Advances in the directed evolution of proteins
Michael D Lane1,2 and Burckhard Seelig1,2

Natural evolution has produced a great diversity of proteins that can be harnessed for numerous applications in biotechnology and pharmaceutical science. Commonly, specific applications require proteins to be tailored by protein engineering. Directed evolution is a type of protein engineering that yields proteins with the desired properties under well-defined conditions and in a practical time frame. While directed evolution has been employed for decades, recent creative developments enable the generation of proteins with previously inaccessible properties. Novel selection strategies, faster techniques, the inclusion of unnatural amino acids or modifications, and the symbiosis of rational design approaches and directed evolution continue to advance protein engineering.

Addresses
1 Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, Minneapolis, MN, USA
2 BioTechnology Institute, University of Minnesota, St. Paul, MN, USA

Corresponding author: Seelig, Burckhard (seelig@umn.edu)

Introduction
Synthetic biology describes the engineering of biological parts and whole systems by either modifying natural organisms or building new biosystems from scratch. Till date, most proteins used as parts in synthetic biology are taken from nature. Utilizing naturally evolved proteins has led to numerous successful applications in biotechnology. Nevertheless, these applications invariably benefit from an optimization of the original natural proteins by protein engineering [1]. In contrast, building entirely artificial proteins that do not resemble natural proteins is still a major challenge [2–4] and therefore much less common than the engineering of natural proteins for new or improved properties.

Protein engineering has developed into a multi-faceted field with hundreds of publications in the last two years alone. This field encompasses a variety of approaches for creating desired protein properties, ranging from purely computational design to selecting proteins from entirely random polypeptide libraries. Because of the incredible breadth of the field, and to enable us to focus on recent advances, we will direct the reader to excellent reviews on the fundamentals of directed evolution technologies [5–9] and computational protein design [10–12]. This review will therefore focus on the latest developments in the directed evolution of proteins (Figure 1).

Advancing selection technologies
In any directed evolution experiment, the isolation of the desired protein from a library of gene variants is the crucial step. Many efforts have been made to push the boundaries of evolution schemes, attempting to create better protein libraries, new selection systems with improved features, and faster selection procedures (Figure 2).

Maximizing library quality
The chance of discovering desired protein variants is directly related to the quality and complexity of the starting library. For example, random mutations that destabilize a protein can be detrimental. Therefore, building libraries with a high potential of containing functional proteins is vital. ‘Smarter’ libraries have been pursued that are less complex but of high-quality [13]. To build those libraries, targeted mutagenesis guided by structural or phylogenetic information, the use of compensatory stabilizing mutations and other approaches have successfully been applied [14,15]. Alternatively, a library maximizing complexity while enriching for well-folded proteins was constructed based on one of nature’s most common enzyme folds, the β/α8 barrel fold. All residues on the catalytic face of the protein scaffold were randomized and, simultaneously, the library was enriched for protease resistance by an mRNA display selection, which has been correlated with well-folded and therefore more likely functional proteins [16].

Refining selection steps
In vitro directed evolution of membrane proteins has been challenging due to toxicity of either the membrane protein or the selection conditions. Liposome display is a new method that has enabled in vitro directed evolution of toxic integral membrane proteins [17**]. This approach creates giant unilamellar liposomes and encapsulates a single DNA molecule along with a cell-free translation system. Each liposome will therefore display many copies of a single variant. Coupling protein activity to a fluorescent signal enables subsequent sorting by fluorescence-activated cell sorting (FACS). This approach was applied to evolve an α-hemolysin mutant with pore-forming
activity 30-fold greater than wild-type. In addition to membrane protein toxicity, selection conditions can be challenging when using lipid-based barriers, for example when selecting for stability in detergent. To overcome this issue, a cellular high-throughput encapsulation, solubilization, and screening method (CHESS) was developed to screen a library of G-protein coupled receptor (GPCR) variants [18]. GPCRs are an important group of drug targets. A library of 10^8 variants was expressed in Escherichia coli and the cells where then encapsulated in a polymer. The cells were lysed, but the ‘nano-container’ trapped the GPCR variants along with their encoding DNA while allowing free diffusion of fluorescent ligands and thereby enabling FACS. With this technique, functional receptors were identified in the presence of the detergent of choice.

The use of bead display for directed evolution has been limited by very few copies of DNA or displayed protein [19–23]. Recently, a ‘megavalent’ bead surface display (BeSD) system was developed to allow the display of protein and its encoding DNA in defined quantities up to a million copies per bead [24]. This method combines advantages of in vitro selections with multivalency of in vivo display systems, enabling the ranking and sorting of the output variants of an in vitro selection by flow cytometry.

Protease enzymes have a tremendous potential in medicine and biotechnology but engineering their activities via directed evolution for altered specificity, instead of simply broadening activity, has been successful until recently in only a few select cases using E. coli cell surface display of the E. coli outer membrane protease T [25]. This system is limited to the relatively few bacterial proteases that can be displayed and active on the prokaryote’s cell surface. To enable the engineering of more complex mammalian proteases, yeast surface display was modified to evolve novel protease specificity. In the revised system, both the protease variants and a yeast adhesion receptor were colocalized inside the endoplasmic reticulum (ER) through attached signal sequences [26**]. Successful proteolytic cleavage of a linker region detached the ER retention signal and enabled the yeast surface display of the adhesion receptor including its FLAG tag, which was then identified by anti-FLAG antibodies. Counter-selection tags were also incorporated to improve the selectivity of resulting protease variants. This method was used to alter the specificity of tobacco etch virus protease, as well as granzyme K and hepatitis C virus protease, and was even modified to demonstrate in principle the selection of kinase activity.

A directed evolution approach was devised to improve the targeting specificity of an engineered methyltransferase. Methylation of only a single site in a target DNA was selected for by digesting with a target site-specific restriction endonuclease and a second, unusual restriction enzyme that digests DNA with two distally methylated sites [27]. This method identified methyltransferase variants that showed 80% methylation at the target site and less than 1% methylation at off-target sites.

Phage assisted continuous evolution (PACE) enables the sustained evolution of protein variants through hundreds of rounds of evolution in a week with little researcher intervention [28]. This method was used to probe evolutionary pathway independence by evolving RNA polymersases for various promoter specificities [29*]. RNA polymersases that initially recognized the T7 promoter were evolved to recognize T3 or SP6 promoters separately, and then a final hybrid promoter of T3 and SP6. The resulting RNA polymersases from the SP6 pathway were ~3–4-fold more active than those from the T3 pathway and further evolution did not diminish this gap. Sequencing at multiple steps along the evolutionary path further illuminated that the divergent populations were unable to converge to the same solution. This suggests that it may be beneficial to evolve through multiple subpopulations instead of a single large population. In additional work, PACE was improved to allow the modulation of selection stringency via engineering phage propagation to be dependent on the small molecule anhydrotetracycline [30*]. Further, the authors enabled
counterselection to refine promoter specificity. The combination of these methods was used to create RNA polymerases with a 10 000-fold net change in promoter specificity. While this method now enables fast selections with advanced features such as counter-selection, it still can only be used to evolve proteins or activities that directly or indirectly involve expression, such as polymerase activity.

The use of chaperonins such as GroEL and GroES during directed evolution has been shown to allow more destabilizing mutations and mutations in the protein core to survive during evolution by stabilizing folding intermediates [31,32]. This chaperonin system was used to characterize the evolutionary pathway for a natural phosphotriesterase to a novel arylesterase [33*]. This study demonstrated for the first time on a molecular level...
how mutations found early during an evolutionary optimization yield larger improvements than later mutations. The results suggest that mutations seem to initially cluster near the active site and then radiate towards the rest of the enzyme to stabilize the early mutations.

**Improving speed of selection**

Two different strategies have significantly expedited the directed evolution process for *in vitro* selections. In the first strategy, many rounds of selection were performed very quickly, followed by sequencing and characterization of relatively few output sequences. For this purpose, a modified version of mRNA display, named ‘TRAP display’ was devised where the puromycin linker was attached simply via base pairing instead of covalent modification, enabling a round of selection in as little as 2.5 hours compared to the traditional 2–3 days [34*]. In just 14 hours and 6 rounds of selection, macrocyclic peptides with low nanomolar affinity against human serum albumin were selected.

In the second strategy, only a single round of stringent selection was performed, but then a large number of selected clones were analyzed by high-throughput sequencing to enable population-level statistical analysis. Following this approach, nanomolar affinity binders were identified after a single round of selection from the small protein scaffold 10Fn3 (10th fibronectin type III domain) with two random sequence loops, using the continuous flow magnetic separation mRNA display technology [35**]. The key to this approach was identifying the clones with the most enriched copy numbers after selection. In another example of the same strategy, multiple rounds of phage display biopanning of a heptapeptide library was performed against target cells, but high-throughput sequencing was performed at each step to assess the value of each round [36**]. Overall, a single round of screening was capable of identifying the best binders when sequenced at sufficient depth; and multiple rounds were only helpful in decreasing the background of non-binders when screening a small pool.

**Expanding the scope of selections to new properties**

**Photoresponsive binders**

Biological systems engineered to use light-sensing components have attracted attention due to their vast potential applications and have been recently reviewed [37]. Similar efforts have developed photo-reactive peptide aptamers for future use with *in vitro* or *in vivo* photo-regulation, immunoassays, or bio-imaging analyses. Ribosome display of peptides that contained azobenzene-modified lysine enabled the selection of UV-responsive streptavidin binders [38]. A different approach used a ribosome display scheme with a benzoazadiazole-modified phenylalanine to select for calmodulin binding peptides with single-digit micromolar affinity that fluoresce upon binding [39]. Two laboratories created cyclized peptides via azobenzene linkers to select for light-responsive peptides by phage display technology. A peptide cyclized via azobenzene-modified cysteines flanking a randomized 7 residue peptide was capable of binding to streptavidin in the dark and cease binding upon irradiation with a 22-fold discrimination [40]. Likewise, a synthesized photo-switchable azobenzene-based cyclization compound enabled the identification of peptide binders with single-digit micromolar affinity and a 3-fold change in affinity upon UV exposure [41]. Photoresponsive peptide ligands can now be selected with large UV-induced binding affinity changes. In the future, *in vivo* activity of evolved ligands needs to be demonstrated to further their use in optogenetics.

**Selecting unnatural peptide binders**

Binding peptides are valuable for a variety of purposes including detection assays and therapeutic applications. mRNA display and cDNA display have been used recently to create short peptide aptamers with unusual structure or composition, creating new diversity and thereby enabling the selection of binders with unique properties. In one approach, up to 12 unnatural amino acids were incorporated into an mRNA-displayed peptide library using a custom-mixed cell-free PURE translation system that was reconstituted from the purified components necessary for *E. coli* translation. The peptides were then cyclized via cysteine residues and the selection yielded unnatural peptides with nanomolar binding affinities for thrombin [42*]. A related approach created mRNA-displayed lantipeptides by using a translation system where lysine was substituted with 4-selenalysine, and inducing post-translational elimination via H$_2$O$_2$ and dehydroalanine. This provided an alternative cyclization mechanism for a drug candidate library and yielded binders with low micromolar affinity for sortase A [43*]. Similarly, a cDNA display library was cyclized [44] and a phage display library was bi-cyclized [45] through disulfide bonds formed in cysteine-rich peptides and used to select for peptide aptamers. Additional work on macrocyclic peptide selections has been recently reviewed [46].

Unnatural amino acids have also been used to decorate peptides and evolve multivalent glycopolypeptides [47]. Alkyne-containing glycine residues were incorporated into an mRNA-displayed peptide library to enable glycosylation via click chemistry. A selection against a broadly neutralizing antibody of HIV identified glycopolypeptides with potential as vaccine candidates.

**Multi-subunit protein selections**

Multi-subunit proteins are common, but directed evolution of these has been limited to *in vitro* methods, as opposed to *in vivo* methods that are capable of screening much larger libraries. To address this issue, an mRNA-displayed Fab fragment was entrapped in emulsion PCR
enabling the in vitro selection of heterodimeric Fab fragments [48]. However, the emulsion step reduced the library complexity to a similar range as in vivo methods. To achieve an in vitro selection of multi-subunit proteins of potentially up to 10^{14} variants, ribosome display was performed where one Fab subunit was randomized while the other subunit was held constant [49]. This was carried out with both heavy chain and light chain libraries yielding tight binders to VEGF and CEA. While cell-surface display of multi-subunit proteins has been performed before, a notable advance is the description of a mammalian cell-surface display that also features a titratable secretion of the same Fab fragments through alternate splicing of the pre-mRNA [50].

**In vitro compartmentalization selections applied to new reactions**

The performance of in vitro compartmentalization (IVC) methods has been improved in recent years with reported screening speeds of 2000 droplets per second for water-in-oil emulsion screening [51]. A number of creative protocols have been developed for this technology and applied to evolve enzymes capable of a range of chemical reactions. This includes a generalizable screen for hydrogenase activity [52], a selection for meganuclease specificity [53], an entirely microfluidic screen for hydrolytic activity of a sulfatase [54], and a quantitative screen for glucose oxidase activity [55]. While the above methods predominantly used custom-made microfluidic chips to sort their water-in-oil emulsions, another study described a generalizable method to produce water-in-oil-in-water emulsions that can be sorted by standard FACS equipment [56]. This protocol enabled the generation of monodisperse double emulsions at 6–12 000 droplets per second, which can be stored for months to years and manipulated to adjust their volume as needed. These droplets can be sorted in a commercial FACS-machine at 10–15 000 droplets per second, while enriching active variants by up to 100 000 fold. The downside to most IVC methods is that fluorescence must be linked to product formation.

**Combining directed evolution with rational design**

The combination of rational design to create informed libraries of variants with directed evolution to refine activity and efficiency has become a recurrent theme in protein engineering. A good example is the optimization of a computationally designed Kemp eliminase by directed evolution. Through rounds of error prone PCR, DNA shuffling, and site-directed mutagenesis, this artificial protein was refined to yield an enzyme that accelerated the reaction 6 × 10^{3}-fold, approaching the efficiency of natural enzymes [57].

In another study evolving an unrelated artificial Kemp eliminase, stabilizing consensus mutations were added during the library generation process. This stabilization facilitated the identification of a variant with >2000-fold improved catalytic efficiency after rounds of DNA shuffling, error prone PCR and selection [58]. Furthermore, an artificial Diels-Alderase was evolved by combining mutations from different rational design variants with rounds of error prone PCR. The modestly active original enzyme was thereby turned into a proficient biocatalyst for this abiological [4 + 2] cycloaddition reaction [59].

Cytochrome P450-derived enzymes can perform a variety of reactions and have been engineered to improve multiple properties [60]. A collection of cytochrome P450 mutants was screened for cyclopropanation of styrenes, and optimized through informed site-directed mutagenesis [61]. This enzymatic activity has not been observed in nature but is very useful to synthetic chemists. The work presents a great example of how the catalytic promiscuity of enzymes can be exploited. Using chemical intuition, the active site of one of the collection of cytochrome P450s was rationally re-designed to change the reduction potential of the heme-bound Fe^{II/III}, which allowed the efficient NAD(P)H-driven cyclopropanation while suppressing the native monooxygenation activity [62]. The modification therefore enabled the use of the P450 variant as a whole-cell catalyst. In another case, high regioselective and stereoselective hydroxylation of unactivated C–H bonds by a cytochrome P450 enzyme was achieved through a creative combination of active site mutagenesis, high-throughput ‘fingerprinting’ to identify functionally diverse variants, and fingerprint-driven reactivity predictions [63].

A mononuclear zinc metalloenzyme was computationally redesigned and then evolved with a combination of saturation mutagenesis and error prone PCR to create a variant that was 2500-fold more efficient than the initial design [64]. Structure-guided design of libraries of paraoxonase 1 and directed evolution led to variants capable of up to 340-fold higher catalytic efficiency for toxic isomers of G-type nerve agents [65]. Similarly, iterative saturation mutagenesis was used to evolve variants of phosphotriesterase capable of hydrolyzing V-type nerve agents with a 230-fold improvement of catalytic efficiency [66]. In a separate effort to create a toxin-neutralizing protein, a unique catalytic triad was computationally designed and subsequently optimized by yeast display. The resulting protein reacted with a fluorophosphonate probe at rates comparable to natural serine hydrolases, yet it was incapable of catalytic turnover [67].

Rational design has also been used to construct whole artificial protein scaffolds. In some cases, directed evolution was applied to these artificial structures to select for desired functions. Proteins from a combinatorial library of artificial four-helix bundle proteins were found to function in vivo by rescuing E. coli strains that lacked a
conditionally essential gene [68]. In addition, select four-helix bundle proteins bound to heme and exhibited peroxidase activity. This activity was improved by random mutagenesis and directed evolution [69]. Using structural principles of natural peroxidase proteins, designed ankyrin repeat proteins (DARPs) have been built and shown to function as artificial antibody mimetics. Recently, a flexibility loop and additional randomized regions were incorporated, creating the LoopDARPin scaffold [70]. A library based on this improved scaffold design yielded picomolar binding proteins after only a single round of selection by ribosome display.

Conclusion
With the investment of sufficient resources and determination, directed evolution generally appears to yield desired improvements of protein properties, sometimes producing remarkable results [71]. Therefore, one might be tempted to consider protein engineering a mature field. But those success stories mainly apply to improving or changing proteins that were provided by natural evolution. In contrast, the generation of novel activities without natural precedent is still in its infancy, although several examples have been reported [2,6,10,61**,72,73]. Achieving the synthetic biology goal of integrating artificial proteins into biological systems will introduce additional challenges, which again can be overcome with the help of directed evolution [9,74].

Acknowledgements
We thank RJ Kazlauskas, JC Haugner, A Morelli, S Dremel and DJ Morrow for critical reading of the manuscript. We gratefully acknowledge financial support from the Biocatalysis Initiative of the BioTechnology Institute at the University of Minnesota, the Minnesota Medical Foundation, the US National Aeronautics and Space Administration (NASA) NNX14AK29G and Agreement NN009AH70A through the NASA Astrobiology Institute–Ames Research Center, the US National Institutes of Health (NIH) (GM108703; MSTP grant T32 GM008244 to MDL), and the American Heart Association (MDL).

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


A library of 1013 soluble monomeric variants of nature’s favorite TIM barrel enzyme fold was generated with all residues on the catalytic face randomized. This library can be used as starting point to evolve new enzyme activities.


A generalizable approach to selecting membrane bound proteins was developed, combining giant unilamellar liposomes with in vitro translation machinery and FACS. This enabled the in vitro selection of membrane bound proteins, which are commonly toxic to living host cells when using in vivo methods, as well as the selection of multimeric proteins.


The “cellular high-throughput encapsulation, solubilization, and screening” (CHESS) method was created to enable screening of solubilized GPCR variants for detergent resistance.


49. In vitro compartmentalization combined with mRNA display enabled the selection of binders from a library of heterodimeric, multi-subunit Fab proteins.


The authors demonstrated that in vitro selection of the multi-subunit Fab protein can be performed by displaying a single randomized subunit and adding a constant synthetic heavy or light chain subunit.


An entirely in vitro compartmentalization system was developed that highlights the broad potential application of utilizing microfluidic systems to evolve proteins. They achieved a screening rate of 2000 events per second, which can be theoretically applied to many different types of reactions.


Highly monodisperse droplets and the ability to lyse cells inside the droplet created essentially a miniaturized cell lyase screening assay. The resulting precise quantitative readout facilitated the differentiation of small differences in hydrolytic activity of a sulfatase.


A generalizable method was created to enable the sorting of in vitro compartmentalized variants through forming water-in-oil-in-water double emulsions and sorting with commercial FACS equipment.


A computationally designed Kemp eliminase was improved through multiple methods of directed evolution, yielding an enzyme approaching the efficiency of natural enzymes.


The authors incorporated stabilizing consensus mutations while generating a library of variants to improve the stability of evolutionary intermediates. After multiple rounds of evolution, this led to identification of an improved variant.


Catalytic promiscuity of cytochrome P450 enabled the engineering of enzymes for the highly diastereo- and enantioselective olefin cyclopropanation. This project was guided by chemical intuition and led to an activity that has not been observed in nature and is synthetically important.


A clever combination of active site mutagenesis, screening, multivariate analysis of screening results, and a second round of screening of a few variants led to enzymes with impressive selectivities. This generalizable method should also work for other substrates.


Expanding upon the capabilities of the designed ankyrin repeat proteins (DARPins), the authors designed an improved scaffold which incorporates additional flexibility and randomized regions.


