# Microdroplets

## A Tool for Protein Engineering

**Directed evolution has arguably superseded protein design as a method to engineer proteins catalysts or binders. Practically directed evolution requires a link between genotype (the coding nucleic acid, DNA or RNA) and phenotype (the observable functional trait of the corresponding protein variant that can be selected). Griffiths and Tawfik have introduced water-in-oil emulsion droplets to achieve this [1,2]. Such microdroplets with diameters between 1 and 100 µm and volumes ranging from femto- to nanolitres are generated either "in bulk" (using a homogenizer, stirrer or extruder) or using a microfluidic setup (fig. 1).**

Miriam Kaltenbach, Yolanda Schaerli and Florian Hollfelder from the Department of Biochemistry at the University of Cambridge

To obtain "monoclonal" droplets in which one gene and protein are unambiguously linked, the gene library is diluted so that each droplet contains no more than one copy of DNA. Genes are transcribed and translated in cells or *in vitro* and subsequently improved for binding are selected via a procedure tailored to their characteristic trait.

Advantages of this technology lie in several areas. First, working in droplets drastically reduces reaction volumes, making this methodology economical and inherently high-throughput. To illustrate this point: One milliliter of emulsion typically contains $10^{10}$ droplets and therefore potentially just as many reactions which take place in parallel. Second, experiments can be performed entirely *in vitro*, which overcomes the limitations of traditional directed evolution experiments in cells such as *E. coli* or yeast. Most important, library sizes are no longer restricted to $<10^9$ by low transformation efficiencies and therefore a higher genetic diversity can be probed.

Third, it is possible to select not only for binding, but also for catalysis. The compartmentalization makes tethering of the substrate to the DNA obsolete; meaning that substrates can be present free in solution and true catalysts can be selected. In contrast, using other *in vitro* systems such as phage or ribosome display for catalysis involves indirect selections for traits such as binding of transition state analogues or inhibitors or of intramolecular, single turnover reactions in which the substrate is attached to the protein-nucleic acid entity [3].

A range of different formats has been developed for droplet-based directed evolution – they are grouped here according to the nature of the selection process.

### Selection by Affinity Panning

Figure 2A illustrates how protein binders can be identified by affinity: following expression, gene and gene product are linked covalently or non-covalently. Subsequently, emulsions can be broken and proteins can be captured by binding to immobilized ligands.

The nucleic acid-protein linkage has been realized in several ways:

(i) Between streptavidin-fusion proteins and biotinylated DNA in Streptavidin – Biotin Linkage in Emulsions (STABLE) [4]. (ii) In M.*Hae*III display, genes code for DNA methyltransferase M.*Hae*III in fusion with the protein of interest and contain a suicide substrate for the methyltransferase (the nucleotide 5-fluoro-C) forming a covalent bond [5]. (iii) SNAP display is based on human $O^6$-alkylguanine DNA alkyltransferase (AGT), also known as the SNAP tag. This DNA repair protein is responsible for repairing alkylated guanine residues by covalently attaching the alkyl moiety to a specific cysteine residue in its active site. In the droplet, the substrate analogue benzylguanine (BG) is attached to the DNA, which enables proteins expressed as AGT fusions to become covalently linked to it [6].

This suite of new methods is gaining popularity and model selections with enrichments between 10- and 150-fold [5,6] as well as actual selection experiments have been reported [4,7].

### Selection of DNA-modifying Enzymes: Gene Equals Substrate

For the selection of DNA-modifying enzymes, the gene can serve as substrate for the protein it encodes as illustrated by the following examples.

(i) In compartmentalized self-replication (CSR), a library of polymerase genes is cloned and expressed in *E.coli*. Single cells are then emulsified with nucleoside triphosphates and appropriate primers

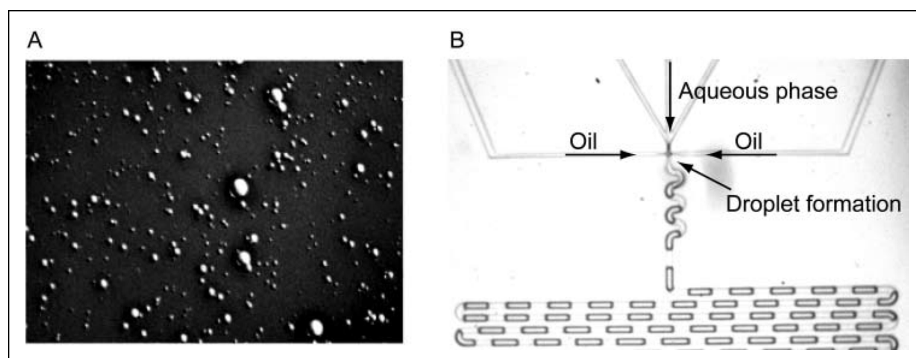


Fig. 1: A. Polydisperse bulk emulsion droplets generated by stirring. Reprinted with permission from Macmillan Publishers [1]. B. Monodisperse droplets generated in a microfluidic device.

and PCR is performed in the microdroplets. Gene and polymerase are released into the droplet by rupture of the cell during the initial denaturation step and self-replication takes place: Active variants replicate their encoding gene, whereas inactive ones cannot produce offspring. Similarly, more active variants produce more offspring. Therefore, the enrichment of a certain variant is directly coupled to its selective advantage (fig. 2B). CSR has been used to evolve polymerases with relaxed substrate specificities capable of processing damaged DNA or unnatural nucleotides [8].

(ii) In the original method developed by Tawfik and Griffiths, the gene is strictly speaking not equal to the substrate, but extended to include the target recognition sequence of DNA methyltransferase *Hae*III. Only if the gene produces an active methyltransferase it will be methylated and protected during subsequent treatment with *Hae*III [2,9].

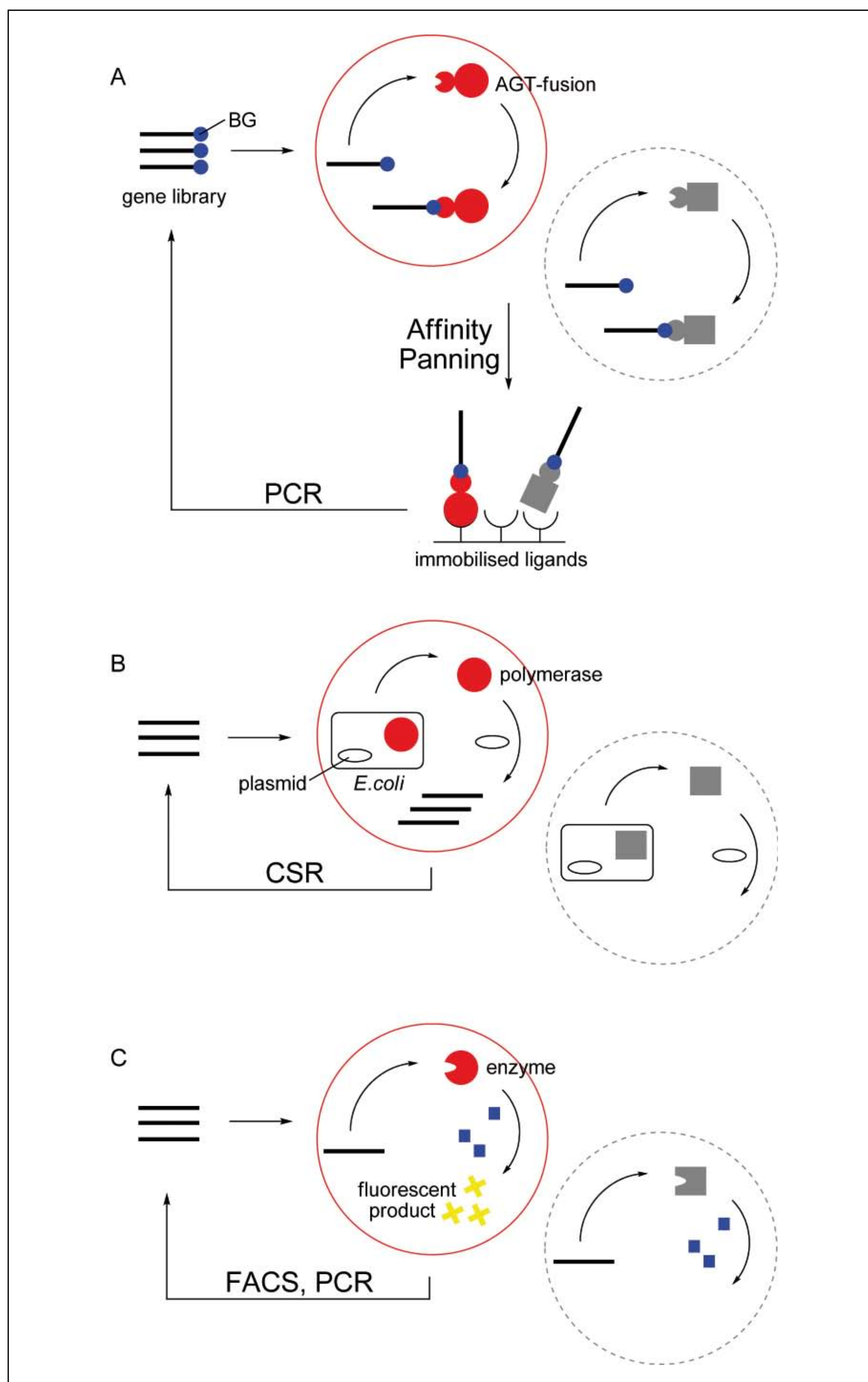


**Fig. 2: Representative directed evolution experiments in microdroplets. A. SNAP display as an example for selections by affinity panning: A gene library is emulsified and *in vitro* expressed AGT fusion proteins covalently attach to their coding DNA via the benzylguanine (BG) tag. Binding proteins are immobilized on a solid support and their DNA amplified via PCR, whereas non-binding proteins are washed away. B. Compartmentalized self-replication. A library of polymerase genes is cloned and expressed in *E. coli*. Single cells are compartmentalized in the presence of nucleotides and primers. During PCR, they release polymerase and plasmid. Only active variants replicate their genes, which then enter into the next selection round. C. Fluorescence-based sorting of double emulsions: A library is transcribed and translated *in vitro* in microdroplets. If the expressed enzyme is active, it generates a fluorescent product that can be sorted by FACS. The corresponding DNA is amplified before the next selection round.**

## Fluorescence-based Selection

A large range of reactions that generate fluorescent product (directly or via a coupled reaction) becomes accessible by fluorescence-assisted sorting (FACS & FADS) [10] of droplets with a throughput of $10^7$ events per hour.

(i) In microbead display, DNA, protein and fluorescent enzymatic product are immobilized on a microbead inside the droplet. Subsequently, the emulsions are broken and the beads sorted [11]. Microbeads can be washed and re-emulsified under different conditions, making this system amenable to perform sequential reactions, e.g. adding an emulsion PCR step before protein expression [12].

(ii) It is also possible to sort the entire microdroplet, which makes immobilization of reactants and products unnecessary (fig. 2C). Direct flow cytometry of the droplet-containing oil phase is incompatible with commercial FACS machines. Therefore the droplets need to be re-emulsified to give water-in-oil-in-water emulsions. The fluorescent signal for any particular enzyme variant is proportional to the number of product molecules formed. This readout of catalytic efficiency enables the discrimination between variants, if their activity is significantly different. In addition, the substrate is presented free in solution. Therefore, selections are for all aspects of catalysis (substrate recognition, product formation, rate acceleration and multiple turnover), as has been shown for several hydrolases [3,13].

## Microfluidics

A downside of conducting experiments in bulk emulsions is that sizes of droplets and therefore compartment volumes and concentrations vary: obtaining quantitative read-outs is clearly difficult. An attractive alternative is the generation of highly monodisperse droplets in microfluidic devices. In addition to controlled droplet formation, their processing is possible on line: from their fusion, split-

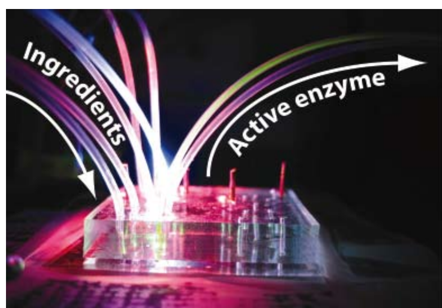


Fig. 3: Future chip-based evolution machine? (Picture taken by Luis Miguel Fidalgo)

ting and incubation to their analysis and sorting. So far, no complete directed evolution experiment has been performed in microfluidic droplets. However, the individual steps have been demonstrated, suggesting that protein engineering in this format will eventually become a reality:

Compartmentalisation of single cells [14,15], kinetic measurements of enzymes expressed by cells [15,16] as well as fluorescence activated sorting of droplets containing active mutants (FADS [10]) on chip have all been demonstrated. Unlike cell colony screening assays on agar plates not only endpoints, but also kinetics can be measured and fluorophores do not need to remain associated with the cell as into FACS assays.

Likewise, the individual steps for in vitro based directed evolution experiments using microfluidic droplets have been reported: *In vitro* transcription and translation from single DNA copies is possible [17], kinetic parameters for compartmentalized enzymes can be determined [18] and the current challenge is to integrate these processes with fluorescence activated sorting of droplets.

Currently, droplets can be formed at rates up to 10 kHz in a microfluidic device. This throughput is three orders of magnitude lower than in bulk emulsion, although multiplexing may narrow this gap. Despite lower throughput, advantages are expected: Quantitative measurements will enable the kinetic characterization of whole libraries. This sets the stage for tracking the progress of evolution for an entire system, e.g. giving rise to protein fitness landscapes. In addition, the ability to build in additional droplet manipulation modules allows multi-step processes: these include fusion of droplets to initiate reactions or delivering the *in vitro* transcription/translation system to droplets in which DNA has been amplified beforehand by PCR [19]. Moreover, the microfluidic setup holds the promise of automation. Once the challenge of integrating the different droplet manipulation modules is solved, microfluidic droplet platforms may become future chip-based evolution machines (fig. 3).

### References

A comprehensive review of the work discussed in this article can be found in:

[20] Schaerli Y. and Hollfelder F.: The potential of microfluidic water-in-oil droplets in experimental biology. Mol. Biosyst. 2009, available online (DOI:10.1039/b907578j)

A list of references [1–19] can be obtained from the authors.

► www.eMagazineBIOforum.com

**Authors:**
Miriam Kaltenbach, Yolanda Schaerli and Dr. Florian Hollfelder

CONTACT:
**Dr. Florian Hollfelder**
Department of Biochemistry
University of Cambridge
Cambridge, United Kingdom
fhsec@cam.ac.uk