A protein-fragment complementation assay to quantify synthetic protein scaffold efficiency

Pascale Lemieux[1-4], Alexandre K Dubé[1-4], and Christian R Landry[1-4]

[1] Institut de Biologie Intégrative et des Systèmes (IBIS), 1030, avenue de la Médecine, Université Laval, Québec (Québec), Canada, G1V 0A6

[2] Regroupement Québécois de Recherche sur la Fonction, l'Ingénierie et les Applications des Protéines, (PROTEO), 1045 Avenue de la Médecine, Université Laval, Québec (Québec), Canada, G1V 0A6

[3] Centre de recherche en données massives (CRDM), 1065, avenue de la Médecine, Université Laval, Québec (Québec), Canada, G1V 0A6

[4] Département de biochimie, microbiologie et bio-informatique, 1045 Avenue de la Médecine, Université Laval, Québec (Québec), Canada, G1V 0A6

Authors for correspondence: pascale.lemieux.4@ulaval.ca

Keywords: synthetic protein scaffold; DHFR PCA; SH3 domain; PDZ domain

Abstract

Scaffolds are powerful tools in synthetic biology for various applications, from increasing yield to optimizing signalling specificity. Scaffolds can be built by fusing peptide binding domains (PBD) and attaching the peptide they bind to the enzymes inducing spatial proximity. Only a few PBD-peptide combinations have been tested in this context, and no combination produced a high yield in yeast, an important chassis in biotechnology. Therefore, there is a need for more exploration of PBD-peptide pairs to be used in yeast. Scaffold characterization is challenging because it is often dependent on a model pathway with an output that is difficult to measure quantitatively. Here, we use a protein-fragment complementation assay (PCA) to study scaffold efficiency in yeast which allows to couple scaffold efficiency with growth rate. First, we characterize PBD-peptide interaction (PPI) strength and their binding availability. Then, we test different scaffold architectures and expression levels to quantify the simultaneous binding of peptide pairs to the scaffold. We show that PPI strength of the weakest binding PDB-peptide pair is critical for scaffold efficiency and that PPI strength is limited by low binding availability of some domains and peptides. Also, we find that slight architectural variations and expression levels have a significant impact on scaffold efficiency detected by DHFR PCA. In the future, DHFR PCA method will enable testing in a high throughput manner scaffold variants in yeast through its easy-to-read scaffolding efficiency signal.

Introduction

One challenge in synthetic biology is the optimization of efficiency and specificity of synthetic pathways[1]. In nature, we often observe highly efficient pathways in which proteins operate in close proximity, thus accelerating signal transduction or inducing metabolic channeling. Such endogenous pathways involve the colocalization of enzymes by direct physical interactions[2,3] or via a scaffold molecule[4–6], creating large protein complexes. In such cases, protein stoichiometry and interaction specificity are controlled to maintain function and correct protein complex assembly[7,8]. The regulated spatial proximity is also important to prevent intermediary metabolites from diffusing in the cytoplasm and slowing down the pathway[9,10]. The multiple layers of control and the spatial organisation we observe for endogenous pathways are absent when we insert a synthetic pathway in a host cell.

Since the proteins of synthetic pathways are often exogenous to the host, they do not have a constrained localization in the cytoplasm encoded by their sequence, which is foreign to the host. Therefore, the enzymes of a synthetic metabolic pathway are scattered in the cytoplasm and depend on the diffusion of intermediary metabolites to produce the end molecule. The crowding of the cellular environment could slow down intermediary metabolite diffusion between the enzymes of the synthetic pathway even more, further limiting its yield[11]. Also, many pathways produce toxic metabolites that impair the growth of the host cell when the metabolites are allowed to diffuse[12,13]. Intermediary metabolite diffusion is contingent on the lack of proximity with the enzyme catalysing the subsequent reaction of the pathway. One approach tested to compensate for the low yield of synthetic pathways was to over-express the exogenous enzymes[14], but this can hinder cell growth by increasing the metabolic burden from the expression of the synthetic pathway[12,15,16]. Inspired by what we find in nature, another solution is the use of a synthetic scaffold, which artificially induces the proximity of enzymes of a pathway to increase its efficiency[17].

Important features to design good synthetic scaffolds are the effect on enzyme activity of the protein inducing colocalization and the modular control over the recruitment of specific enzymes. One possibility to limit altering enzyme activity and ensure specificity is to construct artificial scaffolds by fusing peptide binding domains (PBD)[17]. Short peptides (~10 a.a.) recognized by the PBDs can be fused to the enzymes of the pathway, which will result in bringing them together in the presence of the scaffold. The short peptides are expected to have a lesser impact on protein function in comparison to large fusion proteins[1], and the intrinsic specificity of the PBDs enables the recognition of their favored peptide fused to enzymes of the pathway ensuring modularity[18,19].

Scaffolding efficiency, mainly determined by the colocalization of enzymes of the synthetic pathway, is achieved when the PBDs simultaneously bind the peptides fused to the enzymes of interest. Therefore, scaffolding efficiency is dependent on the binding strength between the PBD-peptide pairs used in the scaffold. In *Escherichia coli*, this scaffolding strategy was shown to increase the efficiency of the heterologous mevalonate biosynthesis pathway by 77-fold and had success with other synthetic pathways[17,20–22]. However, the same approach

showed a maximal yield increase of 5-fold in *Saccharomyces cerevisiae* (resveratrol biosynthesis), an important model for eukaryotic chassis in synthetic biology[23]. It is hard to identify a better scaffold sequence when very few PBD-peptide pairs have been tested *in vivo* for this application, and it remains challenging to characterize scaffolding efficiency without an easy-to-read output[24]. Previous studies generally used a heterologous pathway to characterize the increase in product yield, i.e pathway efficiency, but this often requires specific metabolite purification methods, and the findings may not be transferable to another pathway[23]. Indeed, this approach does not directly characterize the colocalization between the scaffold components, but rather the effect of the scaffold on a specific pathway efficiency. Thus, there is a need for a method that can directly quantify the ability of a scaffold to bring components together (scaffolding efficiency) without relying on pathway efficiency as a readout.

The choice of heterologous sequences, such as PBDs and peptides, contributing to the scaffold is crucial for its efficiency. The strength of the PBD-peptide interaction *in vivo* is a determinant for the enzymes' colocalization, but these heterologous proteins may also be degraded[25] or interact with the proteome of the cell host in unexpected ways[26]. Both of these mechanisms would reduce the amount of protein able to effectively bind other proteins in the cytosol. We refer to the free amount of protein or peptide able to bind as binding availability. An in-depth characterization of PBD-peptide scaffolds *in vivo*, while taking into account the binding strength and binding availability of each PBD-peptide pair, is crucial for the development and exploration of more synthetic scaffolds, particularly in species used in synthetic biology applications such as yeast.

Here, we present an approach to measure scaffolding efficiency in *S. cerevisiae* with an easy-to-read output. It uses the dihydrofolate reductase protein-fragment complementation assay (DHFR PCA), which is a system with a growth-based read-out that is broadly used for protein-protein interaction quantification[19,27–29]. We use this reporter to characterize scaffolds built on two PBD-peptide pairs. Using heterologous PBD-peptide pairs described previously *in vitro*, we assayed their protein-peptide interaction (PPI) strength and their binding availability *in vivo* with DHFR PCA[30]. Then, we used DHFR PCA to assess the scaffolding efficiency for scaffolds designed with weak and strong binding PBD-peptide pairs. We also tested different scaffold architectures and expression levels. With these methods, we found that protein binding availability can profoundly impact PPI strength estimated *in vivo*. Most importantly, we found that PPI strength, scaffold architecture and scaffold expression levels significantly impact the scaffolding efficiency. Our findings show that this approach is able to capture scaffolding efficiency variation caused by small differences in sequence and scaffold architecture. In the future, it could be applied to a diverse library of PBD-peptide to characterize a large toolbox of sequences for scaffold design in yeast.

Results

PPI strength *in vivo* differs from affinity values.

We searched the literature for *in vitro* characterized PBD-peptide pairs with a wide range of affinity to understand how affinity impacts scaffold efficiency. We were interested primarily in PBDs previously used in synthetic scaffolds for which the affinity to one or multiple peptides was characterized. We selected PBD sequences that include a SH3, a PDZ and a GBD domain, which were used in synthetic scaffolds and shown to increase yield of synthetic pathways in bacteria and in yeast[17,20,23]. We added the well-characterized C-terminal SH3 domain of GRB2p, referred to as GRB2[31,32]. We use capital letter abbreviations to name PBD sequences and lower case abbreviations to name their binding peptides. We selected a binding peptide for the GBD and GRB2 domain, named respectively gbd1 and gab2[31,33]. For the SH3 and PDZ domains, multiple peptides were selected and numbered from the highest to lowest affinity as previously characterized *in vitro*, i.e. sh3.1 is the strongest binder tested, and sh3.4 is the weakest[34,35].

In order to understand how affinity measurements transfer to *in vivo* interaction strength, we used the DHFR PCA to quantify the interaction strength of the selected PBDs and peptides. In this assay, the dihydrofolate reductase (DHFR) sequence is split into two fragments, F[1,2] and F[3]. Both fragments are encoded by a vector on opposite strands. DHFR F[1,2] and F[3] are under the control of the CYC promoter and are fused to a flexible linker (F[1,2]-$(GGGGS)_4$, F[3]-$(GGGGS)_3$-$(GGGAS)_1$). The CYC promoter produces a low constitutive expression[36] of the DHFR fragments, preventing spurious background signal (Supplementary Fig. 1). The PBD and peptide sequences were codon-optimized for expression in yeast and synthesized. The PBD and peptide sequences were cloned, paired or alone, at the C-terminus end of the DHFR fragments-linker in both orientations, i.e. F[1,2]-PBD-F[3]-peptide or F[1,2]-peptide-F[3]-PBD. When both DHFR fragments are colocalized in *S. cerevisiae*, they fold into an active enzyme. Under methotrexate (MTX) selection, the cells grow proportionally to the number of active DHFR reconstituted, which is dependent on the interaction strength between the PBD and peptide fused to the fragments (Fig. 1A)[37,38].

To quantify the interaction strength of each binding pair, we assessed growth rate of yeast strains transformed with the PBD-peptide containing plasmids under MTX selection (Fig. 1A)[27,32]. We computed the area under the curve (AUC) obtained from the growth measurements of each PBD-peptide combination and calculated a PPI score using the comparisons of the constructions with the negative control (empty plasmid, see methods). This relative growth represents the strength of the PCA signal[38].

The strains expressing either a PBD or a peptide alone show a PPI score very similar to the empty plasmid. As expected, a higher score is measured when a PBD and a peptide are both present (Fig. 1B). Since we cloned every PBD-peptide pair in both DHFR orientations, we were able to assess the effect of the DHFR fragments fused to each sequence on the PCA

signal. We observed that the PCA signal obtained with the peptides fused to F[1,2] is significantly lower than when fused to F[3] for most of the PBD-peptide pairs, with the exception of pairs SH3-sh3.1 and SH3-sh3.3 (Supplementary Fig. 2A). Also, we detected that the PDZ-pdz.2 pair has a lower signal than the PDZ-pdz.3 pair in the F[1,2]-peptide-F[3]-PBD orientation, which is surprising because the K_d of the PDZ-pdz.2 interaction is an order of magnitude lower than the K_d of the PDZ-pdz.3 pair (Supplementary Fig. 2A and B). In fact, the DHFR fragment effect is particularly strong on pdz.2 since the F[1,2]-pdz.2-F[3]-PDZ signal is as low as the empty plasmid, but F[1,2]-PDZ-F[3]-pdz.2 signal is similar to F[1,2]-PDZ-F[3]-pdz.3 (Supplementary Fig. 2A).

Using the PPI scores obtained in the F[1,2]-PBD-F[3]-peptide orientation, we compared the PCA signal with affinity values characterized in previous studies[31,33–35]. We detected a wide range of interaction intensity with DHFR PCA, corresponding to signals over multiple orders of magnitude of K_d values (Fig. 1C). However, we did not observe a significant correlation between the PPI scores and the K_d values obtained *in vitro* (Spearman's rank correlation, r = -0.34, p-value= 0.37). The GBD-gbd1 interaction is a visible outlier that causes the lack of correlation (Fig. 1C). Indeed, there is a trend between PPI score and affinity for the SH3 and PDZ domains with their respective binding peptides. One limitation of this analysis is that the K_d values for each PBD were retrieved from different studies, which could cause this discrepancy. Another factor that could create a disparity between PPI scores and the *in vitro* affinity is the cellular environment, which may affect *in vivo* protein abundance and potentially lead to spurious interaction with yeast endogenous proteins or other effects that would diminish the availability of the PDBs and peptides for binding to each other. For this reason, we performed an assay to quantify the binding availability of each PBDs and peptides *in vivo*.

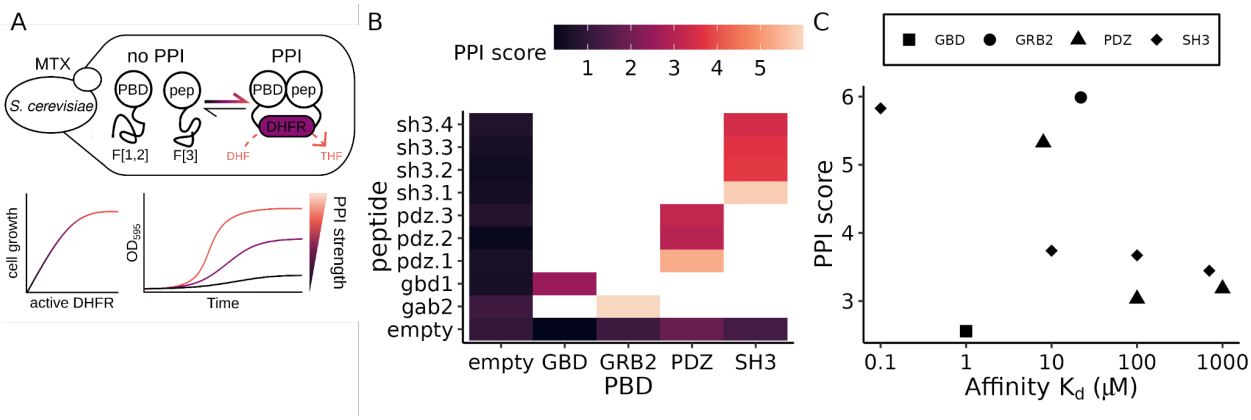


Figure 1. Experimental design and measure of interactions between PBDs and peptides *in vivo*. A. Schematic representation of the DHFR PCA to measure the interaction strength between a PBD and a peptide in living cells[27]. An interaction between DHFR F[1,2]-PBD and DHFR F[3]-peptide, or vice-versa, allows the DHFR to fold into its active form. Cell growth is proportional to the number of active DHFRs, up to signal saturation. We performed growth measurements under MTX selection and computed the area under the curves to quantify PPI strength. B. Pairs of F[1,2]-PBD and F[3]-peptide tested. The color scale indicates the PPI score of the PBD-peptide pairs. White spaces are untested PBD-peptide pairs. C. Comparison of PPI scores obtained with F[1,2]-PBD and F[3]-peptide orientation with the *in vitro* affinities measured

in previous studies[32–35]. The shape of the data points specifies the PBD involved in the binding pair.

PBDs and peptides show different binding availability.

Here, we define the binding availability of a protein as the amount of this protein that is able to effectively bind other proteins in the cytosol[30]. Binding availability can be reduced by the degradation of the protein or by spurious binding with endogenous proteins[25,26]. A strong spurious endogenous binding partner would sequester the heterologous protein and prevent contact with other proteins. In our assays, if a large fraction of a PBD is sequestered by a spurious interaction partner, there is less PBD available to bind with the peptide fused with the complementary DHFR fragment, or vice-versa. This situation could result in a weak PCA signal even if the PBD-peptide interaction is strong. To measure binding availability, we used a binding assay with a non-specific prey fused to DHFR fragments[30]. A yellow fluorescent protein (YFP) fused to a DHFR fragment was expressed under a strong promoter inserted in the genome. We transformed this strain with the plasmids expressing either a PBD or a peptide, fused to the complementary DHFR fragment (Fig. 2A). The resulting PCA signal reflects non-specific binding events between the YFP and the protein of interest, which is proportional to protein abundance (Fig. 2A)[30].

We measured binding availability for all the domains and peptides described previously. We computed the AUC from the growth measurements, and transformed it into an availability score by comparing the AUC values with those from the empty plasmid expressing the DHFR fragments alone, which acts as a positive control (see methods). We found that most proteins show similar, or higher, binding availability compared to the positive control (Fig. 2B). One interesting observation is that the pdz.2 peptide showed a similar availability to the control when fused to the F[3], but a very low availability when fused to the F[1,2] (Fig. 2B). The low availability of the F[1,2]-pdz.2 can explain the weak PPI score obtained when tested against F[3]-PDZ (Supplementary. Fig. 2A). We also observed a small but significant difference in the effect of DHFR fragments on the GBD domain, where F[3]-GBD show less binding availability than F[1,2]-GBD (Student's t-test, p-value = 1.87×10^{-3}, Supplementary Fig. 3). However, the peptide gbd1 fused with both fragments showed similar binding availability to the control. Thus, we were not able to explain the lack of PCA signal in the interaction assay for the F[1,2]-GBD-F[3]-gbd1 pair with the low binding availability of either the PBD or peptide.

The discrepancy between PPI scores and affinity values is affected by binding availability, but it is insufficient to explain every difference. These results highlight the relevance of characterizing heterologous sequences *in vivo* before using them as tools for synthetic biology applications. Following these results, we focused on the SH3 and PDZ domains and their peptides showing similar binding availability to the control and consistent PPI scores with affinity values.

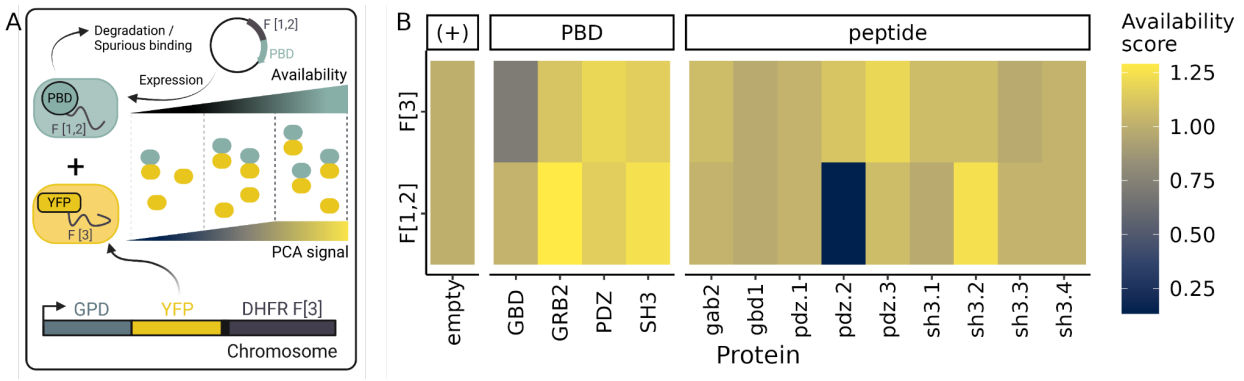


Figure 2. Detection of binding availability of the PDBs and peptides to be used for scaffolding. A. Schematic representation of the experimental approach to assay binding availability with DHFR PCA[30]. The PCA signal depends on the non-specific binding between the heterologous protein expressed from the plasmid and the YFP encoded in the genome. B. For heterologous proteins fused to either DHFR fragments, the availability score is shown with the color scale (see methods). The empty plasmid is considered as positive control (+) in this experiment.

<u>DHFR PCA to detect changes in scaffold efficiency.</u>

After the characterization of PPI strength and binding availability of heterologous proteins *in vivo*, our aim was to use this data to better understand their roles in scaffolding efficiency. We constructed scaffolds by fusing together the SH3 and PDZ domains for which PPIs and binding availability were characterized. We created two scaffold architectures by changing the relative PBD position in the scaffold, namely the orientations SH3-PDZ and PDZ-SH3. We also tested different linker lengths between the PBDs : direct fusion, a flexible linker $(GGGGS)_1$-$(GGGAS)_1$ and a longer flexible 2xlinker $(GGGGS)_3$-$(GGGAS)_1$. We inserted the scaffold variants at the *GAL1* locus. The expression of the scaffold is regulated by a synthetic transcription factor (GEM), which translocates in the nucleus and induces transcription of the gene following the *GAL1* promoter when bound to β-estradiol. The synthetic expression system enables the scaffold to be expressed in a linear fashion with the concentration of β-estradiol in the media[28,39] (Fig. 3A). Then, we used the same expression vector as above, but this time we inserted pairs of peptides in frame with the DHFR fragments. Scaffolding would be apparent by the DHFR fragments being brought together and folding into its active state through the scaffold binding to the two peptides. We transformed each plasmid in the strains containing the different scaffold constructs.

In PCA selection media supplemented with β-estradiol, we expect yeast growth only when the F[1,2]-peptide and F[3]-peptide are simultaneously bound by the scaffold, which allows the complete DHFR to fold (Fig. 3A). To test how PCA signal depends on the abundance of the scaffold, we can modulate its expression level by changing the β-estradiol concentration while preserving expression within the dynamic range of the assay. As with the PPI and binding availability assay, we performed growth measurements to quantify the scaffold efficiency and transformed the AUC into a scaffold score, which is a relative value compared to the empty plasmid.

We found that PBD-peptide affinity, scaffold architecture and scaffold expression level all impact scaffold efficiency. Unsurprisingly, scaffold scores generally increase as expression levels rise by adding β-estradiol to the media, but only when two high-affinity PBD-peptide pairs are used in the scaffold (Fig. 3B, Supplementary Fig. 4). We observe that only the peptide pair with the strongest affinity to the PBDs present in the scaffold (sh3.1-pdz.1) in both fragment orientations generated a high scaffold score at 20nM β-estradiol (Fig. 3C). The pairs including at least one lower affinity peptide all showed similar signals to the empty plasmid, indicating that the scaffold efficiency is limited by the weakest interaction involved in the system. In addition, we tested combinations of peptides binding to the same PBD (ex. F[1,2]-sh3.1-F[3]-sh3.1) and they all showed PCA signals similar to the empty plasmid (Supplementary Fig. 5). This result indicates that two strongly binding and distinct PBD-peptide pairs are necessary for the proper function of the scaffold.

We also found that the scaffold architecture impacts the scaffold score. Indeed, the PDZ-SH3 is the most efficient architecture but, when the positions of the PBDs are swapped (SH3-PDZ), scaffold efficiency is significantly lower for the same peptide pair (F[1,2]-sh3.1-F[3]-pdz.1, Supplementary Fig. 5, Student's t-test, p-value = 0.031). The length of the linker also unpredictably affects the scaffolding for the F[1,2]-sh3.1-F[3]-pdz.1 pair. We observe an intermediate score for SH3-PDZ architecture, then an increase in signal for SH3-linker-PDZ and a lower score for the SH3-2xlinker-PDZ. The opposite pattern is observed for the reverse PBD position, i.e. decrease in signal for PDZ-linker-SH3 compared to PDZ-SH3 (Fig. 3C). All of the design features tested impact the scaffolding efficiency, which highlights the relevance to characterizing in depth variants of a scaffold before choosing one for practical applications.

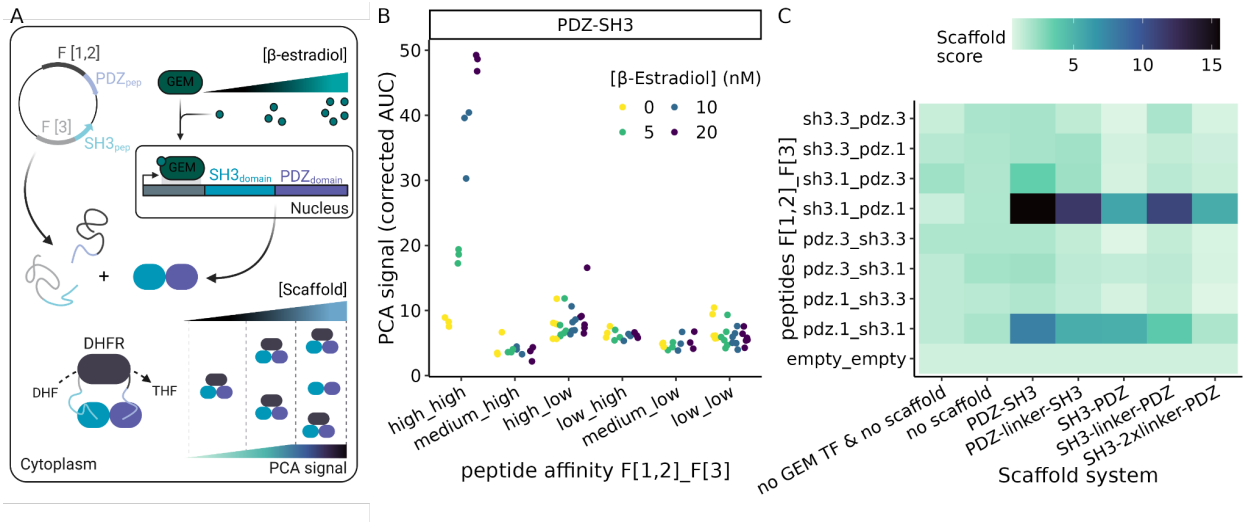


Figure 3. Scaffold efficiency as measured by DHFR PCA. A. Schematic representation of the experimental system to measure scaffold efficiency. The scaffold, a fusion of the SH3 and the PDZ domains, is expressed from the genome at an inducible locus controlled by the concentration of β-estradiol in the media. β-estradiol binds the GEM transcription factor, allowing its translocation in the nucleus and the induction of the scaffold expression[39]. From a plasmid, the DHFR fragments fused to peptides are expressed. PCA signal reflects the ability of the scaffold to colocalize both DHFR fragments by simultaneously binding the peptides. B. For the PDZ-SH3 scaffold architecture, the relation between the PBD-peptide affinity used in the scaffold system (high : K_d < 10 µM, medium : 10 µM <= K_d < 100 µM, low : 100 µM <= K_d) and the PCA signal (corrected AUC, see methods) is displayed for different concentrations of β-estradiol. The colors show the β-estradiol concentration used. C. For different combinations of peptides and scaffold architectures, the scaffold score obtained at 20nM β-estradiol is shown (see methods).

Discussion

Many experimental approaches depend on a pathway output to quantify a scaffold's ability to induce colocalization of the targeted proteins (scaffold efficiency). Such an approach has limited the throughput to test various scaffold architectures and sequences[24]. Thus, there was a need to develop a strategy to study scaffold efficiency with an easy-to-read output. Using DHFR PCA, we characterized the PPI strength and binding availability of PBD-peptide pairs. Additionally, we showed that the position of the PBDs, linker length, PBD-peptide affinity and scaffold expression level play a significant role in the scaffolding efficiency.

A growing number of synthetic protein scaffolds are available due to this strategy's high potential. PBD based scaffolds are only one of the designs available. Other protein scaffold designs use relatively large chimeric proteins, such as the cohesin-dockerin systems (~70 a.a.)[40], that may be damaging for enzyme activity[1,41]. A limitation of the cohesin-dockerin systems is its calcium dependency[42], which makes it less efficient for intracellular assembly because cytosolic calcium concentration is low[43]. Novel designs use inducible protein condensate, which do not, yet, offer control on interaction specificity between the proteins involved in the condensate formation[44]. Thus, the PBD-peptide scaffold strategy is an interesting

option since the small peptide (~10 a.a.) fusion on the enzyme is expected to have limited to no damaging effect on its activity and folding, and the intrinsic specificity of the PBDs ensures specific recruitment in the cytosol. However, as we have seen with F[1,2]-pdz.2, peptides can affect protein availability and impact interaction strength. We also observed that the protein fused to a PBD can change its binding availability. For instance, we have seen that the GBD domain's availability can change depending on the attached DHFR fragment and that its interaction strength with the gbd1 peptide in yeast is lower than expected based on its K_d. These observations could explain why the synthetic scaffold using the pair GBD-gbd1 showed limited yield increase in yeast[23].

One approach to increase synthetic pathway efficiency often compared with synthetic scaffold is the direct fusion of proteins of a pathway. At times, directly fusing enzymes showed increases in product yield in comparison with the separate enzymes and even, in some cases, with a scaffold approach[21,45,46]. However, directly fusing the enzymes can be unfavorable for protein folding or enzymatic activity. The PBD scaffold approach was shown to be more efficient than enzyme fusions, for instance, in the biosynthesis of resveratrol[23,47]. Indeed, depending on the pathway of interest, a scaffold could be more efficient than a fusion, or vice versa. However, a scaffold strategy is more versatile than a direct fusion, because we can change the PBD content and the enzyme ratio is not limited by sequence fusion to 1:1[48]. First, we can adapt the number of PBDs in a scaffold to optimize the enzyme ratio. The effect of the ratio of enzymes involved in a pathway is a main determinant of pathway efficiency, since the enzymes in the pathway may have different kinetics and could cause flux imbalance leading to intermediate accumulation[1]. If a toxic intermediate accumulates, it may result in the diffusion of the intermediate in the cytoplasm and cause damage to the host. We can modulate the number of PBD in the scaffold to address the flux imbalance by repeating the PBD sequence, recruiting the flux limiting enzyme and locally increasing the rate of its reaction[17,23]. Then, we can change PBDs contributing to the scaffold for better compatibility with the pathway of interest. For instance, a specific peptide fusion could compromise the enzymatic function or protein availability of one enzyme, but a change of peptide sequence could resolve the issue. Since very few PBD-peptide pairs have been characterized in a scaffold so far, optimizing a scaffold by changing the PBD and peptide sequences is hard. With the DHFR PCA approach, we could resolve this bottleneck by easily characterizing other pairs of PBD-peptide, providing alternative PBD-peptide sequences. Once the toolbox for scaffold design is expanded to more PBD-peptide pairs, the scaffold strategy might outperform the fusion design for most pathways by fully taking advantage of its versatility.

Some limitations of our approach are connected directly with the number of DHFR fragments available and with the expression system we used. As shown by Dueber et al. 2009[17], we found that scaffold expression levels can be used to optimize the output signal induced by a scaffold. To minimize background signal, we used a constitutive low-expression promoter to express the peptides fused to the DHFR fragments (Supplementary Fig. 1). In addition, a low-expression PCA system is favorable because it assures detection below signal saturation and has a limited metabolic burden. If the expression level of the DHFR fragment-peptides fusion was higher, we might have detected scaffolding signals for weaker binding PBD-peptide pairs. In the future, higher expression of the scaffold and DHFR fragment-peptides

fusion could be tested to evaluate the lower limit of detection of the assay and the effect of metabolic burden on cell growth. Another potential limit to our assays is that the DHFR PCA approach to characterize scaffold efficiency is constrained to two-component scaffolds, which is restrictive because most synthetic pathways have more than two enzymes. However, our approach can be used to test if two PBDs are compatible for scaffold design, and then various pairs could be combined together. Indeed, the folding of some PBD fusion could hinder one binding site of the scaffold and lead to an inefficient scaffold even if the PDB-peptides pairs showed strong interactions. Also, the effect of the linker properties[49] between the PBDs could be further investigated using our method, which has not been done, to our knowledge, for PBD-based scaffolds.

To take the PBD scaffold strategy further, the next step would be to characterize a large set of PBD-peptide pairs and different scaffold architectures. In 2018, Li and collaborators made an effort to create a method to construct a DNA library of PBD scaffolds[50]. First, they validated the interaction strength of PBD-peptide pairs using a split-GFP and yeast two-hybrid assay, which is also one measure we performed with DHFR PCA. They measured the effect of different scaffold architecture and composition on the violacein biosynthetic pathway. However, they used the pathway output as measurement of scaffold efficiency, limiting the number of sequences they were able to test simultaneously. Combining their DNA assembly method with DHFR PCA for sequence characterization of PPI strength, binding availability, and scaffold efficiency could unlock the throughput needed to identify more efficient scaffolds.

In the future, the DHFR PCA for scaffold characterization could be adapted to test libraries of PBD-peptide pairs and scaffolds in a high throughput fashion. A screening approach may help identify stronger PBD-peptide binding pairs *in vivo* and more efficient scaffold architectures with the aim of expanding the toolbox for synthetic scaffold design in yeast. Although many parameters of synthetic pathway design may come into play for pathway efficiency, the use of an efficient synthetic scaffold has the potential to be an important building block for effective bioprocesses.

Author contributions

P.L., A.K.D. and C.R.L. were involved in the experimental design. C.R.L. supervised this work. P.L. and A.K.D. designed molecular biology material for this study. P.L. performed the experiments. P.L. completed the analysis and wrote the first draft of the manuscript. P.L., A.K.D. and C.R.L. contributed to writing, revising and approving the final version of the manuscript. C.R.L. and P.L. acquired the funds to support this project.

Acknowledgments

We would like to thank Guillaume Diss who provided key materials and comments for the experimental design. We are thankful for the comments of Soham Dibyachintan and David F. Jordan on the manuscript, and for the support of all other members of the Landry Lab. Some figures were produced using BioRender.

Funding

This work was funded by a Canadian Institutes of Health Research Foundation grant number 387697. C.R.L. holds the Canada Research Chair in Cellular Systems and Synthetic Biology. P.L. was supported by an Alexander Graham Bell PhD scholarship from NSERC (BESC D) and a Leadership and Sustainable Development Scholarship from Université Laval.

Data availability

The DHFR PCA assay results, code for growth curve analysis and figure generation, and all supplementary material is available on the Github repository (https://github.com/Landrylab/Lemieux_et_al2025). The strains and plasmids are available upon request.

Conflict of interest

The authors declare no conflict of interest.

Material & Methods

Plasmid construction

The destination plasmid (pGD110), kindly donated by Guillaume Diss[32], expresses the murine dihydrofolate reductase fragments, DHFR F[1,2] and DHFR F[3], on opposite DNA strands. Both fragments are under the control of the CYC promoter and CYC terminator and fused to a flexible linker in C-terminal (DHFR F[1,2]-$(GGGGS)_4$, DHFR F[3]-$(GGGGS)_3$-$(GGGAS)_1$). pGD110 was digested with BamHI-HF or HindIII-HF (NEB) for the insertion of sequences in frame with, respectively, the F[1,2] or F[3], in N-terminal. The PBDs, gbd1 and gab2 DNA sequences were codon optimized for expression in *S. cerevisiae* and synthesized (Integrated DNA Technology, IDT, Supplementary Table 1). They were amplified with 40-pb homology arms to pGD110 for both DHFR fragment fusion and purified using magnetic beads. Gibson assembly was performed using 50 ng of digested plasmids with an insert molar ratio of 1:3 (backbone : insert) in a 10uL reaction volume[51]. The reactions were incubated at 50°C for 60 minutes and 5uL of the reactions were transformed in MC1061 chemo-competent cells. All transformations were plated on solid 2YT medium with ampicillin (100 µg/ml, see all media composition in Supplementary File 1). The other peptide sequences were codon optimized for expression in *S. cerevisiae* and synthesized as oligonucleotides (Eurofins). They were inserted downstream of both DHFR fragments paired with a PBD or alone, by directed mutagenesis or using fusion PCR combined with Gibson assembly. The mutagenesis products were digested with DpnI and 5uL were used for transformation in MC1061 chemo-competent cells. All the synthetic DNA used for the plasmid construction is described in the Supplementary Tables 1 and 2. All plasmids were validated by Sanger sequencing (CHUL sequencing platform) and they are listed in Supplementary Table 3.

Strain construction

The strains used for the binding availability assay express a YFP-linker$(GGGGS)_2$-DHFR fragment under GPD promoter control. They were constructed by plasmid integration at the HO

locus. Each fragment was PCR amplified with 40-pb homology to each neighbor fragment. The GPD promoter sequence was amplified from pAG416-GPD-ccdb[52], the linker-DHFR fragments were amplified from pAG25-DHFR F[1,2]-NATMX and pAG32-DHFR F[3]-HPHMX[27], the YFP was amplified from 34a_mVenusNB (Michael Springer Lab). The backbone was amplified from pKB1 (See detailed methods, Supplementary File 1). Gibson assembly was performed as described above. The resulting plasmids, pPL1 (pKB1-YFP-linker$(GGGGS)_2$-DHFR F[1,2]-NATMX) and pPL2 (pKB1-YFP-linker$(GGGGS)_2$-DHFR F[3]-HPHMX), were validated by Sanger sequencing. pPL1 and pPL2 were transformed using standard lithium acetate yeast transformation[53] in BY4741 (his3 leu2 ura3 met15) competent cells and plated on selective media (YPD, 100 µg/mL nourseothricin or 250 µg/mL hygromycin) creating the strains PL0017 and PL0019, respectively. To create the strains to perform the scaffold assays, we retrieved a strain expressing the GEM transcription factor from a previous study (PL0001)[28,39]. The scaffold sequences were synthesized (IDT) and PCR amplified with 40-pb homology arms to the *GAL1* locus in 5'-terminal and to the NATMX cassette in 3'-terminal. The resulting DNA fragments were integrated in strain PL0001 by homologous recombination using standard lithium acetate yeast transformation[53] and plated on selective media (YPD, 100 µg/mL nourseothricin). The scaffold strains were validated by PCR and Sanger sequencing (See detailed methods, Supplementary File 1). All the synthetic DNA and plasmids used for the strains construction are described in Supplementary Tables 1-3 and the constructed strains are listed in Supplementary Table 4.

<u>DHFR PCA assay and analysis</u>

The DHFR PCA protocol is based on Tarassov et al. 2008[27]. DHFR PCA selection condition was liquid synthetic media (PCA medium, 0.67% yeast nitrogen base without amino acids and without ammonium sulfate, 2% glucose, drop-out without adenine and uracil) containing 200 µg/ml methotrexate (MTX, BioBasic) diluted in dimethyl sulfoxide (DMSO, Bioshop Canada). Yeast strains were grown in liquid synthetic media (0.174 % yeast nitrogen base without amino acids and without ammonium sulfate, 2% glucose, 0.1% MSG, drop-out without uracil), then diluted to an optical density (OD) at 595 nm of 0.1 in DHFR PCA selection media and incubated in transparent polystyrene 96-well plates (Greiner Bio-One). When needed, β-estradiol (Sigma-Aldrich) diluted in ethanol (GreenField Global) was added to the DHFR PCA selection media at concentrations ranging from 0nM to 20nM. Growth rates were measured in three or four replicates without shaking by measuring the OD at 595 nm in 15-minute intervals over two days of incubation at 30°C in a BioTek Epoch 2 plate reader (Agilent). The median of the first 5 measurements was subtracted from all the following measurements to correct for variation in initial cell inoculate density. Empirical area under the curve (corrected AUC) was computed using the first 40 hours of growth. The corrected AUC metric was chosen over the derivative growth rate because the corrected AUC showed weaker background signal. The PCA scores (PPI, availability and scaffold scores) were obtained by calculating the ratio of the sample average corrected AUC on the control average corrected AUC (F[1,2]-empty-F[3]-empty) in the same strain background and growth condition.

References

(1) Vanderstraeten, J.; Briers, Y. Synthetic Protein Scaffolds for the Colocalisation of Co-Acting Enzymes. *Biotechnol. Adv.* **2020**, *44*, 107627.

(2) Bulutoglu, B.; Garcia, K. E.; Wu, F.; Minteer, S. D.; Banta, S. Direct Evidence for Metabolon Formation and Substrate Channeling in Recombinant TCA Cycle Enzymes. *ACS Chem. Biol.* **2016**, *11* (10), 2847–2853.

(3) Fleming, J. R.; Schupfner, M.; Busch, F.; Baslé, A.; Ehrmann, A.; Sterner, R.; Mayans, O. Evolutionary Morphing of Tryptophan Synthase: Functional Mechanisms for the Enzymatic Channeling of Indole. *J. Mol. Biol.* **2018**, *430* (24), 5066–5079.

(4) Lamson, R. E.; Winters, M. J.; Pryciak, P. M. Ste5 Membrane Localization Allows MAPK Pathway Signaling in Trans Between Kinases on Separate Scaffold Molecules. *bioRxiv*, 2019, 673855. https://doi.org/10.1101/673855.

(5) Zarrinpar, A.; Bhattacharyya, R. P.; Nittler, M. P.; Lim, W. A. Sho1 and Pbs2 Act as Coscaffolds Linking Components in the Yeast High Osmolarity MAP Kinase Pathway. *Mol. Cell* **2004**, *14* (6), 825–832.

(6) Perham, R. N. Swinging Arms and Swinging Domains in Multifunctional Enzymes: Catalytic Machines for Multistep Reactions. *Annu. Rev. Biochem.* **2000**, *69* (1), 961–1004.

(7) Good, M. C.; Zalatan, J. G.; Lim, W. A. Scaffold Proteins: Hubs for Controlling the Flow of Cellular Information. *Science* **2011**, *332* (6030), 680–686.

(8) Taggart, J. C.; Li, G.-W. Production of Protein-Complex Components Is Stoichiometric and Lacks General Feedback Regulation in Eukaryotes. *Cell Syst.* **2018**, *7* (6), 580–589.e4.

(9) Miles, E. W.; Rhee, S.; Davies, D. R. The Molecular Basis of Substrate Channeling. *J. Biol. Chem.* **1999**, *274* (18), 12193–12196.

(10) Zeke, A.; Lukács, M.; Lim, W. A.; Reményi, A. Scaffolds: Interaction Platforms for Cellular Signalling Circuits. *Trends Cell Biol.* **2009**, *19* (8), 364–374.

(11) Zotter, A.; Bäuerle, F.; Dey, D.; Kiss, V.; Schreiber, G. Quantifying Enzyme Activity in Living Cells. *J. Biol. Chem.* **2017**, *292* (38), 15838–15848.

(12) Nicolaou, S. A.; Gaida, S. M.; Papoutsakis, E. T. A Comparative View of Metabolite and Substrate Stress and Tolerance in Microbial Bioprocessing: From Biofuels and Chemicals, to Biocatalysis and Bioremediation. *Metab. Eng.* **2010**, *12* (4), 307–331.

(13) Gao, L.; Wu, X.; Xia, X.; Jin, Z. Fine-Tuning Ethanol Oxidation Pathway Enzymes and Cofactor PQQ Coordinates the Conflict between Fitness and Acetic Acid Production by Acetobacter Pasteurianus. *Microb. Biotechnol.* **2021**, *14* (2), 643–655.

(14) Le Vo, T. D.; Ko, J.-S.; Lee, S. H.; Park, S. J.; Hong, S. H. Overexpression of Neurospora Crassa OR74A Glutamate Decarboxylase in Escherichia Coli for Efficient GABA Production. *Biotechnol. Bioprocess Eng.* **2013**, *18* (6), 1062–1066.

(15) Wu, G.; Yan, Q.; Jones, J. A.; Tang, Y. J.; Fong, S. S.; Koffas, M. A. G. Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends Biotechnol.* **2016**, *34* (8), 652–664.

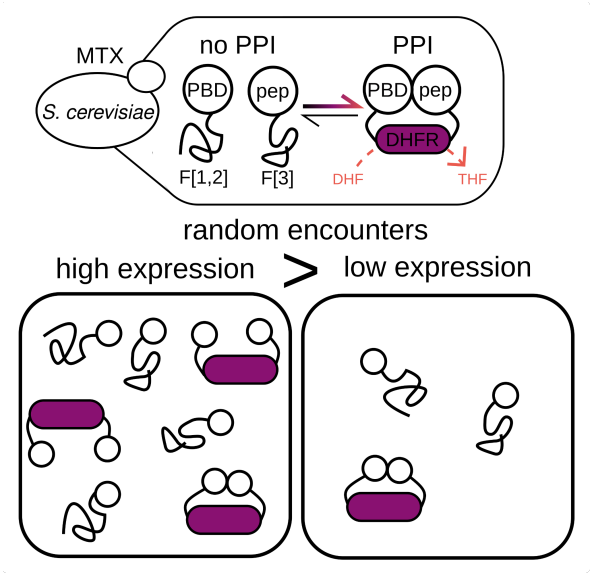
(16) Rajacharya, G. H.; Sharma, A.; Yazdani, S. S. Proteomics and Metabolic Burden Analysis to Understand the Impact of Recombinant Protein Production in E. Coli. *Sci. Rep.* **2024**, *14* (1), 12271.

(17) Dueber, J. E.; Wu, G. C.; Malmirchegini, G. R.; Moon, T. S.; Petzold, C. J.; Ullal, A. V.; Prather, K. L. J.; Keasling, J. D. Synthetic Protein Scaffolds Provide Modular Control over Metabolic Flux. *Nat. Biotechnol.* **2009**, *27* (8), 753–759.

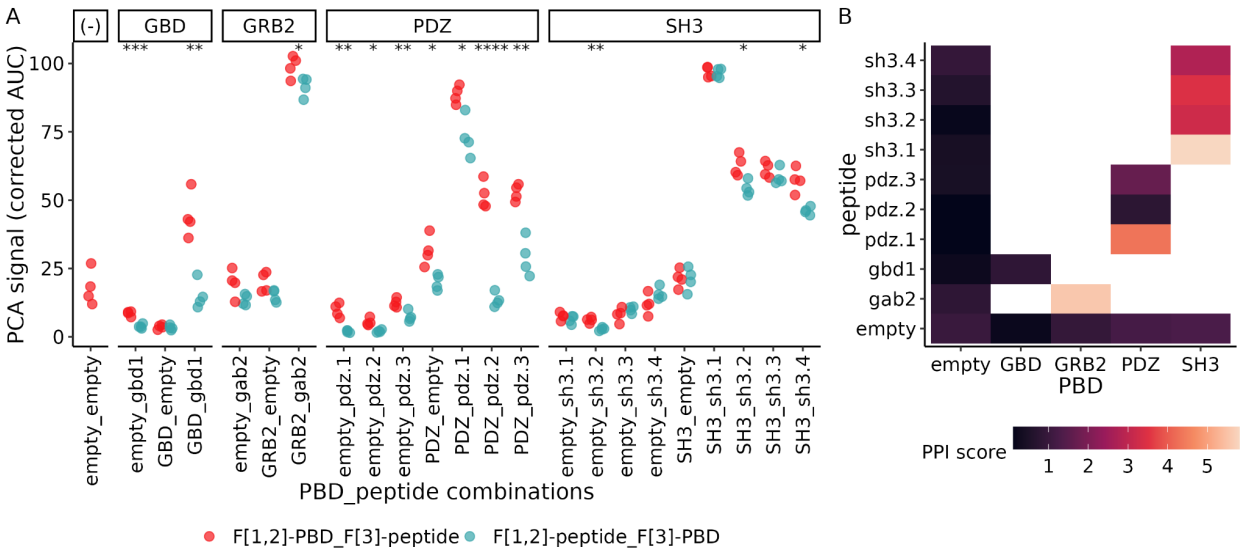
(18) Tonikian, R.; Xin, X.; Toret, C. P.; Gfeller, D.; Landgraf, C.; Panni, S.; Paoluzi, S.; Castagnoli, L.; Currell, B.; Seshagiri, S.; Yu, H.; Winsor, B.; Vidal, M.; Gerstein, M. B.; Bader, G. D.; Volkmer, R.; Cesareni, G.; Drubin, D. G.; Kim, P. M.; Sidhu, S. S.; Boone, C. Bayesian Modeling of the Yeast SH3 Domain Interactome Predicts Spatiotemporal Dynamics of Endocytosis Proteins. *PLoS Biol.* **2009**, *7* (10), e1000218.

(19) Dionne, U.; Bourgault, É.; Dubé, A. K.; Bradley, D.; Chartier, F. J. M.; Dandage, R.; Dibyachintan, S.; Després, P. C.; Gish, G. D.; Pham, N. T. H.; Létourneau, M.; Lambert, J.-P.; Doucet, N.; Bisson, N.; Landry, C. R. Protein Context Shapes the Specificity of SH3 Domain-Mediated Interactions in Vivo. *Nat. Commun.* **2021**, *12* (1), 1597.
(20) Moon, T. S.; Dueber, J. E.; Shiue, E.; Prather, K. L. J. Use of Modular, Synthetic Scaffolds for Improved Production of Glucaric Acid in Engineered E. Coli. *Metab. Eng.* **2010**, *12* (3), 298–305.
(21) Agapakis, C. M.; Ducat, D. C.; Boyle, P. M.; Wintermute, E. H.; Way, J. C.; Silver, P. A. Insulation of a Synthetic Hydrogen Metabolism Circuit in Bacteria. *J. Biol. Eng.* **2010**, *4*, 3.
(22) Baek, J.-M.; Mazumdar, S.; Lee, S.-W.; Jung, M.-Y.; Lim, J.-H.; Seo, S.-W.; Jung, G.-Y.; Oh, M.-K. Butyrate Production in Engineered Escherichia Coli with Synthetic Scaffolds. *Biotechnol. Bioeng.* **2013**, *110* (10), 2790–2794.
(23) Wang, Y.; Yu, O. Synthetic Scaffolds Increased Resveratrol Biosynthesis in Engineered Yeast Cells. *J. Biotechnol.* **2012**, *157* (1), 258–260.
(24) Wang, S.-Z.; Zhang, Y.-H.; Ren, H.; Wang, Y.-L.; Jiang, W.; Fang, B.-S. Strategies and Perspectives of Assembling Multi-Enzyme Systems. *Crit. Rev. Biotechnol.* **2017**, *37* (8), 1024–1037.
(25) Zhang, Y.; Liu, R.; Wu, X. The Proteolytic Systems and Heterologous Proteins Degradation in the Methylotrophic yeastPichia Pastoris. *Ann. Microbiol.* **2007**, *57* (4), 553–560.
(26) Zarrinpar, A.; Park, S.-H.; Lim, W. A. Optimization of Specificity in a Cellular Protein Interaction Network by Negative Selection. *Nature* **2003**, *426* (6967), 676–680.
(27) Tarassov, K.; Messier, V.; Landry, C. R.; Radinovic, S.; Serna Molina, M. M.; Shames, I.; Malitskaya, Y.; Vogel, J.; Bussey, H.; Michnick, S. W. An in Vivo Map of the Yeast Protein Interactome. *Science* **2008**, *320* (5882), 1465–1470.
(28) Lemieux, P.; Bradley, D.; Dubé, A. K.; Dionne, U.; Landry, C. R. Dissection of the Role of a Src Homology 3 Domain in the Evolution of Binding Preference of Paralogous Proteins. *Genetics* **2024**, *226* (1), iyad175.
(29) Michnick, S. W.; Levy, E. D.; Landry, C. R.; Kowarzyk, J.; Messier, V. The Dihydrofolate Reductase Protein-Fragment Complementation Assay: A Survival-Selection Assay for Large-Scale Analysis of Protein-Protein Interactions. *Cold Spring Harb. Protoc.* **2016**, *2016* (11). https://doi.org/10.1101/pdb.prot090027.
(30) Levy, E. D.; Kowarzyk, J.; Michnick, S. W. High-Resolution Mapping of Protein Concentration Reveals Principles of Proteome Architecture and Adaptation. *Cell Rep.* **2014**, *7* (4), 1333–1340.
(31) Malagrinò, F.; Coluccia, A.; Bufano, M.; Regina, G. L.; Puxeddu, M.; Toto, A.; Visconti, L.; Paone, A.; Magnifico, M. C.; Troilo, F.; Cutruzzolà, F.; Silvestri, R.; Gianni, S. Targeting the Interaction between the SH3 Domain of Grb2 and Gab2. *Cells* **2020**, *9* (11). https://doi.org/10.3390/cells9112435.
(32) Faure, A. J.; Domingo, J.; Schmiedel, J. M.; Hidalgo-Carcedo, C.; Diss, G.; Lehner, B. Mapping the Energetic and Allosteric Landscapes of Protein Binding Domains. *Nature* **2022**, *604* (7904), 175–183.
(33) Kim, A. S.; Kakalis, L. T.; Abdul-Manan, N.; Liu, G. A.; Rosen, M. K. Autoinhibition and Activation Mechanisms of the Wiskott-Aldrich Syndrome Protein. *Nature* **2000**, *404* (6774), 151–158.
(34) Harris, B. Z.; Brian J. Hillier, ‡. And; Lim*, W. A. Energetic Determinants of Internal Motif Recognition by PDZ Domains. **2001**. https://doi.org/10.1021/bi010142l.
(35) Dueber, J. E.; Yeh, B. J.; Chak, K.; Lim, W. A. Reprogramming Control of an Allosteric Signaling Switch through Modular Recombination. *Science* **2003**, *301* (5641), 1904–1908.
(36) Peng, B.; Williams, T. C.; Henry, M.; Nielsen, L. K.; Vickers, C. E. Controlling Heterologous Gene Expression in Yeast Cell Factories on Different Carbon Substrates and across the Diauxic Shift: A Comparison of Yeast Promoter Activities. *Microb. Cell Fact.* **2015**, *14* (1),

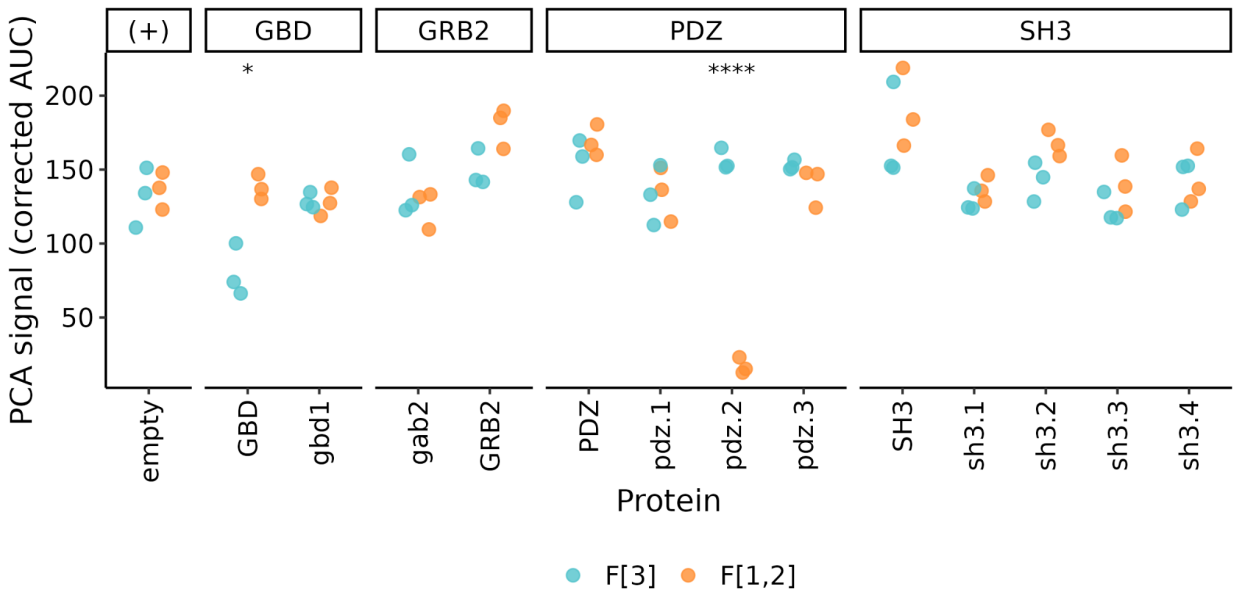
91.
(37) Pelletier, J. N.; Arndt, K. M.; Plückthun, A.; Michnick, S. W. An in Vivo Library-versus-Library Selection of Optimized Protein-Protein Interactions. *Nat. Biotechnol.* **1999**, *17* (7), 683–690.
(38) Freschi, L.; Torres-Quiroz, F.; Dubé, A. K.; Landry, C. R. qPCA: A Scalable Assay to Measure the Perturbation of Protein-Protein Interactions in Living Cells. *Mol. Biosyst.* **2013**, *9* (1), 36–43.
(39) Aranda-Díaz, A.; Mace, K.; Zuleta, I.; Harrigan, P.; El-Samad, H. Robust Synthetic Circuits for Two-Dimensional Control of Gene Expression in Yeast. *ACS Synth. Biol.* **2017**, *6* (3), 545–554.
(40) Kim, S.; Hahn, J.-S. Synthetic Scaffold Based on a Cohesin–dockerin Interaction for Improved Production of 2,3-Butanediol in Saccharomyces Cerevisiae. *J. Biotechnol.* **2014**, *192 Pt A*, 192–196.
(41) Siddiqui, M. S.; Thodey, K.; Trenchard, I.; Smolke, C. D. Advancing Secondary Metabolite Biosynthesis in Yeast with Synthetic Biology Tools. *FEMS Yeast Res.* **2012**, *12* (2), 144–170.
(42) Stahl, S. W.; Nash, M. A.; Fried, D. B.; Slutzki, M.; Barak, Y.; Bayer, E. A.; Gaub, H. E. Single-Molecule Dissection of the High-Affinity Cohesin-Dockerin Complex. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (50), 20431–20436.
(43) Miseta, A.; Kellermayer, R.; Aiello, D. P.; Fu, L.; Bedwell, D. M. The Vacuolar Ca2+/H+ Exchanger Vcx1p/Hum1p Tightly Controls Cytosolic Ca2+ Levels in S. Cerevisiae. *FEBS Lett.* **1999**, *451* (2), 132–136.
(44) Qian, Z.-G.; Huang, S.-C.; Xia, X.-X. Synthetic Protein Condensates for Cellular and Metabolic Engineering. *Nat. Chem. Biol.* **2022**, *18* (12), 1330–1340.
(45) Ding, Y.-W.; Lu, C.-Z.; Zheng, Y.; Ma, H.-Z.; Jin, J.; Jia, B.; Yuan, Y.-J. Directed Evolution of the Fusion Enzyme for Improving Astaxanthin Biosynthesis in Saccharomyces Cerevisiae. *Synth Syst Biotechnol* **2023**, *8* (1), 46–53.
(46) Aalbers, F. S.; Fraaije, M. W. Enzyme Fusions in Biocatalysis: Coupling Reactions by Pairing Enzymes. *Chembiochem* **2019**, *20* (1), 20–28.
(47) Zhang, Y.; Li, S.-Z.; Li, J.; Pan, X.; Cahoon, R. E.; Jaworski, J. G.; Wang, X.; Jez, J. M.; Chen, F.; Yu, O. Using Unnatural Protein Fusions to Engineer Resveratrol Biosynthesis in Yeast and Mammalian Cells. *J. Am. Chem. Soc.* **2006**, *128* (40), 13030–13031.
(48) Qiu, X.-Y.; Xie, S.-S.; Min, L.; Wu, X.-M.; Zhu, L.-Y.; Zhu, L. Spatial Organization of Enzymes to Enhance Synthetic Pathways in Microbial Chassis: A Systematic Review. *Microb. Cell Fact.* **2018**, *17* (1), 120.
(49) Chen, X.; Zaro, J. L.; Shen, W.-C. Fusion Protein Linkers: Property, Design and Functionality. *Adv. Drug Deliv. Rev.* **2013**, *65* (10), 1357–1369.
(50) Li, T.; Chen, X.; Cai, Y.; Dai, J. Artificial Protein Scaffold System (AProSS): An Efficient Method to Optimize Exogenous Metabolic Pathways in Saccharomyces Cerevisiae. *Metab. Eng.* **2018**, *49*, 13–20.
(51) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A., 3rd; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat. Methods* **2009**, *6* (5), 343–345.
(52) Alberti, S.; Gitler, A. D.; Lindquist, S. A Suite of Gateway Cloning Vectors for High-Throughput Genetic Analysis in Saccharomyces Cerevisiae. *Yeast* **2007**, *24* (10), 913–919.
(53) Gietz, R. D.; Schiestl, R. H. High-Efficiency Yeast Transformation Using the LiAc/SS Carrier DNA/PEG Method. *Nat. Protoc.* **2007**, *2* (1), 31–34.



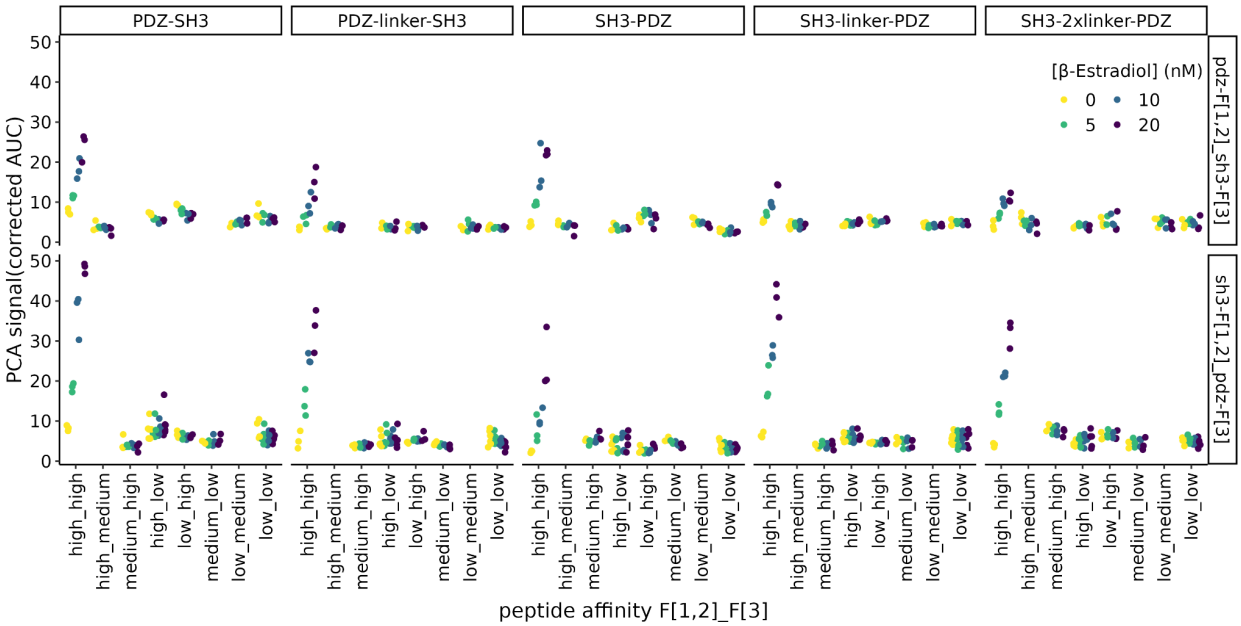
Supplementary Figure 1. Effect of the expression level of the DHFR fragments on PCA background signal. Schematic representation of the DHFR PCA to measure the interaction strength between a PBD and a peptide. The random encounters between the proteins fused to the DHFR fragments can form an active DHFR, generating PCA signal. If the expression level of the DHFR fragments is high, more random encounters happen, thus increasing the PCA signal even if the interaction between the two proteins fused is weak. Also, DHFR PCA can reach signal saturation faster when the background signal is high, reducing the dynamic range of the assay. With a lower level of expression of the DHFR fragments, there is less random encounter and the signal range is reflecting the interaction strength between the proteins fused.



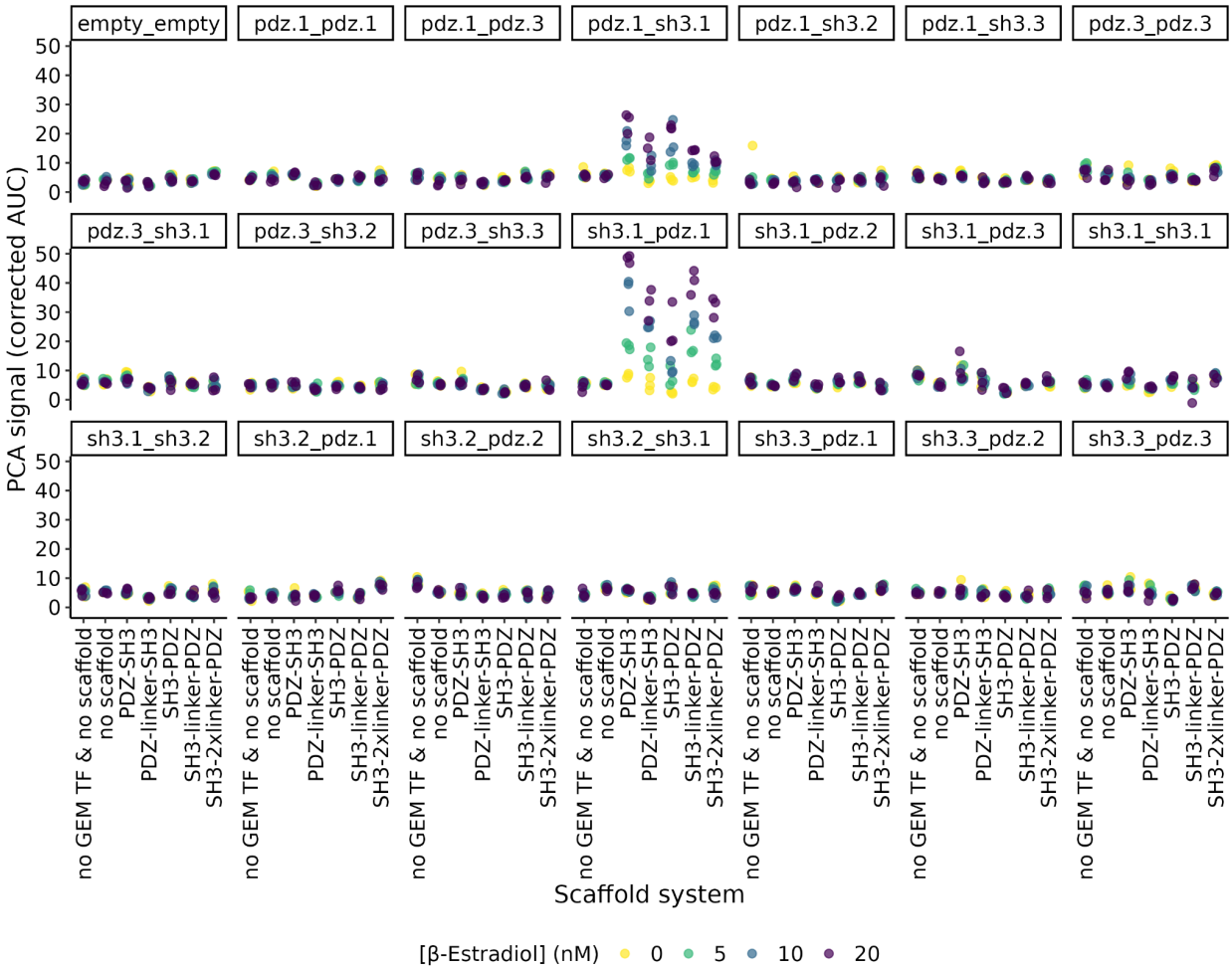
616 Supplementary Figure 2. Detailed DHFR PCA results of the PPI assay. A. For each PBD and
617 peptide, paired or alone, the corrected area under the curve of DHFR PCA assay is shown for
618 each replicate. Empty refers to no sequence inserted as the PBD or the peptide. The color
619 represents the DHFR fragment orientation of the combination. Student's t-tests were performed
620 between the orientations for each sequence combination and the p-values significance levels
621 are labeled above each comparison (* : p <= 0.05, ** : p <= 0.01, *** : p <= 0.001, **** : p <=
622 0.0001). B. Pairs of F[3]-PBD and F[1,2]-peptide tested with DHFR PCA. The color intensity
623 indicates the interaction strength between the pairs compared to the negative control (PPI
624 score, see methods).



Supplementary Figure 3. Detailed DHFR PCA results of the binding availability assay. The corrected area under the curve of the availability assay is shown for each sequence tested in triplicate. The color shows the DHFR fragment fused to the proteins. Student's t-tests were performed to assay the DHFR fragment effect for each sequence and the p-values significance levels are labeled above each comparison (* : p <= 0.05, **** : p <= 0.0001). The empty plasmids are considered as positive controls (+) in this experiment.



631 Supplementary Figure 4. PBD-peptide affinity effect on scaffold efficiency at different scaffold
632 expression levels. The corrected area under the curve is shown for each scaffold variant,
633 peptide pairs and β-estradiol concentration tested in triplicates. The affinity of the PBD-peptide
634 pairs were classified by affinity strength (high : K_d < 10 µM, medium : 10 <= K_d < 100 µM, low :
635 100 µM <= K_d) and the different peptides combinations are labelled on the x-axis. The top
636 panels described the peptides in the F[1,2]-pdz.x_F[3]-sh3.x DHFR fragments orientation and
637 the opposite orientation is shown by the bottom panels, i.e F[1,2]-sh3.x_F[3]-pdz.x. The color of
638 the data point indicates the β-estradiol concentration in the PCA media.



Supplementary Figure 5. Detailed DHFR PCA results of the scaffold efficiency assay. The corrected area under the curve is shown for each scaffold system, peptide pairs and β-estradiol concentration tested in triplicates. The peptide combinations are labeled on top of each plot in F[1,2]-peptide_F[3]-peptide orientation. The color of the data point indicates the β-estradiol concentration in the PCA media.