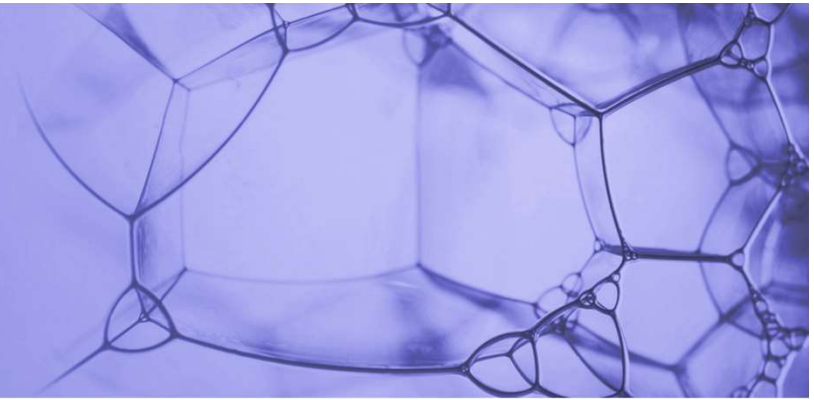


**LOSCHMIDT  
LABORATORIES**



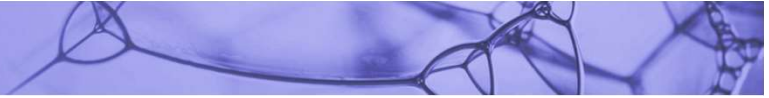
# **Synthetic protein biology II.**

**Directed evolution, combined approaches & screening  
and selection technologies**

---

Modifying protein structure and function for biotechnology,  
biomedicine and basic research

Dr. Martin Marek  
Loschmidt Laboratories  
Faculty of Science, MUNI  
Kamenice 5, bld. A13, room 332  
[martin.marek@recetox.muni.cz](mailto:martin.marek@recetox.muni.cz)



## Short recapitulation of previous lesson

---

- What is protein engineering? Concept, strategies etc...
- Why protein engineering is important for synthetic biology?
- How rational protein design works? Workflow, pros and cons
- Computational *de novo* protein design...?
- Computational design of new enzyme...?
- What to do after the design is finished...?

# What will we talk about

---

- Introduction to protein engineering and design
  - definition, goals and applications
- Rational protein design (knowledge-based strategies)
  - concepts, methodology, limitations, success stories
- Directed evolution (lab-based brute force engineering)
  - strategies, methodology, disadvantages, success stories
- Integrative (combined) approaches
  - the best of both approaches, beneficial synergy, examples
- Selection and screening technologies
  - classical versus emerging technologies, unmet challenges



# The two major strategies of protein engineering

## RATIONAL DESIGN

1. Computer aided design



2. Site-directed mutagenesis



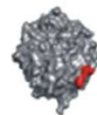
Individual mutated gene

3. Transformation

4. Protein expression

5. Protein purification

6. *not applied*



Constructed mutant enzyme

## DIRECTED EVOLUTION

1. *not applied*

2. Random mutagenesis



Library of mutated genes  
( >10,000 clones )

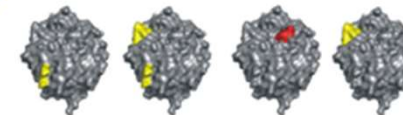
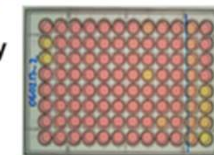
3. Transformation

4. Protein expression

5. *not applied*

6. Screening and selection

- stability
- selectivity
- affinity
- activity



Selected mutant enzymes

IMPROVED  
ENZYME

7. Biochemical testing

# Directed evolution

---

Concepts  
Methods  
Applications

## Expanding the synthetic protein universe by guided evolutionary concepts

---

- The genetic information of a cell is maintained by the sequence composition of the DNA
- The changes in the nucleotide content may potentially alter its transcriptional and translational events thus influencing the properties of the newly synthesized proteins
- The nature's alterations can be helpful in the evolution of proteins with novel or improved functionalities
- Unravelling the principles of such molecular evolutionary process is resourceful to implement it for the benefit of the mankind through laboratory techniques
- The laboratory process of synthesizing new proteins in a constructive way through evolutionary guided principles is called “**directed evolution**”



## What it is directed evolution and why did it win the Nobel prize

---

The Nobel prize  
in Chemistry  
(2018)



Prof. Frances Arnold herself summed it up nicely:

**Directed evolution**

*In directed evolution we provide a new niche in the laboratory, so to speak, and encourage evolution of enzymes to catalyse commercially useful reactions.*

*Enzymes can speed up reactions dramatically while carrying out their work in water at room temperature. They are also really good at making one specific bond – without messing around with other functional groups – and are often enantioselective too.*

*So it's easy to understand why chemists like using enzymes to catalyse reactions. However, many bonds chemists are interested in aren't made by any natural enzyme. This is simply because organisms have never needed to evolve the ability to make, for example, carbon-silicon bonds.*

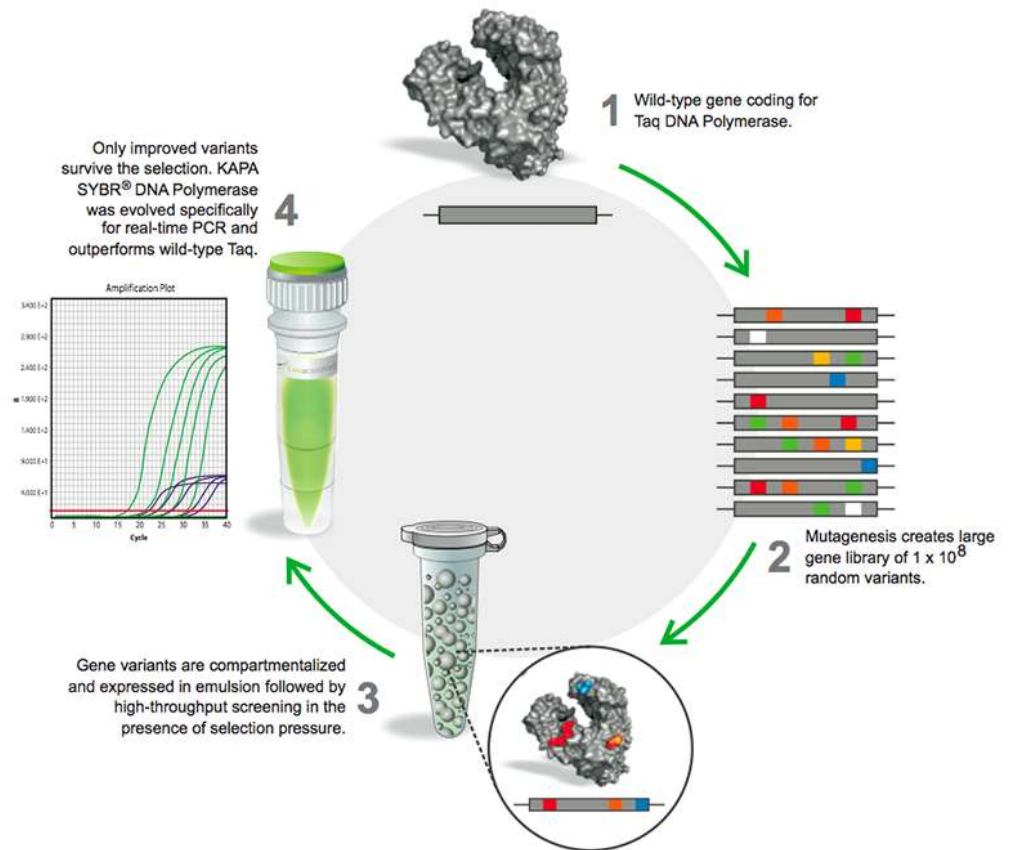
BIOCATALYSIS

## An enzymatic route to carbon-silicon bonds

Protein demonstrated the first example of 'natural' organosilicon chemistry

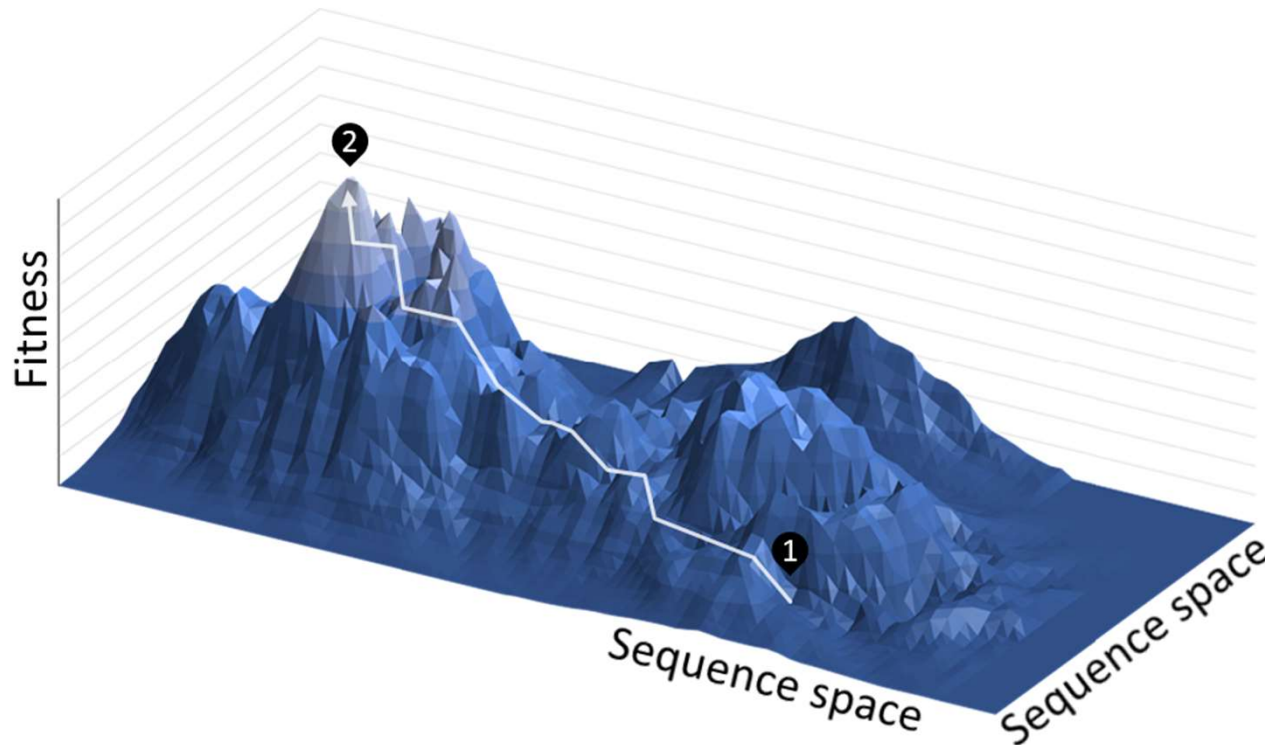
## Directed evolution workflow

- Directed evolution (DE) is a method of protein engineering that simulates natural selection in the lab
- The process starts with a gene coding for a “wild-type”, or unmodified, protein of interest
- Random variation is introduced into the gene through a process of mutagenesis, **generating a library** of millions of genes each coding for a unique protein variant
- A functional **selection** pressure is then applied to the library, and only the genes that coded for the highest performing proteins “survive”
- This process of **random mutation and selection is repeated** until the desired enzyme function evolves





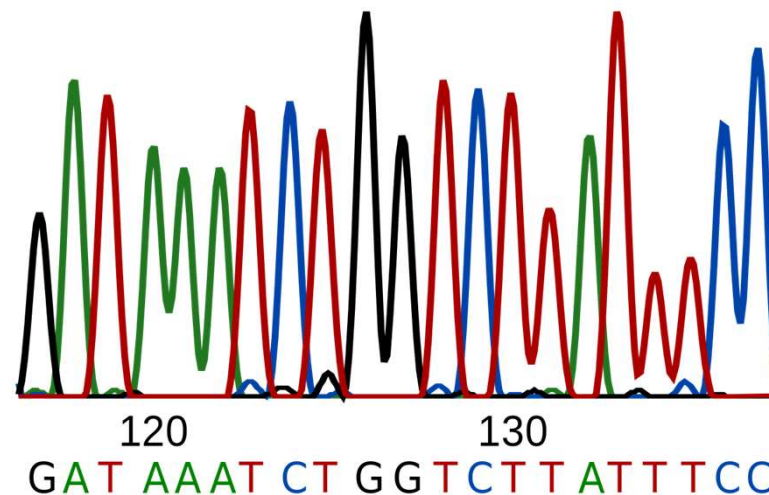
## Directed evolution is a mimic of the natural evolution cycle in a laboratory setting

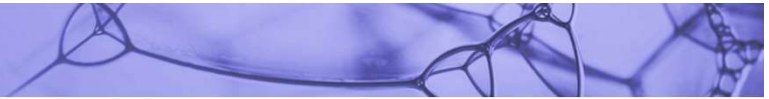


Directed evolution is analogous to climbing a hill on a „**fitness landscape**“ where elevation represents the desired property. Each round of selection samples mutants on all sides of the starting template (1) and selects the mutant with the highest elevation, thereby climbing the hill. This is repeated until a local peak is reached (2).

## Advantages of directed evolution

- Rational design of a protein relies on an in-depth knowledge of the protein structure, as well as its mode of action (e.g. catalytic mechanism)
- Specific changes are then made by site-directed mutagenesis in an attempt to change the function of the protein
- A drawback of this is that even when the structure and mechanism of action of the protein are well known, the change due to mutation is still difficult to predict
- Therefore, an advantage of directed evolution is that there is **no need to understand the mechanism of desired activity** or how mutations would affect it





# Methods for gene diversification and library generation

---

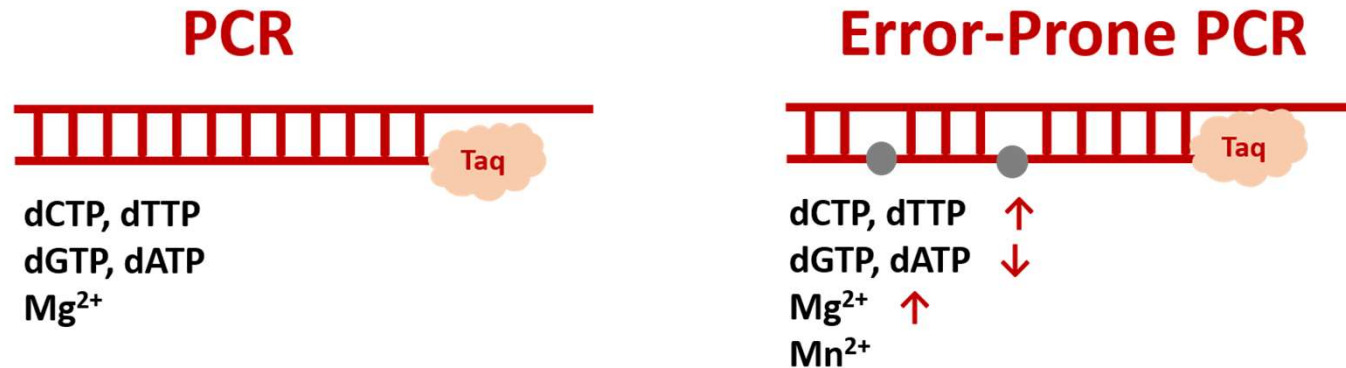
## Asexual methods

- **Random mutagenesis**
  - Error-prone PCR
  - Chemical mutagenesis
  - Mutator strains
  - Insertion and deletion mutagenesis
  - Rolling circle error-prone PCR
- **Focused mutagenesis**
- **Site saturation mutagenesis (SSM)**
  - Casette mutagenesis
  - Sequence saturation mutagenesis
  - Megaprimered and ligase-free

## Sexual methods

- **Homologous recombination**
- **In vitro homologous recombination**
  - DNA shuffling
  - Random priming in vitro recombination
  - Random chimeragenesis on transient templates (RACHITT)
  - Synthetic shuffling
- **In vivo homologous recombination**
  - Cloning preformed in yeast (CLERY)
  - Phage assisted continuous evolution (PACE)
  - Mutator strains
  - Insertion and deletion mutagenesis
  - Rolling circle error-prone PCR
- **In vitro non-homologous recombination**
  - Exon shuffling
  - Golden Gate Shuffling recombination
  - Sequence homology-independent protein recombination (SHIPREC)
  - Megaprimered and ligase-free

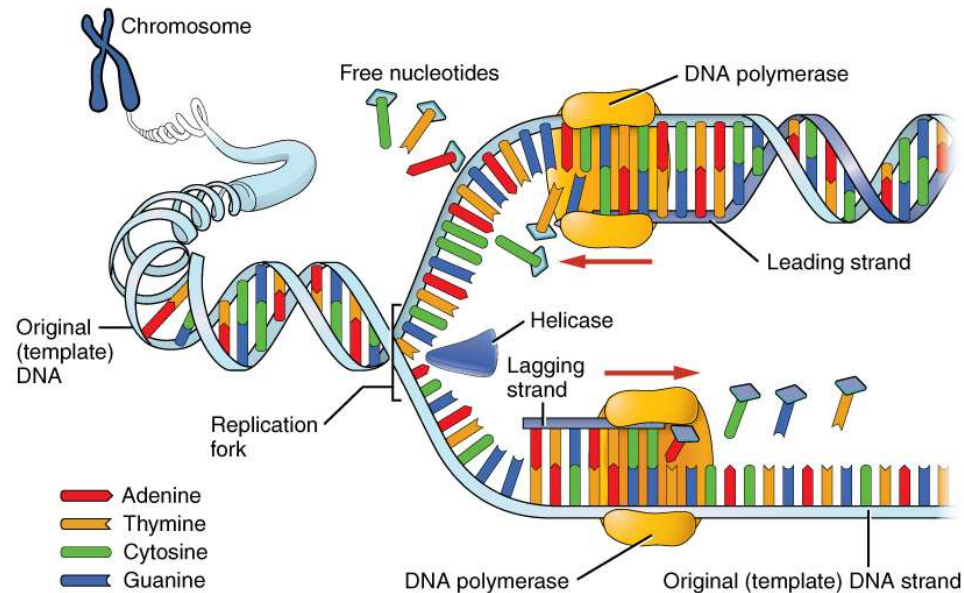
## Error-prone PCR (epPCR)



- The error rate of *Taq* DNA polymerase is 0.001-0.002 % per nucleotide per replication cycle under standard conditions which is sufficient to create mutant libraries of large genes but not for small genes
- Error-prone PCR (epPCR) takes advantage of the inherently low fidelity of *Taq* DNA Polymerase, which may be further decreased by the addition of  $Mn^{2+}$ , increasing the  $Mg^{2+}$  concentration, and using unequal dNTP concentrations.
- The rate of mutagenesis achieved by error-prone PCR is in the range of 0.6-2.0 %

## Mutator strains

- Mutator strains of *E. coli* are deficient in one or more of DNA repair genes, leading to single base **substitutions** at a rate of approximately 1 mutation per 1000 base pairs
- Generation of mutant libraries
- Process is simple

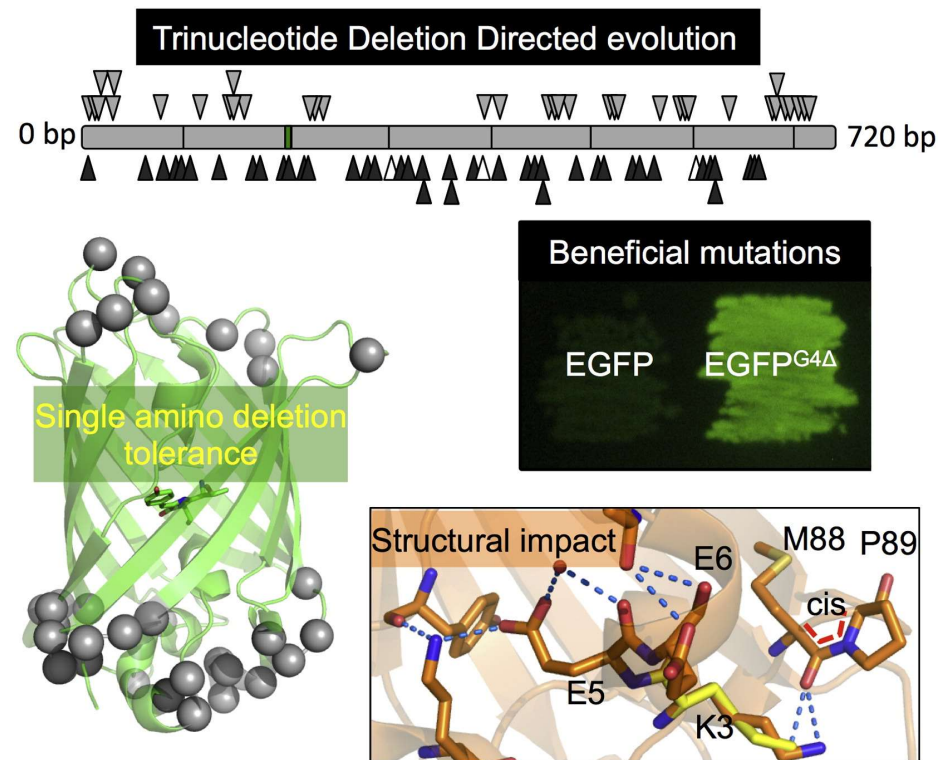


# Insertion and deletion (InDel) mutagenesis

- Gain or lost of one or more nucleotides produces frameshift mutations (triplet reading frame)
- Triplet InDel mutagenesis may trigger protein backbone changes essential for evolvability
- Insertion and deletion mutations can enhance proteins through **structural rearrangements** not possible by substitution mutations alone



Using directed evolution, green fluorescent protein (GFP) was observed to tolerate residue deletions, particularly within short and long loops, helical elements, and at the termini of strands. A variant with G4 removed from a helix (EGFPG4 $\Delta$ ) conferred significantly higher cellular fluorescence.

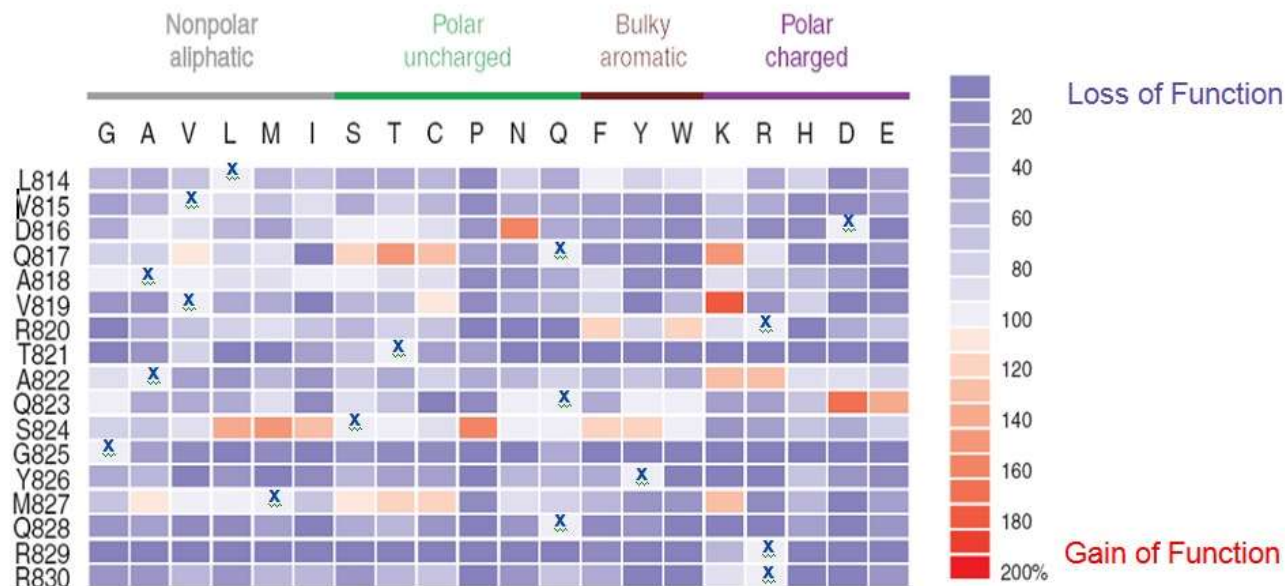


Arpino et al., Structure 22: 889-898 (2014)

## Site saturation mutagenesis (SSM)

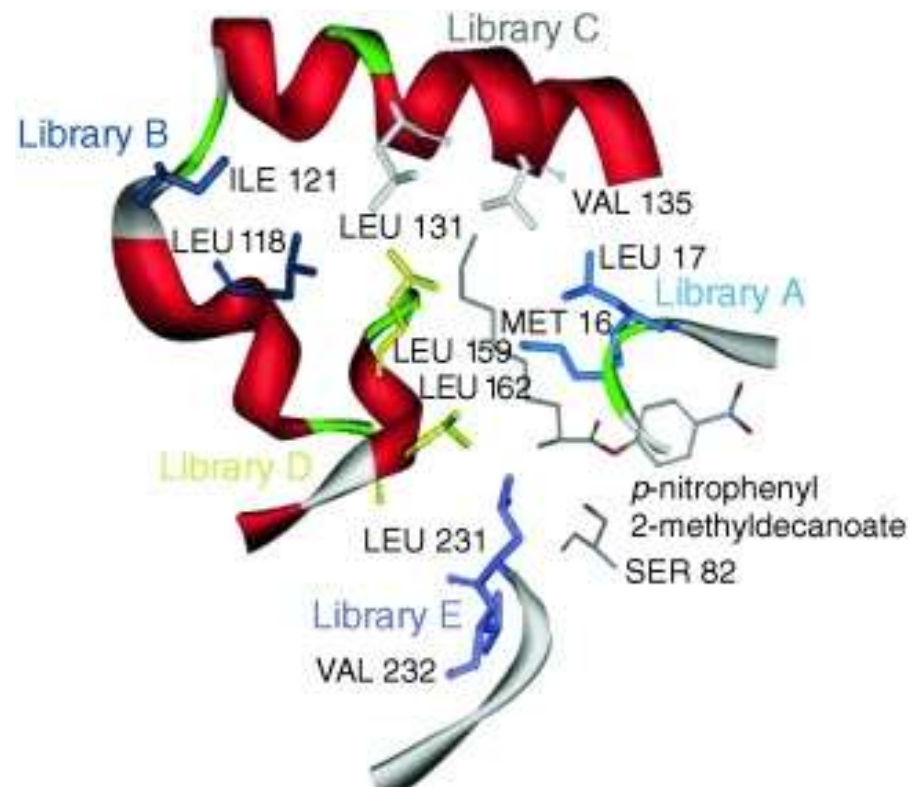


- Site saturation mutagenesis is used to substitute targeted residues to any other naturally occurring amino acid
- The core of a SSM experiment lies in the codon degeneracy or randomness. A completely randomized codon (NNN, where N=A, C, G or T) results in a library size of 64 different sequences encoding all 20 amino acids and 3 stop codons
- When an experiment targets multiple codons, the library size can be considerably higher, making it difficult to perform a complete screening (e.g. targeting three NNN codons has 262,144 unique codon configurations)



## Combinatorial active-site saturation test (CAST)

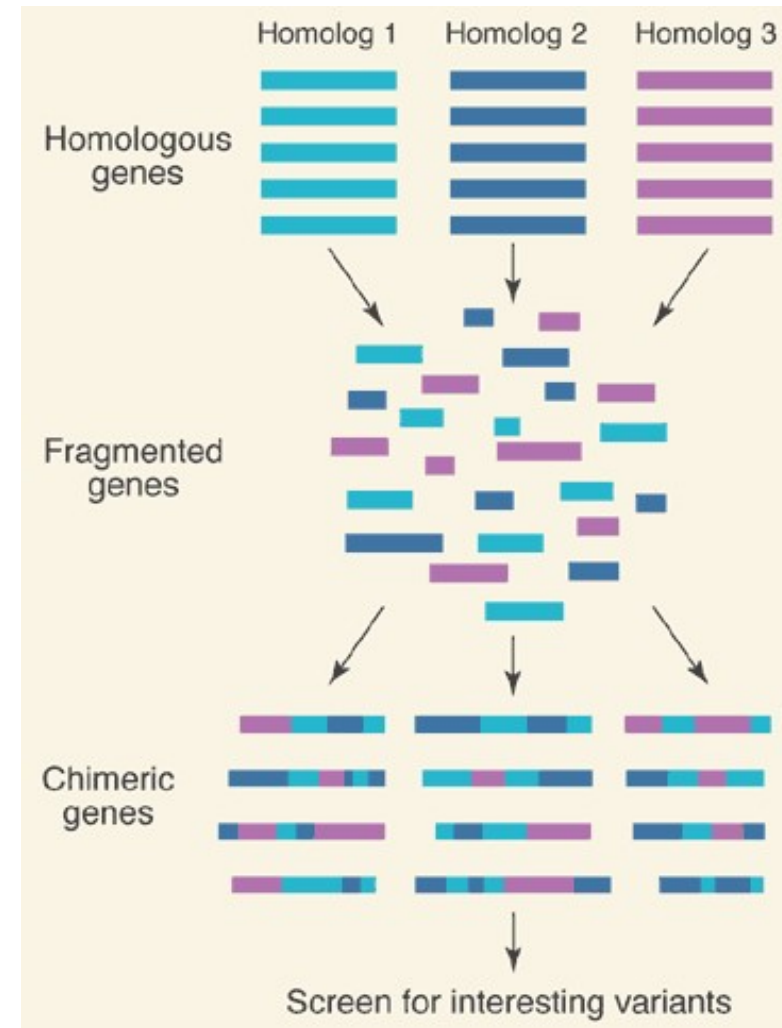
- The Combinatorial Active-site Saturation Test (**CAST**) was developed to increase the enantioselectivity and/or the substrate specificity of enzymes
- The basis of the method is the generation of small libraries of mutant enzymes that are easy to screen for activity
- The mutants are produced by simultaneous randomization of sets of two or three spatially close amino acids, whose side chains form part of the substrate-binding pocket





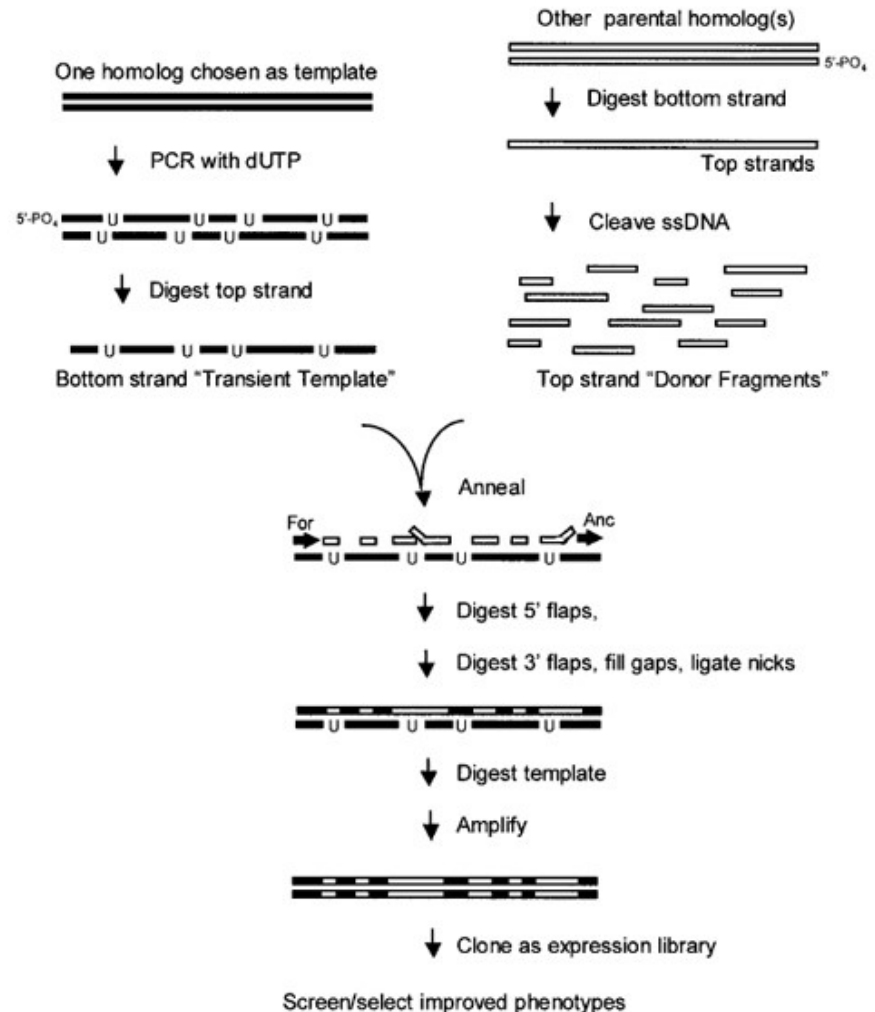
## DNA shuffling

- DNA shuffling is a method for *in vitro* recombination of homologous genes
- The genes to be recombined are randomly fragmented by DNaseI, and fragments of the desired size are purified from an agarose gel
- These fragments are then reassembled using cycles of denaturation, annealing, and extension by a polymerase
- Recombination occurs when fragments from different parental templates anneal at a region of high sequence identity
- Following this reassembly reaction, PCR amplification with primers is used to generate full-length chimeras suitable for cloning into an expression vector
- Moving from DNA shuffling to whole genome shuffling is known as **GENOME SHUFFLING**



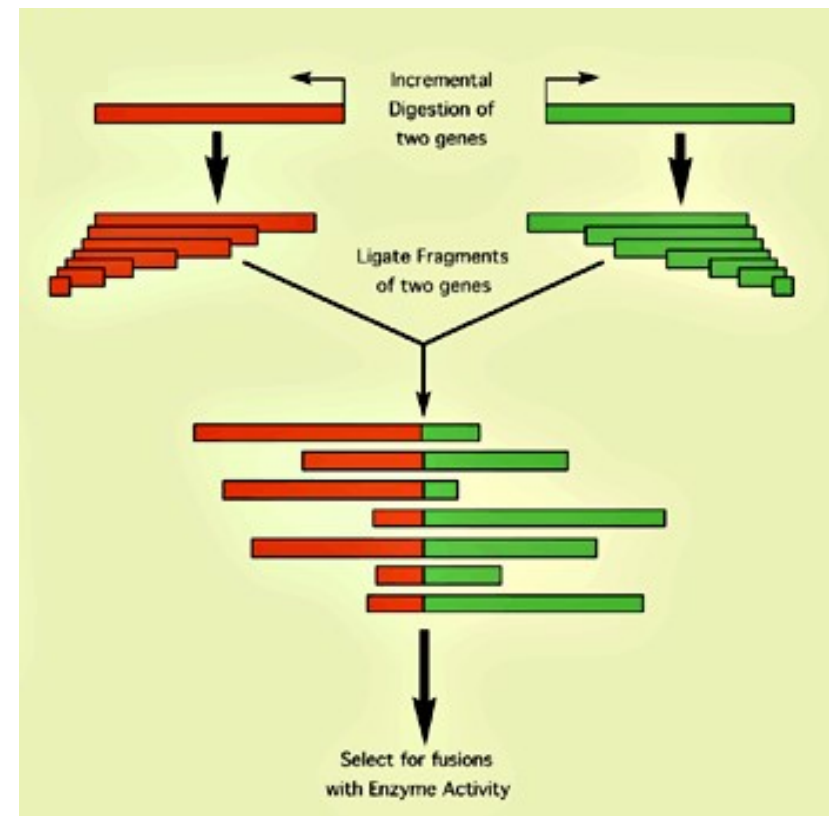
# Random chimeragenesis on transient templates (RACHITT)

- RACHITT creates chimeric genes by aligning parental gene “donor” fragments on a full-length DNA template
- The heteroduplexed top strand fragments are stabilized on the template by a single, long annealing step
- Fragments containing unannealed 5' or 3'-termini are incorporated after flap trimming using the endo and exonucleolytic activities of *Taq* DNA polymerase and *Pfu* polymerase, respectively
- After gap filling and ligation, the template, which was synthesized with uracils in place of thymidine, is rendered non-amplifiable by uracil-DNA glycosylase (UDG) treatment



## Incremental truncation for the creation of hybrid (ITCHY)

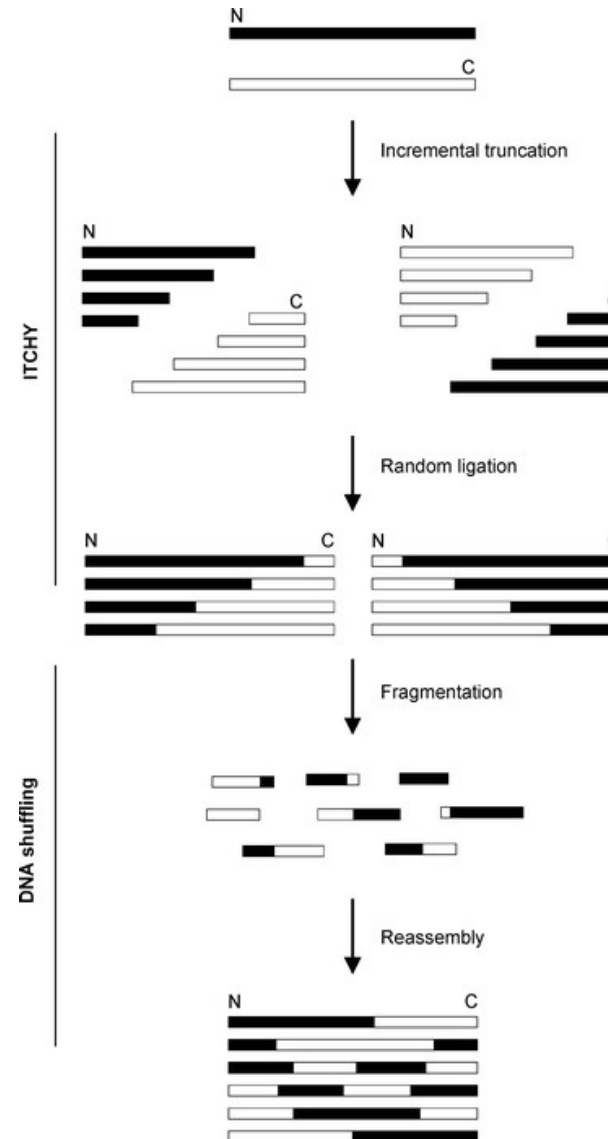
- ITCHY is a directed evolution technique for randomly recombining two genes
- The chief advantage of ITCHY is that there is no requirement for the two genes to share any sequence similarity
- This distinguishes ITCHY from directed evolution methods that are based on homologous recombination, such as DNA shuffling
- In ITCHY, *Escherichia coli* exonuclease III is used to incrementally truncate one of the parental genes from its 3' end and the other from its 5' end
- Ligation of the randomly truncated gene fragments yields a combinatorial library of chimeras



# Combining ITCHY and DNA shuffling

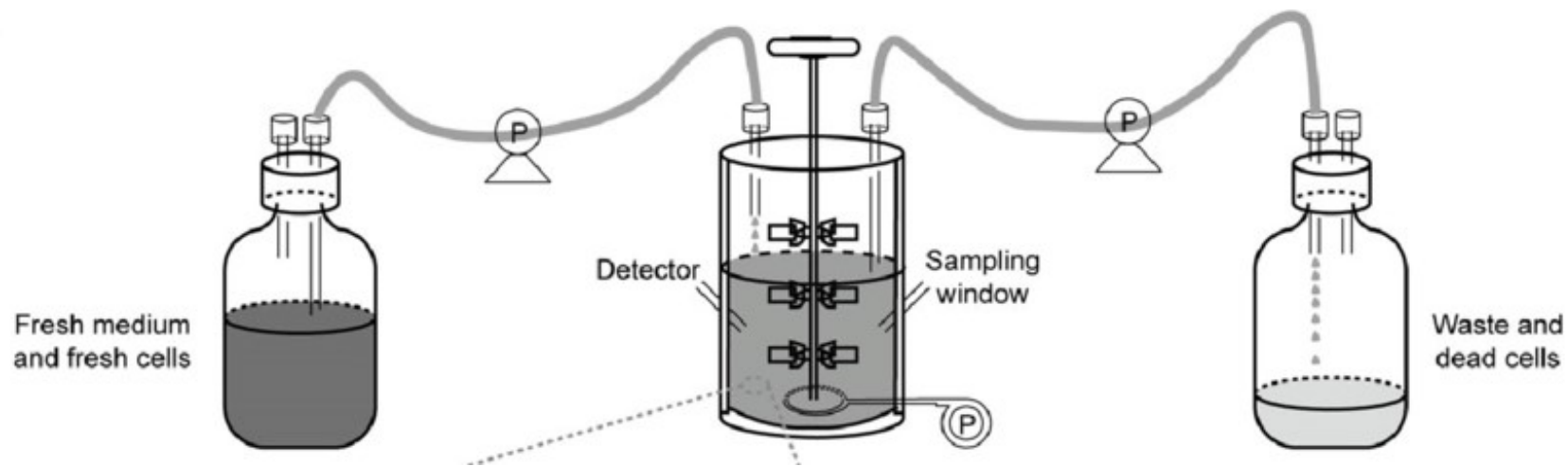
## 1. ITCHY

## 2. DNA shuffling



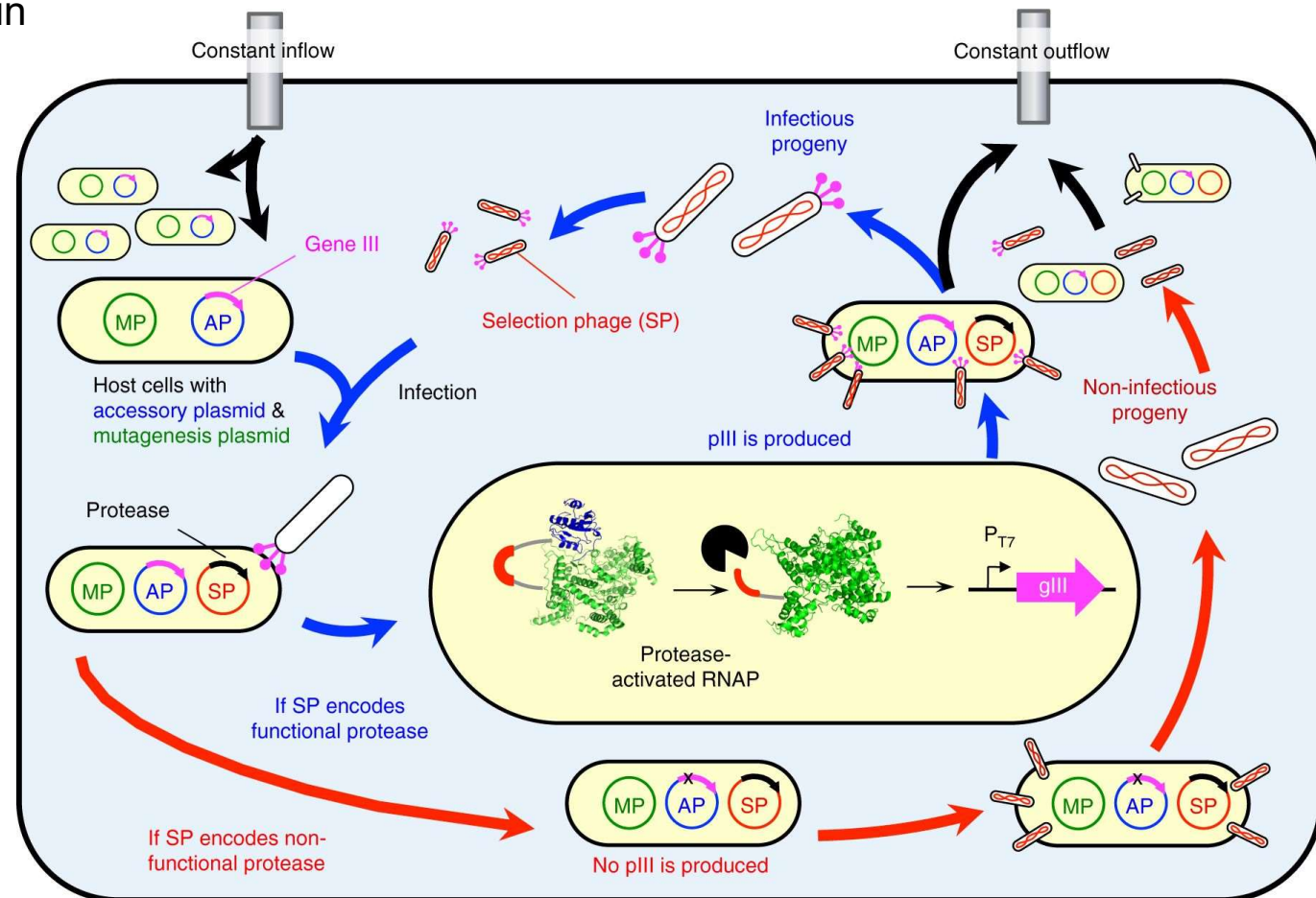
## Continuous directed evolution strategies

- **Continuous evolution systems couple mutagenesis and selection**
- The optimal continuous evolution system is performed in a turbidostat which consists of a medium supply container, fermentation tank, and a waste collection container
- Selection pressures can be imposed in the fermentation tank to enable growth-based enrichment of improved mutants

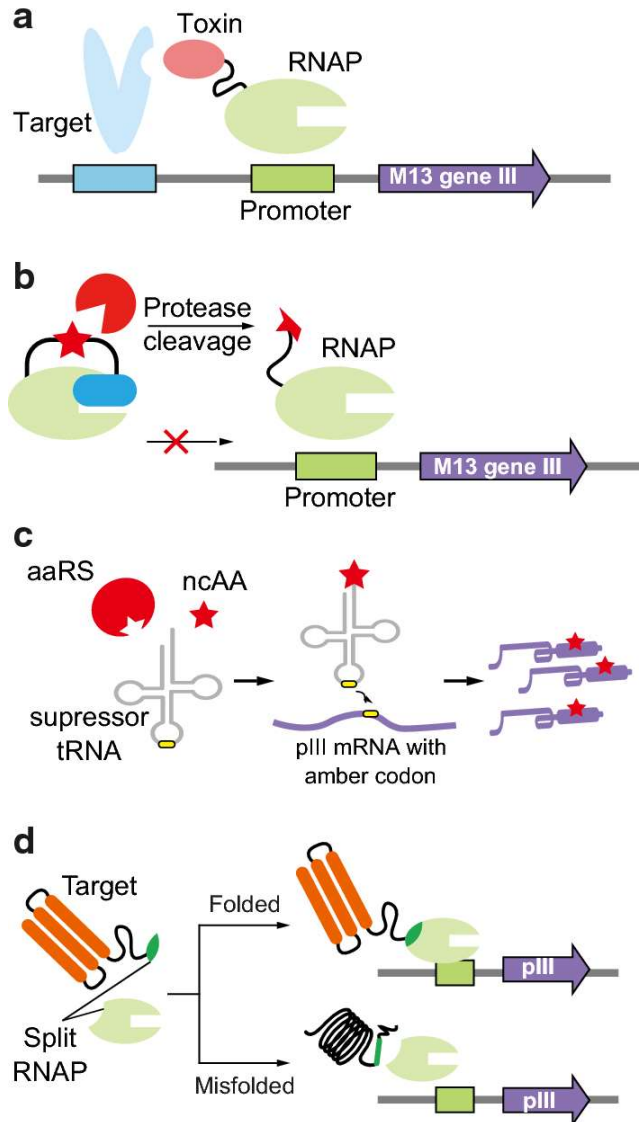


# Phage-assisted continuous evolution (PACE)

- The **PACE** is a form of directed evolution, a laboratory process that uses the phage virus bacterial infection cycle to generate multiple rounds of DNA sequence changes and selection for DNA changes in a target gene that result in a desired structure or activity in the encoded protein



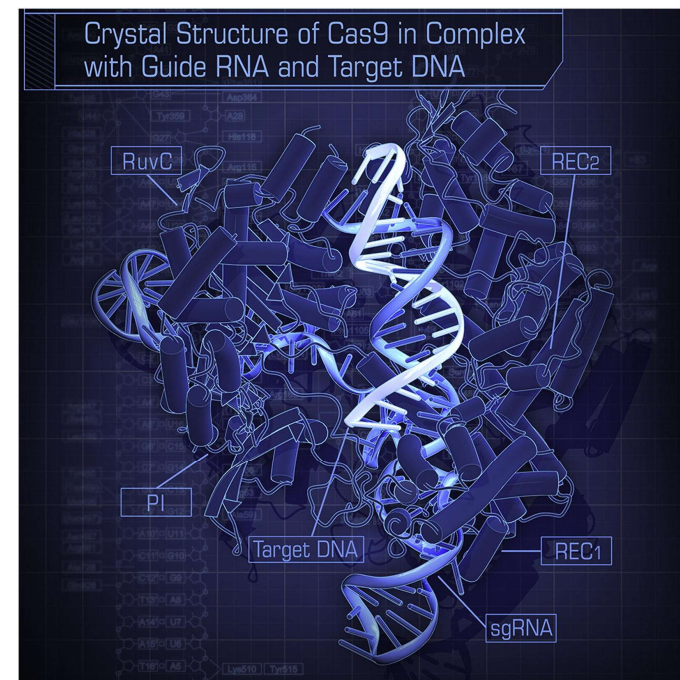
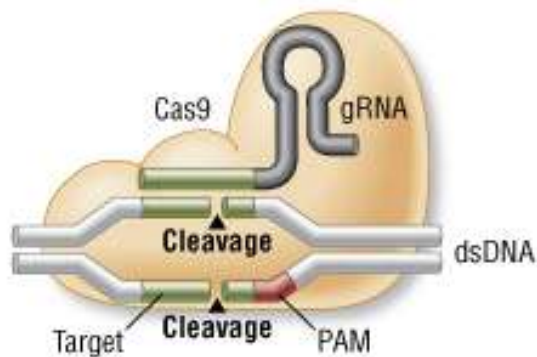
# Evolution of protein functions with PACE: examples



- Bt toxin evolution to targeting new receptor **(a)**.
- TEV protease evolution to target new cleavage site **(b)**.
- aaRS evolution to get specificity with ncAA and suppressor tRNA **(c)**.
- Eukaryotic protein evolution to improve solubility in *E. coli* **(d)**.
- Cas9 evolution to recognize new PAM **(e)**

## The CRISPR/Cas9 system

- The **Cas9** (CRISPR associated protein 9) is a protein which plays a vital role in the immunological defense of bacteria against DNA viruses, and which is used in genetic engineering. Its main function is to cut DNA and therefore it can alter a cell's genome
- Structurally, Cas9 is an **RNA-guided DNA endonuclease enzyme** associated with CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) adaptive immunity system in *Streptococcus pyogenes*
- Cas9 performs this by unwinding foreign DNA and checking for sites complementary to the 20 bp spacer region of the guide RNA
- If the DNA substrate is complementary to the guide RNA, Cas9 cleaves the invading DNA



Crystal structure of *S. pyogenes* Cas9 in complex with sgRNA and its target DNA at 2.5 Å resolution, Nishimazu *et al.*, Cell 156: 935–49 (2014)

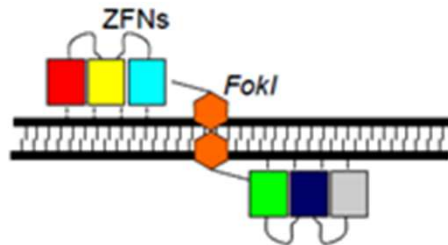




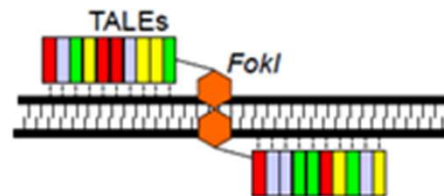


## Advantages of CRISPR/Cas9-mediated mutagenesis

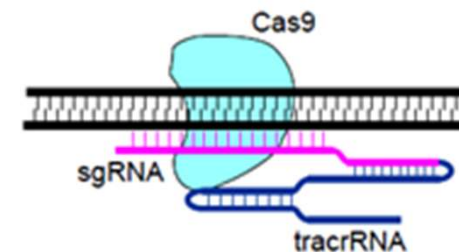
- The CRISPR/Cas9 system requires only the redesign of the crRNA to change target specificity
- This contrasts with other genome editing tools, including zinc finger and TALENs, where redesign of the protein-DNA interface is required
- Furthermore, CRISPR/Cas9 enables rapid genome-wide interrogation of gene function by generating large gRNA libraries for genomic screening



Zinc-finger nucleases



TALENs



Cas9



# Multiplex Automated Genome Engineering (MAGE)

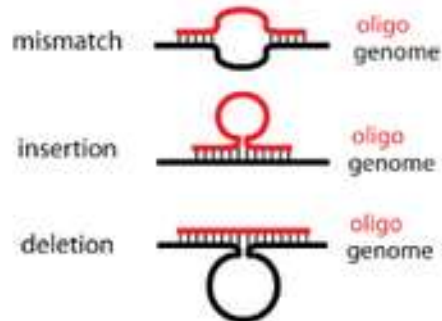


- MAGE is a technique capable of editing the genome by making small changes in existing genomic sequences
- These properties make MAGE a highly useful tool for synthetic biology, allowing researchers to easily modify the bacterial genome and generate diversity within a population

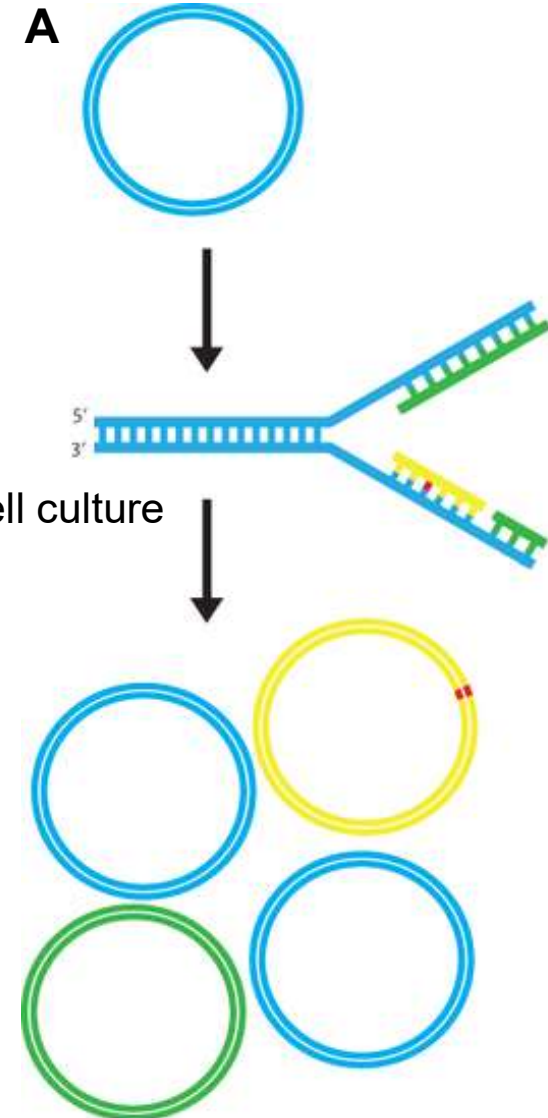
## How MAGE works

- The MAGE is inserting single-stranded oligos that contain the desired mutations into the cell, and more than one gene can be targeted at a time simply by using multiple oligos
- Mediated by  $\lambda$ -Red ssDNA-binding protein  $\beta$ , the oligos are incorporated into the lagging strand of the replication fork during DNA replication, creating a new allele
- The efficiency of oligo incorporation depends on several factors ( $\Delta mutS$ ), but the frequency of the allele can be increased by performing multiple MEGA rounds on the same cell culture
- The process can be automated allowing to run many rounds of MAGE in succession

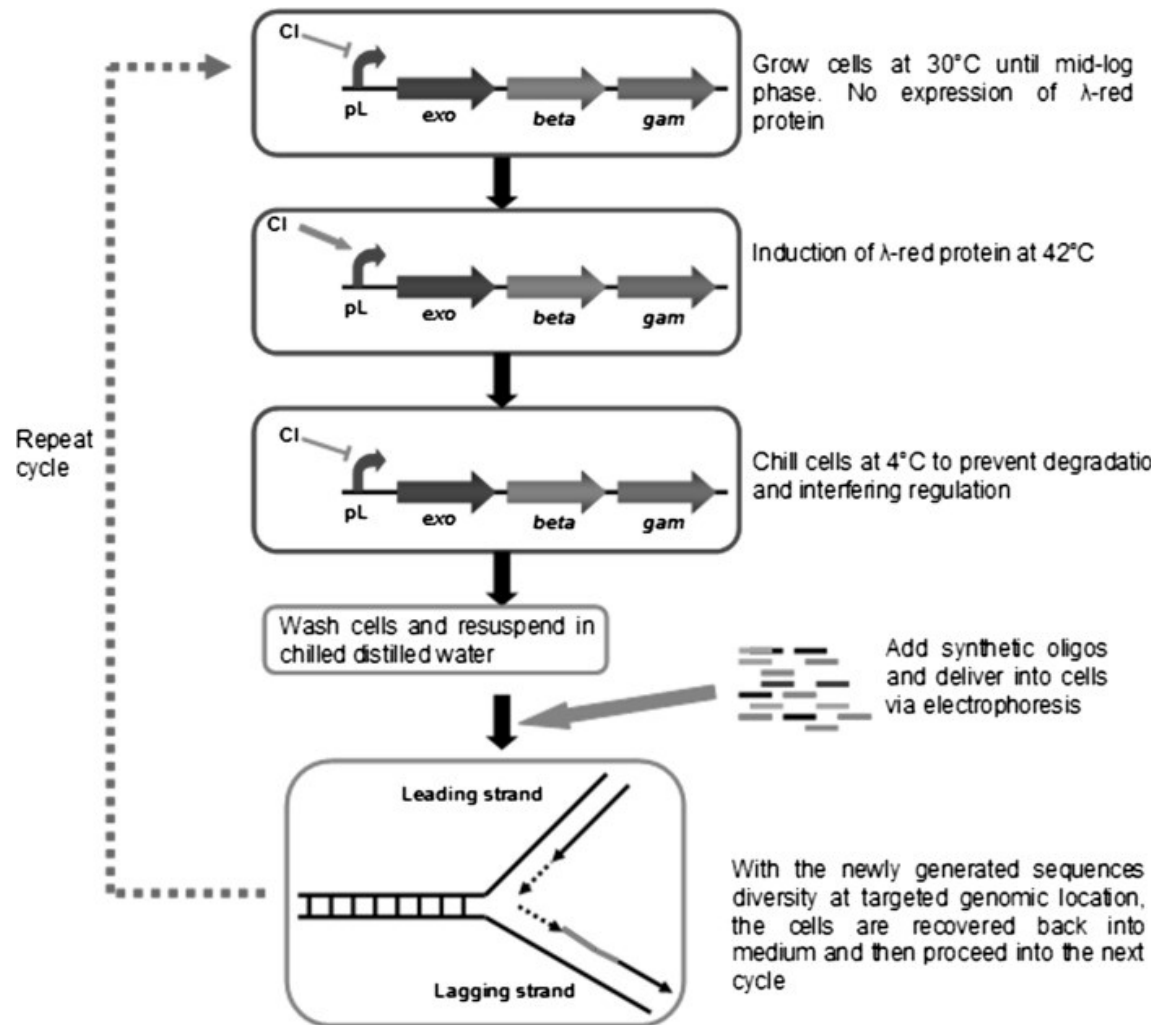
**B**



**Figure 1: Schematic of incorporation of MAGE oligo** (adapted from Wang et al).  
**(A)** The MAGE oligo (in yellow with a red mutation) is incorporated into the lagging strand, similar to an Okazaki fragment, when the genome is replicated, producing a mixture of cells with and without the mutations.  
**(B)** The mutations MAGE is capable of producing are mismatches, insertions, and deletions.



## Basic design and cycling of MAGE

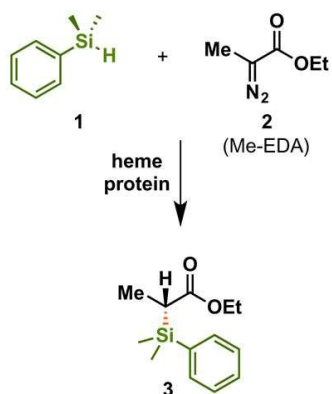


- The MAGE starts with growing the cells at 30 °C until cell density reaches the mid-log phase and lambda ( $\lambda$ ) red proteins are expressed under the control of pL promoter which is regulated by temperature sensitive Cl
- Then cells are moved to 42 °C for 15 min for the heat shock induction of  $\lambda$ -red proteins.
- Cells are moved to 4 °C to repress the  $\lambda$ -red and prevent degradation. Cells were subsequently washed and resuspended in chilled distilled water
- Single stranded oligos were introduced into cells via electroporation and incorporated into the lagging strand of the replication fork during DNA replication
- The cells were kept at 30 °C for 2–3 h for recovery of generated sequences diversity before proceeding into the next MAGE cycle

# DE of cytochrome c for carbon–silicon bond formation

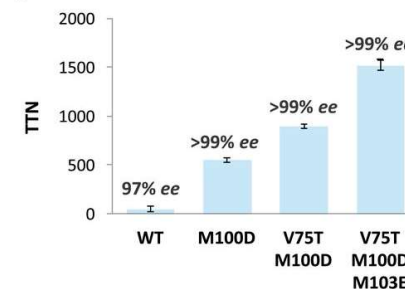
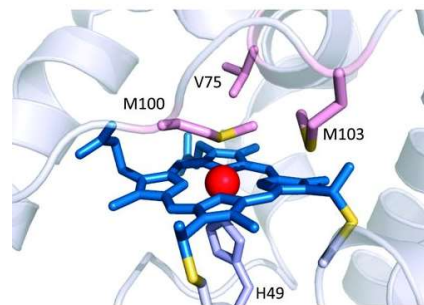
## An example of directed evolution I.

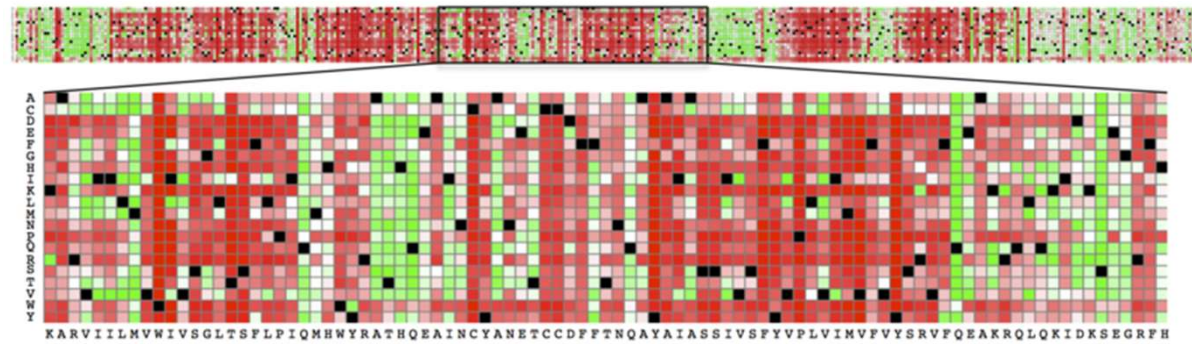
- Enzymes that catalyze carbon–silicon bond formation are unknown in nature, despite the natural abundance of both elements. Such enzymes would expand the catalytic repertoire of biology, enabling living systems to access chemical space previously only open to synthetic chemistry
- Discovery that heme proteins catalyze the formation of organosilicon compounds under physiological conditions via carbene insertion into silicon–hydrogen bonds. The reaction proceeds both in vitro and in vivo, accommodating a broad range of substrates with high chemo- and enantioselectivity
- Using directed evolution the catalytic function of cytochrome c from *Rhodothermus marinus* achieved more than 15-fold higher turnover than state-of-the-art synthetic catalysts. This carbon–silicon bond-forming biocatalyst offers an environmentally friendly and highly efficient route to producing enantiopure organosilicon molecules



TTN	catalyst	% ee
0	None	-
0	Hemin	0
0	Hemin + BSA	0
<5	Horse heart cyt c	<5
6	Bovine heart cyt c	6
<5	<i>S. cerevisiae</i> cyt c	<5
97	<i>R. marinus</i> cyt c	97
16	<i>H. thermophilus</i> cyt c	16
<5	<i>R. globiformis</i> cyt c	<5

Various P450s and myoglobin also catalyzed the formation of carbon–silicon bonds, but the reactions were not enantioselective (see Supplementary Materials).





# Screening and selection technologies

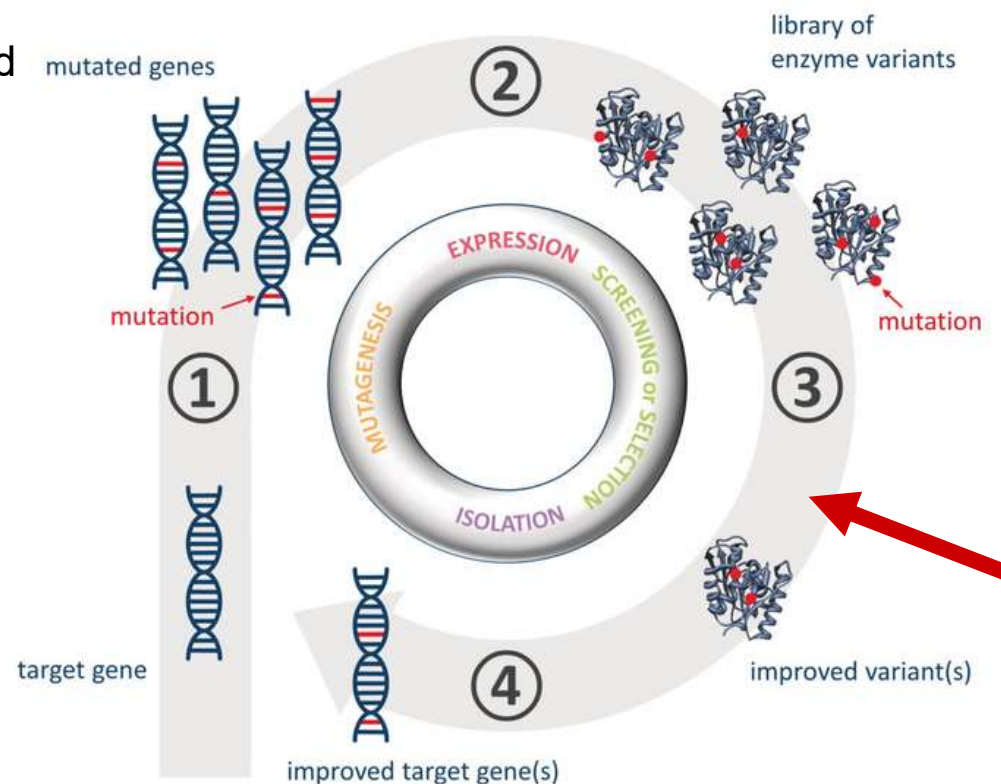
---

Concepts  
Methods  
Applications



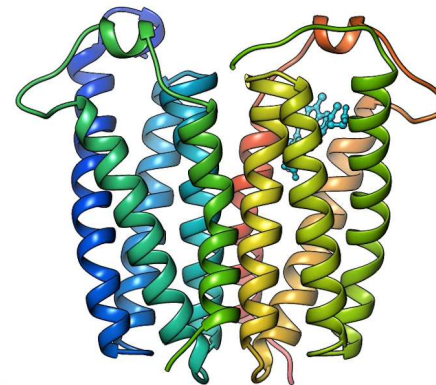
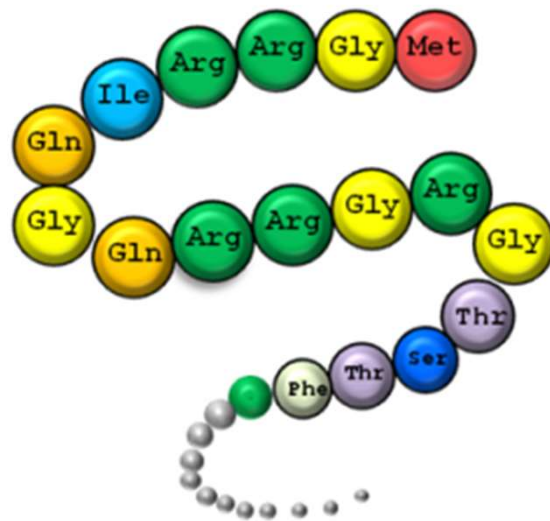
## High-throughput screening is important part of DE cycle

1. Random mutations are introduced into a target gene (1), e.g. error-prone PCR
2. The mutated genes are transferred into a suitable host organism and expressed, thereby resulting in a large library of protein or enzyme variants
3. Improved variants are then identified either by **high-throughput screening** or by **selection** for a desired property
4. The gene(s) encoding such improved variant(s) are isolated and used for another cycle of directed evolution



## The size of sequence space

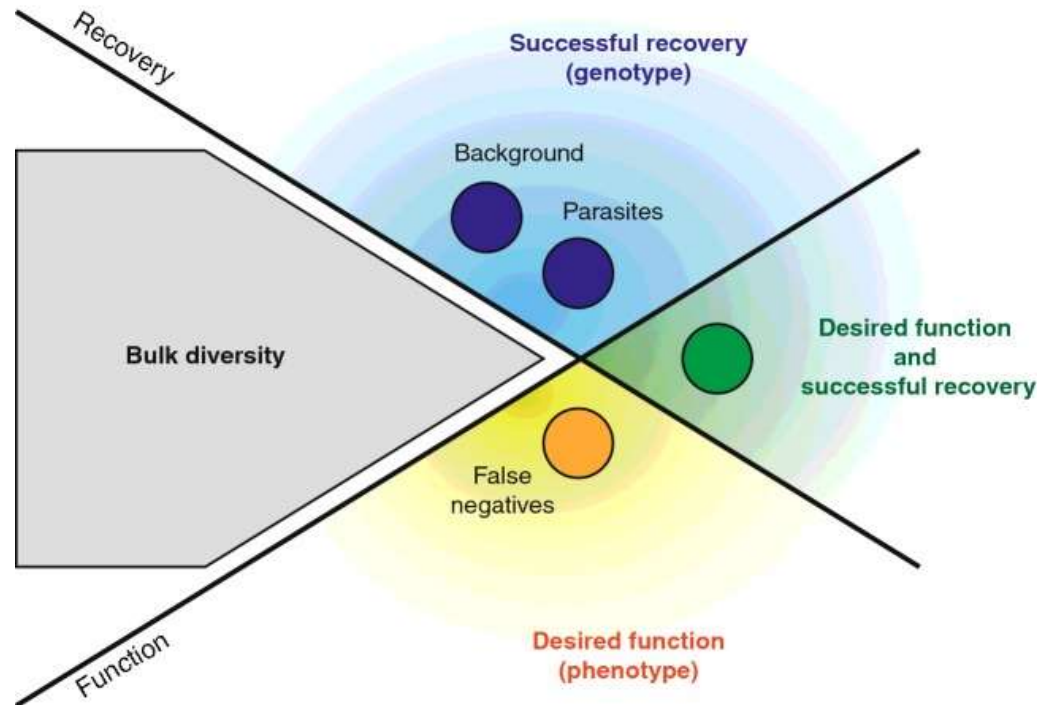
- Important concept when considering a protein's amino acid sequence is that of (its) **sequence space**, *i.e.* the number of variations of that sequence that can possibly exist
- Straightforwardly, for a protein that contains just the 20 main natural amino acids, a sequence length of  $N$  residues has a total number of possible sequences of  $20^N$ . For  $N = 100$  (a rather small protein) the number  $20^{100}$  is already far greater than the number of atoms in the known universe



Amino Acids

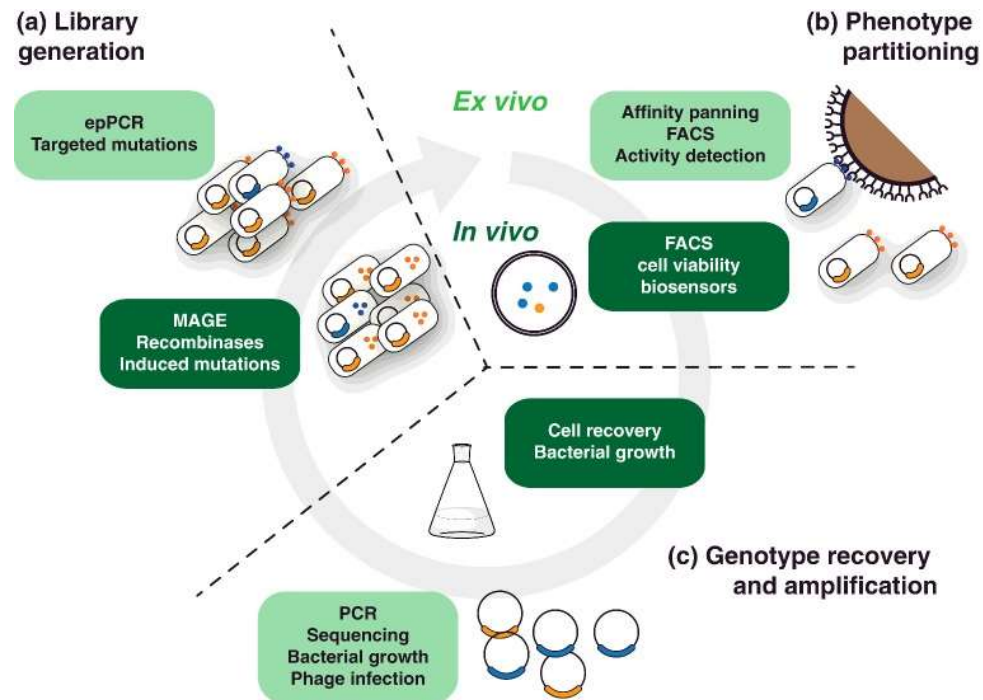
Ala: Alanine	Gln: Glutamine	Leu: Leucine	Ser: Serine
Arg: Arginine	Glu: Glutamic acid	Lys: Lysine	Thr: Threonine
Asn: Asparagine	Gly: Glycine	Met: Methionine	Trp: Tryptophane
Asp: Aspartic acid	His: Histidine	Phe: Phenylalanine	Tyr: Tyrosine
Cys: Cysteine	Ile: Isoleucine	Pro: Proline	Val: Valine

## Principles of selection in directed evolution



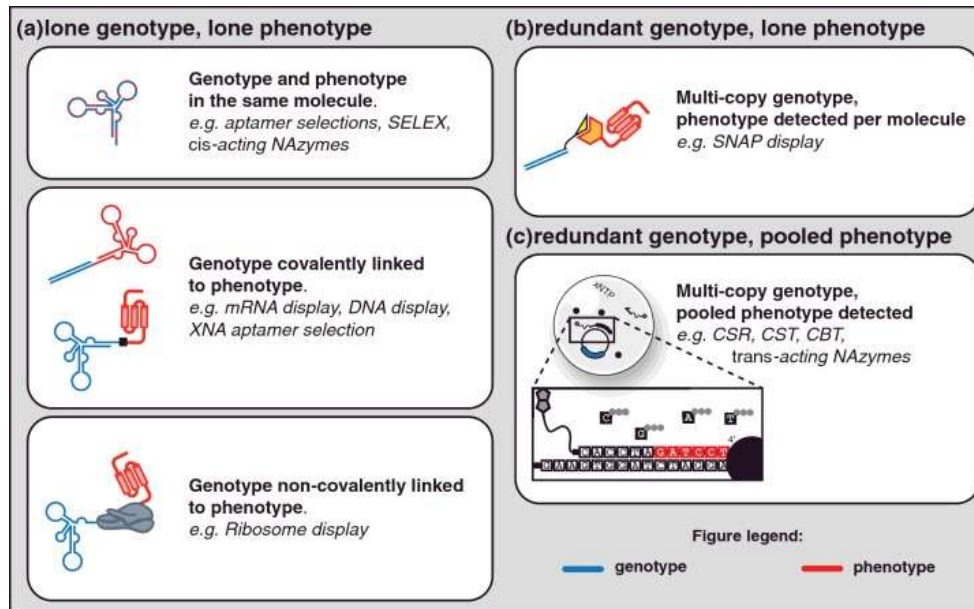
- The goal of all selection (and screening) platforms is to partition a potentially large population (shown in grey as the bulk diversity) by function (phenotype) ensuring the recovery of the genetic information that accounts for that phenotype.
- **Strong phenotype–genotype linkages** allow efficient isolation of mutants with the desired function (green).
- Breakdown of that linkage results in **false negatives** (variants that have the desired function but that are not efficiently recovered–yellow) and **false positives** (variants that are recovered independently of the desired function–blue), which are integral aspects of all selection strategies.

## In vivo and ex vivo directed evolution



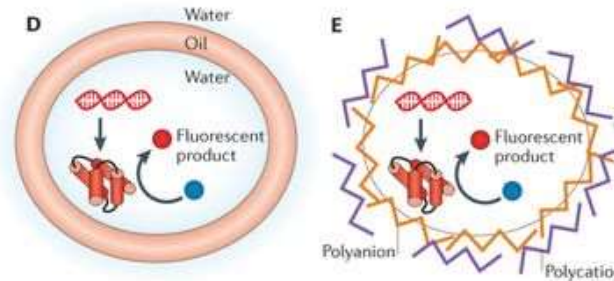
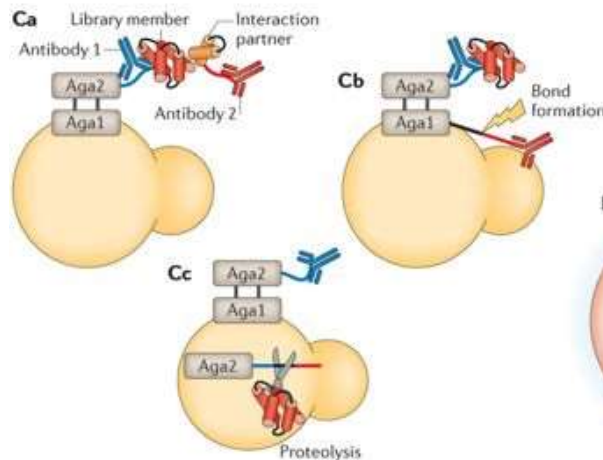
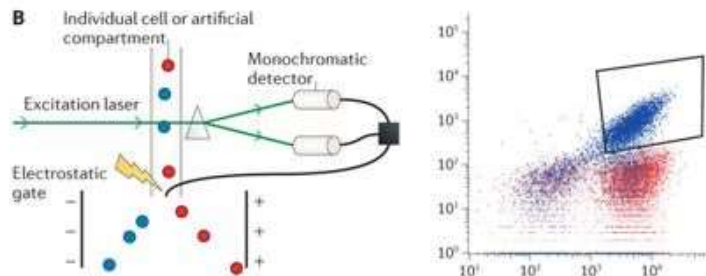
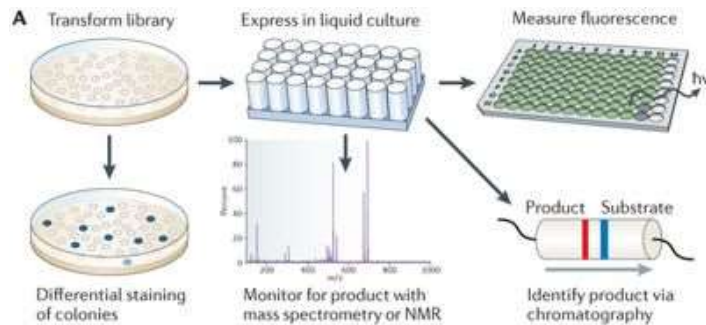
- Both strategies use **the cell (or virus particle) as the physical linkage between genotype and phenotype** through the directed evolution process.
- *Ex vivo* platforms tend to focus diversity (a) on to a single target gene, whereas *in vivo* platforms can extend that to metabolic pathways or even whole genomes.
- Once generated, the diverse repertoires are partitioned (b) with active (blue) variants preferentially recovered over inactive variants (orange).
- Partition by phenotype is linked to genotype recovery and amplification (c) which can take place in a single step if cells are still viable (as is the norm for *in vivo* methodologies). Alternatively, as shown for the *ex vivo* selection (light green boxes), genotype recovery and amplification can be separated, introducing different limitations to the process. The amplified recovered genotypes are the starting point of a subsequent round of selection.

# In vitro selection



- Platforms for *in vitro* selection can be broadly divided by the available redundancy of phenotype and genotype linkages.
- In a number of selection strategies, the link is unique – **a lone genotype molecule is linked to a lone molecule that may have the phenotype being selected (a)**.
- Compartmentalization strategies enable redundancy in the system with one-to-many [redundant genotype to lone phenotype (b) or lone genotype to pooled phenotype (not shown)] and many-to-many [redundant genotype to pooled phenotype (c)] mappings between phenotype and genotype available.

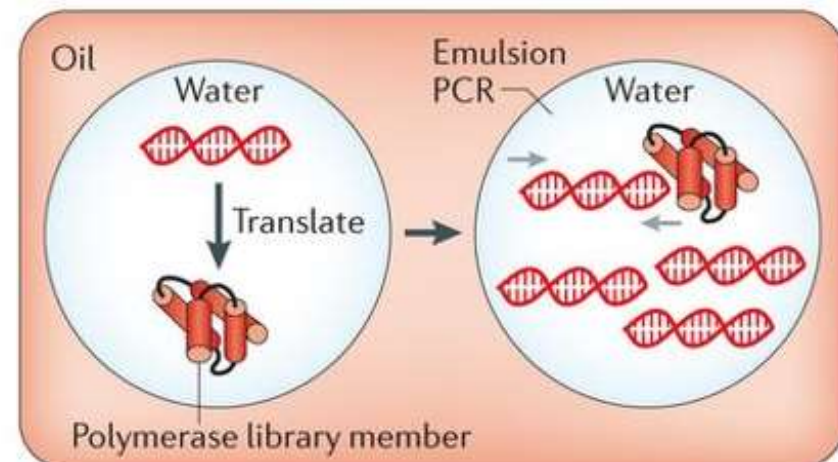
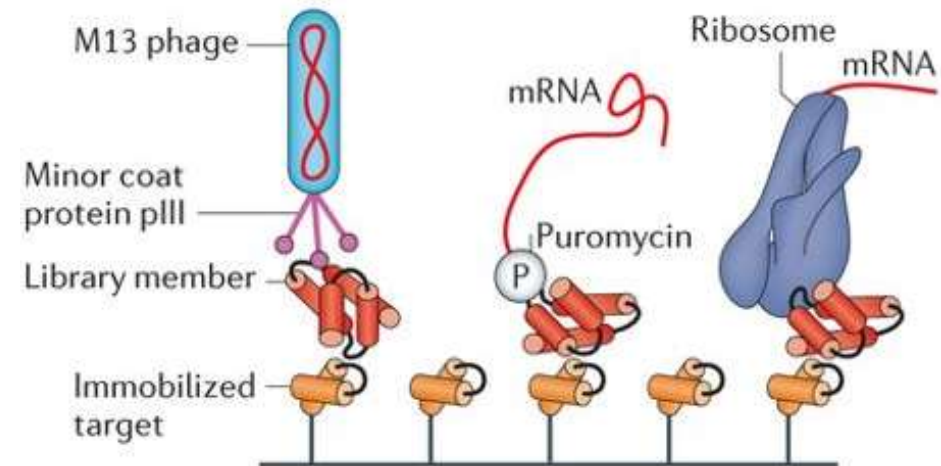
# Overview of screening and selection strategies in DE cycle



- Clonally isolated variants can be screened as colonies on solid media or as wells in liquid culture. Fluorescent or colorimetric reporters are measured by automated microtitre plate readers, chromatography, MS or NMR
- Fluorescence-activated cell sorting (FACS) enables the fluorescence measurement of individual cells and the separation of distinct subpopulations by electrostatic deflection
- Yeast display techniques enable FACS screens of protein–protein interactions, bond formation and peptide bond cleavage
- *In vitro* compartments entrap DNA, translated proteins and fluorogenic substrates, allowing the fluorescence-activated sorting of functional variants

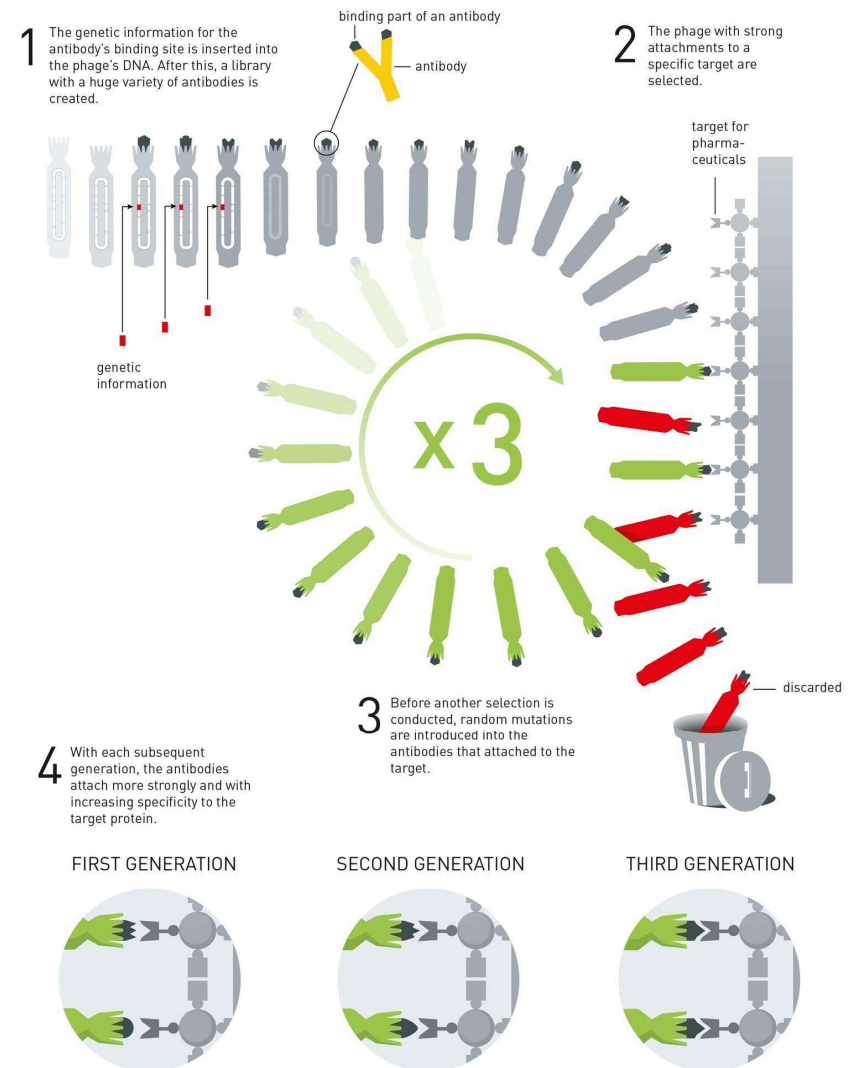
## Screening and selection methods

- **Virus display methods**  
phage display, baculovirus display
- **Cell surface display systems**  
bacteria, insect and mammalian cells
- **Cell-free display systems**  
mRNA display, ribosome display, covalent and non-covalent display, in vitro compartmentalization



# Phage display as the tool for directed evolution

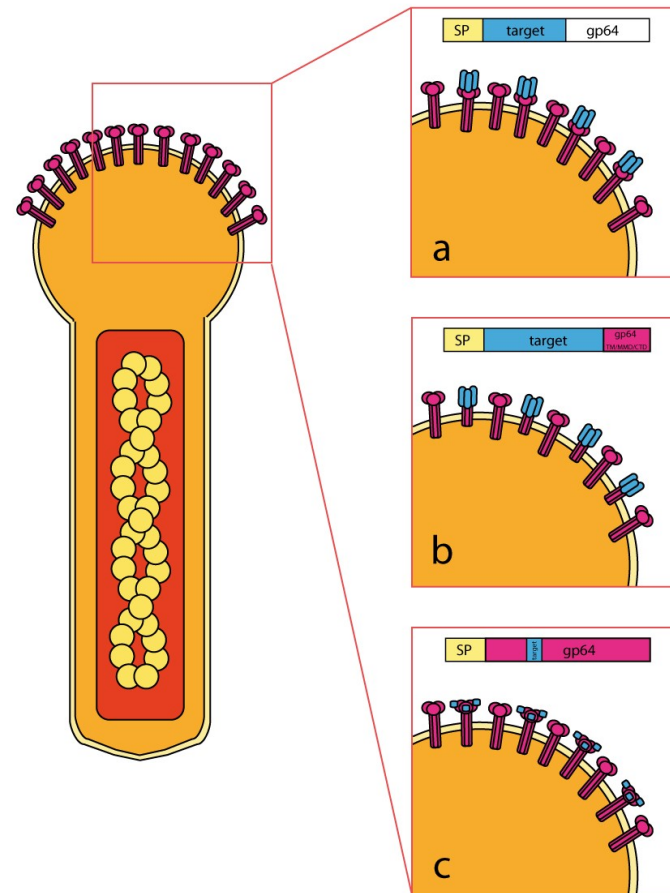
- Since its introduction in 1985, phage display has had a tremendous impact on the discovery of peptides that bind to a variety of receptors, the generation of binding sites within predefined scaffolds, and the creation of high-affinity antibodies without immunization
- Its application to enzymology has required the development of techniques that couple enzymatic activity to selection protocols based on affinity chromatography
- The protein of interest is fused to the N-terminal of g3p. The foreign gene is inserted between the sequences encoding the signal peptide and the mature coat protein g3p
- On phage morphogenesis, the fusion protein is displayed on the phage surface





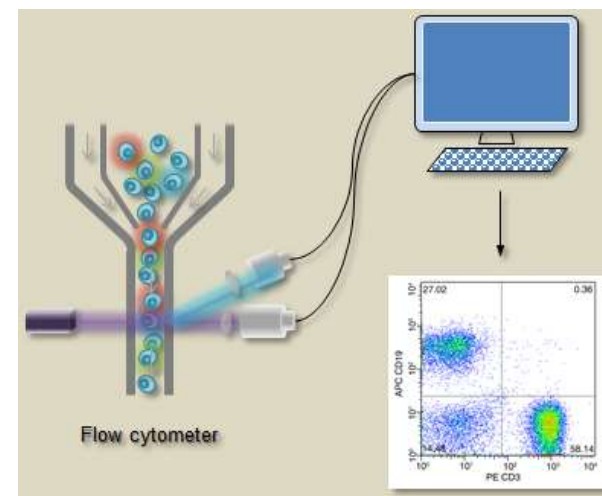
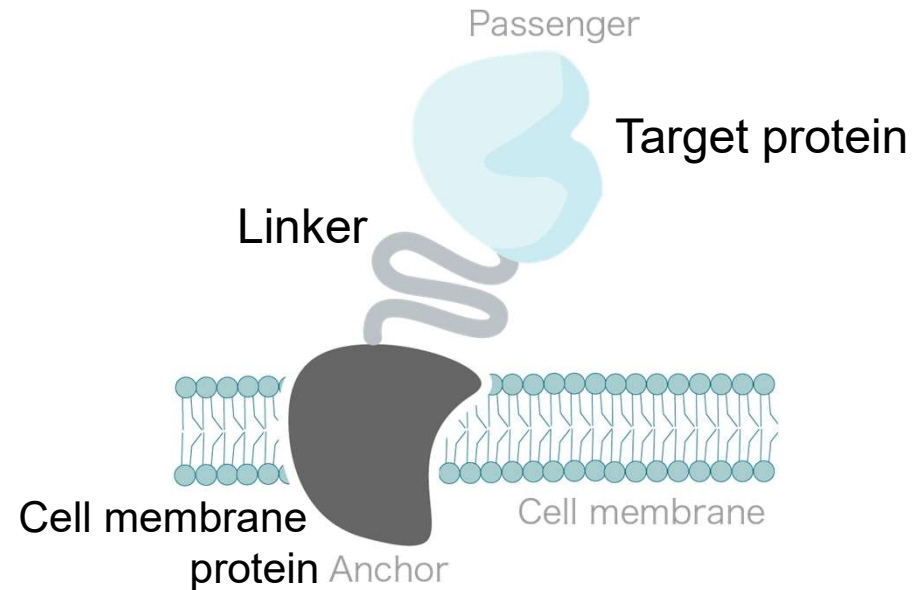
## Baculovirus display system

- Baculovirus is a large DNA insect-infecting virus
- Baculovirus surface glycoprotein Gp64 is expressed early and late in the infection of an insect cell
- It is a 64 kDa protein which forms trimers and locates in the BV envelope with a polarized distribution
- As Gp64 is a transmembrane protein that exposes an outer domain, it can be used to display a selected protein on the BV surface
- A chimeric Gp64 can be constructed to contain the protein of interest allowing it to be incorporated in the BV structure upon infection of insect cells



# Cell surface display systems

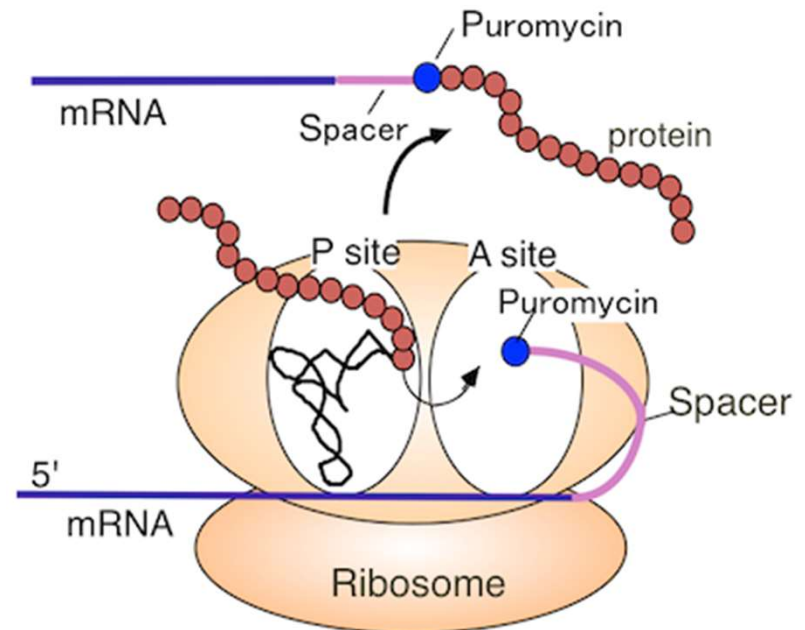
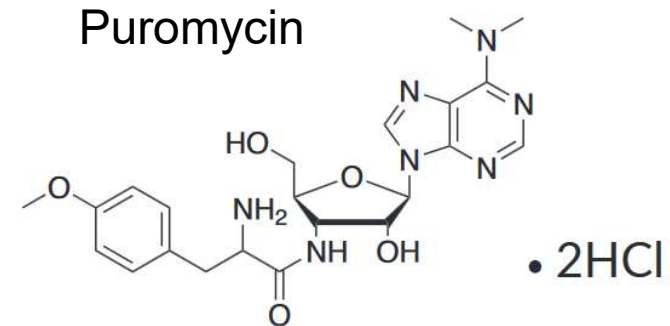
- Cell surface display (bacteria, yeasts, insect or mammalian cells) is a strategy used for in vitro protein evolution
- Libraries of proteins displayed on the surface of cells can be screened using flow cytometry
- This technique allows us to link the function of a protein with the gene that encodes it
- Cell surface display can be used to find target proteins with desired properties, e.g. to make high-affinity ligands, creation of novel vaccines, identification of enzyme substrates



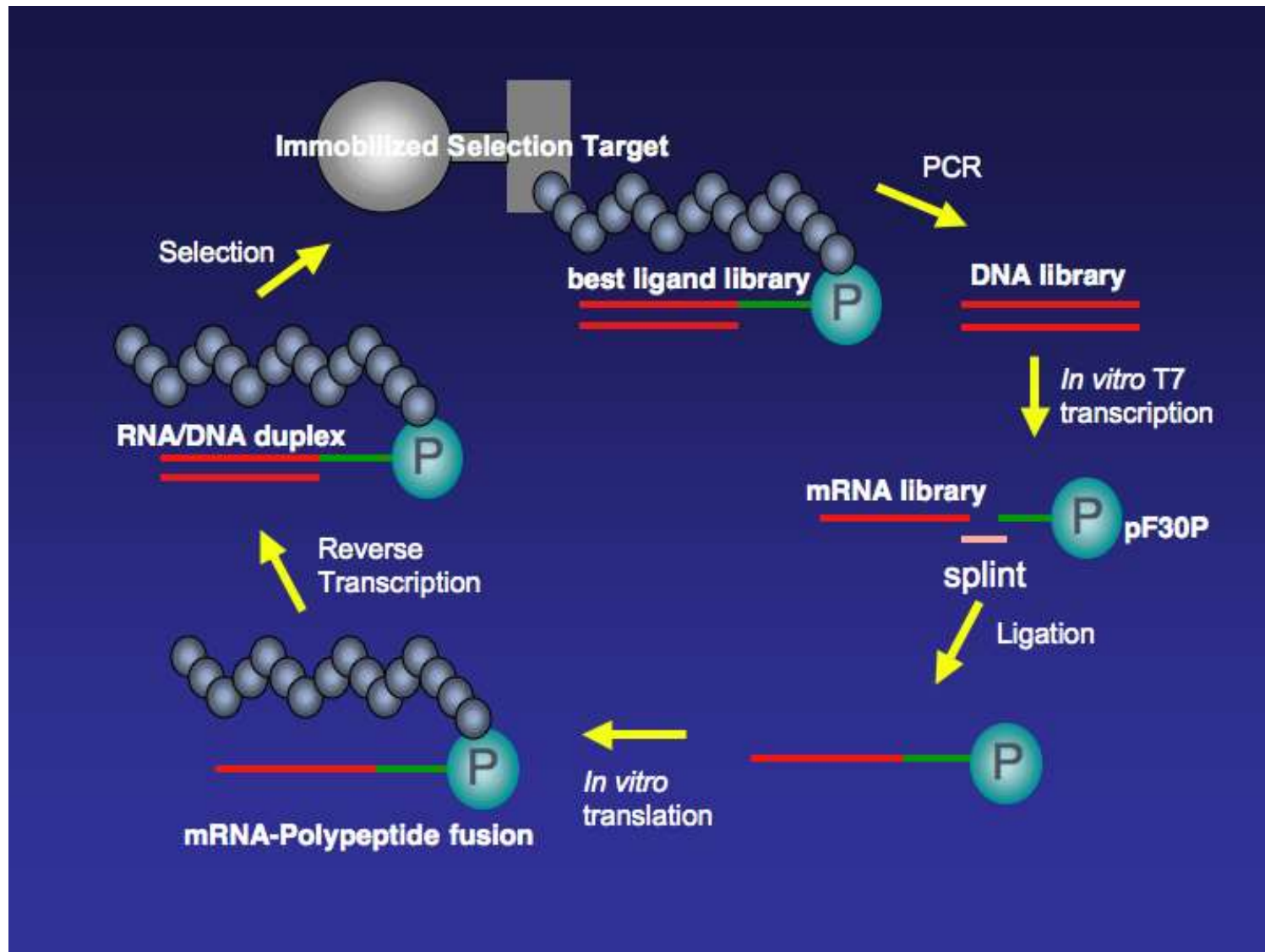
## mRNA display



- **mRNA display** is a display technique used for *in vitro* protein evolution
- The process results in translated proteins that are linked with their mRNA via a puromycin linkage
- **Puromycin** is an analogue of the 3' end of a tyrosyl-tRNA with a part of its structure mimics a molecule of and the other part mimics a molecule of tyrosine
- Compared to the cleavable ester bond in a tyrosyl-tRNA, puromycin has a non-hydrolysable amide bond. As a result, puromycin interferes with translation, and causes premature release of translation products
- The complex then binds to an immobilized target in a selection step, and mRNA-protein fusions that bind well are then reverse transcribed to cDNA and their sequence amplified via a PCR
- The result is a nucleotide sequence that encodes a peptide with high affinity for the molecule of interest

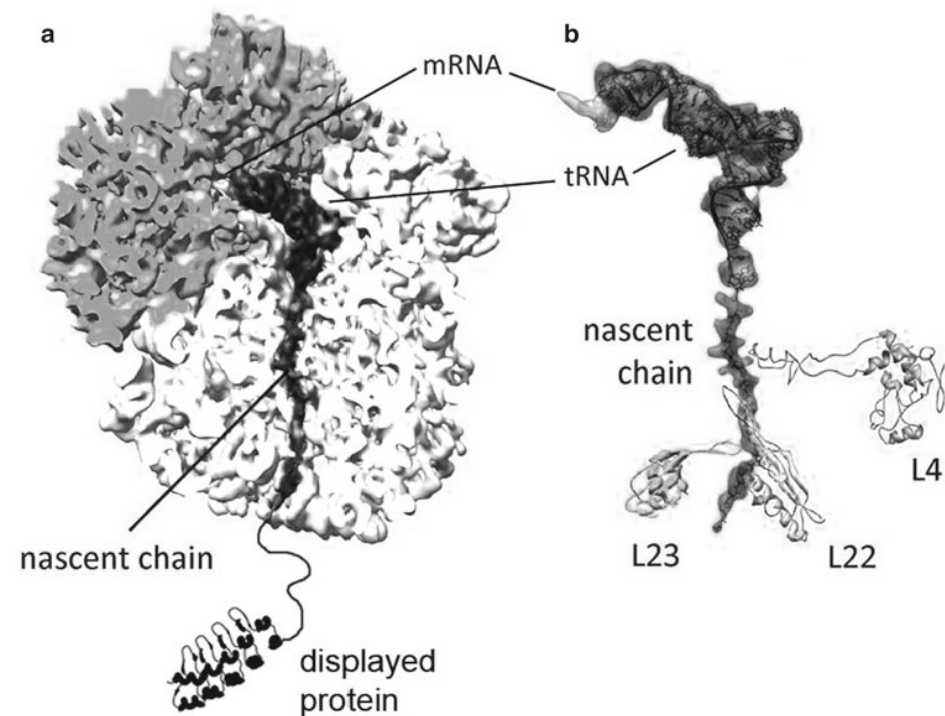


## mRNA display workflow



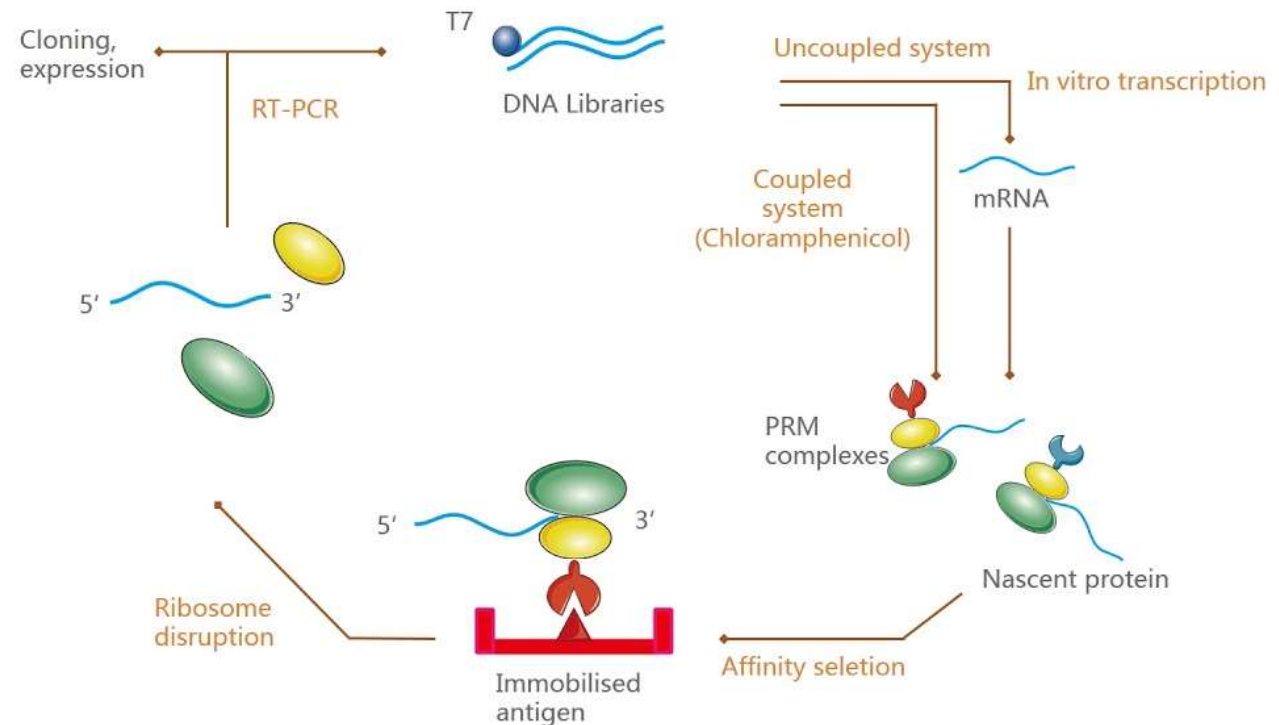
## Ribosome display

- **Ribosome display** is an *in vitro* evolution technology for proteins
- It is based on *in vitro* translation, but prevents the newly synthesized protein and the mRNA encoding it from leaving the ribosome
- It thereby **couples phenotype and genotype**
- Since no cells need to be transformed, very large libraries can be used directly in selections, and the *in vitro* amplification provides a very convenient integration of random mutagenesis that can be incorporated into the procedure.

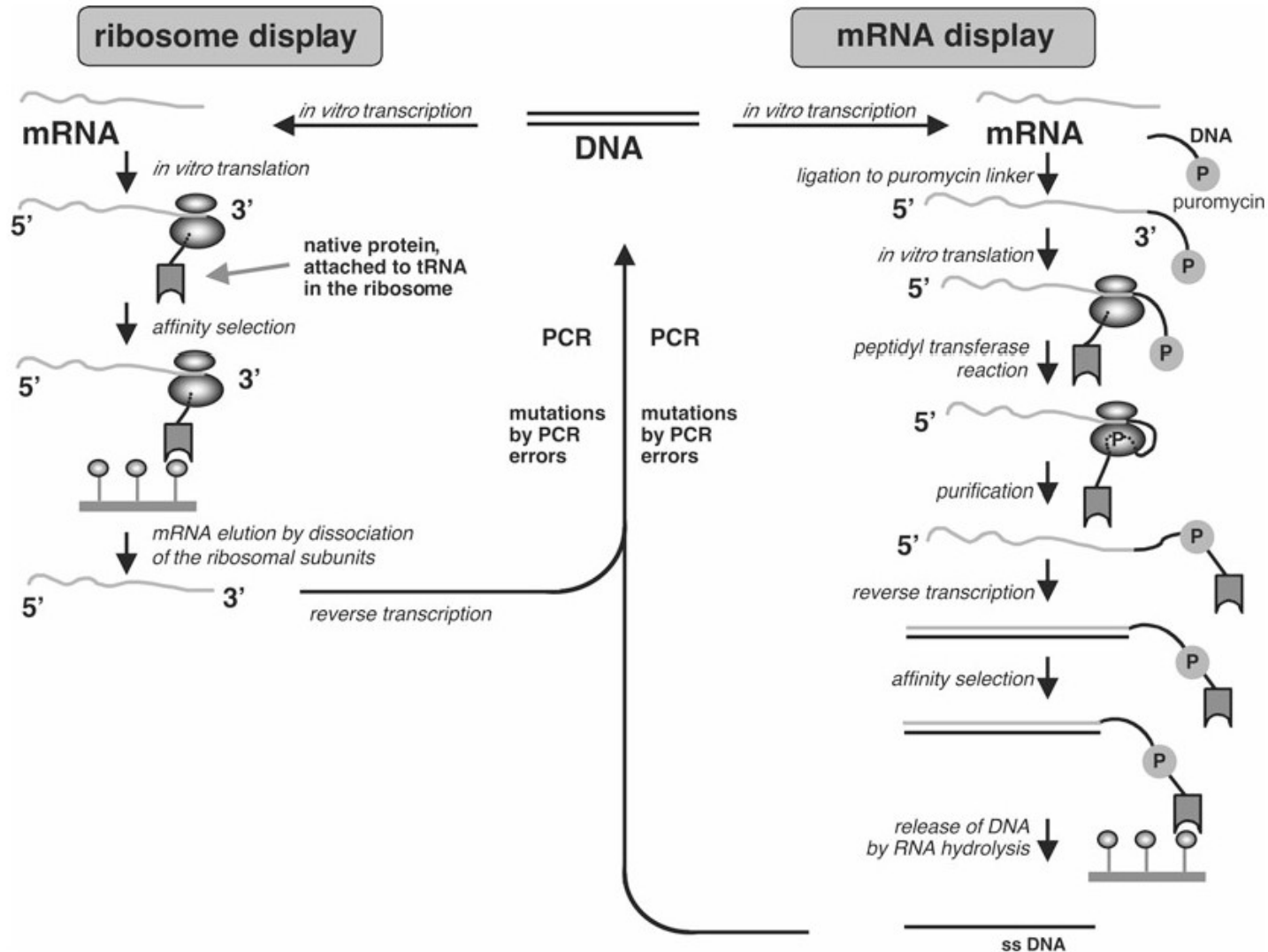


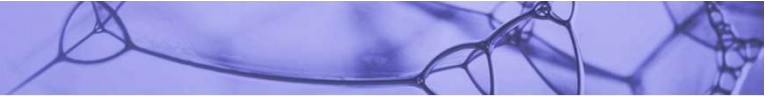
## Ribosome display: step by step

- Ribosome display starts with a DNA library
- Each sequence is transcribed, and then translated *in vitro* into polypeptide
- However, the DNA library coding for a particular library of binding proteins is genetically fused to a spacer sequence lacking a stop codon before its end
- The **lack of a stop codon** prevents release factors from binding and triggering the disassembly of the translational complex
- It results in a **complex of mRNA, ribosome, and protein** which can bind to surface-bound ligand or substrate



# Ribosome display vs. mRNA display



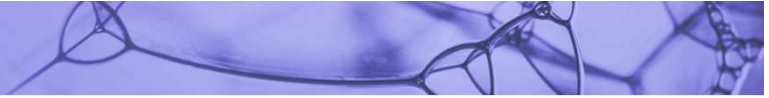


# Semi-rational protein design

---

Concepts  
Methods  
Applications



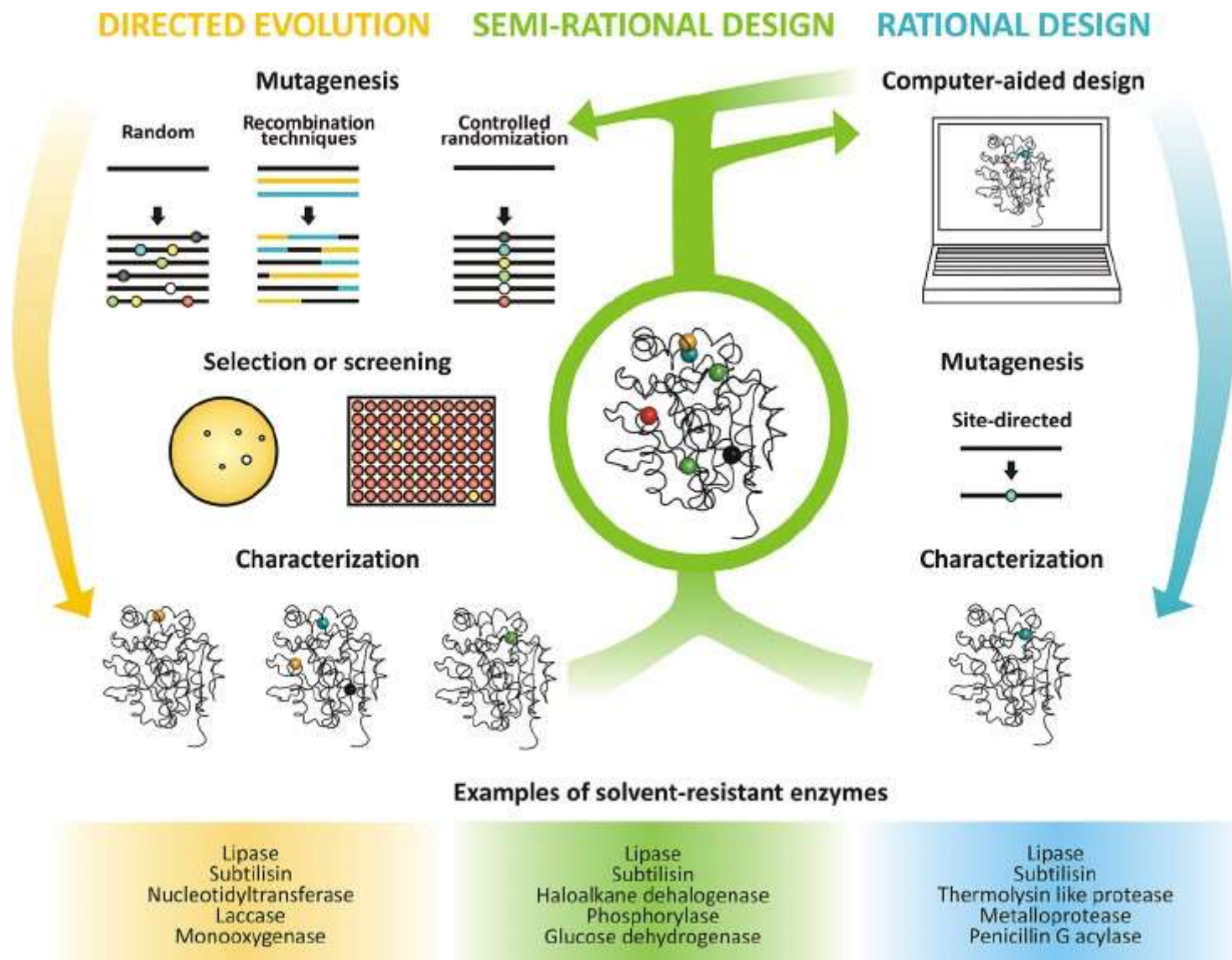


## Combining rational design and directed evolution

---

- Combined, '**semi-rational**' approaches are being investigated to address the limitations of both rational design and directed evolution
- Beneficial mutations are rare, so large numbers of random mutants have to be screened to find improved variants
- '**Focused libraries**' concentrate on randomising regions thought to be richer in beneficial mutations for the mutagenesis step of DE. A focused library contains fewer variants than a traditional random mutagenesis library and so does not require such high-throughput screening
- Creating a focused library requires some knowledge of which residues in the structure to mutate. For example, knowledge of the active site of an enzyme may allow just the residues known to interact with the substrate to be randomised.
- Alternatively, knowledge of which protein regions are variable in nature can guide mutagenesis in just those regions.

# Combining protein engineering approaches



# Discovery of PET-breaking enzyme (PETase)

PET bottles



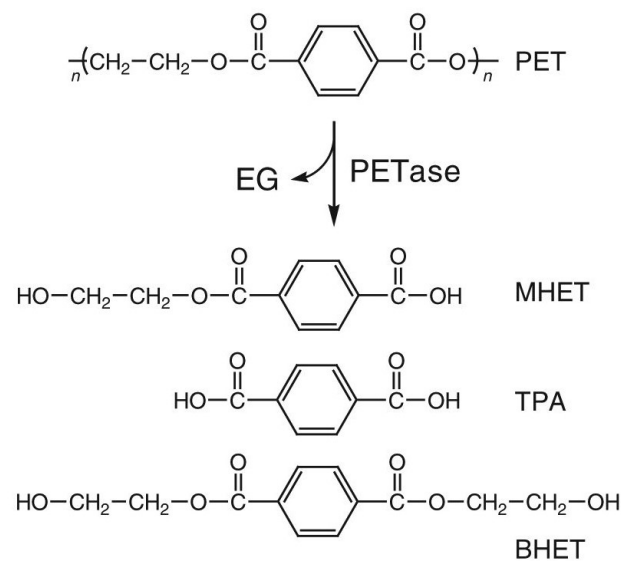
PET art



Plastic waste & pollution



- PET plastic, short for polyethylene terephthalate, is the fourth most-produced plastic, used to make things such as beverage bottles and carpets, most of which are not being recycled
- **PETase** is an esterase class of enzymes that catalyze the hydrolysis of polyethylene terephthalate (PET) plastic to monomeric mono-2-hydroxyethyl terephthalate (MHET)



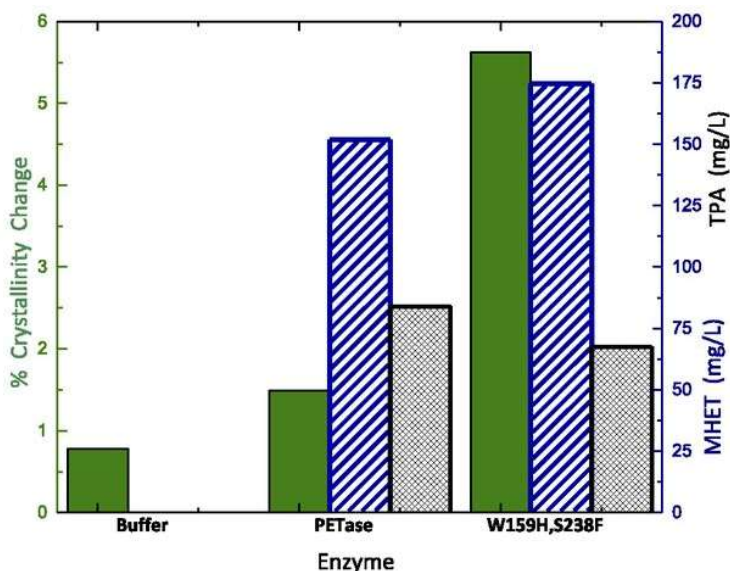
# Engineering of PET-breaking enzyme (PETase)



NEW RESEARCH IN Physical Sciences Social Sciences

## Characterization and engineering of a plastic-degrading aromatic polyesterase

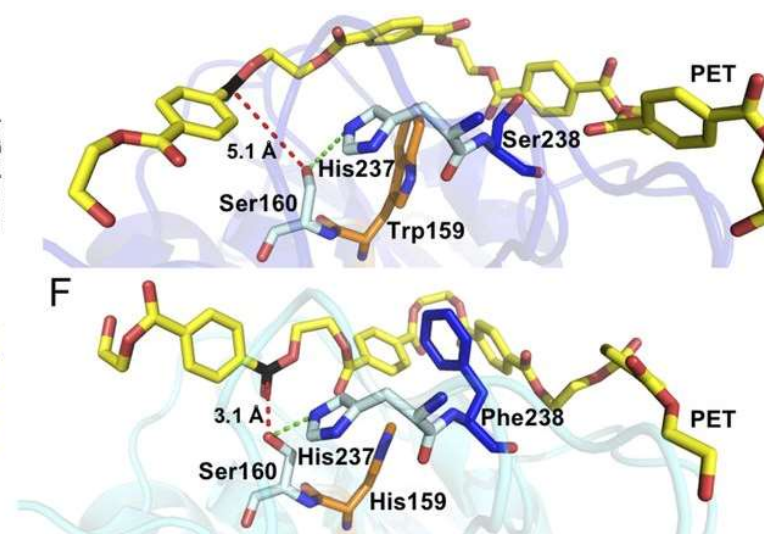
Harry P. Austin, Mark D. Allen, Bryon S. Donohoe, Nicholas A. Rorrer, Fiona L. Kearns, Rodrigo L. Silveira, Benjamin C. Pollard, Graham Dominick, Ramona Duman, Kamel El Omari, Vitaliy Mykhaylyk, Armin Wagner, William E. Michener, Antonella Amore, Munir S. Skaf, Michael F. Crowley, Alan W. Thorne, Christopher W. Johnson, H. Lee Woodcock, John E. McGeehan, and Gregg T. Beckham



Research Synthetic Biology—Article

## Enhanced Poly(ethylene terephthalate) Hydrolase Activity by Protein Engineering

Yuan Ma<sup>a, b</sup>, Mingdong Yao<sup>a, b</sup>, Bingzhi Li<sup>a, b</sup>, Mingzhu Ding<sup>a, b</sup>, Bo He<sup>a, b</sup>, Si Chen<sup>a, b</sup>, Xiao Zhou<sup>a, b</sup>, Yingjin Yuan<sup>a, b</sup>



The S238F mutation provides new  $\pi$ -stacking and hydrophobic interactions to adjacent terephthalate moieties, while the conversion to His159 from the bulkier Trp allows the PET polymer to sit deeper within the active-site channel.



CARBIOS TO PRESENT AT THE XXV ENZYME  
ENGINEERING CONFERENCE IN WHISTLER, BRITISH  
COLUMBIA, CANADA

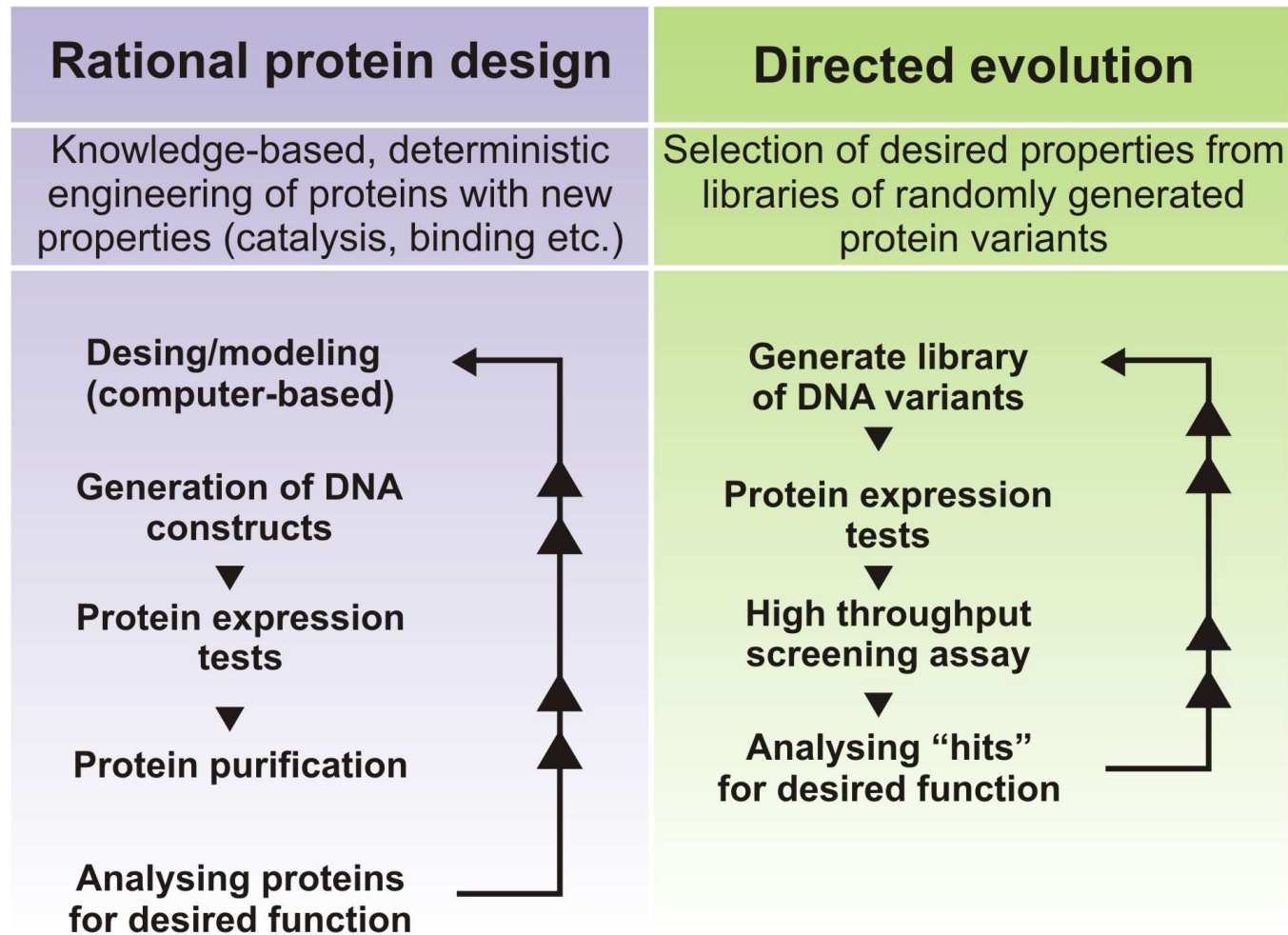
CLERMONT-FERRAND ON SEPTEMBER 16, 2019 AT 6:45AM

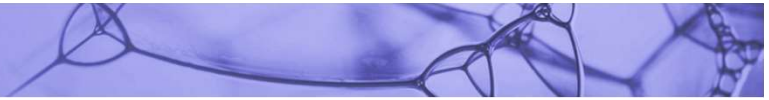
- **"Addressing the problem of plastic waste: development of an enzymatic process for PET recycling"**
- **by Alain MARTY, Chief Scientific Officer of CARBIOS**

Clermont-Ferrand, France, September 16, 2019 (06:45 AM CEST) – CARBIOS (Euronext Growth Paris: ALCRB), a company pioneering new, bio-industrial solutions to reinvent the lifecycle of plastic and textile polymers, will present the company's innovative enzymatic process for PET plastic recycling at the *XXV Enzyme Engineering Conference* in Whistler, British Columbia, Canada.

***CARBIOS Chief Scientific Officer, Prof. Alain Marty will explain the development of novel enzymes allowing PET plastics such as bottles, packaging and textiles to be recycled in an eco-responsible way. Drawing on the company's proprietary enzyme engineering technologies and the extensive experience of its team and partners, CARBIOS has developed a breakthrough solution to infinitely recycle all kind of PET plastics and polyester fibers.***

## Rational protein design vs. directed evolution

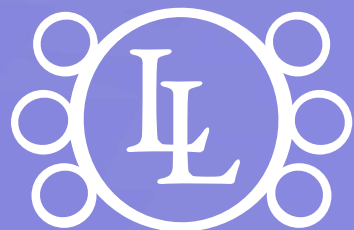




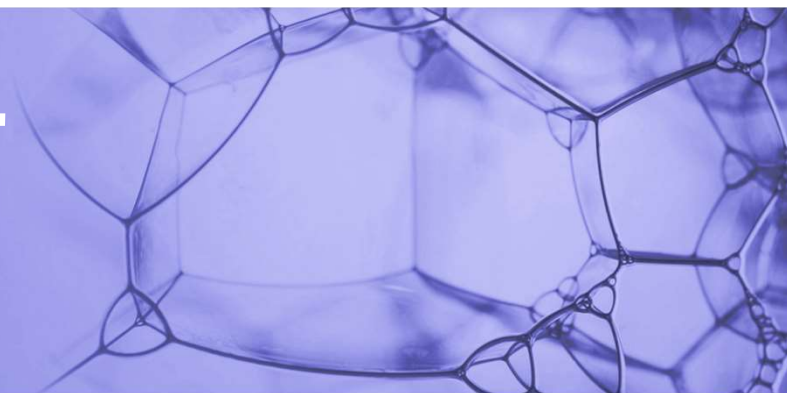
# Questions

---

Dr. Martin Marek  
Loschmidt Laboratories  
Faculty of Science, MUNI  
Kamenice 5, bld. A13, room 332  
[martin.marek@recetox.muni.cz](mailto:martin.marek@recetox.muni.cz)



**LOSCHMIDT**  
**LABORATORIES**





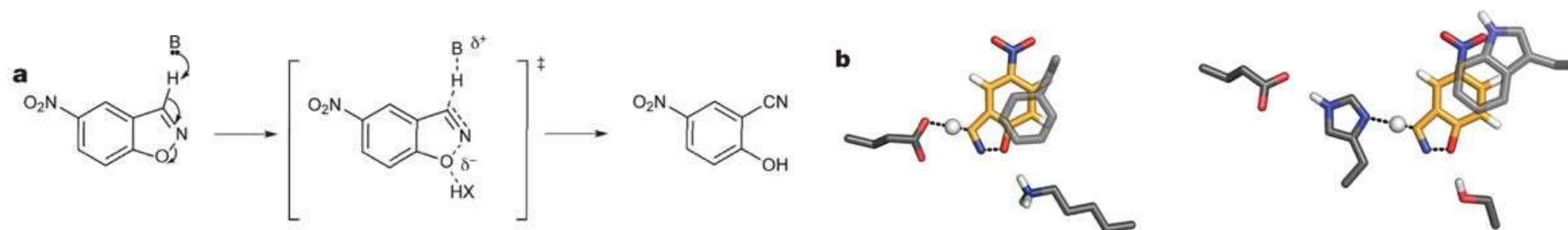


# Supplementary materials

---

## An example of coupling design and experiment

### The Kemp elimination: a model reaction for proton transfer from carbon atom

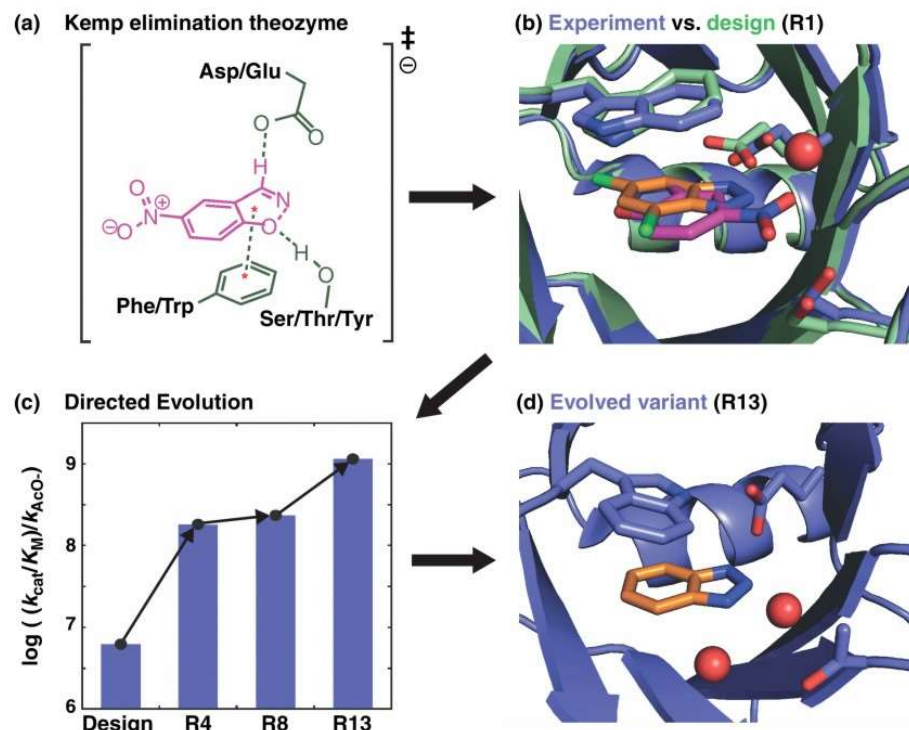


- a)** The Kemp elimination proceeds by means of a single transition state, which can be stabilized by a base deprotonating the carbon and the dispersion of the resulting negative charge; a hydrogen bond donor can also be used to stabilize the partial negative charge on the phenolic oxygen.
- b)** Examples of active site motifs highlighting the two choices for the catalytic base (a carboxylate (left) or a His–Asp dyad (right)) used for deprotonation, and a  $\pi$ -stacking aromatic residue for transition state stabilization. For each catalytic base, all combinations of hydrogen bond donor groups (Lys, Arg, Ser, Tyr, His, water or none) and  $\pi$ -stacking interactions (Phe, Tyr, Trp) were input as active site motifs into RosettaMatch.

# An example coupling design and experiment

Generation of an improved Kemp eliminase with a  $k_{\text{cat}}/K_{\text{M}}$  of  $570.000 \text{ M}^{-1} \text{ s}^{-1}$

1. Theozyme containing a carboxylate base, an aromatic side-chain for substrate binding, and a hydrogen bond donor to stabilize developing negative charge in the transition state was used as the template for design
2. The crystal structure of the first-generation enzyme (blue) complexed with the transition state analog shows good qualitative agreement with the computational design model (green). However, the ligand in the crystal structure (orange) is flipped relative to the designed orientation (pink)
3. The catalytic efficiency of the starting enzyme (R1) relative to the nonenzymic acetate-promoted reaction was increased by more than two orders of magnitude over thirteen rounds of mutagenesis and screening (R2–R13)
4. The structure of the best evolved variant R13 (blue) cocrystallized with benzotriazole (orange)



Current Opinion in Chemical Biology

Khersonsky *et al.*, PNAS 109:10358-10363 (2012)

Kries *et al.*, Curr. Opin. in Chem. Biol. 17:221-8 (2013)

## Directed evolution as a tool for synthetic biology

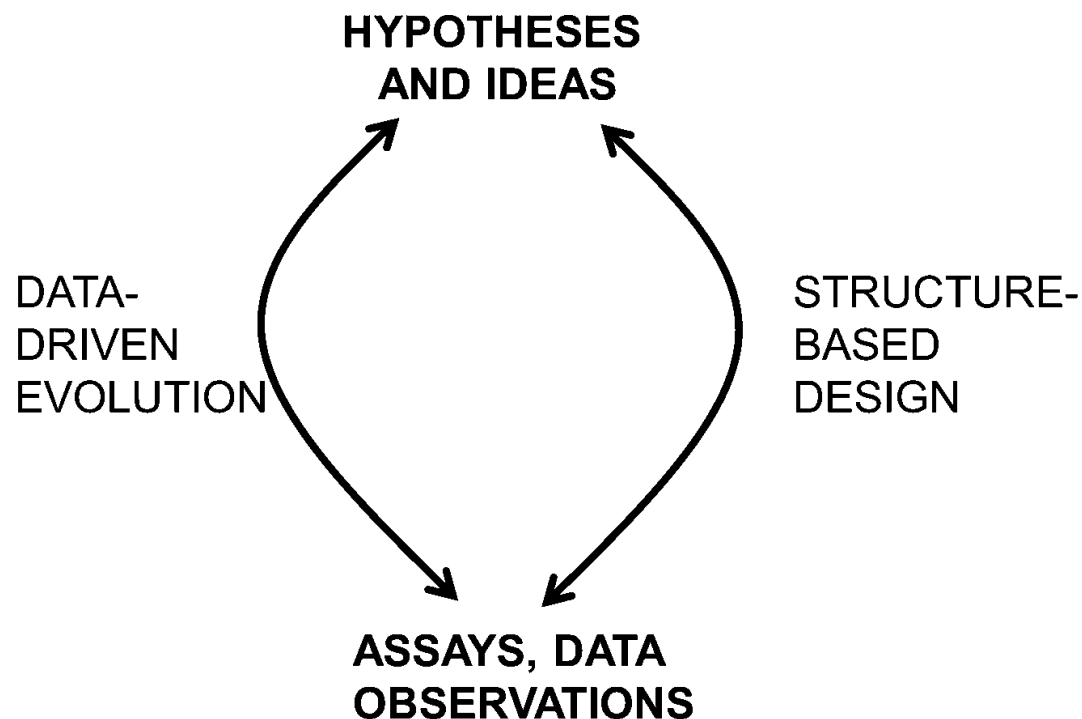
---

- Despite the diversity and versatility of selection platforms available, novel ones are regularly being developed—delivering custom solutions to ever growing challenges. As molecular biology methods and technologies develop, novel strategies to diversify and partition a biopolymer population become available, increasing experimental control, throughput and pace.
- Directed evolution performs the design, build and test cycle of synthetic biology on a scale that is unnatural in engineering: it would be the equivalent of building millions (or even trillions) of slightly different machines (e.g. watches) in search of a specific improvement (e.g. more precise time keeping). On an engineering scale, such approach would be prohibitive, if even possible. However, on a biological scale, millions are still small numbers, barely able to cover the immediate sequence neighbourhood of even a small protein.
- Directed evolution has successfully been used to isolate novel and optimize existing function on natural and synthetic biopolymers. But its key strength lies on how it deals with uncertainty. Even in the absence of complete understanding of complex biological systems, directed evolution is a powerful tool to re-engineer even the most central truths of life on our planet—that life is based on DNA and RNA, and that life requires (or is optimal with only) 20 amino acids.



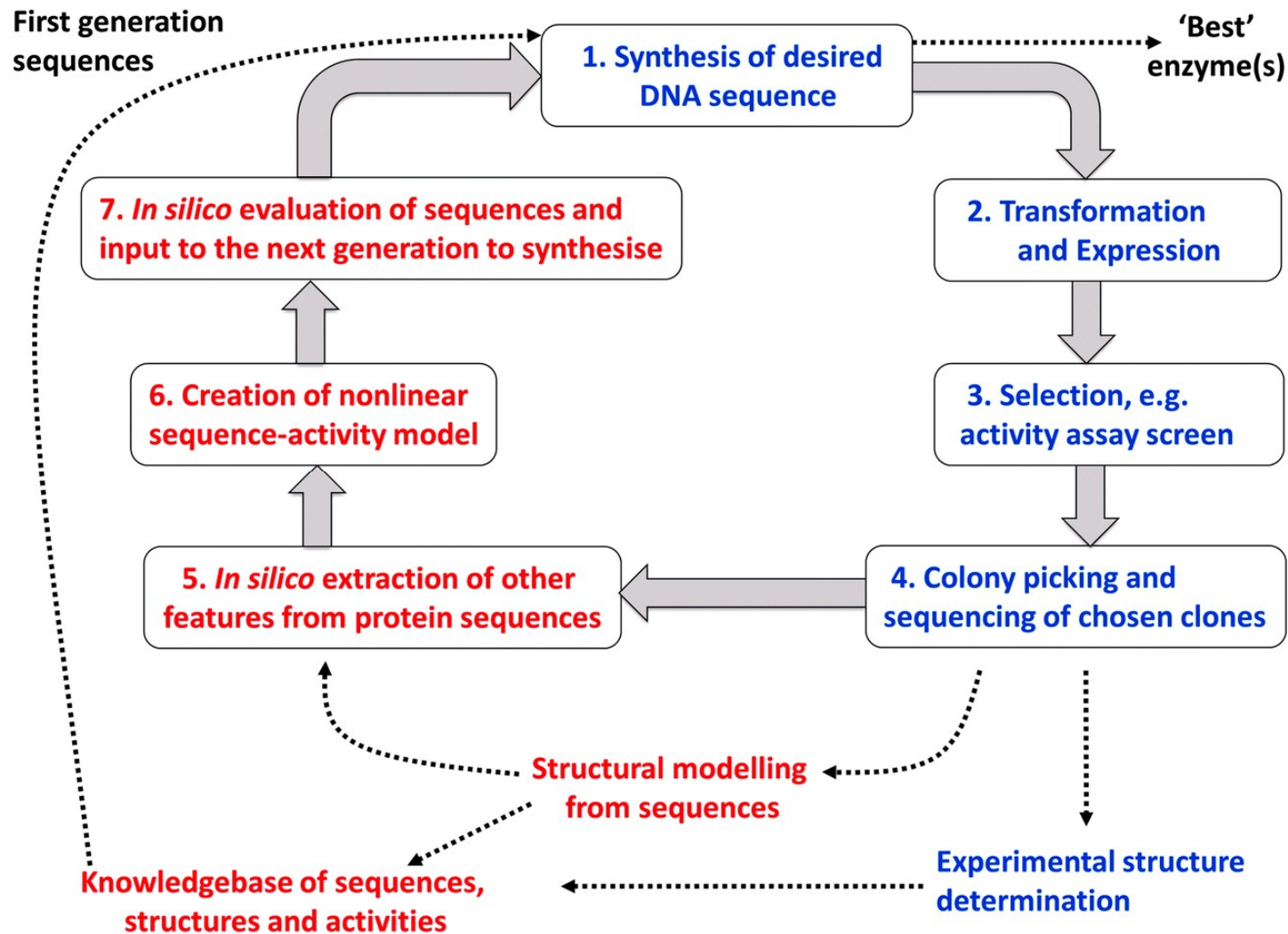
## The cycle of knowledge in directed evolution

---



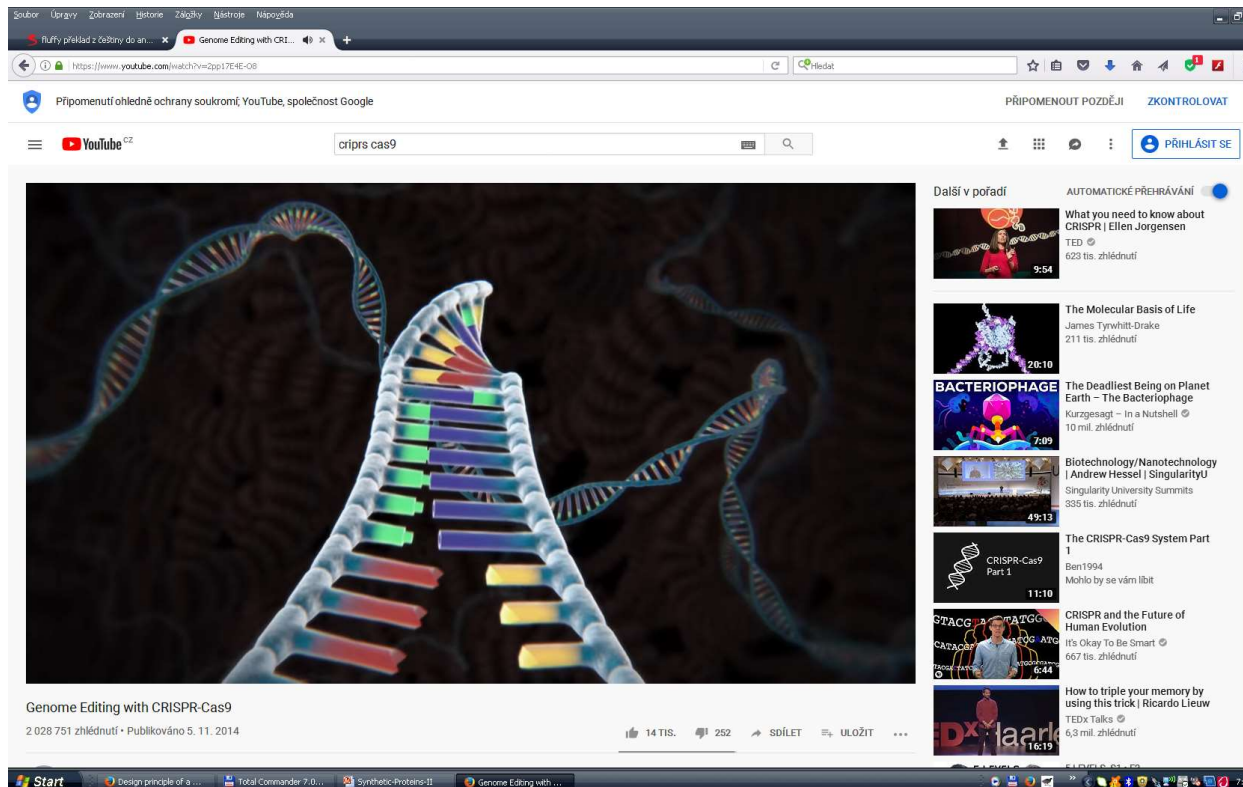
- Both structure-based design and a more empirical data-driven approach can contribute to the evolution of a protein with improved properties, in a series of iterative cycles.

# Basic elements of a mixed computational and experimental programme in directed evolution



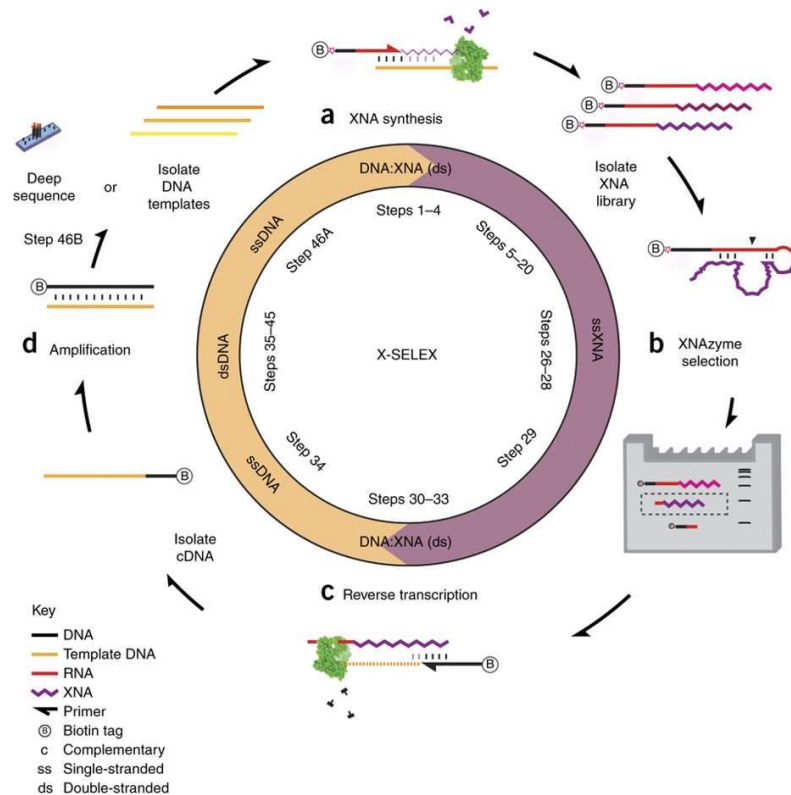
# The CRISPR/Cas9 system on YouTube

<https://www.youtube.com/watch?v=bXnWIk8FgKc>  
<https://www.youtube.com/watch?v=OjNrbPMXyMA>  
<https://www.youtube.com/watch?v=0dRT7slyGhs>  
<https://www.youtube.com/watch?v=2pp17E4E-O8>



# SELEX

## Systematic Evolution of Ligands by EXponential enrichment



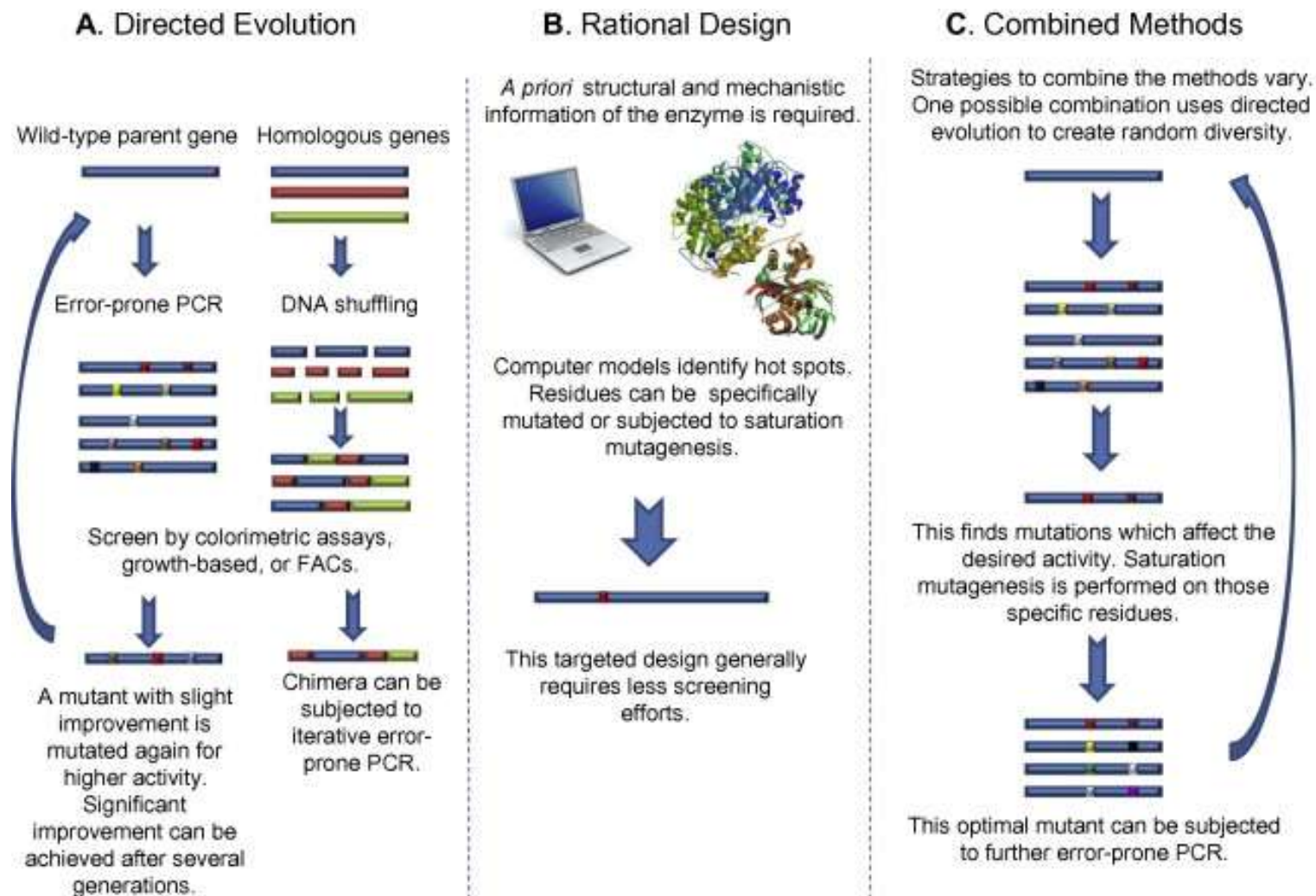
- Systematic evolution of ligands by exponential enrichment (SELEX) is well established allowing not only directed evolution of functional RNA molecules, but also functional DNA and synthetic nucleic acids (XNAs).
- In most cases, selection involves isolating functional nucleic acids, converting them to DNA (to allow efficient amplification by PCR), and the regeneration of the functional nucleic acid repertoire for further rounds of selection.

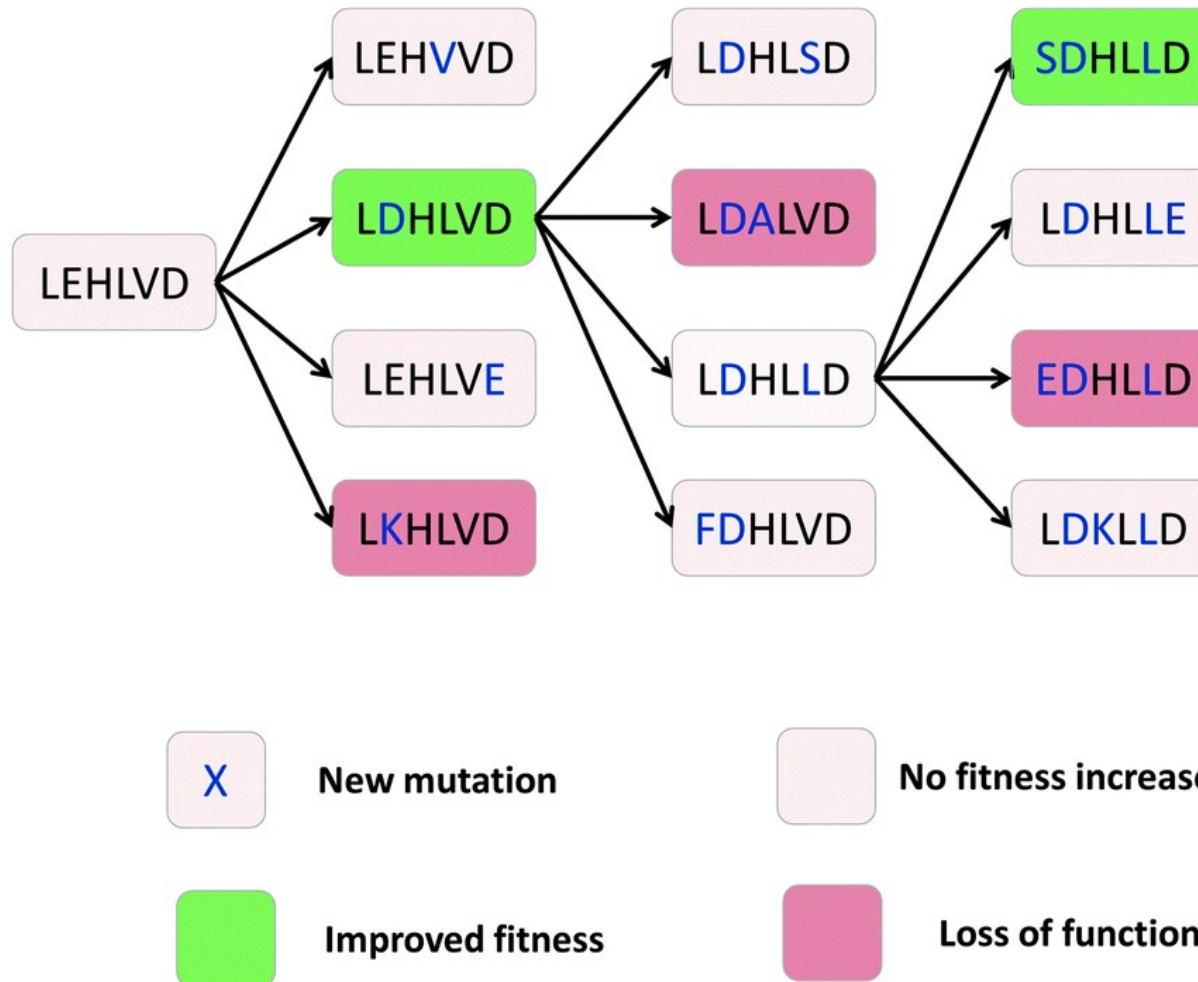
**(a)** Diverse repertoires of XNA molecules may be synthesized (using engineered DNA-dependent XNA polymerases). **(b)** Catalytic XNAs (XNAzymes) may be selected by reacting libraries of XNAs tagged with substrates of interest (e.g., a disease-related RNA) and isolating on the basis of change in substrate (e.g., urea-PAGE gel shift upon RNA hydrolysis). **(c,d)** XNAzymes may be subsequently reverse-transcribed using engineered XNA-dependent DNA polymerases **(c)**, yielding cDNA that may be amplified to enable either deep sequencing or generation of templates for XNA synthesis **(d)**, and further rounds of X-SELEX.





# Rational protein design workflow





**Additivity:** Additivity implies simple continuing fixing of improved mutations, and follows from a model in which selection in natural evolution quite badly disfavors lower fitnesses, a circumstance known from Gillespie as ‘strong selection, weak mutation’. For small changes (close to neutral in a fitness or free energy sense), additivity may indeed be observed, and has been exploited extensively in DE. If additivity alone were true, however (and thus there is no epistasis for a given protein at all) then a rapid strategy for DE would be to synthesise all 20L amino acid variants at each position (of a starting protein of length L) and pick the best amino acid at each position. However, the very existence of convergent and divergent evolution implies that landscapes are rugged (and hence epistatic), so at the very least additivity and epistasis must coexist.

**Epistasis:** The term ‘epistasis’ in DE covers a concept in which the ‘best’ amino acid at a given position depends on the amino acid at one or more other positions. In fact, we believe that one should start with an assumption of rather strong epistasis, as did Wright. Indeed the rugged fitness landscape is itself a necessary reflection of epistasis and vice versa. Thus, epistasis may be both cryptic and pervasive, the demonstrable coevolution goes hand in hand with epistasis, and “to understand evolution and selection in proteins, knowledge of coevolution and structural change must be integrated”.

Promiscuity. The concept of enzyme promiscuity mainly implies that some enzymes may bind, or catalyse reactions with, more than one substrate, and this is inextricably linked to how one can traverse evolutionary landscapes. It clearly bears strongly on how we might seek to effect the directed evolution of biocatalysts.

# SELEX

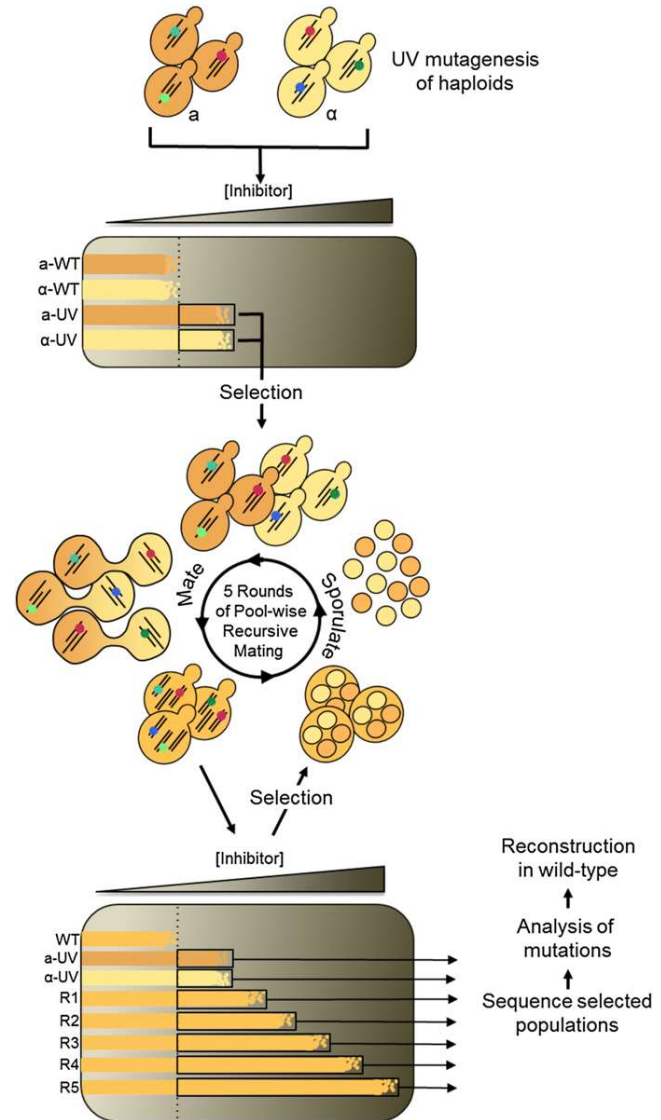
---

**SELEX—A (r)evolutionary method to generate high-affinity nucleic acid ligands**

<https://experiments.springernature.com/articles/10.1038/nprot.2015.104>

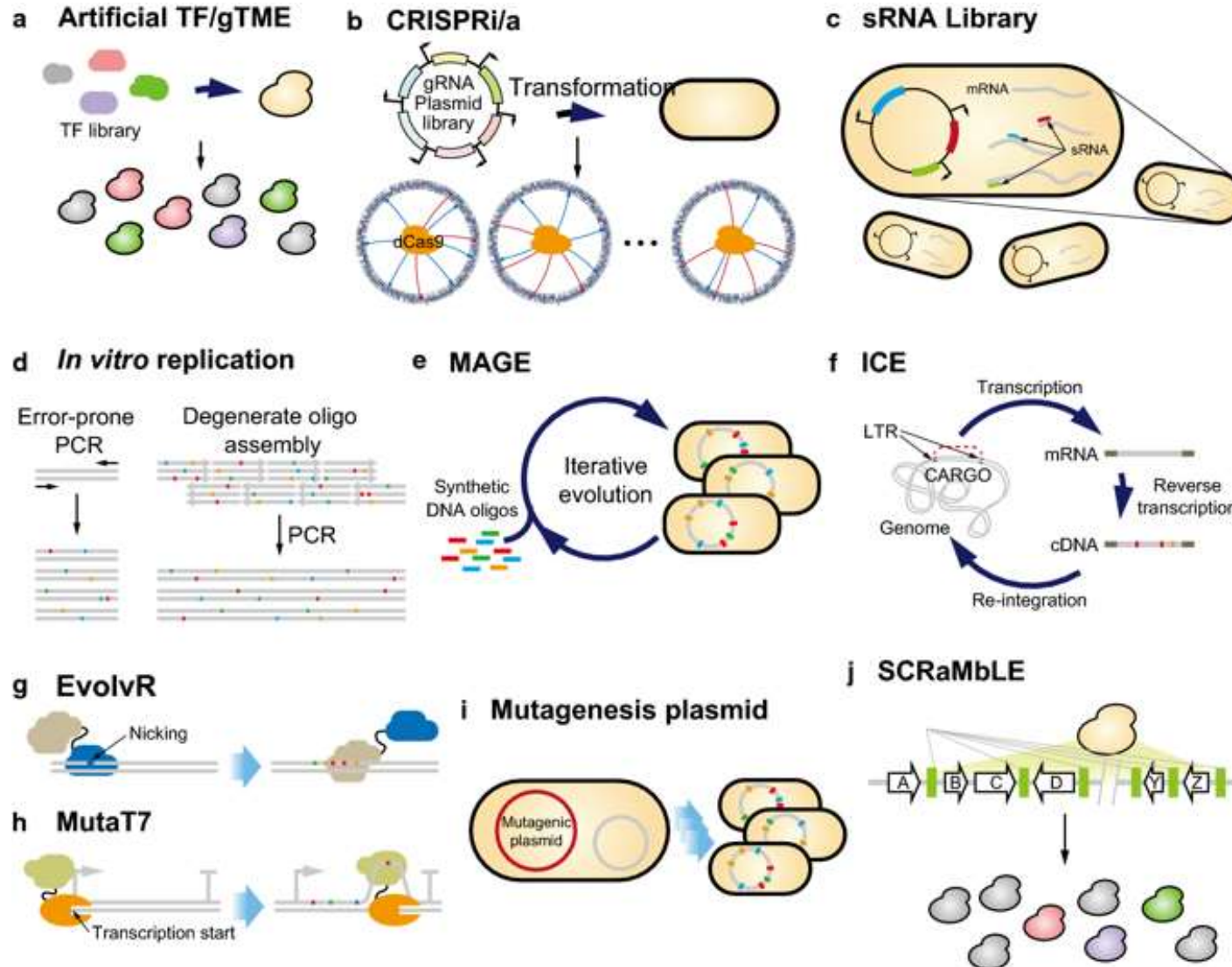
<https://www.nature.com/articles/nprot.2015.104/figures/1>

# Genome shuffling





# Continuous directed evolution techniques

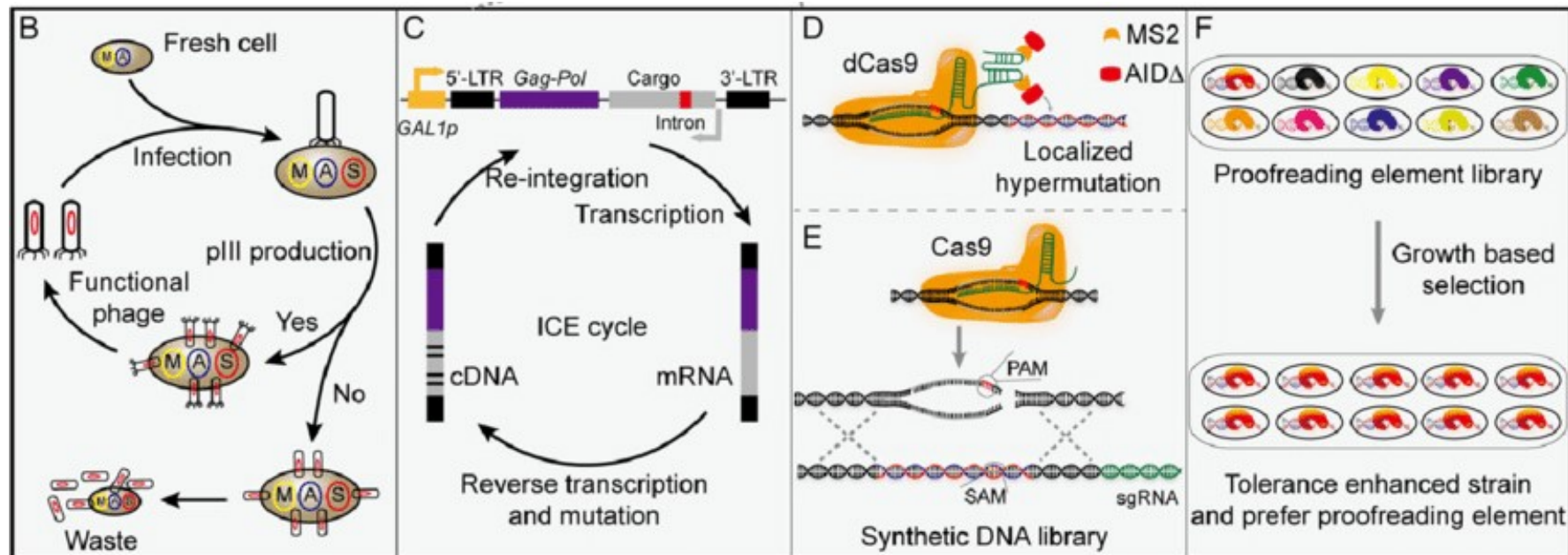




## Continuous directed evolution strategies

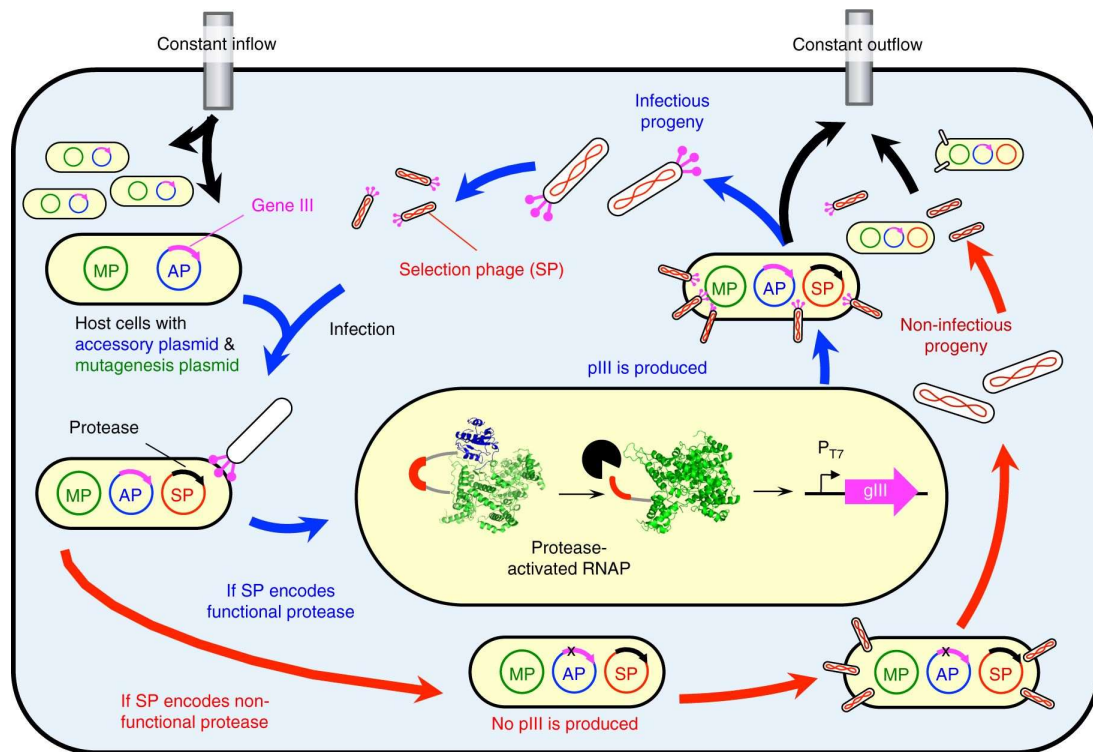
Strategies for *in vivo* continuous directed evolution approaches include:

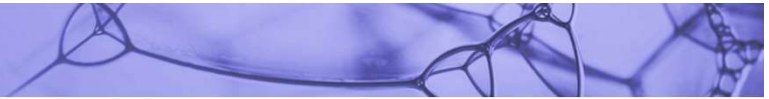
- Phage-assisted continuous evolution– PACE (B)
- In vivo continuous evolution – ICE (C)
- CRISPR/Cas9-based hypermutation strategies – CRSPR-X (D) and CREATE (E)
- GREACE (F)



# Overview of phage-assisted continuous evolution (PACE)

- A culture of host *E. coli* continuously dilutes a fixed-volume vessel containing an evolving population of selection phage (SP) in which essential phage gIII has been replaced by a protease gene.
- These host cells contain an arabinose-inducible mutagenesis plasmid (MP) and an accessory plasmid (AP) that supplies gIII.
- The expression of gIII is made protease-dependent through the use of a protease-activated RNA polymerase (PA-RNAP) consisting of T7 RNA polymerase fused through a cleavable substrate linker to T7 lysozyme, a natural inhibitor of T7 RNAP transcription.
- If an SP encodes a protease capable of cleaving the substrate linker, then the resulting liberation of T7 RNAP leads to the production of pIII and infectious progeny phage encoding active proteases.
- Conversely, SP encoding proteases that cannot cleave the PA-RNAP yield non-infectious progeny phage





## MAGE: the principle

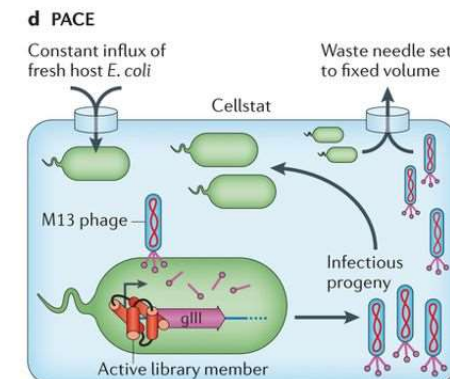
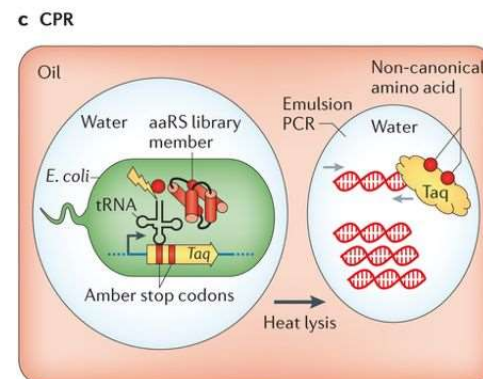
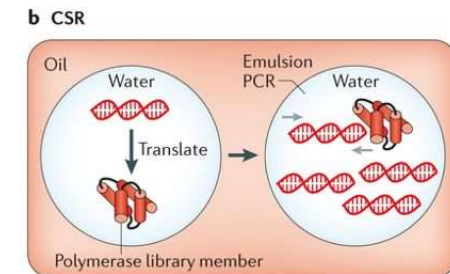
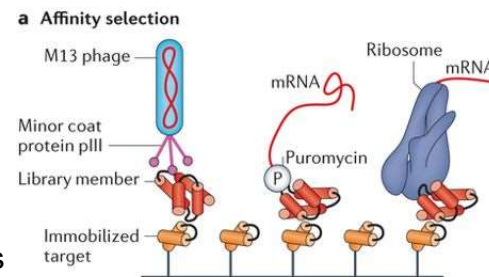
---

The technique of MAGE is practically begun by growing the cells at 30 °C until the cell density reaches mid-log phase. In this process, the  $\lambda$ -red proteins are expressed under the control of pL promoter which is regulated by temperature sensitive  $\lambda$ CI. The  $\lambda$ CI represses the expression of  $\lambda$ -red proteins. The cells are transferred to 42 °C for 15 min for heat shock induction where  $\lambda$ CI is inactivated due to high temperature and pL promoter expresses the  $\lambda$ -red proteins. The Exo protein degrades dsDNA in the 5'–3' direction. The Beta protein binds to ssDNA and generates recombination while Gam plays a key role in binding to the RecBCD protein complex and subsequently preventing this complex from binding to dsDNA ends. It also helps to induce the high efficiency of ssDNA recombination. The cells are then moved to 4 °C to repress  $\lambda$ -red and prevent degradation, and then washed with chilled distilled water thrice. A pool of single stranded oligos is introduced into cells via electroporation and these oligos become incorporated into the lagging strand of the replication fork during DNA replication.

Growth medium is added to the culture, which is then transferred to 30 °C for 2–3 h for the recovery of cells with different sequence diversity. MAGE cycling should be repeated many times as required by the experimental design. Each cell of the generated heterogeneous population contains a different set of mutations. There are a number of applications of MAGE which facilitates the rapid and continuous generation of a diverse set of genetic changes including for example, mismatches, insertions, and deletions. The oligo mediated allelic replacement is the capability of introducing a number of genetic changes at high efficiency. With MAGE, up to 30 bp mismatch mutations and insertion could be possible while up to the 45 kbp may be chromosomally sequences deleted and two-state hybridization free energy  $\Delta G$  between oligo and the targeted complement region in the genome was also predicted. It indicates that a pool of oligos with generate sequences have lesser frequency of incorporation than highly homologous sequences. MAGE also provides a highly efficient, inexpensive and automated solution to concurrently modify many genomic locations across different length scales from the nucleotide to genome level.

# Selection strategies

- Affinity selection identifies library members that bind to an immobilized target. Methods for covalently linking proteins with their corresponding genes during selection include display on phage particles via protein fusion to the coat protein pIII (left), covalent attachment to their encoding mRNA transcript via a puromycin linkage (middle) and the non-covalent attachment of both mRNA and nascent polypeptide to stalled ribosomes (right). **b** | Compartmentalized self-replication (CSR) selects for DNA and RNA polymerases that can amplify, by PCR, their own genes within water emulsion droplets (blue circle) isolated from one another by an oil phase (brown rectangle). **c** | In compartmentalized partnered replication (CPR), the evolving activity must trigger expression of Taq polymerase. For example, aminoacyl tRNA synthetase (aaRS) activity promotes amber stop codon suppression, leading to the expression of full-length Taq polymerase. Individual *Escherichia coli* cells are then isolated in water–oil emulsion droplets and lysed by heat. Higher Taq expression leads to better PCR amplification of the active library members. **d** | During phage-assisted continuous evolution (PACE), host *E. coli* cells continuously dilute an evolving population of ~1010 filamentous bacteriophages in a fixed-volume vessel (cell stat; blue rectangle). Phage encoding active variants trigger host cell expression of the missing phage protein (pIII) in proportion to the desired activity and consequently produce infectious progeny, whereas phage with inactive variants produce progeny that are not infectious and are diluted out of the vessel.



## Screening protocols

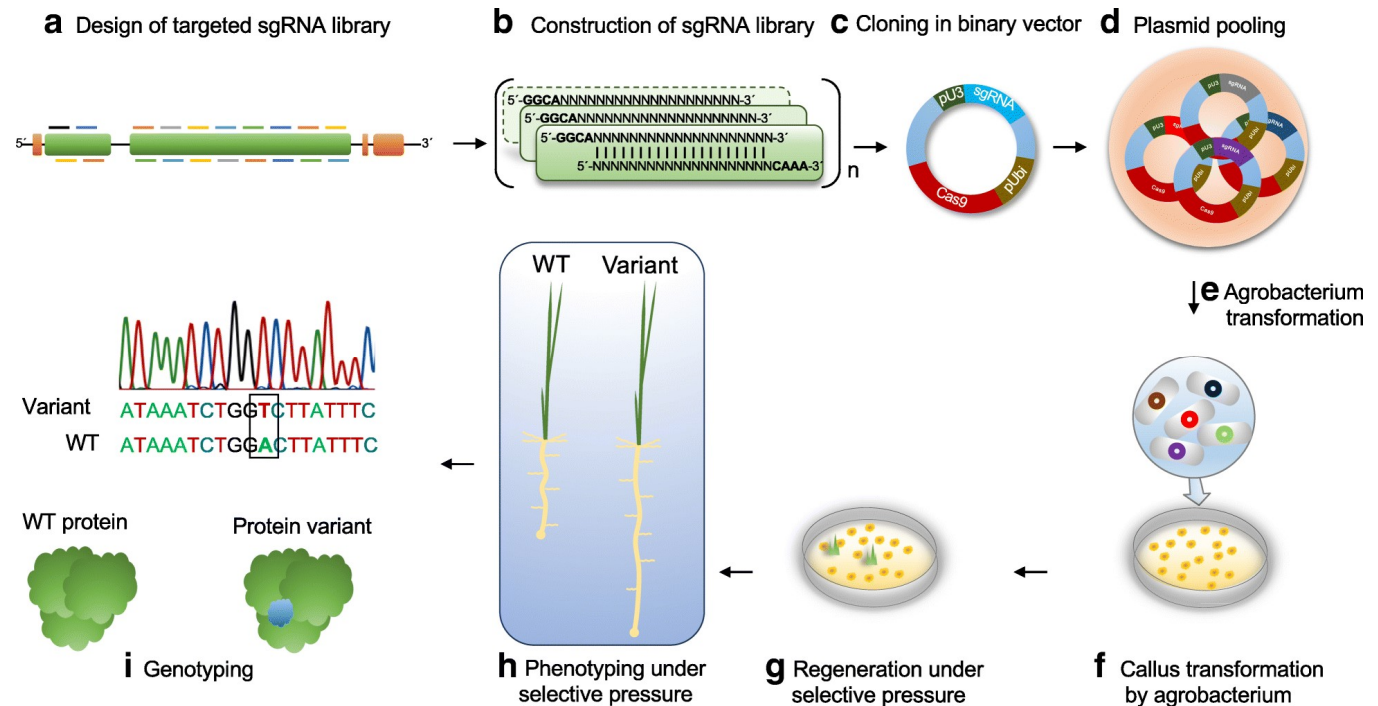
---

- In screening protocols, individual clones are examined in small liquid cultures or on Petri dishes to determine the properties of an enzyme that they produce. Their efficiency depends on the concentration and specific activity of the catalyst, the sensitivity of the test, the rate of the corresponding uncatalysed reaction, and possible interference with other enzymes. The term selection is reserved for situations where only clones producing a certain level of activity survive. *In vivo* selection protocols are the simplest but require the expression of the gene to confer a significant biological advantage. *In vitro* selection techniques are based on a physical link between a protein and its encoding gene – the best known system is phage display. In principle, selection processes are more efficient because they can handle larger libraries (frequently  $\geq 10^7$  clones), whereas screening protocols are better suited for libraries of a few-thousand clones. In both strategies, however, success depends on the quality of the designed protocols.

# CRISPR/Cas-directed evolution (CDE) platform



- All possible sgRNAs targeting the whole coding sequence of a gene are designed
- The sgRNA library is constructed via oligo synthesis and annealing
- The annealed oligos are cloned with sgRNA scaffold in the binary vector
- The sequences are confirmed by Sanger sequencing.
- All the plasmids are pooled and transformed into *Agrobacterium*
- The *Agrobacterium* cells are washed from plates with transformation medium and used for callus transformation
- After two consecutive selections on hygromycin, the callus is regenerated under selection pressure (e.g., inhibitor)
- The resistant seedlings are recovered and the resistant plants are further analyzed by exhaustive phenotyping under selection pressure
- The plants are genotyped by amplicon sequencing, and protein variants are identified



# Examples of computationally designed enzymes

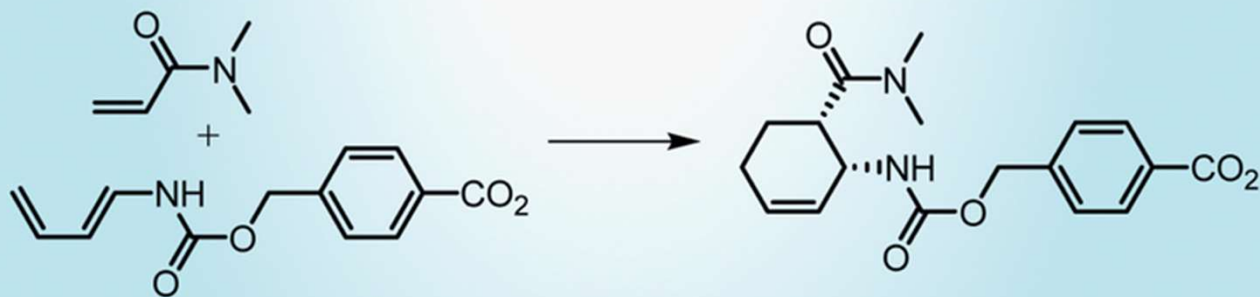
## Kemp Elimination Enzyme



## Retro-Aldolase



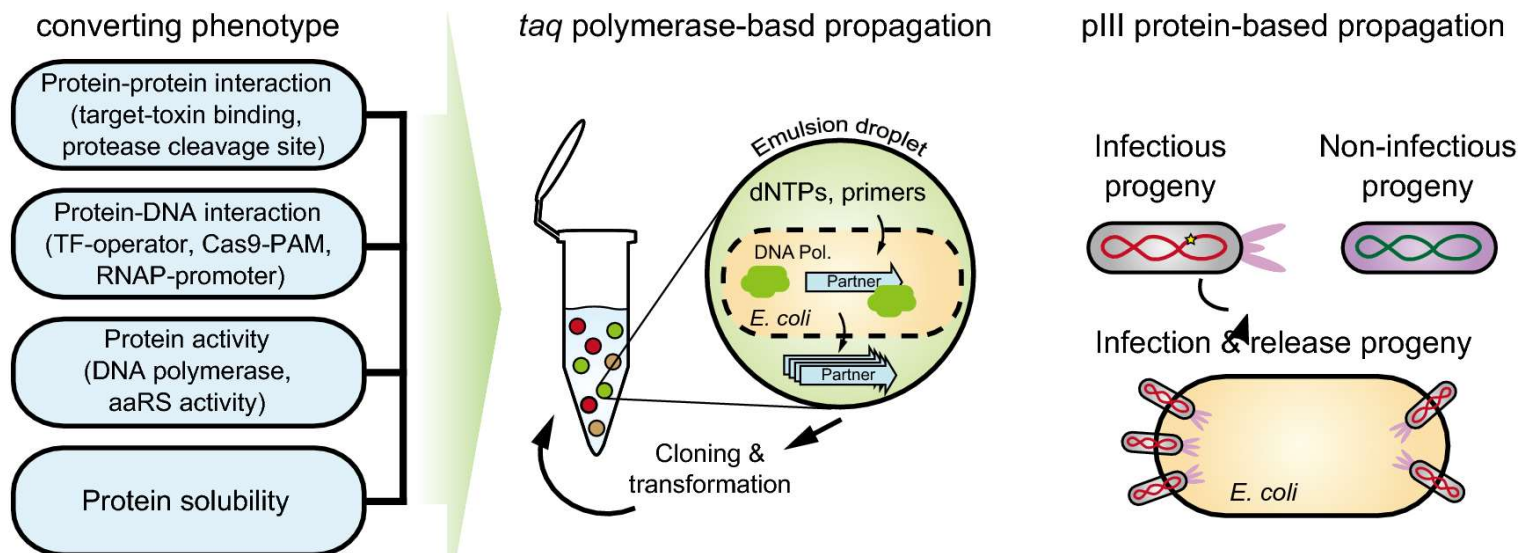
## Diels-Alderase



# Phage-assisted continuous evolution (PACE)

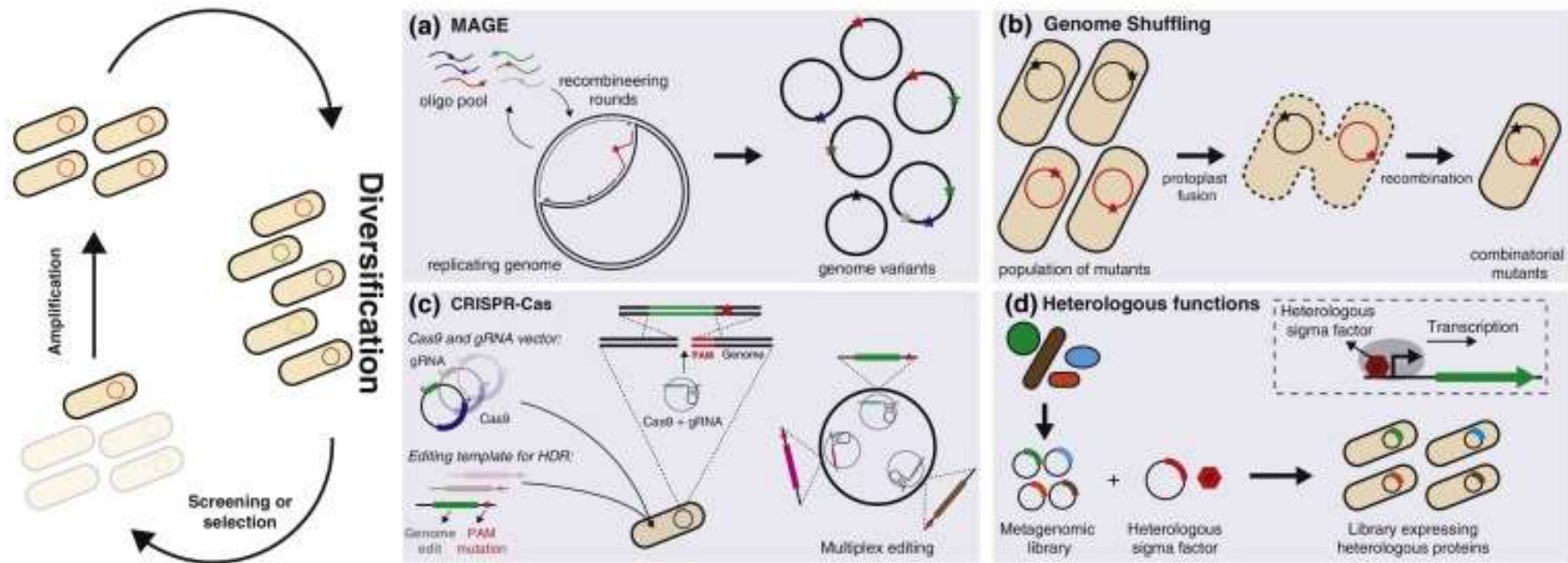
DNAs harboring beneficial mutations are propagated by compartmentalization technique and PACE. Genetic cassettes convert diverse protein functions such as protein–protein binding, protein–DNA interaction, protein specific activity, and protein solubility to changes in the expression levels of *taq*-polymerase or of phage infection protein. In the *taq*-polymerase case, the amount of amplified target product depends on *taq*-polymerase content which is expressed in a cell. Each cell containing a plasmid carrying target DNA and *taq*-polymerase is encapsulated together with PCR mixture (PCR buffer, dNTPs, primers). During emulsion PCR, cells are disrupted and expose plasmid as template and expressed *taq*-polymerase. In M13 protein III (pIII) case, expression of pIII is regulated. When beneficial mutations are occurred and increasing pIII expression, phage carrying these mutations can generate more progeny

## Biomolecule enrichment



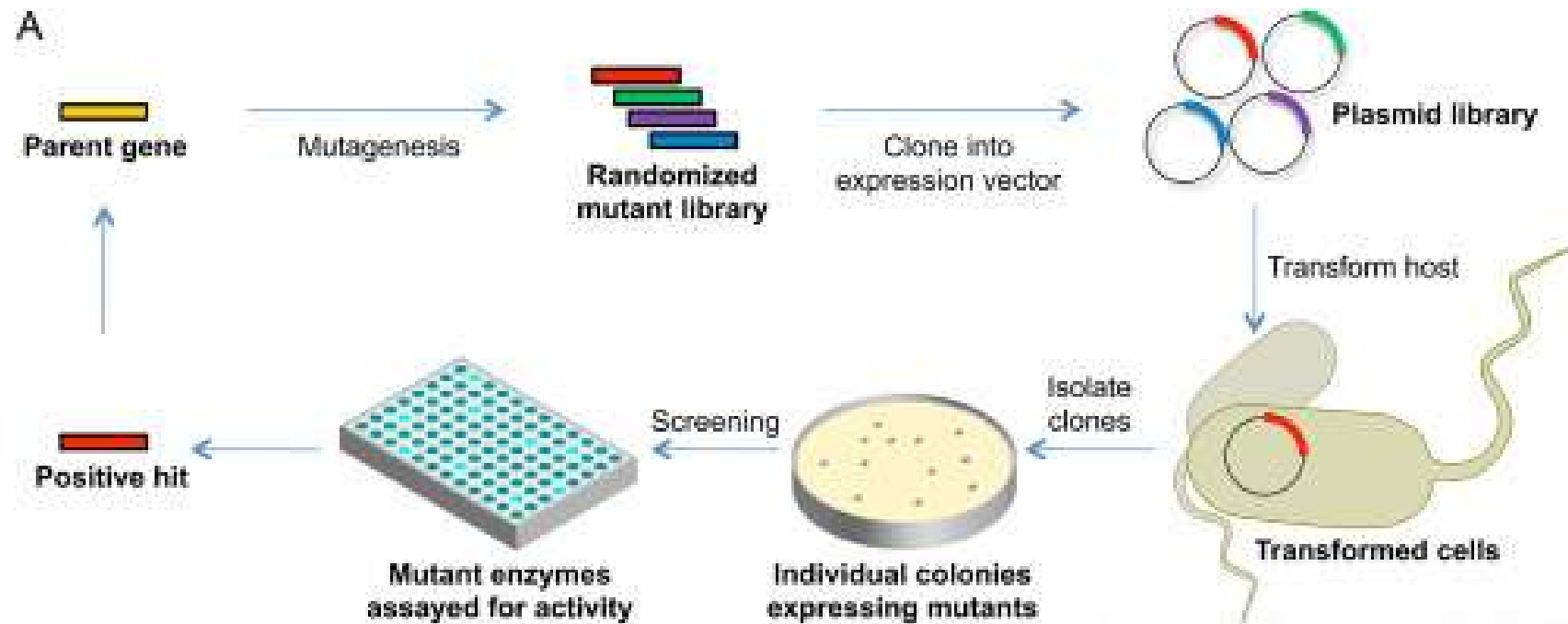


# Continuous directed evolution techniques

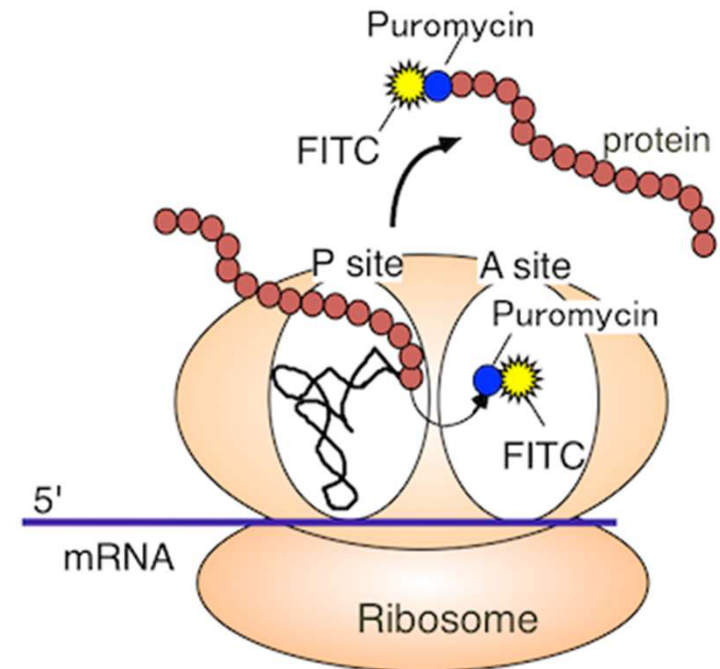
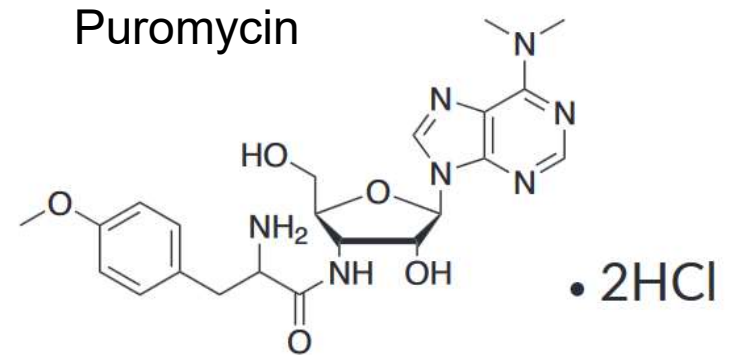


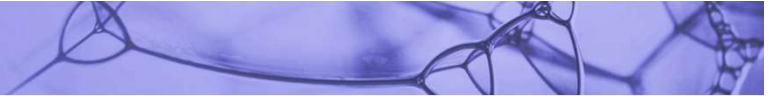
Current Opinion in Biotechnology

## Cell surface display systems



## Protein C-terminal end labelling





## Protein synthesis using unnatural elements

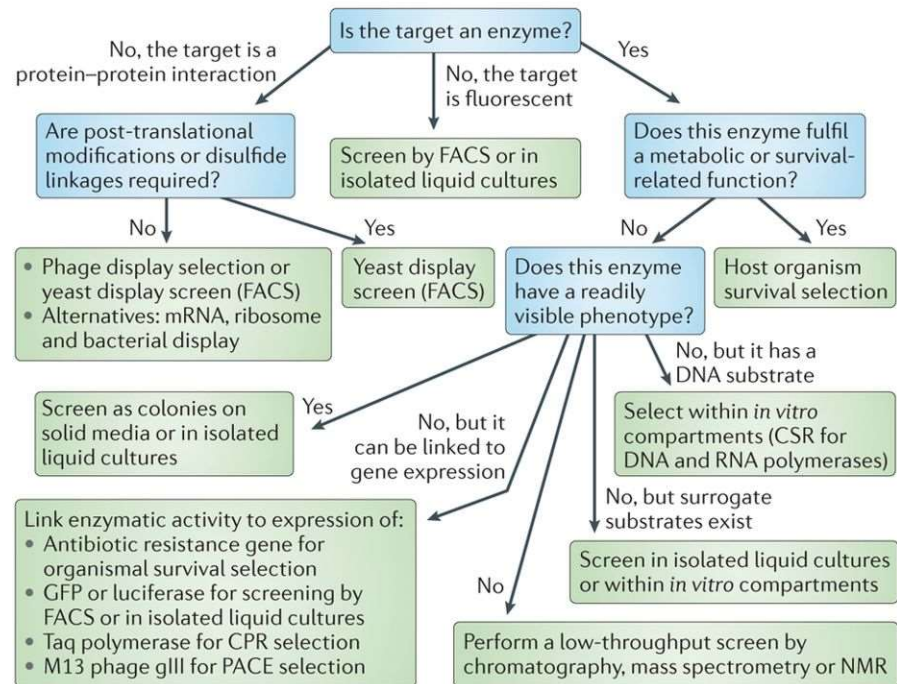
---

- The *in vitro* translation can also be done in a PURE (protein synthesis using recombinant elements) system. PURE system is an *E. Coli* cell-free translation system in which only essential translation components are present. Some components, such as amino acids and aminoacyl-tRNA synthases (AARSs) can be omitted from the system. Instead, chemically acylated tRNA can be added into the PURE system. It has been shown that some unnatural amino acids, such as N-methyl-amino acid acylated tRNA can be incorporated into peptides or mRNA-polypeptide fusions in a PURE system

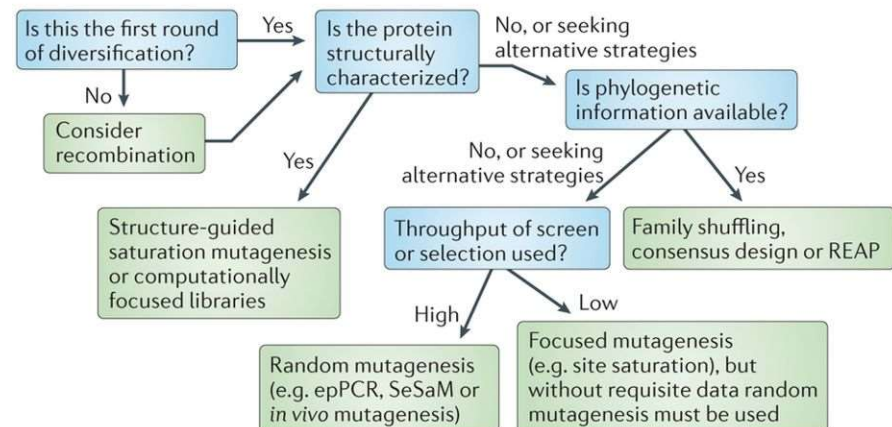
## The choice of screening strategy

- The choice of a screening or selection method can be depicted as a decision tree that operates primarily on the properties of the protein and phenotype to be evolved. Although many techniques can be extended to alternative phenotypes, this figure focuses on the most popular methods for each set of conditions. **b** | Diversification strategies must be chosen both at the outset of an evolution project and between rounds of screening or selection. Considerations can and should change over the course of a project due to the phenotypes and genotypes within the evolving population. This decision tree attempts to distil these considerations with an emphasis on focused mutagenesis methods that have the maximum potential to identify functional variants. epPCR, error-prone PCR; CPR, compartmentalized partnered replication; CSR, compartmentalized self-replication; FACS, fluorescence-activated cell sorting; gIII, gene III; NMR, nuclear magnetic resonance; PACE, phage-assisted continuous evolution; REAP, reconstructed evolutionary adaptive path; SeSaM, sequence saturation mutagenesis.

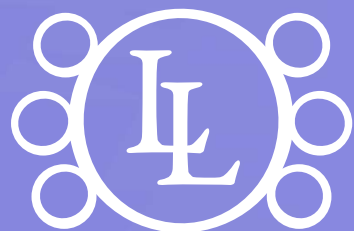
### a Choosing a screening or selection method



### b Choosing a library diversification strategy



Dr. Martin Marek  
Loschmidt Laboratories  
Faculty of Science, MUNI  
Kamenice 5, bld. A13, room 332  
[martin.marek@recetox.muni.cz](mailto:martin.marek@recetox.muni.cz)



**LOSCHMIDT**  
**LABORATORIES**

