CellPress
OPEN ACCESS

Trends in Chemistry

Focus issue: Protein engineering and chemoenzymatic synthesis

Review

High-throughput directed evolution: a golden era for protein science

Romany J. McLure,[1] Sheena E. Radford,[1,*] and David J. Brockwell[1,*]

Directed evolution is a robust and powerful tool for engineering new and/or improved functions in biomolecules for therapeutic and industrial applications, as well as to uncover fundamental insights into protein behaviour. It works by exploiting the principle of natural evolution and accelerating it through multiple rounds of gene diversification and selection. To evolve the desired property, an appropriate assay for the property of interest must be chosen. Here, we describe recent advances in the development of *in vitro* and *in vivo* diversification methods, as well as high-throughput assays for protein directed evolution. Using recent examples, we discuss the drawbacks and challenges of the array of diversification methods and selection assays and consider future challenges in the field.

Highlights

Different *in vivo* diversification methods are emerging to enable targeted mutagenesis of both genomic and plasmid DNA.

Split protein biosensors can be utilised for directed evolution of optimal biophysical properties in proteins relevant to medicine and biotechnology.

Innovative screening technologies are emerging for *in vivo* continuous evolution to enable engineering of proteins with enhanced biophysical properties, as well as new or improved functions.

Screening technologies can be utilised to enable high-throughput identification of beneficial mutants, as well as to study the relationship between sequence and function for disease- and biopharmaceutically-relevant proteins.

Screens can produce rich datasets that can inform computational algorithms to better understand and predict the fundamental mechanisms governing protein behaviour and thereby facilitate protein design.

The coming (of) age of directed evolution

Over the past 3.5 billion years, organisms have been adapting and evolving to increase their competitiveness. Many cellular processes are carried out by proteins for which evolution has generated functionality that is often beyond our current ability to rationally design. Consequently, protein engineers have been working to exploit and expedite Nature's evolutionary processes to evolve and improve different protein functions since the advent of recombinant DNA technology in the 1970s (Figure 1). Directed evolution utilises the principles of Darwinian evolution, whereby genetic diversity is introduced into the test protein, which is then subjected to a selective pressure (Figure 2A). Compared with natural evolution, directed evolution has higher mutation rates to accelerate the process. By using an appropriate genotype–phenotype screen, rare beneficial mutations are enriched and can be identified.

Advances in molecular biology, such as the discovery of restriction endonucleases and the invention of PCR, have facilitated the specific and rational engineering of proteins (Figure 1). These advances have permitted the study of proteins by enabling the effect of specific amino acid substitutions on a protein's biological function and stability to be explored, as well as allowing the creation of proteins with improved biophysical properties [1–5]. Over the past 50 years, there have been enormous advances in molecular biology and DNA sequencing technologies (Figure 1), allowing the field to advance from experiments limited to assessing the effect of a single amino acid substitution in a single protein-of-interest (POI), to deep mutational scanning (DMS), which allows simultaneous characterisation of millions of variants. For small proteins (e.g., up to ~150 residues in length), DMS allows measurement of the functional consequence of every possible amino acid substitution at every position of the protein sequence in a single experiment [2,3,6].

Directed evolution has been used to engineer proteins with improved functional and biophysical properties (catalytic activity [7], binding affinity [8], thermal [9], and thermodynamic [10] stability); to evolve novel functions [11]; to enable characterisation and understanding of the function and

[1]Astbury Centre for Structural Molecular Biology, School of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, Leeds, LS2 9JT, UK

*Correspondence:
s.e.radford@leeds.ac.uk (S.E. Radford) and d.j.brockwell@leeds.ac.uk (D.J. Brockwell).

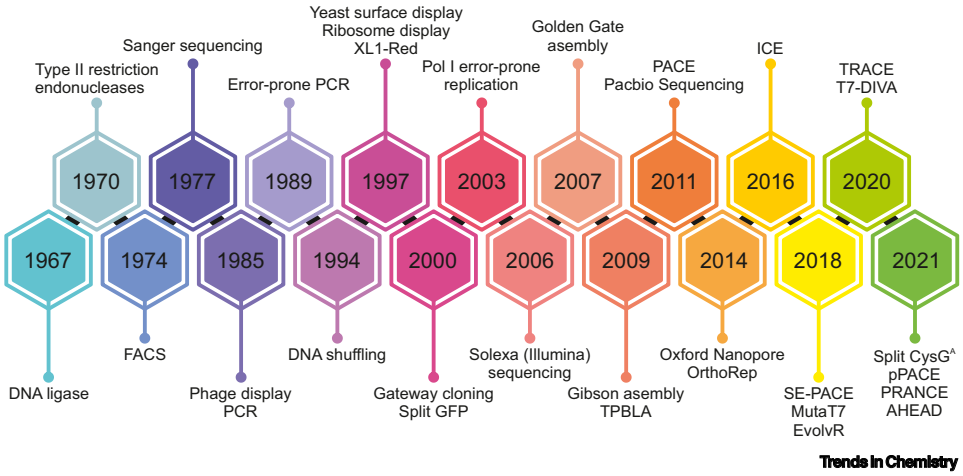

Figure 1. Major advances in molecular biology, next-generation sequencing, and protein engineering technologies. Early advances in molecular biology, such as the discovery of DNA ligase [99], type II restriction enzymes [100,101], and the invention of PCR [102,103], have paved the way for directed evolution and protein engineering as we know it today, subsequently enabling the invention of *in vitro* mutagenesis techniques such as error-prone PCR (epPCR) [97] and DNA shuffling [29]. Development of cloning methods such as Gateway cloning [104], Golden Gate [105], and Gibson assembly [106] facilitated creation of mutagenic libraries. DNA sequencing has advanced from low-throughput Sanger sequencing [107] to high-throughput next-generation technologies such as Illumina [108], Pacbio [109], and Oxford Nanopore [109], enabling rapid identification of evolved mutants at a massive scale. *In vivo* diversification has evolved from increasing the global mutagenesis rate with mutator strains such as XL1-Red [34], to targeted techniques such as OrthoRep [44], Ty1 retrotransposon mutagenesis or *in vivo* continuous evolution (ICE) [43], MutaT7 [37], EvolvR [40], T7 polymerase-driven continuous editing (TRACE) [39], and T7-targeted dCas9-limited *in vivo* mutagenesis (T7-DIVA) [42]. To accompany these diversification techniques, creative selection assays have been developed to enrich beneficial mutants. Fluorescence-activated cell sorting (FACS) [110] facilitated high-throughput identification of evolved proteins correlated to a fluorescent output. Early examples of evolving improved binding affinity include display technologies, such as phage [51], yeast surface [52], or ribosome [53] display. Protein reporter bioassays such as split green fluorescent protein (GFP) [64], tripartite β-lactamase assay (TPBLA) [80], and tripartite $CysG^A$ [10], have been developed to evolve beneficial biophysical properties. Selection platforms incorporating *in vivo* mutagenesis to enable continuous evolution include autonomous hypermutation yeast surface display (AHEAD) [49], phage-assisted continuous evolution (PACE) [11], soluble-expression PACE (SE-PACE) [4], periplasmic PACE (pPACE) [86], and phage-and-robotics-assisted near-continuous evolution (PRANCE) [14].

Glossary

Affibody: small protein scaffold designed as an antibody mimetic to bind to a variety of target proteins.
Framework: subdivision of the variable regions of an antibody. The part of the variable region that does not contain the complementarity determining regions (CDRs). The variable regions of an antibody contain three CDRs and three framework regions per domain.
Retrotransposon: transposable genetic elements that copy and paste themselves into different locations of a genome. The cycle works via transcription to mRNA, reverse-transcription into cDNA, then re-integration into the genome at a different location.
scFv: single-chain variable fragment. Artificial antibody fragment made from a fusion of the heavy and light variable domains of an antibody via a short linker.
Somatic hypermutation: part of the natural process of antibody maturation whereby large numbers of mutations accumulate within variable region genes of both antibody heavy and light chains.
Split intein: an intein is a segment within a larger protein that exhibits the ability to catalyse its own excision and simultaneous ligation of the flanking N- and C- terminal residues via a peptide bond. A split intein is where the intein segment is split between two proteins, thereby facilitating the ligation of two polypeptides into one.

stability of natural proteins [2,3]; as well as to engineer entire organisms [12]. These techniques have subsequently allowed research to extend beyond the confines of Nature; the directed evolution of novel tRNAs has enabled engineering of proteins beyond the 20 canonical amino acids [13–18], and the evolution of quadruplet tRNAs has the potential to further expand the genetic code with up to 255 unique amino acids possible (in principle) [14,19]. Furthermore, directed evolution has been invaluable in aiding in the optimisation of computationally designed proteins [20–22], as well as facilitating *de novo* protein design pipelines [23,24]. In this review, we outline different diversification methods, including *in vitro* and *in vivo* approaches, as well as novel selection strategies for evolving proteins with optimised functional and biophysical properties. Throughout we draw on recent examples that exemplify the power of these approaches to answer fundamental biological questions, as well as to address problems in medicine and biotechnology.

Methods of creating DNA libraries

In vitro mutagenesis

The first step in any directed evolution experiment is the creation of genetic diversity upon which selection pressures can be applied. Early work developing directed evolution techniques for engineering enzymes in the 1990s used random mutagenesis technologies to create genetic diversity [5,25,26]. Error-prone PCR (epPCR; Box 1) is by far the most popular of these

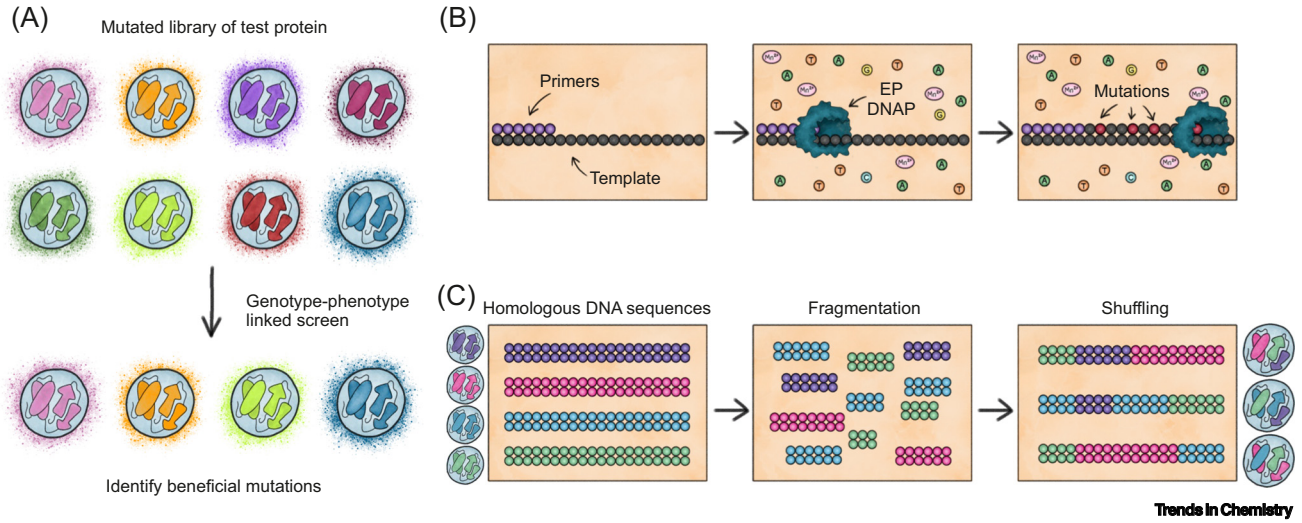

Figure 2. Directed evolution and *in vitro* mutagenesis. (A) A directed evolution experiment works by creating a library of gene variants of a protein-of-interest and subjecting them to a selective pressure to identify beneficial mutations. (B) Error-prone PCR (epPCR) uses an error-prone DNA polymerase (EP DNAP) to amplify a gene of interest and introduce mutations. Alternatively, the buffer conditions can be modified to increase the mutation rate of a standard DNAP, such as by adding magnesium ions (pink) or having unbalanced dNTP concentrations (A, green; T, orange; G, yellow; C, blue). (C) DNA shuffling allows mixing of homologous sequences, such as variants of the same protein with single point mutants, to create hybrid genes combining different mutations. Libraries are created by random fragmentation of genes, which are then joined together using primer-free PCR.

techniques owing to its ease of use (Figure 2B). epPCR is still widely employed and has been used successfully to engineer the properties of proteins, such as to increase aggregation-resistance [1], to improve enzyme activity [27], to evolve complex proteins useful for biotechnology [28], and to determine protein fitness landscapes using DMS [6,10]. Another often-used method for *in vitro* gene diversification is DNA shuffling (Box 1), wherein libraries are created by random fragmentation and recombination of homologous DNA sequences (Figure 2C) [29]. Since its invention, DNA shuffling has been widely used and adapted to engineer a range of properties, including improved thermostability [30] or catalytic activity [27], and, most recently, to develop chemogenetic fluorescent reporters with tuneable fluorescent properties [31].

Targeted gene mutagenesis methods have been developed to overcome the limitations of classic random mutagenesis methods and have been reviewed at length elsewhere [26]. In short, recent

Box 1. Traditional approaches to create libraries

Error-prone PCR (epPCR) works by using an error-prone DNA polymerase (DNAP) to randomly generate mutations during PCR amplification, or by modifying the buffer components to decrease the fidelity of a standard DNAP (see Figure 2B in main text) [97,98]. Reaction components can be modified to increase the mutation rate, such as using unbalanced dNTP concentrations, increasing the concentration of magnesium ions, increasing the number of PCR cycles, or adding manganese ions [98]. However, epPCR has limitations: often the DNAP has a bias for certain nucleotide substitutions over others, which can affect the amino acids available for a particular codon and, as a result, there is high chance for synonymous substitutions, therefore reducing library diversity. Mutations acquired in early PCR cycles can become dominant in the library compared with those acquired in later cycles, thereby biasing the library towards mutants acquired in early cycles. Furthermore, in epPCR, consecutive mutation of two bases is rare, which can further reduce the possible amino acids available; it requires large amounts of screening in order to sample the entire library; and can result in insertions and deletions (although at low frequencies), as well as the introduction of stop codons [26,98].

DNA shuffling makes use of fragmentation and recombination of homologous genes (see Figure 2C in main text) [29]. Genes are fragmented using DNase I and recombined using primer-free PCR where fragments with sufficient overlap will anneal to each other and be amplified. This approach is especially useful for mixing and combining a library of mutants that have already been evolved and selected as beneficial, in order to combine advantageous characteristics and improve them further.

advances in solid-phase DNA synthesis methods allow tight control over designed libraries to eliminate amino acid bias, allowing the effect of defined sets of amino acid substitutions in focussed regions of interest [e.g., the complementarity determining regions (CDRs) of antibodies] or the entire primary sequence to be determined [26]. Such libraries are particularly useful for DMS experiments as they allow understanding of protein functional landscapes and can be used to uncover the contribution of the identity (e.g., amino acid side chain chemistry) and individual residues to protein function, stability, and/or aggregation [32].

In vivo mutagenesis

In vivo mutagenesis approaches involve altering the genome sequence of an organism via the addition of mutagens (such as chemicals or UV light), or the use of hypermutator strains that contain deletions or modifications in genes for enzymes involved in proofreading, mismatch-repair, and base-excision (such as XL1-Red) [33,34]. Alternatively, various examples of mutagenic plasmids expressing different mutagenic enzymes involved in mismatch repair, translesion synthesis, and proof-reading have been developed with a wide range of induced mutagenic potency to globally increase the mutation rate in *Escherichia coli* [33]. These strategies have the potential to yield high mutation rates (up to 322 000-fold over wild type *E. coli*). Such methods can be problematic as the accumulation of mutations throughout the *E. coli* genome can result in toxic mutations if they occur within essential regions of the genome. Alternatively, these mutations accumulating outside of the gene-of-interest (GOI) could allow the bacteria to circumvent the selection pressure.

To overcome these limitations, targeted *in vivo* mutagenesis strategies have been developed. An early example of this strategy is the use of a mutated *E. coli* polymerase I (pol I) that selectively mutates genes on a ColE1 plasmid (although mutations are limited to within a few kb of the ColE1 origin) [35,36]. Furthermore, pol I still replicates parts of the genome, which can result in off-target mutations [35].

A popular method of *in vivo* mutagenesis is fusing specific DNA binding proteins to DNA-mutating enzymes. An example of this is MutaT7, wherein a cytidine deaminase is fused to T7 RNA polymerase (RNAP) to continuously direct mutations to specific, well-defined, DNA regions of any length in *E. coli* [37]. This allows targeted mutagenesis of genes under the control of the T7 promoter (Figure 3A). However, this approach has the potential to accumulate off-target effects, which can be problematic, particularly in the promoter regions. For example, they can potentially inhibit expression of the GOI, or lead to escape mutations, which allow the cells to evade the selection pressure applied without evolving the GOI. Furthermore, as this method utilises cytidine deaminases, their specific activity is limited to C>T and G>A mutations. Alternative cytidine deaminases have been employed to increase the mutation rate and expand the applicability of this method [38], and MutaT7 has also been adapted for use in eukaryotic cells (TRACE; T7 polymerase-driven continuous editing) [39].

A similar method (EvolvR), developed for use in both yeast and bacteria, utilises a fusion between an error-prone DNA polymerase (DNAP) and a nickase-Cas9 (nCas9), which allows mutations within a region adjacent to the Cas9 nick site (Figure 3B) [40,41]. The mutation rate can be tuned by using polymerases with different fidelities (~10^{-7}–10^{-3} per base) and this method enables all possible nucleotide substitutions, unlike those utilising cytidine deaminases. The approach is limited due to elevated (~10^{11}–10^{8} per base) off-target mutation rates and the narrow mutation window within the sequence (most mutations occur within 50 bp of the nick site).

T7-targeted dCas9-limited *in vivo* mutagenesis (T7-DIVA) utilises a similar method whereby T7 RNAP fused to a cytidine deaminase is used to introduce mutations (Figure 3C) [42]. The GOI

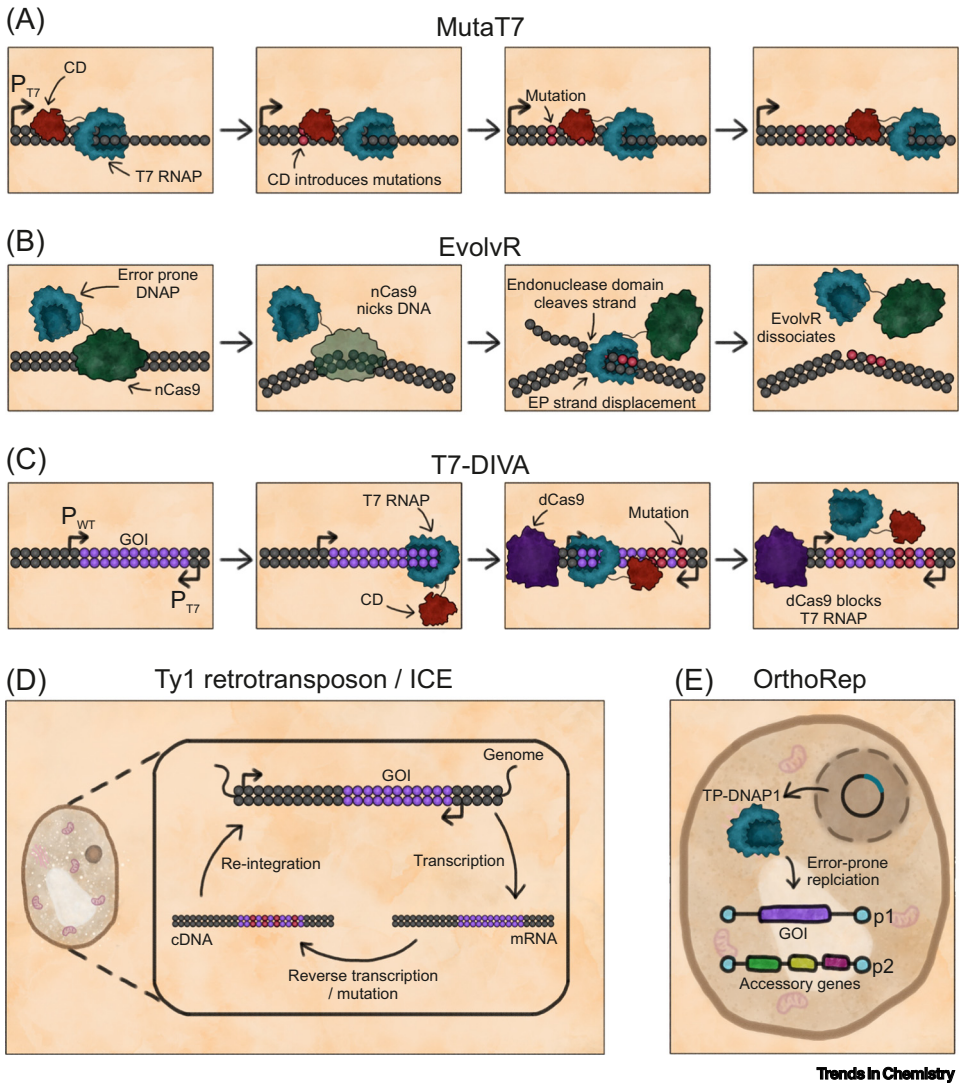

Figure 3. Schematic overview of *in vivo* diversification techniques. (A) MutaT7 uses a T7 RNA polymerase (RNAP) fused to a cytidine deaminase (CD), which enables targeted mutations to genes under the T7 promoter (P_{T7}). (B) EvolvR uses an error-prone DNA polymerase (DNAP) fused to a nickase Cas9 (nCas9), which enables targeted mutation within regions adjacent to the nick site via error-prone (EP) strand displacement. (C) T7-targeted dCas9-limited *in vivo* mutagenesis (T7-DIVA) enables targeted mutagenesis of genes without altering their genomic promoter (P_{WT}). By introducing an antisense P_{T7}, a T7 RNAP fused to a cytidine deaminase (CD) is able to introduce mutations. A catalytically dead Cas9 (dCas9) is used as a 'roadblock' to demarcate the boundaries of mutagenesis. (D) Ty1 retrotransposon mutagenesis, or *in vivo* continuous evolution (ICE), uses native yeast retrotransposon Ty1. The replication cycle of Ty1 is error-prone, so by introducing an inducible gene-of-interest (GOI), each time Ty1 is replicated mutations will accumulate. (E) OrthoRep uses an orthogonal plasmid/polymerase pair (TP-DNAP1/p1) whereby the error-prone TP-DNAP1 (expressed from a nuclear plasmid) replicates p1 and introduces mutations. All accessory genes required for the replication of p1 are encoded on a second plasmid (p2) to spare them from mutagenesis.

can remain under the control of its genomic promoter and a T7 promoter is inserted downstream of the GOI on the antisense strand. This allows the T7 RNAP to translocate along the GOI and to introduce mutations without altering the endogenous 5′ promoter. A catalytically dead Cas9 (dCas9) is used as a 'roadblock' demarcating the boundaries of the mutagenesis, enabling

targeted *in vivo* mutagenesis of specific genes. However, as this method requires introduction of a downstream T7 promoter, it is unable to mutate specific regions of a GOI.

Error-prone DNA replication utilising the native yeast **retrotransposon** (see Glossary) Ty1 has been developed for selective mutation of genes inserted between long terminal repeats (Figure 3D) [43]. The replication cycle of Ty1 occurs via an RNA intermediate that is converted into complementary DNA through an encoded reverse transcriptase and re-integrated back into the genome. Heterologous gene expression from Ty1 has previously been demonstrated and the replication cycle has been shown to be error-prone [43]. This enables random mutations to accumulate within a GOI expressed off Ty1, without any bias towards transitions or transversions over lengths of 5 kb [43]. However, as the diversification occurs across the whole length of the Ty1 retrotransposon element, mutations can accumulate within regulatory elements of the GOI or retrotransposon leading to escape mutations to evade the selection pressure [43]. Nonetheless, its large mutagenesis window makes this approach a powerful tool for *in vivo* continuous evolution of entire biosynthetic pathways [43].

OrthoRep is an extranuclear replication system in *Saccharomyces cerevisiae* consisting of an orthogonal DNA polymerase-DNA plasmid (TP-DNAP1/p1) pair [44]. It involves an engineered error-prone DNA polymerase (TP-DNAP1) that selectively replicates a specific plasmid (p1) encoding the GOI and introduces mutations (Figure 3E). TP-DNAP1 is expressed *in trans* from a yeast nuclear plasmid and a second polymerase/plasmid (TP-DNAP2/p2) pair encodes all the essential accessory genes for replication, transcription, and maintenance of p1 and p2, sparing them from error-prone replication and reducing off-target effects [45]. This method was developed further to adapt the TP-DNAP2/p2 pair for error-prone replication, which would allow two mutationally orthogonal DNA replication systems within the same cell, each with different custom mutation rates [46]. However, this method still does not enable targeted mutagenesis of specific regions of a GOI, as the polymerase replicates the entire plasmid. Nevertheless, OrthoRep has been used to evolve a wide range of proteins, including enzymes with promiscuous activities [47], small molecule biosensors [48], and antibody fragments [49].

Recent developments in screening technologies

Display technologies

The basic rule of directed evolution is that 'you get what you screen for' and selecting an appropriate screen is paramount [50]. Typically, screens are used to evolve proteins to have particular biophysical behaviours, such as aggregation-resistance and/or enhanced stability, or a new/improved function, such as improved enzyme performance or increased binding affinity. Commonly used methods to evolve binding affinity include phage display [51], yeast surface display [52], ribosome display [53], bacterial display [54], and mammalian display [55]. In all of these approaches, a POI is displayed on the surface of an organism and screened for binding affinity to a target antigen (for a recent review see [56]) (Figure 4). Each of these assays has been used extensively and adapted over the years to include selections for additional beneficial properties, such as increased solubility and aggregation-resistance [57,58], enhanced thermal stability [9], and resistance to acid-induced aggregation [59]. This is typically achieved by modifying the screening conditions (e.g., temperature, pH) to select for variants stable under the chosen conditions. These display technologies have been combined with next-generation sequencing (NGS) to allow high-throughput identification of specific binders, often utilising cell sorting technologies such as magnetic- or fluorescence-activated cell sorting (MACS or FACS, respectively) to separate and enrich for cells expressing proteins with the desired, improved properties [60–62].

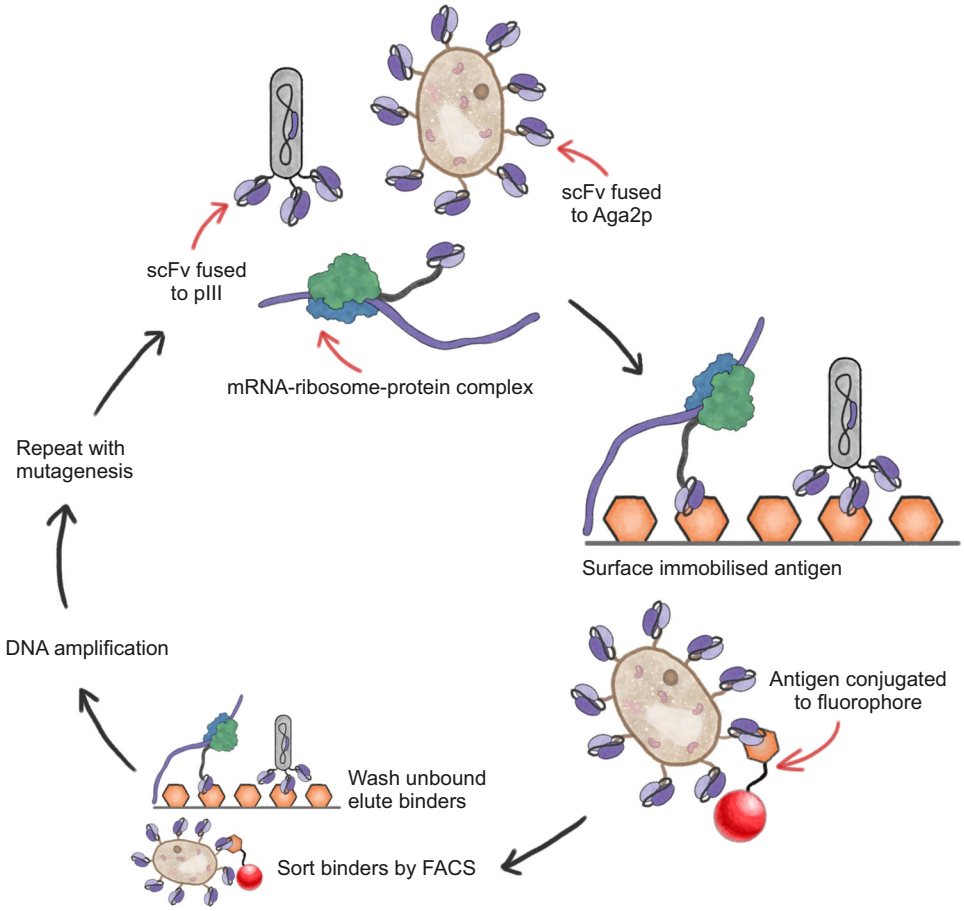

Figure 4. Schematic of display technologies. In phage display, the protein-of-interest (POI) (here displayed as a single-chain variable fragment (scFv)) is displayed on the surface of a filamentous bacteriophage via fusion to the phage coat protein III (pIII). Phage are screened for binding to a surface-immobilised antigen. During this step the conditions can be modified (pH, temperature, etc.) to alter the selection pressure. Unbound phage are washed away, bound phage are eluted and used to reinfect *Escherichia coli*. The DNA for the binders is isolated and used in a second round of selection to improve the affinity. In yeast surface display, scFvs are presented on the surface of *Saccharomyces cerevisiae* by fusion with the Aga2p protein. Cells are screened for binding by incubating with antigens conjugated to magnetic beads or a fluorophore, allowing magnetic-activated cell sorting (MACS) or fluorescence-activated cell sorting (FACS), respectively. The DNA of binders is isolated, and repeated rounds of mutation and cell sorting can be used to isolate variants with improved binding affinity. For ribosome display, an scFv DNA library lacking a stop codon is PCR amplified and transcribed to form mRNA *in vitro*. The lack of a stop codon causes the mRNA transcript to stall on the ribosome, forming an mRNA–ribosome–protein complex. These are screened against an immobilised antigen and binders are isolated. The mRNA of binders is isolated and reverse transcribed back to cDNA for sequencing. Repeated cycles of each of these display techniques using mutated libraries can improve the affinity for the target.

Protein reporter biosensors

Green fluorescent protein (GFP) is able to be split into two halves that can generate fluorescence via their noncovalent reassembly [63,64]. Making use of this property of GFP, an array of different systems has been developed enabling fluorescence to be correlated with protein stability, solubility, or the ability of a POI to interact with a target protein (Figure 5) [64–67]. However, the main issue with using fluorescent proteins as reporters is that GFP itself, which is added to the POI via a short linker, may alter the properties of the POI. Issues can arise because the fluorescent protein itself can remain fluorescent even when the POI aggregates if the rates of aggregation

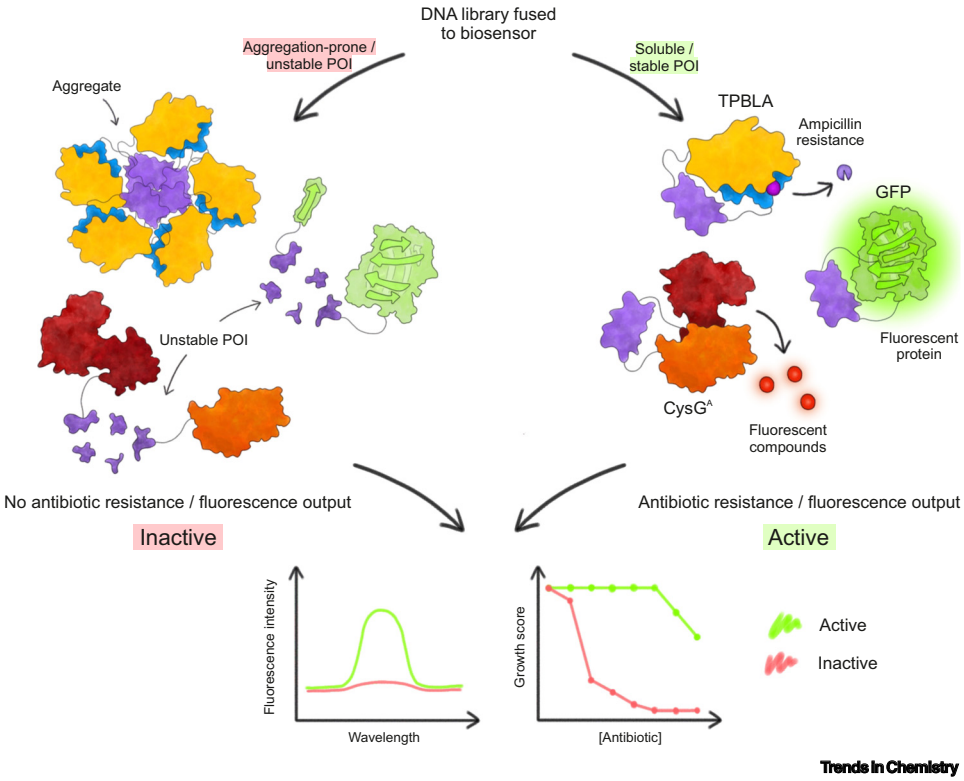

Figure 5. Overview of biosensors for evolving protein stability. DNA libraries are fused to a tripartite β-lactamase [tripartite β-lactamase assay (TPBLA)], green fluorescent protein (GFP), or CysG^A biosensor. Soluble and stable protein-of-interest (POI) (purple) will result in an antibiotic resistance (TPBLA) or fluorescence (GFP, CysG^A) readout, whereas unstable or aggregation-prone variants will form an inactive protein and give no antibiotic resistance or fluorescence readout.

of the POI are slower than the rate of formation of the chromophore [68]. Early examples of using a split GFP assay for evolving stability were relatively low throughput as they involved individually picking colonies of *E. coli* displaying increased fluorescence levels [65]. More recently, the split GFP assay has been expanded into a high-throughput assay by utilising FACS and deep sequencing to identify soluble variants of Gp2 (an **affibody**) [67].

A number of alternative split fluorescent proteins have been developed to expand the usefulness of these systems as they encompass a range of excitation and emission wavelengths [69–71]. Recently, a split luciferase-based biosensor was developed for detecting anti-severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) antibodies in patient sera [72,73]. This method has not yet been used for directed evolution, but it is similar in concept to the split GFP assay for evolving binding affinity and has the potential to be utilised in the same way [66,74].

Recently a tripartite biosensor using *E. coli* uroporphyrinogen-III methyltransferase CysG^A was developed in which the POI is inserted into a loop of CysG^A and used to evolve protein stability (Figure 5) [10]. CysG^A catalyses the formation of fluorescent compounds [10]. Therefore, by inserting a POI within a permissive site in CysG^A protein stability can be correlated with a fluorescence readout. The assay was first evaluated using variants of the *E. coli* immunity protein 7 (Im7), along with maltose binding protein and acylphosphatase. It was then used in a deep mutational scan to unpick the contribution of individual residues of the catalytic domain of a histone H3K4 methyltransferase to

understand its stability landscape [10]. As a result of the reducing environment of the *E. coli* cytoplasm, proteins that require disulfide bonds, such as antibody fragments and many enzymes, cannot be analysed using this system. In the first report of the CysG^{A} system the authors used manual inspection to select bacteria with increased fluorescence. Combining the assay with FACS followed by deep sequencing, however, has the potential to expand its capabilities so that variants with improved properties can be selected in a high-throughput manner.

A number of groups have demonstrated that β-lactamase can be used as a selectable reporter for engineering improved properties into proteins [75–79]. In one example, a tripartite β-lactamase construct [tripartite β-lactamase assay (TPBLA)] was used to correlate antibiotic resistance of *E. coli* with protein folding, thermodynamic stability, and aggregation-resistance of a POI [1,80,81]. To achieve this, a POI is inserted between two domains of TEM-1 β-lactamase and the resistance to β-lactam antibiotics of bacteria expressing the fusion protein is measured (Figure 5). This allows an *in vivo* measurement of successful protein folding and to rank proteins by their aggregation propensity, including Aβ_{40} and Aβ_{42}, wild type/D76N β_2-microglobulin (β_2m), and human/rat islet amyloid polypeptide (hIAPP/rIAPP) [81], as well as antibody fragments [single domains and single-chain variable fragment (**scFv**)] relevant to the biopharmaceutical industry [1]. Making use of the porosity of the *E. coli* outer membrane to small molecules (<600 Da), the assay has also been used as a screening method for identifying excipients [82] and small molecules [81] that inhibit protein aggregation. Furthermore, by introducing diversity into the POI by epPCR, the system has been utilised as a directed evolution assay to selectively engineer protein thermodynamic stability [80] and aggregation-resistance in a model scFv [1]. As this assay is carried out in the oxidising environment of the *E. coli* periplasm, it permits the formation of disulfide bonds, allowing analysis and evolution of proteins such as peptide hormones (e.g., hIAPP), immunoglobulin domains (i.e., β_2m), or antibody fragments (such as scFvs). Additionally, modification of the antibiotic concentration allows tight control over the level of selective pressure. Using deep sequencing to identify fitter variants that enable bacterial growth at increasingly high antibiotic concentrations, the TPBLA has the potential to unpick the complex relationship between sequence, thermodynamic stability, and aggregation for intrinsically disordered proteins, as well as globular POIs.

There is a limit to the sensitivity of these reporter protein assays to detect small increases in the folded fraction of a POI above a certain threshold (e.g., once >90% of the POI is correctly folded at 37°C). For example, in the TPBLA once a certain level of antibiotic resistance is achieved, the minimal inhibitory concentration (MIC) of antibiotic required to further select for stability is beyond the range of the experiment. Similarly, for fluorescence readouts, small increases in fluorescence may be masked by the broad distribution of the fluorescence intensity in the background. Additionally, as these systems select for correct protein folding and/or aggregation-resistance they neglect to select for function. This could be problematic when evolving a protein for enhanced biophysical properties that needs to maintain a function, such as an antibody fragment or an enzyme, as it could result in stability:function trade-off costs. Indeed, functional residues were found most often to be mutated when evolving protein stability using TPBLA, consistent with the concept of protein frustration (stability:function trade-off) [80]. This highlights the importance of choosing an appropriate selective assay for the system under investigation. Consequently, it can be necessary to develop orthogonal selection platforms with the ability to simultaneously evolve function and biophysical properties.

In vivo continuous evolution

One of the benefits of *in vivo* mutagenesis is that the approach allows continuous directed evolution of proteins, wherein continuous cycles of gene diversification and selection occur within a cell without any need for intervention. One popular example of this approach is phage-assisted

continuous evolution (PACE), which uses filamentous bacteriophage [selection phage (SP)] to evolve the desired property (Figure 6A) [11]. In this method, a population of SP is continuously diluted in a fixed volume of *E. coli*, known as the 'lagoon'. The GOI replaces the gene for the minor coat protein III (pIII) within the SP, a protein that is required for a phage to be infectious. The gene for pIII (gIII) is therefore supplemented on an accessory plasmid (AP), where its expression is linked to the function of the GOI being evolved. Variants will only persist if the resultant

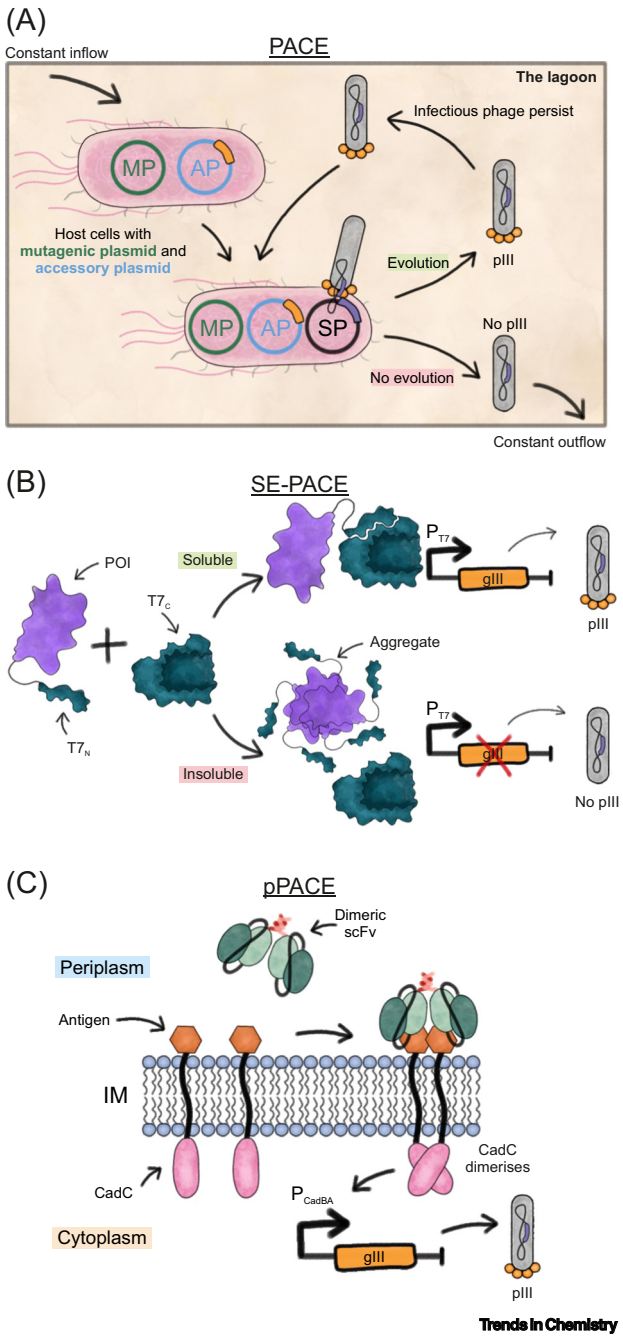

Figure 6. Schematic of phage-assisted continuous evolution (PACE) and its related screens. (A) In PACE, expression of phage protein pIII (via the gene gIII) is linked to the property of the protein-of-interest (POI) being evolved. Successful variants will be able to express pIII, and the resulting phage will be able to infect a new host. Unsuccessful variants will not be able to express pIII and the resulting phage will be diluted out of the fixed volume 'lagoon'. (B) Soluble expression PACE (SE-PACE) splits T7 RNAP into two parts ($T7_N$ and $T7_C$) and fuses a POI to $T7_N$. By expressing gIII under a T7 promoter (P_{T7}), only the expression of a soluble POI will result in gIII expression and infectious phage. (C) Periplasmic PACE (pPACE) replaces the periplasmic sensor domain of CadC with an antigen and puts gIII under the control of the P_{CadBA} promoter, which is switched on in response to dimerisation of CadC. By expressing a dimerising scFv, gIII will be expressed if the dimerising scFv binds to the antigen. This figure is adapted, with permission, from [4,11,86]. Abbreviations: AP, accessory plasmid; IM, inner membrane; MP, mutagenesis plasmid; SP, selection phage.

phage is able to become infectious before being diluted from the lagoon [11]. As the error rate is elevated using an inducible mutagenesis plasmid, mutations accumulate in the gene being evolved [11,33]. Since its conception, PACE has been used to evolve a diverse range of proteins, including polymerases with new recognition sites [11], dehydrogenases with improved activity [7], proteases with novel specificities [83,84], biosynthetic pathways [85], antibody fragments [4,86], proteins for DNA binding and manipulation [87–89], and novel quadruplet tRNAs for genetic code expansion [14,19].

As well as evolving protein function, PACE has been utilised to evolve protein solubility [soluble expression PACE (SE-PACE)] by linking pIII expression to the soluble expression levels of a POI [4]. SE-PACE has been adapted to use an 'AND gate' that utilises a **split intein** pIII to include a selection for binding affinity of antibody fragments to avoid loss of this property when evolving solubility (Figure 6B). This method has been utilised successfully to evolve scFvs with up to fivefold greater soluble expression yields and comparable binding affinities to the wild type. However, the complexity of this assay makes it difficult to use as it requires technical expertise and it is challenging to design appropriate genetic circuits to link the property being evolved to pIII expression. Furthermore, the fact that screening for binding occurs in the *E. coli* cytoplasm could be problematic for assessing proteins that contain disulfide bonds. Consequently, PACE has recently been adapted to carry out evolution in the oxidising environment of the *E. coli* periplasm, termed periplasmic PACE (pPACE) [86]. This approach uses the natural *E. coli* protein CadC, which is part of a two-component sensor that transduces signals in the periplasm to the cytoplasm (Figure 6C). CadC senses acidic pH and high lysine levels in the periplasm, causing the periplasmic sensor domain to dimerise and bind two motifs on the CadBA promoter and initiate gene transcription [90]. By replacing the periplasmic sensor domains of CadC with antigens and expressing a dimerising scFv, binding of these two proteins in the periplasm can be linked to gene expression of pIII, which is under the control of CadBA promoter [86]. pPACE has been used to evolve novel protein–protein interactions and restore binding between subunits of the homodimeric YibK; to restore binding affinity of a non-binding mutant of an anti-GCN4 Ω-graft antibody, as well as improve its soluble expression levels approximately eightfold; and to evolve an approximately twofold improvement in binding affinity and approximately fivefold improvement in soluble expression of the scFv of trastuzumab [86].

Despite its advantages, a number of challenges remain to be overcome with PACE: experiments have a high failure rate whereby phage expressing the evolving protein frequently 'wash out', meaning the selection pressure is too high, and experiments are difficult to multiplex [14]. To overcome this, PACE has been as miniaturised and extended as phage-and-robotics-assisted near-continuous evolution (PRANCE), which automates the process of continuous evolution utilising a liquid handling robot and a 96-well plate format to enable multiplexing [14]. To reduce the failure rate, PRANCE uses real-time monitoring of phage activity by expressing luciferase alongside pIII to give a read-out of phage propagation and to trigger a feedback control whereby selection pressure is modified depending on luminescence. PRANCE was used recently to characterise the evolutionary fitness landscape of T7 RNAP to recognise the foreign T3 promoter by conducting 90 simultaneous evolutions [14]. As small volumes are required, PRANCE also allows the evolution of aminoacyl-tRNA synthetases to incorporate non-natural amino acids, as well as allowing multiplexed evolution of quadruplet tRNAs [14], both of which require expensive reagents.

The gene diversification method OrthoRep, discussed earlier, has also been adapted for continuous evolution experiments and used to evolve drug-resistant malarial dihydrofolate reductases (DHFR) [91]. DHFR was encoded on the OrthoRep plasmid and dhfr1 (gene for DHFR) containing yeast were grown in the presence of a DHFR inhibitor (pyrimethamine) in a 96-well plate format to enable multiple evolution experiments in parallel [91]. The high-throughput of this method allows analysis of

evolution pathways and convergent evolution to assess protein fitness landscapes. Additionally, the OrthoRep diversification method has been combined with yeast surface display to develop 'autonomous hypermutation yeast surface display' (AHEAD) [49]. AHEAD has been used to identify nanobodies able to bind to the receptor binding domain of SARS-CoV-2. Starting with an individual clone, OrthoRep was used to introduce variance and FACS used to identify binders. Exploiting a computationally designed diverse naive nanobody library made up of 200 000 clones that encode key features of camelid immune repertoires, AHEAD has the potential to generate nanobody binders to a wide range of targets [49]. In this way, AHEAD is analogous to **somatic hypermutation**. The main difference between somatic hypermutation and *in vivo* mutagenesis within AHEAD is that somatic hypermutation often occurs within 'hotspots' of the gene, such as within the CDR regions, whereas mutations in AHEAD occur throughout the entire gene. However, this could be beneficial for binding, as the importance of **framework** mutations for both binding and stability have been described [92–94]. One of the disadvantages of AHEAD is the potential for accumulating mutations outside of the GOI, such as within the linker region or yeast anchor protein Aga2, which could affect the levels of display and allow the cells to circumvent the selection and result in escape mutants.

A similar method adapting yeast surface display for continuous evolution reduced the potential for off-target effects by using a SpyTag/SpyCatcher pair [95], where SpyTag is fused to the C-terminal end of a nanobody and SpyCatcher is fused to an anchor protein displaying it on the yeast surface [96]. The nanobody is displayed on the surface via post-translational protein ligation via isopeptide bond formation between SpyTag/Catcher. As SpyTag is only 16 amino acids in length compared with the 87 amino acid Aga2, the potential for accumulating off-target escape mutants is reduced. However, this method has not yet been combined with *in vivo* mutagenesis and successfully used in a continuous evolution experiment.

Concluding remarks

Directed evolution is a powerful tool for improving the biophysical properties of proteins for biopharmaceutical and industrial processes, for engineering new functions such as enzymes with new activity or tRNAs that incorporate noncanonical amino acids, as well as DMS experiments to uncover protein fitness landscapes and understand proteins in disease models. While there are many potential avenues that remain to be explored (see Outstanding questions), the palette of selection techniques now available to researchers, combined with the advent of low-cost NGS to allow high-throughput identification and analysis of the variants unmasked by these screens, is democratizing access to this incredible tool. It is now up to researchers to unleash this power onto new and exciting targets, but all without forgetting the golden rule of directed evolution: 'You get what you screen for'.

Outstanding questions

Residues involved in enzyme active sites are often localised in structure but not in sequence. Can we develop more accurate *in vivo* mutagenesis techniques that are able to diversify multiple specific regions at the same time?

Can we develop accurate targeted *in vivo* mutagenesis techniques to specifically diversify internal regions of proteins with a set length?

Can we create less biased targeted and global mutagenesis methods, both *in vivo* and *in vitro*, that equally sample all possible nucleotide substitutions?

A directed evolution screen must maintain a link between genotype and phenotype as DNA sequencing is utilised to identify positive mutants. Can we develop accurate protein sequencing methods that eliminate the need for this linkage?

Can we develop high-throughput long-read sequencing techniques for assessing coevolution? For the techniques that already exist (PacBio, Nanopore), can we make these cheaper and more accurate?

Can we harness the power of high-throughput directed evolution to enhance prediction of stability, solubility, and/or function of a new protein sequence?

Acknowledgments

R.J.M. is funded by the Biotechnology and Biological Sciences Research Council (BBSRC) and UCB Celltech (BB/M011151/1). S.E.R. is a Royal Society Research Professor (RSRP/RI/211057). We thank all the members of the Radford and Brockwell groups for their helpful discussions, particularly Jim Horne for reading this manuscript in draft form and providing many useful suggestions.

Declaration of interests

The authors declare they have no conflicts of interest.

References

1. Ebo, J.S. *et al.* (2020) An in vivo platform to select and evolve aggregation-resistant proteins. *Nat. Commun.* 11, 1816
2. Bolognesi, B. *et al.* (2019) The mutational landscape of a prion-like domain. *Nat. Commun.* 10, 4162
3. Newberry, R.W. *et al.* (2020) Deep mutational scanning reveals the structural basis for α-synuclein activity. *Nat. Chem. Biol.* 16, 653–659
4. Wang, T. *et al.* (2018) Continuous directed evolution of proteins with improved soluble expression. *Nat. Chem. Biol.* 14, 972–980

5. Chen, K. and Arnold, F.H. (1993) Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl. Acad. Sci. U. S. A.* 90, 5618–5622
6. Seuma, M. *et al.* (2021) The genetic landscape for amyloid beta fibril nucleation accurately discriminates familial Alzheimer's disease mutations. *eLife* 10, e63364
7. Roth, T.B. *et al.* (2019) Phage-assisted evolution of *Bacillus methanolicus* methanol dehydrogenase 2. *ACS Synth. Biol.* 8, 796–806
8. Julian, M.C. *et al.* (2019) Nature-inspired design and evolution of anti-amyloid antibodies. *J. Biol. Chem.* 294, 8438–8451
9. Pavoor, T.V. *et al.* (2012) An enhanced approach for engineering thermally stable proteins using yeast display. *Protein Eng. Des. Sel.* 25, 625–630
10. Ren, C. *et al.* (2021) An enzyme-based biosensor for monitoring and engineering protein stability in vivo. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2101618118
11. Esvelt, K.M. *et al.* (2011) A system for the continuous directed evolution of biomolecules. *Nature* 472, 499–503
12. Mans, R. *et al.* (2018) Under pressure: evolutionary engineering of yeast strains for improved performance in fuels and chemicals production. *Curr. Opin. Biotechnol.* 50, 47–56
13. Chin, J.W. *et al.* (2002) Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 99, 11020–11024
14. DeBenedictis, E.A. *et al.* (2022) Systematic molecular evolution enables robust biomolecule discovery. *Nat. Methods* 19, 55–64
15. Wang, L. *et al.* (2001) Expanding the genetic code of *Escherichia coli*. *Science* 292, 498–500
16. Park, H.-S. *et al.* (2011) Expanding the genetic code of *Escherichia coli* with phosphoserine. *Science* 333, 1151–1154
17. Reddington, S.C. *et al.* (2015) Directed evolution of GFP with non-natural amino acids identifies residues for augmenting and photoswitching fluorescence. *Chem. Sci.* 6, 1159–1166
18. Dumas, A. *et al.* (2015) Designing logical codon reassignment – expanding the chemistry in biology. *Chem. Sci.* 6, 50–69
19. DeBenedictis, E.A. *et al.* (2021) Multiplex suppression of four quadruplet codons via tRNA directed evolution. *Nat. Commun.* 12, 5706
20. Bunzel, H.A. *et al.* (2019) Emergence of a negative activation heat capacity during evolution of a designed enzyme. *J. Am. Chem. Soc.* 141, 11745–11748
21. Tinberg, C.E. *et al.* (2013) Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* 501, 212–216
22. Khersonsky, O. *et al.* (2011) Optimization of the in-silico-designed kemp eliminase KE70 by computational design and directed evolution. *J. Mol. Biol.* 407, 391–412
23. Cao, L. *et al.* (2020) De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science* 370, 426–431
24. Chevalier, A. *et al.* (2017) Massively parallel de novo protein design for targeted therapeutics. *Nature* 550, 74–79
25. Chen, K. and Arnold, F.H. (1991) Enzyme engineering for non-aqueous solvents: random mutagenesis to enhance activity of subtilisin E in polar organic media. *Bio/Technology* 9, 1073–1077
26. Currin, A. *et al.* (2021) The evolving art of creating genetic diversity: from directed evolution to synthetic biology. *Biotechnol. Adv.* 50, 107762
27. Nearmnala, P. *et al.* (2021) An in vivo selection system with tightly regulated gene expression enables directed evolution of highly efficient enzymes. *Sci. Rep.* 11, 11669
28. Tetter, S. *et al.* (2021) Evolution of a virus-like architecture and packaging mechanism in a repurposed bacterial protein. *Science* 372, 1220–1224
29. Stemmer, W.P. (1994) Rapid evolution of a protein in vitro by DNA shuffling. *Nature* 370, 389–391
30. Hao, J. and Berry, A. (2004) A thermostable variant of fructose bisphosphate aldolase constructed by directed evolution also shows increased stability in organic solvents. *Protein Eng. Des. Sel.* 17, 689–697
31. Benaissa, H. *et al.* (2021) Engineering of a fluorescent chemogenetic reporter with tunable color for advanced live-cell imaging. *Nat. Commun.* 12, 6989
32. Fowler, D.M. and Fields, S. (2014) Deep mutational scanning: a new style of protein science. *Nat. Methods* 11, 801–807
33. Badran, A.H. and Liu, D.R. (2015) Development of potent in vivo mutagenesis plasmids with broad mutational spectra. *Nat. Commun.* 6, 8425
34. Greener, A. *et al.* (1997) An efficient random mutagenesis technique using an *E. coli* mutator strain. *Mol. Biotechnol.* 7, 189–195
35. Allen, J.M. *et al.* (2011) Roles of DNA polymerase I in leading and lagging-strand replication defined by a high-resolution mutation footprint of ColE1 plasmid replication. *Nucleic Acids Res.* 39, 7020–7033
36. Camps, M. *et al.* (2003) Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9727–9732
37. Moore, C.L. *et al.* (2018) A processive protein chimera introduces mutations across defined DNA regions in vivo. *J. Am. Chem. Soc.* 140, 11560–11564
38. Park, H. and Kim, S. (2021) Gene-specific mutagenesis enables rapid continuous evolution of enzymes in vivo. *Nucleic Acids Res.* 49, e32
39. Chen, H. *et al.* (2020) Efficient, continuous mutagenesis in human cells using a pseudo-random DNA editor. *Nat. Biotechnol.* 38, 165–168
40. Halperin, S.O. *et al.* (2018) CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature* 560, 248–252
41. Tou, C.J. *et al.* (2020) Targeted diversification in the *S. cerevisiae* genome with CRISPR-guided DNA polymerase I. *ACS Synth. Biol.* 9, 1911–1916
42. Álvarez, B. *et al.* (2020) In vivo diversification of target genomic sites using processive base deaminase fusions blocked by dCas9. *Nat. Commun.* 11, 6436
43. Crook, N. *et al.* (2016) In vivo continuous evolution of genes and pathways in yeast. *Nat. Commun.* 7, 13051
44. Ravikumar, A. *et al.* (2014) An orthogonal DNA replication system in yeast. *Nat. Chem. Biol.* 10, 175–177
45. Javanpour, A.A. and Liu, C.C. (2019) Genetic compatibility and extensibility of orthogonal replication. *ACS Synth. Biol.* 8, 1249–1256
46. Arzumanyan, G.A. *et al.* (2018) Mutually orthogonal DNA replication systems in vivo. *ACS Synth. Biol.* 7, 1722–1729
47. Rix, G. *et al.* (2020) Scalable continuous evolution for the generation of diverse enzyme variants encompassing promiscuous activities. *Nat. Commun.* 11, 5644
48. Javanpour, A.A. and Liu, C.C. (2021) Evolving small-molecule biosensors with improved performance and reprogrammed ligand preference using OrthoRep. *ACS Synth. Biol.* 10, 2705–2714
49. Wellner, A. *et al.* (2021) Rapid generation of potent antibodies by autonomous hypermutation in yeast. *Nat. Chem. Biol.* 17, 1057–1064
50. You, L. and Arnold, F.H. (1996) Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide. *Protein Eng.* 9, 77–83
51. Smith, G.P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 228, 1315–1317
52. Boder, E.T. and Wittrup, K.D. (1997) Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* 15, 553–557
53. Hanes, J. *et al.* (1997) In vitro selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. U.S.A.* 94, 4937–4942
54. Bessette, P.H. *et al.* (2004) Rapid isolation of high-affinity protein binding peptides using bacterial display. *Protein Eng. Des. Sel.* 17, 731–739
55. Dyson, M.R. *et al.* (2020) Beyond affinity: selection of antibody variants with optimal biophysical properties and reduced immunogenicity from mammalian display libraries. *mAbs* 12, 1829335
56. Valldorf, B. *et al.* (2021) Antibody display technologies: selecting the cream of the crop. *Biol. Chem.* Published online March 23, 2021. https://doi.org/10.1515/hsz-2020-0377
57. Dudgeon, K. *et al.* (2012) General strategy for the generation of human antibody variable domains with increased aggregation resistance. *Proc. Natl. Acad. Sci. U. S. A.* 109, 10879–10884

58. Jespers, L. *et al.* (2004) Aggregation-resistant domain antibodies selected on phage by heat denaturation. *Nat. Biotechnol.* 22, 1161–1165
59. Famm, K. *et al.* (2008) Thermodynamically stable aggregation-resistant antibody domains through directed evolution. *J. Mol. Biol.* 376, 926–931
60. Christiansen, A. *et al.* (2015) High-throughput sequencing enhanced phage display enables the identification of patient-specific epitope motifs in serum. *Sci. Rep.* 5, 12913
61. Heyduk, E. and Heyduk, T. (2014) Ribosome display enhanced by next generation sequencing: a tool to identify antibody-specific peptide ligands. *Anal. Biochem.* 464, 73–82
62. McMahon, C. *et al.* (2018) Yeast surface display platform for rapid discovery of conformationally selective nanobodies. *Nat. Struct. Mol. Biol.* 25, 289–296
63. Baird, G.S. *et al.* (1999) Circular permutation and receptor insertion within green fluorescent proteins. *Proc. Natl. Acad. Sci. U.S.A.* 96, 11241–11246
64. Ghosh, I. *et al.* (2000) Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein. *J. Am. Chem. Soc.* 122, 5658–5659
65. Lindman, S. *et al.* (2010) In vivo protein stabilization based on fragment complementation and a split GFP system. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19826–19831
66. Magliery, T.J. *et al.* (2005) Detecting protein–protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism. *J. Am. Chem. Soc.* 127, 146–157
67. Golinski, A.W. *et al.* (2021) High-throughput developability assays enable library-scale identification of producible protein scaffold variants. *Proc. Natl. Acad. Sci. U.S.A.* 118, e2026658118
68. Kothawala, A. *et al.* (2012) Quantitative analysis of α-synuclein solubility in living cells using split GFP complementation. *PLoS One* 7, e43505
69. Feng, S. *et al.* (2017) Improved split fluorescent proteins for endogenous protein labeling. *Nat. Commun.* 8, 370
70. Feng, S. *et al.* (2019) Bright split red fluorescent proteins for the visualization of endogenous proteins and synapses. *Commun. Biol.* 2, 344
71. Tamura, R. *et al.* (2021) Multiplexed labeling of cellular proteins with split fluorescent protein tags. *Commun. Biol.* 4, 1–8
72. Elledge, S.K. *et al.* (2021) Engineering luminescent biosensors for point-of-care SARS-CoV-2 antibody detection. *Nat. Biotechnol.* 39, 928–935
73. Yao, Z. *et al.* (2021) A homogeneous split-luciferase assay for rapid and sensitive detection of anti-SARS CoV-2 antibodies. *Nat. Commun.* 12, 1806
74. Rozbeh, R. and Forchhammer, K. (2021) Split NanoLuc technology allows quantitation of interactions between PII protein and its receptors with unprecedented sensitivity and reveals transient interactions. *Sci. Rep.* 11, 12535
75. Galarneau, A. *et al.* (2002) Beta-lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein protein interactions. *Nat. Biotechnol.* 20, 619–622
76. Guntas, G. and Ostermeier, M. (2004) Creation of an allosteric enzyme by domain insertion. *J. Mol. Biol.* 336, 263–273
77. Guntas, G. *et al.* (2005) Directed evolution of protein switches and their application to the creation of ligand-binding proteins. *Proc. Natl. Acad. Sci. U.S.A.* 102, 11224–11229
78. Edwards, W.R. *et al.* (2010) Regulation of beta-lactamase activity by remote binding of heme: functional coupling of unrelated proteins through domain insertion. *Biochemistry* 49, 6541–6549
79. Edwards, W.R. *et al.* (2008) Linking the functions of unrelated proteins using a novel directed evolution domain insertion method. *Nucleic Acids Res.* 36, e78
80. Foit, L. *et al.* (2009) Optimizing protein stability in vivo. *Mol. Cell* 36, 861–871
81. Saunders, J.C. *et al.* (2016) An in vivo platform for identifying inhibitors of protein aggregation. *Nat. Chem. Biol.* 12, 94–101
82. Hailu, T.T. *et al.* (2013) In vivo detection and quantification of chemicals that enhance protein stability. *Anal. Biochem.* 434, 181–186
83. Packer, M.S. *et al.* (2017) Phage-assisted continuous evolution of proteases with altered substrate specificity. *Nat. Commun.* 8, 956
84. Blum, T.R. *et al.* (2021) Phage-assisted evolution of botulinum neurotoxin proteases with reprogrammed specificity. *Science* 371, 803–810
85. Johnston, C.W. *et al.* (2020) Continuous bioactivity-dependent evolution of an antibiotic biosynthetic pathway. *Nat. Commun.* 11, 4202
86. Morrison, M.S. *et al.* (2021) Disulfide-compatible phage-assisted continuous evolution in the periplasmic space. *Nat. Commun.* 12, 5959
87. Miller, S.M. *et al.* (2020) Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nat. Biotechnol.* 38, 471–481
88. Thuronyi, B.W. *et al.* (2019) Continuous evolution of base editors with expanded target compatibility and improved activity. *Nat. Biotechnol.* 37, 1070–1079
89. Richter, M.F. *et al.* (2020) Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* 38, 883–891
90. Kuper, C. and Jung, K. (2005) CadC-mediated activation of the cadBA promoter in *Escherichia coli*. *J. Mol. Microbiol. Biotechnol.* 10, 26–39
91. Ravikumar, A. *et al.* (2018) Scalable, continuous evolution of genes at mutation rates above genomic error thresholds. *Cell* 175, 1946–1957
92. Fernández-Quintero, M.L. *et al.* (2021) Mutation of framework residue h71 results in different antibody paratope states in solution. *Front. Immunol.* 12, 630034
93. Ovchinnikov, V. *et al.* (2018) Role of framework mutations and antibody flexibility in the evolution of broadly neutralizing antibodies. *eLife* 7, e33038
94. Warszawski, S. *et al.* (2019) Optimizing antibody affinity and stability by the automated design of the variable light-heavy chain interfaces. *PLoS Comput. Biol.* 15, 1–24
95. Zakeri, B. *et al.* (2012) Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci. U.S.A.* 109, E690–E697
96. Kajiwara, K. *et al.* (2021) Development of a yeast cell surface display method using the SpyTag/SpyCatcher system. *Sci. Rep.* 11, 11059
97. Leung, D. *et al.* (1989) A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* 1, 11–15
98. Wang, Y. *et al.* (2021) Directed evolution: methodologies and applications. *Chem. Rev.* 121, 12384–12444
99. Lehman, I.R. (1974) DNA ligase: structure, mechanism, and function. *Science* 186, 790–797
100. Kelly, T.J. and Smith, H.O. (1970) A restriction enzyme from *Hemophilus influenzae*. II. *J. Mol. Biol.* 51, 393–409
101. Smith, H.O. and Wilcox, K.W. (1970) A restriction enzyme from *Hemophilus influenzae*. I. purification and general properties. *J. Mol. Biol.* 51, 379–391
102. Saiki, R.K. *et al.* (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230, 1350–1354
103. Mullis, K. *et al.* (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* 51, 263–273
104. Hartley, J.L. *et al.* (2000) DNA cloning using in vitro site-specific recombination. *Genome Res.* 10, 1788–1795
105. Engler, C. *et al.* (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* 3, e3647
106. Gibson, D.G. *et al.* (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345
107. Sanger, F. *et al.* (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463–5467
108. Liu, L. *et al.* (2012) Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* 2012, 251364
109. van Dijk, E.L. *et al.* (2018) The third revolution in sequencing technology. *Trends Genet.* 34, 666–681
110. Picot, J. *et al.* (2012) Flow cytometry: retrospective, fundamentals and recent instrumentation. *Cytotechnology* 64, 109–130