Full title

Engineering of biomolecules by bacteriophage directed evolution

Authors

Andreas K. Brödel[1], Mark Isalan[1] and Alfonso Jaramillo[2,3,4*]

Affiliations

[1] Department of Life Sciences, Imperial College London, London SW7 2AZ, UK.

[2] Warwick Integrative Synthetic Biology Centre and School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK.

[3] CNRS-UMR8030, Laboratoire iSSB and Université Paris-Saclay and Université d'Évry and CEA, DRF, IG, Genoscope, Évry 91000, France.

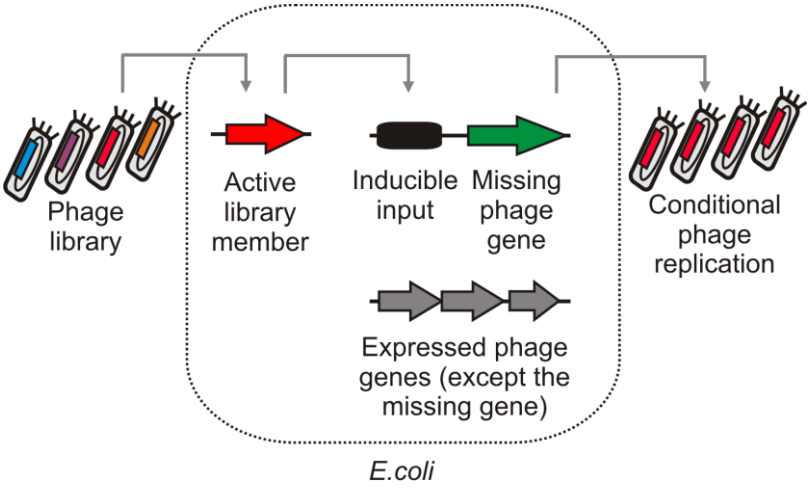
[4] Institute for Integrative Systems Biology (I2SysBio), University of Valencia-CSIC, 46980 Paterna, Spain.

***** To whom correspondence should be addressed: Prof. Alfonso Jaramillo, School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK. Alfonso.Jaramillo@synth-bio.org.

Abstract

Conventional *in vivo* directed evolution methods have primarily linked the biomolecule's activity to bacterial cell growth. Recent developments instead rely on the conditional growth of bacteriophages (phages), viruses that infect and replicate within bacteria. Here we review recent phage-based selection systems for *in vivo* directed evolution. These approaches have been applied to evolve a wide range of proteins including transcription factors, polymerases, proteases, DNA-binding proteins, and protein-protein interactions. Advances in this field expand the possible applications of protein and RNA engineering. This will ultimately result in new biomolecules with tailor-made properties, as well as giving us a better understanding of basic evolutionary processes.

Graphical abstract (submitted in separate TIFF file)



Highlights (submitted in separate word file)

- Directed evolution systems based on conditional phage replication (CPR) expand the potential of protein engineering.
- CPR platforms function in both batch and continuous culture.
- CPR systems bypass key limitations of conventional phage display.
- Directed evolution of a wide range of proteins can be achieved by CPR.

Introduction

Protein engineering enables the development of valuable biomolecules for pharmaceutical and biotechnological purposes. There are generally two strategies to guide protein engineering: rational design or directed evolution (**Figure 1**). Rational design usually uses computational tools and structural considerations to identify beneficial mutations in the protein of interest [1]. Recent advances in this strategy even allow the design of proteins completely *de novo* [2-6]. In comparison, directed evolution mimics natural evolution and starts with a population of genotype(s) and then proceeds with the iterative generation of genotype diversity and a selection based on linked phenotype activity. It is applied when too little structural or biochemical information is available to guide engineering. In many cases, these two strategies can be combined in a semi-rational approach to improve the activity of biomolecules [7,8]. This illustrates how the method must be chosen to fit the particular problem.

A variety of directed evolution techniques have been developed that employ customized gene circuits [9-12]. One commonly used approach is to link the target protein's activity to cell growth, which is particularly suitable when the evolving gene directly improves cellular fitness [13-15]. Alternatively, the use of phage particles offers a convenient way to uncouple the target protein's activity from the fitness function of a cell. Instead, an artificial genetic circuit couples the evolving protein's function to increasingly efficient production of phage packaging the gene of interest [16].

Directed evolution requires genotypic diversity in the gene of interest and this can either be achieved *in vivo* or *in vitro*. *In vivo* mutagenesis relies on intracellular modification of the target gene [17-19] whereas *in vitro* mutagenesis can be achieved extracellularly by chemical modification [20], ultraviolet irradiation [21], or polymerase chain reaction (PCR) [22]. PCR-based methods generally employ an error-prone polymerase or oligonucleotides that contain randomized bases at the desired positions. Chemical mutagenesis and irradiation are less commonly-used methods because of the lack of uniform mutational spectra [20,23]. By making randomized libraries or using a progressive series of mutations, it is possible to explore the 'design space' of a target gene, ultimately enabling the engineering of new proteins.

In this review, we first discuss the requirements for using phages to evolve biomolecules. We then focus on new directed evolution methods based on conditional phage replication that have been developed thanks to advances in molecular and synthetic biology.

Re-engineering phage-host genetic interactions to select functional biomolecules

When evolving a target gene from either a gene library or mutation system, the phenotype selection can either be performed outside a living cell (*in vitro*) or inside (*in vivo*). *In vivo* evolution systems allow selecting for more complex functions than *in vitro* methods (e.g. phage display) which are only suitable for binary protein-molecule interactions [24] (**Figure 2a**). By contrast, intracellular evolution potentially allows selection for multi-step processes, as long as they can be linked to genotype survival [25]. For example, intracellular processes can facilitate the simultaneous mutation and selection of the gene of interest. Furthermore, it enables the use of counterselections against an undesired biomolecule function [26]. Another advantage of intracellular evolution is the subsequent compatibility of evolved genes or complex gene networks with the entire host cell machinery, as these have to function in a host cell context. To exploit these advantages, alternative phage-assisted directed evolution platforms have been developed.

To allow enrichment of functional genes, phage selection systems require a link between the desired phenotype and conditional phage replication. This can be achieved by removing an essential gene required for phage replication from the phage genome and linking its expression to the function of the evolving biomolecule. Alternatively, this gene (or genes) could be a host co-factor required by the phage replication but dispensable to the cell (to allow cell survival in the uninfected cells that are required as a host reservoir). However, the only approaches developed so far rely on moving essential genes from the phage to the host cell or its associated plasmids [16,27] (**Figure 2b,c**). These systems may be classified according to the degree of phage engineering involved, where only a single gene may be moved or practically all of them.

The evolving biomolecule has to be encoded in the phage and a genetic system has to be designed to allow a functional molecule to activate the expression of the essential gene (**positive selection**). When the evolving biomolecule is able to induce the expression of the missing gene, infectious virions will package the DNA encoding the biomolecule, promoting its survival. The conditional expression of the essential gene can be done at the

transcriptional or post-transcriptional levels, depending on the biomolecule to be evolved (e.g. a transcription factor or a riboregulator).

Alternatively, selection may consist in designing a conditional interference with phage replication if a biomolecule is functional (**negative selection**). This is used to penalize any unwanted activity such as the original parental function of the biomolecule. The selection can also be complex or variable, where the stringency of positive and negative selection can be modulated exogenously [26].

Many alternative phage-host systems can in principle be chosen for the evolution of biomolecules depending on the application. For instance, if one wanted to evolve a photosynthetic protein, one might choose a cyanobacterium and one of its known phages. The disadvantage of such approaches is that the phage biology is not well characterized. Consequently, in this article we will focus on *E. coli* due to the lack of reported works with other organisms. The *E. coli* phages M13 [28], T4 [29,30], T7 [31] or λ [32] have been used to optimize protein function and stability with phage display, although M13 has been the only phage vector used to evolve biomolecules *in vivo* thus far.

Evolving biomolecules through positive selection

Recently, a new method to evolve biomolecules using M13 was developed, using a redesign of the host to implement a positive selection: Phage-Assisted Continuous Evolution (PACE) describes a general approach for the directed evolution of proteins *in vivo* [16]. Using PACE, new T7 RNA polymerase (RNAP) variants against a T3 promoter have been evolved, which are not bound by the wild-type T7 RNAP. For this, the minor coat protein pIII is replaced by the evolving gene of interest on the packaged M13 genome and the activity of the evolving protein is linked to conditional expression of pIII on a second plasmid, named an accessory plasmid (**Figure 2b**). Only phage particles assembled with pIII are infectious and propagate fast enough to stay in continuous culture. Mutations only accumulate within the packaged phage genome containing the target gene, and not in the *E. coli* strain, as bacteria are discarded (new uninfected bacteria are continuously provided). Enhanced genetic variation is obtained by a third mutagenesis plasmid (MP) that increases the mutation rate of *E. coli* cells [17]. All mutator genes on this MP are under an arabinose-inducible P_{BAD} promoter allowing conditional mutagenesis only at the phage replication stage. In this way, a protein with desired characteristics can be evolved after dozens of reinfections within the continuous flow chamber.

Evolving biomolecules through negative selection

In many cases, the requirements for evolved proteins not only include target activity but also the avoidance of potential off-target effects. This can be achieved by engineering a negative selection to remove variants with unwanted properties, which can be implemented by down-regulating a gene required for phage replication [27,33]. Alternatively, one may exploit any of the known mechanisms by which a bacterium can counteract a phage infection [34]. PACE has been adapted for negative selection pressures by choosing an abortive infection mechanism, where the undesired activity (activation of the original promoter) was linked to the inhibition of phage propagation using a non-functional pIII variant [26].

Modulating selection stringency for new functions

An important challenge is the ability to maintain phage replication when there is a lack of initial function for the biomolecule to be evolved. In the original PACE approach, an intermediate selection system was used where the T7 RNAP was initially evolved to transcribe a hybrid T3-T7 promoter, which had some activity, to later switch the selection to the full-target T3 promoter [16]. This is actually very difficult to achieve because it requires engineering a hybrid promoter that is still active with the original polymerase. Therefore this cannot be easily generalised to other cases. Fortunately, an alternative method was proposed that does not require re-engineering the target promoter [26] and instead relies on adding a second complementary copy of the gene used for selection (here gIII). This is similar to the hypothesis for the natural evolution of new functions *de novo* by gene duplication, where one gene duplicate maintains the original function, while the second copy is allowed to drift [35]. In the directed evolution case, the first gIII copy is under the control of a T3 promoter. The second copy is under the control of a T7 promoter, but the expression of this gIII is regulated (“stringency modulation”) to ensure this additional copy will cease to complement the original as the evolution progresses and the T7 RNAP acquires activity for the T3 promoter. Thus, the selection pressure is gradually increased over time to select the new function.

Tackling complex evolution pressures

Since the initial development of PACE, the platform has been adapted for the directed evolution of many different classes of proteins. For example, protease-PACE links the proteolysis of a target peptide to phage replication using a protease-activated RNA polymerase [36]. The system was used in the presence of two hepatitis C virus (HCV) protease inhibitor drug candidates (danoprevir and asunaprevir) to evolve HCV protease variants that possess up to 30-fold drug resistance. Strikingly, the predominant mutations

obtained in the HCV protease were consistent with the mutations observed in human patients treated with danoprevir or asunaprevir. Alternatively, DNA-binding PACE is a general method for the directed evolution of DNA-binding activity and specificity [25]. The platform was used to engineer transcription activator-like effector nucleases (TALENs) with improved DNA cleavage specificity [25]. On the other hand, protein-binding PACE enables the directed evolution of protein-protein interactions [37]. The authors evolved variants of the Bt toxin CrylAc against a cell receptor from the insect pest *Trichoplusia ni* with novel binding affinity that can ultimately overcome insect toxin resistance. PACE was also employed to continuously evolve T7 split RNA polymerases for downstream biosensor applications [38]. PACE has even been combined with high-throughput sequencing methods to improve downstream analysis which allows the characterization of whole protein populations as they adapt to selection pressures over time [39].

Evolution using phagemids

Phagemids can provide an alternative to classic full-phage selection systems. They have specific advantages, such as large library sizes and avoiding the mutation of phage genes. Consequently, we developed a phagemid selection system [27,33] where only the phagemid (PM) containing a library member and one essential phage gene (gIII) is packaged, while all the other phage components (except gVI) are provided on a modified helper phage (HP). To complete the system, an accessory plasmid (AP) contains a conditional gene VI circuit (**Figure 2c**). After infection, a protein with desired activity upregulates gene VI expression and therefore increases phage production. In this way, a protein with desired activity can be selected after several rounds of reinfection. Notably, our recently described system [27,33] uses conditional production of the minor coat protein pVI instead of pIII used in PACE. This is particularly useful for the directed evolution of transcription factors against basally-active promoters as expressed gIII in the starter culture would otherwise cause infection resistance resulting in a significantly decreased selection efficiency [40,41].

Phagemid selection has been applied for the directed evolution of a set of orthogonal transcription factors based on λ cI against synthetic promoters [27]. Negative selection against wild-type (WT) activity via repression has been achieved by putting the WT DNA sequence between the -35 and -10 regions of each synthetic promoter. The resulting toolkit contains 12 transcription factors, operating as activators, repressors, dual activator-repressors or dual repressor-repressors for the use in gene network engineering. Moreover, this evolution strategy functions in batch mode and therefore requires no special equipment

for reactor assembly, although it does rely on daily researcher interventions during selection [33].

Conclusion and perspectives

Recently developed directed evolution methods based on conditional phage replication further emphasize the strengths of phage-assisted protein engineering. These systems are particularly useful as they bypass key limitations of the widely-used phage display technology such as the simultaneous mutation and selection of complex biological functions. When choosing the most suitable method, various aspects including desired protein activity, available structural information, selection pressure and required selection efficiency need to be considered. Intracellular phage-assisted systems can, in principle, be used for all types of proteins, as long as their activity can be linked to conditional phage production (**Figure 3**). Notably, this is easier to achieve for cytosolic proteins than it is for complex proteins (e.g. membrane proteins). Furthermore, general limitations of bacterial expression over mammalian expression such as protein solubility, posttranslational modifications and disulfide bond formation have to be taken into account when using any phage-assisted technology.

Phages may also be used to evolve non-coding RNAs provided that their function can be linked to gene expression. This is particularly useful to complement computational designs of riboregulators [42], where a cognate regulatory sequence has to be added in the 5'UTR of the gene used for selection (for instance gene VI in [27,33]). Protein or RNA-based sensors (activating gene expression under the presence of a target chemical inducer) may also be encoded in the phage, provided one designs cycles of selections composed of two steps. The first step consists of a positive selection where the sensor may activate the infectious virion packaging in the presence of the chemical inducer. The second step occurs in the absence of the chemical inducer, where only sensors that do not activate the negative selection gene would be able to produce infectious virions. Similarly, negative selections may also be used to evolve the targeted function in the case of a negative regulator of gene expression (e.g., repressor). A negative selection would here act as an inverter such that constitutive phage replication could be used for evolving a repressor.

Advances in the fields of DNA sequencing, gene synthesis and genome engineering will likely reduce costs and improve the efficiency of current phage-assisted systems as well as drive the development of new technologies based on bacteriophages other than M13 [43].

These advances will also impact new mutagenesis strategies, in particular ones that enable targeted mutagenesis with improved mutation rates *in vivo*. The mutation of only the target gene(s) while not affecting any other genetic information is desirable in order to reduce the probability of selecting false positive variants in any directed evolution approach. As a consequence, phage-assisted evolution technologies will continue to play a key role in protein engineering efforts for basic as well as applied research.

Acknowledgments

AB, AJ and MI were supported by the European Commission grant FP7-ICT-2013-10 (no 610730, EVOPROG) and the BBSRC grant EVO-ENGINE BB/P020615/1. AJ is funded by FP7-KBBE (no 613745, PROMYS), H2020 Marie Sklodowska-Curie (no 642738, MetaRNA) and EPSRC-BBSRC (no BB/M017982/1, WISB center). MI is funded by New Investigator award no WT102944 from the Wellcome Trust U.K. The authors like to thank Tina Bartels for her critical reading of the manuscript.

Competing financial interests

The authors declare no competing financial interests.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- •
- •• outstanding

• Badran AH, Liu DR: **Development of potent in vivo mutagenesis plasmids with broad mutational spectra.** *Nat Commun* (2015) **6**(8425).
This study describes a plasmid-based mutagenesis system in *E. coli* that significantly enhances mutational efficiency compared to conventional *in vivo* methods.

• Hubbard BP, Badran AH, Zuris JA, Guilinger JP, Davis KM, Chen L, Tsai SQ, Sander JD, Joung JK, Liu DR: **Continuous directed evolution of DNA-binding proteins to improve talen specificity.** *Nature Methods* (2015).
This study describes a general approach for the phage-assisted continuous evolution of DNA-binding proteins. The authors used this method to generate transcription activator-like effectors nucleases (TALENs) with improved DNA cleavage specificity.

• Brödel AK, Jaramillo A, Isalan M: **Engineering orthogonal dual transcription factors for multi-input synthetic promoters.** *Nat Commun* (2016) **7**(13858).
This study describes the development of a selection system based on conditional phage replication in batch mode. The system was used to select a set of orthogonal transcription factors based on λ cI against synthetic promoters.

•• Esvelt KM, Carlson JC, Liu DR: **A system for the continuous directed evolution of biomolecules.** *Nature* (2011) **472**(7344):499-503.
This was the first study on phage-assisted continuous evolution (PACE). The authors used a general approach to evolve new T7 RNA polymerase variants that recognize distinct promoters.

•• Badran AH, Guzov VM, Huai Q, Kemp MM, Vishwanath P, Kain W, Nance AM, Evdokimov A, Moshiri F, Turner KH, Wang P *et al*: **Continuous evolution of bacillus thuringiensis toxins overcomes insect resistance.** *Nature* (2016) **533**(7601):58-63.
This study describes a general method for the phage-assisted continuous evolution of protein-protein interactions. The authors used a two-hybrid approach to evolve variants of the Bt toxin CrylAc that bind a cadherin-like receptor from the insect pest *Trichoplusia ni*.

1. Baker P, Seah SYK: **Rational approaches for engineering novel functionalities in carbon-carbon bond forming enzymes.** *Computational and Structural Biotechnology Journal* (2012) **2**(3):e201204003.

2. Huang P-S, Boyken SE, Baker D: **The coming of age of de novo protein design.** *Nature* (2016) **537**(7620):320-327.

3. Bale JB, Gonen S, Liu Y, Sheffler W, Ellis D, Thomas C, Cascio D, Yeates TO, Gonen T, King NP, Baker D: **Accurate design of megadalton-scale two-component icosahedral protein complexes.** *Science* (2016) **353**(6297):389.

4. Boyken SE, Chen Z, Groves B, Langan RA, Oberdorfer G, Ford A, Gilmore JM, Xu C, DiMaio F, Pereira JH, Sankaran B *et al*: **De novo design of protein homo-oligomers with modular hydrogen-bond network–mediated specificity.** *Science* (2016) **352**(6286):680.

5. Strauch E-M, Bernard SM, La D, Bohn AJ, Lee PS, Anderson CE, Nieusma T, Holstein CA, Garcia NK, Hooper KA, Ravichandran R *et al*: **Computational design of trimeric influenza-neutralizing proteins targeting the hemagglutinin receptor binding site.** *Nat Biotech* (2017) **advance online publication**(

6. Jaramillo A, Wernisch L, Héry S, Wodak SJ: **Folding free energy function selects native-like protein sequences in the core but not on the surface.** *Proceedings of the National Academy of Sciences* (2002) **99**(21):13554-13559.

7. Lutz S: **Beyond directed evolution - semi-rational protein engineering and design.** *Current opinion in biotechnology* (2010) **21**(6):734-743.

8. Chica RA, Doucet N, Pelletier JN: **Semi-rational approaches to engineering enzyme activity: Combining the benefits of directed evolution and rational design.** *Current Opinion in Biotechnology* (2005) **16**(4):378-384.

9. Ellefson JW, Meyer AJ, Hughes RA, Cannon JR, Brodbelt JS, Ellington AD: **Directed evolution of genetic parts and circuits by compartmentalized partnered replication.** *Nat Biotech* (2014) **32**(1):97-101.

10. Meyer AJ, Ellefson JW, Ellington AD: **Directed evolution of a panel of orthogonal t7 rna polymerase variants for in vivo or in vitro synthetic circuitry.** *ACS Synthetic Biology* (2014) 141007170011008.

11. Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z, Gonzales APW, Li Z, Peterson RT, Yeh J-RJ, Aryee MJ *et al*: **Engineered crispr-cas9 nucleases with altered pam specificities.** *Nature* (2015) **523**(7561):481-485.

12. Yokobayashi Y, Weiss R, Arnold FH: **Directed evolution of a genetic circuit.** *Proceedings of the National Academy of Sciences* (2002) **99**(26):16587-16591.

13. Fisher MA, McKinley KL, Bradley LH, Viola SR, Hecht MH: **De novo designed proteins from a library of artificial sequences function in escherichia coli and enable cell growth.** *PLOS ONE* (2011) **6**(1):e15364.

14. Kim Y-S, Jung H-C, Pan J-G: **Bacterial cell surface display of an enzyme library for selective screening of improved cellulase variants.** *Applied and Environmental Microbiology* (2000) **66**(2):788-793.

15. Digianantonio KM, Hecht MH: **A protein constructed de novo enables cell growth by altering gene regulation.** *Proceedings of the National Academy of Sciences of the United States of America* (2016) **113**(9):2400-2405.

16. Esvelt KM, Carlson JC, Liu DR: **A system for the continuous directed evolution of biomolecules.** *Nature* (2011) **472**(7344):499-503.

17. Badran AH, Liu DR: **Development of potent in vivo mutagenesis plasmids with broad mutational spectra.** *Nat Commun* (2015) **6**(8425).

18. Bridges BA, Woodgate R: **Mutagenic repair in escherichia coli: Products of the reca gene and of the umud and umuc genes act at different steps in uv-induced mutagenesis.** *Proceedings of the National Academy of Sciences of the United States of America* (1985) **82**(12):4193-4197.

19. Muteeb G, Sen R: **Random mutagenesis using a mutator strain.** In: *In vitro mutagenesis protocols: Third edition.* Braman J (Ed) Humana Press, Totowa, NJ (2010):411-419.

20. Lai Y-P, Huang J, Wang L-F, Li J, Wu Z-R: **A new approach to random mutagenesis in vitro.** *Biotechnology and Bioengineering* (2004) **86**(6):622-627.

21. Pfeifer GP, You Y-H, Besaratinia A: **Mutations induced by ultraviolet light.** *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* (2005) **571**(1–2):19-31.

22. Isalan M: **Construction of semi-randomized gene libraries with weighted oligonucleotide synthesis and pcr.** *Nat Protocols* (2006) **1**(1):468-475.

23. Myers RM, Lerman LS, Maniatis T: **A general method for saturation mutagenesis of cloned DNA fragments.** *Science* (1985) **229**(4710):242.

24. Badran AH, Liu DR: **In vivo continuous directed evolution.** *Current Opinion in Chemical Biology* (2015) **24**(1-10.

25. Hubbard BP, Badran AH, Zuris JA, Guilinger JP, Davis KM, Chen L, Tsai SQ, Sander JD, Joung JK, Liu DR: **Continuous directed evolution of DNA-binding proteins to improve talen specificity.** *Nature Methods* (2015).

26. Carlson J, Badran AH, Guggiana-Nilo DA, Liu DR: **Negative selection and stringency modulation in phage-assisted continuous evolution.** *Nature Chemical Biology* (2014) **10**(3):216-222.

27. Brödel AK, Jaramillo A, Isalan M: **Engineering orthogonal dual transcription factors for multi-input synthetic promoters.** *Nat Commun* (2016) **7**(13858).

28. Salmond GPC, Fineran PC: **A century of the phage: Past, present and future.** *Nat Rev Micro* (2015) **13**(12):777-786.

29. Efimov VP, Nepluev IV, Mesyanzhinov VV: **Bacteriophage t4 as a surface display vector.** *Virus Genes* (1995) **10**(2):173-177.

30. Gamkrelidze M, Dąbrowska K: **T4 bacteriophage as a phage display platform.** *Archives of Microbiology* (2014) **196**(7):473-479.

31. Danner S, Belasco JG: **T7 phage display: A novel genetic selection system for cloning rna-binding proteins from cdna libraries.** *Proceedings of the National Academy of Sciences* (2001) **98**(23):12954-12959.

32. Gamage LNA, Ellis J, Hayes S: **Immunogenicity of bacteriophage lambda particles displaying porcine circovirus 2 (pcv2) capsid protein epitopes.** *Vaccine* (2009) **27**(47):6595-6604.

33. Brödel AK, Jaramillo A, Isalan M: **Intracellular directed evolution of proteins from combinatorial libraries based on conditional phage replication.** *Nature Protocols* (2017) **12**(9):1830-1843.

34. Labrie SJ, Samson JE, Moineau S: **Bacteriophage resistance mechanisms.** *Nat Rev Micro* (2010) **8**(5):317-327.

35. Andersson DI, Jerlström-Hultqvist J, Näsvall J: **Evolution of new functions de novo and from preexisting genes.** *Cold Spring Harbor Perspectives in Biology* (2015) **7**(6).

36. Dickinson BC, Packer MS, Badran AH, Liu DR: **A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations.** *Nat Commun* (2014) **5**(5352).

37. Badran AH, Guzov VM, Huai Q, Kemp MM, Vishwanath P, Kain W, Nance AM, Evdokimov A, Moshiri F, Turner KH, Wang P *et al*: **Continuous evolution of bacillus thuringiensis toxins overcomes insect resistance.** *Nature* (2016) **533**(7601):58-63.

38. Pu J, Zinkus-Boltz J, Dickinson BC: **Evolution of a split rna polymerase as a versatile biosensor platform.** *Nat Chem Biol* (2017) **13**(4):432-438.

39. Dickinson BC, Leconte AM, Allen B, Esvelt KM, Liu DR: **Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution.** *Proceedings of the National Academy of Sciences of the United States of America* (2013) **110**(22):9007-9012.

40. Boeke J, Model P, Zinder N: **Effects of bacteriophage f1 gene iii protein on the host cell membrane.** *Mol Gen Genet* (1982) **186**(2):185-192.

41. Rakonjac J, Model P: **Roles of piii in filamentous phage assembly.** *Journal of Molecular Biology* (1998) **282**(1):25-41.

42. Rodrigo G, Landrain TE, Jaramillo A: **De novo automated design of small rna circuits for engineering synthetic riboregulation in living cells.** *Proceedings of the National Academy of Sciences* (2012) **109**(38):15271-15276.

43. Sagona AP, Grigonyte AM, MacDonald PR, Jaramillo A: **Genetically modified bacteriophages.** *Integrative Biology* (2016) **8**(4):465-474.

44. Stano NM, Patel SS: **T7 lysozyme represses t7 rna polymerase transcription by destabilizing the open complex during initiation.** *Journal of Biological Chemistry* (2004) **279**(16):16136-16143.

Figures

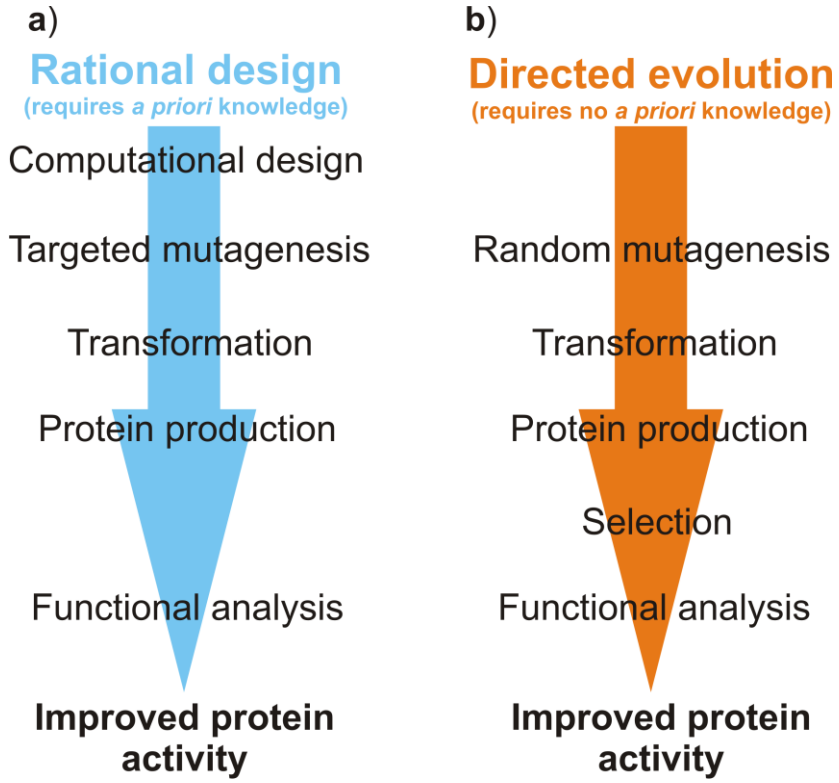


Figure 1. Protein engineering by rational design or directed evolution. a) Rational design uses computational tools as well as structural or other biochemical knowledge to identify beneficial mutations in the protein of interest. These mutations are inserted into the gene of interest (targeted mutagenesis) which is then expressed in host cells. Functional analysis for each protein variant is performed to confirm improved activity. **b)** Directed evolution is applied when too little structural or biochemical information is available to guide engineering. Mutations in the gene of interest are inserted randomly or by targeting specific positions in the gene sequence leading to a library of gene variants. Functional library members are then selected via a suitable selection system (e.g. phage-assisted evolution) against a target function. The activity of the selected protein is finally confirmed by functional analysis. Rational design and directed evolution are often combined to obtain the best results (semi-rational approach).

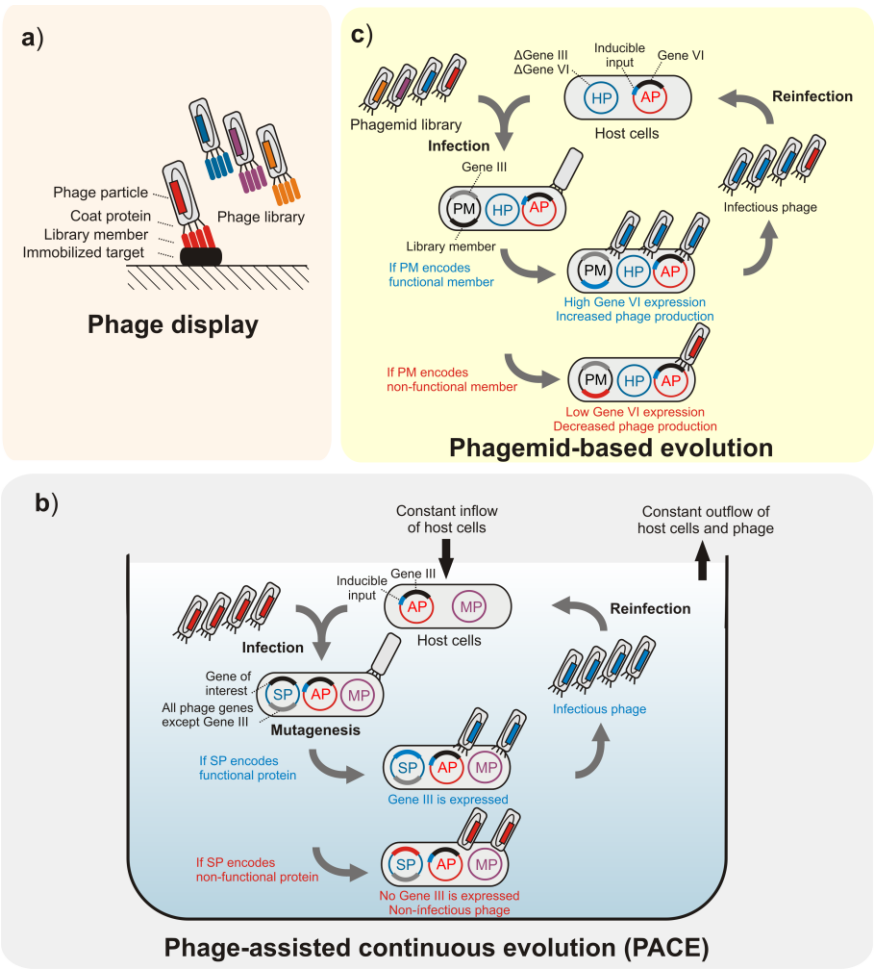


Figure 2. Phage-assisted directed evolution methods. a) Affinity selection of library members by phage display. Protein variants are fused to a phage coat protein and are displayed on phage particles providing a physical connection between genotype and phenotype. **b**) Phage-assisted continuous evolution (PACE) is based on conditional M13 phage replication. The activity of the evolving protein on the selection phage (SP; contains the gene of interest and all phage genes except gene III) is linked to gIII expression on the accessory plasmid (AP; contains a conditional gene III expression circuit). Only phage particles assembled with pIII are infectious and propagate fast enough to stay in continuous culture. The system uses a mutagenesis plasmid (MP) that increases the mutation rate of *E. coli* cells to generate target gene diversity. Mutations only accumulate within the packaged phage genome containing the target gene and not in the *E. coli* strain due to the continuous nature of the system. In this way, a protein with desired characteristics can be evolved after dozens of rounds of reinfection. **c**) Phagemid-based evolution from combinatorial libraries in batch mode. The library members are located on a packaged phagemid (PM) which also contains one essential phage gene (gIII). All the other phage genes are located on a modified helper phage (HP; contains all phage genes except genes III and VI) and an accessory plasmid (AP; contains a conditional gene VI expression circuit). After infection, a protein with desired activity upregulates gene VI expression and therefore increases phage production. In this way, a protein with desired activity can be selected after several rounds of reinfection.

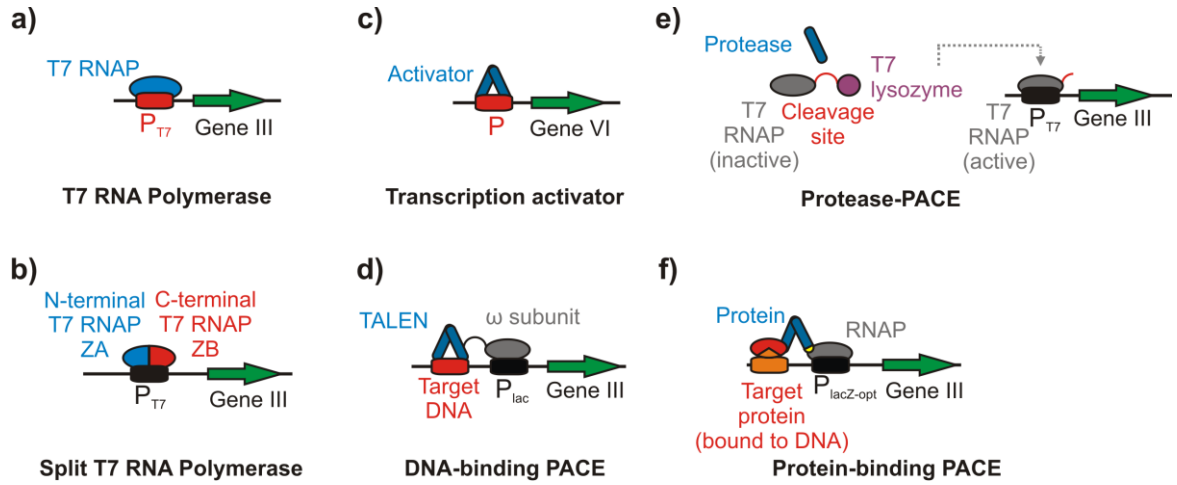


Figure 3. Directed evolution of different classes of proteins based on conditional M13 phage replication. a) An evolving T7 RNA polymerase upregulates gene III expression in an activity-dependent manner [16]. **b**) An evolving N-terminal T7 RNA polymerase fused to a leucine zipper ZA assembles with a C-terminal T7 RNA polymerase variant fused to leucine zipper ZB leading to gene III expression in an activity-dependent manner [38]. **c**) An evolving transcription activator (e.g. λ cI) upregulates gene VI expression downstream of a specific promoter (e.g. λ P_{RM}) [27]. **d**) DNA-binding PACE enables the evolution of transcription activator-like effector nucleases (TALENs) [25]. The evolving DNA-binding protein is linked to the ω subunit of bacterial RNA polymerase III and binding to a target DNA sequence upstream of a minimal lac promoter enables gene III expression in an activity-dependent manner. **e**) Protease-PACE enables the evolution of proteases against desired cleavage sites [36]. The T7 polymerase is inhibited when bound to T7 lysozyme as it inhibits transcription initiation and the transition from initiation to elongation [44]. Proteolysis of the target cleavage site by an evolving protease activates the T7 RNA polymerase leading to gene III expression in an activity-dependent manner. **f**) Protein-binding PACE allows the evolution of protein-protein interactions [37]. The target protein is bound to the DNA upstream the promoter $P_{lacZ\text{-}opt}$ via a fused DNA-binding domain (orange) and the RNA polymerase omega subunit (RpoZ; yellow) is fused to the evolving protein. The binding of the evolving protein to the target protein enables the transcription of gene III from the $P_{lacZ\text{-}opt}$ promoter. The evolving protein is highlighted in blue and the target sequence is depicted in red for each individual example.