Strategies for protein synthetic biology

Raik Grünberg[1,*] and Luis Serrano[1,2]

[1]EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG), UPF, 08003 Barcelona and
[2]ICREA (Institució Catalana de Recerca i Estudis Avan cats), 08010 Barcelona, Spain

ABSTRACT

Proteins are the most versatile among the various biological building blocks and a mature field of protein engineering has lead to many industrial and biomedical applications. But the strength of proteins—their versatility, dynamics and interactions—also complicates and hinders systems engineering. Therefore, the design of more sophisticated, multi-component protein systems appears to lag behind, in particular, when compared to the engineering of gene regulatory networks. Yet, synthetic biologists have started to tinker with the information flow through natural signaling networks or integrated protein switches. A successful strategy common to most of these experiments is their focus on modular interactions between protein domains or domains and peptide motifs. Such modular interaction swapping has rewired signaling in yeast, put mammalian cell morphology under the control of light, or increased the flux through a synthetic metabolic pathway. Based on this experience, we outline an engineering framework for the connection of reusable protein interaction devices into self-sufficient circuits. Such a framework should help to 'refacture' protein complexity into well-defined exchangeable devices for predictive engineering. We review the foundations and initial success stories of protein synthetic biology and discuss the challenges and promises on the way from protein- to protein systems design.

INTRODUCTION

Dynamic networks of interacting proteins are the nuts, bolts, sensors and microprocessors of any cellular machinery. Networks of protein assemblies give cells their structure, provide energy, convert chemicals, sense, integrate and process information, and build or break down most other components of a cell. So when the (arguably) first generation of synthetic biologists set out to construct artificial feedback loops (1,2), oscillators (3) and toggle switches (4), why did they not tap into this rich repertoire of protein signaling? Why was the first synthetic oscillator constructed from an energy-hungry and slow network of mutually repressive transcription factors (3) —so slow, in fact, that a single period could span several cell divisions? Why, for example, was it not based on protein circuitry from neurons which fire with millisecond frequencies?

For a long time now, a large community of researchers has been studying chemistry, structure and function of proteins as well as their complexes and interactions. This includes a growing body of experience in protein design and engineering with a multitude of biotechnological applications. Evidently, we should thus be ready to jump from the manipulation of individual proteins to the design of protein systems—larger assemblies or protein networks that combine different functions. Protein circuits that integrate sensing and information processing with biochemical effectors could have enormous impact on medicine, biotechnology and the way we study and understand life. Yet, protein engineering has so far been restricted to an only auxiliary role in the design of synthetic gene circuits (5–7). The design of evenly matched, self-contained protein systems appears still out of reach. What is holding us back?

There are good reasons why the design of increasingly sophisticated gene networks was—and still is—more feasible than the development of protein circuits. The basic rules for the regulation of gene expression are rather well understood. Ideally, regulative sequences such as promoters, operators or ribosomal binding sites are more or less independent both from each other and from the protein coding region that they control. In engineering terms, they are (or can be made) 'uncoupled'. The logic of gene circuits can therefore be stitched together from linear pieces of DNA.

In contrast, the complexity and dynamics of proteins and protein networks still puzzles us. Large-scale screens continue to turn out long lists of potentially interacting

*To whom correspondence should be addressed. Tel: +34 933 160 186; Fax: +34 933 160 099; Email: raik.gruenberg@crg.es
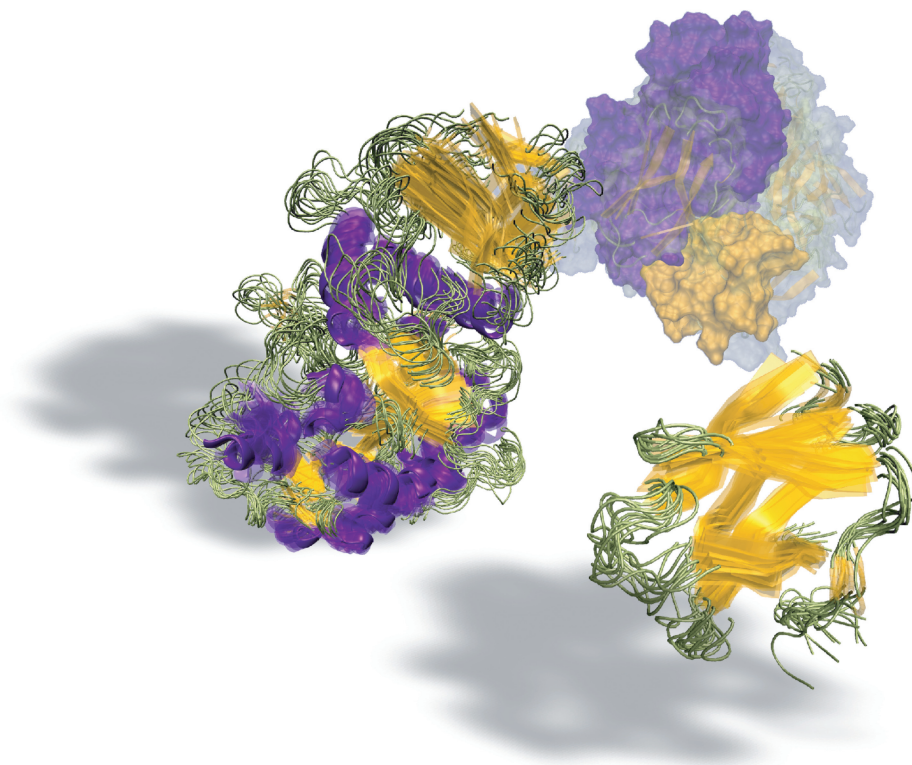
Figure 1. Already single proteins are complex dynamic systems but they are open to scrutiny by experimental and computational methods. Simplified structures of an enzyme (glycosyltransferase, left) and its inhibitor (right) are shown as ensembles of snapshots taken from molecular dynamics simulations. The specific complex of the two proteins is shown in the background together with alternative non-native orientations from a docking calculation. Binding is governed by diffusion but may also require the correct matching of quickly interchanging conformational states. The stability of the complex is then influenced by the redistribution of dynamics between different protein regions as well as the surrounding solvent [simulation and docking data taken from (11)].

proteins, often with little overlap between experiments (8). Furthermore, many reproducibly verified physical interactions may still turn out to be 'noise' without functional relevance (9). Our understanding of even the best studied signaling pathways is still far from complete. In fact, the very concept of cascading pathways may be misleading (10). Information is often processed through the cooperative re-arrangement and modification of pre-assembled protein complexes (10) and 'cross-talk' (at least in eukaryotes) is the rule not the exception. Adding to the puzzle is the complexity of individual proteins. The stability and kinetics of their interactions is governed by a complex interplay of atomic structure and dynamics spanning several scales of length and time (11,12) (Figure 1).

On the surface, all this complexity appears to leave no hope for the rational design of sophisticated protein circuitry, at least, not in the near future. Yet, here we show that efforts in this direction are well underway and progress is being made. Several recent studies have utilized the natural modularity of proteins and managed to rewire signaling networks by the clever exchange and transfer of individual protein domains. Many more have fused unrelated domains into synthetic protein switches. Missing, however, are conceptual frameworks (13,14) for the design of 'plug-and-play' protein devices—devices that would be mutually compatible and reusable for the

construction of sophisticated multicomponent protein systems. We briefly review the foundational technologies that will help us to reach this next level of protein systems design. We will then document initial success stories in the rewiring of signaling networks and the construction of modular protein switches. Our second purpose is to outline an engineering framework for protein synthetic biology as it is emerging from these works. The framework is based on the modularity of specific interactions, and we discuss its possible applications and challenges.

FOUNDATIONS FOR PROTEIN SYNTHETIC BIOLOGY

Synthetic Biology aims to prepare the ground for the routine engineering of complex biological systems (13,15). The foundations for a protein synthetic biology are, in fact, more solid than for many other areas in this young field. A whole industry supports biochemists in the manipulation and production of recombinant proteins. Small-and large-scale initiatives provide atomic structures (16), electron microscopy (17) and other methods yield pictures of large assemblies and a wide range of biophysical methods are dedicated to the detailed study of protein function and dynamics. The experimental methods are complemented by a rich set of modeling tools. Quantum mechanic calculations describe fast

reaction mechanisms at the subatomic level (18). Molecular mechanics strategies push the simulation of atomic dynamics into the microsecond time range (19). Higher order approximations (18) support rational design (20), virtual screening for binding partners (21) or the prediction of structures (22) and assembly geometries (23). Granted, none of this is easy. On the other hand, synthetic biologists have the luxury to cherry-pick well-characterized systems for which these methods actually work. A protein systems engineer can thus establish a near-complete chain of information from macroscopic quantities such as rate constants or stabilities down to subatomic detail. In contrast, most synthetic biology projects currently rely on the art of 'black box engineering' with only partial understanding of the systems they are dealing with. Synthetic gene networks, for example, depend on complex transcription and translation machineries and are subject to cell-state variation and other 'side effects'. Protein-only circuits would be amenable to a more rational design approach—they could be optimized *in vitro* and be tested in solutions or extracts of increasing complexity before being employed to actual cells. RNA-based devices (24) or DNA computation systems (25) may offer similar levels of control and, like in natural cells, DNA, RNA and protein devices could in some future complement each other in synthetic systems (15).

The engineering of individual proteins has matured into a full-fledged scientific discipline with important applications. Traditionally, this field had been dominated by directed evolution methods which pan large pools of proteins with partly randomized sequence (26). More recently, computational protein design methods are becoming increasingly successful at the structure-based engineering of protein folds, interactions and activities (20). A combination of both approaches has recently culminated in the *de-novo* design of two enzymes (27,28) with novel activities that are not found in nature. Increasingly though, protein engineers shift their attention from the manipulation of residues within individual globular proteins to the recombination and fusion of whole protein domains (29–32).

THE FIRST STEPS INTO PROTEIN SYNTHETIC BIOLOGY

Cells process information through networks of large dynamic protein assemblies. The complexity of these systems is kept in check by natural modularity. Catalytic activities, their inhibition, conditional localization and interactions with other proteins are often split up between independently folding protein domains (10,33) which are interspersed by unstructured regions and linear motifs (34,35). Rational tinkering with this domain composition has led to some surprisingly straightforward cases of cellular re-programming. Most of this work has already been very well analyzed in dedicated reviews (36,37). Here, we aim to extract common themes that may reveal the contours of a general engineering framework for protein synthetic biology.

Pathway rewiring with adapters and scaffolds

Scaffold or adapter proteins co-recruit signaling components, for example, kinases and their substrates, into functional assemblies (33,37,38). They thus channel signals through networks of overlapping and cross-reacting 'pathways'. A series of domain swapping experiments established the crucial function of scaffolding within the yeast MAP kinase signaling networks (39–42). This seminal work is summarized in the Appendix and the most recent experiment is described in Figure 2A.

The success of these studies seemingly confirms that scaffold proteins act by physically tethering each kinase to its subsequent substrate, as shown in Figure 2A. Yet, this intuitive '*cis* model' is challenged by theoretical and experimental data (43,44). Natural MAPK signaling involves the membrane recruitment of the activated scaffold (45–47). Synthetic domain swaps established that the relocalization of the scaffold (48) as well as of another upstream kinase (49) are, in fact, a prerequisite for signal transduction. Furthermore, the cytosolic scaffold protein appears only partially occupied and incapable of promoting processive phosphorylation in *cis* (44). The scaffold is thus more likely operating in *trans*, by enriching signaling components in membrane-associated clusters (43,44). This would also explain why, in neither case, there appeared any need for optimizing the spatial arrangement and orientation of the synthetic protein fusions. Rather than depending on exact positioning and timing, the various domain swaps may have benefited from a simpler, but more robust, colocalization effect (Figure 2B).

Also a prominent success in the rewiring of mammalian signaling relied on relatively unspecific colocalization: Howard *et al.* (52) used a chimeric adaptor to recruit an apoptosis signaling protein (caspase 8) to active growth hormone receptors. At least under certain conditions, the growth signal was thus indeed rewired into the opposing apoptosis. The completely unrelated growth receptor cannot quite be expected to activate a caspase by any direct means. Yet, the clustering around the activated receptor was sufficient to trigger caspase dimerization and activation.

Most, if not all, examples of modular signal network rewiring published to date, thus appear to have followed the same strategy: (i) identify an adapter protein that changes localization and/or clustering due to some natural input signal, (ii) identify an unrelated signaling protein that is activated by the recruitment to the same compartment, (iii) introduce a specific protein–protein interaction that connects the signaling intermediate to the alien adapter. Instead of relying on a natural scaffold, membrane recruitment of individual proteins can also be controlled by a drug-induced (53–56) or a light-triggered (57) protein–protein interaction and this alone is often sufficient to trigger various signaling responses (53,58–63) with high temporal (60) or even spatiotemporal (63) control.

Another lesson from these studies concerns modularity itself. Rather than swapping domains, engineers were, in fact, swapping interaction pairs. That means, the actual
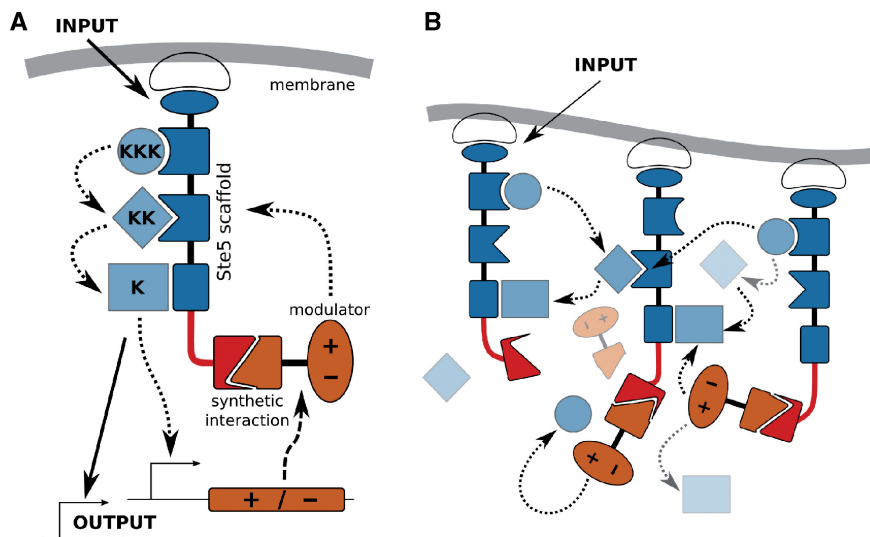
Figure 2. Rewiring of MAPK signaling in yeast. (**A**) *cis* model of scaffold action: the scaffold protein Ste5 channels the signal from upstream activators through a phosphorylation cascade of three kinases (MAPKKK, MAPKK, MAPK) to the activation of mating response genes. Natural scaffold and kinases are colored in blue. A synthetic extension of this scaffold is shown in red. Bashor and colleagues used this extension for the recruitment of positive or negative modulator proteins to the scaffold complex. Modulators were expressed from a mating response promoter and were thus closing a positive or negative feedback loop. (**B**) *trans* or cluster model of scaffold action: signal transduction depends on the relocalization of Ste5 to the plasma membrane (45,48) and kinase activation seems to propagate through clusters of only partially occupied scaffolds rather than within individual complexes (44). The synthetic recruitment would increase the local concentration of modulator proteins within these signaling clusters. See text for details [simplified; partially adapted from (36,42,44)].

unit of engineering were pairs of specifically interacting domains or pairs of a domain and its cognate binding peptide. Such 'interaction devices' were either rewired within a pathway (40,52) or transferred from entirely different contexts (40,42,48,63). Specific synthetic interactions increase local concentrations of kinase substrates, metabolic intermediates (64), or receptor ligands (65,66). Such co-recruitment effects are often enough to re-route signaling but can also accelerate metabolic pathways (64).

Building modular protein switches

Many enzymes, in particular, kinases and phosphatases are inactive by default and they get switched on only for signal processing. A common natural 'design pattern' for this kind of regulation is modular autoinhibition (67). Autoinhibitory domains establish intramolecular interactions that block the activity of another domain within the same molecule. The inhibiting interaction may, for example, sterically occlude the active site of a kinase domain or may inactivate its catalytic activity due to conformational strain. Autoinhibition is then relieved by covalent modifications (e.g. de-/phosphorylation) of the interaction region, by proteolysis, or by a higher affinity binding partner arriving in *trans*. The autoinhibited protein thus turns into a switch with built-in signal processing which may be amenable to modular engineering.

Dueber and coworkers (68,69) swapped the autoinhibitory interaction module of the yeast kinase N-WASP for several domain-peptide interactions from unrelated signaling proteins. A pair of phosphorylation-dependent input interactions put the N-WASP output

(actin polymerization) under the control of two unrelated kinases (68). The fusion to constitutively interacting domain–peptide pairs rendered N-WASP responsive to competing peptide ligands (69). Different combinations and arrangements of input interactions lead to various gating behaviors (including AND, OR) and switching dynamics. The same strategy and, in fact, some of the very same heterologous interaction domains, were later also applied to re-program guanine nucleotide exchange factors (70).

Natural systems sometimes conserve the same modular domain architecture and similar structural mechanisms for the processing very different signals in different cells or contexts. Signal rewiring can then be a relatively simple matter of swapping homologous domains, even across kingdoms (71). An example is the replacement of a non-light-sensitive LOV (light, oxygen, voltage) domain by a light sensitive homolog, which converted a voltage-dependent histidine kinase into a light-triggered one (72).

Systems where regulation and activity are naturally separated into domains, as in the examples above, are evidently prime candidates for domain-based engineering. Nevertheless, a large number of modular protein switches have also been engineered without co-opting natural regulation [see (73) for a comprehensive review]. A common success strategy is the mutual coupling of overlapping protein domains, which means two domains are tightly fused or inserted into each other so that the folding of one domain restricts (or, less commonly, assists) the functioning of the other. Small ligand, peptide or protein binding partners then stabilize one domain and reduce (or increase) the activity of the other. Protein domain or domain–peptide interactions are therefore

again important building blocks in many of these constructs. Ligand-sensing domains have been inserted into loops of dihydrofolate reductase (DHFR) (74) and β-lactamase (75) producing enzymes with artificial allosteric regulation. Similar effects were reached by inserting, vice versa, lactamase variants into ligand-binding proteins (76,77). Sallee *et al*. (78) searched databases for small sequence overlaps between unrelated interaction domains and constructed several two-domain fusion proteins (or peptides) with mutual exclusive binding to either one or another partner. Last but not least, the careful overlapping with a photo-sensitive LOV2 domain made DNA binding of *Escherichia coli trp* repressor depending on light (79).

Conceptually, building switches by domain replacement, insertion or overlap appears straightforward. Practically however, there are issues of folding, stability and dynamics. Operational constructs are therefore often picked from screens of many variants with different insertion sites and linker lengths. Techniques and tools from structure-based computational protein design (20) have not yet been applied to this problem but could probably facilitate the effort. However, domain fusions do not necessarily compromise protein function. Insertion into the loop of a thermostable protein (80) or the fusion to well-known solubility enhancers such as maltose-binding protein or glutathione *S*-transferase are, in fact, strategies to stabilize a protein fold.

CHALLENGES

Many labortories have started engineering proteins at the level of domains rather than single residues. A few have also ventured into the rerouting of well-studied signaling networks. Yet, the design of more complex systems, comprising more than one or two synthetic proteins, is long in coming. Progress in this area is impeded not only by technical but also conceptual issues.

From parts to DNA

The need to routinely recombine a protein from several unrelated domains and linker segments is quite different from classic cloning tasks. Traditional methods streamline the transfer of single DNA fragments into various vectors for expression or purification. In contrast, protein synthetic biologists need to assemble several DNA fragments without or with only very short intervening sequences. Gene synthesis has become attractive for obtaining codon-optimized single 'protein parts' but remains expensive when it comes to the building of numerous whole fusion constructs, which typically measure between 1000 and 2000 bp. Most of the time, these DNA templates will be mere recombinations of large recurring fragments. Paradoxically though, gene synthesis—considered the driving technology behind synthetic biology—is not at all adapted to this typical synthetic biology work flow and commercial providers re-synthesize every large construct from scratch. Until more suitable, for example recursive (81), synthesis becomes widely available, researchers are evaluating

various technologies (82) ranging from overlap extension PCR (83) or sequence-independent cloning (84) to customized restriction/ligation methods. The iterative restriction/ligation-based BioBrick assembly protocol (85,86) could serve, at least, as a temporary solution and would allow to build a collection of 'protein parts' within the Registry of Standard Biological Parts (http://partsregistry.org). Unfortunately though, the original BioBrick cloning standard (85,86) is incompatible with protein fusions. Two follow-up standards have been proposed early on (87,88) and were later formally registered as BioBrick Foundation Request For Comments (BBF RFC 23 and 25). The two formats retain some, although imperfect, compatibility with each other (89) as well as with the old standard (referred to as BBF RFC 10). The BioBrick community has not settled on either of these formats and new proposals continue to be made and used. The standardization framework of the BioBrick Foundation is thus put to a first serious test, right after inception. The consistent documentation and naming of new standard proposals can be considered an initial success. Hopefully, the community will now face up to the challenge and agree on a common solution for the growing number of innovative protein parts that have been entering the Registry for several years already.

From DNA to protein

One and the same protein sequence can be encoded by a large number of synonymous DNA sequences. The actual codon choice often has a strong effect on protein expression. Yet, for a long time, exact rules for the rational optimization of codon usage have remained elusive (90). Studies on large cohorts of synonymous sequences have now quantified the importance of mRNA secondary structure around the translation start site (91,92) and highlight the large effect of synonymous codon usage throughout the sequence (93). Interestingly, these very recent results contradict previous *ad hoc* models of optimal codon usage. While the situation is improving, our current data remain anecdotal in the sense that they are based on a very limited number of actual proteins. More work is needed to broaden this data and to explore other factors like the relation between translation speed and proper folding (94,95).

Folding or rather misfolding, aggregation and toxicity pose a perhaps more difficult problem for the completely rational engineering of synthetic proteins. Their complex chemistry, three-dimensional structure and dynamics bestows 'personality' on individual proteins—exactly what synthetic biologists would like to avoid. Especially the overexpression and purification of proteins for *in vitro* work often require individually adapted protocols. However, while there is no one-fits-all solution, a very limited set of protocols covers most of the cases and has been compiled into a consensus strategy (96) of 'what to try first'. A standardization of these protocols would help the exchange of protein parts and would make experimental data more comparable. This, in turn, could aid the development of computational tools that predict solubility and other parameters from structural

information. Current sequence-based methods (97,98) work well for some systems but are not robust and accurate enough to substitute for the trial and error routine of protein biochemistry.

Moreover, synthetic biologists can and should focus on a limited subset of well-behaved and characterized building blocks. Such a trend is already apparent from the current literature. A small set of interaction partners or interfaces to cellular signaling and gene expression have been used and reused in different combinations: Many studies work with the same drug- or light-induced protein complexes or with certain peptide binding SH3 and PDZ domains. Such 'input interactions' are often wired to the same signaling proteins like, for example, protein kinase A, N-WASP or GEF DH. Also the synthetic triggering of gene expression often relies on the same yeast-two-hybrid constructs. Such reusability across laboratories depends on careful documentation of experimental conditions and experiences. A physical or virtual parts registry could help to collate and expand this information which until now remains spread throughout literature and laboratory notebooks.

Last but not least, synthetic systems require means to balance the expression levels and concentrations of two or more proteins within a cell. This is currently best achieved by regulated expression from genomically integrated genes in simple host organisms like yeast or *E. coli*. In mammalian cell lines, such stable integration is relatively difficult. Vectors that express multiple proteins from a single plasmid using independent or bidirectional promoters (99) or from Internal Ribosome Entry Sites (IRES) (100) may offer an alternative.

From proteins to systems

Perhaps the biggest obstacle on our way to a flourishing ecosystem of 'plug-and-play' protein systems is the lack of a universal abstraction and interfacing strategy (13,14,24). A prototypical synthetic protein circuit may, for example, evaluate different molecular health sensors and send malignant cells into apoptosis or initiate self-destruction otherwise. Synthetic systems thus need to integrate a sensory input layer with an information processing network in order to, ultimately, trigger some useful output. In principle, all three layers could be realized with proteins and there are countless natural examples of this architecture. However, unlike gene regulatory networks, protein networks process signals by a complex combination of mutual modification, allosteric regulation, active transport and many other mechanisms. What keeps cell biologists excited and on their toes, is more akin to a nightmare for synthetic biologists. Every case seems to be special and entangled with everything else.

Can we, nonetheless, extract reusable protein modules from natural networks? How can we formulate devices that are cross-compatible? Can we hide protein regulation complexity behind some standard functional interface? Can we thus formulate an abstraction hierarchy (13,14) that allows us to rewire refined protein devices into more and more sophisticated systems? Here we argue that this may indeed be feasible. The key is to focus on modular molecular interactions.

A PROTEIN DEVICE FRAMEWORK

Definitions and abstraction hierarchy

Synthetic biologists aim to 'redesign' or 'reformulate' nature. First, however, they tend to reformulate the way we talk about biological systems. Many new terms are borrowed from (electrical) engineering or programming and then applied to molecules and cells that are actually quite unlike the screws, wires and circuits from the engineering catalogs. The new language does thus not necessarily compensate for lack of words but is itself part of the experiment. Some engineering terms have taken on new meaning and inspired new experiments in synthetic biology. In particular, these include 'part', 'device', and 'system' (13). Exact definitions are still evolving and here we re-iterate what we think is the emerging consensus.

A '**Part**' is a component of some functional interest. Protein parts may, for example, be single domains, reusable linker sequences or signal peptides and purification tags. Such basic parts can be recombined into composite fusion proteins, which are still considered parts because they form single molecules. Parts define the *physical units* within an engineered system.

A '**Device**' is a collection of one or more parts that operate together and expose a defined (standardized) functional interface. Unlike individual parts, the different components of a device may or may not be physically connected. Importantly though, devices guarantee to interoperate with other devices according to rules of 'functional composition'. Devices thus define the *functional units* within an engineered system.

This idea of a device goes beyond Endy's original definition (13). A simple illustration is given in Figure 3: a protease cleaving a specific peptide sequence qualifies as a part but not as a device because detailed knowledge or customization is needed for its application. However, the same protease *together* with its cognate peptide would qualify: This 'proteolysis device' would take its own transcription as an input and would split any two protein parts that are fused right and left of its cognate peptide. The complexity of substrate recognition and specificity is thus encapsulated and of no concern to the engineer. Different proteolysis devices could be optimized for different catalytic efficiencies or different environments. Each of them could be combined with various regulatory transcription or translation (24,101) devices on the input side to attack any fusion of protein parts on the output. Standardization of device interfaces creates a 'functional composition framework' (13,14,24). This framework tells engineers about connection rules and characterization data that they can expect from any device falling into the same class.

'**Systems**' are combinations of devices that realize a final application. We do not make an attempt at a detailed definition as we have yet to see examples of full-fledged synthetic protein systems.

Protein interaction devices

Many proteins are modular (33) and almost all are embedded in a web of interactions with other proteins (8). In particular, signal processing networks are held together by multiple specific protein–protein contacts. Dynamic changes of these interactions are a common means for propagating and integrating signals (10,102,103). Variations on a set of binding partners are
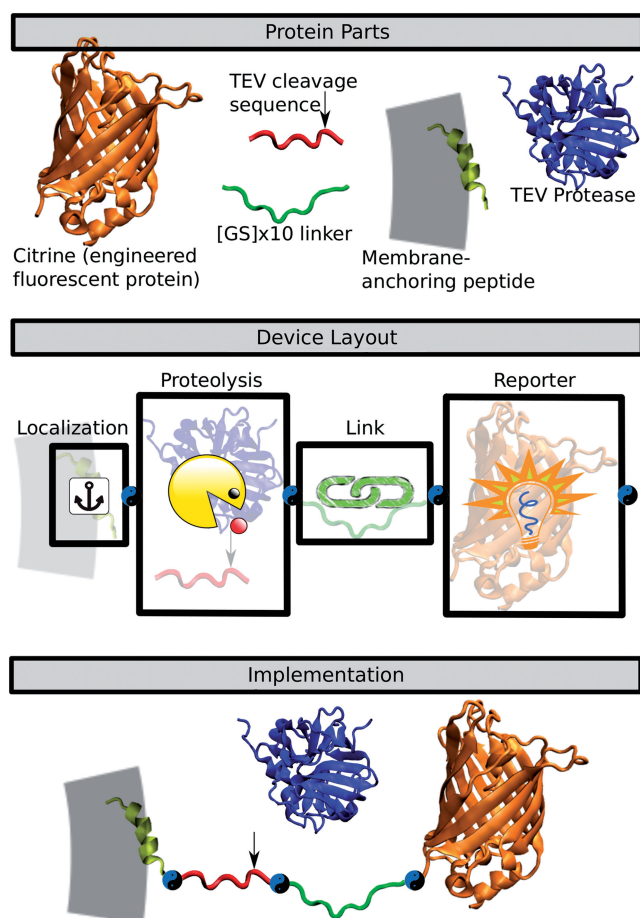

Figure 3. Protein synthetic biologists assemble artificial fusion proteins from reusable segments—or parts. Our very simple example makes the localization of a reporter dependent on the expression level of a protease. The design is simplified by the definition of 'devices' that group recurring patterns of cross-reacting parts into functional units with defined input and output. The proteolysis device in our example comprises both a protease and the peptide with its specific cleavage site. An engineer can swap different implementations of proteolysis devices (for example, using different proteases) and still expect his overall system to work.

recurring in many different contexts throughout the proteome. In other words, interactions often show a high degree of modularity. So modular that, as we have seen, they can sometimes be swapped for entirely unrelated binding pairs. Such 'interaction swapping' has emerged as the almost universal success strategy for both protein pathway and protein switch engineering (102,103).

A set of well-characterized, standardized and interchangeable protein interaction devices may therefore be the best foundation for the design of sophisticated protein systems. Protein interaction devices communicate via the creation or disruption of transient interactions. A prototypical device consists of two disconnected parts that are either creating or responding to a physical interaction between each other. The functional connection to another device occurs through a pair of protein fusions. We can organize protein interaction devices into three global classes, according to their input and output:

(1) A interaction input device (or sensor) converts some signal (environmental cues, ligands, cellular states, etc.) into the change of an interaction. Example: the drug induced interaction between FKBP and FRB (56) is a widely used device that puts the corecruitment of any two proteins under chemical control. Further examples of well-tested interaction input devices are listed in Table 1. A schematic data sheet for an interaction input device is given in Figure 4.
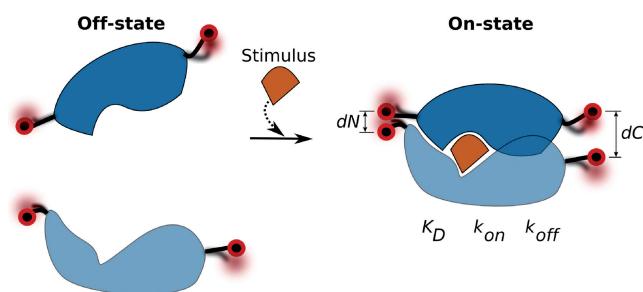(2) A interaction output device (or actuator) converts the interaction change from a connected input

Figure 4. Schematic data sheet for a protein interaction input device. This device converts a chemical signal into the corecruitment of two proteins. A popular implementation would be the rapamycin-induced interaction between FKBP12 and FRB. The device is characterized in two states—Off (unbound, without stimulus) and On (bound, after stimulus). Engineers would need to know about possible connection points for protein fusions (red dots), structural information like, for example, mean distances between N and C termini (dN, dC), as well as the binding equilibrium (K_D) and kinetics (k_{on}, k_{off}) of the fully stimulated state.

Table 1. Examples of protein interaction input devices

Device	Description	Input	References
Jun:Fos	Engineered variants of a constitutive leuzine zipper interaction	None	(111,112)
FKBP:FRB	Drug-induced heterodimerization	Rapamycin	(53–56,59–62,108,113)
FKBP:FKBP	Drug-induced homodimerization	Synthetic dimerizers	(55,114)
Gyrase B	Drug-induced and -reversible homodimerization	Coumermycin, Novobiocin	(115,116)
PIF3:PhyB	Light-induced and -reversible binding	Light	(57,63,117)

Table 2. Examples of protein interaction output devices

Device	Description/input	Output	References
Yeast-two-hybrid	Reconstitution of a transcription factor	Gene expression	(104)
MAPPIT	Reconstitution of a cytokine signaling pathway	Gene expression	(118,119)
Split DHFR	Reconstitution of DHFR	(Color) Reaction	(120)
Split lactamase	Reconstitution of β-lactamase	Antibiotic resistence; (color) reaction	(105,107)
Split luciferase	Reconstitution of different luciferases	Light	(121,122)
Split GFP	Reconstitution of green fluorescent protein variants (BiFC)	Fluorescence	(123–125)
Split ubiquitine	Reconstitution of ubiquitine	Proteolysis	(126,127)
Split TEVP	Reconstitution of tEV protease	Proteolysis	(128,129)
Split intein	Reconstitution of intein domain	Protein splicing	(113,117)
GEF	Activation of guanine nucleotide exchange factors by competition with autoinhibitory interactions	Cell morphology	(70)
Rho:membrane	Membrane recruitment of Rho GTPases	Cell morphology	(59,60,63)
Fas:Fas	Homodimerization of membrane-tethered Fas intracellular domain	Apoptosis	(55,58,114)
FRET	Fluorescence resonance energy transfer between variants of GFP	Fluorescence (different signals)	(130)

device into a useful biological action (enzymatic activity, cell signaling, reporter readout, gene expression, etc.). Example: the yeast two hybrid system (104). Further examples are listed in Table 2.

(3) A interaction transmission device uses changes of interactions both as input and output. Corecruitment of its input domains triggers, disrupts or modifies corecruitment at its output interface. No examples have been put forth yet. Many natural protein signaling networks could be logically decomposed into chains of interaction transmission devices but synthetic variants have yet to be realized.

All examples given in Tables 1 and 2 have been applied in several studies, often in combination with different fusion partners. Many of the output devices in Table 1 were developed for protein–protein interaction assays and have therefore been tested with many different input interactions in various environments. Other devices, with high-level physiological output, are of course context specific and will only work in certain cell types. The more general-purpose output devices were often constructed as protein complementation assays (PCA) (105–107). This technique of 'split protein' engineering is far from trivial but has the advantage to work for many different proteins. PCA-style engineering may thus allow us to put pretty much any protein activity under the control of interaction devices.

However, the technically most straightforward output of an interaction device would be the simple relocalization of a target protein, for example, from cytosol to membrane or from nucleus to cytosol (108). As we have discussed above, membrane recruitment was probably a key factor during the rewiring of MAP kinase and apoptosis pathways. Many proteins are spatially regulated. Relocalization between cytosol and plasma membrane can, for example, modify the activity of phosphatases (109) or even change the specificity of metabolic enzymes (110). Cases like these are the low-hanging (but nevertheless juicy) fruit for synthetic protein network engineers.

The obvious gap in the proposed device framework is the current lack of any genuine interaction transmission devices. That means we are still missing the kind of information-processing capabilities that are driving the design of synthetic gene regulatory networks and set the stage for functional RNA device frameworks (24). Although we have already seen some examples of sophisticated protein-based information processing (68–70,42), these systems relied on natural pathway responses and cannot be easily transferred or recombined. Filling this gap should become one of the primary goals of protein systems engineers.

Device characterization

In an ideal world, functional composition frameworks should allow engineers to mix and match biological devices into higher order systems with ease and reliability. Details on the inner workings of a device should be hidden behind standard interfaces. The properties of a specific device should be quantified in standardized measurements that should be directly comparable between different implementations of the same functionality. Design software could then feed this standardized information into meaningful predictive models and aid the engineering of complexity. Obviously, we are still very far from this ideal situation. One of the difficulties with the messy substrates of synthetic biology is the definition of measurement units that make devices comparable across laboratories. The activity of gene regulation devices, for example, is highly dependent on the cellular environment and synthetic biologists are therefore evaluating 'measurement kits' that provide characterization data relative to an internal standard (131).

Quite the contrary, protein interactions can be rigorously characterized by established biophysical methods. Binding affinities, kinetics, as well as enzymatic activities can be measured *in vitro* as well as *in vivo* (109,132). Most

characteristics of protein interaction devices could therefore be quantified in meaningful absolute numbers. The perfect 'data sheet'(14) of such a device would describe positions and rules for protein fusions (the physical interface) and then define inputs and outputs (the functional interface). Interaction output should be quantified thermodynamically in terms of dissociation constant (K_D) and, even more important, kinetically by on- and off-rate (k_{on}, k_{off}) for binding in the different states of the device (e.g. before and after stimulus). Figure 4 sketches a model data sheet for a well-characterized interaction input device.

The response of interaction output devices may be more difficult to characterize in a consistent manner. One could quantify activity at zero and at full recruitment. This degree of recruitment could be predicted from the K_D of the input interaction. Nevertheless, the activity of corecruited proteins may also depend on more subtle binding kinetics and absolute concentrations. Corecruitment increases the relative local concentrations of, for example, enzyme and substrate, but it can also lower the entropy penalty for secondary interactions or have more complex steric effects. Moreover, the length and composition of peptide linkers between coupled devices will often influence the transfer of information. It will therefore be interesting to see to which extend we can predict higher level device and system characteristics from hard biophysical measurements on individual parts.

WHAT WILL WE LEARN?

Synthetic protein circuits will provide an acid test for systems biology methods and our understanding in general. A system that has been built from well-characterized parts according to human specifications leaves little excuse for failed predictions. In fact, we should be able to reconstitute synthetic protein circuits *in vitro* and study them with hardly any gaps in knowledge. Sequences and structures should be known, molecular dynamics can be simulated, rates and equilibrium constants can be measured and reactions can be modeled. Carefully controlled synthetic protein systems could therefore allow us to venture deep into the *Terra incognita* between structural and systems biology and study the interdependence of protein architecture, molecular dynamics and cellular signal processing.

Synthetic multicomponent protein systems may also become valuable research tools. A first generation of simple two-component protein interaction devices have found wide-spread use as sensors and controls throughout laboratories: yeast-two-hybrid (104) and related methods convert protein binding into gene expression and have revealed millions of physical interactions. Protein complementation devices (105–107) provide alternative interaction readouts. The latest generation of drug- or even light-inducible interaction input devices (55,57,63) now allow researchers to intercept and manipulate cellular dynamics at high temporal and even spatial resolution. A few of these interaction input devices have already been combined with reusable output devices to give, for example, fine control over expression (57), proteolysis (127,129) or intein splicing (113,117). Examples are given in Tables 1 and 2.

APPLICATIONS OF PROTEIN SYSTEMS ENGINEERING

Beyond the study of signal processing, synthetic scaffold proteins are now being evaluated for biotechnological and medical applications. The recruitment of three heterologous enzymes to a single synthetic scaffold protein increased the flux through the 'synthetic metabolic pipeline' by a factor of 77 (64). Interestingly, this corecruitment of bacterial enzymes was realized through modular domain–peptide interactions that were borrowed from metazoan signaling networks. The same approach should be applicable to many other metabolic engineering projects. In a recent example of a biomedically oriented application, Cironi *et al.* (65) fused epidermal growth factor (EGF) with interferon-alpha-2a (IFNalpha-2a). The recruitment of EGF to its receptor increased the local concentration of the interferon and allowed them to weaken the interaction with the interferon receptor. Their engineered chimeric protein was therefore now targeting the interferon signal only to cells also bearing EGF receptors. The same method helped direct erythropoietin to red blood cells (66). More generally, the fusion of independent interaction domains has already been used for several other protein-based therapeutics (133).

Biosensing is an obvious application area for synthetic biology in general. Protein-based biosensors are already used in a wide range of practical settings from *in vivo* diagnosis (134) to the detection of explosives (135). Current sensors are usually based on single proteins, albeit often heavily engineered. While there would be no need to trade something simple for anything more complicated, a modular device-oriented approach could probably speed up the design of new sensors and add versatility to existing ones. One could envision a layered approach with standardized interfaces between varying sensor modules, signal processing devices, and reporters. Carefully refined transmission devices for signal amplification or noise filtering could then be reused for different input sensors and could be mounted on various reporting platforms. Noisy signals from multiple sensors with overlapping specificity could be integrated directly on the chip. Protein-based components could of course also be combined with RNA or gene regulation devices into self-regenerating and self-organizing cellular biosensors.

More importantly, synthetic protein systems are positively predestined for therapeutic applications. Development costs, safety and regulatory issues, combined with sobering experiences from initial attempts at gene therapy led most synthetic biologists shy away from direct medical applications. Yet, a modular device-oriented approach to the development of therapeutic protein systems could, in fact, slash costs, shorten development and improve safety. Several waves of

protein-based therapeutics have entered the market. Proteins now represent the majority of approvals for novel drugs and the medical application of proteins is becoming routine (136). Virtually all these new drugs are single proteins, usually antibodies. As it stands, each new development starts from scratch as a single molecule needs to be optimized for safety, delivery and therapeutic effect. It is not that difficult to imagine a different approach where we separate the development of specific targeting and delivery modules from the design of protein effector and regulation devices. Components from the different layers could be tested and perhaps also approved separately, speeding up development, lowering costs, and improving safety. The same effector, for example, a trigger of apopoptosis, could then be re-used for different diseases in different tissues. Moreover, cell type-specific domains could be used for the delivery of, first, a diagnostic marker and, later, for therapeutic cargo. Viral vectors may be refactored into ferries for small protein circuits or encoding mRNA. Protein-based circuits would not compromise genomic DNA, yet, could very specifically interfere with cellular signaling and be cleanly disposed afterwards. The use of multiple components and simple information processing devices would, without doubt, increase specificity and reduce side-effects. The development of functionally compatible protein parts thus holds great promise for a new modular medicine.

CONCLUSION

Superficially, the field of synthetic biology is currently dominated by the manipulation of gene regulatory networks. However, speed, versatility and a large body of knowledge all point to proteins as an optimal substrate for biological systems engineering. In fact, a string of recent studies have illustrated this potential. Nevertheless, the bewildering complexity of proteins remains to be tamed by a robust engineering framework. Such a framework, based on natural modularity and specific interactions, appears now within reach and may allow the assembly of synthetic networks from reusable protein (and non-protein) devices. Just as in natural cells, protein interaction devices are poised to take center stage in future systems that integrate synthetic RNA and gene networks with non-natural chemistry and metabolic engineering.

Conflict of interest statement. None declared.

REFERENCES

1. Becskei,A. and Serrano,L. (2000) Engineering stability in gene networks by autoregulation. *Nature*, **405**, 590–593.
2. Becskei,A., Seraphin,B. and Serrano,L. (2001) Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J.*, **20**, 2528–2535.
3. Elowitz,M. and Leibler,S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature*, **403**, 335–338.
4. Gardner,T.S., Cantor,C.R. and Collins,J.J. (2000) Construction of a genetic toggle switch in Escherichia coli. *Nature*, **403**, 339–342.
5. Kramer,B.P. and Fussenegger,M. (2005) Hysteresis in a synthetic mammalian gene network. *Proc. Natl Acad. Sci. USA*, **102**, 9517–9522.
6. Ajo-Franklin,C.M., Drubin,D.A., Eskin,J.A., Gee,E.P.S., Landgraf,D., Phillips,I. and Silver,P.A. (2007) Rational design of memory in eukaryotic cells. *Genes Dev.*, **21**, 2271–2276.
7. Buchler,N.E. and Cross,F.R. (2009) Protein sequestration generates a flexible ultrasensitive response in a genetic network. *Mol. Syst. Biol.*, **5**, 272–272.
8. Yu,H., Braun,P., Yildirim,M.A., Lemmens,I., Venkatesan,K., Sahalie,J., Hirozane-Kishikawa,T., Gebreab,F., Li,N., Simonis,N. *et al.* (2008) High-quality binary protein interaction map of the yeast interactome network. *Science*, **322**, 104–110.
9. Levy,E.D., Landry,C.R. and Michnick,S.W. (2009) How perfect can protein interactomes be? *Sci. Signal.*, **2**, pe11.
10. Gibson,T.J. (2009) Cell regulation: determined to signal discrete cooperation. *Trends Biochem. Sci.*, **34**, 471–482.
11. Grünberg,R., Leckner,J. and Nilges,M. (2004) Complementarity of structure ensembles in protein-protein binding. *Structure*, **12**, 2125–2136.
12. Grünberg,R., Nilges,M. and Leckner,J. (2006) Flexibility and conformational entropy in protein-protein binding. *Structure*, **14**, 683–693.
13. Endy,D. (2005) Foundations for engineering biology. *Nature*, **438**, 449–453.
14. Canton,B., Labno,A. and Endy,D. (2008) Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol.*, **26**, 787–793.
15. Andrianantoandro,E., Basu,S., Karig,D.K. and Weiss,R. (2006) Synthetic biology: new engineering rules for an emerging discipline. *Mol. Syst. Biol.*, **2**, 2006.0028.
16. Terwilliger,T.C., Stuart,D. and Yokoyama,S. (2009) Lessons from structural genomics. *Ann. Rev. Biophy.*, **38**, 371–383.
17. Leis,A., Rockel,B., Andrees,L. and Baumeister,W. (2009) Visualizing cells at the nanoscale. *Trends Biochem. Sci.*, **34**, 60–70.
18. Sherwood,P., Brooks,B.R. and Sansom,M.S. (2008) Multiscale methods for macromolecular simulations. *Curr. Opin. Struct. Biol.*, **18**, 630–640.
19. Klepeis,J.L., Lindorff-Larsen,K., Dror,R.O. and Shaw,D.E. (2009) Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.*, **19**, 120–127.
20. Van der Sloot,A.M., Kiel,C., Serrano,L. and Stricher,F. (2009) Protein design in biological networks: from manipulating the input to modifying the output. *Protein Eng. Des. Sel.*, **22**, 537–542.
21. Totrov,M. and Abagyan,R. (2008) Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr. Opin. Struct. Biol.*, **18**, 178–184.
22. Moult,J., Fidelis,K., Kryshtafovych,A., Rost,B. and Tramontano,A. (2009) Critical assessment of methods of protein structure prediction – round VIII. *Proteins*, **77(Suppl. 9)**, 1–4.
23. Ritchie,D.W. (2008) Recent progress and future directions in protein-protein docking. *Curr. Protein Pept. Sci.*, **9**, 1–15.
24. Win,M.N., Liang,J.C. and Smolke,C.D. (2009) Frameworks for programming biological function through RNA parts and devices. *Chem. Biol.*, **16**, 298–310.
25. Zhang,D.Y., Turberfield,A.J., Yurke,B. and Winfree,E. (2007) Engineering entropy-driven reactions and networks catalyzed by DNA. *Science*, **318**, 1121–1125.

26. Bershtein,S. and Tawfik,D.S. (2008) Advances in laboratory evolution of enzymes. *Curr. Opin. Chem. Biol.*, **12**, 151–158.

27. Röthlisberger,D., Khersonsky,O., Wollacott,A.M., Jiang,L., DeChancie,J., Betker,J., Gallaher,J.L., Althoff,E.A., Zanghellini,A., Dym,O. *et al.* (2008) Kemp elimination catalysts by computational enzyme design. *Nature*, **453**, 190–195.

28. Jiang,L., Althoff,E.A., Clemente,F.R., Doyle,L., Röthlisberger,D., Zanghellini,A., Gallaher,J.L., Betker,J.L., Tanaka,F., Barbas,C.F. *et al.* (2008) De novo computational design of retro-aldol enzymes. *Science*, **319**, 1387–1391.

29. McDaniel,R., Ebert-Khosla,S., Hopwood,D.A. and Khosla,C. (1995) Rational design of aromatic polyketide natural products by recombinant assembly of enzymatic subunits. *Nature*, **375**, 549–554.

30. Binz,H.K. and Plückthun,A. (2005) Engineered proteins as specific binding reagents. *Curr. Opin. Biotechnol.*, **16**, 459–469.

31. Heyman,A., Barak,Y., Caspi,J., Wilson,D.B., Altman,A., Bayer,E.A. and Shoseyov,O. (2007) Multiple display of catalytic modules on a protein scaffold: nano-fabrication of enzyme particles. *J. Biotechnol.*, **131**, 433–439.

32. Parmeggiani,F., Pellarin,R., Larsen,A.P., Varadamsetty,G., Stumpp,M.T., Zerbe,O., Caflisch,A. and Plückthun,A. (2008) Designed armadillo repeat proteins as general peptide-binding scaffolds: consensus design and computational optimization of the hydrophobic core. *J. Mol. Biol.*, **376**, 1282–1304.

33. Pawson,T. and Nash,P. (2003) Assembly of cell regulatory systems through protein interaction domains. *Science*, **300**, 445–452.

34. Neduva,V. and Russell,R.B. (2005) Linear motifs: evolutionary interaction switches. *FEBS Lett.*, **579**, 3342–3345.

35. Diella,F., Haslam,N., Chica,C., Budd,A., Michael,S., Brown,N.P., Trave,G. and Gibson,T.J. (2008) Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front. Biosci. A Journal and Virtual Library*, **13**, 6580–6603.

36. Pryciak,P.M. (2009) Designing new cellular signaling pathways. *Chem. Biol.*, **16**, 249–254.

37. Zeke,A., Lukács,M., Lim,W.A. and Reményi,A. (2009) Scaffolds: interaction platforms for cellular signalling circuits. *Trends Cell Biol.*, **19**, 364–374.

38. Pawson,T. and Scott,J.D. (1997) Signaling through scaffold, anchoring and adaptor proteins. *Science*, **278**, 2075–2080.

39. Harris,K., Lamson,R.E., Nelson,B., Hughes,T.R., Marton,M.J., Roberts,C.J., Boone,C. and Pryciak,P.M. (2001) Role of scaffolds in MAP kinase pathway specificity revealed by custom design of pathway-dedicated signaling proteins. *Curr. Biol.*, **11**, 1815–1824.

40. Park,S., Zarrinpar,A. and Lim,W.A. (2003) Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms. *Science*, **299**, 1061–1064.

41. Grewal,S., Molina,D. and Bardwell,L. (2006) Mitogen-activated protein kinase (MAPK)-docking sites in MAPK kinases function as tethers that are crucial for MAPK regulation in vivo. *Cell. Signal.*, **18**, 123–134.

42. Bashor,C.J., Helman,N.C., Yan,S. and Lim,W.A. (2008) Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics. *Science*, **319**, 1539–1543.

43. Pincet,F. (2007) Membrane recruitment of scaffold proteins drives specific signaling. *PLoS ONE*, **2**, e977.

44. Takahashi,S. and Pryciak,P.M. (2008) Membrane localization of scaffold proteins promotes graded signaling in the yeast MAP kinase cascade. *Curr. Biol. CB*, **18**, 1184–1191.

45. Pryciak,P.M. and Huntress,F.A. (1998) Membrane recruitment of the kinase cascade scaffold protein ste5 by the gbetagamma complex underlies activation of the yeast pheromone response pathway. *Genes Dev.*, **12**, 2684–2697.

46. vanDrogen,F., Stucke,V.M., Jorritsma,G. and Peter,M. (2001) MAP kinase dynamics in response to pheromones in budding yeast. *Nat. Cell Biol.*, **3**, 1051–1059.

47. Maeder,C.I., Hink,M.A., Kinkhabwala,A., Mayr,R., Bastiaens,P.I.H. and Knop,M. (2007) Spatial regulation of fus3 MAP kinase activity through a reaction-diffusion mechanism in yeast pheromone signalling. *Nat. Cell Biol.*, **9**, 1319–1326.

48. Winters,M.J., Lamson,R.E., Nakanishi,H., Neiman,A.M. and Pryciak,P.M. (2005) A membrane binding domain in the ste5 scaffold synergizes with Gβγ binding to control localization and signaling in pheromone response. *Mol. Cell*, **20**, 21–32.

49. Takahashi,S. and Pryciak,P.M. (2007) Identification of novel membrane-binding domains in multiple yeast cdc42 effectors. *Mol. Biol. Cell*, **18**, 4945–4956.

50. Chen,R.E. and Thorner,J. (2007) Function and regulation in MAPK signaling pathways: lessons learned from the yeast saccharomyces cerevisiae. *Biochim. Biophys. Acta*, **1773**, 1311–1340.

51. Ingolia,N.T. and Murray,A.W. (2007) Positive-feedback loops as a flexible biological module. *Curr. Biol. CB*, **17**, 668–677.

52. Howard,P.L., Chia,M.C., Rizzo,S.D., Liu,F. and Pawson,T. (2003) Redirecting tyrosine kinase signaling to an apoptotic caspase pathway through chimeric adaptor proteins. *Proc. Natl Acad. Sci. USA*, **100**, 11267–11272.

53. Belshaw,P.J., Ho,S.N., Crabtree,G.R. and Schreiber,S.L. (1996) Controlling protein association and subcellular localization with a synthetic ligand that induces heterodimerization of proteins. *Proc. Natl Acad. Sci. USA*, **93**, 4604–4607.

54. Liberles,S.D., Diver,S.T., Austin,D.J. and Schreiber,S.L. (1997) Inducible gene expression and protein translocation using nontoxic ligands identified by a mammalian three-hybrid screen. *Proc. Natl Acad. Sci. USA*, **94**, 7825–7830.

55. Clackson,T., Yang,W., Rozamus,L.W., Hatada,M., Amara,J.F., Rollins,C.T., Stevenson,L.F., Magari,S.R., Wood,S.A., Courage,N.L. *et al.* (1998) Redesigning an FKBP-ligand interface to generate chemical dimerizers with novel specificity. *Proc. Natl Acad. Sci. USA*, **95**, 10437–10442.

56. Banaszynski,L.A., Liu,C.W. and Wandless,T.J. (2005) Characterization of the FKBP.rapamycin.FRB ternary complex. *J. Am. Chem. Soc.*, **127**, 4715–4721.

57. Shimizu-Sato,S., Huq,E., Tepperman,J.M. and Quail,P.H. (2002) A light-switchable gene promoter system. *Nat. Biotechnol.*, **20**, 1041–1044.

58. Spencer,D.M., Belshaw,P.J., Chen,L., Ho,S.N., Randazzo,F., Crabtree,G.R. and Schreiber,S.L. (1996) Functional analysis of fas signaling in vivo using synthetic inducers of dimerization. *Curr. Biol. CB*, **6**, 839–847.

59. Castellano,F., Montcourrier,P., Guillemot,J.C., Gouin,E., Machesky,L., Cossart,P. and Chavrier,P. (1999) Inducible recruitment of cdc42 or WASP to a cell-surface receptor triggers actin polymerization and filopodium formation. *Curr. Biol. CB*, **9**, 351–360.

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

61. Suh,B., Inoue,T., Meyer,T. and Hille,B. (2006) Rapid chemically induced changes of PtdIns(4,5)P2 gate KCNQ ion channels. *Science*, **314**, 1454–1457.

62. Inoue,T., Meyer,T. and Insall,R. (2008) Synthetic activation of endogenous PI3K and rac identifies an AND-Gate switch for cell polarization and migration. *PLoS ONE*, **3**, e3068.

63. Levskaya,A., Weiner,O.D., Lim,W.A. and Voigt,C.A. (2009) Spatiotemporal control of cell signalling using a light-switchable protein interaction. *Nature*, **461**, 997–1001.

64. Dueber,J.E., Wu,G.C., Malmirchegini,G.R., Moon,T.S., Petzold,C.J., Ullal,A.V., Prather,K.L.J. and Keasling,J.D. (2009) Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.*, **27**, 753–759.

65. Cironi,P., Swinburne,I.A. and Silver,P.A. (2008) Enhancement of cell type specificity by quantitative modulation of a chimeric ligand. *J. Biol. Chem.*, **283**, 8469–8476.

66. Taylor,N.D., Way,J.C., Silver,P.A. and Cironi,P. (2010) Anti-glycophorin single-chain fv fusion to low-affinity mutant erythropoietin improves red blood cell-lineage specificity. *Protein Eng. Des. Sel.*, doi: 10.1093/protein/gzp085 [Epub ahead of print, 18, January 2010].

67. Pufall,M.A. and Graves,B.J. (2002) Autoinhibitory domains: modular effectors of cellular regulation. *Annu. Rev. Cell Dev. Biol.*, **18**, 421–462.

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

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

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

71. Antunes,M.S., Morey,K.J., Tewari-Singh,N., Bowen,T.A., Smith,J.J., Webb,C.T., Hellinga,H.W. and Medford,J.I. (2009) Engineering key components in a synthetic eukaryotic signal transduction pathway. *Mol. Syst. Biol.*, **5**, doi:10.1038/msb2009.28.

72. Möglich,A., Ayers,R.A. and Moffat,K. (2009) Design and signaling mechanism of Light-Regulated histidine kinases. *J. Mol. Biol.*, **385**, 1433–1444.

73. Ostermeier,M. (2009) Designing switchable enzymes. *Curr. Opin. Struct. Biol.*, **19**, 442–448.

74. Tucker,C.L. and Fields,S. (2001) A yeast sensor of ligand binding. *Nat. Biotechnol.*, **19**, 1042–1046.

75. Edwards,W.R., Busse,K., Allemann,R.K. and Jones,D.D. (2008) Linking the functions of unrelated proteins using a novel directed evolution domain insertion method. *Nucl. Acids Res.*, **36**, e78.

76. Guntas,G. and Ostermeier,M. (2004) Creation of an allosteric enzyme by domain insertion. *J. Mol. Biol.*, **336**, 263–273.

77. Guntas,G., Mansell,T.J., Kim,J.R. and Ostermeier,M. (2005) Directed evolution of protein switches and their application to the creation of ligand-binding proteins. *Proc. Natl Acad. Sci. USA*, **102**, 11224–11229.

78. Sallee,N.A., Yeh,B.J. and Lim,W.A. (2007) Engineering modular protein interaction switches by sequence overlap. *J. Am. Chem. Soc.*, **129**, 4606–4611.

79. Strickland,D., Moffat,K. and Sosnick,T.R. (2008) Light-activated DNA binding in a designed allosteric protein. *Proc. Natl Acad. Sci. USA*, **105**, 10709–10714.

80. Kim,C., Pierre,B., Ostermeier,M., Looger,L.L. and Kim,J.R. (2009) Enzyme stabilization by domain insertion into a thermophilic protein. *Protein Eng. Des. Sel.*, **22**, 615–623.

81. Linshiz,G., Yehezkel,T.B., Kaplan,S., Gronau,I., Ravid,S., Adar,R. and Shapiro,E. (2008) Recursive construction of perfect DNA molecules from imperfect oligonucleotides. *Mol. Syst. Biol.*, **4**, doi:10.1038/msb2008.26.

82. Lu,Q. (2005) Seamless cloning and gene fusion. *Trends Biotechnol.*, **23**, 199–207.

83. Heckman,K.L. and Pease,L.R. (2007) Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat. Protocols*, **2**, 924–932.

84. Li,M.Z. and Elledge,S.J. (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Meth.*, **4**, 251–256.

85. Knight, T. Idempotent vector design for standard assembly of biobricks. Available at http://dspace.mit.edu/handle/1721.1/21168?show = full (2003) (15 February 2010, date last accessed).

86. Shetty,R., Endy,D. and Knight,T. (2008) Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.*, **2**, 5.

87. Phillips, I. and Silver, P. BBF RFC 23: A new biobrick assembly strategy designed for facile protein engineering. Available at http://dspace.mit.edu/handle/1721.1/32535 (15 February 2010, date last accessed).

88. Grünberg, R., Arndt, K. and Müller, K. (2009) BBF RFC 25: Fusion protein (Freiburg) biobrick assembly standard. Available at http://dspace.mit.edu/handle/1721.1/45140 (15 February 2010, date last accessed).

89. Grünberg, R. (2009) BBF RFC 24: conversion of freiburg (Fusion) biobricks to the silver (BioFusion) format. Available at http://dspace.mit.edu/handle/1721.1/44961 (15 February 2010, date last accessed).

90. Wu,G., Dress,L. and Freeland,S.J. (2007) Optimal encoding rules for synthetic genes: the need for a community effort. *Mol. Syst. Biol.*, **3**, 134.

91. Kudla,G., Murray,A.W., Tollervey,D. and Plotkin,J.B. (2009) Coding-sequence determinants of gene expression in escherichia coli. *Science*, **324**, 255–258.

92. Salis,H.M., Mirsky,E.A. and Voigt,C.A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.*, **27**, 946–950.

93. Welch,M., Govindarajan,S., Ness,J.E., Villalobos,A., Gurney,A., Minshull,J. and Gustafsson,C. (2009) Design parameters to control synthetic gene expression in Escherichia coli. *PLoS ONE*, **4**, e7002.

94. Komar,A.A., Lesnik,T. and Reiss,C. (1999) Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. *FEBS Lett.*, **462**, 387–391.

95. Zalucki,Y.M. and Jennings,M.P. (2007) Experimental confirmation of a key role for non-optimal codons in protein export. *Biochem. Biophys. Res. Commun.*, **355**, 143–148.

96. Gräslund,S., Nordlund,P., Weigelt,J., Bray,J., Gileadi,O., Knapp,S., Oppermann,U., Arrowsmith,C., Hui,R., Ming,J. *et al.* (2008) Protein production and purification. *Nat. Methods*, **5**, 135–146.

97. Rousseau,F., Schymkowitz,J. and Serrano,L. (2006) Protein aggregation and amyloidosis: confusion of the kinds? *Curr. Opin. Struct. Biol.*, **16**, 118–126.

98. Pechmann,S., Levy,E.D., Tartaglia,G.G. and Vendruscolo,M. (2009) Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. *Proc. Natl Acad. Sci. USA*, **106**, 10159–10164.

99. Baron,U., Freundlieb,S., Gossen,M. and Bujard,H. (1995) Co-regulation of two gene activities by tetracycline via a bidirectional promoter. *Nucleic Acids Res.*, **23**, 3605–3606.

100. Jang,S.K., Krausslich,H.G., Nicklin,M.J., Duke,G.M., Palmenberg,A.C. and Wimmer,E. (1988) A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J. Virol.*, **62**, 2636–2643.

101. Isaacs,F.J., Dwyer,D.J. and Collins,J.J. (2006) RNA synthetic biology. *Nat. Biotechnol.*, **24**, 545–554.

102. Crabtree,G.R. and Schreiber,S.L. (1996) Three-part inventions: intracellular signaling and induced proximity. *Trends Biochem. Sci.*, **21**, 418–422.

103. Klemm,J.D., Schreiber,S.L. and Crabtree,G.R. (1998) Dimerization as a regulatory mechanism in signal transduction. *Annu. Rev. Immunol.*, **16**, 569–592.

104. Fields,S. and Song,O. (1989) A novel genetic system to detect protein-protein interactions. *Nature*, **340**, 245–246.

105. Galarneau,A., Primeau,M., Trudeau,L. and Michnick,S.W. (2002) [beta]-Lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein-protein interactions. *Nat. Biotechnol.*, **20**, 619–622.

106. Kerppola,T.K. (2006) Complementary methods for studies of protein interactions in living cells. *Nat. Methods*, **3**, 969–971.

107. Remy,I. and Michnick,S.W. (2007) Application of protein-fragment complementation assays in cell biology. *BioTechniques*, **42**, 137, 139, 141.

108. Haruki,H., Nishikawa,J. and Laemmli,U.K. (2008) The Anchor-Away technique: rapid, conditional establishment of yeast mutant phenotypes. *Mol. Cell*, **31**, 925–932.

109. Yudushkin,I.A., Schleifenbaum,A., Kinkhabwala,A., Neel,B.G., Schultz,C. and Bastiaens,P.I.H. (2007) Live-cell imaging of enzyme-substrate interaction reveals spatial regulation of PTP1B. *Science*, **315**, 115–119.

110. Heilmann,I., Pidkowich,M.S., Girke,T. and Shanklin,J. (2004) Switching desaturase enzyme specificity by alternate subcellular targeting. *Proc. Natl Acad. Sci. USA*, **101**, 10266–10271.

111. Mason,J.M. and Arndt,K.M. (2004) Coiled coil domains: stability, specificity and biological implications. *Chembiochem*, **5**, 170–176.

112. Mason,J.M., Hagemann,U.B. and Arndt,K.M. (2007) Improved stability of the Jun-Fos activator protein-1 coiled coil motif: a stopped-flow circular dichroism kinetic analysis. *J. Biol. Chem.*, **282**, 23015–23024.

113. Mootz,H.D., Blum,E.S., Tyszkiewicz,A.B. and Muir,T.W. (2003) Conditional protein splicing: A new tool to control protein structure and function in vitro and in vivo. *J. Am. Chem. Soc.*, **125**, 10561–10569.

114. Amara,J.F., Clackson,T., Rivera,V.M., Guo,T., Keenan,T., Natesan,S., Pollock,R., Yang,W., Courage,N.L. and Holt,D.A. (1997) A versatile synthetic dimerizer for the regulation of protein-protein interactions. *Proc. Natl Acad. Sci. USA*, **94**, 10618–10623.

115. Gormley,N.A., Orphanides,G., Meyer,A., Cullis,P.M. and Maxwell,A. (1996) The interaction of coumarin antibiotics with fragments of the DNA gyrase b protein. *Biochemistry*, **35**, 5083–5092.

116. Zhao,H.F., Boyd,J., Jolicoeur,N. and Shen,S.H. (2003) A coumermycin/novobiocin-regulated gene expression system hum. *Gene Ther.*, **14**, 1619–1629.

117. Tyszkiewicz,A.B. and Muir,T.W. (2008) Activation of protein splicing with light in yeast. *Nat. Meth.*, **5**, 303–305.

118. Eyckerman,S., Verhee,A., der Heyden,J.V., Lemmens,I., Ostade,X.V., Vandekerckhove,J. and Tavernier,J. (2001) Design and application of a cytokine-receptor-based interaction trap. *Nat. Cell Biol.*, **3**, 1114–1119.

119. Lemmens,I., Lievens,S., Eyckerman,S. and Tavernier,J. (2006) Reverse MAPPIT detects disruptors of protein-protein interactions in human cells. *Nat. Protocols*, **1**, 92–97.

120. Pelletier,J.N., Campbell-Valois,F. and Michnick,S.W. (1998) Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proc. Natl Acad. Sci. USA*, **95**, 12141–12146.

121. Luker,K.E., Smith,M.C.P., Luker,G.D., Gammon,S.T., Piwnica-Worms,H. and Piwnica-Worms,D. (2004) Kinetics of regulated protein–protein interactions revealed with firefly luciferase complementation imaging in cells and living animals. *Proc. Natl Acad. Sci. USA*, **101**, 12288–12293.

122. Stefan,E., Aquin,S., Berger,N., Landry,C.R., Nyfeler,B., Bouvier,M. and Michnick,S.W. (2007) Quantification of dynamic protein complexes using renilla luciferase fragment complementation applied to protein kinase a activities in vivo. *Proc. Natl Acad. Sci. USA*, **104**, 16916–16921.

123. Hu,C. and Kerppola,T.K. (2003) Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. *Nat. Biotechnol.*, **21**, 539–545.

124. Grinberg,A.V., Hu,C. and Kerppola,T.K. (2004) Visualization of Myc/Max/Mad family dimers and the competition for dimerization in living cells. *Mol. Cell. Biol.*, **24**, 4294–4308.

125. Robida,A.M. and Kerppola,T.K. (2009) Bimolecular fluorescence complementation analysis of inducible protein interactions: Effects of factors affecting protein folding on fluorescent protein fragment association. *J. Mol. Biol.*, **394**, 391–409.

126. Johnsson,N. and Varshavsky,A. (1994) Split ubiquitin as a sensor of protein interactions in vivo. *Proc. Natl Acad. Sci. USA*, **91**, 10340–10344.

127. Stankunas,K. and Crabtree,G.R. (2007) Exploiting protein destruction for constructive use. *Proc. Natl Acad. Sci. USA*, **104**, 11511–11512.

128. Wehr,M.C., Laage,R., Bolz,U., Fischer,T.M., Grünewald,S., Scheek,S., Bach,A., Nave,K. and Rossner,M.J. (2006) Monitoring regulated protein-protein interactions using split TEV. *Nat. Methods*, **3**, 985–993.

129. Williams,D.J., Puhl,H.L. and Ikeda,S.R. (2009) Rapid modification of proteins using a rapamycin-inducible tobacco etch virus protease system. *PLoS ONE*, **4**, e7474.

130. Kentner,D. and Sourjik,V. (2009) Dynamic map of protein interactions in the escherichia coli chemotaxis pathway. *Mol. Syst. Biol.*, **5**, doi: 10.1038/msb.2008.77.

131. Kelly,J., Rubin,A., Davis,J., Ajo-Franklin,C., Cumbers,J., Czar,M., deMora,K., Glieberman,A., Monie,D. and Endy,D. (2009) Measuring the activity of BioBrick promoters using an in vivo reference standard. *J. Biol. Eng.*, **3**, 4.

132. Haustein,E. and Schwille,P. (2007) Fluorescence correlation spectroscopy: novel variations of an established technique. *Annu. Rev. Biophys. Biomol. Struct.*, **36**, 151–169.

133. Cochran,J.R. (2010) Engineered proteins pull double duty. *Sci. Transl. Med.*, **2**, 17ps5.

134. Wilson,G.S. and Hu,Y. (2000) Enzyme-Based biosensors for in vivo measurements. *Chem. Rev.*, **100**, 2693–2704.

135. Smith,R.G., D'Souza,N. and Nicklin,S. (2008) A review of biosensors and biologically-inspired systems for explosives detection. *Analyst*, **133**, 571–584.

136. Walsh,G. (2006) Biopharmaceutical benchmarks 2006. *Nat. Biotechnol.*, **24**, 769–776.

Appendix

Synthetic rewiring of MAPK (mitogen-activated protein kinase) circuits.

Yeast uses two related MAP kinase signaling networks for the mating response to pheromone or the stress response to high osmolarity, respectively (50). Signal transduction for the pheromone response is organized around the scaffold protein Ste5 (see Figure 2A). The high-osmolarity response is mediated by scaffold protein Pbs2. One intermediate kinase, MAPKK Ste11, is shared between the two pathways. Harris *et al.* (39) locked this promiscuous kinase into either of the two specific signaling routes, simply by covalently fusing it to Ste5 or Pbs2. Park *et al.* (40) created a chimera of the two scaffold proteins that rewired the pheromone signal input to the osmolarity response output. Grewal and colleagues (41) later swapped the docking site between a MAPK kinase and its substrate MAPK for an unrelated protein–protein interaction. Several teams then reshaped the response characteristic of MAPK signaling by introducing transcriptional feedback loops; that is they put the expression of regulating proteins under the control of the pathway itself. Ingolia and Murray (51) established a positive feedback leading to bi-stability. Bashor and colleagues (42) extended the Ste5 scaffold with a synthetic leucine zipper interaction interface (Figure 2A). They expressed positive or negative MAPK regulators from pathway-specific promoters and used the new, modular, protein interaction interface to recruit them to the scaffold (42). Different combinations of recruitment strength, binding decoys and positive or negative feedback accelerated or delayed the signal output or switched it from a graded response to a hypersensitive or pulse-generating regime.