Universal strategies in research and drug discovery based on protein-fragment complementation assays

Stephen W. Michnick, Po Hien Ear, Emily N. Manderson, Ingrid Remy and Eduard Stefan

Abstract | Changes in the interactions among proteins that participate in a biochemical pathway can reflect the immediate regulatory responses to intrinsic or extrinsic perturbations of the pathway. Thus, methods that allow for the direct detection of the dynamics of protein–protein interactions can be used to probe the effects of any perturbation on any pathway of interest. Here we describe experimental strategies — based on protein-fragment complementation assays (PCAs) — that can achieve this. PCA-based strategies can be used with or instead of traditional target-based drug discovery strategies to identify novel pathway-component proteins of therapeutic interest, to increase the quantity and quality of information about the actions of potential drugs, and to gain insight into the intricate networks that make up the molecular machinery of living cells.

Perturbation
Any treatment of a cell, including chemicals, gene deletion or ablation with a short-interfering RNA that may produce an observable effect on an assay that reports activity of a pathway.

Module
A group of genes that are shown to interact with each other to produce a common phenotype, or proteins that are functionally linked and have more physical interactions between themselves than with other proteins.

Département de Biochimie, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, Québec H3C 3J7, Canada. Correspondence to S.W.M. e-mail: stephen.michnick@umontreal.ca.
doi:10.1038/nrd2311

Over the past century, the concept of selective chemotherapy has dominated the discovery and development of novel therapeutics[1]. The standard goal is the identification of compounds that are selective for a target within the biochemical pathway that underlies a physiological process or pathology of interest. In practice, however, it is a daunting task to identify small molecules that bind to a therapeutic target, which also do not have unexpected effects when applied to living cells or organisms. The application of genome-wide approaches — such as microarray gene-expression profiling — to cellular pharmacology has highlighted this problem[2–6].

In this Review, we focus on the potential of protein–protein interactions to be universal reporters of cellular processes. Strategies to investigate protein–protein interactions have become a crucial component of efforts to define the gene function, the information flow and organization of biochemical networks and the actions of molecular perturbations on these networks. The protein-fragment complementation assay (PCA) is a method that allows insights beyond the static representations of protein interactions. After illustrating the rationale and key concepts of this technique, we describe how the PCA has allowed for the evolution of applications to targeted therapeutic development and a universal strategy to link a gene to its function on a genome-wide scale. We also describe how protein-interaction data can be combined with other approaches (such as genome-wide gene expression or phenotypic analysis, and genetic manipulations) to define the organization of biochemical networks. Last, we outline the applications of PCAs to the drug discovery process, focusing on their use as sensors for the actions of organic molecules on biochemical networks, as well as their use in model systems.

The rationale behind PCAs

Small-molecule perturbations of gene expression can, in some instances, link small molecules to specific cellular mechanisms of action, and it has been proposed that gene-expression analysis could be used to identify cellular processes that drugs act on[4,5,7–14]. For example, drawing upon systematic genome-wide analyses of *Saccharomyces cerevisiae* gene expression, several groups have suggested that biochemical networks may be organized as modules of physically or genetically linked genes that contribute to a particular cellular function[7,15–21]. Specifically, if the knockout of a gene or set of genes resulted in the same change of pattern of gene expression as that observed when cells are treated with a small molecule, then one might infer that the target(s) of the molecule include one or several of the genes that were knocked out. However, a key limitation of gene-expression profile and drug–gene interaction analyses is that the effects of perturbations of biochemical networks on

gene expression, such as those caused by small molecules binding to specific proteins, may be distant in time and be obscured by extensive cross-talk between pathways. For instance, the inhibition of an enzyme in a biochemical pathway may trigger a whole series of parallel changes in several pathways, ultimately resulting in the initiation of a complex remodelling of gene expression.

To pinpoint the small-molecule-targeted pathway to a specific location within the complex biochemical networks of a living cell, a method to directly probe the pathways that comprise those networks is desirable. As dynamic protein–protein interactions are a universal mechanism of regulated biochemical pathways, a general way to probe the effects of small molecules on individual pathways would be to detect such changes as a result of pathway activation or inhibition. The reasoning is that any perturbation, for example binding of a hormone to a receptor or a small molecule to an enzyme, would propagate through or between biochemical pathways, inducing spatial and temporal changes in protein complexes (protein–protein interactions) downstream of the target of a perturbation (FIG. 1a). These changes could be due to any biochemical perturbation of the pathway including post-translational modification, allosteric transition, protein degradation or *de novo* protein synthesis. Temporal, perturbation-induced changes in the quantity or location of a known protein complex within a pathway and at different steps within pathways would be detected by directly measuring protein–protein interactions in the cells of humans or model organisms.

PCA is a simple method that allows one to link genes to each other and in turn to their function. This is achieved by first, measuring the interactions among the proteins that the genes of interest encode and second, by observing what happens to these interactions when perturbed by a drug, a hormone or by knocking out other genes.

In the PCA strategy protein–protein interactions are measured by fusing each of the proteins of interest to two fragments of a 'reporter' protein that has been rationally dissected into two fragments using protein-engineering strategies[22–25] (BOX 1; FIG. 1b). The reporter-protein fragments are brought into proximity through the association of the two interacting proteins of interest, which allows them to fold together into the unique three-dimensional structure of the reporter protein and reconstitute its activity. As folding is a universal phenomenon of proteins, PCAs can be created with many reporter proteins and thus provide for different types of readouts depending on the desired application (FIG. 1b, right panel). This means that PCAs are a tool-kit from which an appropriate assay can be picked for specific applications. PCAs based on a number of reporter proteins have been developed, including murine dihydrofolate reductase (DHFR)[23,24,26–28], glycinamide ribonucleotide transformylase[25], aminoglycoside kinase[25], hygromycin B kinase[25], TEM1 β-lactamase[29–31], green fluorescent protein (GFP)[32] and colour and behavioural variants[33–37] of these fluorescent proteins, and firefly[38], *Renilla*[39,40] and *Gaussia* luciferases[41] (see TABLE 1 for a comprehensive list of currently developed PCAs). Other assays similar to PCAs use naturally occurring subunits of enzymes, such as β-galactosidase, or indirect reporter activity such as split inteins[42,43].

The unique features of the PCA strategy are summarized in BOX 2. A crucial point is that PCA fragments are designed so that they cannot fold spontaneously, a situation that would result in a false-positive signal and would confound the interpretation of library or systematic screens of protein–protein interactions[23–25,44].

A concern with the PCA strategy is that protein complexes become trapped when the refolded PCA reporter enzyme prevents the dissociation of the complex, and this may confound the interpretation of protein-complex localization or turnover. However, this is generally not observed. Reversibility of PCAs is based on the principles of protein folding (BOX 1; FIG. 2) and the reversibility of two PCAs based on the *Gaussia* and *Renilla* luciferases has been demonstrated[41,45]. It is known that PCAs that are based on GFP and variants are irreversible, which can be useful (for example, for trapping and visualizing rare or transient complexes), but interpretation of the kinetics of complex turnover or localization of interacting proteins must be done with caution because trapping could lead to significant changes in the duration of residence in an individual compartment[33,46,47]. Furthermore, these assays are generally not suitable for studies of dynamics or kinetics of protein–protein interactions in the time frame of seconds or minutes. However, it has been demonstrated that the folding of certain PCA reporters can be rapid enough to detect protein associations that occur in the time-frame of seconds[40,41,45]. In addition, despite their irreversibility, fluorescent-protein PCAs have proved useful to study the relocalization and the prevention or induction of protein–protein interactions by small molecules, which is described below. It is equally possible that attaching unfolded fragments of the PCA reporter protein could perturb the function of the proteins to which they are fused. As with any method that relies on measuring protein function in which that protein is fused to another molecule, appropriate controls must be evaluated. However, for the applications described in this Review, it is important to note that only a simple question is being asked: does a particular protein complex assemble or is its assembly prevented by the actions of a perturbation (for example, a small molecule acting on an enzyme upstream of the complex)?

PCAs for targeted therapeutics

Here we focus on the development and utility of PCAs to identify the function of proteins, and as probes of pathway or network responses to drugs and other perturbations. However, it is worth mentioning that PCAs can also be useful for studying two further classes of therapeutic agents: antibodies or other protein therapeutics, and proteins that activate their targets by an allosteric mechanism.

Among the first applications of the PCA technology was the design of protein–protein interaction interfaces, for example to engineer more specific leucine-zipper

Cross-talk
Interactions between two pathways that are thought to mediate common or different cellular processes.

Allosteric
A change in the spatial orientation of subunits within a protein complex that is shown to regulate the function of the individual subunits. For example, a change in the affinity of haemoglobin subunits for molecular oxygen on changing from one configuration to another.

Reversibility
In the context of protein-fragment complementation assays, the complete unfolding and separation of protein-reporter fragments when the proteins to which they are attached dissociate.

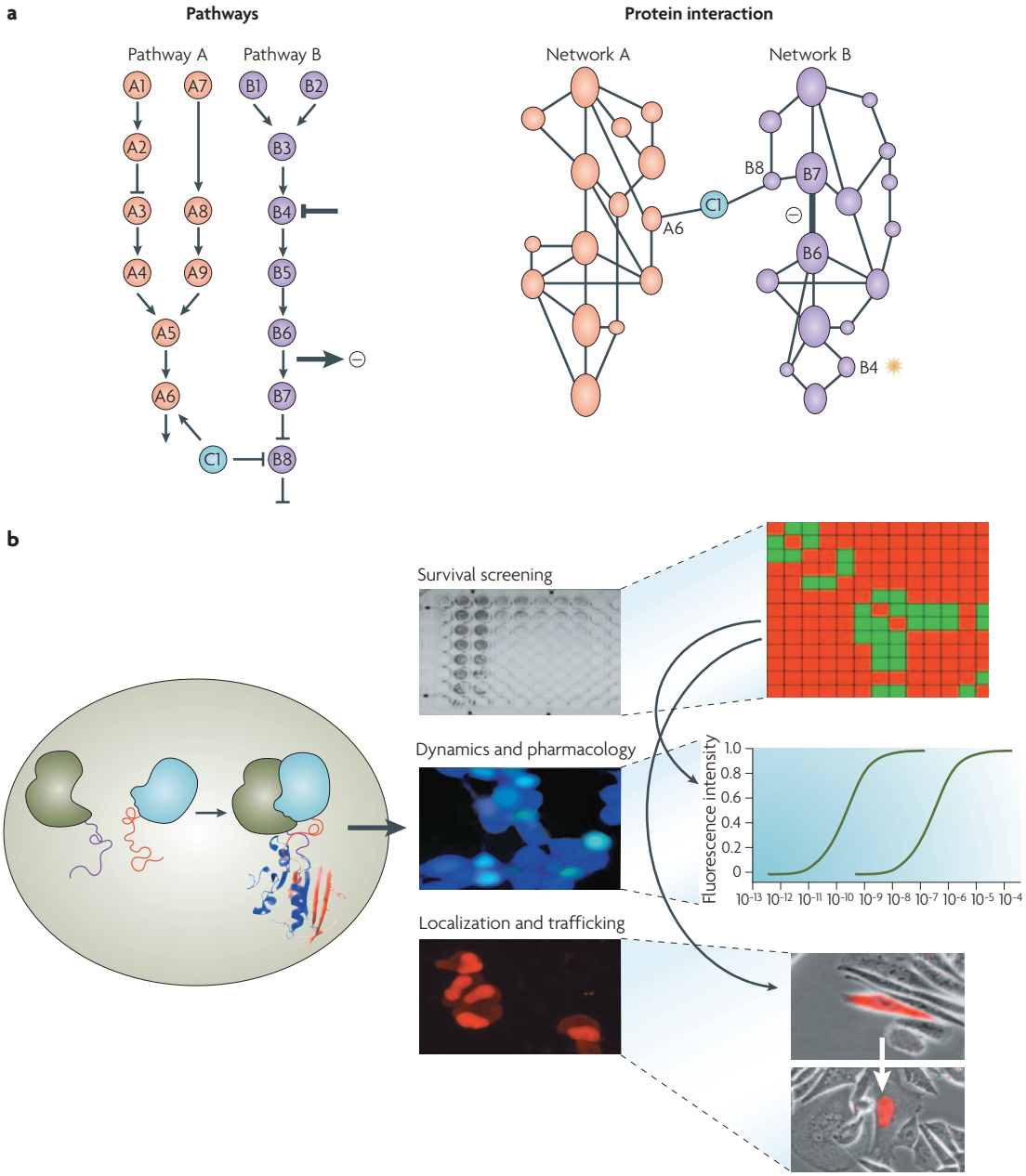

Figure 1 | **Linking genes to function. a** | Two pathways and their component enzymes (for example, A1, B7) are represented on the left, and the interactions between the pathway enzymes are depicted on the right. Perturbation of one enzyme (B4) with, for example, a drug, changes the interactions of other enzymes coupled to B4 (B6:B7). This way it can be inferred that these enzymes have a role in the same biochemical pathway. Equally an enzyme of unknown function (C1) can be identified to play a role in a particular pathway by demonstrating that it interacts with components of the pathways (A6 and B8), and by showing that these interactions are perturbed by agents that affect the pathways. **b** | The left panel represents the protein-fragment complementation assay (PCA) strategy . Interacting proteins are fused to either of the two complementary fragments of a reporter protein (red and blue). Interaction of the two proteins brings the unfolded reporter-protein fragments into proximity allowing them to fold into their active conformation. In the middle panel, examples of PCA reporter-protein readouts of increasing complexity are shown. These readouts can be simple survival selection (using reporters that allow for growth under selective conditions); the conversion of a substrate to a fluorescent, coloured or luminescent product by an enzyme; or intrinsically fluorescent or fluorophore-binding proteins that allow for the study of changes in the localization of protein complexes. The right panel gives examples of how the results of different types of PCAs can be used to learn about protein–protein interactions in detail. The top panel represents the results of a simple systematic screen for interactions among individual proteins using a simple survival-selection PCA (green = positive interactions; red = no interaction). This could be followed by tests of positive-interaction responses to a cellular perturbation (for example, hormone stimulation of a pathway) using an enzyme-reporter PCA. In some cases the response of a protein complex to a perturbation of a cell could be, for example, the translocation of a complex from the cytosol to nucleus.

Phage-display
A method to identify or design proteins with the specific ability to bind to a molecule by expressing proteins as fusions to filamentous bacterial viruses, testing binding to a molecule arrayed on a solid surface, identifying those phage particles that bind and then replicating them in bacteria.

Box 1 | **General concepts of the PCA**

The crucial feature of the protein-fragment complementation assay (PCA) is that it relies on protein folding rather than on the docking of preformed subunits. Polypeptides have evolved to code for all of the chemical information necessary to fold into a stable, unique three-dimensional structure[84–86]. In limited cases, a polypeptide can fold from complementary *N*-terminal and *C*-terminal peptides as demonstrated in the classic experiments of Richards[87], and Taniuchi and Anfinsen[88]. In practice this is more complicated; the major driving force for protein folding is the hydrophobic effect, but this is also the driving force for nonspecific aggregation. However, the principle of the PCA is that if one fuses interacting proteins to the fragments that, by interacting, increase the effective concentration of the fragments, correct folding could be favoured over any other non-productive process[23,24,79] (FIG. 2a,b). One can think of the PCA as a way to recreate the conditions under which a protein folds as an intact full-length polypeptide.

The binary and highly cooperative folding event of the reporter protein in a PCA provides for a specific measure of protein interactions that are dependent on not just proximity, but the absolute requirement that the peptides must be organized precisely in space to allow for folding of the protein from the polypeptide chain. The PCA strategy benefits from this all-or-none nature of protein folding: if two proteins to which the PCA fragments are fused interact, folding occurs and the activity of the folded reporter protein is detected. Otherwise, no signal is detected. This means that the generation of a signal by complementation has an enormous dynamic range of response: no interaction= no signal; interaction= maximum signal proportional to the number of complexes, which is unlike fluorescence or bioluminescence resonance energy transfer techniques[89] (FIG. 2a).

Folding, PCA and paradoxical reversibility

Although intuition might suggest that the folding of a PCA reporter protein would create a kinetic trap of the complex, this is not necessarily true. A key criterion of PCA design is that these assays should be completely reversible so that accurate rates of association and dissociation can be measured. We propose two models to explain how PCA could be reversible (FIG. 2c).

Before considering the models, one must first consider the possible states of the reporter-protein structure. PCA is a process of folding of the reporter protein from fragments that are unfolded before being brought together by the interacting proteins to which they are fused. Generally we think of folded proteins as being stable, and indeed for intact proteins this is largely the case. However, this may not be true of PCA reporter proteins. FIGURE 2c includes the possible equilibrium distribution states of the PCA in a complete cycle from unassociated with unfolded fragment (state 1) to associated but unfolded (state 2) to associated-folded (state 3) and dissociated-folded (state 4). We assume that in the starting state (state 1) the fragments are unfolded and remain so until they are brought together by the association of the two proteins (state 2). Direct evidence of such behaviour was demonstrated for the dihydrofolate reductase (DHFR) PCA, in which the individual fragments were shown to be unfolded and did not fold when the complementary fragments were mixed together[90]. State 4 is assumed to be highly unlikely, as it would require the spontaneous folding of fragments.

In the first model of reversibility, the reconstituted reporter protein is not stable, but is in equilibrium between the folded and unfolded state. In this scenario, the reporter protein may be stable because of a remaining free-energy barrier to unfolding; however it is reversible if the complex is disrupted when the barrier is small. In a second model, the free-energy landscape changes when the protein complex is disrupted, resulting in a rapid and cooperative unfolding. In this scenario we might expect that the kinetics of dissociation of the protein complex would be unaffected by the PCA, provided that the PCA reporter enzyme unfolds faster than the complex dissociates. We have recently shown that the latter scenario is the most likely for two PCAs based on the *Gaussia* and *Renilla* luciferases[41,45].

transcription factors. The success of these efforts suggested the general application of PCAs to screening libraries of proteins for those that specifically interact with a therapeutic target, or for small molecules that disrupt a target interaction. For example, a PCA strategy based on the DHFR survival assay in *Escherichia coli* was used to screen two libraries of complementary-designed leucine-zipper-forming sequences with 1×10^{10} potential interacting pairs of proteins, of which roughly 1×10^{6} could be tested. The features in the resulting leucine-zipper pairs were consistent with the selection for more stable pairs than before competition, with a large increase in heterospecificity versus homospecificity[44,48]. This landmark study suggests that PCAs can be used as an engineering tool for designing specific protein molecules that bind only to their target of interest. In a recent study, the same strategy was used to explore how varied a protein sequence can be and still retain the same structure and ability to bind to a target protein[49–51].

PCAs have also shown promise as a method to screen for specific diagnostic or therapeutic antibodies or alternative binding proteins. For instance, PCA screening has been used recently to select novel scaffold proteins called designed ankyrin-repeat proteins (DARPins) that bind to specific mitogen-activated protein (MAP) kinase homologues[52]. The DHFR PCA was used in combination with the ribosome-display selection strategy to engineer a DARPin that binds to c-Jun N-terminal kinase 2 (JNK2) in the nanomolar range, and not its homologue JNK1, which shares 80% sequence identity. Disrupting specific protein–protein interactions in an experimental setting has advantages over deleting the protein of interest as it allows for a better dissection of the various functions that a protein can play in the cell. In addition, the DHFR PCA was used to engineer specific, well expressed and stable antibodies against capsid protein D (gpD) of the bacteriophage lambda and JNK2 (REF. 53). The scFv (single-chain variable-fragment) antibody fragments that were selected by PCA screening are soluble, can be purified with high yield and bind to their specific antigen under *in vivo* and *in vitro* conditions. Library screening by PCA has accelerated the discovery of many novel protein binders, and has facilitated a better understanding of the biological functions of their target proteins.

The studies described above implied that the PCA strategy can rapidly select for optimal properties of interacting sequences along with critical stereospecific and regiospecific requirements for such complexes. The PCA strategy can also select for optimal *in vivo* characteristics, such as solubility and stability to proteolysis. The simplicity of this approach and the specific nature of the information obtained from this design strategy suggest broad utility of the DHFR PCA in protein design and directed evolution of therapeutic targets. PCA is complementary, and in some cases preferable to phage-display strategies, as the entire selection, optimization and stringency tests are carried out *in vivo*.

Another feature of the PCA strategy is that by knowing the three-dimensional structure of the reporter enzyme, it can be accurately predicted how close together the fragments must be to ensure that the enzyme will fold correctly and have a measurable activity. This was exploited to test a structural allosteric model for the activation of the dimeric erythropoietin receptor (EpoR) using the DHFR PCA in mammalian cells[27]. This approach could be extended to study other allosteric transitions in dimeric or multimeric protein interfaces. In the EpoR case, the receptor dimer transmembrane domains were known to be separated by 73 Å, as observed in the crystal structure of unligated EpoR. It was reasoned

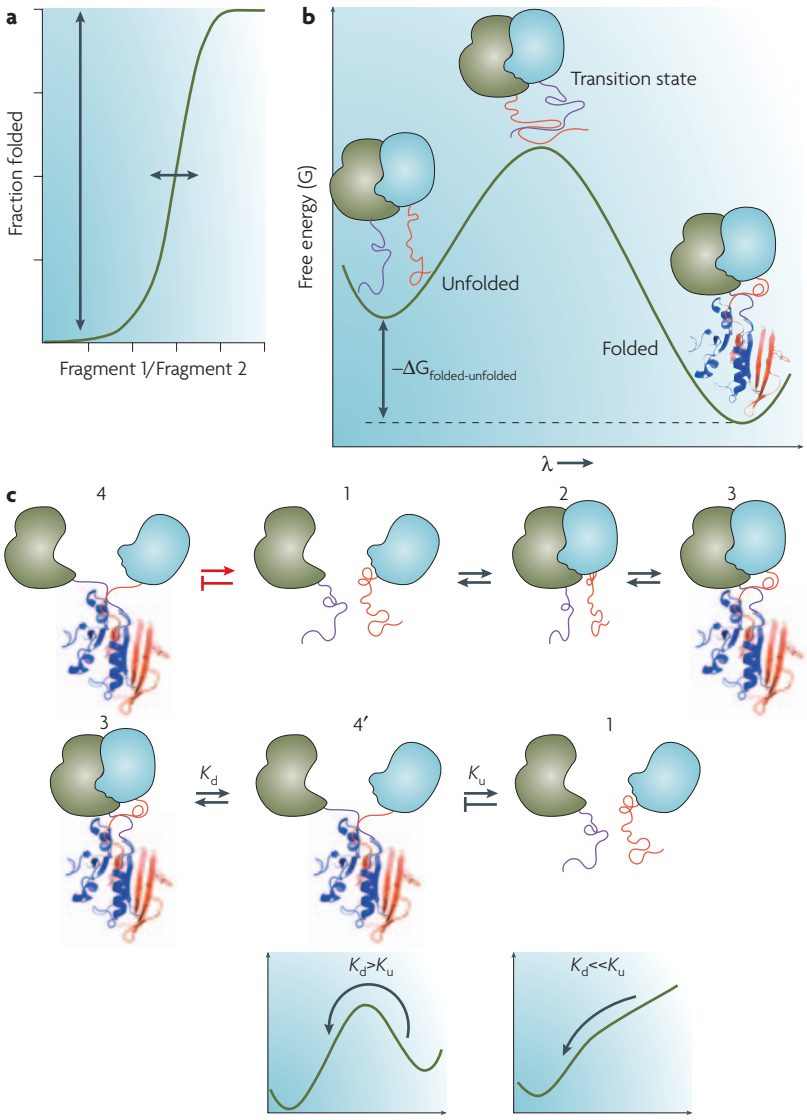

Figure 2 | **Principles of PCA. a** | Protein folding in a protein-fragment complementation assay (PCA) is an all-or-none process. The fraction of PCA reporter-proteins that are folded increases cooperatively as the ratio of the concentration of one fragment to the other increases, resulting in a high dynamic range (vertical arrow) over a narrow range (horizontal arrow) of protein concentration, which results in a switch-like behaviour. **b** | This shows the free energy of folding of the reporter protein in a PCA. The process proceeds because the difference in free energy between the unfolded and folded states ($\Delta G_{folded\text{-}unfolded}$) is negative. **c** | Reversible PCAs allow for the detection of the dissociation of protein complexes. Four states of the PCA-reporter system are depicted. Folding of fragments to the active reporter proteins occurs only if the interacting proteins bring the fragments together in space (1→2→3). Dissociation of a protein complex at a rate K_d will be detected if either the free energy of the transition state (4′) is greater than that of the denatured state (1) and therefore the rate of unfolding K_u is faster than the rate of folding (bottom, left). Alternatively, the free energy landscape is completely altered so that there is no barrier to unfolding (bottom, right), and so unfolding occurs extremely rapidly.

Expression-cloning
The identification of a gene that performs a specific cellular function by expressing a library of cDNAs, and screening for resulting proteins that perform the function.

that if this inactive state existed on the membrane of a living cell, then DHFR fragments fused to the C-termini of the transmembrane domains would fold only if a ligand induced a conformation change that allowed the fragments to come close enough together to ensure that the precise three-dimensional structure of DHFR could be formed. This would require that the *N*-termini of the fragments be 8 Å apart[24,54]. Insertion of flexible linker peptides between the transmembrane domain and DHFR fragments allowed us to probe the distance between the insertion points of the extracellular domain dimer and confirm that linkers long enough to span 73 Å were required for DHFR to fold from its fragments[27]. As this is the defining process of EpoR activation, the use of the EpoR PCA could be particularly useful for screening small molecules that specifically induce this change.

Applications of PCAs to identify gene function

A key challenge in biomedical research is to delineate the function of proteins encoded by genes of interest, particularly those that are linked to diseases or have potential as targets of therapeutic drugs. Such targets are, in most cases, enzymes that if inhibited could alleviate a pathological process. More often they may be individual genes or a locus of genes linked to a disease. This latter case represents the classic problem of mechanistically linking genes to specific biological processes and in the case of drug discovery, determining whether chemical or other targeting of a gene is likely to yield a desired phenotype.

One of the goals in the development of PCA was to devise a way to attribute genes to specific cellular processes without the need for specific screening strategies for each individual case. We call this the generalized expression-cloning problem. In the past, many ingenious strategies have been devised to simultaneously screen cDNA libraries using a protein-specific or enzyme-specific assay that allows for both selection of the clones and validation of their biological relevance[55–59] (FIG. 3a). However, there are many examples of classes of genes for which there is no obvious and specific functional assay that can be combined with cDNA library screening. Particularly difficult are classes of important enzymes such as kinases, phosphatases and proteases that have broad substrate specificity and bind to many proteins or protein domains when studied out of their appropriate context of intact living cells[60,61].

In the absence of simple and specific assays, researchers have turned to strategies that use some of the general functional properties of proteins as readout. The best example of this is the screening for protein–protein interactions between a 'bait' protein of known function against cDNA libraries of 'prey' as exemplified by the yeast two-hybrid strategy[62–68] (FIG. 3b). However, although the genius of the two-hybrid approach is in reducing the problem of linking gene to function through a universal mechanism, the linkage of a bait to prey is tentative and often ambiguous. The question that had a crucial role in the development of the PCA some years ago was how to combine the essential and general observation of protein interactions with assays that would provide some immediate validation of the interaction as biologically relevant.

The generality of two-hybrid and specificity of functional screening can be achieved with a PCA strategy that consists of two steps (FIG. 3c). First, a large-scale screening of bait proteins (X) fused to one PCA reporter fragment against a cDNA library fused to the complementary reporter fragment is performed. In the second step,

Table 1 | **Features and applications of protein-fragment complementation assays (PCAs)**

PCA reporter	Organisms tested	Applications					Detected readout and demonstrated applications
		Identification*	Localization‡	Quantification§	Perturbation‖	Dynamics¶	
Fluorescent proteins and enzymes that bind to fluorescent molecules							
Murine dihydrofolate reductase (DHFR)	Plants, mammalian cells	+	+	+	+	–	• Binding to fluorescein-congugated methotrexate • Localization of protein complexes and pharmacological modulation of protein–protein interaction[26–28,91,92]
Green fluorescent protein (GFP; *Aequoria* jellyfish) and variants including monomeric red fluorescent protein (*Discosoma* coral)	*In vitro*, *Escherichia coli*, plants, *Saccharomyces. cerevisiae*, *Caenorhabditis elegans*, *Drosophila melangaster*, mammalian cells	++	++	+	+	–/+	• Fluorescent protein PCAs[32–36,93,94] • Localization of protein complexes[1,2,35,36,47,76,78,95–98] • Pharmacological modulation of protein–protein interaction and dynamics of subcellular localization[36,70,99,100]
Enzyme reporters with visible products							
TEM1 β-lactamase (*E. coli*)	*In vitro*, mammalian cells	+	–	+	+	–/+	• Fluorescent or coloured products • Pharmacological modulation of protein–protein interactions[29–31]
Firefly, *Gaussia* and *Renilla* luciferases	*In vitro*, mammalian cells, *Musculus musculus*	+	–/+	++	++	++	• Quantitative comparison of bioluminescence intensities[40,41,71] • Pharmacological modulation of protein–protein interactions[38,41,45,73,74,101] • Quantitative dynamics (association/dissociation) of protein complexes[41,45]
Survival selection							
Murine DHFR	*E.coli*, plants, mammalian cells	++	–	–	+	–	• Resistance to trimethoprim or methatrexate, or complements in *Dhfr* knockout[23,24,26,44,53,102]
Hygromycin B phosphotransferase (*Streptomyces hygroscopicus*)	Mammalian cells	+	–	–	+	–	• Resistance to hygromycin B[25]
N-(5′-phosphoribosyl)-anthranilate isomerase (Trp1p) (*S. cerevisiae*)	*S. cerevisiae*	+	–	–	+	–	• Complements in *Trp1* knockout[103]
Glycinamide ribonucleotide (GAR) transformylase (*E. coli*)	*E. coli*	+	–	–	+	–	• Complements in GAR transformylase knockout[25]
Aminoglycoside phosphotransferase (*Streptomyces fradiae*)	*E. coli*, mammalian cells	+	–	–	+	–	• Resistance to neomycin[25,104]
Gene manipulation							
CRE-recombinase (Bacteriophage P1)	Mammalian cells	+	–	–	–	–	• Chemically induced gene knockout[105]
Indirect transcriptional reporter-gene activation#							
Split-ubiquitin system (*S. cerevisiae*)	*S. cerevisiae*	++	–	–	–	–	• Reporter-gene activation[79,80,82,106]
Adenylate cyclase CyaA (*Bordetella pertussis*)	*E. coli*	+	–	–	–	–	• Reporter-gene activation[107]
NIa TEV (tobacco etch virus) protease	Mammalian cells	+	–	–	–	–	• Reporter-gene activation[108]

*Identifying novel protein interactions. ‡Localization of protein complexes. §Quantifying protein complexes. ‖Pharmacological or genetic perturbations. ¶Quantitative temporal/spatial dynamics. #These are examples that involve activation of downstream reporters in response to PCA (indirect readout) . – not applicable; –/+ applicable but limited; + applicable but not fully demonstrated; ++ most fully demonstrated PCA for this application.

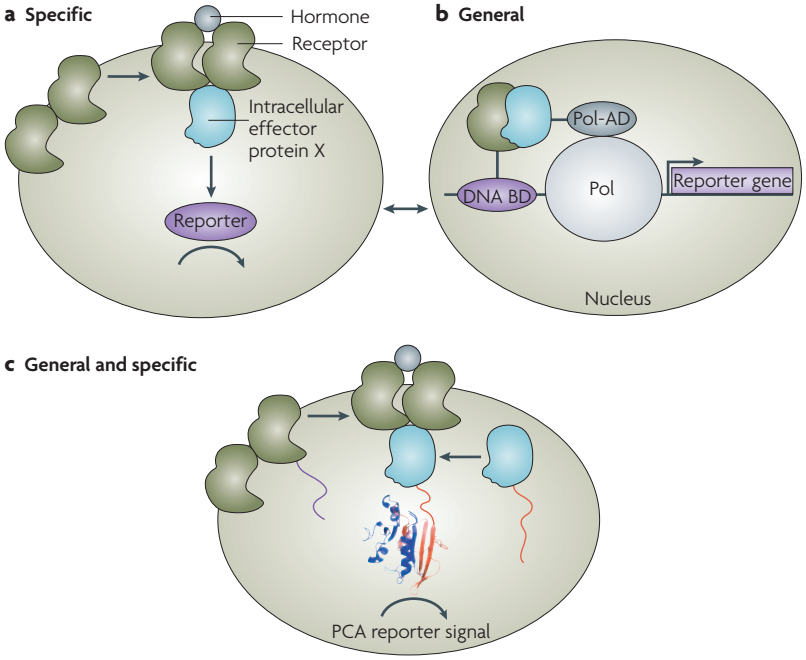

Figure 3 | **Solving the expression-cloning problem. a** | Classic expression cloning. In this hypothetical example, a cDNA library is screened for proteins that could be intracellular effectors (protein X) of a signal-transduction pathway of the known receptor for a hormone. A specific reporter-enzyme assay detects whether a protein can mediate the signal transduction on the binding of the hormone to receptor. **b** | In the absence of a specific reporter assay, one can screen for proteins that interact with the receptor with a two-hybrid screen, but the biological significance of any interaction cannot be evaluated within the screen itself. **c** | Protein-fragment complementation assays (PCAs) provide a specific measurement. For example, the hormone-induced interaction of protein X with the receptor provides both generality (the reporter system) and specificity (induction of the interaction of X with Y on binding of hormone to receptor). DNA BD, DNA-binding domain of two-hybrid gene reporter transcription factor; Pol, RNA polymerase; Pol-AD, RNA polymerase-activation domain of two-hybrid gene reporter transcription factor.

positive hits from the screen are then directly functionally validated by testing for perturbations of the interaction, as measured by the PCA, using agents that act on the biochemical network in which the bait protein is known to participate. For example, several novel substrates and regulators of the serine/threonine protein kinase PKB/Akt were identified in a cDNA library screening strategy using fluorescent-protein-based PCAs[34–36,69]. We were able to link one of these regulators to a specific pathway that controls the proliferation of $CD4^+$ T-cells, and more generally to the regulation of PKB/Akt anti-apoptotic function in both development and disease processes[35]. In another more systematic screen of interactions, an essential regulatory link between growth factor-mediated and transforming growth factor-β (TGFβ)-mediated signal transduction pathways was discovered[35,36].

Sentinel
An assay that provides a direct readout for activity in a specific biochemical pathway.

Box 2 | **Advantages of PCA-based strategies**

- Molecular interactions are detected directly.
- Proteins are expressed in the relevant cellular context, reflecting native post-translational modifications.
- Perturbation of interactions induced by hormones, drugs, gene knockouts or environmental conditions can be detected, providing target validation by linking specific interactions to specific biological processes.
- Subcellular localizations and translocations of protein complexes can be determined.
- Different protein-fragment complementation assays (PCAs) that are based on different reporters allow for multiple modes of detection of protein complexes.
- PCAs are inexpensive and require no specialized reagents.
- PCAs can be automated for high-throughput screening.
- PCAs are designed at the level of the atomic structure of the enzymes used, and because of this there is additional flexibility in designing the probe fragments to control the sensitivity and stringency of the assay.

PCA applications in drug discovery

Instead of seeking the unknown in a pair of protein interactions as described above, PCAs can also be used to examine the consequences of a perturbation on a specific pathway. FIGURE 4 illustrates three examples of how chemical or potential drug actions can be mapped to specific processes. These consist of formulating hypotheses from among three starting points to map a compound's action on a specific pathway. These strategies provide chemical genetics and therapeutic discovery with a focus on network-based targeting. In this view, the target of a molecule is not another molecule but a pre-defined pathway or series of pathways for which PCA sentinels at distinct pathway positions are created. For example, in a signal-transduction pathway, PCAs can be designed to report on an initial docking event, such as the association of a signalling effector protein with a receptor that is activated by a hormone; an intermediary interaction that reports on a secondary event; and last a reporter for the ultimate response such as the activation of a specific transcription factor. Thus, the set of PCA sentinels can report on different steps of a specific pathway, and capture the effects of perturbing any component protein of that pathway or any other pathway that acts on it.

The feasibility of the approach described above was demonstrated for a simple set of well-known signal-transduction pathways several years ago[28]. Three general observations resulted from these studies. First, it was shown that pathway organization could be inferred from patterns of changes in protein complexes under specific perturbations. Second, novel pathway linkages could be inferred that had not been obvious before, and third, the importance of considering both quantitative (changes in signal) and spatial-subcellular (changes in locations) effects on cellular complexes as a means to evaluate effects of drugs was established. This high content aspect to the quantitation of pathway responses illustrates an important point about PCAs as a dynamic reporter strategy. As some PCAs can unambiguously capture cellular locations of protein complexes, it is possible to simultaneously study distinct functions of the same protein in different subcellular compartments (FIG. 5).

A recent study attempted to encompass a larger set of assays covering several biochemical pathways that were either of general pharmacological interest or were involved in cell fate, including assays for cell-cycle checkpoints, apoptosis, mitogenesis, ubiquitin-mediated proteolysis, G-protein–coupled receptors (GPCRs), chaperones, cytoskeletal proteins, stress and inflammation, DNA-damage response and nuclear-hormone-receptor pathways[70]. These pathways were screened against a set

of 107 different drugs from 6 therapeutic areas (cancer, inflammation, cardiovascular disease, diabetes, neurological disorders and infectious disease) and cellular mechanisms (for example, protein transport) with a fluorescent-protein PCA based on the Venus variant of yellow fluorescent protein (YFP). As noted above, this PCA has the advantage of capturing changes in the localization of complexes in response to stimuli as well

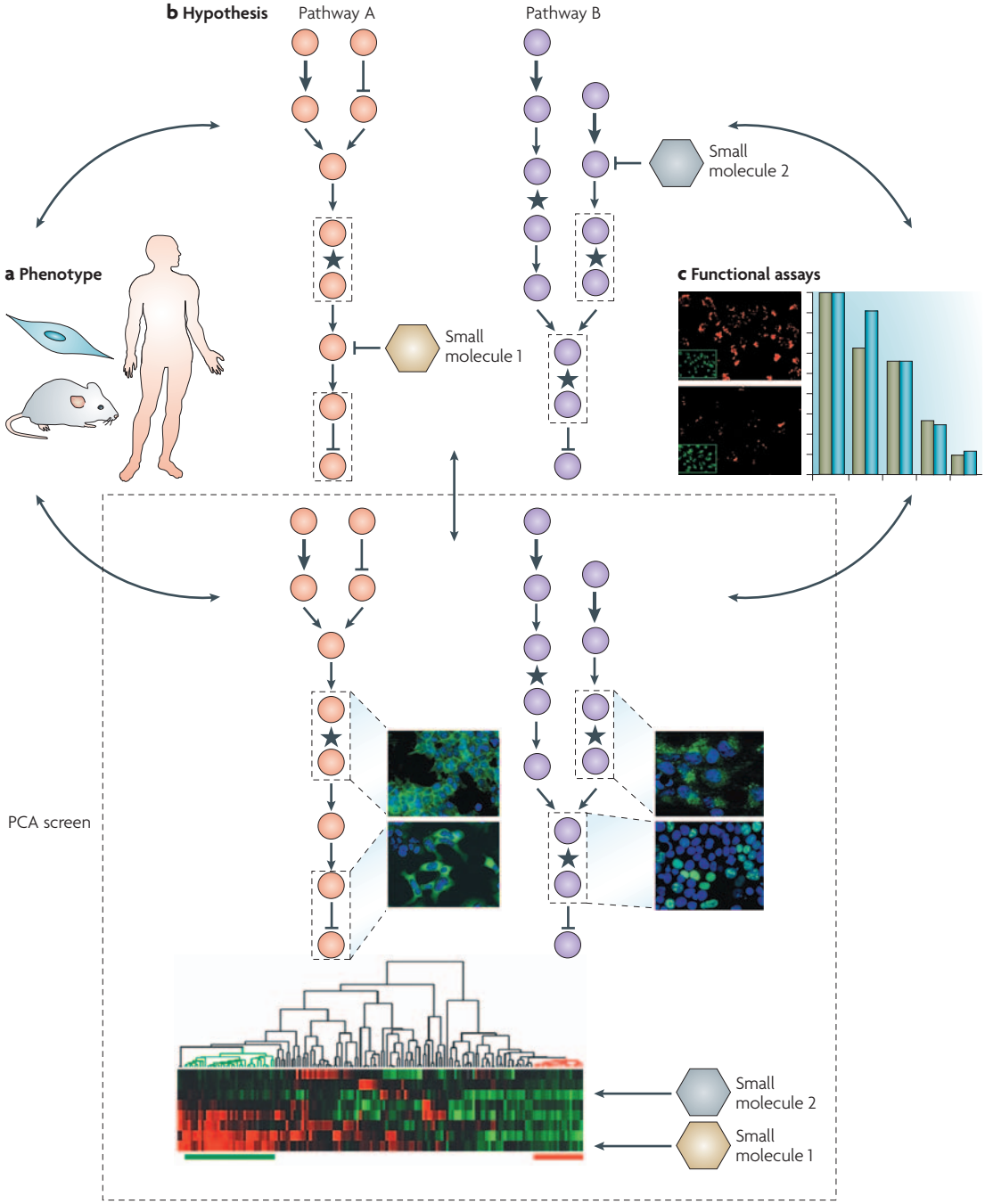

Figure 4 | **Linking small molecules to specific pathways using PCA screens.** Effects of small molecules on cells, model organisms or humans (**a**) could start with a hypothesis that specific pathways and target proteins are affected by the molecules (**b**) followed by testing the compounds with functional assays (**c**) that report on the actions on the targets or pathways of interest. This hypothesis-testing cycle can be augmented or replaced with a protein-fragment complementation assay (PCA) screen that provides sets of sentinels that report changes in protein–protein interactions in specific pathways that may be affected by small molecules. Pathways are represented as proteins (coloured circles) connected by arrows. Stars indicate pairs of interacting proteins that can be detected by a PCA (images) in which the intensity or localization changes of complexes are measured in response to compounds in living cells. Clustering of the pathway sentinel PCA responses can suggest effects on a pathway that is hypothesized to contain a small-molecule target (green cluster) as well as unpredicted effects on other pathways (red cluster).

Off-pathway effect
Effects of a perturbation on pathways that are not predicted to be affected by the perturbation.

High-content imaging
Generally a quantitative analysis of morphological changes or spatial and temporal changes of proteins or protein complexes. Usually performed with fluorescent molecules (for example, antibodies covalently coupled to fluorescent molecules) or fluorescent proteins genetically fused to proteins of interest. These can be imaged by epifluorescence or confocal microscopy.

Hidden phenotypes
A potential outcome to treatment of a cell with some perturbation such as death, differentiation state or morphology that is predicted by pathway sentinel assays, but not necessarily observed in the cells in which the assays were performed.

Hierarchical clustering
A general term for many mathematical strategies used to establish relationships between objects according to common features.

as quantitative changes (FIG. 6a,b). However, as the assays are irreversible their applications are limited to the study of induction or prevention of the formation of protein complexes, and only over time frames longer than the folding and maturation of the fluorophore (>30 minutes). Despite these limitations, the temporal and spatial analysis of the responses of the 49 assays were remarkably informative for predicting basic structure–function relationships among small molecules, on and off-pathway effects and their mechanisms of action, unpredicted and potentially therapeutic actions, and the power of individual or sets of assays to predict the utility of a drug in a specific therapeutic area[70].

An important observation was that temporal high-content imaging analysis was effective at predicting structure–function relationships among molecules. The results were remarkable in that these relationships emerged with so few assays and even for compounds for which there was no reporter for the pathways that these compounds affect. For instance, statins, which act on HMG CoA reductase (3-hydroxy-3-methylglutaryl coenzyme A), an enzyme implicated in lipid synthesis, induced consistent responses in a range of assays, most of which reported on signal-transduction pathways (FIG. 6c). There are two reasons why these results are not surprising. First, it is known that lipids synthesized in the HMG CoA reductase pathway modify proteins in the signalling cascades to which they are linked, including those probed by the assay set. Second, as has been observed in expression profiles for drug treatment, a data-rich approach that can capture the general response of a cell to a chemical or other perturbations will probably reflect some predictable pattern. There are notable differences between these results compared with those obtained from microarray data. First, microarray data provide thousands of readouts, and PCAs cover only a small fraction of cellular responses. Second, although gene expression may be temporally and spatially far removed from the site of action of a drug, PCA reporters can be placed close to or actually include the target and detect responses to pathway pertubations within seconds. We discuss this feature of the PCA strategy below, and also how assay sets can be expanded to more completely cover potential mechanisms of drug action across a broader range of pathways. The advanced state of microarray technology and its accessibility and broad application in cells, tissue and organisms means that it will continue to be the most general approach used for predicting drug actions. Nevertheless, the approaches described here complement expression profiling by providing novel information about mechanisms of action.

In addition to simple structure–function relationships, PCAs have been used to demonstrate hidden phenotypes of drugs, that may be missed by microarray analysis. For example, a compound may induce a pattern of change in pathway sentinel protein interactions but not induce an obvious measurable phenotype in the test cell. Indeed, in addition to observing that sentinel responses can be grouped together according to common drug structures and to targets of the drugs, grouping of drugs based on common phenotypic outcome has been observed[70]. For instance, drugs that prevent cell proliferation tended to perturb assays that report on mechanisms of cell-cycle arrest and apoptosis. These are mechanisms by which cell division would be affected or cell death would be initiated. The obvious question was whether drugs not known to be antiproliferative but grouped with those known to have this outcome would have similar effects. It was not only observed that all drugs in the cluster had similar affects, but four compounds were identified that had similar or greater activities than any other compound in a series of cancer cell lines[70]. Notably, the molecular actions of these compounds were also found to be similar. Of course, antiproliferative activity is a general response and could be mediated by a number of mechanisms. It remains to be seen whether the PCA strategy could afford predictable responses to other therapeutic or toxic effects of drugs or lead compounds.

To fully explore the potential of the PCA strategy, it should be tested against a broader spectrum of compounds and a specific disease model. However, the rich nature of the data afforded by PCAs could allow for a simplification of the task of predicting the site of action of the drug, potential alternative therapeutic applications or toxic effects. There are three basic (and no doubt many more) ways that PCA-screening data can be treated in such a way as to classify the information content of individual or groups of assays (FIG. 7).

The first is to perform hierarchical clustering of temporal PCA-sentinel responses to drugs as described above. In a second strategy, assay responses to a series of compounds are simply ordered by rank from the most negative to the most positive perturbations of all assays in response

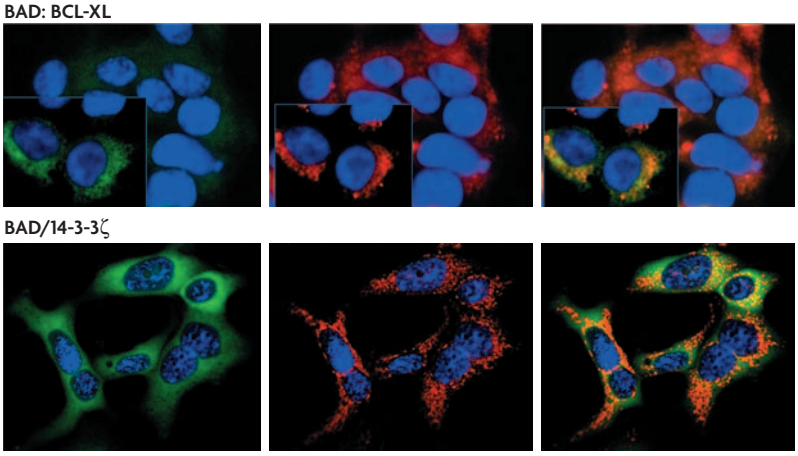

Figure 5 | **Unique subcellular locations of two different functional states of the pro-apoptotic protein BAD revealed by two PCA sentinels.** Depicted are images of cells with fluorescent protein PCA (protein-fragment complementation assay) reporters for complexes BAD (BCL2-antagonist of cell death):BCL-XL in the upper panels and BAD:14-3-3ζ protein in lower panels. In the left panels, PCA signals are green and the nuclei are stained blue. In the middle panels, mitochondria are stained red. The right panels show the overlay of PCA and mitochondrial signals. Yellow–red staining suggests the mitochondrial localization of the BAD:BCL-XL complex in its active form, but cytosolic staining of the BAD:14-3-3ζ complex is consistent with the sequestering of phosphorylated BAD by several 14-3-3 isoforms.

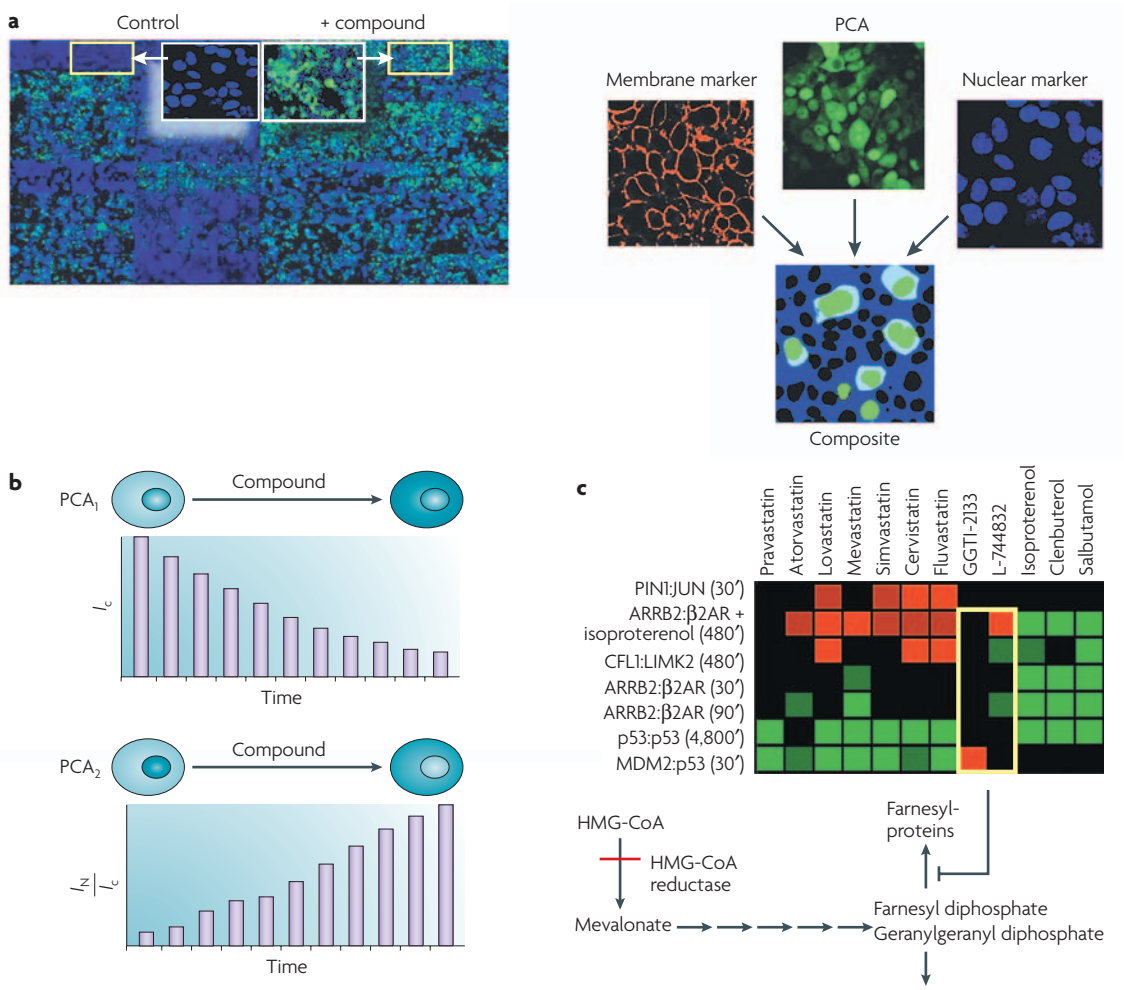

Figure 6 | **Spatial and temporal dynamics of PCAs. a** | Fluorescent protein-fragment complementation assay (PCA) reporters for individual pathways discussed in FIG. 4 are used to capture multiple images of control and compound-treated cells. The PCA reporter signal is captured from specific regions by counterstaining cells with stains for different compartments[70]. **b** | Changes in the intensity or compartmental localization of PCA signals following compound treatment are quantified versus time. For example, the upper panel shows results for a PCA sentinel (PCA_1) that reports a decrease in the quantity of a protein complex in the cytosol following compound treatment as interpreted by intensity of fluorescence (I_c). A second (bottom panel) PCA sentinel (PCA_2) reports a change in the subcellular localization of a protein complex following compound treatment. For the example shown here the relative quantity of the complex located in the nucleus increases after treatment (I_N) compared with the quantity of the complex in the cytosol (I_c). **c** | This is an example of the indirect effects of a clustered drug class (statins) on PCA sentinel reporters of pathways (top), due to indirect inhibition of lipid synthesis. This inhibition results in a decrease in lipid covalent modifications of proteins (farnesylation or geranyl-geranylation), which leads to a change in signalling through the pathways reported on by PCA sentinels. β2AR, β2 adrenergic receptor; ARRB2, arrestin β2; CFL1, cofilin 1 (non-muscle); HMG-CoA, 3-hydroxy-3-methylglutaryl coenzyme A; JUN, jun oncogene; LIMK2, LIM kinase 2; MDM2, mouse double minute 2 homologue; PIN1, protein (peptidylprolyl *cis*/*trans* isomerase) NIMA-interacting 1.

to a single drug. This or more sophisticated approaches allows one to assess how a series of compounds with common enzyme targets (for example, kinases or proteases) affect a number of pathways and, therefore, how specific individual compounds are inhibiting potential targets (FIG. 7, middle panel). This approach can be used to separate drugs that act on a target class in general from those most likely to act specifically on the fewest members of that class. For example, we examined the effects of several protein-kinase inhibitors on our assay panel and the results ranged from drugs having a broad set of significant effects versus others that only affected a small number of assays (unpublished results). Two interpretations are possible. The molecule with a small number of effects may be more specific than those with many potential targets, or those with fewer effects could simply bind more poorly or have lower bioavailability than the more promiscuous drugs. One way or another, the information is valuable for assessing the specificity of a molecule's action while not requiring that an individual assay be created for each and every potential target. Last, assays or combinations of assays are classified according

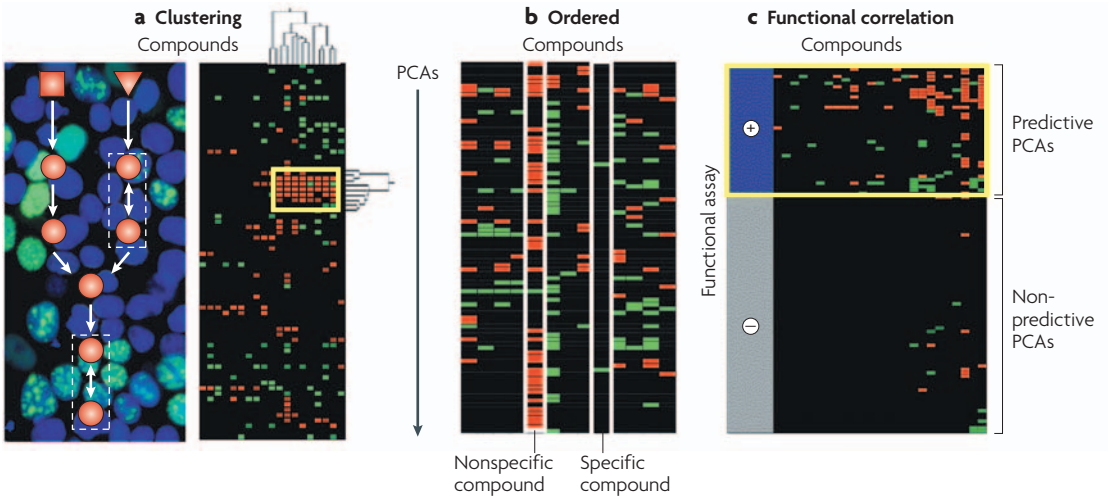

Figure 7 | **Alternative ways to extract information about drug actions on different pathways and cellular processes. a** | Hierarchical clustering of compounds and protein-fragment complementation assay (PCA) responses to compounds can identify groups of compounds that act on a common set of pathways and structure–function relationships among compounds. **b** | PCA responses to compounds that are designed to act on a protein target class (for example, protein kinases) can be used to assess whether an individual compound is specific (PCA responses only for assays that report on a pathway containing the presumed target) or nonspecific (large set of PCA responses). **c** | Grouping of PCA responses according to positive (+) or negative (–) correlation with functional assays that predict an outcome. Individual or sets of PCAs for which responses correlate with functional assays can be predictive of expected effects.

to how well they predict a biological outcome to a drug treatment by ordering the assay responses according to how they correlate with a functional assay (FIG. 7, right panel). For example, we performed this analysis on a panel of 107 drugs for antiproliferative activity and observed that we could obtain more than 90% predictive power for this outcome[70].

PCAs in model cell lines and organisms

PCAs can be used in any cell line or organism in which the reporter assays can be introduced with plasmid or viral vectors, or in organisms in which PCA-reporter fragments could be integrated directly into the genome. Several studies have used PCAs in whole living organisms, which demonstrate the broad potential of this approach, as well as its applicability to target discovery and drug screening. For example, luciferase PCAs have been shown to work in live mice that have been implanted with cells transiently transfected with the fusion-protein constructs. Specific PCAs have also been used to monitor and visualize responses of protein interactions to drugs administered to mice[38,71–74]. These studies suggest the utility of PCAs in screening for small-molecule drugs that either induce or disrupt protein interactions in live animals, as well as to assess how and to what extent drugs can reach their site of action.

The nematode worm (*Caenorhabditis elegans*), the fruitfly (*Drosphila melanogaster*) and yeast (*S. cerevisiae*) are often used for drug discovery experiments, such as large-scale forward genetic screens, because of their cost effectiveness, small size, rapid development and the ability to easily manipulate their genomes[75]. A particular feature of the worm is that its body is transparent, allowing visualization of fluorescent or luminescent reporters. Zhang *et al.* demonstrated the ability to perform PCAs with GFP, YFP and cyan fluorescent protein (CFP) in this model organism to identify cells in which two different promoters are activated, and to identify specific cells in which a particular gene is expressed[76]. The object of interest in these studies was not the protein interactions but an ingenious application of the PCA as a reporter for binary and differential expression of genes in different cell lineages. We can also imagine that panels of PCAs could be introduced into *C. elegans* to link a chemically induced phenotype to potential modes of action, or to evaluate potential pathway targets for drugs or short-interfering RNAs[77]. The feasibility of PCAs in live fruit flies has also been demonstrated in one study that used the YFP PCA to visualize the interactions of the transmembrane odorant receptor OR83b, and to show the unique membrane topology of odorant receptors by taking advantage of the absolute requirement that PCA fragments be oriented in the same compartment[78]. Although the use of PCAs in yeast has so far been limited, we have recently adapted the Venus YFP PCA to yeast in order to visualize and localize protein interactions, as well as the DHFR survival PCA to screen for novel and known protein interactions. At the same time, a growing body of literature exists on an approach from which the generalized PCA strategy evolved, called the ubiquitin split protein sensor[79–82]. Most recently, this approach has been used in *S. cerevisiae* to identify novel protein interactions that are implicated in the function of ATP-binding cassette (ABC) transporters, and the first large-scale analysis of membrane protein–protein interactions in intact living cells[83]. The method is not limited to yeast and further applications to the study of protein interaction networks in other eukaryotes are anticipated.

Perspectives

There are several features of network targeting with PCAs that can be summarized as follows: optimal target(s) for a network of interest can be identified in a simple cell-based assay; the evaluation of potential off-pathway, nonspecific or toxic effects can be done with the same assays, thereby circumventing expensive mid-stage to late-stage clinical failures; the development of PCAs is relatively simple, and their is no need for the production of recombinant proteins or protein purification, which is often an intractable problem for the development of large scale *in vitro*-based assays of drug targets; and PCAs enable the analysis of combinations of drugs, which may affect multiple targets simultaneously. The drug cocktail is becoming particularly relevant in the oncology arena, and recent data have demonstrated the highly synergistic effects of kinase inhibitors, for example, in combination with traditional cytotoxic chemotherapeutic agents.

The biochemical network-based strategy could achieve enormous economies of scale in the drug discovery process. By judiciously creating PCA sentinel probes for specific pathways, one can achieve the monitoring of many other potential protein targets with only a few assays, as the sentinels report what is happening to all the other members of the network to which each sentinel is linked (directly or indirectly). So, for example, a screen for 10 networks with 20 proteins per network would capture the effects of a compound on 200 potential druggable targets. This results in more validated hits from fewer screens in an effort in which prior knowledge of the most druggable step in any specific network is not a prerequisite. Furthermore, because the primary screens are conducted in living cells, small-molecule hits are shown to be membrane permeable and therefore more likely to be bioavailable. Last, as we described above, this approach enables an understanding of the mechanism of action of the compound in the context of the complex networks of the living cell.

The favoured paradigm of modern drug discovery is the target-based, high-throughput screening (HTS) approach. The ability to discover large numbers of drug candidates from HTS has far out-stripped our ability to identify which compounds will be both efficacious and safe at later stages of the discovery process. A rapid exclusion of compounds with potential deleterious effects (the so-called fail-fast strategy) is equally important. The PCA-based approach is unique in its ability to exclude compounds that have undesirable properties by profiling HTS hits against networks that indicate unanticipated or secondary off-pathway effects. By pin-pointing off-pathway effects early on in the discovery process, compounds can be flagged for careful re-evaluation. Thus, PCAs can be used to aid our understanding of the action of a drug in human cells and have the potential to enhance the productivity of therapeutic drug discovery.

1. Drews, J. Drug discovery: a historical perspective. *Science* **287**, 1960–1964 (2000).
2. Chee, M. *et al.* Accessing genetic information with high-density DNA arrays. *Science* **274**, 610–614 (1996).
3. Stoughton, R. B. & Friend, S. H. How molecular profiling could revolutionize drug discovery. *Nature Rev. Drug Discov.* **4**, 345–350 (2005).
An outstanding review on the applications of microarray profiling to predict on- and off-pathway actions of drugs and genetic manipulation. References 4–6 present interesting examples of such applications.
4. Marton, M. J. *et al.* Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nature Med.* **4**, 1293–1301 (1998).
5. Hardwick, J. S., Kuruvilla, F. G., Tong, J. K., Shamji, A. F. & Schreiber, S. L. Rapamycin-modulated transcription defines the subset of nutrient-sensitive signaling pathways directly controlled by the Tor proteins. *Proc. Natl Acad. Sci USA* **96**, 14866–14870 (1999).
6. Hughes, T. R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).
7. Giaever, G. *et al.* Genomic profiling of drug sensitivities via induced haploinsufficiency. *Nature Genet.* **21**, 278–283 (1999).
Together with references 8, 9, 13 and 14, illustrate how chemical–gene interactions in yeast can be used to map small molecules to potential protein targets or pathways.
8. Giaever, G. *et al.* Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. *Proc. Natl Acad. Sci. USA* **101**, 793–798 (2004).
9. Hartwell, L. H., Szankasi, P., Roberts, C. J., Murray, A. W. & Friend, S. H. Integrating genetic approaches into the discovery of anticancer drugs. *Science* **278**, 1064–1068 (1997).
10. Guo, Z., Kumagai, A., Wang, S. X. & Dunphy, W. G. Requirement for Atr in phosphorylation of Chk1 and cell cycle regulation in response to DNA replication blocks and UV-damaged DNA in *Xenopus* egg extracts. *Genes Dev.* **14**, 2745–2756 (2000).
11. Kuruvilla, F. G., Shamji, A. F., Sternson, S. M., Hergenrother, P. J. & Schreiber, S. L. Dissecting glucose signalling with diversity-oriented synthesis and small-molecule microarrays. *Nature* **416**, 653–657 (2002).
12. Bracha-Drori, K. *et al.* Detection of protein–protein interactions in plants using bimolecular fluorescence complementation. *Plant J.* **40**, 419–427 (2004).
13. Parsons, A. B. *et al.* Integration of chemical–genetic and genetic interaction data links bioactive compounds to cellular target pathways. *Nature Biotech.* **22**, 62–69 (2004).
14. Parsons, A. B. *et al.* Exploring the mode-of-action of bioactive compounds by chemical–genetic profiling in yeast. *Cell* **126**, 611–625 (2006).
15. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402** (Suppl. 6761), C47–C52. (1999).
A concise and thoughtful essay on the meaning of modularity in biological processes.
16. Tong, A. H. *et al.* Global mapping of the yeast genetic interaction network. *Science* **303**, 808–813 (2004).
The most comprehensive analysis of gene interactions on a large scale.
17. Yook, S. H., Oltvai, Z. N. & Barabasi, A. L. Functional and topological characterization of protein interaction networks. *Proteomics* **4**, 928–942 (2004).
18. Ihmels, J. *et al.* Revealing modular organization in the yeast transcriptional network. *Nature Genet.* **31**, 370–377 (2002).
19. Snel, B., Bork, P. & Huynen, M. A. The identification of functional modules from the genomic association of genes. *Proc. Natl Acad. Sci. USA* **99**, 5890–5895 (2002).
20. Rives, A. W. & Galitski, T. Modular organization of cellular networks. *Proc. Natl Acad. Sci. USA* **100**, 1128–1133 (2003).
21. Yao, G., Craven, M., Drinkwater, N. & Bradfield, C. A. Interaction networks in yeast define and enumerate the signaling steps of the vertebrate aryl hydrocarbon receptor. *PLoS Biol.* **2**, e65 (2004).
22. Johnsson, N. & Varshavsky, A. Split ubiquitin as a sensor of protein interactions *in vivo*. *Proc. Natl Acad. Sci. USA* **91**, 10340–10344 (1994).
The first description of a detector of protein–protein interactions based on *de novo* dissection and refolding of a functional protein.
23. Pelletier, J. N. & Michnick, S. W. A protein complementation assay for detection of protein–protein interactions *in vivo*. *Protein Eng.* **10**, 89 (1997).
Together with references 24 and 25, a first description of the design principles used in creating PCAs.
24. Pelletier, J. N., Campbell-Valois, F. & Michnick, S. W. Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proc. Natl Acad. Sci. USA* **95**, 12141–12146 (1998).
25. Michnick, S. W., Remy, I., Campbell-Valois, F. X., Vallee-Belisle, A. & Pelletier, J. N. Detection of protein–protein interactions by protein fragment complementation strategies. *Methods Enzymol.* **328**, 208–230 (2000).
26. Remy, I. & Michnick, S. W. Clonal selection and *in vivo* quantitation of protein interactions with protein fragment complementation assays. *Proc. Natl Acad. Sci. USA* **96**, 5394–5399 (1999).
27. Remy, I., Wilson, I. A. & Michnick, S. W. Erythropoietin receptor activation by a ligand-induced conformation change. *Science* **283**, 990–993 (1999).
Describes the use of a PCA as a 'molecular ruler' and detector of a novel cytokine-receptor activation mechanism.
28. Remy, I. & Michnick, S. W. Visualization of biochemical networks in living cells. *Proc. Natl Acad. Sci. USA* **98**, 7678–7683 (2001).
The first description of the protein–protein interaction sentinel strategy to map the organization and effects of small molecules on signal-transduction pathways. An extension of the approach for identifying small molecule on- and off-pathway effects is described in reference 70.
29. Galarneau, A., Primeau, M., Trudeau, L. E. & Michnick, S. W. β-Lactamase protein fragment complementation assays as *in vivo* and *in vitro* sensors of protein protein interactions. *Nature Biotech.* **20**, 619–622 (2002).

30. Spotts, J. M., Dolmetsch, R. E. & Greenberg, M. E. Time-lapse imaging of a dynamic phosphorylation-dependent protein–protein interaction in mammalian cells. *Proc. Natl Acad. Sci. USA* **99**, 15142–15147. (2002).
Describes the use of single-cell imaging to quantitate induced dynamic protein–protein interactions by PCA.
31. Wehrman, T., Kleaveland, B., Her, J. H., Balint, R. F. & Blau, H. M. Protein–protein interactions monitored in mammalian cells via complementation of β-lactamase enzyme fragments. *Proc. Natl Acad. Sci. USA* **99**, 3469–3474 (2002).
32. Ghosh, I., Hamilton, A. D. & Regan, L. Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein. *J. Am. Chem. Soc.* **122**, 5658–5659 (2000).
33. Hu, C. D., Chinenov, Y. & Kerppola, T. K. Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol. Cell* **9**, 789–798 (2002).
34. Remy, I. & Michnick, S. A cDNA library functional screening strategy based on fluorescent protein complementation assays to identify novel components of signaling pathways. *Methods* **32**, 381–388 (2004).
Togther with references 35, 36 and 69, this reference describes the application of PCA as a solution to the generalized expression-cloning problem.
35. Remy, I. & Michnick, S. W. Regulation of apoptosis by the Ft1 protein, a new modulator of protein kinase B/Akt. *Mol. Cell. Biol.* **24**, 1493–1504 (2004).
36. Remy, I., Montmarquette, A. & Michnick, S. W. PKB/Akt modulates TGF-β signalling through a direct interaction with Smad3. *Nature Cell Biol.* **6**, 358–365 (2004).
37. Remy, I. & Michnick, S. W. Mapping biochemical networks with protein-fragment complementation assays. *Methods Mol. Biol.* **261**, 411–426 (2004).
38. Luker, K. E. *et al.* 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 (2004).
39. Kaihara, A., Kawai, Y., Sato, M., Ozawa, T. & Umezawa, Y. Locating a protein–protein interaction in living cells via split *Renilla* luciferase complementation. *Anal. Chem.* **75**, 4176–4181 (2003).
40. Paulmurugan, R. & Gambhir, S. S. Monitoring protein–protein interactions using split synthetic *Renilla* luciferase protein-fragment-assisted complementation. *Anal. Chem.* **75**, 1584–1589. (2003).
41. Remy, I. & Michnick, S. W. A highly sensitive protein–protein interaction assay based on *Gaussia* luciferase. *Nature Methods* **3**, 977–979 (2006).
Together with reference 45, the first report that PCAs can be applied to study the disruption of protein–protein interactions.
42. Rossi, F., Charlton, C. A. & Blau, H. M. Monitoring protein–protein interactions in intact eukaryotic cells by β-galactosidase complementation. *Proc. Natl Acad. Sci. USA* **94**, 8405–8410 (1997).
43. Ozawa, T., Nogami, S., Sato, M., Ohya, Y. & Umezawa, Y. A fluorescent indicator for detecting protein–protein interactions *in vivo* based on protein splicing. *Anal. Chem.* **72**, 5151–5157 (2000).
44. Pelletier, J. N., Arndt, K. M., Pluckthun, A. & Michnick, S. W. An *in vivo* library-versus-library selection of optimized protein–protein interactions. *Nature Biotech.* **17**, 683–690 (1999).
Describes the application of a PCA to the directed evolution of pairwise protein–protein interactions.
45. Stefan, E., Aquin, S., Berger, N. & Michnick, S. W. Quantification of dynamic protein complexes using *Renilla* luciferase-fragment complementation applied to PKA activities *in vivo*. *Proc. Natl Acad. Sci. USA* (in the press).
46. Magliery, T. J. *et al.* Detecting protein–protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism. *J. Am. Chem. Soc.* **127**, 146–157 (2005).
47. Nyfeler, B., Michnick, S. W. & Hauri, H. P. Capturing protein interactions in the secretory pathway of living cells. *Proc. Natl Acad. Sci. USA* **102**, 6350–6355 (2005).
A demonstration that a PCA can capture transient interactions of cargo protein–lectin interactions in secretory pathways. Such interactions are not observable by other techniques.
48. Arndt, K. M. *et al.* A heterodimeric coiled–coil peptide pair selected *in vivo* from a designed library-versus-library ensemble. *J. Mol. Biol.* **295**, 627–639 (2000).
49. Campbell-Valois, F. X., Tarassov, K. & Michnick, S. W. Massive sequence perturbation of a small protein. *Proc. Natl Acad. Sci. USA* **102**, 14988–14993 (2005).
The most dramatic example of an application of a PCA to directed evolution of a protein. The entire natural evolution of a protein was recreated in this experiment.
50. Campbell-Valois, F. X. & Michnick, S. W. The transition state of the ras binding domain of Raf is structurally polarized based on Phi-values but is energetically diffuse. *J. Mol. Biol.* **365**, 1559–1577 (2007).
51. Campbell-Valois, F. X., Tarassov, K. & Michnick, S. W. Massive sequence perturbation of the Raf Ras binding domain reveals relationships between sequence conservation, secondary structure propensity, hydrophobic core organization and stability. *J. Mol. Biol.* **362**, 151–171 (2006).
52. Amstutz, P., Koch, H., Binz, H. K., Deuber, S. A. & Pluckthun, A. Rapid selection of specific MAP kinase-binders from designed ankyrin repeat protein libraries. *Protein Eng. Des. Sel.* **19**, 219–229 (2006).
This and reference 53 describe elegant applications of PCAs to select for specific binding proteins from naive libraries.
53. Koch, H., Grafe, N., Schiess, R. & Pluckthun, A. Direct selection of antibodies from complex libraries with the protein fragment complementation assay. *J. Mol. Biol.* **357**, 427–441 (2006).
54. Cody, V., Luft, J. R., Ciszak, E., Kalman, T. I. & Freisheim, J. H. Crystal structure determination at 2.3 Å of recombinant human dihydrofolate reductase ternary complex with NADPH and methotrexate-γ-tetrazole. *Anticancer Drug Design* **7**, 483–491 (1992).
55. Aruffo, A. & Seed, B. Molecular cloning of a CD28 cDNA by a high-efficiency COS cell expression system. *Proc. Natl Acad. Sci. USA* **84**, 8573–8577 (1987).
This, together with references 56–58, is a classic example of expression cloning in which cDNAs were identified based on specific protein–protein interactions.
56. D'Andrea, A. D., Fasman, G. D. & Lodish, H. F. Erythropoietin receptor and interleukin-2 receptor β-chain: a new receptor family. *Cell* **58**, 1023–1024 (1989).
57. Sako, D. *et al.* Expression cloning of a functional glycoprotein ligand for P-selectin. *Cell* **75**, 1179–1186 (1993).
58. Lin, H. Y., Wang, X. F., Ng-Eaton, E., Weinberg, R. A. & Lodish, H. F. Expression cloning of the TGF-β type II receptor, a functional transmembrane serine/threonine kinase. *Cell* **68**, 775–785 (1992).
59. Grimm, S. The art and design of genetic screens: mammalian culture cells. *Nature Rev. Genet.* **5**, 179–189 (2004).
An outstanding review on the expression-cloning problem. See also references 34 and 55.
60. Pawson, T. & Nash, P. Protein–protein interactions define specificity in signal transduction. *Genes Dev.* **14**, 1027–1047 (2000).
61. Weston, C. R. & Davis, R. J. Signal transduction: signaling specificity — a complex affair. *Science* **292**, 2439–2440. (2001).
62. Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA* **98**, 4569–4574 (2001).
63. Ito, T. *et al.* Toward a protein–protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc. Natl Acad. Sci. USA* **97**, 1143–1147. (2000).
64. Drees, B. L. Progress and variations in two-hybrid and three-hybrid technologies. *Curr. Opin. Chem. Biol.* **3**, 64–70 (1999).
65. Walhout, A. J. *et al.* Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287**, 116–122 (2000).
66. Vidal, M. & Legrain, P. Yeast forward and reverse 'n'-hybrid systems. *Nucleic Acids Res.* **27**, 919–929 (1999).
67. Fields, S. & Song, O. A novel genetic system to detect protein–protein interactions. *Nature* **340**, 245–246 (1989).
68. Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
69. Ding, Z. *et al.* A retrovirus-based protein complementation assay screen reveals functional AKT1-binding partners. *Proc. Natl Acad. Sci. USA* **103**, 15014–15019 (2006).
70. Macdonald, M. L. *et al.* Identifying off-target effects and hidden phenotypes of drugs in human cells. *Nature Chem. Biol.* **2**, 329–337 (2006).
The first detailed study of pathway targeting of small molecules as detected by PCA, PCA and high-content quantitative imaging of protein–protein interactions, and the ability to predict unintended (off-pathway) effects of small molecules.
71. Paulmurugan, R., Umezawa, Y. & Gambhir, S. S. Noninvasive imaging of protein–protein interactions in living subjects by using reporter protein complementation and reconstitution strategies. *Proc. Natl Acad. Sci. USA* **99**, 15608–15613. (2002).
The first report of detecting protein–protein interactions in nude mice injected with cells harbouring a PCA. Imaging of PCA reporters could be applied to analysing pharmacodynamics and effects of anticancer drugs in living subjects. See also references 38, 72–74 and 101.
72. Paulmurugan, R., Massoud, T. F., Huang, J. & Gambhir, S. S. Molecular imaging of drug-modulated protein–protein interactions in living subjects. *Cancer Res.* **64**, 2113–2119 (2004).
73. Massoud, T. F., Paulmurugan, R. & Gambhir, S. S. Molecular imaging of homodimeric protein–protein interactions in living subjects. *FASEB J.* **18**, 1105–1107 (2004).
74. Paulmurugan, R. & Gambhir, S. S. Firefly luciferase enzyme fragment complementation for imaging in cells and living animals. *Anal. Chem.* **77**, 1295–1302 (2005).
75. Kaletta, T. & Hengartner, M. O. Finding function in novel targets: *C. elegans* as a model organism. *Nature Rev. Drug Discov.* **5**, 387–398 (2006).
76. Zhang, S., Ma, C. & Chalfie, M. Combinatorial marking of cells and organelles with reconstituted fluorescent proteins. *Cell* **119**, 137–144 (2004).
A particularly interesting application of the GFP PCA to study the role of pairs of genes in animal development.
77. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
78. Benton, R., Sachse, S., Michnick, S. W. & Vosshall, L. B. Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors *in vivo*. *PLoS Biol.* **4**, e20 (2006).
Describes the application of a PCA to study membrane topology and secretory processing of an odorant receptor in transgenic flies.
79. Johnsson, N. & Varshavsky, A. Ubiquitin-assisted dissection of protein transport across membranes. *EMBO J.* **13**, 2686–2698 (1994).
80. Stagljar, I., Korostensky, C., Johnsson, N. & te Heesen, S. A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins *in vivo*. *Proc. Natl Acad. Sci. USA* **95**, 5187–5192. (1998).
81. Laser, H. *et al.* A new screen for protein interactions reveals that the *Saccharomyces cerevisiae* high mobility group proteins Nhp6A/B are involved in the regulation of the GAL1 promoter. *Proc. Natl Acad. Sci. USA* **97**, 13732–13737 (2000).
82. Paumi, C. M. *et al.* Mapping protein–protein interactions for the yeast ABC transporter Ycf1p by integrated split-ubiquitin membrane yeast two-hybrid analysis. *Mol. Cell* **26**, 15–25 (2007).
83. Miller, J. P. *et al.* Large-scale identification of yeast integral membrane protein interactions. *Proc. Natl Acad. Sci. USA* **102**, 12123–12128 (2005).
The first large-scale application of a PCA to identify novel membrane protein–protein interactions in living cells using the ubiquitin split protein sensor. See also reference 82.
84. Anfinsen, C. B., Haber, E., Sela, M. & White Jr. F. H. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl Acad. Sci. USA* **47**, 1309–1314 (1961).
References 84–88 are must-reads for those interested in the history and implications of protein-fragment complementation to modern theories of protein folding.
85. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).
86. Gutte, B. & Merrifield, R. B. The synthesis of ribonuclease A. *J. Biol. Chem.***246**, 1922–1941 (1971).

87. Richards, F. M. On the enzymatic activity of subtilisin-modified ribonuclease. *Proc. Natl Acad. Sci. USA* **44**, 162–166 (1958).
88. Taniuchi, H. & Anfinsen, C. B. Simultaneous formation of two alternative enzymology active structures by complementation of two overlapping fragments of staphylococcal nuclease. *J. Biol. Chem.* **246**, 2291–2301 (1971).
89. Zhang, J., Campbell, R. E., Ting, A. Y. & Tsien, R. Y. Creating new fluorescent probes for cell biology. *Nature Rev. Mol. Cell. Biol.* **3**, 906–918. (2002).
90. Gegg, C. V., Bowers, K. E. & Matthews, C. R. Probing minimal independent folding units in dihydrofolate reductase by molecular dissection. *Protein Sci.* **6**, 1885–1892 (1997).
91. Remy, I. & Michnick, S. W. Dynamic visualization of expressed gene networks. *J. Cell. Physiol.* **196**, 419–429 (2003).
92. Subramaniam, R., Desveaux, D., Spickler, C., Michnick, S. W. & Brisson, N. Direct visualization of protein interactions in plant cells. *Nature Biotech.* **19**, 769–772 (2001).
93. Hu, C. D. & Kerppola, T. K. Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. *Nature Biotech.* **21**, 539–545 (2003).
94. Jach, G., Pesch, M., Richter, K., Frings, S. & Uhrig, J. F. An improved mRFP1 adds red to bimolecular fluorescence complementation. *Nature Methods* **3**, 597–600 (2006).
95. Mervine, S. M., Yost, E. A., Sabo, J. L., Hynes, T. R. & Berlot, C. H. Analysis of G protein βγ dimer formation in live cells using multicolor bimolecular fluorescence complementation demonstrates preferences of β1 for particular γ subunits. *Mol. Pharmacol.* **70**, 194–205 (2006).
96. de Virgilio, M., Kiosses, W. B. & Shattil, S. J. Proximal, selective, and dynamic interactions between integrin αIIbβ3 and protein tyrosine kinases in living cells. *J. Cell Biol.* **165**, 305–311 (2004).
97. Fang, D. & Kerppola, T. K. Ubiquitin-mediated fluorescence complementation reveals that Jun ubiquitinated by Itch/AIP4 is localized to lysosomes. *Proc. Natl Acad. Sci. USA* **101**, 14782–14787 (2004).
98. Walter, M. *et al.* Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. *Plant J.* **40**, 428–438 (2004).
99. Grinberg, A. V., Hu, C. D. & Kerppola, T. K. Visualization of Myc/Max/Mad family dimers and the competition for dimerization in living cells. *Mol. Cell. Biol.* **24**, 4294–4308 (2004).
100. Hynes, T. R., Mervine, S. M., Yost, E. A., Sabo, J. L. & Berlot, C. H. Live cell imaging of Gs and the β2-adrenergic receptor demonstrates that both alphas and β1γ7 internalize upon stimulation and exhibit similar trafficking patterns that differ from that of the β2-adrenergic receptor. *J. Biol. Chem.* **279**, 44101–44112 (2004).
101. Paulmurugan, R. & Gambhir, S. S. Novel fusion protein approach for efficient high-throughput screening of small molecule-mediating protein–protein interactions in cells and living animals. *Cancer Res.* **65**, 7413–7420 (2005).
102. Phillips, K. J., Rosenbaum, D. M. & Liu, D. R. Binding and stability determinants of the PPARγ nuclear receptor-coactivator interface as revealed by shotgun alanine scanning and *in vivo* selection. *J. Am. Chem. Soc.* **128**, 11298–11306 (2006).
103. Tafelmeyer, P., Johnsson, N. & Johnsson, K. Transforming a (β/α)8-barrel enzyme into a split-protein sensor through directed evolution. *Chem. Biol.* **11**, 681–689 (2004).
104. Paschon, D. E., Patel, Z. S. & Ostermeier, M. Enhanced catalytic efficiency of aminoglycoside phosphotransferase (3′)-IIa achieved through protein fragmentation and reassembly. *J. Mol. Biol.* **353**, 26–37 (2005).
105. Jullien, N., Sampieri, F., Enjalbert, A. & Herman, J. P. Regulation of Cre recombinase by ligand-induced complementation of inactive fragments. *Nucleic Acids Res.* **31**, e131 (2003).
106. Thaminy, S., Miller, J. & Stagljar, I. The split-ubiquitin membrane-based yeast two-hybrid system. *Methods Mol. Biol.* **261**, 297–312 (2004).
107. Karimova, G., Pidoux, J., Ullmann, A. & Ladant, D. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc. Natl Acad. Sci. USA* **95**, 5752–5756 (1998).
108. Wehr, M. C. *et al.* Monitoring regulated protein–protein interactions using split TEV. *Nature Methods* **3**, 985–993 (2006).

Acknowledgements
The work published from our group was supported by the CIHR, NSERC, NFSP and NIH. S.W.M. holds the Canada Research Chair in integrative genomics.

Competing interests statement
The authors declare competing financial interests: see web version for details.

FURTHER INFORMATION
Stephen Michnick's Laboratory:
http://michnick.bcm.umontreal.ca
Access to this links box is available online.