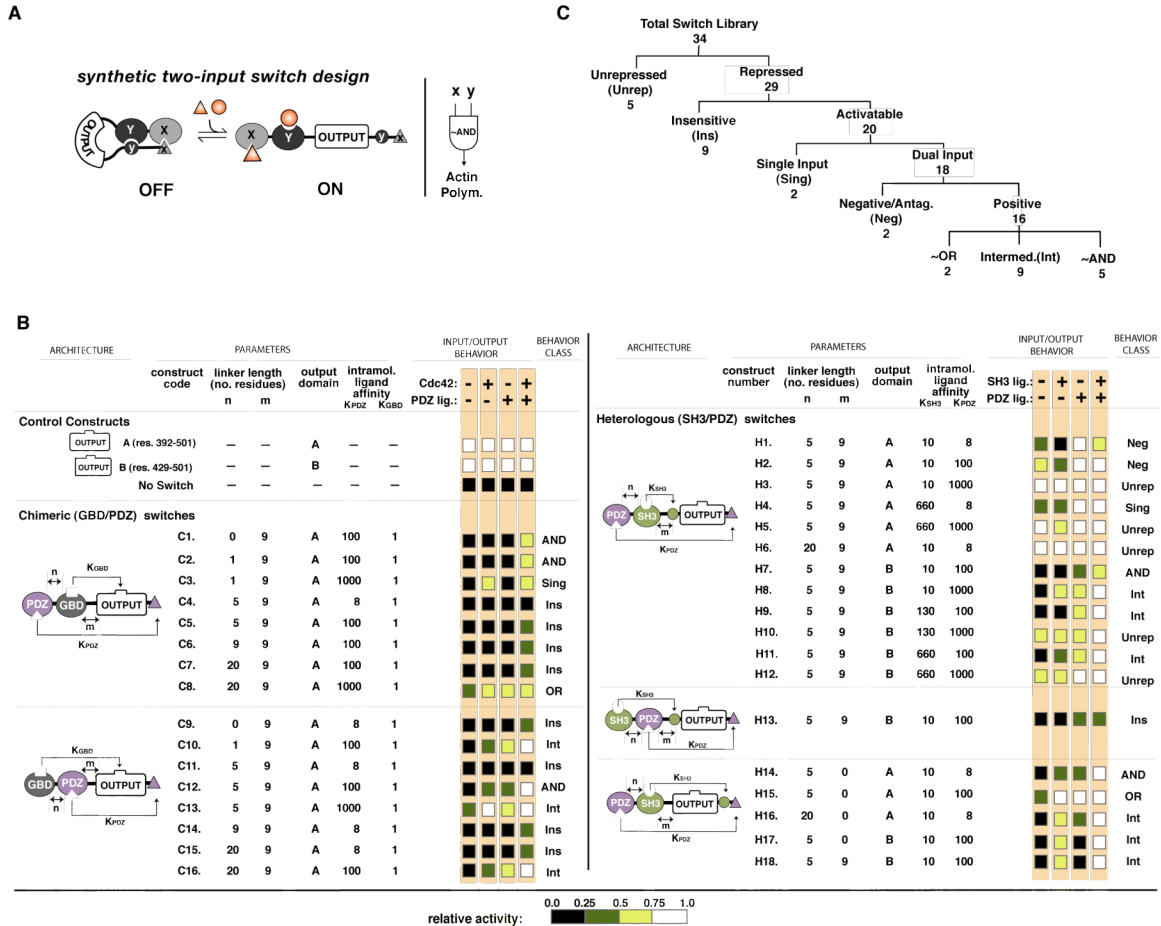| ARCHITECTURE	PARAMETERS					INPUT/OUTPUT BEHAVIOR	BEHAVIOR CLASS
	construct number	linker length (no. residues) n	m	output domain	intramol. ligand affinity K_{SH3} K_{PDZ}	SH3 lig.: − + − + / PDZ lig.: − − + +	
Heterologous (SH3/PDZ) switches							
	H1.	5	9	A	10 8		Neg
	H2.	5	9	A	10 100		Neg
	H3.	5	9	A	10 1000		Unrep
	H4.	5	9	A	660 8		Sing
	H5.	5	9	A	660 1000		Unrep
	H6.	20	9	A	10 8		Unrep
	H7.	5	9	B	10 100		AND
	H8.	5	9	B	10 1000		Int
	H9.	5	9	B	130 100		Int
	H10.	5	9	B	130 1000		Unrep
	H11.	5	9	B	660 100		Int
	H12.	5	9	B	660 1000		Unrep
	H13.	5	9	B	10 100		Ins
	H14.	5	0	A	10 8		AND
	H15.	5	0	A	10 100		OR
	H16.	20	0	A	10 8		Int
	H17.	5	0	B	10 100		Int
	H18.	5	9	B	10 100		Int

relative activity: 0.0 0.25 0.5 0.75 1.0

A

synthetic two-input switch design

OFF ⇌ ON

x y → AND → Actin Polym.

C

Total Switch Library — 34
- Unrepressed (Unrep) — 5
- Repressed — 29
  - Insensitive (Ins) — 9
  - Activatable — 20
    - Single Input (Sing) — 2
    - Dual Input — 18
      - Negative/Antag. (Neg) — 2
      - Positive — 16
        - ~OR — 2
        - Intermed.(Int) — 9
        - ~AND — 5

Figure 2-2 Design of synthetic dual-input switch library
A, Two-input switch design strategy. **B,** Switch library constructed by domain recombination. Components used (**Table 2-S1**) are: two output domains of N-WASP (output A and B) which differ in length; three different input domains (PDZ, SH3, GBD); from one to three intramolecular ligands of differing affinities for each of the input domains (ligand for GBD is contained within the output domain); and four different length interdomain linkers (Gly-Ser repeats). Switch architecture and design parameters are listed at left. Observed gating behavior is listed at right. Activity of library members was screened in the presence of no inputs, each individual input, and both inputs simultaneously using a standard set of input concentrations (Cdc42•GTPγS: 10 µM; PDZ ligand: 200 µM; SH3 ligand: 10 µM; all of these concentrations are 20-100-fold greater than the K_d for input binding to its isolated recognition domain). Relative activity (measured as in **Figure 2-S1**) under these conditions is indicated by a color code (from low to high: black, green, yellow, white). **C,** Classes of gating behavior observed in the library (see **Figure 2-S4** for class definitions).
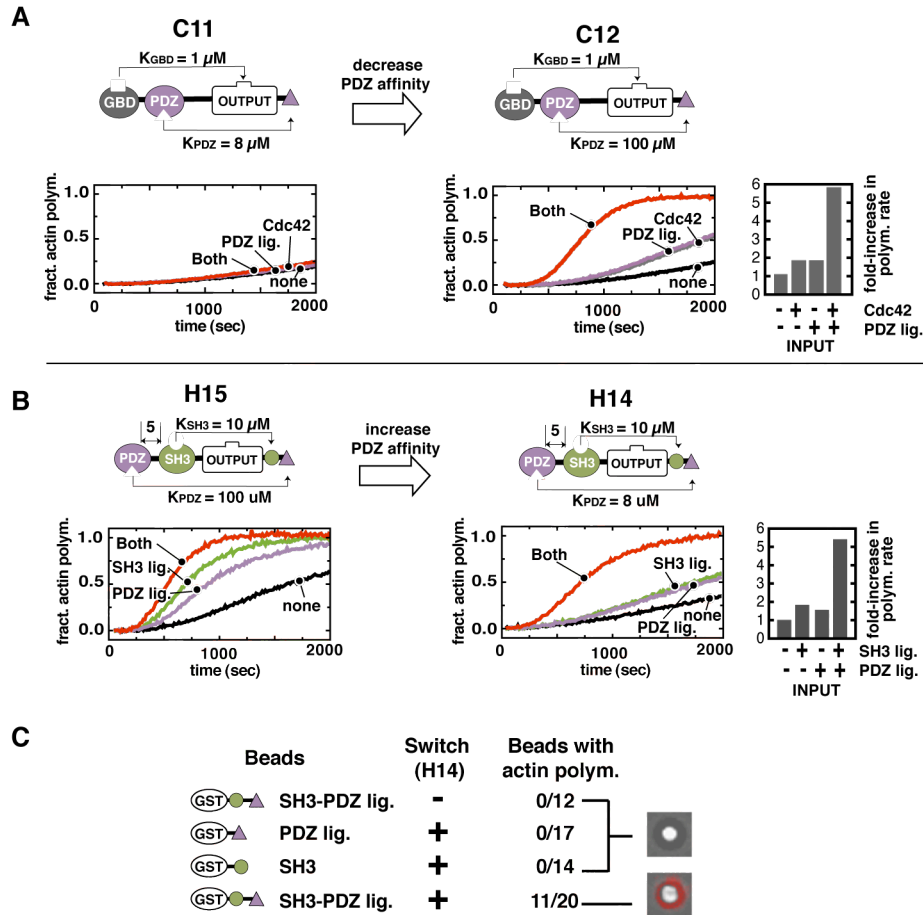
Figure 2-3 Synthetic switches that resemble AND-gates
A, Chimeric switch C12 (right) resembles an AND-gate — it shows strong actin polymerization only in the presence of both PDZ ligand and Cdc42. Adjacent bar graph shows maximal polymerization rates under each condition normalized to the basal rate (no input). Related switch (C11) with identical architecture but a higher affinity intramolecular PDZ ligand (left) is insensitive/over-repressed. **B,** Heterologous switch H14 (right) resembles an AND-gate that responds to SH3 and PDZ ligand. A related switch (H15) with identical architecture but a weaker affinity intramolecular PDZ ligand (left) resembles an OR-gate — individual ligands yield relatively strong activation. **C,** Switch H14, which resembles an AND-gate, can spatially target actin polymerization in a *Xenopus* oocyte extract. Polystyrene beads were coated with GST-fusions to either no ligand, SH3 ligand, PDZ ligand, or a tandem SH3-PDZ ligand (see Methods). Tandem ligand was used at half concentration relative to monovalent ligands. Only beads coated with the tandem ligand and incubated with switch H14 (100 μM) nucleated polymerization of rhodamine labeled actin (red). Merge of brightfield and fluorescence images are shown. Fraction of beads displaying actin polymerization is given.
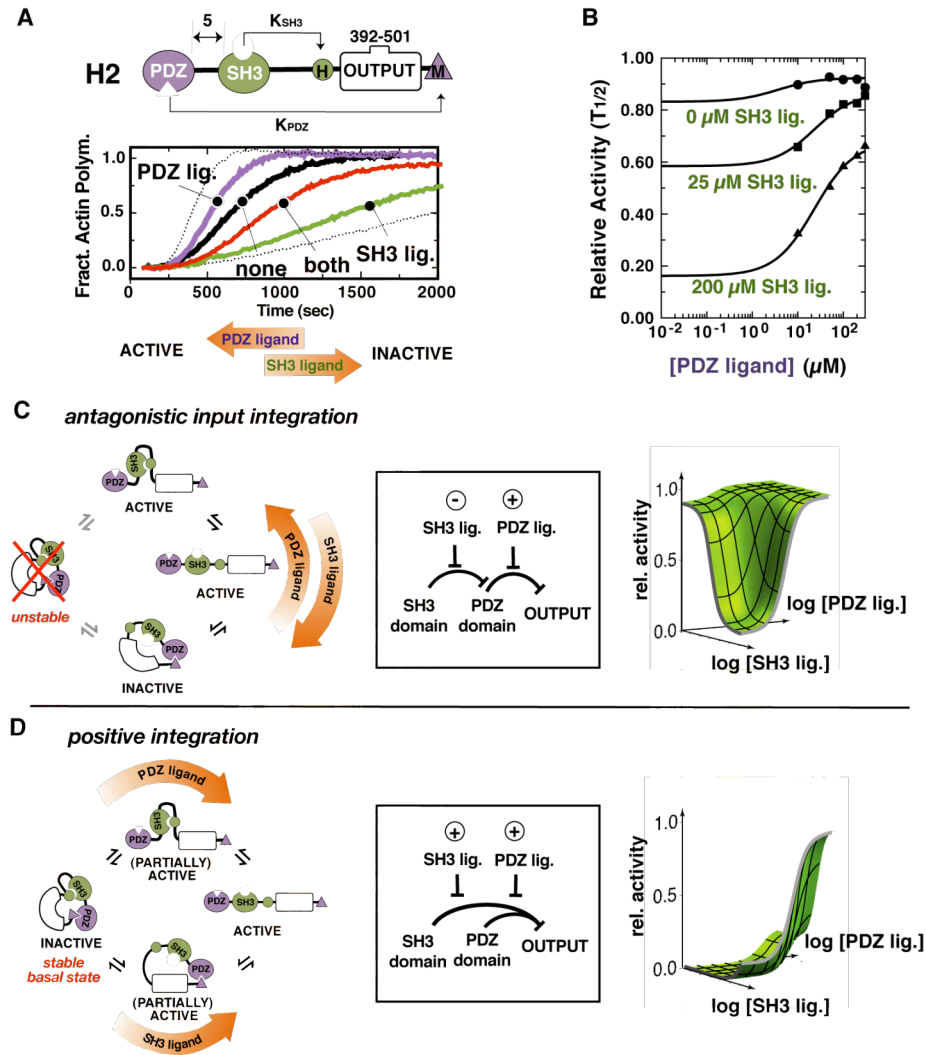
Figure 2-4 Mechanism of antagonistic switch
A, PDZ and SH3 ligands have opposing effects on the activity of switch H2 — PDZ ligand is an activator, SH3 ligand is a repressor. **B,** Effect of PDZ ligand on switch H2 activity in the presence of different, constant concentrations of SH3 ligand. **C,** Antagonistic behavior of switch H2 can be explained by a model in which the SH3 and PDZ intramolecular interactions are anti-cooperative (i.e. the state with both intramolecular interactions is unfavorable and unpopulated), and the intramolecular PDZ interaction mediates autoinhibition. Although the intramolecular SH3 interaction is neutral, it indirectly relieves repression by opposing the intramolecular PDZ interaction. A simple circuit diagram shows how this nested series of regulatory interactions yields antagonistic input control. Positive and negative net effects of inputs are indicated. Modeling of such a switch predicts an activation surface (green) consistent with experimental behavior (**Figure 2-S5**). **D,** Model of switches that resemble AND-gates. Both intramolecular domain interactions contribute to autoinhibition. Thus, both ligands are positive regulators. Modeling yields an activation surface (green) consistent with more potent activation in the presence of both ligands simultaneously (**Figure 2-S5**).

51
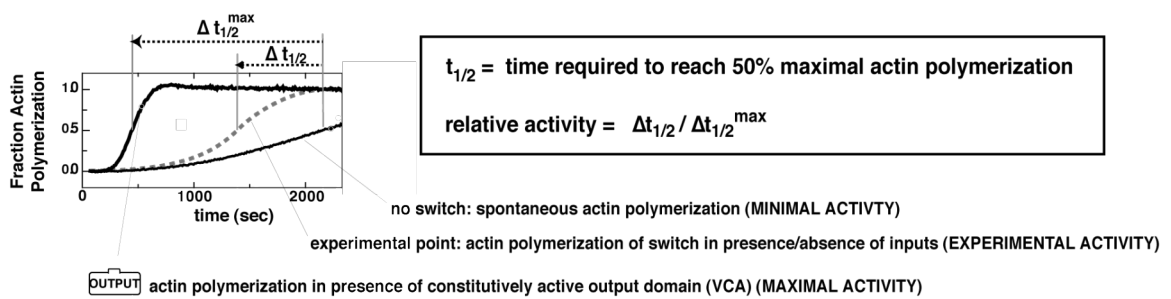
Figure 2-S1 Metric for relative activity of N-WASP switches based on half-time of polymerization

Activity of switch proteins was determined using a fluorescence-based actin polymerization assay[12] (see Methods). Time required to reach 50% polymerization ($t_{1/2}$) was used as a metric for activity. Minimal activity was defined as the $t_{1/2}$ observed with spontaneous actin polymerization under these conditions in the presence of Arp2/3 but no nucleation promoting factors. Maximal activity was defined as the $t_{1/2}$ in the presence of the isolated output domain. Relative activities of individual constructs were scored by measuring the change in $t_{1/2}$ relative to the difference between maximum and minimum activities.
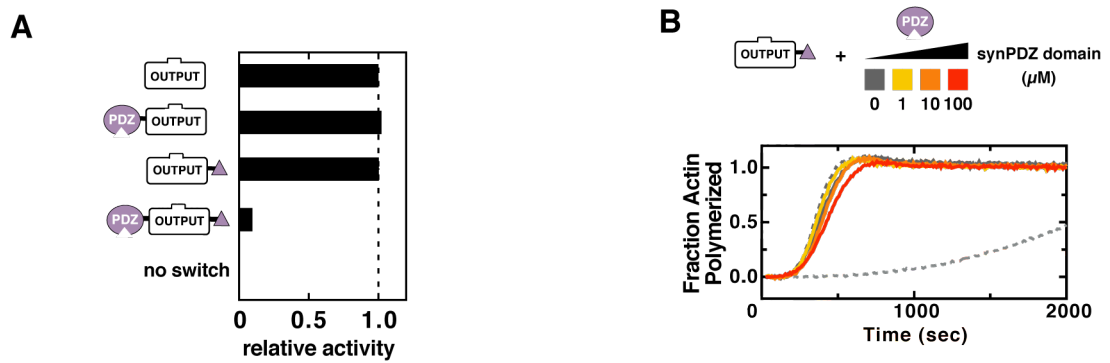
Figure 2-S2 Autoinhibition requires intramolecular interaction
A, Control constructs containing either the PDZ domain or ligand alone did not exhibit significant repression. **B,** Titration of PDZ domain to the output domain with the PDZ ligand covalently attached did not exhibit appreciable repression even at a concentration >10 fold above the K_d.
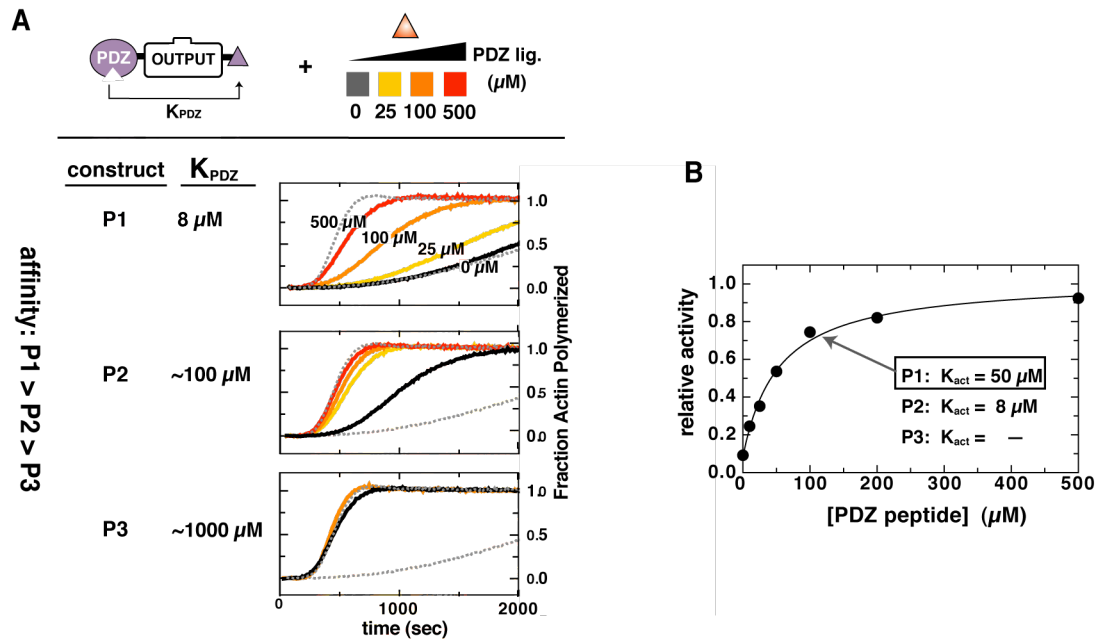
Figure 2-S3 Dependence of single input switch behavior on autoinhibitory ligand affinity
A, Three different affinity PDZ domain/ligand pairs where examined within the same single input architecture. Reduced affinity of this autoinhibitory ligand resulted in reduced basal repression. **B,** Titration of free PDZ ligand (K_{PDZ} = 8 μM) relieved activity in a manner consistent with a single-ligand binding event. Reduced affinity of this autoinhibitory ligand resulted in increased sensitivity to free ligand.

Poorly Repressed:
Relative activity >0.5 under all input conditions

Insensitive:
Relative activity <0.5 under all input conditions

Single Input:
One input is an activator; other input has no effect

Negative dual input:
one input is activator; other input is repressor

Positive dual input:
both inputs are activators

~OR:
inputs either individually or in combination yield approx. equivalent level of activation

~AND:
inputs in combination yield much stronger activation than individually

Intermed:
remaining positive dual input switches

Total Switch Lbrary

Poorly Repressed   **Repressed**

Insensitive, over-repressed   **Activatable**

Single Input   **Dual Input**

Negative (Antagonistic)   **Positive**

~OR   **Intermed.**   **~AND**
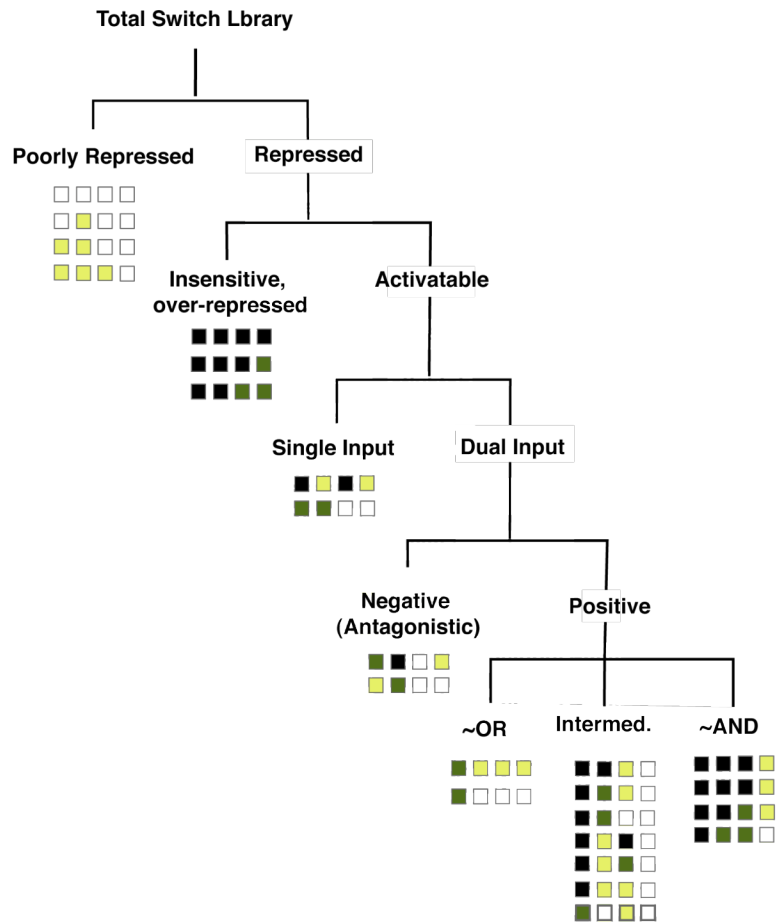
All experimentally observed switch profiles are classified above.
Color code relative actiivty levels are defined in Figure 2-2.

Figure 2-S4 Switch behavior class definitions and examples from two-input library
All experimentally observed switch profiles shown in **Figure 2-2** are classified above in addition to definitions used for each of these behavior classes.

55

A. Framework for switch modeling

Equilibria describing two-input switch states and effect of ligand binding:

$$X_A \cdot B + A \overset{\longleftarrow}{\underset{K_b}{\longrightarrow}} X_A + A + B \overset{\frac{K_B}{c}}{\underset{}{\longleftrightarrow}} X_{AB} + A + B$$

$$K_A \updownarrow \qquad\qquad K_A \updownarrow \qquad\qquad \frac{K_A}{c} \updownarrow$$

$$X \cdot B + A \overset{\longleftarrow}{\underset{K_b}{\longrightarrow}} X + A + B \overset{K_B}{\underset{}{\longleftrightarrow}} X_B + A + B$$

$$\updownarrow K_a \qquad\qquad \updownarrow K_a \qquad\qquad \updownarrow K_a$$

$$X \cdot A \cdot B \overset{\longleftarrow}{\underset{K_b}{\longrightarrow}} X \cdot A + B \overset{K_B}{\underset{}{\longleftrightarrow}} X_B \cdot A + B$$

$$X_A = \frac{X}{K_A} \qquad\qquad X_B = \frac{X}{K_B} \qquad\qquad X_{AB} = \frac{X_B}{\frac{K_A}{c}} = \frac{c\,X}{K_A K_B}$$

$$X \cdot A = \frac{X\ A}{K_a} \qquad X \cdot B = \frac{X\ B}{K_b} \qquad X \cdot A \cdot B = \frac{X \cdot A\ B}{K_b} = \frac{X\ A\ B}{K_a K_b}$$
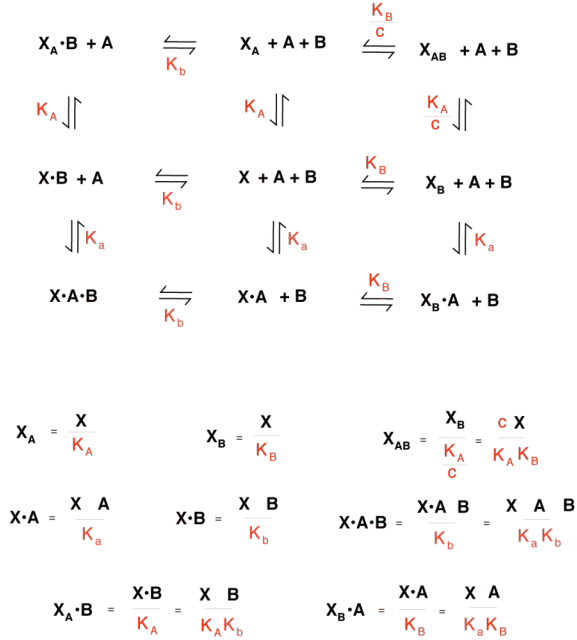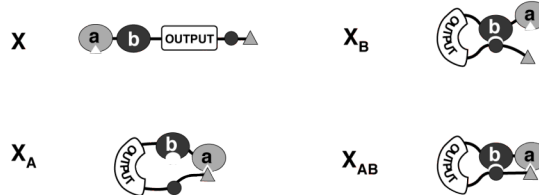
$$X_A \cdot B = \frac{X \cdot B}{K_A} = \frac{X\ B}{K_A K_b} \qquad X_B \cdot A = \frac{X \cdot A}{K_B} = \frac{X\ A}{K_a K_B}$$

Switch States:

X

X_B

X_A

X_{AB}

A and B are external ligands for domains a and b, respectively

Constants:

K_A , K_B : "dissociation" constant for intramolecular binding by domain a or b, respectively. This term encompasses both the inherent affinity of the isolated domain for ligand and their effective concentration within the intramolecular context.

K_a , K_b : dissociation constant for intermolecular ligand binding to domain a or b, respectively.

c : effective concentration linking intramolecular binding of domain a with intramolecular binding of domain b; i.e. fold change in intramolecular association of one domain when the other is already engaged in intramolecular interaction.

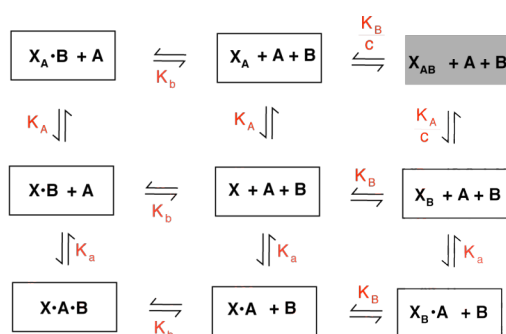**Figure 2-S5** Modeling of two input switch behavior
(continued on following page)

56

B. Modeling specific cases

Specific cases are modeled below, including AND-gates and antagonistic switches.

There are two extreme models that generate AND-gate like behavior[33]. In model I, both intramolecular interactions are required for repression (i.e. states with single intramolecular interaction are fully active). However, the two intramolecular interactions are highly cooperative (c high). Thus, in ligand mediated activation, the two intermolecular ligands act in a highly cooperative manner. In model II, the states with single intramolecular interactions are also inactive. Thus, in the absence of cooperativity both inputs are required for full activation. Many other intermediate models (i.e. hybrids of the two extremes; partial activity of single intramolecular states, etc.) are possible and also yield general AND-gate-like behavior.

   i. "AND-gate," model I - highly cooperative:
   · c is large
   · only X_{AB} is inactive (shaded); remaining states of X are active (white)

$$\text{Rel. Act.} = \frac{X + X{\cdot}A + X{\cdot}B + X{\cdot}A{\cdot}B + X_A + X_A{\cdot}B + X_B + X_B{\cdot}A}{X + X{\cdot}A + X{\cdot}B + X{\cdot}A{\cdot}B + X_A + X_A{\cdot}B + X_B + X_B{\cdot}A + X_{AB}}$$

$$= \frac{1 + A' + B' + A'B' + 1/K_A + B'/K_A + 1/K_B + A'/K_B}{1 + A' + B' + A'B' + 1/K_A + B'/K_A + 1/K_B + A'/K_B + c/K_A K_B}$$
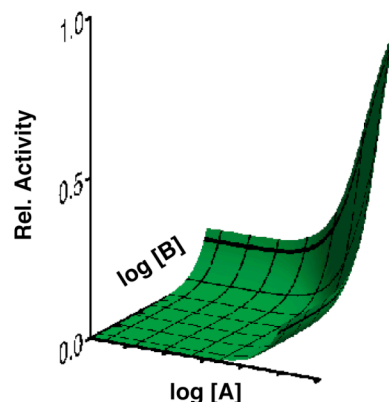
$(A' = A/K_A \; ; \; B' = B/K_B)$

$K_A = K_B = 0.05\ \mu M$

$c = 300$

Figure 2-S5 Modeling of two input switch behavior
(continued on following page)

57

II. "AND-gate," model II semi-cooperative switch:
- X_A, X_B, and X_{AB} are inactive (shaded); only X is active (white)

$$X_A \cdot B + A \quad \underset{K_b}{\rightleftharpoons} \quad X_A + A + B \quad \overset{\frac{K_B}{c}}{\rightleftharpoons} \quad X_{AB} + A + B$$

$$K_A \updownarrow \qquad\qquad K_A \updownarrow \qquad\qquad \frac{K_A}{c} \updownarrow$$

$$X \cdot B + A \quad \underset{K_b}{\rightleftharpoons} \quad X + A + B \quad \overset{K_B}{\rightleftharpoons} \quad X_B + A + B$$

$$\updownarrow K_a \qquad\qquad \updownarrow K_a \qquad\qquad \updownarrow K_a$$

$$X \cdot A \cdot B \quad \underset{K_b}{\rightleftharpoons} \quad X \cdot A + B \quad \overset{K_B}{\rightleftharpoons} \quad X_B \cdot A + B$$
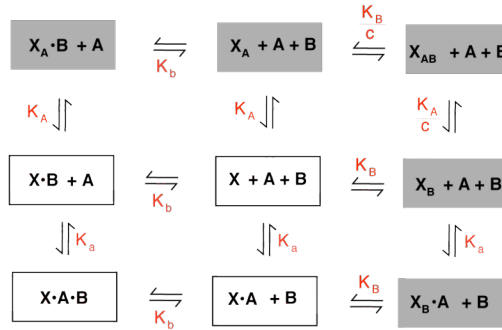
$$\text{Rel. Act.} = \frac{X + X \cdot A + X \cdot B + X \cdot A \cdot B}{X + X \cdot A + X \cdot B + X \cdot A \cdot B + X_A + X_A \cdot B + X_B + X_B \cdot A + X_{AB}}$$

$$= \frac{1 + A' + B' + A'B'}{1 + A' + B' + A'B' + 1/K_A + B'/K_A + 1/K_B + A'/K_B + c/K_A K_B} \qquad (A' = A/K_A \; ; \; B' = B/K_B)$$
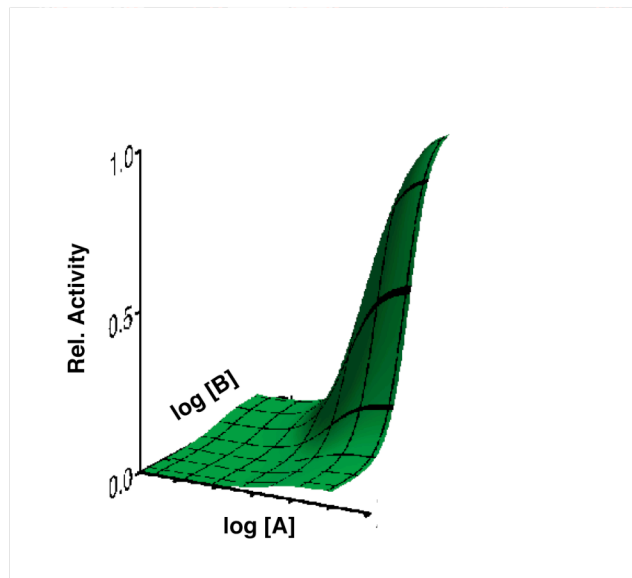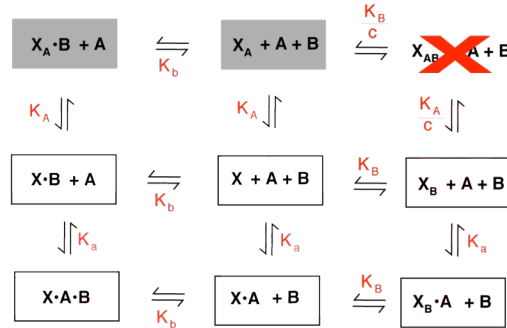
$$K_A = K_B = 0.05 \ \mu M$$

$$c = 1$$

Figure 2-S5 Modeling of two input switch behavior
(continued on following page)

iii. Antagonistic switch:
· **c ~ 0 (anticooperative intramolecular interactions); X_{AB} state is unoccupied**
· **only X_A is inactive (shaded); X and X_B are active (white)**

$$X_A \cdot B + A \underset{K_b}{\rightleftharpoons} X_A + A + B \overset{\frac{K_B}{c}}{\rightleftharpoons} X_{AB} \cdot A + B$$

$$K_A \updownarrow \qquad\qquad K_A \updownarrow \qquad\qquad \frac{K_A}{c} \updownarrow$$

$$X \cdot B + A \underset{K_b}{\rightleftharpoons} X + A + B \overset{K_B}{\rightleftharpoons} X_B + A + B$$

$$\updownarrow K_a \qquad\qquad \updownarrow K_a \qquad\qquad \updownarrow K_a$$

$$X \cdot A \cdot B \underset{K_b}{\rightleftharpoons} X \cdot A + B \overset{K_B}{\rightleftharpoons} X_B \cdot A + B$$

Rel. Act. $= \dfrac{X + X \cdot A + X \cdot B + X \cdot A \cdot B + X_B + X_B \cdot A}{X + X \cdot A + X \cdot B + X \cdot A \cdot B + X_B + X_B \cdot A + X_A + X_A \cdot B}$

$= \dfrac{1 + A' + B' + A'B' + 1/K_B + A'/K_B}{1 + A' + B' + A'B' + 1/K_B + A'/K_B + 1/K_A + B'/K_A}$

$(A' = A/K_A; \ B' = B/K_B)$

$K_A = K_B = 0.05 \ \mu M$

$c \sim 0$
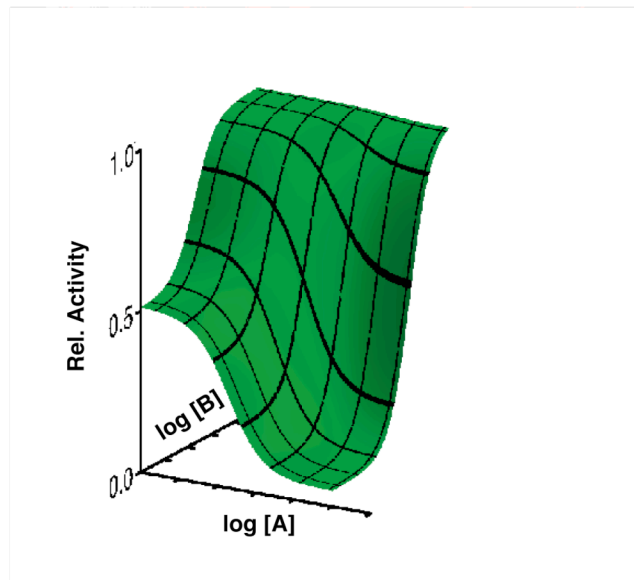
(graph shown in Figure 2-4D was calculated with $K_A = 0.02 \ \mu M$ and $K_B = 0.2 \ \mu M$)

Figure 2-S5 Modeling of two input switch behavior
A, Framework for switch modeling related as equilibria describing two-input switch states and effect of ligand binding. **B,** Modeling of specific cases including AND-gates and antagonistic switches.

59

References

1.   Pawson, T. & Nash, P. Protein-protein interactions define specificity in signal transduction. *Genes Dev* **14**, 1027-1047 (2000).

2.   Pawson, T. & Nash, P. Assembly of cell regulatory systems through protein interaction domains. *Science* **300**, 445-452 (2003).

3.   Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).

4.   Lim, W.A. The modular logic of signaling proteins: building allosteric switches from simple binding domains. *Curr Opin Struct Biol* **12**, 61-68 (2002).

5.   Pufall, M.A. & Graves, B.J. AUTOINHIBITORY DOMAINS: Modular Effectors of Cellular Regulation. *Annu Rev Cell Dev Biol* **18**, 421-462 (2002).

6.   Mullins, R.D. How WASP-family proteins and the Arp2/3 complex convert intracellular signals into cytoskeletal structures. *Curr Opin Cell Biol* **12**, 91-96 (2000).

7.   Higgs, H.N. & Pollard, T.D. Regulation of actin filament network formation through ARP2/3 complex: activation by a diverse array of proteins. *Annu Rev Biochem* **70**, 649-676 (2001).

8.   Kim, A.S., Kakalis, L.T., Abdul-Manan, N., Liu, G.A. & Rosen, M.K. Autoinhibition and activation mechanisms of the Wiskott-Aldrich syndrome protein. *Nature* **404**, 151-158 (2000).

9. Prehoda, K.E., Scott, J.A., Mullins, R.D. & Lim, W.A. Integration of multiple signals through cooperative regulation of the N-WASP-Arp2/3 complex. *Science* **290**, 801-806 (2000).

10. Rohatgi, R., Ho, H.Y. & Kirschner, M.W. Mechanism of N-WASP activation by CDC42 and phosphatidylinositol 4, 5-bisphosphate. *J Cell Biol* **150**, 1299-1310 (2000).

11. Rohatgi, R. *et al.* The interaction between N-WASP and the Arp2/3 complex links Cdc42-dependent signals to actin assembly. *Cell* **97**, 221-231 (1999).

12. Mullins, R.D. & Machesky, L.M. Actin assembly mediated by Arp2/3 complex and WASP family proteins. *Methods Enzymol* **325**, 214-237 (2000).

13. Zalevsky, J., Lempert, L., Kranitz, H. & Mullins, R.D. Different WASP family proteins stimulate different Arp2/3 complex-dependent actin-nucleating activities. *Curr Biol* **11**, 1903-1913 (2001).

14. Mayer, B.J. SH3 domains: complexity in moderation. *J Cell Sci* **114**, 1253-1263 (2001).

15. Zarrinpar, A., Bhattacharyya, R.P. & Lim, W.A. The structure and function of proline recognition domains. *Sci STKE* **2003**, RE8 (2003).

16. Buck, M., Xu, W. & Rosen, M.K. Global disruption of the WASP autoinhibited structure on Cdc42 binding. Ligand displacement as a novel method for monitoring amide hydrogen exchange. *Biochemistry* **40**, 14115-14122 (2001).

17. Young, M.A., Gonfloni, S., Superti-Furga, G., Roux, B. & Kuriyan, J. Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell* **105**, 115-126 (2001).

18.   Monod, J., Changeux, J.-P. & Jacob, J. Allosteric Proteins and Cellular Control Systems. *J Mol Biol* **6**, 306-329 (1963).

19.   Hillier, B.J., Christopherson, K.S., Prehoda, K.E., Bredt, D.S. & Lim, W.A. Unexpected modes of PDZ domain scaffolding revealed by structure of nNOS-syntrophin complex. *Science* **284**, 812-815 (1999).

20.   Harris, B.Z., Hillier, B.J. & Lim, W.A. Energetic determinants of internal motif recognition by PDZ domains. *Biochemistry* **40**, 5921-5930 (2001).

21.   Posern, G. *et al.* Development of highly selective SH3 binding peptides for Crk and CRKL which disrupt Crk-complexes with DOCK180, SoS and C3G. *Oncogene* **16**, 1903-1912 (1998).

22.   Nguyen, J.T. *et al.* Improving SH3 domain ligand selectivity using a non-natural scaffold. *Chem Biol* **7**, 463-473 (2000).

23.   Knudsen, B.S. *et al.* Affinity and specificity requirements for the first Src homology 3 domain of the Crk proteins. *Embo J* **14**, 2191-2198 (1995).

24.   Pardee, J.D. & Spudich, J.A. Purification of muscle actin. *Methods Enzymol* **85 Pt B**, 164-181 (1982).

25.   MacLean-Fletcher, S. & Pollard, T.D. Mechanism of action of cytochalasin B on actin. *Cell* **20**, 329-341 (1980).

26.   Egile, C. *et al.* Activation of the CDC42 effector N-WASP by the Shigella flexneri IcsA protein promotes actin nucleation by Arp2/3 complex and bacterial actin-based motility. *J Cell Biol* **146**, 1319-1332 (1999).

27. Machesky, L.M. *et al.* Scar, a WASp-related protein, activates nucleation of actin filaments by the Arp2/3 complex. *Proc Natl Acad Sci U S A* **96**, 3739-3744 (1999).

28. Cameron, L.A., Footer, M.J., van Oudenaarden, A. & Theriot, J.A. Motility of ActA protein-coated microspheres driven by actin polymerization. *Proc Natl Acad Sci U S A* **96**, 4908-4913 (1999).

29. Taunton, J. *et al.* Actin-dependent propulsion of endosomes and lysosomes by recruitment of N-WASP. *J Cell Biol* **148**, 519-530 (2000).

30. Schultz, J. *et al.* Specific interactions between the syntrophin PDZ domain and voltage-gated sodium channels. *Nat Struct Biol* **5**, 19-24 (1998).

31. Wu, X. *et al.* Structural basis for the specific interaction of lysine-containing proline-rich peptides with the N-terminal SH3 domain of c-Crk. *Structure* **3**, 215-226 (1995).

32. Nguyen, J.T., Turck, C.W., Cohen, F.E., Zuckermann, R.N. & Lim, W.A. Exploiting the basis of proline recognition by SH3 and WW domains: design of N-substituted inhibitors. *Science* **282**, 2088-2092 (1998).

33. Prehoda, K.E., & Lim, W.A. How signaling proteins integrate multiple inputs: a comparison of N-WASP and Cdk2. *Curr Opin Cell Biol* **14**, 149-154 (2002).

Chapter 3

Rewiring Cellular Morphology Pathways With Synthetic

Guanine Nucleotide Exchange Factors

Summary

Eukaryotic cells mobilize the actin cytoskeleton to generate a remarkable diversity of morphological behaviors, including motility, phagocytosis, and cytokinesis. Much of this diversity is mediated by guanine nucleotide exchange factors (GEFs) that activate Rho family GTPases — the master regulators of the actin cytoskeleton[1-3] (**Figure 3-1a**). There are over 80 Rho GEFs in the human genome (compared to only 22 genes for the Rho GTPases themselves), and the evolution of new and diverse GEFs is thought to provide a mechanism for linking the core cytoskeletal machinery to a wide range of novel control inputs. Here we test this hypothesis and ask if we can systematically reprogram cellular morphology by engineering synthetic GEF proteins. We focused on Dbl-family Rho GEFs, which have a highly modular structure common to many signaling proteins[4,5]: they contain a catalytic Dbl homology (DH) domain linked to diverse regulatory domains, many of which autoinhibit GEF activity[2,3] (**Figure 3-1b**). We show that by recombining catalytic GEF domains with novel regulatory modules, we can generate synthetic GEFs that are activated by non-native inputs. We have used these synthetic GEFs to reprogram cellular behavior in diverse ways. The GEFs can be used to link specific cytoskeletal responses to normally unrelated upstream signaling pathways. In addition, multiple synthetic GEFs can be linked as components in series to form an artificial cascade with improved signal processing behavior. These results show the high degree of evolutionary plasticity of this important family of modular signaling proteins, and suggest that it may be possible to use synthetic biology approaches to manipulate the complex spatiotemporal control of cell morphology.

Introduction

Rho family GTPases are central signaling molecules in the regulation of the actin cytoskeleton[1]. These proteins are conformational switches that exist in GDP- and GTP-bound states; only the GTP-bound state actively transduces signal to downstream effectors. Cycling between states is primarily controlled by opposing enzymes: GTPase activating proteins (GAPs) promote hydrolysis of bound GTP to GDP (inactivation), while GEFs promote exchange of bound GDP for GTP (activation). The three canonical members of the Rho family — Cdc42, Rac1, and RhoA — stimulate distinct morphological outputs: protrusive filopodia (thin actin microspikes), protrusive lamellipodia (broad membrane ruffles) and contractile actin:myosin filaments, respectively.

Results and Discussion

As an initial target for rewiring GTPase signaling, we attempted to reprogram Dbl-family GEFs so that their activity was controlled by Protein Kinase A (PKA), a well-characterized prototypical kinase[6] (**Figure 3-2a**). We first designed a PKA-sensitive autoinhibitory module inspired by natural examples[7]: a PDZ domain-peptide interaction pair that could be disrupted by PKA phosphorylation. The syntrophin PDZ domain recognizes short C-terminal peptide motifs (consensus sequence: R/K-E-S/T-x-ψ-COOH; ψ denotes aliphatic residues)[8], which are close in sequence to the ideal PKA substrate

(RRRRSIIFI)[9]. A hybrid sequence (RRRESIV-COOH) could serve both as an interaction ligand for the syntrophin PDZ domain and a PKA substrate. Most importantly, we found that phosphorylation by PKA disrupted binding to the PDZ domain (**Figure 3-S1**).

To build a Cdc42 GEF that could be activated by PKA, we fused this PKA-sensitive PDZ-peptide interaction module to the DH-PH catalytic core from Intersectin — a Cdc42-specific Dbl-family member whose catalytic activity is normally regulated by autoinhibitory SH3 domains[10, 11] (**Figure 3-2b**). We refer to this construct as **GEF1** (see **Tables 3-S1** and **3-S2** for details of all synthetic GEFs). In an *in vitro* Cdc42 nucleotide exchange assay, **GEF1** was repressed relative to the constitutively-active DH-PH fragment ($< 20\%$ activity), indicating that the intramolecular PDZ interaction sterically occluded or conformationally disrupted the DH-PH domain (**Figures 3-2c,d; 3-S2**). Phosphorylation of **GEF1** by PKA relieved repression, increasing Cdc42 exchange activity (**Figures 3-2c,d; 3-S2**). As a control, we mutated the peptide to a sequence that could still bind the PDZ domain but could not be phosphorylated by PKA. A construct bearing this mutation (**GEF1***) was still repressed, but was not activated by PKA (**Figures 3-2d, 3-S2**).

To test if the PKA regulatory module could be transferred to another GEF, we replaced the Intersectin DH-PH with the N-terminal DH domain of Trio, which preferentially activates Rac1 (**GEF2**)[12]. **GEF2** was also repressed *in vitro* (relative to the Trio DH domain alone), and could be activated by PKA (**Figures 3-2d, 3-S2**). A control construct

bearing a non-phosphorylatable peptide (**GEF2***) could not be activated by PKA (**Figure 3-S3**).

In total, we fused the PDZ-peptide module to the DH and/or DH-PH fragments of five Dbl-family members with varying GTPase specificities (including Intersectin and Trio), and tested their activity *in vitro*. All seven constructs tested showed some degree of repression under basal conditions, and four of seven were activated by PKA (**Table 3-S3**). No attempts were made to optimize autoinhibitory affinity, domain orientation, or interdomain linker lengths, and it is likely that such efforts would improve activation of the three remaining synthetic GEFs[13].

To test if these synthetic GEF proteins could create new functional signaling linkages *in vivo*, we introduced **GEF1** and **GEF2** into cells by microinjection. We first tested the effect of microinjecting the unregulated catalytic GEF modules into the REF52 fibroblast cell line (**Figure 3-3a**). As expected, microinjection of the Trio DH domain led to a constitutive Rac1-associated lamellipodial phenotype. Microinjection of the Intersectin DH-PH module yielded a constitutive Cdc42-associated filopodial phenotype in a large fraction of cells; however, a significant but inconsistent fraction of these cells showed an alternative rounded phenotype that is distinct from filopodia and lamellipodia (**Figure 3-S4**). Co-injection of additional Cdc42 with the Intersectin DH-PH resulted only in cells with filopodia, alleviating this dual phenotype problem. Thus, to simplify phenotypic scoring, we co-injected the relevant GTPases in all of the following experiments. We estimate that we are increasing the cellular concentration of the specific GTPases by

approximately two-fold, which has no morphological effect in the absence of the GEF domain. This method has previously been used to clarify scoring of GEF-induced phenotypes[14].

Injection of **GEF1** into REF52 cells resulted in a novel PKA-activated filopodial response. After microinjecting the purified proteins, we tested the cellular response to stimulation with increasing doses of forskolin, a pharmacological activator of PKA[15] (**Figure 3-3b**). When **GEF1** was injected into cells, even in the absence of PKA stimulation, there was a weak background activity; 14% of the cells showed filopodia, probably due to somewhat leaky repression of GEF activity. However, this phenotype was much weaker than that observed with injection of an equivalent amount of the unregulated DH-PH module (> 95% of cells with filopodia). Most importantly, filopodia were stimulated in a dose-dependent fashion as a function of forskolin concentration: > 60% of the cells had filopodia at the highest forskolin concentrations tested (**Figure 3-3c**). Furthermore, induction of filopodia was observed within minutes of forskolin addition to cells pre-injected with **GEF1** (Supplementary Movie), demonstrating the rapid timescale of response with protein-based networks that do not require transcription and translation. Forskolin treatment of cells lacking **GEF1** (injected only with Cdc42) led to a small background stimulation of filopodia (~ 20%), indicating that there is only a weak endogenous linkage between PKA and filopodia formation in REF52 cells. As an important control, we observed no significant stimulation of filopodia in cells microinjected with **GEF1***, which is autoinhibited but cannot be activated by PKA.

These results imply that the strong stimulation of filopodia is the result of a new, functional signaling connection mediated directly by the engineered **GEF1** protein.

Similarly, injection of **GEF2** into REF52 cells resulted in a PKA-inducible lamellipodial response. Injection of **GEF2** and Rac1 had little basal effect on the cells (4% of cells with lamellipodia); however, treatment with forskolin resulted in a dose-dependent increase in the number of cells with lamellipodia (> 60%) (**Figure 3-3d**). Activation of lamellipodia also occured within minutes of stimulation (data not shown). Cells injected with **GEF2*** showed no significant lamellipodial response to forskolin (**Figure 3-S3**). Thus, both synthetic GEFs are capable of mediating linkages between the endogenous PKA signaling pathway and Rho GTPase-mediated morphological rearrangements in live cells.

Many complex behaviors observed in living cells are mediated by multiple signaling proteins that do not function alone, but which instead are linked into more complex multistep pathways[16]. For example, the canonical GTPase Ras can activate multiple effectors, including the Rac1 GEF Tiam1[17]. Thus, we asked whether we could link synthetic GEFs with specifically engineered input-output linkages into a two GTPase cascade in which PKA would activate Cdc42, and Cdc42 would in turn activate Rac1 (**Figure 3-4a**). **GEF1** could provide the connection between PKA and Cdc42; however, the second step required a Cdc42-responsive autoinhibitory module, which we extracted from the signaling protein N-WASP: a GTPase-binding domain (GBD) that recognizes a short cofilin (C) peptide. The GBD-C interaction is normally involved in autoinhibition of N-WASP, and can be disrupted by activated Cdc42[18-21]. We fused the GBD-C module

70

to the Trio DH domain, producing a Rac1-specific GEF that is activated by Cdc42 (**GEF3**). *In vitro* analysis of **GEF3** showed that its Rac1 exchange activity was regulated by Cdc42, as expected (data not shown), providing further evidence for the flexibility of this overall framework for engineering diverse signaling linkages.

Co-injection of **GEF1** and **GEF3** (along with Cdc42 and Rac1) into REF52 cells resulted in a novel signaling cascade: PKA stimulation by forskolin ultimately led to Rac1 activation and a lamellipodial phenotype (**Figure 3-4b**). Almost no filopodial response was observed, perhaps because lamellipodia tend to be dominant over filopodia, and because much of the activated Cdc42 may be sequestered by binding to **GEF3**, instead of other effectors. To confirm that signal is passing through both synthetic GEFs, we disrupted each individual component. **GEF3** was selectively disrupted by a small deletion in the GBD that blocks binding to Cdc42(GTP) but does not affect autoinhibition (**GEF3***)[20, 21]. When **GEF1** and **GEF3*** were injected into REF52 cells, forskolin treatment led only to the activation of Cdc42, resulting in robust formation of filopodia (**Figure 3-4b**). Similarly, **GEF1*** is a variant of **GEF1** that cannot be phosphorylated and activated by PKA. As predicted, cells injected with **GEF1*** and **GEF3** showed no activation of either Cdc42 or Rac1 (no significant filopodial or lamellipodial response) upon forskolin treatment. Together, these results imply that **GEF1** and **GEF3** form a functional signaling cascade that links PKA to a lamellopodial response.

The **GEF1**-**GEF3** cascade demonstrated several properties that distinguished it from the direct, single GEF circuits (**Figures 3-4c, 3-S5**). First, the synthetic cascade had

dampened noise: there was a ~2-fold reduction in basal response (no forskolin stimulation), both in terms of filopodial and lamellipodial output, when **GEF3** was introduced downstream of **GEF1**. Second, the cascade appeared to amplify response within a certain range of stimulation. In the direct PKA→Cdc42 circuit (**GEF1** only), the amount of Cdc42 activated by 1-2 μM forskolin was insufficient to mount a significant filopodial response. However, in the **GEF1**-**GEF3** cascade, this low Cdc42 activation was sufficient to activate **GEF3**, producing a stronger Rac1-mediated lamellipodial response. Third, the cascade is ultrasensitive — it had a sharp activation threshold with an apparent Hill coefficient (n_H) of > 4, despite the fact that its individual components respond in a linear (Michaelian) fashion. The increased ultrasensitivity of the cascade is consistent with theoretical and experimental studies that compared pathways with increasing number of steps[16, 22]. While individual signaling proteins can exhibit non-linear behaviors[13], these simple synthetic GEFs can be linked into higher order architectures that begin to show complex emergent properties.

Here we demonstrate that Rho GEFs provide a flexible framework for engineering novel signaling pathways. Modular recombination allows the expansion of GTPase control relationships beyond those generated through evolution. GTPases regulate many biological processes (nuclear trafficking, endocytosis, etc.)[23]; thus, such approaches could be applied to manipulate these processes. These findings, along with related studies, demonstrate that modular protein signaling components can be engineered in a relatively facile manner, suggesting that it may be possible to apply synthetic biology approaches to generate cells with precisely engineered target behaviors[5, 24-27]. Although there has been

significant progress in engineering transcriptional networks in living cells[28, 29], there are comparatively fewer examples of synthetic signal transduction networks. These protein-based networks are important because they mediate many of the rapid and spatially precise responses in cells, including complex properties such as cell shape and movement. The ability to manipulate these properties will be critical for engineering cells with diverse therapeutic and biotechnological applications.

Acknowledgements

Author Contributions

B.J.Y. and R.J.R. contributed equally to this work. B.J.Y., R.J.R., D.B.-S., and W.A.L. conceived the experiments. B.J.Y. and A.D. designed and purified the constructs and performed the *in vitro* experiments. R.J.R. performed the *in vivo* experiments.

Materials and Methods

Protein construction, expression, and purification

Synthetic GEFs. GEF domains (DH and/or DH-PH) were cloned by polymerase chain reaction from human or mouse first-strand cDNA. pET-19b (Novagen) derived plasmids encoding synthetic GEFs were constructed using conventional restriction-enzyme molecular biology (sequence details in **Tables 3-S1**, **3-S2**, and **3-S3**). Proteins were expressed as hexahistidine fusions in *Escherichia coli* BL21(DE3)RIL, and purified by chromatography on Ni-NTA resin (Qiagen). Hexahistidine tags were removed by incubation with tobacco etch virus (TEV) protease at room temperature. Uncleaved protein, free hexahistidine tag, and protease were removed by subsequent incubation with Ni-NTA resin. Proteins were dialyzed into 20 mM Tris, 50 mM NaCl, 2 mM DTT, pH 7.5 (for *in vitro* nucleotide exchange assay) or Microinjection Buffer (PBS, 200 mM NaCl, 20 mM HEPES, 5 mM $MgCl_2$, pH 7.4), flash frozen in liquid nitrogen, and stored at -80°C.

GTPases. For nucleotide exchange assays, fragments of human Cdc42 (residues 1-179), human Rac1 (residues 1-177), and human RhoA (residues 1-190, C190S) lacking C-terminal prenylation sequences were expressed as hexahistidine fusions and purified as described above for synthetic GEFs. GTPases were further purified on a Source Q column (Amersham). Residual bound nucleotide was removed by dialysis in 20 mM Tris, 50 mM NaCl, 5 mM EDTA, 2 mM DTT, pH 7.5. GTPases were loaded with GDP or methylanthraniloyl-GDP (mant-GDP, Molecular Probes) by incubation with excess

74

nucleotide. Nucleotide exchange was quenched by addition of 50-fold molar excess of MgCl$_2$, and excess nucleotide was removed by dialysis into GEF Assay Buffer (20 mM Tris, 50 mM NaCl, 10 mM MgCl$_2$, 1% glycerol, 1 mM DTT, pH 7.5). GTPases loaded with GDP were flash frozen in liquid nitrogen, and stored at -80°C. GTPases loaded with mant-GDP were stored at 4°C, and used within 1 week of nucleotide loading.

For microinjection experiments, full-length human Cdc42 and Rac1 were expressed as Glutathione S-Transferase (GST) fusions in BL21(DE3)RIL (pGEX-4T1 vector, Amersham). GTPases were purified on glutathione-agarose resin (Sigma). GST tags were removed by incubation with thrombin (Calbiochem) at room temperature. Uncleaved fusion protein and free GST were removed by subsequent incubation with glutathione-agarose, and thrombin was removed by incubation with benzamidine-sepharose (Amersham). GTPases were further purified on a Source Q column, dialyzed into Microinjection Buffer, flash frozen in liquid nitrogen, and stored at -80°C.

PKA. The catalytic (C$_\alpha$) subunit of PKA was cloned from mouse first-strand cDNA, and expressed as a hexahistidine fusion in BL21(DE3)pLysS. PKA was purified on Ni-NTA resin, the hexahistidine tag was removed by incubation with TEV protease, and PKA was further purified on a Source S (Amersham) column. PKA was dialyzed into 20 mM Tris, 50 mM NaCl, 1 mM EDTA, 2 mM DTT, pH 7.5, flash frozen, and stored at -80°C.

Verification of PKA-sensitive interaction module

The candidate (RRRESIV) and mutant (GVKESLV) ligands (**Figure 3-S1**) were

expressed as GST-fusions in BL21(DE3)RIL, and immobilized on glutathione-agarose

resin. GST-peptides were tested for phosphorylation by incubation with PKA in the

presence of 200 µM ATP at 30°C. Phosphorylated and unphosphorylated GST-peptides

were incubated with 25 µM His6-syntrophin PDZ (expressed and purified as previously

described[31]) for 15 minutes at 4°C. Glutathione-agarose beads were washed and

resuspended in SDS-PAGE loading buffer. Samples were separated by SDS-PAGE and

transferred to nitrocellulose for Western blotting. To assess phosphorylation,

nitrocellulose membranes were visualized with Phospho-(Ser/Thr) PKA Substrate

Antibody (Cell Signaling Technology). To assess binding to syntrophin PDZ domain,

membranes were visualized with His-probe (Santa Cruz Biotechnology).

in vitro *nucleotide exchange assays*

Qualitative assays. Dissociation of mant-GDP from Cdc42 was measured using a

SpectraMax Gemini XS (Molecular Devices) fluorescence multi-well plate reader (25°C,

excitation: 360 nm, emission: 440 nm). Solutions were pre-equilibrated at 25°C for 10

minutes, and the reaction was initiated by mixing solutions of GEF/GDP and

Cdc42(mant-GDP). Final concentrations were 1 µM Cdc42(mant-GDP), 200 nM GEF,

200 µM GDP in GEF Assay Buffer.

Quantitative assays. Relative activities of synthetic GEFs were quantified using a

variation of the above assay in which an increase in fluorescence was observed following

the incorporation of mant-GDP into GTPases[10] (see **Figure 3-S2** for raw data). Solutions were pre-equilibrated at 25°C for 10 minutes, and the reaction was initiated by mixing solutions of GEF/mant-GDP and GDP-loaded GTPase. Final concentrations were 1 μM Cdc42(GDP), 25 nM GEF, 400 nM mant-GDP in GEF Assay Buffer for reactions involving synthetic GEFs based on Intersectin. For reactions involving GEFs based on Trio, final concentrations were 1 μM Rac1(GDP), 250 nM GEF, 400 nM mant-GDP.

Activity was quantified by determining the slope of the initial linear phase of the exchange reaction, and normalized to reactions involving no GEF and DH or DH-PH alone: relative activity = $(\text{slope}_{\text{experimental}} - \text{slope}_{\text{no GEF}})/(\text{slope}_{\text{DH/DH-PH alone}} - \text{slope}_{\text{no GEF}})$. Relative activity measured in this fashion, under these conditions, was linear with respect to GEF concentration. We prefer quantifying activity using initial slope (as opposed to fitting to exponential functions) because it does not require making any assumptions about reaction mechanism.

Activation by PKA. GEFs were pre-incubated with PKA at a 1:10 kinase:GEF molar ratio, in GEF Assay Buffer supplemented with 200 μM ATP, at 30°C for 30 minutes. (The relatively high kinase:GEF ratio was chosen to ensure complete or near-complete phosphorylation.) GEF/kinase mixture was then used in exchange assays as described above. For quantitative assays, activity was normalized to equivalently treated samples containing no GEF or DH/DH-PH alone.

Microinjection experiments

Cell culture. Rat embryo fibroblasts (REF52) were grown at 37°C and cultured in Dulbecco's modified Eagles's media (DMEM, Gibco) supplemented with 10% fetal bovine serum (FBS, Gibco) in a humidified incubator with 6.3% CO_2. Cells were cultured and plated at low passage number.

Microinjection. For microinjection experiments, cells were plated on ethanol washed non-gridded glass coverslips (Bellco) and grown as sub-confluent monolayers overnight and serum-starved in DMEM medium containing 0.5% FBS for 24 hours prior to injection. Microinjection was performed using an Eppendorf 5246 pressure system and an Eppendorf 5171 microinjector. Proteins were diluted in Microinjection Buffer and injected into the cytosol of cells. 3000 MW, anionic, lysine fixable, fluorescein-labeled dextran (Molecular Probes) was used as an injection marker at a concentration of 1 mg/mL (**Figure 3-S6**). Constitutively-active Intersectin DH-PH was titrated to give maximal filopodial output at 2 mg/mL (in the microinjection needle). Subsequently, all other GEFs were injected at the molar equivalent of 2 mg/mL Intersectin DH-PH. We estimate the final cellular concentration of injected GEFs to be 0.1-1 μM, which is consistent with experimentally measured concentrations of Rho GEFs[32]. Additionally, the GEFs were co-injected with their associated GTPases (0.5 mg/mL) to facilitate scoring. Importantly, filopodia and lamellipodia are not induced in cells microinjected with Cdc42 or Rac1 alone.

Fixed cell experiments. Injected cells were incubated for 30 minutes at 37°C, treated

with indicated concentration of forskolin (Alexis Biochemicals)[33], and incubated for an

additional 30 minutes. Cells were washed twice with PBS and replaced in pre-incubated

DMEM containing 0.5% FBS for 30 minutes at 37°C and 6.3% CO_2 to allow for recovery

from any deleterious effects of the drug. Cells were then fixed in 3.7%

formaldehyde/PBS for 1 hour at room temperature. After fixation, cells were washed

three times for 5 minutes with PBS, permeabilized with 0.1% Triton X-100 (Sigma) in

PBS for 3 minutes at room temperature followed by three PBS washes for 5 minutes

each. Non-specific binding was blocked using 1% bovine serum albumin (BSA, Sigma)

in PBS for 5 minutes. Filamentous actin was stained by incubation with rhodamine-

phalloidin (0.2 mg/mL in block solution, Molecular Probes) for 1 hour at 37°C in a

humidified incubator. Coverslips were washed three times for 5 minutes in PBS, rinsed in

distilled water, and mounted onto glass slides using 10 µL Immu-mount (Shandon)

containing 0.04% paranitrodiphenylene (PADA, Sigma) as an anti-fade agent. Slides

were visualized on Axiovert 200M or S100 microscopes (Zeiss); fluorescent micrographs

were captured using the Axiovision software (Zeiss). Morphological phenotypes were

scored in a blind fashion (without knowledge of the experimental condition). Cells

displaying at least 5 protrusive spikes were scored positive for filopodia, and cells that

displayed dense peripheral actin staining were scored positive for lamellipodia (no cells

were observed with both filopodia and lamellipodia). The percentage of cells with each

phenotype was calculated by dividing the number of cells with the scored phenotype by

the total number of cells scored.

Time-lapse microscopy. REF52 cells were plated on glass bottom dishes (MatTek) in DMEM supplemented with 10% FBS, and serum-starved 24 hours prior to injection. Cells were microinjected with indicated protein constructs and placed in a humidified thermo chamber ventilated with a heated air-CO_2 mixture (6.3% CO_2, 37˚C) (Zeiss Tempcontrol 37-2 digital, Heating Insert, CTI-Controller). Live injected cells were identified by fluorescein-labeled dextran and images were captured every 20 seconds with a 63X DIC oil lens (Zeiss). For stimulation with forskolin, the chamber was opened, media was aspirated, and 2 mL pre-incubated DMEM with 0.5% FBS supplemented with 10 µM forskolin in DMSO was added. Cells were refocused and imaging resumed at 20 second intervals. Cells were observed using an Axiovert 200M microscope (Zeiss); micrographs were captured using the Axiovision software (Zeiss).

Table 3-S1 Regulatory modules used for construction of synthetic GEFs

Domain	Parent Protein	Residues
PDZ	mouse α-syntrophin	77-171
GBD	rat N-WASP	196-274
GBD*	rat N-WASP	209-274
C	rat N-WASP	458-489

Table 3-S2 Synthetic GEFs described in this study

GEF	Composition	Substrate	Input
GEF1	[PDZ]-GT-[Intersectin DHPH]-TGRRRESIV	Cdc42	Activated by PKA
GEF1*	[PDZ]-GT-[Intersectin DHPH]-TGVKESLV	Cdc42	Not activated by PKA
GEF2	[PDZ]-GT-[Trio N-term. DH]-TGRRRESIV	Rac1	Activated by PKA
GEF2*	[PDZ]-GT-[Trio N-term. DH]-TGVKESLV	Rac1	Not Activated by PKA
GEF3	[GBD]-GT-[Trio N-term. DH]-TG-[C]	Rac1	Activated by Cdc42(GTP)
GEF3*	[GBD*]-GT-[Trio N-term. DH]-TG-[C]	Rac1	Not activated by Cdc42(GTP)

Table 3-S3 GEFs tested for regulation by PKA-sensitive interaction module
(Qualitative activities are relative to most-active GEF tested for each GTPase)

Parent Protein	Frag-ment	Residue Numbers	Activity of DH(PH) alone vs.			Activity of PDZ-DH(PH)-RRRESIV		
			Cdc42	Rac1	RhoA	vs.	- PKA	+ PKA
Intersectin 1L (Human)[10]	DH	1229-1445	+++	+	-	Cdc42	+	+++
	DHPH	1229-1580	++++	+	-	Cdc42	++	++++
Trio (Human)[12]	DH	1284-1477	-	++	+	Rac1	+	++
	DHPH	1284-1594	-	++++	+	Rac1	+++	++++
Tiam1 (Mouse)[34]	DH	1033-1259	-	+	-	not tested		
	DHPH	1033-1406	-	++++	+	Rac1	+	+
Prex1 (Human)[35]	DH	38-258	++++	++++	+	Cdc42	+	+
	DHPH	38-415	insoluble			not tested		
Tim (Human)[10]	DH	1166-1367	-	-	++++	RhoA	++	++
	DHPH	1166-1527	insoluble			not tested		

Table 3-S4 Specificities of **GEF1** and **GEF2**
(Qualitative activities scaled as in Table S3)

Synthetic GEF	- PKA			+ PKA		
	Cdc42	*Rac1*	*RhoA*	*Cdc42*	*Rac1*	*RhoA*
GEF1	++	+	-	++++	+	-
GEF2	-	+	+	-	++	+
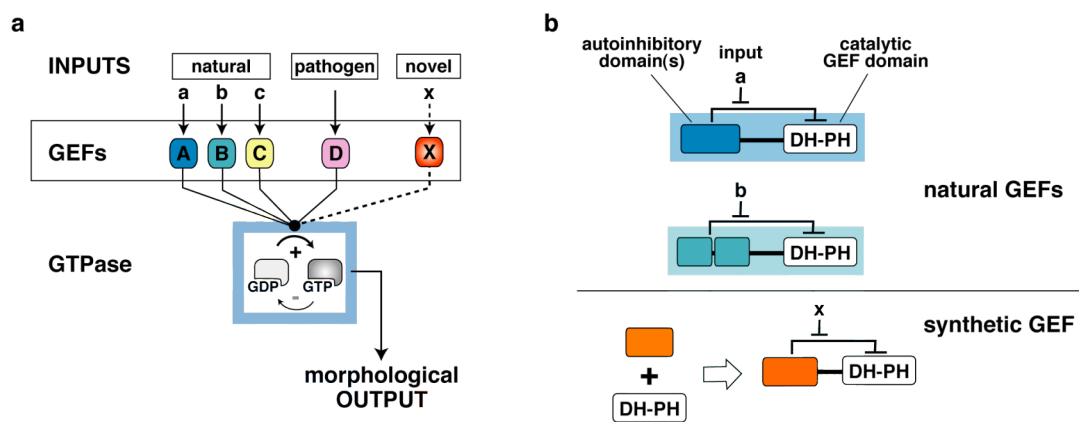
Figure 3-1 GEFs link diverse inputs to Rho GTPase modules that control cell morphology

a, GEFs functionally connect signaling inputs to activation of Rho GTPases, which regulate morphology of the actin cytoskeleton. Some bacterial pathogens encode GEFs that activate host GTPases[30]. Synthetic GEFs could, in principle, mediate novel connections in living cells. **b**, The largest family of Rho GEFs are Dbl-related proteins, which share a catalytic DH-PH core. In many cases, adjacent modular domains mediate autoinhibitory interactions that can be disrupted by specific inputs. Here we exploit this modular structure to construct synthetic GEFs.

Figure 3-2 Modular recombination yields PKA-responsive synthetic GEFs
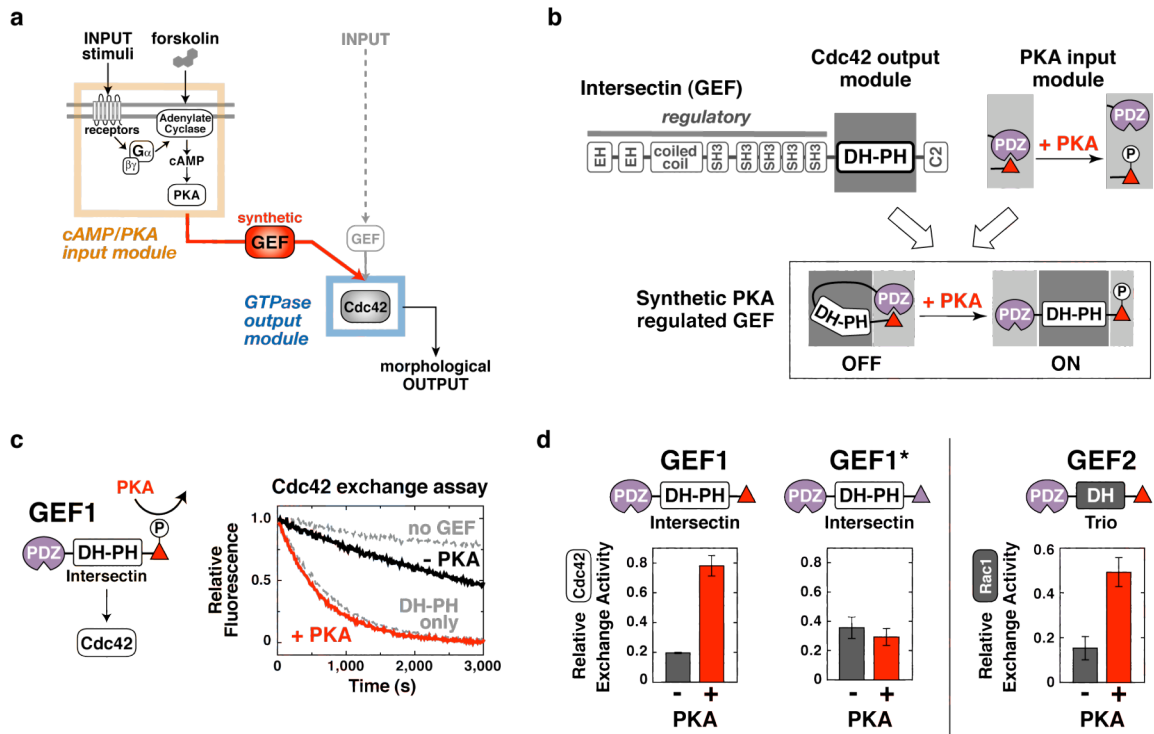**a,** We attempted to engineer GEFs that link PKA signaling to specific cytoskeletal changes. **b,** PKA-sensitive GEFs were constructed by fusing DH GEF output modules with a PKA input module composed of the syntrophin PDZ domain and a peptide that binds the PDZ domain and is a PKA substrate. **c,** *in vitro* assay of **GEF1** showing activation by PKA. Dissociation of fluorescent mant-GDP from Cdc42 was measured in the presence of no GEF or constitutively-active Intersectin DH-PH (dotted lines), **GEF1** (solid black line), or **GEF1** pre-treated with PKA (red line). **d,** Activities of synthetic GEFs (relative to Intersectin DH-PH or Trio DH). **GEF1** and **GEF2** were basally repressed, but were activated by PKA. **GEF1*** contains a mutation that abolishes PKA phosphorylation, but retains binding to the syntrophin PDZ. Error bars represent SD of three experiments. Substrate specificities of **GEF1** and **GEF2** were identical to those of their respective parental DH proteins, Intersectin and Trio (**Table 3-S4**).
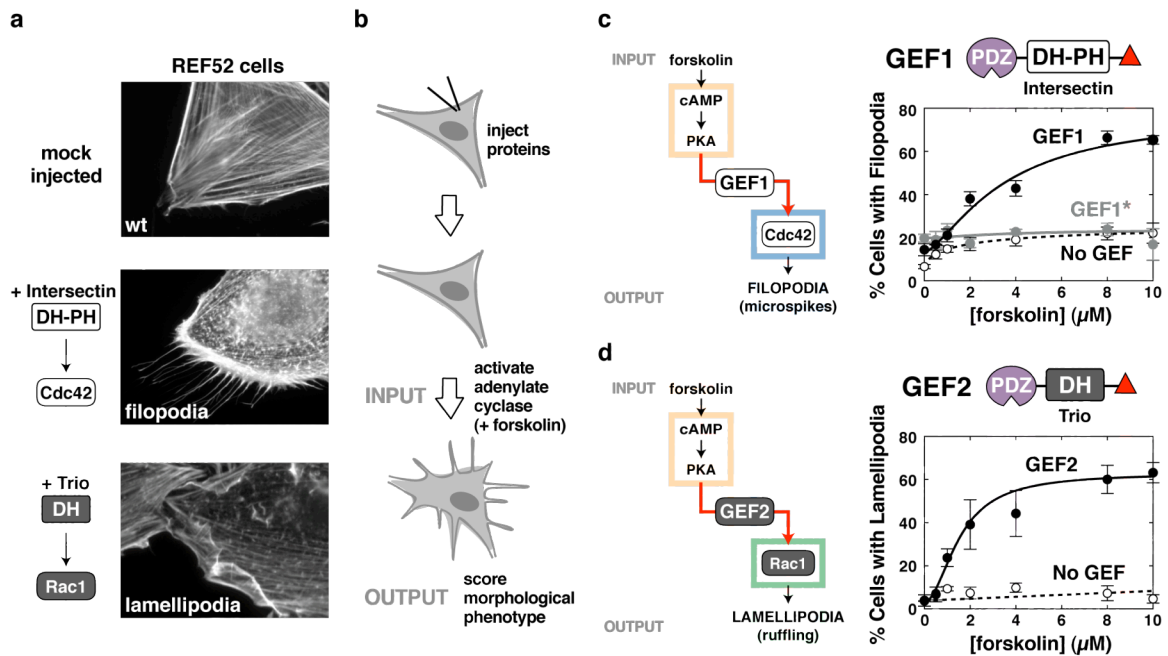
Figure 3-3 Synthetic GEFs generate novel PKA-dependent morphological changes in cells

a, Microinjection of constitutively-active Intersectin DH-PH (and Cdc42) induced filopodia in REF52 cells. Constitutively-active Trio DH (and Rac1) induced lamellipodia. **b**, After microinjecting synthetic GEFs, cells were treated with forskolin, which activates endogenous PKA. Morphological response was scored by counting cells exhibiting filopodia or lamellipodia. **c**, Filopodia were stimulated by forskolin treatment of cells injected with **GEF1** (solid black line). Forskolin treatment of cells lacking **GEF1** led only to a weak background stimulation of filopodia (dashed line). Filopodia were not significantly stimulated in cells injected with **GEF1*** (gray line), which is autoinhibited but cannot be activated by PKA. Data points represent mean ± SD of three experiments (> 50 cells scored per experiment), and were fit to a conventional Hill equation. **d**, Injection of **GEF2** allowed stimulation of lamellipodia by forskolin (solid line). Little or no response was observed in cells lacking **GEF2** (dashed line).
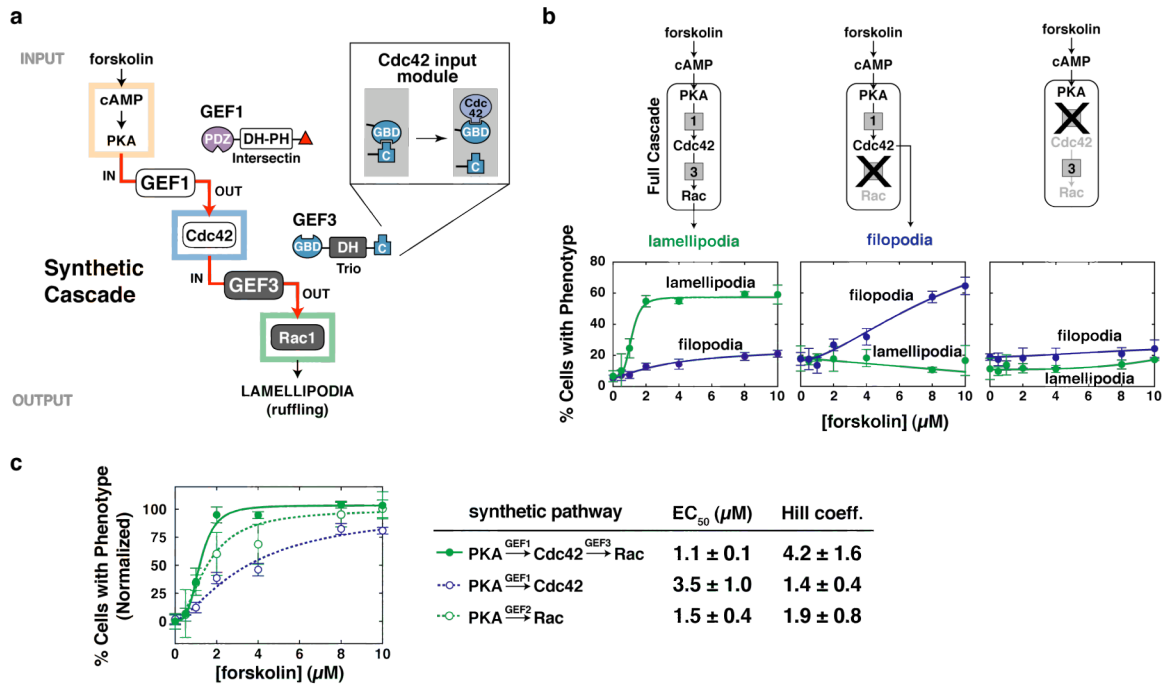
Figure 3-4 Two synthetic GEFs can be linked in series to form a higher order cascade **a**, **GEF1** and **GEF3** form a cascade in which PKA activates Cdc42, which in turn activates Rac1. **GEF3** is composed of the Trio DH domain and a Cdc42 input module extracted from N-WASP. **b**, Co-injection of **GEF1** and **GEF3** resulted in a functioning cascade (left panel). Forskolin induced lamellipodia (green curve) with very little induction of filopodia (blue curve). Forskolin treatment of cells co-injected with **GEF1** and **GEF3*** (which cannot respond to Cdc42) resulted only in filopodia (middle panel). Cells co-injected with **GEF1*** and **GEF3** showed no significant filopodial or lamellipodial response (right panel). Data points represent mean ± SD of three experiments (> 50 cells scored per experiment), and were fit to a conventional Hill equation. **c**, Comparison of **GEF1-GEF3** cascade (solid green line) to direct single-GEF circuits mediated by **GEF1** (dashed blue line) or **GEF2** (dashed green line). Data were normalized to lower and upper baselines obtained from fits to the Hill equation.

a,

syntrophin PDZ ligand consensus sequence:

$$\begin{matrix} R \\ K \end{matrix} \; E \; \begin{matrix} S \\ T \end{matrix} \; x \; \varphi \; \text{- COOH}$$

optimal PKA substrate: R R R R S I I F I

candidate ligand (▲):   R R R E S I V - COOH

mutant ligand (▲):   G V K E S L V - COOH

b,

candidate ligand   mutant ligand

▲   ▲

− +   − +   **PKA**

anti-pS/T PKA substrate
(phosphorylation)   ◀ GST-△

anti-His
(syntrophin PDZ binding)   ◀ His6-syntrophin PDZ

Figure 3-S1 Design of PKA-sensitive interaction module
a, The syntrophin PDZ domain binds ligands that are close in sequence to optimal PKA substrates. A hybrid sequence (candidate ligand, red triangle) should be a PDZ ligand as well as a PKA substrate. The mutant ligand (purple triangle) acts only as a PDZ ligand. **b,** The candidate peptide was phosphorylated by PKA, and interacted with the syntrophin PDZ, but only in its non-phosphorylated form. The mutant ligand interacted with the PDZ domain, and was not phosphorylated by PKA. Phosphorylation was visualized using anti-pS/T PKA substrate Western blotting, and PDZ binding was assessed using GST pulldown experiments.
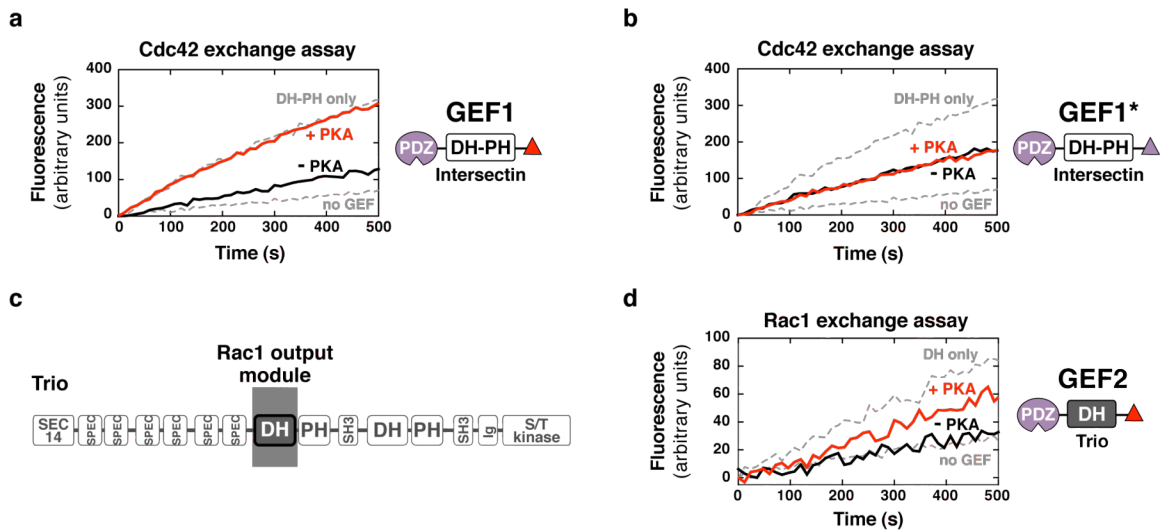
Figure 3-S2 Sample raw fluorescence data used to quantify synthetic GEF activity
a, Loading of mant-GDP into Cdc42 catalyzed by **GEF1** (solid black line) or **GEF1** pre-treated with PKA (red line). Dotted lines represent spontaneous exchange (no GEF) or exchange catalyzed by Intersectin DH-PH. **b,** Cdc42 nucleotide exchange catalyzed by **GEF1***. **c,** Schematic of Trio domains. The N-terminal DH domain was used to construct **GEF2**. **d**, Loading of mant-GDP into Rac1 catalyzed by **GEF2** (solid black line) or **GEF2** pre-treated with PKA (red line). Dotted lines represent spontaneous exchange or exchange catalyzed by Trio DH. All curves are representative of at least three independent experiments.

Figure 3-S3 GEF2* is repressed, but not activated by PKA
a, Loading of mant-GDP into Rac1 catalyzed by **GEF2*** (solid black line) or **GEF2*** pre-treated with PKA (red line). Dotted lines represent spontaneous exchange or exchange catalyzed by Trio DH. All curves are represen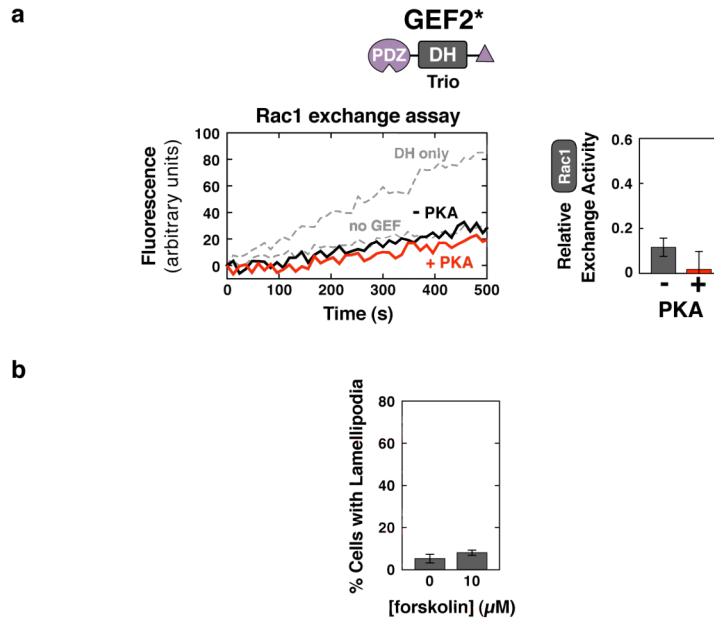tative of at least three independent experiments. Bar graph shows quantitated activities. Error bars represent SD of three experiments. **b,** Lamellipodia were not significantly stimulated by forskolin in cells injected with **GEF2***. Error bars represent SD of three experiments (at least 50 cells scored per experiment).

a wild-type
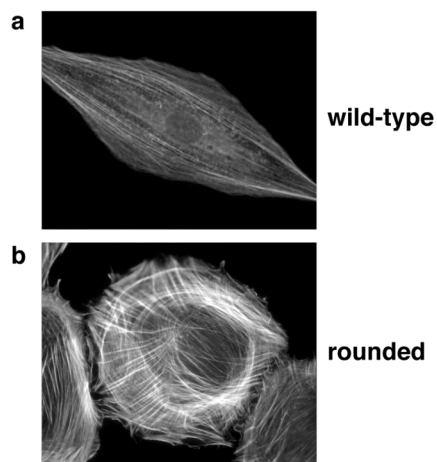
b rounded

Figure 3-S4 Rounded phenotype observed after microinjection of Intersectin DH-PH without co-injection of Cdc42

a, Wild-type morphological phenotype. REF52 cell was mock-injected with fluorescein-dextran only. **b,** Microinjection of Intersectin DH-PH induced filopodia or the rounded phenotype (but never both) in the majority of cells. The relative distribution of cells exhibiting the two phenotypes was inconsistent between experiments.
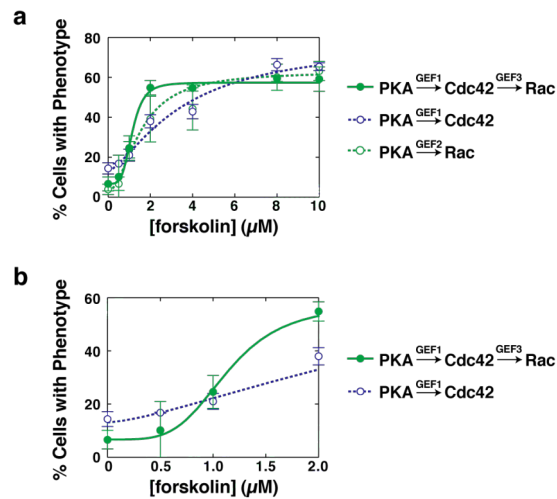
Figure 3-S5 Comparison of **GEF1-GEF3** cascade to direct single-GEF circuits
a, Morphological responses of cells injected with **GEF1** and **GEF3** (lamellipodia, solid green line), **GEF1** (filopodia, dashed blue line), and **GEF2** (lamellipodia, dashed green line). Data points and error bars represent mean and SD of three experiments (at least 50 cells scored per experiment). **b,** Comparison of **GEF1-GEF3** cascade (solid green line) to direct single-GEF circuit mediated by **GEF1** (dashed blue line). The cascade had lower basal response (no forskolin stimulation), but was more sensitive to low concentrations (1-2 μM) of forskolin.

fluorescein-dextran (injection marker)	rhodamine-phalloidin (F-actin)	merge	
a			wild-type
b			filopodia
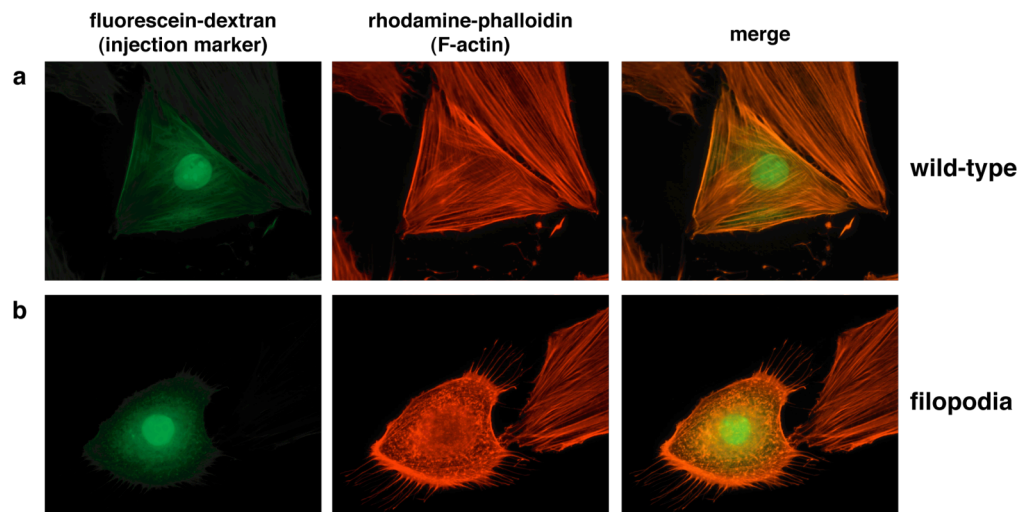
Figure 3-S6 Methodology for scoring fixed cells
Microinjected REF52 cells were identified by the presence of a fluorescein-labeled dextran marker (left panel). Filamentous actin was stained with rhodamine-conjugated phalloidin to visualize the morphology of the cytoskeleton (middle panel) and allow for scoring of injected cells. **a,** Microinjection of Cdc42 alone did not induce any morphological phenotype. **b,** Microinjection of constitutively-active Intersectin DH-PH and Cdc42 induced filopodia.

Supplementary Movie

Forskolin induced filopodia in cells injected with **GEF1**. A REF52 cell injected with **GEF1** was visualized before (10 min) and during (60 min) forskolin treatment. Stimulation of filopodia can be observed within minutes of addition. Quicktime; 5.1 MB.

http://www.nature.com/nature/journal/v447/n7144/extref/nature05851-s2.mov

References

1.   Jaffe, A.B. & Hall, A. Rho GTPases: biochemistry and biology. *Annu Rev Cell Dev Biol* **21**, 247-269 (2005).

2.   Hoffman, G.R. & Cerione, R.A. Signaling to the Rho GTPases: networking with the DH domain. *FEBS Lett* **513**, 85-91 (2002).

3.   Rossman, K.L., Der, C.J. & Sondek, J. GEF means go: turning on RHO GTPases with guanine nucleotide-exchange factors. *Nat Rev Mol Cell Biol* **6**, 167-180 (2005).

4.   Pawson, T. & Nash, P. Assembly of cell regulatory systems through protein interaction domains. *Science* **300**, 445-452 (2003).

5.   Bhattacharyya, R.P., Remenyi, A., Yeh, B.J. & Lim, W.A. Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu Rev Biochem* **75**, 655-680 (2006).

6.   Walsh, D.A. & Van Patten, S.M. Multiple pathway signal transduction by the cAMP-dependent protein kinase. *Faseb J* **8**, 1227-1236 (1994).

7.   Kim, E. & Sheng, M. PDZ domain proteins of synapses. *Nat Rev Neurosci* **5**, 771-781 (2004).

8.   Wiedemann, U. *et al.* Quantification of PDZ domain specificity, prediction of ligand affinity and rational design of super-binding peptides. *J Mol Biol* **343**, 703-718 (2004).

9.   Songyang, Z. *et al.* Use of an oriented peptide library to determine the optimal substrates of protein kinases. *Curr Biol* **4**, 973-982 (1994).

10.   Snyder, J.T. *et al.* Structural basis for the selective activation of Rho GTPases by Dbl exchange factors. *Nat Struct Biol* **9**, 468-475 (2002).

11.   Zamanian, J.L. & Kelly, R.B. Intersectin 1L guanine nucleotide exchange activity is regulated by adjacent src homology 3 domains that are also involved in endocytosis. *Mol Biol Cell* **14**, 1624-1637 (2003).

12.   Debant, A. *et al.* The multidomain protein Trio binds the LAR transmembrane tyrosine phosphatase, contains a protein kinase domain, and has separate rac-specific and rho-specific guanine nucleotide exchange factor domains. *Proc Natl Acad Sci U S A* **93**, 5466-5471 (1996).

13.   Dueber, J.E., Mirsky, E.A. & Lim, W.A. Engineering synthetic signaling proteins with ultrasensitive input/output control. *Nat Biotechnol* **25**, 660-662 (2007).

14.   Nimnual, A.S., Yatsula, B.A. & Bar-Sagi, D. Coupling of Ras and Rac guanosine triphosphatases through the Ras exchanger Sos. *Science* **279**, 560-563 (1998).

15.   Seamon, K.B. & Daly, J.W. Forskolin: a unique diterpene activator of cyclic AMP-generating systems. *J Cyclic Nucleotide Res* **7**, 201-224 (1981).

16.   Ferrell, J.E., Jr. Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. *Trends Biochem Sci* **21**, 460-466 (1996).

17.   Yamauchi, J., Miyamoto, Y., Tanoue, A., Shooter, E.M. & Chan, J.R. Ras activation of a Rac1 exchange factor, Tiam1, mediates neurotrophin-3-induced Schwann cell migration. *Proc Natl Acad Sci U S A* **102**, 14889-14894 (2005).

18.   Miki, H., Sasaki, T., Takai, Y. & Takenawa, T. Induction of filopodium formation by a WASP-related actin-depolymerizing protein N-WASP. *Nature* **391**, 93-96 (1998).

19.   Rohatgi, R. *et al.* The interaction between N-WASP and the Arp2/3 complex links Cdc42-dependent signals to actin assembly. *Cell* **97**, 221-231 (1999).

20.   Kim, A.S., Kakalis, L.T., Abdul-Manan, N., Liu, G.A. & Rosen, M.K. Autoinhibition and activation mechanisms of the Wiskott-Aldrich syndrome protein. *Nature* **404**, 151-158 (2000).

21.   Prehoda, K.E., Scott, J.A., Mullins, R.D. & Lim, W.A. Integration of multiple signals through cooperative regulation of the N-WASP-Arp2/3 complex. *Science* **290**, 801-806 (2000).

22.   Hooshangi, S., Thiberge, S. & Weiss, R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc Natl Acad Sci U S A* **102**, 3581-3586 (2005).

23.   Colicelli, J. Human RAS superfamily proteins and related GTPases. *Sci STKE* **2004**, RE13 (2004).

24.   Dueber, J.E., Yeh, B.J., Chak, K. & Lim, W.A. Reprogramming control of an allosteric signaling switch through modular recombination. *Science* **301**, 1904-1908 (2003).

25.   Howard, P.L., Chia, M.C., Del Rizzo, S., Liu, F.F. & Pawson, T. Redirecting tyrosine kinase signaling to an apoptotic caspase pathway through chimeric adaptor proteins. *Proc Natl Acad Sci U S A* **100**, 11267-11272 (2003).

26. Park, S.H., Zarrinpar, A. & Lim, W.A. Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms. *Science* **299**, 1061-1064 (2003).

27. Inoue, T., Heo, W.D., Grimley, J.S., Wandless, T.J. & Meyer, T. An inducible translocation strategy to rapidly activate and inhibit small GTPase signaling pathways. *Nat Methods* **2**, 415-418 (2005).

28. Sprinzak, D. & Elowitz, M.B. Reconstruction of genetic circuits. *Nature* **438**, 443-448 (2005).

29. Voigt, C.A. Genetic parts to program bacteria. *Curr Opin Biotechnol* **17**, 548-557 (2006).

30. Patel, J.C. & Galan, J.E. Manipulation of the host actin cytoskeleton by Salmonella--all in the name of entry. *Curr Opin Microbiol* **8**, 10-15 (2005).

31. Harris, B.Z., Hillier, B.J. & Lim, W.A. Energetic determinants of internal motif recognition by PDZ domains. *Biochemistry* **40**, 5921-5930 (2001).

32. Wu, J.Q. & Pollard, T.D. Counting cytokinesis proteins globally and locally in fission yeast. *Science* **310**, 310-314 (2005).

33. Schoenwaelder, S.M. & Burridge, K. Evidence for a calpeptin-sensitive protein-tyrosine phosphatase upstream of the small GTPase Rho. A novel role for the calpain inhibitor calpeptin in the inhibition of protein-tyrosine phosphatases. *J Biol Chem* **274**, 14359-14367 (1999).

34. Worthylake, D.K., Rossman, K.L. & Sondek, J. Crystal structure of Rac1 in complex with the guanine nucleotide exchange region of Tiam1. *Nature* **408**, 682-688 (2000).

35.   Welch, H.C. *et al.* P-Rex1, a PtdIns(3,4,5)P3- and Gbetagamma-regulated

guanine-nucleotide exchange factor for Rac. *Cell* **108**, 809-821 (2002).

Chapter 4

Synthetic Biology: Lessons from the History of Synthetic

Organic Chemistry

Introduction

In 1828, the German chemist Friedrich Wöhler could hardly contain his excitement as he wrote to his former mentor, Jöns Jakob Berzelius, of a new finding[1,2]: "I must tell you that I can prepare urea without requiring a kidney of an animal, either man or dog." At the beginning of the 19th Century, the synthesis of this small organic molecule was earth-shattering news. At that time, chemists believed there was a clear distinction between molecules from living beings (referred to as "organic") and those from non-living origin ("inorganic"). It was known that organic substances could be easily converted to inorganic compounds through heating or other treatments; however, chemists could not perform the reverse transformation. Surely, a "vital force" present only in living organisms was required to convert the inorganic into organic. Wöhler's discovery that ammonium cyanate could be converted to urea in the laboratory was a key nail in the coffin of vitalism, and in the next few decades, chemists began to synthesize hundreds of other organic molecules. In a particularly interesting example in 1854, the French chemist Marcellin Berthelot synthesized the fat molecule tristearin from glycerol and stearic acid, a common naturally occurring fatty acid. Taking this a step further, he realized that he could replace stearic acid with similar acids not found in natural fats, thus generating non-natural fats that had properties similar to natural fats[3]. These and other early syntheses demonstrated that chemists could indeed make "living" molecules as well as new compounds not known to previously exist, thus giving birth to synthetic organic chemistry. It was unclear where this field would lead, and many feared these advances could lead to goals such as the creation of living beings.

Advances in our ability to build and modify "organic" molecules on increasingly larger

scales have continued to push the frontier of our understanding of the physical principles

underlying living systems. For example, chemical synthesis of DNA oligonucleotides

(first performed by Gobind Khorana) led directly to the elucidation of the genetic code[4].

Bruce Merrifield's complete synthesis of RNase A demonstrated that chemical structure

(primary sequence) is sufficient to confer tertiary structure and the mysterious activity of

enzymes[5]. More recently, complete chemical synthesis of poliovirus cDNA was a vivid

demonstration that genetic instructions are sufficient to specify an active biological

system[6].

Over the last several years, this line of research has culminated in the emergence of a

field known as "Synthetic Biology". Whether synthetic biology represents a truly new

field is open to debate, but the boldness of the stated goals — to learn how to precisely

and reliably engineer and build self-organizing systems that both recapitulate biological

function and show new functions — is unquestionably novel. These goals hold promise

for harnessing the efficiency and precision of living systems for diverse purposes:

microbial factories that manufacture drugs, materials, or biofuels[7]; cells that seek and

destroy tumors[8]; cells that can carry out rapid tissue repair and regeneration; cells that can

direct the assembly of nanomaterials; even living systems that can compute. Synthetic

biology, however, is more than a broad set of visionary applications. Much effort is going

into developing a common toolkit of well-defined biological parts and devices as well as

strategies to link them together into predictable systems[9-12]. These foundational efforts

are aimed at one day making engineering biology as reliable and predictable as engineering an electronic device.

At this threshold, where our view of biology as something to be observed is transitioning into something that can be engineered, there are many important questions. Why even attempt synthetic biology when our understanding of biological systems is still incomplete? And should we choose to proceed, how should we go about it? Many reviews have compared synthetic biology to electrical engineering[9,11-13]. While this comparison is useful, in some respects a comparison with the development of synthetic organic chemistry may be more appropriate[10]. In this commentary, we consider the historical role of synthetic approaches in the development of modern organic chemistry in order to extract some lessons that might help guide the development of synthetic biology.

The Importance of Synthesis: a Necessary Complement to Analysis

Before the time of Wöhler and Berthelot, the understanding of even simple molecules was as naïve as our current understanding of complex biological systems. How the composition of organic compounds determined their properties and reactivity was unknown, and the concept of molecules having defined structures was still undeveloped. Ultimately, analytical and synthetic approaches synergized to produce an explosive growth in our knowledge of chemical principles, and fundamental theories of chemical structure would develop concurrently with the explosion of synthesis (**Figure 4-1**).

A critical early advance at the beginning of the 19th Century was precise measurement and analysis of compounds as they reacted[3]. For example, by collecting and weighing the carbon dioxide and water that formed upon combustion of organic molecules, it became possible to determine the atomic compositions of these molecules, and therefore, their empirical formulas. This careful analysis led to the discovery of isomers — the shocking finding that compounds with very different physical and chemical properties could have identical empirical formulas. Clearly, a more sophisticated understanding of chemical structure would be required.

It was not until after the mid-19th Century — after synthesis of small compounds was already becoming routine — that chemists began to develop models to explain bonding between atoms[3]. In 1852, Edward Frankland proposed that each atom had an ability to combine with a fixed number of other atoms. Kekulé used this "theory of valence" to propose structures for many simple organic molecules in 1858, and in the 1860's, Alexander Bulterov pointed out that these structural formulas could explain the majority of isomers. This notion that atoms were held in fixed arrangements was a critical advance. In 1865, Kekulé proposed the structural formula for benzene. Although it had already been synthesized by Berthelot in the 1850's, benzene's stability could not be adequately explained until Kekulé's ring structure, cementing the usefulness of this paradigm. Later, these structures were extended to three dimensions based on the notion that carbon bonds are arranged in a tetrahedral geometry, proposed by Jacobus Van't Hoff in 1874.

Eventually, these structural models would inform more rational chemical syntheses; however, synthetic methods themselves were also critical to the understanding of molecular structure. At the simplest level, synthesizing a molecule was often the ultimate proof of the proposed molecular structure. Perhaps more significantly, the ability to synthesize variants of known molecules with diverse functional groups allowed a truly rigorous exploration of the principles underlying chemical properties and reactivity, thus beginning the field of physical organic chemistry.

Biology has historically been a field based almost entirely on observation and analysis, and it is currently undergoing exponential growth in the accuracy and throughput of measurement. Modern experimental techniques — genome sequencing, DNA microarrays, molecular structure determination, and high-throughput microscopy coupled with *in vivo* biosensors — represent major analytical advances, giving us extensive parts lists and descriptions of biological systems and their behaviors. These developments are akin to the advances in analytical chemistry of the early 19th Century. However, the history of organic chemistry suggests that synthesis will be a necessary complement to analysis in order for biologists to truly understand the mechanism of complex living systems.

In many ways, synthetic biology can be viewed as *in vivo* reconstitution — an intellectual descendant of simple biochemistry. Reconstitution methods (which essentially apply the synthetic philosophy to non-covalent systems) allow us to determine not only what is

necessary, but also what is sufficient to build a system that carries out a particular function. The ability to build and systematically modify biological systems will allow one to explore their design principles in much deeper ways. Thus, synthetic biological systems will allow experimentation not possible with extant living systems, which are encumbered by eccentric evolutionary histories and constraints. How does phenotype precisely vary as individual components, their network linkages, and biochemical parameters are altered?

What can we expect to learn? Clearly, there will be more than a few simple rules explaining how living systems work. As in physical organic chemistry, however, it is likely that patterns will emerge, and an understanding of the logic underlying biological systems will develop. It may even be possible to construct something analogous to a "periodic table" of biology that facilitates the systematic understanding of network "elements" and their properties[14].

Diverse and Unexpected Driving Applications

If one important goal of synthetic biology is to create useful systems, then what applications should we be targeting? Again, it is useful to consider the early applications of synthetic organic chemistry. In today's world, we (particularly readers of this journal) tend to link synthetic chemistry with the production of drugs. Indeed, it was abundantly clear to early chemists that synthetic products could improve human health, but their

initial efforts actually led to an industrial explosion in an unexpected direction. August von Hofmann and his student William Perkin discussed whether it would be possible to synthesize the anti-malarial agent quinine from aniline, which could be obtained in abundance from coal tar[3]. Although a synthetic supply of quinine would be extremely valuable, its structure was unknown; therefore, such a synthesis was doomed to failure. However, in attempting to synthesize quinine from aniline in 1856, Perkin unexpectedly produced a brilliant purple compound — a dye. He opened a factory to synthesize this molecule, which he called "Aniline Purple", and founded the synthetic dye industry. Along with other examples, such as the synthesis of indigo by Adolf von Baeyer in 1867, these advances led to the explosive growth of the German and Swiss dye industry, while simultaneously dismantling the import of indigo and other natural dyes from distant tropical locales. In fact, synthetic indigo remains a major commercial product, and is used in today's blue jeans (Levi's was founded in 1873).

Although dyes were the earliest economically important synthetic compounds, the development of the European dye industry would ultimately lead to successes in chemotherapy. Indeed, nearly all of the modern big pharmaceutical companies are in part descended from German or Swiss dye manufacturers. For example, the first effective antibacterials were the sulfa drugs[3,15]. The first of these molecules, sulfanilamide, was synthesized by IG Farbenindustrie in 1908 because of its potential as a dye. In 1932, Gerhard Domagk discovered that sulfanilamide and related compounds had bactericidal activity, and was aided in his studies by chemists that could make a variety of related compounds.

The lesson here is very clear: synthetic biologists (and their funding agencies) must be open to a variety of potential industrial and therapeutic applications including those we have not yet foreseen. Most of the successes in synthetic biology to date have been in so-called "toy systems"; however, we should not underestimate the importance of these achievements. Advances in any one area will be applicable to many others. Developing protecting-group strategies or specific classes of reactions to make carbon-carbon bonds was applicable to the synthesis of a wide array of organic products. Similarly, learning how to link a set of molecules into a specific type of regulatory circuit module, such as an ultrasensitive positive-feedback loop, could be useful for a diverse range of synthetic biology applications[9,11,16].

Focusing on highly specific goals and applications in the context of biological systems is an innovation of synthetic biology. This philosophy may seem strange to many academic biologists, but it is actually quite appropriate. A defining feature of evolution is the constant selection of organisms that achieve the best fitness in a particular niche. Understanding how to build biological systems to achieve well-defined performance specifications will force us to understand biology at a far more quantitative level. Organizing inter-disciplinary teams of scientists around clear, practical goals will also have important effects on the sociology of biology. Furthermore, striving to make systems that perform more efficiently will introduce competitive pressures, and undoubtedly increase creativity and innovation.

Conclusions and Perspectives

It is important to remember that early synthetic chemists did not always know what to expect in their reactions, only that something interesting *could* happen[15]. Most importantly, they were prepared to follow up these experiments to understand what *did* happen. Similarly, in these early days of synthetic biology, it will be very difficult to predict the behavior of novel biological systems. Therefore, directed evolution and combinatorial methods will be useful[17]; analyzing libraries of synthetic circuits and systems will maximize the probability of obtaining the targeted biological behavior. Furthermore, systematically varying many parameters will produce structure-activity relationships at the biological network level that will improve future designs.

Developing educational initiatives will continue to be an important emphasis in synthetic biology. Perkin was a teenager when he initially synthesized Aniline Purple — during Easter vacation in a small laboratory in his home[3]. Similarly, it is young scientists who are likely to look at biology from a new perspective, and who will shape the yet unforeseen "killer applications" of synthetic biology. The spirit of Perkin lives on today in the teams of undergraduates and high school students participating in the International Genetically Engineered Machine (iGEM) competition each summer. In 2007, 56 teams from North America, Europe, and Asia have registered[18].

Like all technologies, synthetic organic chemistry also introduced its own set of problems. In addition to the plethora of beneficial drugs and polymers that have

significantly increased our standard of living, harmful or "dual-use" compounds, including explosives and chemical weapons, have also been created. There is no question that synthetic biological systems will also bring a mixed array of potential applications. Synthetic biologists are not ignoring this possibility; discussions of biosafety and security have been a major component of major synthetic biology conferences[19].

In the coming decades, we are likely to see a revolution in biology akin to the revolution in chemistry that occurred in the latter half of the 19th Century. The development of increasingly sophisticated methods to alter and build biological systems will provide essential synthetic tools that will synergize with analytical methods, which together will ultimately lead to a far deeper understanding of the physical principles underlying the behavior and design of cellular systems. The applications of synthetic biology will be highly varied, and progress and innovation is likely to come from unexpected areas. We might also expect that understanding the engineering principles of biological systems will have a transformative effect on other fields of science. For example, man-made materials, even at the nanoscale, are currently templated or built using directed assembly, exactly the opposite of how biological molecules create structure and function. Biological molecules are self-assembling systems that can adapt to change, display robust homeostasis, and can self-repair. There may come a day when man-made materials also have these properties. Universities, industry, governmental agencies, and scientists will have to work together in an open-minded and responsible way to foster productive growth of this exciting field.

Acknowledgements

MILESTONES OF ORGANIC CHEMISTRY

SYNTHESIS | THEORY

early 1800's
empirical formulas

1828
urea (Wöhler)

1845
acetic acid (Kolbe)

1850's
methanol, ethanol, benzene (Berthelot)

1852
valence (Frankland)

1854
tristearin (Berthelot)

1856
Aniline Purple (Perkin)

1858
structural formulas (Kekulé)

1865
ring structure for benzene (Kekulé)

1867
indigo (Baeyer)

1874
tetrahedral geometry for carbon (Van't Hoff)

1908
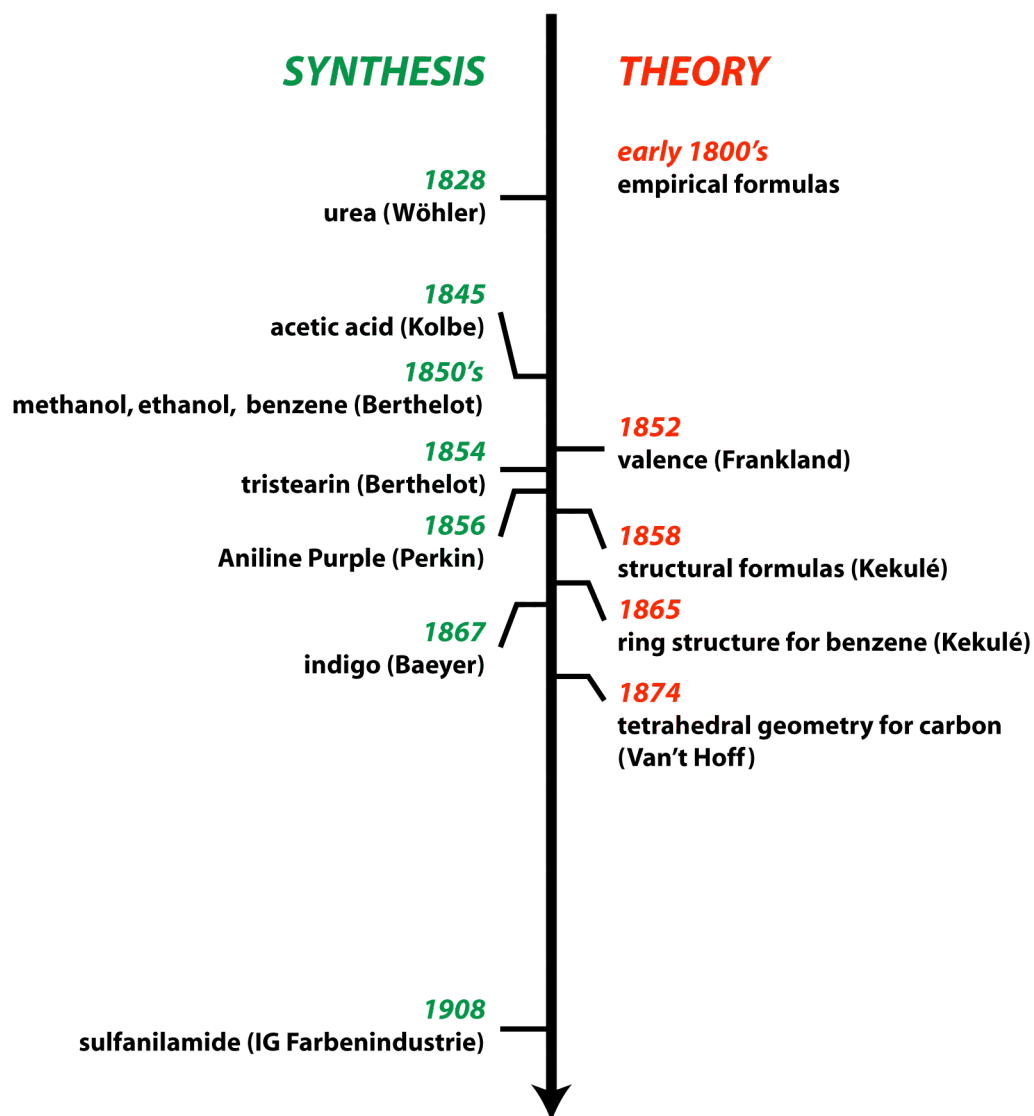**sulfanilamide (IG Farbenindustrie)**

Figure 4-1 Chemical synthesis and theories of structure emerged concurrently. Significant milestones in chemical synthesis (left of timeline, dates shown in green) and theories of chemical structure (right of timeline, dates shown in red).

References

1.   Jaffe, B. *Crucibles: the story of chemistry from ancient alchemy to nuclear fission*. (Dover Publications, New York, 1976).

2.   Wohler, F. *Poggendorff's Ann. Phys.* **12**, 253-256 (1828).

3.   Asimov, I. *A short history of chemistry*. (Anchor Books, Garden City, N. Y., 1965).

4.   Khorana, H.G. Polynucleotide synthesis and the genetic code. *Fed Proc* **24**, 1473-87 (1965).

5.   Gutte, B. & Merrifield, R.B. The synthesis of ribonuclease A. *J Biol Chem* **246**, 1922-41 (1971).

6.   Cello, J., Paul, A.V. & Wimmer, E. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science* **297**, 1016-8 (2002).

7.   Khosla, C. & Keasling, J.D. Metabolic engineering for drug discovery and development. *Nat Rev Drug Discov* **2**, 1019-25 (2003).

8.   Anderson, J.C., Clarke, E.J., Arkin, A.P. & Voigt, C.A. Environmentally controlled invasion of cancer cells by engineered bacteria. *J Mol Biol* **355**, 619-27 (2006).

9.   Andrianantoandro, E., Basu, S., Karig, D.K. & Weiss, R. Synthetic biology: new engineering rules for an emerging discipline. *Mol Syst Biol* **2**, 2006 0028 (2006).

10.   Benner, S.A. & Sismour, A.M. Synthetic biology. *Nat Rev Genet* **6**, 533-43 (2005).

11.   Endy, D. Foundations for engineering biology. *Nature* **438**, 449-53 (2005).

12.   Voigt, C.A. Genetic parts to program bacteria. *Curr Opin Biotechnol* **17**, 548-57 (2006).

13.   Arkin, A.P. & Fletcher, D.A. Fast, cheap and somewhat in control. *Genome Biol* **7**, 114 (2006).

14.   Alon, U. Network motifs: theory and experimental approaches. *Nat Rev Genet* **8**, 450-61 (2007).

15.   Hoffmann, R. *The same and not the same*, xvi, 294 (Columbia University Press, New York, 1995).

16.   Sprinzak, D. & Elowitz, M.B. Reconstruction of genetic circuits. *Nature* **438**, 443-8 (2005).

17.   Haseltine, E.L. & Arnold, F.H. Synthetic gene circuits: design with directed evolution. *Annu Rev Biophys Biomol Struct* **36**, 1-19 (2007).

18.   http://www.igem2007.com.

19.   http://syntheticbiology.org/SB2.0/Biosecurity_resolutions.html.

Chapter 5

Conclusions and Future Directions

In the past decade, the observation that many eukaryotic signaling proteins have distinct functional domains has led many people to hypothesize that this modular structure may have facilitated the evolution of diverse input-output relationships. We tested this hypothesis by attempting to engineer novel signaling proteins containing new combinations of catalytic (output) and interaction (input) modules. We first performed this type of modular recombination using the output domain from N-WASP and a series of input domains, producing a family of N-WASP variants (Chapter 2). The usefulness of this approach was highlighted by two key observations: a majority of the constructs we engineered were regulated in some fashion by input ligand, and the quantitative relationships between input and catalytic activity were highly diverse. We then extended this work by demonstrating that a similar methodology could be used to engineer synthetic Rho GEFs that responded to phosphorylation or activated GTPases themselves (Chapter 3).

How generally applicable is this method of reprogramming signaling proteins? Our success with N-WASP and Rho GEFs — coupled with accumulating evidence supporting the prevalence of autoinhibition as a regulatory mechanism — suggests that other catalytic activities may be amenable to this type of regulation (which we refer to as *modular allostery*). However, it is important to note that we have also tried unsuccessfully to regulate a number of other catalytic domains including metabolic enzymes (dihydrofolate reductase and staphylococcal nuclease) and activities involved in signal transduction — PKA and Sos (a Ras GEF). Whether these catalytic domains are fundamentally refractory to regulation by modular allostery, or require parameters that

we have not yet sampled, is unclear. In collaboration with Shalini Singh and Todd Miller (SUNY Stony Brook), we have also designed variants of Hck, a prototype for this type of modular autoinhibition. To date, the results with these experiments have been inconclusive; there is some evidence that some of these constructs are repressed in cells, but attempts to activate them have been largely unsuccessful. Ultimately, it would be useful to have a set of guidelines that could estimate the likelihood that a catalytic activity could be reprogrammed by modular allostery. However, deriving these guidelines could prove to be a monumental undertaking, and in the short term, researchers may simply have to address this question empirically.

Our work with Dbl-family GEFs also demonstrates that synthetic signaling proteins can rewire cellular pathways, providing a foundation for engineering cells with useful industrial or therapeutic properties. In particular, we focused on Rho-family GTPases because they are important regulators of the actin cytoskeleton, and therefore, cell morphology. One long-term goal of the lab is to engineer synthetic pathways that can recapitulate polarization and directed movement. As might be expected from simple linear circuits, we have not yet observed any spatial asymmetries with the current panel of synthetic GEFs. To move towards circuits that can support polarization, we believe that it will be necessary to engineer positive-feedback linkages that reinforce protrusion (at the front of the cell) and retraction (at the back), as well as inhibitory linkages between the front and the back. Rho GTPases are indeed central to these pathways, but it will be important to also consider other important intermediates such as PIP_2/PIP_3 and F-actin structures. One way to engineer these linkages is through engineering modular

allostery, as described in this work; however, a simpler method will be to exploit specificity resulting from localization by fusing binding domains (input) to catalytic domains (output) in new combinations.

One initial attempt at engineering positive-feedback involved combining the Cdc42 input module (GBD-C) with the Cdc42 GEF activity of the Intersectin DH-PH. In principle, such a construct would activate Cdc42 in the presence of activated Cdc42 itself, thus producing a positive-feedback loop. When we microinjected this construct into REF52 cells, the results were difficult to interpret. Although its activity was repressed *in vitro*, the construct appeared to be constitutively-active in cells: a large proportion of cells had filopodia. This is not necessarily surprising; even an initially low level of Cdc42 activation would activate the positive-feedback loop. We also observed that cells microinjected with this putative positive-feedback construct appeared to have a different morphological phenotype, which we refer to as "enhanced", compared to cells microinjected with the Intersectin DH-PH alone or **GEF1**. On average, there were more filopodia per cell, as well as more lamellipodia-like protrusions in cells displaying the "enhanced" phenotype. Clearly, this class of constructs (including variants with two and three copies of the GBD-C module) bears further investigation.

In these initial studies, the cycle of design and testing was quite laborious, requiring protein expression and purification, *in vitro* assays, and cellular experimentation involving microinjection. In the future, this process must unquestionably be streamlined. One solution may be to avoid purifying proteins (and *in vitro* experiments), and proceed

directly to cellular assays involving transfection of DNA constructs. For this approach to be effective, however, we will need to develop protocols for inducing protein expression at defined levels with minimal variation between cells. Perhaps more importantly, it will also be critical to develop more meaningful (and high-throughput) quantitative assays. The primary experimental readout in this work was simply the absence or presence of filopodia or lamellipodia, the final morphological output. Measuring intermediates in these pathways would provide enormous insight. For example, we still have very little understanding of the quantitative relationship between active Cdc42 or Rac1 and the resulting morphological response. Querying multiple steps in the pathways of interest will allow better evaluation of circuit designs, and therefore, facilitate subsequent designs.

Appendices

Appendix A. Published Plasmids

Yeh, B.J., Rutigliano, R.J., Deb, A., Bar-Sagi, D., and Lim, W.A. Rewiring cellular morphology pathways with synthetic guanine nucleotide exchange factors. *Nature* 447, 596-600 (2007).

GEFs in Main Paper
pBY601	Intersectin DHPH
pBY619	**GEF1** [PDZ-ITSN DHPH-RRRESIV]
pBY602	**GEF1*** [PDZ-ITSN DHPH-VKESLV]
pAD8	Trio N-term. DH
pAD10	**GEF2** [PDZ-(Trio N-DH)-RRRESIV]
pBY787	**GEF2*** [PDZ-(Trio N-DH)-VKESLV]
pAD26	**GEF3** [GBD-(Trio N-DH)-C]
pBY794	**GEF3*** [GBD*-(Trio N-DH)-C]

Other GEFs (Table S3)
pBY605	Intersectin DH
pBY623	PDZ-ITSN DH-RRRESIV
pBY791	PDZ-Tim DH-RRRESIV
pAD1	mouse Tiam1 DHPH
pAD2	human Trio N-term. DHPH
pAD5	PDZ-mouse Tiam1 DHPH-RRRESIV
pAD6	PDZ-(human Trio N-DHPH)-RRRESIV
pAD7	mouse Tiam1 DH
pAD12	human Prex1 DH
pAD22	PDZ-Prex1 DH-RRRESIV
pAD24	human Tim DHPH
pAD25	human Tim DH

Figure S1
pBH7	syntrophin PDZ (His-tagged)
pBY251	GST-RRRESIV (red triangle)
GST-GVKESLV	GST-GVKESLV (purple triangle)

GTPases
Cdc42(1-179)/pBH4	Cdc42 for in vitro exchange assay
Rac1(1-177)/pBH4	Rac1 for in vitro exchange assay
pAD15	RhoA(1-190, C190S) for in vitro exchange assay
GST-Cdc42	Full-length Cdc42 for microinjection
pBY650	GST-Rac1 (full-length) for microinjection

PKA
pBY401	Mouse PKA (Calpha-subunit)

Appendix B. Protocol for *in vitro* Nucleotide Exchange Assays

(Note: every experiment has an associated spreadsheet, which contains raw and analyzed data, as well as the setup for that experiment.)

Association Assay

Reagents

- mant-GDP: Molecular Probes M-12414; 5 mM stock solution; store aliquots at -20°C

- GEF assay buffer: 20 mM Tris, 50 mM NaCl, 10 mM $MgCl_2$, 1% glycerol, 1 mM DTT, pH 7.5; add DTT fresh (e.g., from 0.5 M stock) daily

- GEFs: purify on Ni-NTA, remove His-tag; store concentrated stock solutions in 20 mM Tris, 50 mM NaCl, 2 mM DTT, pH 7.5 (or GEF Assay Buffer) at -80°C; use within 1 week after thawing (stored at 4°C)

- GTPases: purify on Ni-NTA and Source Q; remove residual nucleotide by dialyzing into 20 mM Tris, 50 mM NaCl, 5 mM EDTA, 2 mm DTT, pH 7.5; load with GDP by incubating with 10-fold molar excess GDP; quench nucleotide exchange by adding 50-fold excess of $MgCl_2$; dialyze into GEF Assay Buffer; store concentrated stock solutions at -80°C; use within 1 week after thawing (stored at 4°C)

Typical concentrations:

- For Intersectin: 1 µM Cdc42(GDP), 25 nM GEF, 400 nM mant-GDP

- For Trio N-term. DH: 1 µM Rac1(GDP), 250 nM GEF, 400 nM mant-GDP

Protocol:

- Prepare appropriate dilutions of mant-GDP and GEFs in GEF assay buffer

- Prepare premix containing mant-GDP and GEF (140 µL + 15% extra) in 96-well plate (translucent plastic with tight lid)

- Prepare solution containing GTPase(GDP) (10 µL; i.e., usually 15 µM) in experimental assay plate (Corning 3693)

- Equilibrate for 10 minutes: mant-GDP + GEF in plate reader set at 25°C, GTPase(GDP) at room temperature

- Add mant-GDP + GEF to GTPase with multichannel pipetter, 1 column at a time

- Remove bubbles, shake plate for a few seconds

- Begin data collection ("mant-GDP Protocol"; 25°C, excitation: 360 nm, emission: 440 nm, auto cutoff, readings: 6, PMT: high, interval: 10-15 s)

Notes:

- GEFs: I always make dilutions of the GEFs at a standard concentration (10-30X final concentration), where I pipet the same volume of stock GEF (and different volumes of buffer). I use that dilution only for the immediate experiment.

- mant-GDP: I dilute the stock solution first to 5 µM (i.e., 1 µL stock + 999 µL buffer), then make a second dilution to a concentration so that the remaining volume is mant-GDP

- For the mant-GDP + GEF premix, I make at least 15% extra in the 96-well plate, so that it is easy to pipet 140 µL with the multichannel pipetter.

Analysis:

- I export the data from SoftMax Pro to a .txt file, which I import into Excel.

- Calculate the initial slope of each reaction (i.e., while the curves are still linear)

- Calculate "relative activity" by normalizing to the spontaneous and appropriate maximal rate (e.g., DH or DH-PH alone): relative activity = ($slope_{experimental}$ − $slope_{no\ GEF}$)/($slope_{DH/DH-PH\ alone}$ − $slope_{no\ GEF}$)

- I always normalize to curves within the same column. (At one point, I noticed that there were significant differences in slope between columns.)

Activation by peptide (e.g., VKESLV)

- Pre-equilibrate peptide with mant-GDP + GEF premix

- Control for volume of water (or peptide's solvent)

Activation by PKA

- Mix GEFs with PKA directly in the 96-well plate in a volume of 11.5 μL (10 μL + 15% extra); for the paper, I used a 1:10 PKA:kinase ratio

- Incubate at 30°C by floating the plate in a water bath; for the paper, I phosphorylated for 30 min

- For the mant-GDP + GEF premix, add mant-GDP directly to each well, and incubate in the plate reader for 10 minutes, as usual

- Normalize to samples that have been treated equivalently; i.e., perform reactions with no GEF and DH/DH-PH with and without PKA

Dissociation Assay

Reagents

- GTPases: load with mant-GDP (as with GDP); use within 1 week of loading

- GDP: 15 mM stock solutions (pH to ~ 7.5); store at -20°C; use day of thawing

Typical concentrations:

- 1 µM Cdc42(mant-GDP), 200 nM GEF, 200 µM GDP

Protocol:

- Prepare appropriate dilutions of GEFs in GEF assay buffer

- Prepare premix containing GEF and free GDP (140 µL + 15% extra) in 96-well plate (translucent plastic with tight lid)

- Prepare solution containing GTPase(mant-GDP) (10 µL) in experimental assay plate

- Equilibrate for 10 minutes: GEF + GDP in plate reader set at 25°C, GTPase(mant-GDP) at room temperature

- Add GEF + GDP to GTPase with multichannel pipetter, 1 column at a time

- Remove bubbles, shake plate for a few seconds

- Begin data collection ("mant-GDP Protocol"; 25°C, excitation: 360 nm, emission: 440 nm, auto cutoff, readings: 6, PMT: high, interval: 10-15 s)
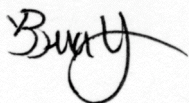
Analysis: the convention in the field is to fit to a single exponential $(y = Ae^{-kx} + \text{const})$, using k_{obs} as the metric of activity. I believe this is incorrect. There is no theoretical basis for using this equation, and empirically, the fits are poor. If absolutely necessary, I would recommend normalizing the data to curves that actually go to completion, and then forcing the fits to go from 1 to 0. (See "060813-3 ITSN" for an example.)

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses and dissertations. Copies of all UCSF theses and dissertations will be routed to the library via the Graduate Division. The library will make all theses and dissertations accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis or dissertation to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

_____   ___14 June, 2007___
Author Signature   Date