FEBS Letters 579 (2005) 1808–1814 FEBS 29346

Minireview

# Synthetic modular systems – reverse engineering of signal transduction

Tony Pawson, Rune Linding*

*Samuel Lunenfeld Research Institute, Mt. Sinai Hospital, 600 University Avenue, Toronto, ON, Canada M5G 1X5*

Accepted 8 February 2005

Available online 16 February 2005

Edited by Robert Russell and Giulio Superti-Furga

**Abstract During the last decades, biology has decomposed cellular systems into genetic, functional and molecular networks. It has become evident that these networks consist of components with specific functions (e.g., proteins and genes). This has generated a considerable amount of knowledge and hypotheses concerning cellular organization. The idea discussed here is to test the extent of this knowledge by reconstructing, or reverse engineering, new synthetic biological systems from known components. We will discuss how integration of computational methods with proteomics and engineering concepts might lead us to a deeper and more abstract understanding of signal transduction systems. Designing and successfully introducing synthetic proteins into cellular pathways would provide us with a powerful research tool with many applications, such as development of biosensors, protein drugs and rewiring of biological pathways.**



*Keywords:* Synthetic biology; Signal transduction; Modular proteins; Rewiring pathways; Dynamics

## 1. Biomolecular interactions and systems biology

Analysing the individual molecular components of the cell, and their immediate interactions, is critical for understanding cellular organization. However, a central tenet of systems biology holds that defining the dynamic behaviour and complex properties of cells requires an exploration of larger biological networks. Here, we will explore how the biochemical mechanisms that underlie physiological regulatory pathways, particularly protein–protein interactions, may be developed into a meaningful system biological analysis. We suggest that by creating artificial biological pathways from known components, we can test our ability to predict biological behaviour, and enhance our understanding of complexity. Such manipulations are also important to learn how disease-causing gene products, such as oncoproteins or pathogenic proteins of microorganisms, affect cellular function. These latter polypeptides commonly re-wire the signaling pathways of the host cell, leading to multifaceted changes in cellular phenotypes. For example, a single aberrant oncogene product, such as the v-Src tyrosine kinase, can elicit alterations in gene expression, cytoskeletal architecture, metabolism, proliferation and migration [1].

Previously, artificial transcriptional networks have been constructed in vivo and in vitro using rational approaches to gene circuit design, thereby creating feed-back [2], stability [3], toggle switch [4] and cell–cell communication devices [5]. Similar methods might be applicable to signaling proteins, since signaling pathways are naturally subject to very complex regulation [6,7].

## 2. Modular nature of protein–protein interactions

The majority of proteins encoded by the human genome have a modular design, in the sense that they contain multiple folded domains, many of which mediate specific protein–protein interactions. Interaction domains are typically structured such that their N- and C-termini are juxtaposed in space, and as a consequence they can potentially be inserted into a loop on an existing protein, while leaving their ligand-binding surface exposed. This modular design may have facilitated the evolution of new biological functions, through the juxtaposition of interaction or catalytic domains in novel combinations. As an example, protein-tyrosine kinases commonly exert their effects by phosphorylating sites that are subsequently recognized by the SH2 domains of downstream targets. With the emergence of phosphotyrosine (pTyr) signals to promote communication between cells and multicellularity, incorporation of an SH2 domain into an existing protein could have provided an immediate physical connection with an activated tyrosine kinase [8,9]. The signature of such domain insertions can be seen in signaling proteins such as phospholipase Cγ, where two adjacent SH2 domains and a SH3 domain (which recognizes proline-rich sequences), are inserted as a unit into a PH domain (which binds phosphoinositides) [10].

A further important feature of interaction domains prevalent in metazoan species is their remarkable flexibility. Any one domain may be present in hundreds of copies in the human proteome, and different members of a particular domain family can show quite distinct binding properties. SH3 domains, for example, typically bind proline-rich motifs that adopt a polyproline type II helix, even prior to SH3 recognition; some SH3 domains, in contrast engage basic sequences that undergo an order–disorder transition upon binding to form a $3_{10}$ helix [11]. In a similar fashion, PH domains, which commonly bind phosphoinositides at the plasma membrane, have the same structural fold as PTB domains (which recognize phosphorylated Asn-Pro-X-Tyr motifs), EVH1 domains

*Corresponding author. Fax: +1 416 586 8869.
*E-mail addresses:* pawson@mshri.on.ca (T. Pawson), linding@mshri.on.ca (R. Linding).

0014-5793/$30.00 
doi:10.1016/j.febslet.2005.02.013

(that associate with proline-rich sequences), a subunit of FERM domains (involved in integrin and cytoskeletal signaling), GRAM domains (components of Myotubularin phosphoinositide phosphatases), the N-terminal region of the p62 subunit of the TFIIH transcription factor (which binds a 3′ endonuclease), and a subunit of BEACH domains [12–14]. In this case, the PH fold appears to be a versatile scaffold that has evolved distinct ligand-binding surfaces and biological functions.

As noted above, modular interaction domains frequently bind short peptide motifs in their targets, and this recognition can depend on post-translational modification of the ligand, such as phosphorylation on tyrosine or serine/threonine, acetylation or methylation of lysine residues, or prolyl hydroxylation [15]. This allows protein interactions to form and dissolve in response to external signals, such as growth factor stimulation, or internal cues such as DNA damage [16,17]. We suggest that our understanding of the modular architecture and dynamic binding properties of regulatory proteins and their constituent interaction domains has now advanced to a stage where these domains can be used as building blocks to develop new synthetic proteins and modular networks.

## 3. Identification and prediction of interaction motifs

A recent analysis has suggested that around 10 000 types of structurally different protein interactions can be anticipated [18]. Many of these interactions will likely fall into one of two major classes, involving either domain–domain interactions with relatively large surface interfaces between globular folded domains, or interactions between globular domains and short peptide motifs. Understanding the molecular properties of these different types of interactions can provide tremendous descriptive and predictive power.

Interactions in which a folded domain binds a short peptide sequence can be of high affinity (low nM), and consequently stable and longlived, exemplified by interactions between the regulatory subunits of cAMP-dependent protein kinase (PKA) and their binding motifs on Akinase anchoring proteins [19]. This particular protein–protein interaction functions to hold PKA in an inactive state close to specific targets, in readiness for a local increase in the concentration of cAMP [20]. However, most domain–peptide interactions are much weaker (in the micromolar range) [21], and frequently dependent on post-translational modification of the peptide ligand, or conformational change to expose the binding surface of the folded interaction domain. Consequently, such interactions tend to be transient, and are well suited to the dynamic regulation of signalling pathways and regulatory networks.

Domain–peptide interactions have been intensively investigated in the context of individual signalling pathways and comprehensively analysed for their binding properties using degenerate peptide libraries [22]. Such complexes also lend themselves to future chip-based approaches. They have also been explored in high throughput (HTP) experiments, for example, by combined yeast 2-hybrid and phage display analysis of interactions mediated by yeast SH3 domains [23,24]. However, their transient and conditional nature, which is the very thing that makes them biologically important, can also render them difficult to study. They tend to be under-represented in the large-scale affinity purification assays reported so far [25,26], and may be missed by yeast two-hybrid methods, for example if they require a modification such as tyrosine phosphorylation which is unlikely to occur in yeast. Thus, most interaction networks derived from HTP data are dominated by the more stable domain–domain interactions, and may therefore lack the biochemical and biological complexity imparted by these more evanescent interactions, which likely confer dynamic responsiveness to cellular networks. A recent advance in the exploration of mammalian signalling networks makes use of protein pairs tagged, respectively, with a Flag epitope and luciferase, and expressed in human cells. This technique (termed LUMIER, [62]) has been automated, and thus many thousands of potential interactions can be interrogated in a physiological setting. Equally important, the effect on intracellular interaction networks of stimulating cells with growth factors can be rapidly tested.

Transient networks may also be under-studied because they are difficult to handle computationally [27–30], since short peptide sequences are statistically insignificant and are often located in disordered or unstructured regions of the host protein [31]. A prime example of this is the C-terminal tail of the p53 oncoprotein, which is disordered in solution, but which also harbours tens of post-translational modification sites and peptide motifs. Since disordered segments in globular proteins often cause difficulties during expression, purification and crystallization of a protein, they are frequently experimentally removed, and are therefore lost for subsequent analysis [32,33].

In higher Eukaryotes 60–80% of the proteome consists of multidomain proteins [34,35]; these can be viewed as being built of modular [1] globular domains, connected as beads on a string by non-globular, often unstructured or disordered, linkers. Regions biased towards a subset of amino acid residues (termed low-complexity) often contains hyper-clusters of linear motifs, the arch example being proline rich sequences which are bound by a large group of modular protein domains (e.g., SH3, WW, EVH1, GYF). These non-globular or disordered segments therefore contain a large family of modular linear motifs, potentially involved in protein interactions.

The comprehensive and sensitive identification of functionally relevant interaction motifs has been difficult, primarily because linear motifs are typically very short (4–8 residues). An improvement in predicting physiological sites has employed algorithms that predict the structural context for linear motifs, and thus their availability to bind an interaction domain. These new tools for improved prediction of non-globular and disordered regions in proteins [31,32,36,37] will be helpful for predicting linear interaction motifs, and also aid in the design of protein expression vectors for biophysical studies [32]. These computational tools can also be applied to the exploration of intrinsically disordered proteins such as Tau, Prions, Bcl-2 and p53 [33]. Not only are these proteins hot-spots for linear motifs, but they are also important for the study of protein folding and diseases, relating to misfolding and aggregation of proteins (such as Alzheimers', Parkinsons' and BSE) [38].

[1] Note: In structural bioinformatics the term "module" is used differently from genetic networks. A module here is a basic structural, functional and evolutionary unit of a polypeptide. Often domains correspond to folding units. Linear modules are peptide stretches whereas domains are larger and contain more structural information.

## 4. Regulated protein–protein interactions

Transient or conditional interactions can potentially lead to the oriented flow of information in signal transduction pathways, and may connect distinct sub-networks in response to specific cues. As noted above, modular protein interactions can be directly induced by modification of the linear peptide ligand. In addition, signalling proteins frequently adopt an autoinhibited conformation, in which interaction domains engage internal peptide ligands. In the prototypic example of the inactive Src tyrosine kinase, both the SH2 and SH3 domains are bound through intramolecular interactions that suppress their ability to interact with exogenous proteins and also inhibit kinase activity [39]. Activation of the Src tyrosine kinase liberates the SH2 and SH3 domains to recruit targets, and potentially re-wire pTyr-dependent networks [8,9].

## 5. Protein interactions can be used to build larger complexes; switch-like functions

The concepts discussed above suggest that protein interaction domains, while rather simple in isolation, can be used in a combinatorial fashion to generate more complex behaviours, such as co-operativity and switch-like functions. In T cells, stimulation of the antigen receptor induces the formation of a pTyr-dependent network, involving two docking proteins, LAT and SLP-76 [40]. Tyrosine phosphorylation of membrane-associated LAT creates binding sites for the SH2 domains of the adaptor protein Gads, and phospholipase Cγ, and these proteins can also interact through their SH3 domains with SLP-76. Interestingly, the C-terminal SH3 domain of Gads induces an order-disorder transition in SLP-76, and this may have a cooperative effect on PLC-γ binding and activity, and on the specificity with which Gads and PLC-γ interact with phoshphorylated sites on LAT [11]. Phosphorylation-dependent protein interactions may also show digital switch-like behaviour, based on a requirement for multiple phosphorylation sites to enforce binding to an interaction domain [41], refer to Fig. 1. In the case of the yeast Cdk inhibitor Sic1, multi-site phosphorylation during the G1 phase of the cell cycle appears important for Sic1 binding to the WD40 repeat domain of Cdc4, a component of an SCF E3 protein-ubiquitin ligase complex [42]. This interaction is required for Sic1 polyubiqutination and proteosome-mediated degradation, which lifts the inhibition of Cdk activity necessary for the cell to enter S-phase [42]. This phosphorlation-dependent interaction network therefore provides directionality to the cell cycle. Such ultrasensitive behaviour is not limited to protein–protein interactions. Interestingly, the polybasic region of the N-Wasp protein binds in a cooperative fashion to the phospholipid phosphatidylinositol (PI)-4,5-bisphosphate, which may allow for a switch-like binding of N-Wasp to membranes in response to a small increase in $PIP_2$ concentrations [43].

We have previously tested the notion that phosphorylation-dependent interactions may be used to couple distinct sub-networks, by employing a chimeric adaptor protein in which a pTyr recognition domain (SH2 or PTB) is fused to a death effector domain (DED) from the adaptor protein FADD [41]. This chimeric polypeptide binds to activated receptor tyrosine kinases, conditional upon their autophosphorylation, and to Caspase-8, an initiator of the apoptotic pathway. In cells expressing such an adaptor, mitogenic signals can recruit and activate the caspase pathway, thereby inducing cell death. In effect, this artificial adaptor creates a phosphorylation-dependent link between two sub-networks that are normally insulated from one another, leading to a novel cellular behaviour.

An interesting challenge would be to design artificial multi-domain proteins which can be activated to re-wire cellular networks by conformational reorganization [44]. Creating such gating function in chimeric proteins would be useful both to provide exquisite control over their activation, and also to explore the properties and evolution of physiological circuits. The first step towards this goal was recently made by Lim and colleagues [45], who created chimeric proteins focused on a region (VCA) from the N-Wasp polypeptide which binds the Arp2/3 complex and thereby promotes branching actin polymerization. By flanking the VCA region with interaction domains (i.e., SH3, PDZ) and their peptide ligands, they created a working, dual input, synthetic protein switch, refer to Fig. 2 [45]. The regulation of this switch is based on autoinhibitory intramolecular interactions that are alleviated by competing binding events, in the form of exogenous peptides. These data have all been obtained in vitro, and it will be of considerable interest to create an artificial switch of this sort that functions inside cells. One precedent is provided by fusion proteins containing the ligand-binding domain of the estrogen-receptor

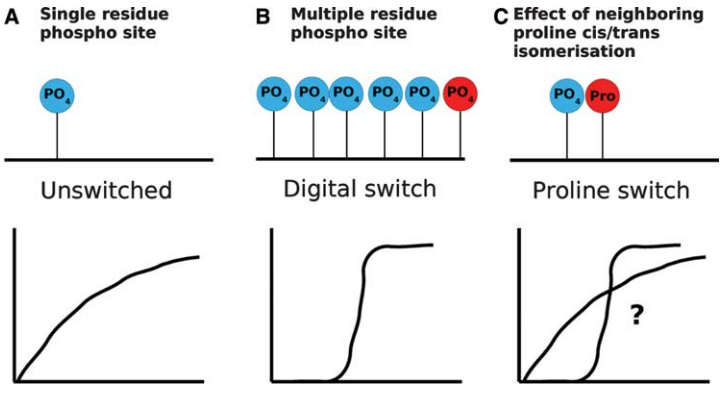


Fig. 1. (A) A single phosphorylation site as seen in many activation loops of kinases. This gives an unswitched response curve. (B) An ultra-sensitive cellular switch (CDK/Sic1 complex) created by a multistep phosphorylation process. (C) There is some evidence [59] that a proline residue in proximity to the phosphorylation site can function as a single residue switch.

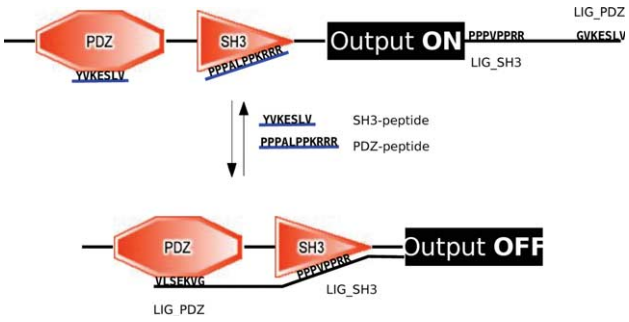


Fig. 2. A dual input switch that behaves as an "AND" device was created using the system shown. The switch function is based on autoinhibitory interactions (PDZ-domain/PDZ-peptide and SH3-peptite) that are alleviated by competing binding events (PDZ-domain/LIG-PDZ and SH3-domain/LIG-SH3).

(ER). In the absence of a steroid ligand, the ER associates with heat shock proteins that can block the activity of a covalently linked domain. Added estradiol or tamoxifen binds the ER, causes release of heat shock proteins, and activation of the polypeptide fused to the ER [46].

## 6. Emergence in protein function

We have previously discussed the possibility that the juxtaposition of interaction domains and motifs in novel combinations may contribute to the evolution of new biological functions. The addition or removal of interaction modules can also be achieved in real time through alternative splicing, which together with post-translational modifications can greatly increase the numbers of protein isoforms in any cell with different binding properties, and potentially with different effects on regulatory networks. It remains unclear how the addition/elimination of domains or motifs might modify the overall complexity and function of a biological system, or how this might give rise to emergence in a cellular system. Informatics studies have shown that a general functional classification of a protein sequence can be performed by looking at its set of "features", which includes post-translational modifications and other derived parameters [47]. However, complexity appears to arise partly as a result of emergent properties of a system, but it is not clear whether our current inventory of modules is sufficiently well defined to describe this in quantitative terms. In this context, building new synthetic systems with chimeric proteins, using the currently known set of signalling components and interaction modules, will allow us to monitor their ability to perturb complexity. Standard network parameters, for instance connectivity and average cluster size, can be measured experimentally.

## 7. Specificity and cross-talk in signalling pathways

Transient interactions between proteins play an important role in preventing aberrant interactions between pathways in normal cells, and in stimulating cross-talk when this is physiologically desirable. Since these interactions frequently involve the recognition of peptide motifs by modular interaction domains, it is critical to understand how short linear peptides can confer specificity towards their cognate domain partners, especially in cases where their binding affinity is modest. In any one cell, many different members of a particular domain family are likely expressed (yeast for example have 28 SH3 domains), which might then compete for related ligands. Several mechanisms likely act in combination to build specificity. One is expression and subcellular localization – only when two proteins are co-expressed and co-localized will an interaction take place. This allows for combinatorial effects, since one domain may localize a protein, for example to a specific membrane site, thereby directing a second domain to specifically recognize a binding partner. Second, selectivity is driven by both permissive and inhibitory forces, and a steric restriction on binding non-physiological binding partners may therefore be just as important as an ability to engage the appropriate target. Thus, a linear proline-rich motif in the yeast Pbs2 protein, a MAP kinase scaffold, engages the SH3 domain of the Sho1 osmosensor, but does not undergo aberrant crosstalk with other yeast SH3 domains. However, the Pbs2 motif does interact promiscuously with SH3 domains from non-yeast species [48]. This has led to the suggestion that an interaction motif need only discriminate between a set of domains that it meets in the context of the cell in which it is expressed, or the subcellular compartment in which it is localized [48]. Although some interaction domains show a very specific interaction with a single target, this usually applies only to proteins with highly specialized functions in one or a few cell types. In a biological setting, most interaction domains likely have numerous partners, which may differ according to the cell type, or within a single cell depending on the environmental conditions. Thus, it may be dangerous to assign an interaction domain to a unique function in a cellular network, when its connectivity is likely undergoing constant flux.

## 8. Topologies and architectures of modular interaction networks

Deriving the topology and dynamics of protein–protein interaction networks in cellular signaling systems under different conditions, such as in the DNA damage response or following the expression of a disease gene product, will be important for understanding robustness, complexity and checkpoint mechanisms of biological systems. Orthogonal sets of data (i.e., protein interactions, protein localization, gene expression arrays), from different cellular conditions are proving useful for network analysis [49,50]. To provide deeper functional insight into protein interaction networks, the relevant modules (domains and linear motifs) within a network can be used to decipher its observed interactions (i.e., domain–domain or motif–domain), and to predict their stability. Deriving network dynamics by analysis of cell cycle dependent interaction data will be of significant interest, for example.

Recent efforts have been made to derive new molecular details from proteome-wide protein interaction data. Typically, these data are provided in the form "protein A interacts with protein B". By applying recently developed algorithms, it is possible to predict *how* proteins A and B interact. This involves looking for protein features that can mediate an interaction, for example through binding of a linear motif to a globular domain, and can lead to the identification of previously unknown interaction sequences [51]. This is potentially important in

defining the organization of networks based on HTP experimental data.

## 9. Dynamics of complexes and networks

We have argued that transient interactions, such as those based on protein phosphorylation or switch-like conformational changes, may influence the topology/architecture and dynamics of cellular signaling networks.

Predictive and experimental analysis of modular interactions can readily be extended to illuminate the dynamics of cellular signaling networks, by exploring time-dependent networks such as those that control the cell cycle. Both normal cell cycle progression and DNA damage response are regulated in large part by the action of protein kinases, including polypeptides such as ATM, ATR, Chk1 and Chk2. The interconnections of these proteins and their substrates involves a series of interaction modules, such as FHA and BRCT domains and their phosphorylated peptide ligands, but the shape of these pathways and networks is only just emerging [16,52]. An important goal for such analysis will be to see how the networks change throughout the cell cycle by obtaining real-time catalytic measurements on these kinases in response to DNA damage, or during normal cell cycle progression, in order to add a temporal component to the signaling network model [53].

Combining biocomputational predictions with large experimental phosphorylation and interaction datasets [54,55] will allow for the development of a complex temporal model for the cell cycle that cannot be obtained from either approach alone. Recently, a computational study showed for the first time how periodically and constitutively expressed subunits can be found in the context of a temporal cell cycle, thereby revealing new networks [56].

## 10. Synthetic biology – reverse engineering networks

The biological potential of modularity in signaling proteins might be explored by designing new proteins in a modular manner (i.e., not at an atomic level). In other words, modules could be joined in an in silico predicted sequence background. This is in contrast with more traditional protein engineering approaches such as ''rational design'', ''de novo'' and ''molecular evolution'', refer to Fig. 3. Assuming proteins are assembled in a modular fashion, it follows that one can apply principles from modular systems engineering. Thus, by redesigning or reverse engineering of known modular proteins, new synthetic, modular and functional polypeptides could be constructed. By attempting this we can fully explore whether modules have served as building blocks in the evolution of protein function, something that has been suggested from bioinformatics and experimental analysis [57], but remains to be substantiated for reverse engineered proteins. Such experiments will address the extent to which biological systems actually conform to this paradigm of modularity. A system with logical, dynamical and programmable behaviour could be assembled from well-defined components/parts (e.g., domains and linear motifs). This approach is different from the classical molecular biology method of creating chimeric proteins by changing an existing system. This would also allow biologists to establish a repertoire of components that could be plugged together in an open-source version of wetlab hacking [58].

An integrated approach of in silico and in vivo design, coupled with directed evolution, might allow us to determine fundamental building blocks of selected interaction networks. Engineered proteins can be expressed and tested for predicted functions, refer to Fig. 3, such as activation of specific signalling pathways using phospho-specific antibodies (for example to MAP kinases and STAT proteins). This could be performed

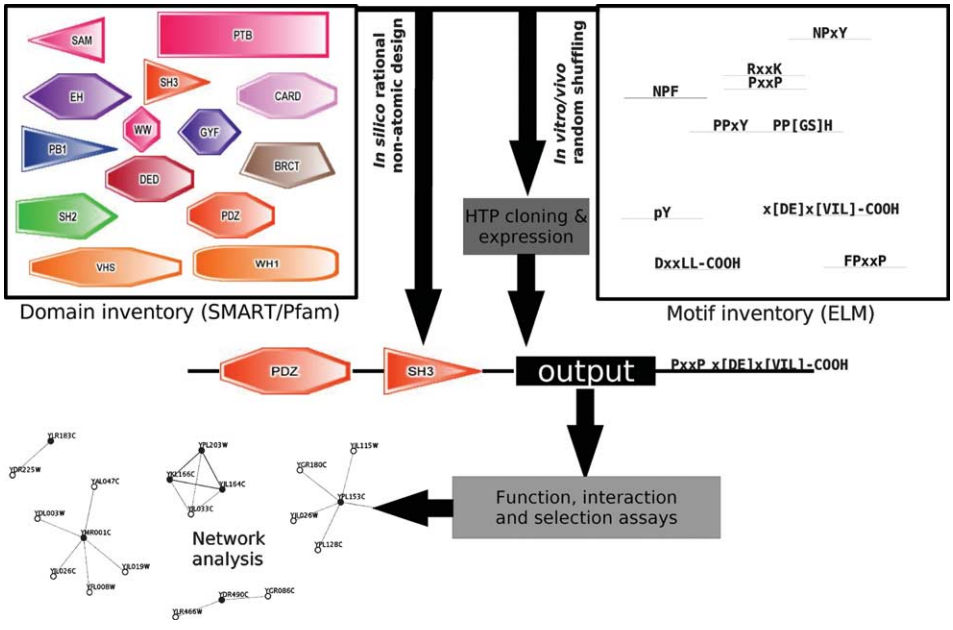


Fig. 3. Modular reverse engineering. Starting from the parts inventories (SMART[60]/pfam[61], ELM[28] and Scansite [30]) one can assemble novel modular proteins. Two strategies should be followed, one based on rational modular (non-automatic) recombination of domains and linear motifs. The second strategy is a directed evolution approach with hight-throughput random synthesis and screening of synthetic proteins. The functional assays can be carried out in vivo but we also envisage in vitro 'network assays' to determine how complexity and connectivity in protein–protein interaction networks are mediated and perturbed by the introduction of modules.

at the single cell level using flow cytometry with phospho-specific antibodies. Another approach would be to look at markers of DNA synthesis/apoptosis. These synthetic proteins could function as control units (gates and switches) that allow for switching and gating in cellular or in vitro systems, thus allowing us to rewire or probe signaling pathways [41].

## References


[1] Brown, M.T. and Cooper, J.A. (1996) Regulation, substrates and functions of src. Biochim. Biophys. Acta 1287 (2-3), 121–149.
[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 (10), 2528–2535.
[3] Becskei, A. and Serrano, L. (2000) Engineering stability in gene networks by autoregulation. Nature 405 (6786), 590–593.
[4] Kobayashi, H., Kaern, M., Araki, M., Chung, K., Gardner, T.S., Cantor, C.R. and Collins, J.J. (2004) Programmable cells: interfacing natural and engineered gene networks. Proc. Natl. Acad. Sci. USA 101 (22), 8414–8419.
[5] You, L., Cox, R.S. 3rd, Weiss, R. and Arnold, F.H. (2004) Programmed population control by cell–cell communication and regulated killing. Nature 428 (6985), 868–871.
[6] Sondermann, H., Soisson, S.M., Boykevisch, S., Yang, S.S., Bar-Sagi, D. and Kuriyan, J. (2004) Structural analysis of autoinhibition in the Ras activator Son of sevenless. Cell 119 (3), 393–405.
[7] Xiong, W. and Ferrell Jr., J.E. (2003) A positive-feedback-based bistable 'memory module' that governs a cell fate decision. Nature 426 (6965), 460–465.
[8] Pawson, T., Gish, G.D. and Nash, P. (2001) SH2 domains, interaction modules and cellular wiring. Trends Cell. Biol. 11 (12), 504–511.
[9] Pawson, T. (2004) Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. Cell 116 (2), 191–203.
[10] Manning, C.M., Mathews, W.R., Fico, L.P. and Thackeray, J.R. (2003) Phospholipase C-gamma contains introns shared by src homology 2 domains in many unrelated proteins. Genetics 164 (2), 433–442.
[11] Liu, Q., Berry, D., Nash, P., Pawson, T., McGlade, C.J. and Li, S.S. (2003) Structural basis for specific binding of the Gads SH3 domain to an RxxK motifcontaining SLP-76 peptide: a novel mode of peptide recognition. Mol. Cell. 11 (2), 471–481.
[12] Blomberg, N., Baraldi, E., Nilges, M. and Saraste, M. (1999) The PH superfold: a structural scaffold for multiple functions. Trends Biochem. Sci. 24 (11), 441–445.
[13] Begley, M.J., Taylor, G.S., Kim, S.A., Veine, D.M., Dixon, J.E. and Stuckey, J.A. (2003) Crystal structure of a phosphoinositide phosphatase, MTMR2: insights into myotubular myopathy and Charcot-Marie-Tooth syndrome. Mol. Cell. 12 (6), 1391–1402.
[14] Gervais, V., Lamour, V., Jawhari, A., Frindel, F., Wasielewski, E., Dubaele, S., Egly, J.M., Thierry, J.C., Kieffer, B. and Poterszman, A. (2004) TFIIH contains a PH domain involved in DNA nucleotide excision repair. Nat. Struct. Mol. Biol. 11 (7), 616–622, Epub 2004 Jun 13.
[15] Pawson, T. and Nash, P. (2003) Assembly of cell regulatory systems through protein interaction domains. Science 300 (5618), 445–452.
[16] Yaffe, M.B. (2002) Phosphotyrosine-binding domains in signal transduction. Nat. Rev. Mol. Cell Biol. 3 (3), 177–186.
[17] Durocher, D., Smerdon, S.J., Yaffe, M.B. and Jackson, S.P. (2000) The FHA domain in DNA repair and checkpoint signaling. Cold Spring Harb. Symp. Quant. Biol. 65, 423–431.
[18] Aloy, P. and Russell, R.B. (2004) Ten thousand interactions for the molecular biologist. Nat. Biotechnol. 22 (10), 1317–1321.
[19] Alto, N.M., Soderling, S.H., Hoshi, N., Langeberg, L.K., Fayos, R., Jennings, P.A. and Scott, J.D. (2003) Bioinformatic design of A-kinase anchoring protein-in silico: a potent and selective peptide antagonist of type II protein kinase A anchoring. Proc .Natl. Acad. Sci. USA 100 (8), 4445–4450.
[20] Wong, W. and Scott, J.D. (2004) AKAP signalling complexes: focal points in space and time. Nat. Rev. Mol. Cell. Biol. 5 (12), 959–970.
[21] Ladbury, J.E., Lemmon, M.A., Zhou, M., Green, J., Botfield, M.C. and Schlessinger, J. (1995) Measurement of the binding of tyrosyl phosphopeptides to SH2 domains: a reappraisal. Proc. Natl. Acad. Sci. USA 92 (8), 3199–3203.
[22] Songyang, Z., Shoelson, S.E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W.G., King, F., Roberts, T., Ratnofsky, S. and Lechleider, R.J., et al. (1993) SH2 domains recognize specific phosphopeptide sequences. Cell 72 (5), 767–778.
[23] Landgraf, C., Panni, S., Montecchi-Palazzi, L., Castagnoli, L., Schneider-Mergener, J., Volkmer-Engert, R. and Cesareni, G. (2004) Protein Interaction Networks by Proteome Peptide Scanning. PLoS Biol 2 (1), E14.
[24] Tong, A.H., Drees, B., Nardelli, G., Bader, G.D., Brannetti, B., Castagnoli, L., Evangelista, M., Ferracuti, S., Nelson, B., Paoluzi, S., Quondam, M., Zucconi, A., Hogue, C.W., Fields, S., Boone, C. and Cesareni, G. (2002) A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. Science 295 (5553), 321–324.
[25] Russell, R.B., Alber, F., Aloy, P., Davis, F.P., Korkin, D., Pichaud, M., Topf, M. and Sali, A. (2004) A structural perspective on protein–protein interactions. Curr. Opin. Struct. Biol. 14 (3), 313–324.
[26] Aloy, P. and Rusell, R.B. (2002) The third dimension for protein interactions and complexes. Trends Biochem. Sci. 27 (12), 633–638.
[27] Puntervoll, P., Linding, R., Gemund, C., Chabanis-Davidson, S., Mattingsdal, M., Cameron, S., Martin, D.M., Ausiello, G., Brannetti, B., Costantini, A., Ferre, F., Maselli, V., Via, A., Cesareni, G., Diella, F., Superti-Furga, G., Wyrwicz, L., Ramu, C., McGuigan, C., Gudavalli, R., Letunic, I., Bork, P., Rychlewski, L., Kuster, B., Helmer-Citterich, M., Hunter, W.N., Aasland, R. and Gibson, T.J. (2003) ELM server: A new resource for investigating short functional sites in modular eukaryotic proteins. Nucleic Acids Res. 31 (13), 3625–3630.
[28] Available from: http://elm.eu.org.
[29] Obenauer, J.C., Cantley, L.C. and Yaffe, M.B. (2003) Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. Nucleic Acids Res. 31 (13), 3635–3641.
[30] Available from: http://scansite.mit.edu.
[31] Linding, R., Russell, R.B., Neduva, V. and Gibson, T.J. (2003) GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res. 31 (13), 3701–3708.
[32] Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J. and Russell, R.B. (2003) Protein disorder prediction: implications for structural proteomics. Structure (Camb) 11 (11), 1453–1459.
[33] Tompa, P. (2002) Intrinsically unstructured proteins. Trends Biochem. Sci. 27 (10), 527–533.
[34] Apic, G., Huber, W. and Teichmann, S.A. (2003) Multi-domain protein families and domain pairs: comparison with known structures and a random model of domain recombination. J. Struct. Funct. Genomics 4 (2-3), 67–78.
[35] Apic, G., Gough, J. and Teichmann, S.A. (2001) An insight into domain combinations. Bioinformatics 17 (Suppl 1), S83–S89.
[36] Available from: http://dis.embl.de.
[37] Available from: http://globplot.embl.de.
[38] Linding, R., Schymkowitz, J., Rousseau, F., Diella, F. and Serrano, L. (2004) A Comparative Study of the Relationship between Protein Structure and ß-Aggregation in Globular and Intrinsically Disordered Proteins. J. Mol. Biol. 342 (1), 345–353.
[39] Sicheri, F., Moarefi, I. and Kuriyan, J. (1997) Crystal structure of the Src family tyrosine kinase Hck. Nature 385 (6617), 602–609.
[40] Jordan, M.S., Singer, A.L. and Koretzky, G.A. (2003) Adaptors as central mediators of signal transduction in immune cells. Nat. Immunol. 4 (2), 110–116.
[41] Howard, P.L., Chia, M.C., Del Rizzo, S., Liu, F.F. and Pawson, T. (2003) Redirecting tyrosine kinase signaling to an apoptotic caspase pathway through chimeric adaptor proteins. Proc. Natl. Acad. Sci. USA 100 (20), 11267–11272.
[42] Nash, P., Tang, X., Orlicky, S., Chen, Q., Gertler, F.B., Mendenhall, M.D., Sicheri, F., Pawson, T. and Tyers, M. (2001) Multisite phosphorylation of a CDK inhibitor sets a

threshold for the onset of DNA replication. Nature 414 (6863), 514–521.
[43] Papayannopoulos, V., Co, C., Prehoda, K.E., Snapper, S., Taunton, J. and Lim, W.A. (2005) A Polybasic Motif Allows N-WASP to Act as a Sensor of PIP(2) Density. Mol. Cell. 17 (2), 181–191.
[44] Lim, W.A. (2002) The modular logic of signaling proteins: building allosteric switches from simple binding domains. Curr. Opin. Struct. Biol. 12 (1), 61–68.
[45] 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 (5641), 1904–1908.
[46] Pritchard, C.A., Samuels, M.L., Bosch, E. and McMahon, M. (1995) Conditionally oncogenic forms of the A-Raf and B-Raf protein kinases display different biological and biochemical properties in NIH 3T3 cells. Mol. Cell. Biol. 15 (11), 6430–6442.
[47] Jensen, L.J., Ussery, D.W. and Brunak, S. (2003) Functionality of system components: conservation of protein function in protein feature space. Genome Res. 13 (11), 2444–2449.
[48] Zarrinpar, A., Park, S.H. and Lim, W.A. (2003) Optimization of specificity in a cellular protein interaction network by negative selection. Nature 426 (6967), 676–680.
[49] von Mering, C., Jensen, L.J., Snel, B., Hooper, S.D., Krupp, M., Foglierini, M., Jouffre, N., Huynen, M.A. and Bork, P. (2005) STRING: known and predicted protein–protein associations, integrated and transferred across organisms. Nucleic Acids Res. 33 (Database Issue), D433–D437.
[50] Wong, S.L., Zhang, L.V., Tong, A.H., Li, Z., Goldberg, D.S., King, O.D., Lesage, G., Vidal, M., Andrews, B., Bussey, H., Boone, C. and Roth, F.P. (2004) Combining biological networks to predict genetic interactions. Proc. Natl. Acad. Sci. USA 101 (44), 15682–15687.
[51] Reiss, D.J. and Schwikowski, B. (2004) Predicting protein–peptide interactions via a network-based motif sampler. Bioinformatics 20 (Suppl 1), I274–I282.
[52] Manke, I.A., Lowery, D.M., Nguyen, A. and Yaffe, M.B. (2003) BRCT repeats as phosphopeptide-binding modules involved in protein targeting. Science 302 (5645), 636–639.
[53] Bettencourt-Dias, M., Giet, R., Sinka, R., Mazumdar, A., Lock, W.G., Balloux, F., Zafiropoulos, P.J., Yamaguchi, S., Winter, S., Carthew, R.W., Cooper, M., Jones, D., Frenz, L. and Glover, D.M. (2004) Genome-wide survey of protein kinases required for cell cycle progression. Nature 432 (7020), 980–987.
[54] Diella, F., Cameron, S., Gemund, C., Linding, R., Brunak, S., Blom, N. and Gibson, T.J. (2004) Phospho. ELM–A new database for exprimentally verified protein phosphorylation sites. BMC Bioinformatics 5 (1), 79.
[55] Available from: http://phospho.elm.eu.org.
[56] de Lichtenberg, U., Jensen, L.J., Brunak, S. and Bork, P. (2005) Dynamic complex formation during the yeast cell cycle. Science 307 (5710), 724–727.
[57] Linding, R., Letunic, I., Gibson, T.J. and Bork, P. (2004) Computational analysis of modular protein architectures. Modular Protein Domains, Wiley-WCH GmbH, London, (Chapeter 21).
[58] Brent, R. (2004) A partnership between biology and engineering. Nat. Biotechnol. 22 (10), 1211–1214.
[59] Brazin, K.N., Mallis, R.J., Fulton, D.B. and Andreotti, A.H. (2002) Regulation of the tyrosine kinase Itk by the peptidyl-prolyl isomerase cyclophilin A. Proc. Natl. Acad. Sci. USA 99 (4), 1899–1904.
[60] Available from: http://smart.embl.de.
[61] Available from: http://www.sanger.ac.uk/Pfam.
[62] Barrios-Rodiles et al. (2005) Science, in press.