Exploring protein interactions by interaction-induced folding of proteins from complementary peptide fragments

Stephen W Michnick

The study of protein–protein interactions is central to understanding the chemical machinery that makes up the living cell. Until recently, facile methods to study these processes in intact, living cells have not existed. Furthermore, the assignment of function to novel proteins relies on demonstrating interactions of these proteins with proteins of known function. This review describes an experimental strategy, devised to study protein–protein interactions in any intact living cells based on protein-fragment complementation assays. Applications to quantitative analysis of interactions, allosteric processes and cDNA library screening are discussed. Recently, the feasibility of employing this strategy in genome-wide biochemical pathway mapping efforts has been demonstrated.

Addresses
Département de Biochimie, Université de Montréal, CP 6128, Succursale centre-ville, Montréal, Québec, H3C 3J7, Canada; e-mail: stephen.michnick@umontreal.ca

Current Opinion in Structural Biology 2001, 11:472–477

0959-440X/01/$ – see front matter

Abbreviations

DHFR	dihydrofolate reductase
EpoR	erythropoietin receptor
fMTX	fluorescein-conjugated methotrexate
GFP	green fluorescent protein
PCA	protein-fragment complementation assay
RTK	receptor tyrosine kinase
USPS	ubiquitin split protein sensor

Introduction

Biochemical 'pathways' are really systems of dynamically assembling and disassembling protein complexes, and thus much of modern biological research is concerned with how, when and where proteins interact with other proteins (and nucleic acids) involved in biochemical processes. The demand for simple approaches to study protein–protein interactions, particularly on a large scale, has grown recently with the progress in genome projects, as the association of unknown with known gene products provides one crucial way of establishing the function of a gene. Yeast two-hybrid screening strategies have been applied systematically across entire genomes and are elegant and robust, but limited to simply 'encoding' protein interactions [1–4]. What we then want to know is where, when, how and in response to what a protein interaction occurs in a living cell. It was with these challenges in mind that my laboratory developed the protein-fragment complementation assay (PCA) strategy. The approach was inspired by a 1994 paper by Johnsson and Varshavsky [5] describing what they called the 'ubiquitin split protein sensor' (USPS). Applications of USPS to studies of transient protein–protein interactions, interaction screening and protein conformation have also been reported more recently [6,7••,8••,9]. In the general PCA strategy, any reporter enzyme can be rationally dissected into two fragments and the fragments are fused to two test proteins that are thought to bind to each other. Folding of the reporter enzyme from its fragments is catalyzed by the binding of the test proteins to each other and is detected as the reconstitution of enzyme activity (Figure 1a). The PCA strategy takes advantage of the spontaneous all-or-none nature of protein folding and the design of fragments follows from basic concepts of protein engineering. In this review, I discuss recent advances in the development and application of the PCA strategy to library screening, quantitative analysis of induced protein–protein interactions and configuration changes and, finally, to an ultimate goal, a way to monitor dynamically evolving biochemical pathways in living cells.

From design to design: putting protein folding to work

The literature is rich in classic experiments in protein chemistry in which protein-fragment complementation was used as the central strategy [10–19]. In practice, however, protein chemists know that reconstituting an enzyme from fragments is not a trivial task; most protein fragment experiments end up in samples of aggregated peptides. To make fragment complementation more successful, we proposed that, if one added soluble oligomerization domains to the fragments that, by interacting, increase the effective concentration of the fragments, correct folding would be favored. My colleagues and I [20] demonstrated the principle with the enzyme dihydrofolate reductase (DHFR). It was also obvious that, if the folding of the enzyme from fragments is absolutely dependent on the binding of the oligomerization domains, then the system described is, in fact, a detector of the interaction. For historical reasons, we decided to call this a protein-fragment complementation assay, or PCA. Others and ourselves have since demonstrated that this principle can be generalized for a number of enzymes, including glycinamide ribonucleotide transformylase, aminoglycoside kinase, hygromycin B kinase, TEM1 β-lactamase and green fluorescent protein (GFP) [20,21••–24••].

A crucial feature of PCA fragments is that they should not fold spontaneously [25]. If this occurred, PCA simply wouldn't work. Spontaneous folding would lead to a false-positive signal; a situation that would hopelessly confound the interpretation of library screens *in vivo* (anticipated to be an important application), for which a more or less positive result could be a result of differences in expression levels, solubility or proteolytic stability,

Figure 1

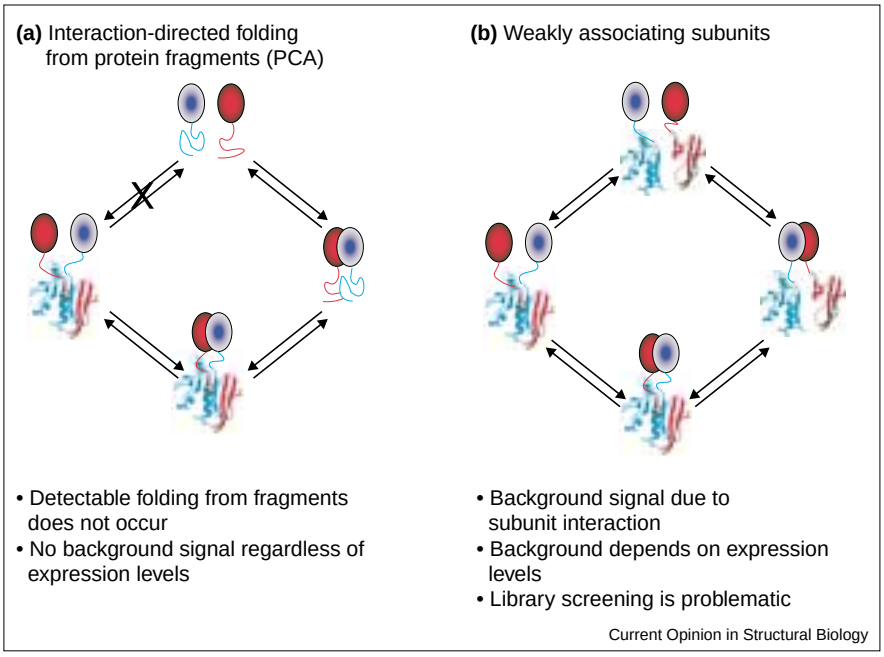


Illustration of two alternative protein-fragment complementation strategies. (a) Oligomerization domain fragment induced folding of complementary peptides (PCA) starts with fragments that have been selected and designed not to fold except when brought into proximity, precluding spontaneous folding that would lead to ambiguous false-positive results. (b) Naturally occurring subunits can be fused to oligomerizing proteins; however, weakly interacting subunits will nevertheless result in some population of spontaneously reconstituted enzyme.

rather than a protein–protein interaction itself. We have shown that, even when expressed at very high levels in bacteria, spontaneous fragment folding does not occur [20,26••]. In contrast to PCAs, there are assay systems based on β-galactosidase and split inteins (self-splicing internal proteins) that resemble PCA in a superficial way, but that are conceptually and practically different [27,28]. In both cases, well-known, naturally occurring and spontaneously associating subunits of the enzymes are fused to interacting proteins. The central problem here is that the subunits are always capable of associating to some extent, even if only weakly, meaning that there is a constant background of spontaneous assembly (Figure 1b).

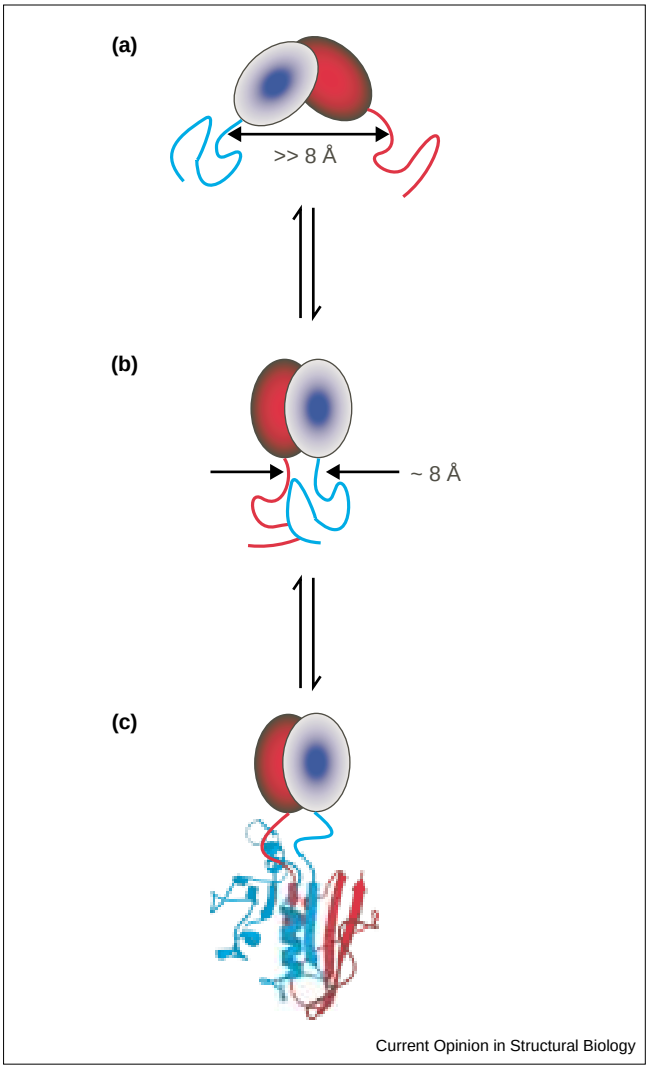
The selection of a protein as a good candidate for designing a PCA and the design steps involved have been discussed recently in some detail and will only be summarized here [21••]. The design of a PCA starts with the known three-dimensional structure of a small, monomeric enzyme to assure that the stoichiometry of resulting complexes is understood (i.e. binary interactions only). Designing fragments and fusions depends on a 'grab bag' of protein engineering ideas. The most critical issue, of course, is where to cut. The best evidence for whether particular fragments will work is to determine whether the polypeptide sequence can be circularly permuted at the proposed points of dissection; this has turned out to be the case for every enzyme that has been used in a PCA. Another crucial issue is on which terminus to fuse the proteins of interest to the fragments. For DHFR, the orientation of fusions was found to make little difference [29••]. For proteins that have clear domains that are contiguous with the polypeptide sequence, such as DHFR, we would predict that any configuration would work, as the domain topology can be formed from any configuration of fusions. In contrast, however, β-lactamase does not have a domain structure that is contiguous with the sequence and GFP is a single domain. In both cases, one configuration is favored over any other.

PCA in protein design: library-versus-library screening for optimally interacting proteins

One of the first applications of the DHFR PCA was to a protein design problem. The DHFR survival assay in *Escherichia coli* was used to screen two libraries of complementary designed leucine-zipper-forming sequences with 10^{10} potential interacting pairs of which 10^6 could practically be covered. The features of the resulting coiled coils were consistent with the selection of more stable pairs, with a large increase in heterospecificity versus homospecificity [26••,30•].

The implications of these results were that the PCA strategy rapidly selects not only for optimal properties of interacting sequences, along with critical stereospecific and regiospecific requirements of such complexes, but also 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 about the design strategy suggest broad utility of the DHFR PCA in protein design and directed evolution experiments. It also shows that PCA rivals 'phage display' strategies, as the entire selection, optimization and stringency tests are done *in vivo*, making this approach easily executed. Recently, the USPS strategy has been demonstrated to show use in

Figure 2



Schematic representation of how a structural model for an intersubunit conformation change can be detected using the DHFR PCA. (a) Protein dimers in one state are in a configuration that prevents the fragments from coming into sufficient proximity (~8 Å) to fold/reassemble into the native structure. (b) In the second state, fragment N termini are brought into proximity and (c) the enzyme can fold.

genomic DNA library screening in yeast and may therefore also show promise as a general cDNA library screening strategy with some advantages over yeast two-hybrid approaches, such as the ability to study interactions in the cytosol instead of the nucleus [7••].

Quantitative pharmacology using PCA in mammalian cells

Recently, we have successfully demonstrated two different types of DHFR PCA in mammalian cells [31••,32••]. In the 'DHFR PCA survival assay', DHFR-negative cells are co-transfected with DHFR complementary fragments fused to two partner proteins. Co-transfectants are then selected for survival in a nucleotide-free medium (selection for DHFR activity). In a second assay, fluorescence PCA, the high-affinity fluorescein-conjugated DHFR inhibitor methotrexate (fMTX) passively diffuses into cells, where it binds in a 1:1 complex with DHFR. Free fMTX is actively transported from the cells, leaving only DHFR-bound fMTX. Fluorescence can then be detected using fluorescence microscopy or fluorescence-activated flow cytometry. Two capabilities of the fluorescence assay have been demonstrated. First, microscopic analysis has demonstrated that the cellular locations of interacting proteins can be observed. No other method allows the locations of protein complexes to be determined unambiguously. Second, quantitative dose-dependence of ligand-induced protein association or conformation changes has been demonstrated and results are consistent with the known pharmacological responses for tested interactions. These results are surprising and suggest that the overall free energy contribution to binding of fragment folding is effectively zero. Similar results for formation of coiled coils with the GFP PCA are also consistent with a null contribution to binding free energy by fragment folding [24••]. It is not yet clear how this occurs and further experimentation will be required. The simplest explanation is that folding from fragments is not particularly efficient, such that the effective association constant for folding is close to one. In this case, the overall equilibrium constant for the entire process would be equal to that of the associating proteins.

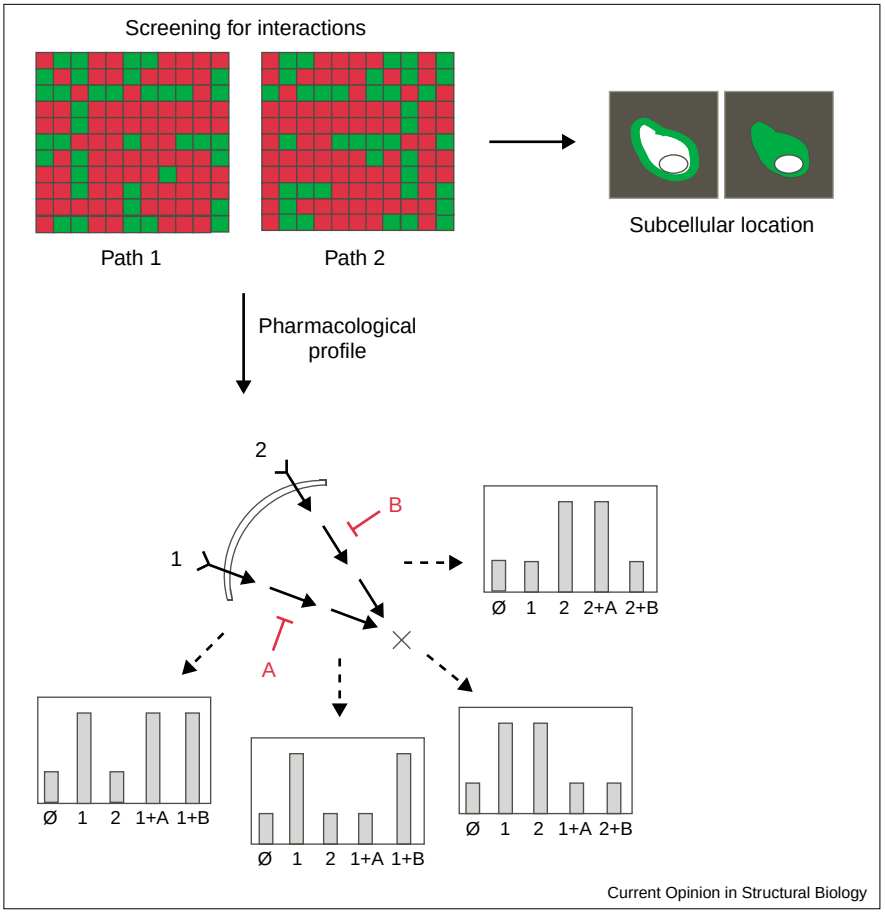
Molecular rulers

A special feature of PCA strategies is that, knowing the three-dimensional structure of the reporter enzyme, it is possible to accurately predict how close together the fragments must be to assure that the enzyme will fold correctly and have a measurable activity (Figure 2). This fact was put to work to test a structural allosteric model for activation of the dimeric erythropoietin receptor (EpoR) using the DHFR PCA. The approach could be extended to studying allosteric transitions in dimeric or multimeric protein interfaces [31••]. In the EpoR case, the dimeric receptor's transmembrane domains were shown to be separated by 73 Å, as observed in the crystal structure of unligated EpoR. It was reasoned that, if this inactive state exists on the membrane of a living cell, then DHFR fragments fused to the C termini of the transmembrane domains fold only if ligand binding induced a conformational change that allowed the fragments to come close enough together to assure that the precise three-dimensional structure of DHFR can be formed [20,33]. This would require that the N termini of the fragments be 8 Å apart [33]. 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 to confirm that linkers long enough to span 73 Å were needed for DHFR to fold from its fragments.

Another consequence of knowing the final structure of the PCA probe enzyme is that linker lengths can be

Figure 3

Schematic representation of the strategy for generating a functional validation profile of a biochemical network using the DHFR PCA. Positive clones are detected with the DHFR survival-selection assay. They correspond to interacting component proteins of two convergent signal transduction pathways (path 1 and path 2). An interaction matrix represents all positive (green) and negative (red) interacting pairs observed in the survival-selection assay. Positive clones from survival selection are propagated and subjected to two functional analyses. Using the DHFR fluorescence assay, interactions are probed with pathway-specific stimulators (1 and 2) and inhibitors (A and B). Pharmacological profiles are established based on the pattern of response of individual interactions to stimulators and inhibitors, represented in the histograms. For example, stimulation of pathway 1 will augment all the interactions comprising that pathway. The inhibitor A will inhibit protein interactions downstream, but not upstream, of its site of action in pathway 1. Cellular locations of the interactions are determined by fluorescence microscopy, also using the DHFR fluorescence assay.



adjusted to allow only certain types of orientations of fused protein interactions to occur. For example, this strategy was employed to assure that only parallel coiled-coil-forming sequences would be selected using the DHFR PCA and a method for selecting only antiparallel coiled-coil sequences has been proposed using the GFP PCA [24••]. Furthermore, intramolecular conformation changes in the yeast protein Sec62p have been demonstrated using USPS. Although no structural model was available, careful comparison of mutants provided a functional interpretation of the conformation change [8••].

Mapping biochemical pathways

At the outset, I noted that the real motivation behind studying protein interactions in living cells was to quantitatively map biochemical pathways. In a recent report, we demonstrated that screening, pharmacological 'profiling' (monitoring effects of pathway-specific drugs and protein hormones) and determining the cellular location of protein–protein interactions can be achieved using the survival-selection and fluorescence DHFR PCA (Figure 3) [29••]. Analysis of these results allows a representation of how biochemical pathways evolve in time and space, and in response to specific stimuli.

As a proof of principle, Remy and Michnick [29••] reported the application of this strategy to the mapping of a signal transduction pathway mediated by receptor tyrosine kinases (RTKs). The pharmacological profiles and cellular location of interactions we observed allowed us to 'place' each gene product at its relevant point in the pathways. From the results of our analysis, a map of the organization of the RTK network emerged that was consistent with existing models, but included several novel interactions. The ability to monitor a network of protein interactions in living cells containing all of the components of the underlying pathway studied revealed hidden connections not observed before, in spite of intense scrutiny of this network. The results presented demonstrate that the PCA strategy has the features necessary for a general gene function validation and pathway mapping strategy.

Conclusions and future perspectives

The development and application of PCA strategies are still at a very early stage. Applications of the existing techniques to problems discussed in this review and improvements in the existing techniques are being explored. For example, instead of artificial libraries described here, the PCA strategy is being applied to large-scale screening of entire genomes. More sophisticated

problems of protein design and protein folding are being explored, including studies of what factors control the selection of sequences for optimal interactions between proteins, proteins and nucleic acids, and proteins and small organic molecules. As discussed above, protein interactions are not merely interesting in themselves, but the dynamic formation of protein complexes is intrinsic to what we call biochemical pathways in the cell. Thus, the 'mapping' of biochemical pathways by PCA should prove a fruitful endeavor. Finally, the focus returns to protein-interaction-assisted fragment folding as a method to study enzymes and design novel catalysts. Kinetic studies of oligomerization-domain-assisted reconstitution of enzymes could be used to quantify the reassembly process and the catalytic properties of reassembled enzymes, to facilitate the alteration of the protein fragments that are responsible for catalytic activity to determine the functional motifs that are essential for substrate binding and catalysis, and to define functional building blocks that can be combined to design new protein catalysts.

Acknowledgements

I wish to acknowledge the work of my students and, in particular, Ingrid Remy who conceived and performed most of the key pioneering studies. Work from my laboratory discussed here was supported by the Burroughs-Wellcome Fund (BWF) and the Canadian Institute of Health Sciences. SWM is a BWF New Investigator, MRC of Canada Scientist and Canada Research Chair in Integrative Genomics.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Vidal M, Legrain P: Yeast forward and reverse 'n'-hybrid systems. *Nucleic Acids Res* 1999, 27:919-929.

2. Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, Brasch MA, Thierry-Mieg N, Vidal M: Protein interaction mapping in ***C. elegans*** using proteins involved in vulval development. *Science* 2000, 287:116-122.

3. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P *et al.*: A comprehensive analysis of protein–protein interactions in ***Saccharomyces cerevisiae***. *Nature* 2000, 403:623-627.

4. Rain JC, Selig L, De Reuse H, Battaglia V, Reverdy C, Simon S, Lenzen G, Petel F, Wojcik J, Schachter V *et al.*: The protein-protein interaction map of ***Helicobacter pylori***. *Nature* 2001, 409:211-215.

5. Johnsson N, Varshavsky A: Split ubiquitin as a sensor of protein interactions ***in vivo***. *Proc Natl Acad Sci USA* 1994, 91:10340-10344.

6. Dunnwald M, Varshavsky A, Johnsson N: Detection of transient ***in vivo*** interactions between substrate and transporter during protein translocation into the endoplasmic reticulum. *Mol Biol Cell* 1999, 10:329-344.

7. •• Laser H, Bongards C, Schuller J, Heck S, Johnsson N, Lehming N: 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* 2000, 97:13732-13737.
The authors demonstrate a survival-selection assay yeast genomic DNA library screening for interacting partner proteins.

8. •• Raquet X, Eckert JH, Muller S, Johnsson N: Detection of altered protein conformations in living cells. *J Mol Biol* 2001, 305:927-938.
This paper describes the use of the USPS to elucidate the effect of a previously described mutation in the protein Sec62p on its conformation in living cells.

9. Wittke S, Lewke N, Muller S, Johnsson N: Probing the molecular environment of membrane proteins ***in vivo***. *Mol Biol Cell* 1999, 10:2519-2530.

10. Richards FM: On the enzymatic activity of subtilisin-modified ribonuclease. *Proc Natl Acad Sci USA* 1958, 44:162-166.

11. Taniuchi H, Anfinsen CB: Simultaneous formation of two alternative enzymically active structures by complementation of two overlapping fragments of staphylococcal nuclease. *J Biol Chem* 1971, 216:2291-2301.

12. Ullmann A, Jacob F, Monod J: Characterization by ***in vitro*** complementation of a peptide corresponding to an operator-proximal segment of the beta-galactosidase structural gene of ***Escherichia coli***. *J Mol Biol* 1967, 24:339-343.

13. Ladurner AG, Itzhaki LS, de Prat Gay G, Fersht AR: Complementation of peptide fragments of the single domain protein chymotrypsin inhibitor 2. *J Mol Biol* 1997, 273:317-329.

14. Ruiz-Sanz J, de Prat Gay G, Otzen DE, Fersht AR: Protein fragments as models for events in protein folding pathways: protein engineering analysis of the association of two complementary fragments of the barley chymotrypsin inhibitor 2 (CI-2). *Biochemistry* 1995, 34:1695-1701.

15. Hartley RW: Complementation of peptides of barnase, extracellular ribonuclease of ***Bacillus amyloliquefaciens***. *J Biol Chem* 1977, 252:3252-3254.

16. Shiba K, Schimmel P: Functional assembly of a randomly cleaved protein. *Proc Natl Acad Sci USA* 1992, 89:1880-1884.

17. Tasayco ML, Carey J: Ordered self-assembly of polypeptide fragments to form nativelike dimeric trp repressor. *Science* 1992, 255:594-597.

18. Bryson JW, Betz SF, Lu HS, Suich DJ, Zhou H, Oneil KT, Degrado WF: Protein design — a hierarchic approach. *Science* 1995, 270:935-941.

19. Braisted AC, Wells JA: Minimizing a binding domain from protein A. *Proc Natl Acad Sci USA* 1996, 93:5688-5692.

20. Pelletier JN, Campbell-Valois F, Michnick SW: Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proc Natl Acad Sci USA* 1998, 95:12141-12146.

21. •• Michnick SW, Remy I, Valois CFX, Belisle A, Pelletier JN: Detection of protein–protein interactions by protein fragment complementation strategies. In *Methods in Enzymology*, vol 328. Edited by Abelson JN, Emr SD, Thorner J. New York: Academic Press; 2000:208-230.
A detailed discussion of the design principles used to make PCAs. Examples are given of PCAs based on several different proteins and their applications, advantages and disadvantages.

22. •• Remy I, Michnick SW: Clonal selection and ***in vivo*** quantitation of protein interactions with protein fragment complementation assays. *Proc Natl Acad Sci USA* 1999, 96:5394-5399.
A description of survival-selection and fluorescence DHFR PCAs in living mammalian cells. Examples are given of applications to quantitative pharmacology by FACS (fluorescence-activated cell sorting).

23. •• Remy I, Pelletier JN, Galarneau A, Michnick SW: Protein fragment complementation strategies. In *Protein–Protein Interactions.* Edited by Golemis E, Serebriiskii I. Cold Spring Harbor, NY: Cold Spring Harbor Laboratories Press; 2001:in press.
Protocols for performing the DHFR and β-lactamase PCAs are described.

24. •• Ghosh I, Hamilton AD, Regan L: Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein. *J Am Chem Soc* 2000, 122:5658-5659.
A short report describing a PCA based on GFP. This is potentially very useful as no substrate is required to observe fluorescence in living cells. The assay works with a fixed orientation of peptides fused to the N and C termini of the fragments, suggesting a strategy to select for antiparallel coiled coils, but there are probably many other applications.

25. Gegg CV, Bowers KE, Matthews CR: Probing minimal independent folding units in dihydrofolate reductase by molecular dissection. *Protein Sci* 1997, 6:1885-1892.

26. •• Pelletier JN, Arndt KM, Plückthun A, Michnick SW: An ***in vivo*** competition strategy for the selection of optimized protein–protein interactions. *Nat Biotechnol* 1999, 17:683-690.
The first example of a true library-versus-library screening strategy based on a survival-selection DHFR PCA in *E. coli*. Stringency and metabolic selection strategies are applied to a rationally designed library of potential parallel coiled-coil-forming sequences (10^{10} potential interactions) and the method was shown to rapidly select for optimally interacting heterodimerizing pairs.

27. Rossi F, Charlton CA, Blau HM: Monitoring protein–protein interactions in intact eukaryotic cells by beta-galactosidase complementation. *Proc Natl Acad Sci USA* 1997, 94:8405-8410.

28. 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* 2000, 72:5151-5157.

29. •• Remy I, Michnick SW: Visualization of biochemical networks in living cells. *Proc Natl Acad Sci USA* 2001, 98:7678-7683.
A demonstration of how an interacting network of proteins comprising a signal transduction pathway in mammalian cells can be mapped and visualized using the DHFR PCA.

30. • Arndt KM, Pelletier JN, Muller KM, Alber T, Michnick SW, Pluckthun A: A heterodimeric coiled-coil peptide pair selected ***in vivo*** from a designed library-versus-library ensemble. *J Mol Biol* 2000, 295:627-639.
The follow up to [26••], in which selected coiled-coil-forming peptides are analyzed physically. Peptides selected from the libraries are shown to be optimal dimerizers, with affinities greater than any coiled coil described to date. The rates of evolution of sequences through the selection process suggest that some types of interactions are optimized faster than others.

31. •• Remy I, Wilson IA, Michnick SW: Erythropoietin receptor activation by a ligand-induced conformation change. *Science* 1999, 283:990-993.
The first demonstration of the DHFR PCA being used as a molecular ruler. This study confirms a structure-based model for the activation of a protein hormone receptor.

32. •• Remy I, Michnick SW: Clonal selection and ***in vivo*** quantitation of protein interactions with protein fragment complementation assays. *Proc Natl Acad Sci USA* 1999, 96:5394-5399.
The first description of the DHFR survival-selection and fluorescence PCAs in mammalian cells.

33. Cody V, Luft JR, Ciszak E, Kalman TI, Freisheim JH: Crystal structure determination at 2.3 Å of recombinant human dihydrofolate reductase ternary complex with NADPH and methotrexate-gamma-tetrazole. *Anticancer Drug Des* 1992, 7:483-491.