

REVIEW

Open Access



Selection and screening strategies in directed evolution to improve protein stability

Chang Ren, Xin Wen, Jun Mencius and Shu Quan* 

Abstract

Protein stability is not only fundamental for experimental, industrial, and therapeutic applications, but is also the baseline for evolving novel protein functions. For decades, stability engineering armed with directed evolution has continued its rapid development and inevitably poses challenges. Generally, in directed evolution, establishing a reliable link between a genotype and any interpretable phenotype is more challenging than diversifying genetic libraries. Consequently, we set forth in a small picture to emphasize the screening or selection techniques in protein stability-directed evolution to secure the link. For a more systematic review, two main branches of these techniques, namely cellular or cell-free display and stability biosensors, are expounded with informative examples.

Keywords: Directed evolution, Protein folding, Protein stability, Cell surface display, Cell-free display, Reporter protein, Stability biosensor

Introduction

Miscellaneous functional and structural properties have made proteins promising for developing novel therapeutics, biocatalysts, and advanced materials. Contrary to the common perception, most proteins are just marginally stable (Taverna and Goldstein 2002). Although marginal stability has been suggested to correlate with protein flexibility and easy adaptation of new functions, it limits industrial applications and scientific research which rely on stable, soluble and functional proteins. Additionally, because most mutations tend to compromise the overall protein stability, purely piping up functionally beneficial mutations during protein engineering may eventually become a zero-sum game for both function and stability (Socha and Tokuriki 2013). Maintaining protein stability is, therefore, the prerequisite of mutational robustness to evolve distinct functions (Bloom et al. 2006). Confronting

these challenges, methods on improving protein folding and stability¹ have been established both in vivo and in vitro during the past few decades.

One main branch of these methods focuses on optimizing the environment in which the protein is synthesized, transported into, or functioning. Conventional approaches include attempting different media, temperatures, promoters, inducer concentrations, and co-expression of molecular chaperones during protein expression (Deller et al. 2016; de Marco et al. 2007). These approaches may buffer the potential deleterious effect of protein overexpression and promote productive protein folding. Another great example is enzyme immobilization, which provides not only solid supports for stabilizing enzyme structure, but also protective micro-environments against distorting agents (heat, organic

*Correspondence: shuquan@ecust.edu.cn

State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai Collaborative Innovation Center for Biomanufacturing (SCICB), 130 Meilong Rd., Shanghai 200237, China

¹ Protein folding and stability are inter-connected terms. Thermodynamic stability (ΔG_{NU}) is defined by the Gibbs free energy change upon unfolding ($\Delta G_{NU} = -RT \ln K_{eq}$) deduced from the equilibrium constant (K_{eq}) of the unfolding reaction ($K_{eq} = \frac{[U]}{[N]}$, where $[U]$ is the concentration of the unfolded species and $[N]$ is the concentration of the native species). Therefore, higher stability implies a more substantial fraction of the native species at equilibrium and better foldability of the protein.

solvents, etc.) during biocatalytic industrial processes (Silva et al. 2018; Homaei et al. 2013). The other branch of these methods illustrates that protein folding may be amended by modulating the intrinsic properties of target proteins, such as optimizing codon usage, fusing with solubility-enhancing tags, and a more targeted way, namely, molecular engineering of the protein sequences. We will continue to discuss the last strategy as others have been covered by recent reviews (Elena et al. 2014; Angov 2011; Kosobokova et al. 2016; Costa et al. 2014; Canchi and Garcia 2013).

Molecular engineering for improving protein folding and stability can be carried out through rational design and directed evolution. In rational design, stabilized mutants can be conveniently screened from a 'smart library' based on the sequence, structure, and mechanism of action of a target protein with considerable success rates (Yang et al. 2015). However, rational design is not reliable when prior knowledge of a protein is deficient. Directed evolution, a mimic of natural evolution on the laboratory scale, can be deployed. Directed evolution mainly encompasses genetic diversification to generate libraries and high-throughput selection or screening techniques to isolate and identify desired variants (Packer and Liu 2015). Although for genetic diversification it is nearly impossible to cover the entire sequence space of a target protein, researchers can generate libraries containing 10^6 – 10^{13} clones with modern mutagenesis techniques (Packer and Liu 2015). The real bottleneck for stability engineering then lies on establishing high-throughput strategies for mutant isolation based on phenotypes that are directly linked to protein folding or stability (Packer and Liu 2015). To match with massive library capacity, the development of such strategy can be a major undertaking. Here, we underline this crucial step by introducing two broad categories of technologies to identify the stabilized protein variants, namely display techniques and stability biosensors.

Main text

Improving protein stability by library-based display technologies

Display technologies are straightforward. In general, through physical linkages, target variants displayed on cell surfaces, ribosomes, or attached with their encoding nucleic acids are exposed to harsh conditions such as high temperature, organic solvents, and aggressive pH. The survival clones with selected properties will be enriched through functional assays. The output of one round of selection is often iteratively used as templates for library-generation in the next round (Fig. 1), following by detection and validation of stabilized mutants.

Display technologies can be divided into two main branches, cellular display, and cell-free display. Cellular display methods such as phage or yeast display, along with cell-free display approaches such as ribosome or mRNA/cDNA display initially designed to screen for affinity now have all been applied for protein stability engineering.

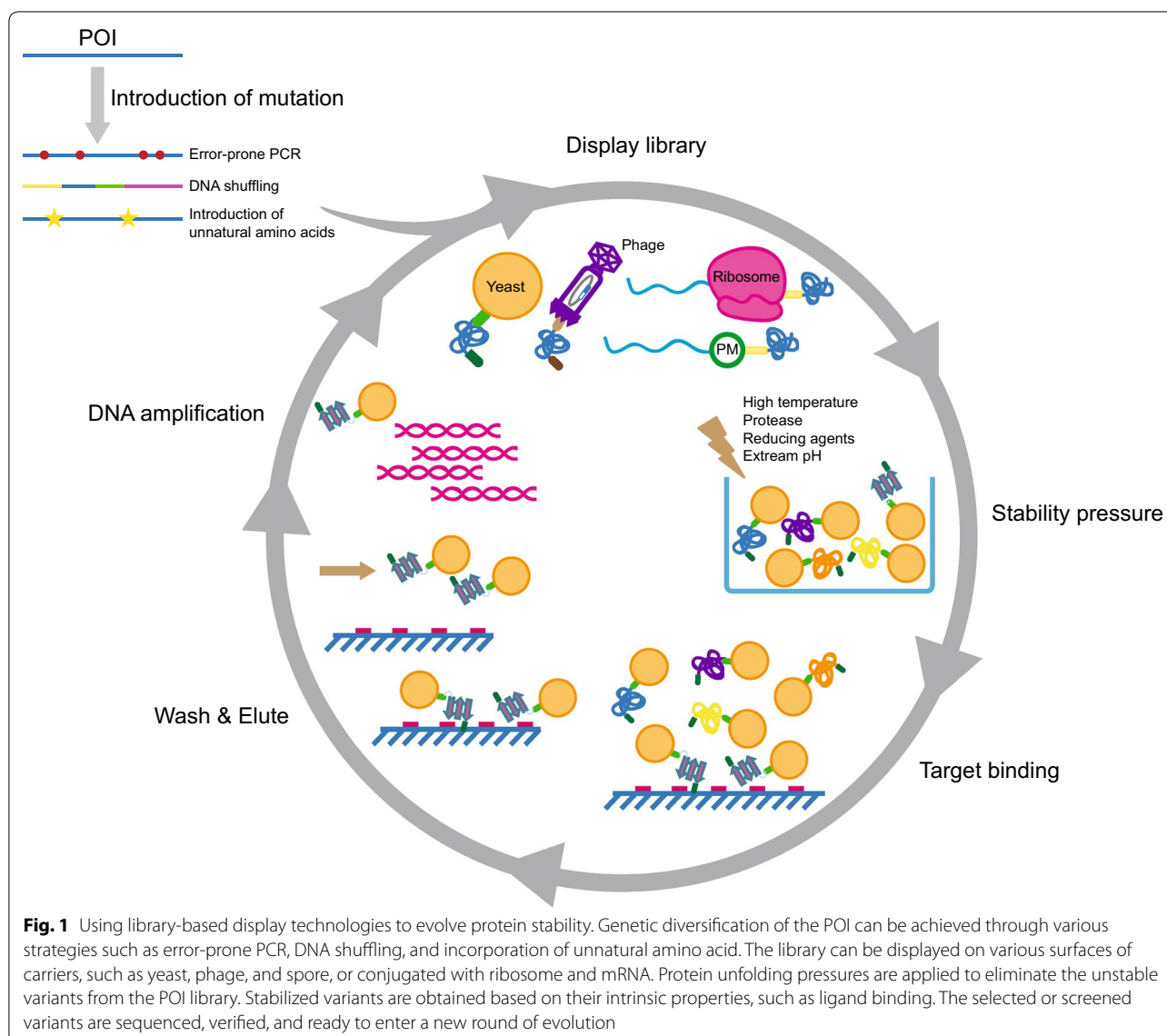
Cellular display approaches

Cell surface display

Cell surface display technology refers to the technique of displaying polypeptides or proteins on the surface of living cells and subsequently analyzing and sorting cells for desired properties. The display of a protein of interest (POI) can be achieved by fusing its gene to a cell-surface anchor protein, for example the outer membrane proteins of bacterial cells or subunits of cell surface appendages (Lee et al. 2003; Galan et al. 2016). Using the signal sequence of the anchor protein, the fusion protein is directed to the cell surface, where it is immobilized and can be accessible from the extracellular space. Typically, each individual cell only expresses a single variant of the POI and the library size is determined by transformation efficiency [for instance, at best 10^8 – 10^{10} for bacterial cells (Galan et al. 2016; Packer and Liu 2015)]. Functional display of the POI can be detected by a fluorescently labeled, POI-specific antibody or ligand so that a screening signal is generated. Alternatively, enzymatic reactions may be exploited to label the POI to generate signals. An important rationale behind this technology is that only fusion proteins with proper folding can reach the cell surface and those containing misfolded POIs are retained and degraded by the cellular protein quality control system. Therefore, the signals of individual cells are correlated with the amount of the fusion proteins on the cell surface, which can be used as an approximate measure for protein stability.

The advantages of cell surface display are as follows. First of all, the screening process can be conveniently performed by using flow cytometric analysis such as fluorescence-activated cell sorting (FACS) or magnetic-activated cell sorting (MACS). Next, eukaryotic cells can be used for display, allowing mammalian-derived proteins go through post-translational modifications (PTM) to ensure proper folding. Finally, thousands of proteins or peptides may be displayed on the cell surface to avoid stochastic fluctuation as seen in phage display, in which only 1–5 copies of POI are displayed (Boder and Wittrup 1997).

As the simplest eukaryotic display system, yeast display is by far the most popular platform for protein engineering (Cherf and Cochran 2015). In *Saccharomyces cerevisiae*, common surface anchor proteins used in



surface display include α -agglutinin (Aga1), a-agglutinin (Aga1p–Aga2p), and the C-terminus region of flocculin (Kondo and Ueda 2004). Kieke et al. displayed a single-chain fragment-variable form of T cell receptor (scTCR) with a library size of 6×10^7 variants on yeast surface through fusion with Aga2p (Kieke et al. 1999). They used TCR antibodies that recognize conformational epitopes of the V_β domain to enrich those displaying properly folded scTCRs. After three rounds of sorting with flow cytometry, stabilized scTCR mutants were obtained with more than half of mutated residues located at the interface of the V_α domain and V_β domain, suggesting that interactions between the two domains may be crucial for scTCR stability. Besides relying on the quality control of the yeast secretory apparatus to prevent the poorly folded target variants from reaching the cell surface, one may

apply extracellular stress conditions to screen for variants more resistant to denaturation. For instance, Shusta et al. applied a 30-min heat shock at 46 °C for the displayed scTCR library before sorting for individual cells (Shusta et al. 2000). After three consecutive sorting cycles, the enriched mutant populations had an order of magnitude higher level of surface-displayed scTCR than the stabilized mutants previously isolated by Kieke et al. that were grown only at the optimal expression temperature (Kieke et al. 1999), demonstrating that the efficiency of stability engineering by yeast display can be improved by extra 'in vitro screening' steps.

Bacterial spores display and gas vesicle display

In recent years, cell surface display has been extended into many new technologies, such as bacterial spores

display (Pan et al. 2012) and gas vesicle nanoparticles display (DasSarma and DasSarma 2015). Spores of *Bacillus subtilis* are coated by two protective layers consisting of more than 70 kinds of distinct proteins (Pan et al. 2012), which provide enough anchoring candidate sites for displaying exogenous protein. The commonly used anchor proteins in *B. subtilis* spore surface display are CotB, CotC, CotE, and CotG (Wang et al. 2017a). Displayed proteins do not undergo secretion process since during spore assembly the coat proteins (and their fusion partners) are directly deposited on the spore surface inside the mother cell. In addition, spores have high tolerance to extreme environments, which makes them suitable for displaying enzymes used in harsh industrial processes (Chen et al. 2017), and potentially allows the screening for more stabilized biocatalysts under these conditions. For instance, Silu et al. displayed a library of the laccase CotA, which itself is a spore coat protein, on the spore surface of *B. subtilis* to screen for pH stable mutants that can be used as an immobilized whole-cell biocatalyst (Sheng et al. 2017). The spore was incubated at low pH before reacting with the substrate of CotA. After seven rounds of screening, an E498G variant was identified to have a 25-fold increase of half-life of inactivation ($t_{1/2}$) at pH 4 compared with wild-type CotA.

Gas vesicle nanoparticles (GVNP) refer to the buoyant organelles produced in halophilic microorganisms that can be purified as protein nanoparticles (DasSarma and DasSarma 2015). Exogenous proteins can be displayed on the surface of GVNP by fusing to the GvpC protein, one of the two major proteins that build up the GVNP membrane (Stuart et al. 2001). These nanoparticles are stable, non-toxic, and bio-engineerable, making them suitable for displaying antigens for vaccine development. Although there is no report by far, we hypothesize that the ability to display multiple proteins by GVNP makes this system very promising for protein stability engineering after applying an in vitro unfolding stress, as demonstrated by other conventional cell surface display systems.

Phage display

Phage display was first proposed by George Smith in 1985. The target protein is inserted into the minor coat protein (pIII) of the filamentous phage to be directed to the surface of phage particles (Smith 1985). Folding of the target protein influences the amount of functional pIII during phage assembly and thus determines the infectivity of the virion, providing a linkage between protein stability and the phenotype. Just as other developed display technologies, phage display has an advantage that a large library size ($\sim 10^9$) can be achieved swiftly and that a wide range of screening conditions may be applied since the phage particles are relatively stable and can be

amplified effectively. Even though the library size can reach up to 10^{12} in theory, the final library size of phage surface display is around 10^9 in practice. This is due to restrictions of transformation efficiency of bacterial cells that are used to produce the phages, as well as phage surface expression level (Galan et al. 2016). Besides pIII, the major coat protein pVIII and a few other structural proteins have also been applied in phage display (Tan et al. 2016). Both lysogenic phages such as M13 and fd and lytic phages such as T7 may be used to display target proteins. Higher library diversity can be achieved with the use of lytic phages (Krumpe et al. 2006).

Once the POI is displayed on phage's surface, the virions are subject to high temperature, extreme pH, chemical denaturing agents or proteases (Pershad and Kay 2013). Phages with unfolded target variants will exhibit reduced binding capacity, causing their elimination during subsequent affinity screening, whereas well-folded variants retaining binding affinity will be enriched. This process usually requires several rounds of binding, washing, elution (collectively referred to as biopanning), and amplification before the mutants with significantly improved properties are finally obtained. Beliën et al. applied heat stress to screen for stabilized endoxylanase (XynA) from *B. subtilis* with three iterative biopanning (Belien et al. 2008). Five single-point mutants were identified, and the half-inactivation temperature of each mutant was increased by 2–3 °C compared to the wild type. Jespers et al. used transient heat stress to induce unfolding of the antibody heavy chain variable domains (dAbs) (Jespers et al. 2004). After cooling, screening for dAbs towards protein A was performed to isolate the variants that could resist aggregation and efficiently refold. This resulted in six mutant dAbs that could be handily purified as soluble proteins by bacterial expression system. Frank et al. sequentially applied pepsin and chymotrypsin pressure to screen for protease resistant hirudin variants displayed on phage M13 (Wirsching et al. 2003). A variant harboring 6 single-point mutations was identified after 2 rounds of biopanning. The IC₅₀ value (inhibitor concentration to achieve 50% inhibition) of this variant increased 100-fold compared to wild-type hirudin. Moreover, the variant still retained partial thrombin inhibition activity.

The fast-replicating nature of the filamentous phage allows automation of the biopanning procedure. As the pioneer, David Liu's group developed phage-assisted continuous evolution (PACE) to evolve proteins or nucleic acids with the capacity of performing more than one round of evolution per hour without manual intervention (Esvelt et al. 2011). In this delicate method, fresh *E. coli* host cells are continuously fed into a lagoon vessel and instantaneously infected by a replicating population

of M13 phages while the old cells escape lagoon at the same speed. Only the phages producing sufficient amounts of functional coat protein pIII are infectious thus their propagation outruns the rate of dilution, resulting in automatic enrichment. To create a linkage between POI activity and pIII production, synthetic circuits were designed. Instead of fusing the POI to pIII, the POI is encoded by a selection phage vector (SP) and pIII is separately encoded by an accessory plasmid (AP), whose transcription is controlled by activity of the POI. The host cells also contain a mutagenesis plasmid (MP) which increases the mutation rate of the POI gene upon infection. To select for higher stability, Wang et al. introduced an extra folding reporter by fusing the POI with the N-terminal half of T7 RNA polymerase on the phage vector while expressing the other half of T7 RNAP from the host (Wang et al. 2018). Soluble expression of the POI permitted reconstitution of T7 RNAP activity, thus initiated the transcription of pIII from its T7 promoter to allow the phage to produce infectious progenies. Via this strategy, a P33T mutant capable of inhibiting the aggregation-prone phenotype of a maltose-binding protein (MBP) variant, G32D I33P, by ~twofold was isolated after a 3-day continuous evolution.

Cell-free approaches

Cell-free display technologies avoid the need of introducing genetic materials into cells. This eliminates laborious transformation steps, facilitates the generation of large libraries and maintenance of library diversity, and also enables users to control the mutational spectrum more precisely. The library capacity can effectively achieve 10^{12} – 10^{14} (Galan et al. 2016). Library diversity can be further expanded by incorporation of unnatural amino acids or nucleotides.

Ribosome display

Ribosome display is based on in vitro translation of a library of mRNA molecules containing no stop codon. The DNA library of the POI is first transcribed into mRNA. During translation, a ternary complex forms between the nascent polypeptide (covalently connected to the peptidyl-tRNA at the P-site), the ribosome, and the mRNA (Pluckthun 2012). When the translation reaches the physical end of the mRNA, absence of the stop codon in the mRNA causes ribosomes to stall and the complex is stabilized with sophisticated control of reaction conditions. Thus, a direct physical link between the genotype molecule (the nucleic acid) and the phenotype molecule (the protein) is established using ribosome as the connector. After one round of screening for desired properties of the protein, the mRNA molecules attached to the protein

variants are reversely transcribed into cDNA products, which can be deployed into the next round of transcription, translation, and subsequent screening.

Displaying protein or polypeptide on the surface of ribosomes is facilitated by fusing the POI library to an unstructured, flexible tether sequence, which allows the POI to fold outside of the ribosome tunnel (Makeyev et al. 1996). Coupling in vitro transcription and translation can be achieved by incubating DNA with cell extracts, either *E. coli* S30 extract or rabbit reticulocyte extract (He and Taussig 2007). To screen for stabilized protein variants, an unfolding pressure is usually applied prior to the functional screening (Buchanan 2012), similar to the strategies exploited in the cell-surface display technologies.

Jermutus et al. applied reducing redox potential during the translation of single-chain variable fragments (scFvs) and used antigen binding as the readout to select for variants with increased stability (Jermutus et al. 2001). DTT reduced the two conserved disulfide bonds in scFvs, causing the proteins to lose activity and aggregate. After five-round selection with increasing concentration of DTT, the most stable scFv mutant increased ΔG_{NU} by 30 kJ/mol, which also possessed a shift of denaturation midpoint by 0.9 M urea compared to the wild type. In another example, Matsuura et al. developed a procedure to screen for properly folded proteins by coupling ribosome display with proteolytic pressure. They also used hydrophobic interaction chromatography (HIC) to exclude aggregation-prone protein variants (Matsuura and Pluckthun 2003). The combination of the two pressures led to prominent enrichment of more stable proteins.

mRNA display/cDNA display

First proposed in 1997, mRNA display is a screening method based on the formation of a binary complex between a protein and its encoding nucleic acid in vitro (Nemoto et al. 1997; Roberts and Szostak 1997). The key to this method is that the mRNA is attached through a DNA linker to a puromycin (an analog of the aminoacyl end of a tyrosyl tRNA), which terminates translation by entering the ribosomal A site. As a result, the growing polypeptide chain is covalently attached to puromycin and its encoding mRNA. Similar to ribosome display, screening for function and stability can be conducted in mRNA display simultaneously. After in vitro translation, the mRNA–protein complex of the POI library undergoes a stability pressure, following by functional screening until the desired variants are obtained.

The *E. coli* S30 extract system commonly used in cell-free translation contains endogenous nucleases and proteases that degrade mRNA and translated proteins. To

bypass these risks, Nagumo et al. introduced a reconstituted *E. coli* protein synthesis system called Protein synthesis Using Recombinant Elements (PURE) into the mRNA display platform (Nagumo et al. 2016). All steps carried out using purified elements lead to larger libraries compared with the conventional S30-based system. Additionally, it offers the feasibility of introducing unnatural amino acids with superior screening efficiency. The library capacity of mRNA display can reach up to 10^{13} , which is 10^6 times of yeast display and 10^4 times of traditional phage display (Takahashi et al. 2003). Nevertheless, both mRNA display and ribosome display are restricted from extreme screening conditions because of the instability of mRNA. To counter this problem, Kurz et al. developed a cDNA display strategy based on conventional mRNA display, with the modification that a branched mRNA–puromycin conjugate was used. A reverse transcription step was immediately performed after generation of the mRNA–protein fusion, leading to covalent attachment of the protein to its encoding DNA (Ueno and Nemoto 2012). The cDNA–protein fusion was then used in the subsequent screening steps, which was demonstrated to be more inert towards degradation with a half-life at least fourfold longer than that of the mRNA–protein fusion (Kurz et al. 2001).

Stephen et al. engineered cycGiBP, a cyclized peptide, for higher stability and target affinity by introducing the codon for unnatural amino acids into its mRNA (Fiacco et al. 2016). The mRNA–cycGiBP complex was incubated with immobilized trypsin, chymotrypsin, proteinase K, and aminopeptidase before binding screening to evolve the stability of cycGiBP. The mutant retrieved in the last round of screening showed a 500-fold stability increase in serum and a 3700-fold improvement in protease resistance compared with the parent sequence.

Improving protein stability by protein stability biosensor-based technologies

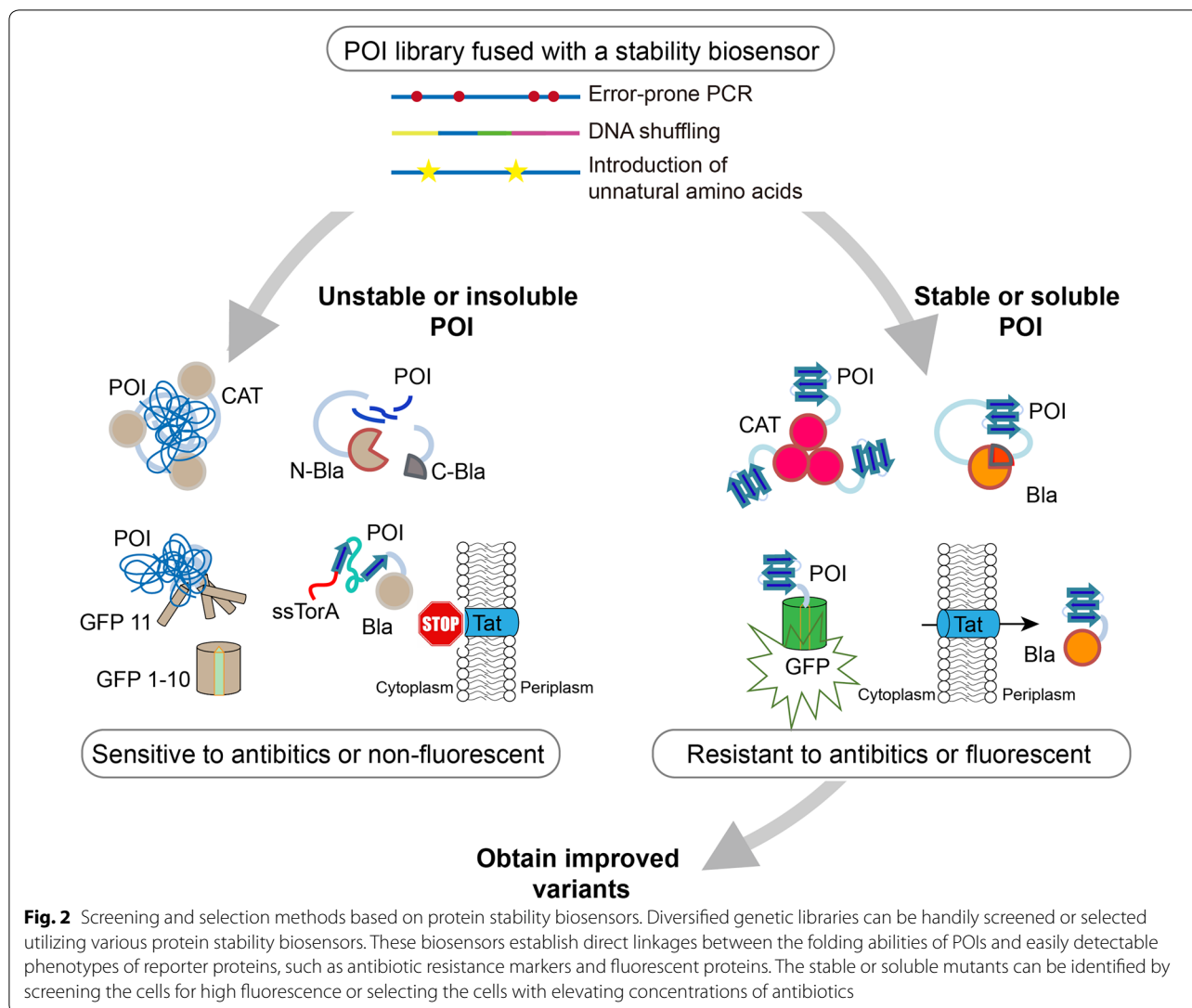
In the display technologies described above, the folding of a specific POI is often inferred by its function, i.e., the ability to bind a ligand or an antibody, or the enzymatic activity leading to colorimetric or fluorescent signals. Protein stability biosensors provide alternative methods for monitoring protein stability when the function of the POI is not easily characterized. A typical protein stability biosensor consists of two basic components: a phenotypical reporter protein and a POI genetically fused to the reporter. Through expression of the fusion, folding of the reporter is influenced by the folding of the POI, resulting in a phenotypical change. Therefore, a direct linkage between the folding ability of the POI and the phenotype of the reporter is established without prior knowledge of the structure or function of the POI. Many of the stability

biosensors provide easy-to-detect phenotypes that allow the selection or screening for stabilized protein variants in a high-throughput manner (Fig. 2). However, since the selection or screening is performed in vivo, the size of the POI library is generally limited by transformation efficiency.

Chloramphenicol acetyltransferase (CAT) reporter

Chloramphenicol acetyltransferase (CAT) is a trimer of identical subunits, which confers chloramphenicol resistance by acetylating chloramphenicol to prevent its binding to ribosomes. CAT is a widely used reporter protein due to its easily detectable enzymatic activity (Dauert et al. 2000). To be used as a biosensor, CAT is fused to the C-terminus of the POIs so that its ability to detoxify chloramphenicol is linked to the stability and solubility of the POIs (Maxwell et al. 1999). Cells expressing CAT fusions with unstable or insoluble proteins exhibit decreased chloramphenicol resistance compared to cells expressing the fusions with stable and soluble proteins, which provides a quantitative measure of the in vivo stabilities of proteins. In directed evolution experiments, stabilized protein variants were isolated by plating the cells expressing the library of the POI in CAT fusion on solid media containing increasing concentrations of chloramphenicol. A number of successful applications of this biosensor include the screening for solubility-enhanced mitochondrial cytochrome P450 (Janocha et al. 2011), stabilization of imidazole glycerol phosphate synthase (Seitz et al. 2007), interleukin-15 (Behar et al. 2011), (+)- δ -cadinene synthase (Yoshikuni et al. 2006), and many more.

It should be noted that this head-to-tail construction has a few drawbacks, not only limited to the CAT biosensor, but also to other reporter systems with terminal fusion designs, such as green fluorescent protein (GFP) (Kawasaki and Inagaki 2001; Cabantous et al. 2005a), kanamycin nucleotidyl transferase (Chautard et al. 2007), and dihydrofolate reductase (DHFR) (Liu et al. 2006). An intact, functional reporter protein can be released by proteolytic cleavage of a poorly folded POI, by alternative translation due to frameshifting or the presence of internal cryptic ribosome binding sites within the POI gene (Seitz et al. 2007; Kawasaki and Inagaki 2001; Cabantous et al. 2005a), or by the incorporation of a stop codon near the 5' terminus of the POI gene (Chautard et al. 2007), all of which lead to increased false-positive rate. Chen et al. also reported that the amino acid sequences of the linker connecting the POI and CAT had significant different effects on the selection (Chen et al. 2009). For example, the WPGSPA linker resulted in slower cell growth on chloramphenicol plates, lower tolerance to larger POIs, and a much lower in-frame ratio compared to the AGSSAAGSGS linker. Thus, the apparent false-positive



rates and the dependence on linker sequence precluded the wide application of the CAT system in protein stability engineering.

β -Lactamase tripartite reporter

One alternative approach to create fusions is to insert the POIs into the reporter protein to create a sandwich construct, which is expected to address the problems caused by proteolysis or frameshift translation. Once the POI is inserted into a permissive site within the reporter, its productive folding determines the association efficiency of the two halves of the reporter and is therefore related to the phenotype conferred by the reporter. If the POI is stable and well folded, the reporter protein can reassemble into a functional conformation. In contrast, unstable or poorly folded POI is prone to proteolysis, splitting the reporter into two halves that are eventually degraded. If

a frameshift mutation occurs within the POI gene, the expression of the reporter will also be disrupted, resulting in no phenotype. In this regard, the tripartite setting will reduce the rate of false positive.

The most representative tripartite reporter system was developed by Foit et al. based on the TEM-1 β -lactamase (β la), which confers resistance to β -lactam antibiotics by hydrolyzing them in the bacterial periplasm (Foit et al. 2009). β la is an ideal reporter protein because it is monomeric, relatively small (29 kDa) and has low cytotoxicity. In addition, it lacks orthologs in eukaryotes and many prokaryotes, so the phenotype can be well correlated with the level of exogenously expressed β la. β la split at a surface-exposed loop has been used for the detection of protein-protein interactions, ligand binding, and interaction networks (Wehrman et al. 2002; Galarneau et al. 2002; Lofdahl et al. 2009; Remy and Michnick 2015). Foit

et al. converted β la into a stability biosensor by inserting the POI between residues 196 and 197, a site previously demonstrated to tolerate splitting. Selection for colonies grown on increasing concentrations of penicillin allowed the isolation of 26 stabilized variants of the immunity protein 7 (Im7). These variants were confirmed to have ΔG_{NU} values increased by 0.6–5.9 kJ/mol and were more kinetically stable and had higher expression levels. Similar approaches were applied by the same group to obtain better folded variants of bovine pancreatic trypsin inhibitor (BPTI) (Foit et al. 2011). Among the 14 BPTI variants, G28W and K26M, which exhibited the most intensive penicillin resistance, significantly increased the steady-state levels in the periplasm by more than twofold compared to the wild type.

Xiong et al. demonstrated the capacity of this system on assessing the foldability of de novo designed proteins spanning 4 structure classes (all- α , all- β , α/β and $\alpha + \beta$) (Xiong et al. 2014). They found that when expressed as a β la fusion, the designed sequence that conferred high penicillin resistance also formed well-folded structure in solution. Through directed evolution approach, the foldability of the initially designed proteins was further optimized, demonstrating the ability of this system to correct the original problematic designs as an effective complement to the rational design approach. Wang et al. also showed that this system can be applied to optimize the folding of a designed chimeric ($\beta\alpha$) 8-barrel protein (TIM-barrel protein) (Wang et al. 2017b). Through six rounds of evolution, the half melting temperature (T_m) of the isolated variant that led to the highest ampicillin resistance has been improved by more than 20 °C.

In addition to TEM-1 β -lactamase, tripartite reporter systems based on antibiotic resistance have been analogously generated in kanamycin, streptomycin/spectinomycin, and nourseothricin resistance markers (Malik et al. 2014).

Split GFP reporter

Green fluorescent protein (GFP) is by far the most commonly used and most studied reporter protein in biochemistry and cell biology. For the application of GFP in protein folding, Waldo et al. first developed a rapid protein folding assay using the head-to-tail construction of GFP and POI (Waldo et al. 1999). The basic principle of sensing protein stability is that GFP chromophore formation requires proper folding of GFP (Ormo et al. 1996), and GFP folds poorly when fused with misfolded proteins. Thus, only fusion protein with proper folding and reasonable solubility can confer GFP with strong fluorescence. By screening emitted fluorescence, the aggregation-prone ferritin H-subunit from bullfrog was successfully optimized to be completely soluble at 37 °C

(Waldo et al. 1999). However, the main problem of using GFP fusions as an aggregation reporter is that insoluble fusion proteins will still be considerably fluorescent if their aggregation rates are slower than the formation rate of the GFP chromophore (Kothawala et al. 2012). Moreover, the terminal fusion of GFP may perturb the folding of target proteins and change their misfolding propensity (Kothawala et al. 2012; Bertens et al. 2003). To overcome these limitations, modifications have been deployed to the original GFP system.

Waldo and coworkers advanced the system in which GFP is split into two inactive, nonfluorescent fragments: GFP1–10 (residues 1–214) expressed on its own and GFP11 (residues 215–230) destined to fuse with POI (Cabantous et al. 2005b). They then evolved the two fragments to improve their solubility and the efficiency of self-assembly. Owing to its small size, the GFP11 tag has minimal effect on the solubility and folding of proteins it fused with. Sequential expression of the GFP11-tagged POI and the large fragment GFP1–10 allows the POI to undergo protein quality control before reconstituting the chromophore. An important rationale of this technique in measurement of protein stability is that processes such as POI misfolding or aggregation that make the GFP11 tag inaccessible can prevent its complementation with the GFP1–10 fragment, resulting in no green fluorescence and only the GFP11 tag fused with well-folded POI can access to the GFP1–10 fragment along with formation of the chromophore of GFP (Cabantous et al. 2005b). Using the split GFP system coupled with fluorescence size exclusion chromatography (FSEC) analysis, Rodriguez-Banqueri et al. successfully improved the stability of a membrane protein, SteT (the *Bacillus subtilis* L-serine/L-threonine exchanger) (Rodriguez-Banqueri et al. 2016). Two best-optimized mutants, I134V/A377T and L210Q/M229V, showed the best-combined improvement of expression, stability, and monodispersity in detergents. This split-GFP system has also been used to screen for protein solubility in an automated and high-throughput way by incorporation of modular liquid-handling platforms (Listwan et al. 2009).

Additionally, Waldo and coworkers further split GFP1–10 into GFP1–9 (residues 1–193) and GFP10 (residues 194–212), together with GFP11 (residues 213–233) to make a tripartite split GFP system (Cabantous et al. 2013). The POI is then inserted between GFP10 and GFP11 to produce a sandwiched hairpin for GFP1–9 to recognize. Recently, Mo et al. inserted the esterase Aaeo1 between GFP10 and GFP11 and screening the Aaeo1 library with a high-throughput flow cytometry and a QPix420 colony picker (Mo et al. 2018). They isolated 55 colonies with brighter fluorescence from a library containing 25,000 variants. After the secondary screening for

esterase activity, they obtained two mutants whose soluble expression was improved by twofold in *E. coli* with k_{cat}/K_m values twofold higher than that of the parent.

If GFP is dissected on a surface loop between residues 157 and 158, the resulting fragments will not spontaneously reassemble when produced simultaneously in bacteria (Magliery et al. 2005). Lindman and coworkers utilized this split system to improve the stability of the 56-residue B1 domain of protein G (PGB1) (Lindman et al. 2010). Here, PGB1 was split between residues 40 and 41 and the resulting N- and C-fragments were genetically fused to GFP_{158–238} and GFP_{1–157}, respectively. The underlining principle is that interactions stabilizing the folded conformation of an intact protein are often strong and specific enough to allow reconstitution of the intact protein fold from fragments of the protein. Therefore, screening for mutations that enhance the binding of the two PGB1 fragments will identify stability-enhancing mutations. From a 'smart library' of 3456 variants of PGB1_{1–40}, colonies exhibiting brighter fluorescence were picked, sequenced, and the mutations were reconstituted on the intact PGB1 to assay for stability. The best mutant showed a 12 °C increase in the midpoint of thermal denaturation and an 8.7 kJ/mol increase in ΔG_{NU} at 25 °C (Lindman et al. 2010).

Since the discovery of GFP, a variety of colorful fluorescent proteins with spectra ranging from blue to red now is available for the development of fluorescence resonance energy transfer (FRET)-based biosensors. In these sensors, a POI is flanked by a fluorescence donor and an acceptor protein. Alterations in POI conformation are reflected by the changes in energy transfer and only well-folded POIs can bring the donor and acceptor into close proximity to produce efficient FRET signals (Tamura and Hamachi 2014; Lindenburg and Merckx 2014). Such properties allow the screening for stabilized protein variants. For example, using a tripartite sensor composed of a GFP, a POI, and a blue fluorescent protein (BFP), Philipps et al. screened a library of immunoglobulin V_L 2 domains for bright green bacterial colonies using the excitation wavelength of BFP (Philipps et al. 2003). They finally identified two immunoglobulin V_L 2 domain mutants whose ΔG_{NU} increased by 4 kJ/mol.

GFP also has the potential to work as a semi-in vitro stability biosensor. Many of the stress conditions (e.g., organic solvents, detergents) applied in the aforementioned display technologies are not practical with living cells because they cause cell lysis. This precludes the deployment of in vivo biosensors, such as GFP, in the engineering of stabilities other than thermodynamic or protease stability. A recent good attempt to address these limitations is to couple GFP to cellular high-throughput encapsulation solubilization and screening (CHESS)

(Yong and Scott 2015), which was originally developed to engineer detergent-resistant G protein-coupled receptors (GPCRs) (Scott and Pluckthun 2013). In CHESS, each cell expressing a POI variant is encapsulated with detergent-stable polymers to produce cell-like microcapsules. Therefore, stability of the POI can be determined using a high-throughput technique under conditions that solubilize cell membranes. The researchers demonstrated that GFP biosensors larger than 70 kDa were well retained inside the microcapsules after detergent treatment. In addition, the fluorescent signal from the encapsulated GFP can be maintained at 90 °C for up to 5 min and in 2% SDS up to 24 h (Yong and Scott 2015). These results provide a decent starting point for screening stable protein variants with the GFP biosensor under harsh industrial conditions. As a proof of principle, an MBP-sfGFP (super-folder GFP) library was iteratively screened with CHESS and sfGFP variants showing enhanced temperature and SDS resistance were obtained (Yong and Scott 2015). It is expected that the availability of detergent-resistant sfGFP makes it possible to report the stability of its fused POI under similar denaturing conditions. It is conceivable that by changing the kind of stressors (such as pH, reactive oxygen species, and redox states), this approach may increase the likelihood of engineering other stabilities in addition to thermodynamic stability.

Tat folding sensor

Cells utilize diverse protein quality control (PQC) systems to monitor whether proteins are appropriately folded. In most bacteria and archaea, the twin-arginine translocation (Tat) protein export system performs protein quality control by delivering only correctly folded proteins across the inner membrane (Palmer and Berks 2012; DeLisa et al. 2003). Based on the fact that the Tat translocase discriminates folded proteins from misfolded proteins and stringently exports only the former, DeLisa and coworkers developed a genetic selection system for protein solubility (Fisher et al. 2006, 2011). The tripartite biosensor consists of an N-terminal Tat signal peptide (e.g., ssTorA) which directs the fusion protein to the Tat translocase, a POI, and a TEM1 β -lactamase reporter fused at C-terminus of POI, which only confers penicillin resistance when translocated into the periplasmic space (Fisher et al. 2011; Waraho-Zhmeyev et al. 2013). Utilizing the effective proofreading capacity of the Tat export pathway and localization-specific function of the β la reporter, the folding status of the target protein is directly coupled with the antibiotic resistance phenotype.

The Tat biosensor was used in directed evolution to identify solubility-enhanced variants of a couple of proteins. Following the identification of the F19S L34P mutant of the Alzheimer's A β 42 peptide, Fisher et al.

revealed that the hydrophobic cluster composed of residues 17–21 and 30–36 was responsible for the solubility of A β 42 peptide (Fisher et al. 2006). In addition, they isolated a soluble scFv after three rounds of laboratory evolution starting from an insoluble parent sequence (Fisher and DeLisa 2009). For the endoglucanase Cel5A from the phytopathogen *Fusarium graminearum* (FgCel5A), Boock et al. improved its heterologous production in *E. coli* and identified a variant whose production was increased by 30-fold (Boock et al. 2015). By coupling a semi-high-throughput screening for function, they also restored the enzymatic activities of the isolated mutants to the same level as the wild-type enzyme. Besides DeLisa's group, Kim et al. identified soluble variants of the human heavy chain variable domain (V_H) from a human germ-line VH library by performing three rounds of solubility selection (Kim et al. 2014). Then, the best variant was used as the backbone into which combinatorial mutations were introduced at seven specific sites using NNK degenerative codons. The selection revealed a strong bias towards the bulky side chain residue, tryptophan, at two adjacent positions localized in a prominent pocket on the surface. These residues together with a third tryptophan made a triad, demonstrating the essential role of hydrophobic packing in maintaining protein stability.

In addition to transporting of a single polypeptide chain, the Tat translocase also support the export of hetero-oligomeric protein complexes. This unique feature permits the application of the Tat-based biosensor for the selection of stable protein complexes. A hitchhiker mechanism is adopted in such cases, in which one component is fused with the Tat signal sequence, while the other is connected to the β la reporter. β la can only be co-translocated into the periplasm if the complex is relatively stable (Rodrigue et al. 1999; Waraho and DeLisa 2009). This strategy allowed the isolation of scFv variants with enhanced affinity towards the 47-residue bZIP domain of the yeast transcription factor Gcn4p. The solubility of these variants was also improved when they were expressed on their own, and the solubility of the best variant enhanced by more than tenfold (Waraho-Zhmayev et al. 2014).

Improving protein stability without using reporter protein fusions

Although researchers have spent every effort to minimize the interference of reporter proteins to the innate properties of target proteins, the presence of fusion tags affects more or less the folding status or in vivo functions of target proteins (Waldo 2003). Therefore, direct screening or selection based on the stability or solubility of the target protein rather than the phenotype conferred by the

fusion partner is the initial and ultimate goal of directed evolution for protein stability.

Colony filtration assay, prototyped by Dreher and coworkers, utilizes the principle that soluble proteins can diffuse through a porous membrane, but protein aggregates are difficult to penetrate the membrane due to their large diameters (Dreher et al. 1991). In their work, antibody fragments expressed in bacterial cells diffused through a Durapore filter, captured by the antigen on the second membrane, and detected by standard immunochemical reagents. This assay was sensitive enough to allow the precise isolation of a single positive bacterial colony expressing antigen-specific scFv from 10,000 negative colonies (Dreher et al. 1991). Conceptually similar, Knaust et al. used filter plates in the 96-well format to allow the separation of the soluble fractions from inclusion bodies in the cell lysates (Knaust and Nordlund 2001). The soluble fractions passed through the filter plates were loaded onto a nitrocellulose membrane. Only soluble recombinant proteins can be detected with target-specific antibodies in subsequent immunoblot assays.

Based on the previous research, Cornvik et al. first applied colony filtration blot (abbreviated as CoFi blot) with directed evolution principles to isolate soluble protein variants (Cornvik et al. 2005). In the screening of leucine carboxyl methyltransferase 1 from a N-terminal deletion library, Cornvik et al. identified the domains that can be expressed in the soluble form. Asial et al. further expanded the variety of target proteins in the CoFi blot (Asial et al. 2013). They improved the T_m of the NUDIX hydrolase domain of human NUDT18 by 26.6 °C, anti-Grb2 scFv by 9.5 °C, interleukin-1 receptor antagonist by 9.5 °C, and TEV protease by 9.8 °C. Moreover, they demonstrated that those stabilized mutants still retained similar activities as their corresponding wild-type proteins. Overall, CoFi blot is easy to manipulate and has the potential to be a robust and reporter-independent screening technique for protein stability engineering.

Concluding remarks

Directed evolution, the concept of which was coined nearly 50 years ago (Francis and Hansche 1972), has become a standard tool to tailor natural-occurring proteins for enhanced or innovative functions. The thermodynamic aspect of protein has gained focus in contemporary protein engineering, because increased stability often promotes the evolution of other desired properties. Also, other aspects, such as stability towards proteases, organic solvents, extreme pH, or oxidative species, are receiving increasing attention because these properties are closely related to enzyme performance in the biocatalytic industry.

For the two essential elements of directed evolution, although approaches for genetic diversification have become standard, challenges remain in optimizing the delicate design of screening or selection techniques that allow for a comprehensive exploration of the sequence space of the library and the effective identification of stabilized protein variants. To date, methods for enhancing enzyme stability in organic solutions have not been matured compared to those operated in aqueous solutions. In addition, the influence of the fusion partner on the folding or solubility of the target requires is not minimized in most settings, and the tradeoff between stability and function also needs to be weighed and balanced. Moreover, only during the last decade, high-throughput operations that allow fully automated, continuous rounds of evolution have emerged, such as PACE. To address all these problems, there is a great need for the development of versatile, practical, sensitive, and fully automatic stability engineering techniques. With these innovative goals in mind, future endeavors will fully embrace the beauty of directed evolution and continue to produce superior medicines, biocatalysts, and advanced materials.

Abbreviations

ΔG_{NU} : thermodynamic stability expressed as the Gibbs free energy change of protein unfolding; POI: protein of interest; FACS: fluorescence-activated cell sorting; MACS: magnetic-activated cell sorting; PTM: post-translational modifications; scTCR: single-chain fragment-variable form of T cell receptor; GVNPs: gas vesicle nanoparticles; dAbs: antibody heavy chain variable domains; IC50: inhibitor concentration to achieve 50% inhibition; PACE: phage-assisted continuous evolution; scFvs: single-chain variable fragments; HIC: hydrophobic interaction chromatography; PURE: protein synthesis using recombinant elements; CAT: chloramphenicol acetyltransferase; GFP: green fluorescent protein; DHFR: dihydrofolate reductase; β la: TEM-1 β -lactamase; Im7: immunity protein 7; BPTI: bovine pancreatic trypsin inhibitor; T_m : half melting temperature; FSEC: fluorescence size exclusion chromatography; PGB1: B1 domain of protein G; FRET: fluorescence resonance energy transfer; BFP: blue fluorescent protein; CHES: cellular high-throughput encapsulation solubilization and screening; GPCRs: G protein-coupled receptors; sfGFP: super-folder GFP; PQC: protein quality control; Tat: twin-arginine translocation; CoFi: colony filtration.

Acknowledgements

We thank Shijia Hu, Alessandro Ruan, and Zixiao Xue for the critical reading of the manuscript and assistance with figure preparation.

Authors' contributions

SQ conceived the study. CR, XW, and JM collected references. CR and XW prepared the figures. All authors contributed to the writing of the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (NSFC) (Grant Nos. 31661143021, 31670802, and 31870054 to S. Quan), the Fundamental Research Funds for the Central Universities (Grant No. 22221818014), and the Research Program of State Key Laboratory of Bioreactor Engineering.

Availability of data and materials

Not applicable.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 4 October 2019 Accepted: 11 December 2019

Published online: 27 December 2019

References

- Angov E (2011) Codon usage: nature's roadmap to expression and folding of proteins. *Biotechnol J* 6(6):650–659. <https://doi.org/10.1002/biot.20100332>
- Asial I, Cheng YX, Engman H, Dollhopf M, Wu BH, Nordlund P, Cornvik T (2013) Engineering protein thermostability using a generic activity-independent biophysical screen inside the cell. *Nat Commun* 4:2901. <https://doi.org/10.1038/ncomms3901>
- Behar G, Sole V, Defontaine A, Maillason M, Quemener A, Jacques Y, Tellier C (2011) Evolution of interleukin-15 for higher *E. coli* expression and solubility. *Protein Eng Des Sel* 24(3):283–290. <https://doi.org/10.1093/protein/gzq107>
- Belien T, Verjans P, Courtin CM, Delcour JA (2008) Phage display based identification of novel stabilizing mutations in glycosyl hydrolase family 11 *B. subtilis* endoxylanase XynA. *Biochem Biophys Res Commun* 368(1):74–80. <https://doi.org/10.1016/j.bbrc.2008.01.047>
- Bertens P, Heijne W, Van der Wel N, Wellink J, Van Kammen A (2003) Studies on the C-terminus of the Cowpea mosaic virus movement protein (journal article). *Adv Virol* 148(2):265–279. <https://doi.org/10.1007/s00705-002-0918-z>
- Bloom JD, Labthavikul ST, Otey CR, Arnold FH (2006) Protein stability promotes evolvability. *Proc Natl Acad Sci USA* 103(15):5869–5874. <https://doi.org/10.1073/pnas.0510098103>
- Boder ET, Wittrup KD (1997) Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol* 15(6):553–557. <https://doi.org/10.1038/nbt0697-553>
- Boock JT, King BC, Taw MN, Conrado RJ, Siu KH, Stark JC, Walker LP, Gibson DM, DeLisa MP (2015) Repurposing a bacterial quality control mechanism to enhance enzyme production in living cells. *J Mol Biol* 427(6):1451–1463. <https://doi.org/10.1016/j.jmb.2015.01.003>
- Buchanan A (2012) Evolution of protein stability using ribosome display. In: Douthwaite JA, Jackson RH (eds) *Ribosome display and related technologies: methods and protocols*. Springer New York, New York, pp 191–212. https://doi.org/10.1007/978-1-61779-379-0_11
- Cabantous S, Pedelacq JD, Mark BL, Naranjo C, Terwilliger TC, Waldo GS (2005a) Recent advances in GFP folding reporter and split-GFP solubility reporter technologies. Application to improving the folding and solubility of recalcitrant proteins from *Mycobacterium tuberculosis*. *J Struct Funct Genomics* 6(2–3):113–119. <https://doi.org/10.1007/s10969-005-5247-5>
- Cabantous S, Terwilliger TC, Waldo GS (2005b) Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat Biotechnol* 23(1):102–107. <https://doi.org/10.1038/nbt1044>
- Cabantous S, Nguyen HB, Pedelacq JD, Koraichi F, Chaudhary A, Ganguly K, Lockard MA, Favre G, Terwilliger TC, Waldo GS (2013) A new protein–protein interaction sensor based on tripartite split-GFP association. *Sci Rep* 3:2854. <https://doi.org/10.1038/srep02854>
- Canchi DR, Garcia AE (2013) Cosolvent effects on protein stability. *Annu Rev Phys Chem* 64:273–293. <https://doi.org/10.1146/annurev-physchem-040412-110156>
- Chautard H, Blas-Galindo E, Menguy T, Grand'Moursel L, Cava F, Berenguer J, Delcourt M (2007) An activity-independent selection system of thermostable protein variants. *Nat Methods* 4(11):919–921. <https://doi.org/10.1038/nmeth1090>
- Chen Y, Li S, Chen T, Hua H, Lin Z (2009) Random dissection to select for protein split sites and its application in protein fragment complementation. *Protein Sci* 18(2):399–409. <https://doi.org/10.1002/pro.42>

- Chen H, Ullah J, Jia J (2017) Progress in *Bacillus subtilis* spore surface display technology towards environment, vaccine development, and biocatalysis. *J Mol Microbiol Biotechnol* 27(3):159–167. <https://doi.org/10.1159/000475177>
- Cherf GM, Cochran JR (2015) Applications of yeast surface display for protein engineering. *Methods Mol Biol* 1319:155–175. https://doi.org/10.1007/978-1-4939-2748-7_8
- Cornvik T, Dahlroth SL, Magnusdottir A, Herman MD, Knaust R, Ekberg M, Nordlund P (2005) Colony filtration blot: a new screening method for soluble protein expression in *Escherichia coli*. *Nat Methods* 2(7):507–509. <https://doi.org/10.1038/nmeth767>
- Costa S, Almeida A, Castro A, Domingues L (2014) Fusion tags for protein solubility, purification and immunogenicity in *Escherichia coli*: the novel Fh8 system. *Front Microbiol* 5:63. <https://doi.org/10.3389/fmicb.2014.00063>
- DasSarma S, DasSarma P (2015) Gas vesicle nanoparticles for antigen display. *Vaccines* 3(3):686–702. <https://doi.org/10.3390/vaccines3030686>
- Daunert S, Barrett G, Feliciano JS, Shetty RS, Shrestha S, Smith-Spencer W (2000) Genetically engineered whole-cell sensing systems: coupling biological recognition with reporter genes. *Chem Rev* 100(7):2705–2738
- de Marco A, Deuerling E, Mogk A, Tomoyasu T, Bukau B (2007) Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. *BMC Biotechnol* 7:32. <https://doi.org/10.1186/1472-6750-7-32>
- DeLisa MP, Tullman D, Georgiou G (2003) Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. *Proc Natl Acad Sci USA* 100(10):6115–6120. <https://doi.org/10.1073/pnas.0937838100>
- Deller MC, Kong L, Rupp B (2016) Protein stability: a crystallographer's perspective. *Acta Crystallogr Sect F Struct Biol Commun* 72:72–95. <https://doi.org/10.1107/S2053230x15024619>
- Dreher ML, Gherardi E, Skerra A, Milstein C (1991) Colony assays for antibody fragments expressed in bacteria. *J Immunol Methods* 139(2):197–205. [https://doi.org/10.1016/0022-1759\(91\)90189-m](https://doi.org/10.1016/0022-1759(91)90189-m)
- Elena C, Ravasi P, Castelli ME, Peiru S, Menzella HG (2014) Expression of codon optimized genes in microbial systems: current industrial applications and perspectives. *Front Microbiol* 5:21. <https://doi.org/10.3389/fmicb.2014.00021>
- Esvelt KM, Carlson JC, Liu DR (2011) A system for the continuous directed evolution of biomolecules. *Nature* 472(7344):U499–U550. <https://doi.org/10.1038/nature09929>
- Fiacco SV, Kelderhouse LE, Hardy A, Peleg Y, Hu B, Ornelas A, Yang P, Gammon ST, Howell SM, Wang P, Takahashi TT, Millward SW, Roberts RW (2016) Directed evolution of scanning unnatural-protease-resistant (SUPR) peptides for in vivo applications. *ChemBioChem* 17(17):1643–1651. <https://doi.org/10.1002/cbic.201600253>
- Fisher AC, DeLisa MP (2009) Efficient isolation of soluble intracellular single-chain antibodies using the twin-arginine translocation machinery. *J Mol Biol* 385(1):299–311. <https://doi.org/10.1016/j.jmb.2008.10.051>
- Fisher AC, Kim W, DeLisa MP (2006) Genetic selection for protein solubility enabled by the folding quality control feature of the twin-arginine translocation pathway. *Protein Sci* 15(3):449–458. <https://doi.org/10.1110/ps.051902606>
- Fisher AC, Rocco MA, De Lisa MP (2011) Genetic selection of solubility-enhanced proteins using the twin-arginine translocation system. In: Evans JTC, Xu M-Q (eds) *Heterologous gene expression in E. coli: methods and protocols*. Humana Press, Totowa, pp 53–67. https://doi.org/10.1007/978-1-61737-967-3_4
- Foit L, Morgan GJ, Kern MJ, Steimer LR, von Hacht AA, Titchmarsh J, Warriner SL, Radford SE, Bardwell JC (2009) Optimizing protein stability in vivo. *Mol Cell* 36(5):861–871. <https://doi.org/10.1016/j.molcel.2009.11.022>
- Foit L, Mueller-Schickert A, Mamathambika BS, Gleiter S, Klaska CL, Ren G, Bardwell JC (2011) Genetic selection for enhanced folding in vivo targets the Cys14-Cys38 disulfide bond in bovine pancreatic trypsin inhibitor. *Antioxid Redox Signal* 14(6):973–984. <https://doi.org/10.1089/ars.2010.3712>
- Francis JC, Hansche PE (1972) Directed evolution of metabolic pathways in microbial populations. I. Modification of the acid phosphatase pH optimum in *S. cerevisiae*. *Genetics* 70(1):59–73
- Galan A, Comor L, Horvatic A, Kules J, Guillemin N, Mrljak V, Bhide M (2016) Library-based display technologies: where do we stand? *Mol BioSyst* 12(8):2342–2358. <https://doi.org/10.1039/c6mb00219f>
- Galarneau A, Primeau M, Trudeau L-E, Michnick SW (2002) β -Lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein-protein interactions. *Nat Biotechnol* 20(6):619–622. <https://doi.org/10.1038/nbt0602-619>
- He M, Taussig MJ (2007) Eukaryotic ribosome display with in situ DNA recovery. *Nat Methods* 4(3):281–288. <https://doi.org/10.1038/nmeth1001>
- Homaei AA, Sariri R, Vianello F, Stevanato R (2013) Enzyme immobilization: an update. *J Chem Biol* 6(4):185–205. <https://doi.org/10.1007/s12154-013-0102-9>
- Janocha S, Bichet A, Zollner A, Bernhardt R (2011) Substitution of lysine with glutamic acid at position 193 in bovine CYP11A1 significantly affects protein oligomerization and solubility but not enzymatic activity. *Biochim Biophys Acta* 1814(1):126–131. <https://doi.org/10.1016/j.bbapa.2010.06.002>
- Jermutus L, Honegger A, Schwesinger F, Hanes J, Pluckthun A (2001) Tailoring in vitro evolution for protein affinity or stability. *Proc Natl Acad Sci USA* 98(1):75–80. <https://doi.org/10.1073/pnas.011311398>
- Jespersen L, Schon O, Famm K, Winter G (2004) Aggregation-resistant domain antibodies selected on phage by heat denaturation. *Nat Biotechnol* 22(9):1161–1165. <https://doi.org/10.1038/nbt1000>
- Kawasaki M, Inagaki F (2001) Random PCR-based screening for soluble domains using green fluorescent protein. *Biochem Biophys Res Commun* 280(3):842–844. <https://doi.org/10.1006/bbrc.2000.4229>
- Kieke MC, Shusta EV, Boder ET, Teyton L, Wittrup KD, Kranz DM (1999) Selection of functional T cell receptor mutants from a yeast surface-display library. *Proc Natl Acad Sci USA* 96(10):5651–5656. <https://doi.org/10.1073/pnas.96.10.5651>
- Kim DS, Song HN, Nam HJ, Kim SG, Park YS, Park JC, Woo EJ, Lim HK (2014) Directed evolution of human heavy chain variable domain (V-H) using in vivo protein fitness filter. *PLoS ONE* 9(6):e98178. <https://doi.org/10.1371/journal.pone.0098178>
- Knaust RK, Nordlund P (2001) Screening for soluble expression of recombinant proteins in a 96-well format. *Anal Biochem* 297(1):79–85. <https://doi.org/10.1006/abio.2001.5331>
- Kondo A, Ueda M (2004) Yeast cell-surface display—applications of molecular display. *Appl Microbiol Biotechnol* 64(1):28–40. <https://doi.org/10.1007/s00253-003-1492-3>
- Kosobokova EN, Skrypnik KA, Kosorukov VS (2016) Overview of fusion tags for recombinant proteins. *Biochemistry* 81(3):187–200. <https://doi.org/10.1134/S0006297916030019>
- Kothawala A, Kilpatrick K, Novoa JA, Segatori L (2012) Quantitative analysis of alpha-synuclein solubility in living cells using split GFP complementation. *PLoS ONE* 7(8):e43505. <https://doi.org/10.1371/journal.pone.0043505>
- Krumpe LR, Atkinson AJ, Smythers GW, Kandel A, Schumacher KM, McMahon JB, Makowski L, Mori T (2006) T7 lytic phage-displayed peptide libraries exhibit less sequence bias than M13 filamentous phage-displayed peptide libraries. *Proteomics* 6(15):4210–4222. <https://doi.org/10.1002/pmic.200500606>
- Kurz M, Gu K, Al-Gawari A, Lohse PA (2001) cDNA-protein fusions: covalent protein-gene conjugates for the in vitro selection of peptides and proteins. *ChemBioChem* 2(9):666–672. <https://doi.org/10.1093/nar/gng119>
- Lee SY, Choi JH, Xu Z (2003) Microbial cell-surface display. *Trends Biotechnol* 21(1):45–52
- Lindenburg L, Merckx M (2014) Engineering genetically encoded FRET sensors. *Sensors* 14(7):11691–11713. <https://doi.org/10.3390/s140711691>
- Lindman S, Hernandez-Garcia A, Szczepankiewicz O, Frohm B, Linse S (2010) In vivo protein stabilization based on fragment complementation and a split GFP system. *Proc Natl Acad Sci USA* 107(46):19826–19831. <https://doi.org/10.1073/pnas.1005689107>
- Listwan P, Terwilliger TC, Waldo GS (2009) Automated, high-throughput platform for protein solubility screening using a split-GFP system. *J Struct Funct Genomics* 10(1):47–55. <https://doi.org/10.1007/s10969-008-9049-4>
- Liu JW, Boucher Y, Stokes HW, Ollis DL (2006) Improving protein solubility: the use of the *Escherichia coli* dihydrofolate reductase gene as a fusion reporter. *Protein Expr Purif* 47(1):258–263. <https://doi.org/10.1016/j.pep.2005.11.019>

- Lofdahl PA, Nord O, Janzon L, Nygren PA (2009) Selection of TNF- α binding affibody molecules using a beta-lactamase protein fragment complementation assay. *N Biotechnol* 26(5):251–259. <https://doi.org/10.1016/j.nbt.2009.06.980>
- Magliery TJ, Wilson CGM, Pan WL, Mishler D, Ghosh I, Hamilton AD, Regan L (2005) Detecting protein–protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism. *J Am Chem Soc* 127(1):146–157. <https://doi.org/10.1021/ja046699g>
- Makeyev EV, Kolb VA, Spirin AS (1996) Enzymatic activity of the ribosome-bound nascent polypeptide. *FEBS Lett* 378(2):166–170. [https://doi.org/10.1016/0014-5793\(95\)01438-1](https://doi.org/10.1016/0014-5793(95)01438-1)
- Malik A, Mueller-Schickert A, Bardwell JC (2014) Cytosolic selection systems to study protein stability. *J Bacteriol* 196(24):4333–4343. <https://doi.org/10.1128/JB.02215-14>
- Matsuura T, Pluckthun A (2003) Selection based on the folding properties of proteins with ribosome display. *FEBS Lett* 539(1–3):24–28. [https://doi.org/10.1016/S0014-5793\(03\)00178-9](https://doi.org/10.1016/S0014-5793(03)00178-9)
- Maxwell KL, Mittermaier AK, Forman-Kay JD, Davidson AR (1999) A simple in vivo assay for increased protein solubility. *Protein Sci* 8(9):1908–1911. <https://doi.org/10.1111/ps.8.9.1908>
- Mo HM, Xu Y, Yu XW (2018) Improved soluble expression and catalytic activity of a thermostable esterase using a high-throughput screening system based on a split-GFP assembly. *J Agric Food Chem* 66(48):12756–12764. <https://doi.org/10.1021/acs.jafc.8b04646>
- Nagumo Y, Fujiwara K, Horisawa K, Yanagawa H, Doi N (2016) PURE mRNA display for in vitro selection of single-chain antibodies. *J Biochem* 159(5):519–526. <https://doi.org/10.1093/jb/mv131>
- Nemoto N, Miyamoto-Sato E, Husimi Y, Yanagawa H (1997) In vitro virus: bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. *FEBS Lett* 414(2):405–408. [https://doi.org/10.1016/s0014-5793\(97\)01026-0](https://doi.org/10.1016/s0014-5793(97)01026-0)
- Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ (1996) Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* 273(5280):1392–1395. <https://doi.org/10.1126/science.273.5280.1392>
- Packer MS, Liu DR (2015) Methods for the directed evolution of proteins. *Nat Rev Genet* 16(7):379–394. <https://doi.org/10.1038/nrg3927>
- Palmer T, Berks BC (2012) The twin-arginine translocation (Tat) protein export pathway. *Nat Rev Microbiol* 10(7):483–496. <https://doi.org/10.1038/nrmicro2814>
- Pan JG, Kim EJ, Yun CH (2012) Bacillus spore display. *Trends Biotechnol* 30(12):610–612. <https://doi.org/10.1016/j.tibtech.2012.09.005>
- Pershad K, Kay BK (2013) Generating thermal stable variants of protein domains through phage display. *Methods* 60(1):38–45. <https://doi.org/10.1016/j.jymeth.2012.12.009>
- Philipps B, Hennecke J, Glockshuber R (2003) FRET-based in vivo screening for protein folding and increased protein stability. *J Mol Biol* 327(1):239–249. [https://doi.org/10.1016/S0022-2836\(03\)00077-9](https://doi.org/10.1016/S0022-2836(03)00077-9)
- Pluckthun A (2012) Ribosome display: a perspective. *Methods Mol Biol* 805:3–28. https://doi.org/10.1007/978-1-61779-379-0_1
- Remy I, Michnick SW (2015) Mapping biochemical networks with protein fragment complementation assays. *Methods Mol Biol* 1278:467–481. https://doi.org/10.1007/978-1-4939-2425-7_31
- Roberts RW, Szostak JW (1997) RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc Natl Acad Sci USA* 94(23):12297–12302. <https://doi.org/10.1073/pnas.94.23.12297>
- Rodrigue A, Chanal A, Beck K, Muller M, Wu LF (1999) Co-translocation of a periplasmic enzyme complex by a hitchhiker mechanism through the bacterial Tat pathway. *J Biol Chem* 274(19):13223–13228. <https://doi.org/10.1074/jbc.274.19.13223>
- Rodriguez-Banqueri A, Errasti-Murugarren E, Bartoccioni P, Kowalczyk L, Peralvarez-Marín A, Palacin M, Vazquez-Ibar JL (2016) Stabilization of a prokaryotic LAT transporter by random mutagenesis. *J Gen Physiol* 147(4):353–368. <https://doi.org/10.1085/jgp.201511510>
- Scott DJ, Pluckthun A (2013) Direct molecular evolution of detergent-stable G protein-coupled receptors using polymer encapsulated cells. *J Mol Biol* 425(3):662–677. <https://doi.org/10.1016/j.jmb.2012.11.015>
- Seitz T, Bocola M, Claren J, Sterner R (2007) Stabilisation of a (beta alpha)8-barrel protein designed from identical half barrels. *J Mol Biol* 372(1):114–129. <https://doi.org/10.1016/j.jmb.2007.06.036>
- Sheng S, Jia H, Topiol S, Farinas ET (2017) Engineering CotA laccase for acidic pH stability using *Bacillus subtilis* spore display. *J Microbiol Biotechnol* 27(3):507–513. <https://doi.org/10.4014/jmb.1608.08026>
- Shusta EV, Holler PD, Kieke MC, Kranz DM, Wittrup KD (2000) Directed evolution of a stable scaffold for T-cell receptor engineering. *Nat Biotechnol* 18(7):754–759
- Silva C, Martins M, Jing S, Fu J, Cavaco-Paulo A (2018) Practical insights on enzyme stabilization. *Crit Rev Biotechnol* 38(3):335–350. <https://doi.org/10.1080/07388551.2017.1355294>
- Smith GP (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 228(4705):1315–1317. <https://doi.org/10.1126/science.4001944>
- Socha RD, Tokuriki N (2013) Modulating protein stability—directed evolution strategies for improved protein function. *FEBS J* 280(22):5582–5595. <https://doi.org/10.1111/febs.12354>
- Stuart ES, Morshed F, Sremac M, DasSarma S (2001) Antigen presentation using novel particulate organelles from halophilic archaea. *J Biotechnol* 88(2):119–128. [https://doi.org/10.1016/S0168-1656\(01\)00267-X](https://doi.org/10.1016/S0168-1656(01)00267-X)
- Takahashi TT, Austin RJ, Roberts RW (2003) mRNA display: ligand discovery, interaction analysis and beyond. *Trends Biochem Sci* 28(3):159–165. [https://doi.org/10.1016/S0968-0004\(03\)00036-7](https://doi.org/10.1016/S0968-0004(03)00036-7)
- Tamura T, Hamachi I (2014) Recent progress in design of protein-based fluorescent biosensors and their cellular applications. *ACS Chem Biol* 9(12):2708–2717. <https://doi.org/10.1021/cb500661v>
- Tan Y, Tian T, Liu W, Zhu Z, Yang C J (2016) Advance in phage display technology for bioanalysis. *Biotechnol J* 11(6):732–745. <https://doi.org/10.1002/biot.201500458>
- Taverna DM, Goldstein RA (2002) Why are proteins marginally stable? *Proteins* 46(1):105–109. <https://doi.org/10.1002/prot.10016>
- Ueno S, Nemoto N (2012) cDNA display: rapid stabilization of mRNA display. *Methods Mol Biol* 805:113–135. https://doi.org/10.1007/978-1-61779-379-0_8
- Waldo GS (2003) Genetic screens and directed evolution for protein solubility. *Curr Opin Chem Biol* 7(1):33–38. [https://doi.org/10.1016/S1367-5931\(02\)00017-0](https://doi.org/10.1016/S1367-5931(02)00017-0)
- Waldo GS, Standish BM, Berendzen J, Terwilliger TC (1999) Rapid protein-folding assay using green fluorescent protein. *Nat Biotechnol* 17(7):691–695. <https://doi.org/10.1038/10904>
- Wang H, Wang Y, Yang R (2017a) Recent progress in *Bacillus subtilis* spore-surface display: concept, progress, and future. *Appl Microbiol Biotechnol* 101(3):933–949. <https://doi.org/10.1007/s00253-016-8080-9>
- Wang J, Zhang T, Liu R, Song M, Wang J, Hong J, Chen Q, Liu H (2017b) Recurring sequence-structure motifs in (beta alpha)8-barrel proteins and experimental optimization of a chimeric protein designed based on such motifs. *Biochim Biophys Acta Proteins Proteom* 1865(2):165–175. <https://doi.org/10.1016/j.bbapap.2016.11.001>
- Wang TN, Badran AH, Huang TP, Liu DR (2018) Continuous directed evolution of proteins with improved soluble expression. *Nat Chem Biol* 14(10):972–980. <https://doi.org/10.1038/s41589-018-0121-5>
- Waraho D, DeLisa MP (2009) Versatile selection technology for intracellular protein–protein interactions mediated by a unique bacterial hitchhiker transport mechanism. *Proc Natl Acad Sci USA* 106(10):3692–3697. <https://doi.org/10.1073/pnas.0704048106>
- Waraho-Zhmayer D, Gkogka L, Yu TY, DeLisa MP (2013) A microbial sensor for discovering structural probes of protein misfolding and aggregation. *Prion* 7(2):151–156. <https://doi.org/10.4161/pri.23328>
- Waraho-Zhmayer D, Meksiriporn B, Portnoff AD, DeLisa MP (2014) Optimizing recombinant antibodies for intracellular function using hitchhiker-mediated survival selection. *Protein Eng Des Sel* 27(10):351–358. <https://doi.org/10.1093/protein/gzu038>
- Wehrman T, Kleaveland B, Her JH, Balint RF, Blau HM (2002) Protein-protein interactions monitored in mammalian cells via complementation of beta-lactamase enzyme fragments. *Proc Natl Acad Sci USA* 99(6):3469–3474. <https://doi.org/10.1073/pnas.062043699>
- Wirsching F, Keller M, Hildmann C, Riester D, Schwienhorst A (2003) Directed evolution towards protease-resistant hirudin variants. *Mol Genet Metab* 80(4):451–462
- Xiong P, Wang M, Zhou X, Zhang T, Zhang J, Chen Q, Liu H (2014) Protein design with a comprehensive statistical energy function and boosted by experimental selection for foldability. *Nat Commun* 5:5330. <https://doi.org/10.1038/ncomms6330>

- Yang H, Liu L, Li J, Chen J, Du G (2015) Rational design to improve protein thermostability: recent advances and prospects. *ChemBioEng Rev* 2(2):87–94. <https://doi.org/10.1002/cben.201400032>
- Yong KJ, Scott DJ (2015) Rapid directed evolution of stabilized proteins with cellular high-throughput encapsulation solubilization and screening (CHESS). *Biotechnol Bioeng* 112(3):438–446. <https://doi.org/10.1002/bit.25451>
- Yoshikuni Y, Martin VJ, Ferrin TE, Keasling JD (2006) Engineering cotton (+)-delta-cadinene synthase to an altered function: germacrene D-4-ol

synthase. *Chem Biol* 13(1):91–98. <https://doi.org/10.1016/j.chembiol.2005.10.016>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
