Protein-fragment complementation assays for large-scale analysis of protein–protein interactions

Ewa Blaszczak1, Natalia Lazarewicz1,2, Aswani Sudevan2, Robert Wysocki1 and Gwenaël Rabut2

1Department of Genetics and Cell Physiology, Faculty of Biological Sciences, University of Wroclaw, Kanonia 6/8, 50-328 Wroclaw, Poland; 2Univ Rennes, CNRS, IGDR (Institute of Genetics and Development of Rennes) – UMR 6290, F-35000 Rennes, France

Correspondence: Ewa Blaszczak (ewa.blaszczak@uwr.edu.pl) or Gwenaël Rabut (gwenael.rabut@univ-rennes1.fr)

Protein–protein interactions (PPIs) orchestrate nearly all biological processes. They are also considered attractive drug targets for treating many human diseases, including cancers and neurodegenerative disorders. Protein-fragment complementation assays (PCAs) provide a direct and straightforward way to study PPIs in living cells or multicellular organisms. Importantly, PCAs can be used to detect the interaction of proteins expressed at endogenous levels in their native cellular environment. In this review, we present the principle of PCAs and discuss some of their advantages and limitations. We describe their application in large-scale experiments to investigate PPI networks and to screen or profile PPI targeting compounds.

Introduction

Protein–protein interactions (PPIs) constitute a complex cellular network and play a central role in nearly all biological processes, including DNA replication, transcription, signal transduction, enzymatic reactions, cell-to-cell communication, and membrane transport [1,2]. Alterations in protein–protein interactions (either by breakdown or the formation of novel PPIs) may lead to various diseases in humans, including cancers [3,4] or neurodegenerative disorders [5]. PPIs are attractive molecular targets with a vast therapeutic potential [6–8].

A wide range of methods are available to investigate cellular PPIs [9,10]. Biochemical methods such as immuno- and affinity purification are widely used. Coupled to mass spectrometry they enable proteome-wide analysis of PPI networks [11]. However, these techniques remove proteins from their native cellular environment, which disturbs numerous physiological interactions. In contrast, several methods allow to monitor PPIs in intact living cells and organisms, with minimal cellular perturbations. These include Förster resonance energy transfer (FRET) [12], bioluminescence resonance energy transfer (BRET) [13], fluorescence cross-correlation spectroscopy (FCCS) [14], proximity labeling [15], and protein-fragment complementation assays (PCAs) [16,17]. With the advent of gene editing and PCR tagging technologies [18–20], these methods now enable exploring the interaction network of proteins expressed at near endogenous levels in virtually any model organism or tissue culture system. Importantly, several of the aforementioned methods are suitable to probe transient PPIs and to quantify their dynamics over time. In this review, we present the principle and development of PCAs and their application in large-scale studies. The recent advances are highlighted and the main advantages and limitations of PCA-based methods are also discussed.

General principles of protein-fragment complementation assays

Development of PCA reporters

The idea of using the functional complementation of protein fragments to probe PPIs in living cells dates back to 1989, when Fields and Song described the yeast two-hybrid system (Y2H) [21]. This
method relies on monitoring the interaction of two proteins of interest (a ‘bait’ and a ‘prey’) via the transcriptional activation of a reporter gene (usually encoding a selectable marker or a colorimetric enzyme). The bait and the prey are genetically fused to two fragments of the Gal4 transcription factor, the DNA binding domain and the activation domain, respectively, and expressed in yeast cells. Interaction of these ‘hybrid’ proteins restores the activity of the transcription factor and controls the expression of the reporter gene [22]. Y2H was optimised for large-scale analysis and largely contributed to the construction of proteome-wide PPI maps in multiple organisms, from yeast to humans [23–27].

Contrary to Y2H, PCAs do not rely on the cellular transcriptional machinery to detect PPIs. They make use of monomeric reporter proteins whose activity can be directly detected (enzymes or fluorescent proteins). The structure of the reporter protein is engineered to split it into two inactive but complementary fragments. When fused to interacting proteins, these fragments are brought into close proximity, which triggers the reconstitution of the reporter and the detection of its activity (Figure 1). Hence, Y2H and PCAs are conceptually different methods and have distinct applications. Johnsson and Varshavsky described a PCA precursor technique using ubiquitin as a reporter [28]. However, in this assay detection of reconstituted ubiquitin is indirect: it requires cellular proteases to cleave it and immunoblotting to reveal the cleavage. The first bona fide PCA was based on a mutated version of murine dihydrofolate reductase (DHFR) [29] (Figure 2). Since then, several other enzymes (e.g. β-lactamase, luciferases) and fluorescent proteins have been engineered into PCA reporters [30–38] allowing multiple modes of PPI detection (Figure 3). Fluorescent protein PCAs are very popular because they enable the visualisation of PPIs in living cells using fluorescence microscopy. The most frequent are based on the yellow fluorescent protein (YFP) or its variant Venus. These assays are often referred to as Bimolecular Fluorescence Complementation (BiFC). Among enzymatic reporters, DHFR, which confers resistance to the anti-folate drug methotrexate, allows probing PPIs using simple and inexpensive survival selection assays. These do not require specialised reagents or equipment and can be easily scaled up. Luciferases are enzymes that catalyse the oxidation of substrate luciferins in a reaction that emits light. Various luciferases with different properties have been used as PCA reporters, including Renilla (RLuc) [37], firefly (FLuc) [34], Gaussia (GLuc) [38], and NanoLuc [36] luciferases. Luciferase PCAs are well suited for large-scale studies because they are performed in microtiter plates and are easily scalable.

Properties, advantages and limitations of PCAs

PCAs constitute attractive approaches for investigating PPIs in living cells. Since they detect PPIs directly, they do not require the use of dedicated host cells (as it is the case for Y2H approaches) and can be used in any genetically amenable model system. They enable to explore the interaction of proteins in their native cellular context, expressed at physiological levels from endogenously tagged genes, and subjected to post-translational modifications and other regulatory mechanisms. Apart from widely used model organisms, such as mammalian cells, plants, fruit fly, or budding yeast, PCAs can serve to explore PPIs in pathogens, e.g. the pathogenic yeast Candida albicans [39] or the parasite Plasmodium falciparum [40].
An important feature of PCAs is that they are exquisitely sensitive to topology. The efficiency of the reconstitution of the PCA reporter depends on the spatial distance and the mobility of the individual fragments fused to the bait and prey proteins. Hence, N-terminal, C-terminal, or internal fusions are not equivalent and can yield different results. The linkers are also important and influence PCA signals. Using the DHFR PCA in yeast, it has been shown that long linkers significantly improve the detection of PPIs and allow capturing indirect interactions in multi-subunit complexes [41]. As a rule of thumb, PCAs are suitable to probe PPIs when the tagged protein extremities are <100 Å apart [41,42]. In addition to distance consideration, fusing proteins to reporter fragments can affect their function in different ways. Like any other tag, PCA fragments can create steric hindrances that impair protein activity, interaction or localisation. Moreover, certain unfolded PCA fragments may destabilise fusion proteins [43,44]. Generally, small PCA reporters that can be split in stable fragments are less likely to perturb protein activities. For example, the NanoBiT PCA reporter (engineered from the small (19 kDa) NanoLuc luciferase) has been optimised for both the size and stability of its fragments [36].

The dynamics of the reporter reconstitution is a critical parameter of PCAs. For most applications, reconstitution of the reporter should not happen spontaneously, should be reversible, and should not interfere with the binding properties of the bait and prey proteins. In other words, the PCA fragments should have low intrinsic affinity so that they only reconstitute the reporter when their ‘local’ concentration is raised due to the interaction of the bait and prey proteins. To our knowledge, the affinity of PCA fragments has been determined for FLuc ($K_D = 21 \mu M$) [36], NanoBiT ($K_D = 190 \mu M$) [36] and the recently developed fluorescent PCA termed splitFAST ($K_D \sim 1 \mu M$) [45]. Still, many other PCA reporters (including β-lactamase, DHFR and luciferases) have been demonstrated to be reversible [31,38,42,46], with the notable exception of GFP-family fluorescent proteins [33,47]. Irreversible PCAs may artificially trap and amplify non-physiological interactions, especially when proteins are highly expressed, locally tethered or constrained in dense compartments. They are thus best suited to probe irreversible protein interactions or for applications that benefit from molecular trapping [47]. Despite this limitation, fluorescent PCAs had a tremendous impact on the study of PPIs in living cells and new generations of reversible fluorescent reporters are actively being developed [45,48,49].

The sensitivity of PCAs depends on the detection method and equipment used to measure the activity of the reporter as well as the level of background produced by the cells. In general, enzymatic reporters are very sensitive because they enable signal amplification. The DHFR PCA has been reported to be sufficiently sensitive to detect as little as 25 complexes per cell [50]. Luciferase reporters are similarly able to detect the interaction of many endogenously expressed proteins, although their detection limit has, to our knowledge, not been rigorously determined. Importantly, luciferase PCAs produce linear luminescence signals over a large dynamic range.
and enable to monitor the assembly and disassembly of PPIs in near real-time [36,38,46]. They are thus especially well-suited to dissect the actions of drugs and for drug screening applications. In contrast, GFP-family reporters have a long maturation time (typically minutes to hours), which prevents time-course studies and may even misinform spatial interpretations.

### PCAs applied to large-scale analysis of PPIs

**Unravelling PPI networks**

The elucidation of PPI networks is essential to understand the organisation and regulation of cellular processes. To investigate PPI networks using PCAs, several genome-wide reporter fragment libraries have been developed in budding yeast, mammalian cells and fruit fly (Table 1). Libraries that have been established with a single fragment enable to systematically probe the interactions of selected bait proteins with an array of preys (as exemplified in Figure 4), while two-fragment libraries enable probing any PPI combination. Examples of studies using these and other resources are presented hereafter.

The first large-scale PCA experiment used the DHFR reporter. Tarassov and colleagues [42] constructed genome-wide yeast libraries of both DHFR fragments and conducted a systematic screen, which detected 2770 PPIs involving 1124 proteins. Interestingly, ~80% of these PPIs were previously unknown. This indicates that PCAs interrogate a different space of the interactome than biochemical and Y2H assays and highlights the need of using complementary approaches for PPI mapping. This system was later applied to analyse the modulation of more than 1000 PPIs in response to DNA damage [51]. The DHFR PCA reporter was also combined with barcoding and next-generation sequencing to investigate PPIs at a high-throughput in pooled

---

**Figure 3. Main PCA reporters with their advantages and limitations.**

Presentation of different types of PCAs (colorimetric, fluorescent, luminescent or survival assays) with examples of their main reporters (abbreviations: GFP — green fluorescent protein, CFP — cyan fluorescent protein, RLuc — Renilla luciferase, FLuc — firefly luciferase, GLuc — Gaussia luciferase, NanoLuc — NanoLuc® luciferase, DHFR — dihydrofolate reductase) and some of their advantages and limitations.
cultures [52]. This approach, now termed protein–protein interaction sequencing (PPiSeq), was used to examine the dynamics of large numbers of yeast protein pairs across multiple environmental conditions [53,54]. Recently, PPiSeq was scaled up to quantify the relative abundance of ∼9% of all possible yeast protein pairs (1.6 million). It revealed that most PPIs detected with this method are regulated and only observed in some specific growth conditions [55]. These studies demonstrate the power of using pooled rather than individual cultures for large-scale interaction studies.

BiFC is a popular approach for large-scale PPI studies. Sung and colleagues [56] used a genome-wide library of yeast strains tagged with the N-terminal fragment of the Venus fluorescent protein to probe the interactome of the small ubiquitin-like modifier (SUMO). SUMO is post-translationally conjugated to lysine residues of target proteins to regulate their activity. By comparing the signals produced by the wild type and a mutated form of SUMO that cannot be conjugated to proteins, they were able to identify 280 putative SUMO conjugates, among which 31 were validated in a band-shift assay. We utilised the same collection of yeast strains to systematically search for ubiquitin ligases interacting with the Ubc6 conjugating enzyme. This resulted in the identification of a novel protein quality control pathway involved in the maintenance of the inner nuclear membrane [57]. Recently, Kim and colleagues [58] constructed a complementary library of yeast strains, tagged with the C-terminal fragment of Venus, which enabled them to investigate protein homodimerisation. They identified 630 proteins producing a detectable fluorescent signal, yet the controls they included led them to classify ∼70% of these proteins as false-positives. We also observed a high rate of false-positives, when we investigated the interaction network of ubiquitin ligases and ubiquitin-conjugating enzymes using a sensitive microscopy setup ([44] and unpublished results). These results highlight the importance of designing

### Table 1 PCA resources for large-scale PPI investigation

<table>
<thead>
<tr>
<th>Organism</th>
<th>PCA reporter and its fragment(s)</th>
<th>Details</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. cerevisiae</td>
<td>mDHFR, 2 fragments</td>
<td>4326 strains with the DHFR F[1,2] fragment</td>
<td>[42]</td>
</tr>
<tr>
<td></td>
<td>(DHFR F[1,2], DHFR F[3])</td>
<td>4604 strains with the DHFR F[3] fragment</td>
<td>Commercially available (Horizon Discovery) [50]</td>
</tr>
<tr>
<td></td>
<td>mDHFR, 2 fragments</td>
<td>1741 strains with the DHFR F[1,2] fragment</td>
<td>Partially commercially available (VN173, Bioneer) [56,58]</td>
</tr>
<tr>
<td></td>
<td>(DHFR F[1,2], DHFR F[3])</td>
<td>1113 strains with the DHFR F[3] fragment</td>
<td>[69] and unpublished results</td>
</tr>
<tr>
<td></td>
<td>Venus, 2 fragments</td>
<td>5809 strains with the VN173 fragment</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(VN173, VC155)</td>
<td>5671 strains with the VC fragment</td>
<td>[56]</td>
</tr>
<tr>
<td></td>
<td>NanoBiT, 2 fragments</td>
<td>5580 strains tagged with the LgBiT fragment</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(LgBiT, SmBiT)</td>
<td>4981 strains tagged with the SmBiT fragment</td>
<td>[61]</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>Venus, 1 fragment (VN173)</td>
<td>450 fly lines focused on transcription factors</td>
<td>[60]</td>
</tr>
<tr>
<td></td>
<td>Cerulean, 2 fragments (CN173, CC155)</td>
<td>N- and C-terminally tagged</td>
<td>-</td>
</tr>
<tr>
<td>Mammalian cells</td>
<td>YFP, 1 fragment (YFPC)</td>
<td>11 880 ORFs cloned in retroviral plasmids</td>
<td>[61]</td>
</tr>
<tr>
<td></td>
<td>(VC159)</td>
<td>N- and C-terminally tagged</td>
<td>[86]</td>
</tr>
<tr>
<td></td>
<td>Gauussia luciferiae, 2 fragments (hGLuc1, hGLuc2)</td>
<td>cDNA library cloned in a retroviral plasmid</td>
<td>[27]</td>
</tr>
<tr>
<td></td>
<td>NanoBiT, 2 fragments (LgBiT, SmBiT)</td>
<td>ORFs from 138 protein pairs cloned in mammalian expression vectors compatible with the human ORFeome (17 408 protein coding genes)</td>
<td>[66]</td>
</tr>
<tr>
<td></td>
<td>NanoLuc, 2 fragments (N2H[F1], N2H[F2])</td>
<td>ORFs from 138 protein pairs cloned in mammalian expression vectors compatible with the human ORFeome (17 408 protein coding genes)</td>
<td>[66]</td>
</tr>
</tbody>
</table>

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY-NC-ND).
appropriate controls for the analysis and interpretation of PCA experiments, especially while using irreversible reporters. Large-scale BiFC studies have also been performed in other organisms including plants [59], fruit fly [60] and mammalian cells [61]. Technological developments are essential to empower high-throughput data acquisition and analysis of large-scale BiFC experiments. To examine the interactions of six core telomeric subunits with 12 000 human proteins Lee and colleagues [61] performed fluorescence measurements with a high-throughput flow cytometer and processed them with a specially developed automated data analysis pipeline. They could identify ~300 proteins associated with the six core telomeric proteins, including possible regulators of telomere biology, such as protein kinases and ubiquitin ligases. Like for the DHFR PCA, BiFC studies may benefit from sequencing technologies. Two preprint publications propose to combine flow cytometry, cell sorting and next-generation sequencing to analyse high-throughput BiFC experiments, which could make it possible to perform BiFC screens in pooled cultures [62,63].

Luminescence-based assays are well suited for large scale studies. The GLuc PCA is one of the most prevalent because it offers bright luminescence with a high signal to background ratio, and it is able to detect a large
fraction of well-described PPIs [64]. Gilad and colleagues [65] used it for a large-scale analysis of PPIs within the programmed cell death network of mammalian cells. They used a library composed of 63 proteins from the autophagy and apoptosis machineries, as well as regulatory proteins. By screening all possible combination pairs, they identified 46 previously unknown interactions within this network. The GLuc PCA also served as an orthogonal assay to validate the human reference interactome [64]. NanoLuc PCAs show very good performances [36,66,67] and are very well suited for high-throughput studies. Moreover, contrary to GLuc, NanoLuc produces a stable luminescence signal, which enables investigating the temporal dynamics of PPIs in near real-time. This property has recently been used to profile the dissociation of a family of heterotrimeric G proteins after stimulation by various GPCR ligands [68]. Thus, NanoLuc PCAs open the perspective of revealing the dynamics of PPI networks in response to varying physiological conditions. To this end, we have constructed genome-wide libraries of yeast strains tagged with the NanoBiT PCA fragments (Table 1) ([69] and unpublished results).

**PPI in diseases, modulation by chemical compounds**

A large fraction of disease-associated mutations have been found to perturb PPIs [70]. PCAs can serve to investigate the functional impact of such mutations and resolve disease mechanisms as exemplified in recent studies of mutations causing mitochondrial diseases [71] and Alport Syndrome [72]. In addition, PPIs are potential drug targets for many diseases. PCAs can be applied at all stages of the drug discovery process from target identification to preclinical validation. One of the assets of PCAs is that they can be applied to probe very

---

**Figure 5. Use of PCAs to screen PPI modulators.**

(a) Reversible PCAs are well-suited to investigate PPI modulators. (b) Screening of a library of compounds to identify molecules that target a given PPI of interest. (c) Screening of an array of PPIs to analyse how PPI networks are modulated by a given compound.
diverse therapeutically relevant PPIs in a single setup. For example, Li and colleagues [73] established a simple platform, named ReBiL, based on the reversible firefly luciferase PCA. They used it to probe five different PPIs (UBE2T/FANCL, p53/Mdm2, p53/Mdm4, Mdm2/Mdm4 and BRCA1/BARD1) and to elucidate the mechanism of action of small-molecule- and peptide-based compounds. Several other drug screening platforms have also been successfully established with various luciferase PCAs [74–79]. In general, luciferase-based PCAs are well-suited to perform large-scale drug screens (Figure 5) because of their sensitivity and reversibility. Sensitivity is important to miniaturise the assay and perform it in a high-throughput format, while reversibility is critical to accurately monitor the inhibitory effects of drugs on PPIs. Using such a platform, a recent study screened ∼45 000 bioactive molecules to identify compounds that directly inhibit dimerisation of the NF-κB protein p65 [80]. This enabled to identify withaferin A, a previously characterised molecule with anti-inflammatory and anticancer activities, as allosteric inhibitor of NF-κB dimerisation. In another study, ∼100 000 pure natural plant molecules were screened to identify inhibitors of Spt5, a disease-associated transcription elongation factor that directly interacts with Rpβ1, the large subunit of RNA polymerase II [81]. The authors identified several compounds that interfere with Spt5/Rpβ1 PCA signal and Spt5 functions in cells. Interestingly, in vitro experiments revealed that these compounds likely do not inhibit Spt5/Rpβ1 interaction but rather induce conformational changes that modulate Spt5 activity. These studies illustrate that PCAs can be used to identify various classes of molecules targeting disease-related PPIs and to accelerate the development of new therapeutic strategies.

PCAs can also be used to profile the cellular response to drugs by examining their effect on a range of PPIs (Figure 5). After the development of the first PCAs, Remy and Michnick [82] examined, in a proof of concept study, the effect of two immunosuppressants, wortmannin and rapamycin, on a set of 35 PPIs involved in a signal transduction pathway controlling translation initiation. A later study [83], selected 49 PPIs involved in ten different cellular processes (e.g. cell cycle control, apoptosis, ubiquitin-mediated proteolysis) and analysed their response to 107 different drugs from six therapeutic areas (cancer, inflammation, cardiovascular disease, diabetes, neurological disorders and infectious disease). The results obtained were remarkably informative: related drugs displayed similar activity profiles and unexpected off-target effects of certain drugs could be revealed. The barcoded DHFR PCA has also been employed in pooled cultures to analyse the impact of 80 small molecules on 384 yeast PPIs, revealing an unexpected effect of the anticancer drug doxorubicin on transcription [52]. More recently, Stynen and colleagues [84] mapped the effect of two drugs, the immunosuppressant rapamycin and the type 2 diabetes drug metformin, on 3500 yeast proteins. Although metformin has no established direct target [85], they observed that it interferes with the distribution of iron in the cell. These results indicate that large-scale analysis of PPIs can enable the discovery of unpredicted effects of drugs, and thereby is useful to facilitate drug development.

Conclusion

PCAs have shown broad applicability in a wide range of PPI detection and drug screening investigations across different cells and organisms. Protocols using PCAs in living cells may not only be applied to study individual PPIs but also a high number of distinct PPIs in large-scale assays. The availability of multiple PCA reporters with complementary properties and different readouts give flexibility in the choice of the most appropriate system according to research needs. Thus PCAs are, together with orthogonal biochemical and other approaches, methods of choice to build high-quality protein interaction networks. This is important, not only to understand how proteins operate collectively to achieve cellular functions, but also to reveal how their perturbation leads to pathological states and how they can be restored by therapies. Hence, large-scale PCAs studies are helpful in deciphering genotype to phenotype relationships and resolving the molecular mechanisms underlying diseases and drug action.

Perspectives

- PPIs play an ubiquitous and fundamental role in most biological processes. They contribute to the development of various diseases and are considered as possible therapeutic targets. Identifying PPIs provides information on the function and regulation of individual proteins and their integration in biological networks.
PCAs can not only detect stable but also weak and transient PPIs in living cells. They can be applied in large-scale studies to profile PPI networks and screen PPI modulators.

Future developments in PCA strategies are expected to facilitate the investigation of numerous PPIs in parallel as well as their dynamics upon physiological perturbations.

Competing Interests
The authors declare that there are no competing interests associated with the manuscript.

Funding
This work was supported by the National Science Centre (NCN), Poland, under project no. 2016/21/D/NZ1/00285 to E.B. Also, E.B. acknowledges the French Government and the Embassy of France in Poland, and the Foundation for Polish Sciences (FNP). N.L. received the French Government Scholarship (BGF) to perform a joint-PhD project. G.R. was supported by the ANR grant ANR-16-CE11-0021-01 and the Institut National de la Santé et de la Recherche Médicale. A.S. was supported by an ARED grant from the Région Bretagne.

Author Contributions
All authors (E.B., N.L., A.S., R.W. and G.R.) contributed to the writing and editing of the manuscript and the figures. E.B. and G.R. conceptualised, wrote the main draft and finalised the manuscript, N.L. prepared the figures.

Abbreviations
BiFC, bimolecular fluorescence complementation; BRET, bioluminescence resonance energy transfer; CFP, cyan fluorescent protein; DHFR, dihydrofolate reductase; FCCS, fluorescence cross-correlation spectroscopy; FLuc, firefly luciferase; FRET, Förster resonance energy transfer; GFP, green fluorescent protein; GLuc, Gaussia luciferase; GPCR, G protein-coupled receptor; LgBiT, large NanoBiT fragment; NanoBiT, NanoLuc binary technology; ORF, open reading frame; PCA, protein-fragment complementation assay; POI, protein of interest; PPI, protein–protein interaction; PPiSeq, protein interaction sequencing; RLuc, Renilla luciferase; SmBiT, small NanoBiT fragment; SUMO, small ubiquitin-like modifier; Y2H, yeast two-hybrid; YFP, yellow fluorescent protein.

References


