UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES

Laboratoire de Biologie Chimique
Institut de Science et d'Ingénierie Supramoléculaire (ISIS)

THÈSE présentée par :

Alexei Godina

soutenue le : **01 février 2013**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**
Discipline/ Spécialité : Chimie

In vivo and *In vitro* Directed Evolution of Enzymes using Droplet-based Microfluidics

THÈSE dirigée par :
Prof. A. D. Griffiths — Institut de Science et d'ingénierie Supramoléculaire(ISIS), Université de Strasbourg

RAPPORTEURS :
Prof. U. T. Bornscheuer — Institute of Biochemistry, Greifswald University, Germany
Prof. P. Tabeling — Microfluidique-MEMS-Nanostructures Laboratory, ESPCI Paris, France

---

AUTRES MEMBRES DU JURY :
Prof. N. Winssinger — Department of Organic Chemistry, University of Geneva, Switzerland

UNIVERSITÉ DE STRASBOURG

Thesis

In vivo and *In vitro* Directed Evolution of Enzymes using Droplet-based Microfluidics

Alexei Godina

Laboratoire de Biologie Chimique

Institut de Science et d'Ingénierie Supramoléculaire (ISIS)

Université de Strasbourg, France

Thèse

Présentée à l'Université de Strasbourg
Ecole Doctorale des Sciences Chimiques

Pour obtenir le grade de

Docteur de l'Université de Strasbourg

Composition du jury

Directeur de thèse (Supervisor) Prof. A. D. Griffiths

Institut de Science et d'Ingénierie Supramoléculaire(ISIS),
Université de Strasbourg, France

Rapporteurs (Referees) Prof. U. T. Bornscheuer

Institute of Biochemistry,
Greifswald University, Germany

Prof. P. Tabeling

Microfluidique-MEMS-Nanostructures Laboratory
ESPCI Paris, France

Examinateur (Examiner) Prof. N. Winssinger

Department of Organic Chemistry,
University of Geneva, Switzerland

Acknowledgement

This dissertation would not be possible without the valuable assistance of many people and therefore I am thankful to all of them who contributed and guided me during the completion of this study.

First and foremost, I would like to express my highest gratitude to my thesis supervisor Prof. Andrew D. Griffiths for providing me with the opportunity to carry out my doctoral work in his laboratory. I deeply appreciate his patient guidance, enthusiastic encouragement and invaluable scientific assistance throughout my research.

I am grateful for the receipt of the Marie Curie Research Training Network "ENEFP" fellowship (European Network on directed Evolution of Functional Proteins). I enjoyed connecting with other network members and sharing experiences and new ideas with them. I greatly appreciate the generous funding making possible to meet and interact with so many people on international conferences and other meetings.

It is a pleasure for me to thank Dr. Michaël Ryckelynck, whose insightful advice on biology and microfluidics created a strong grounding for my development. His willingness to give his time so generously has been very much appreciated.

I would like to offer my special thanks to Richard Obexer, from the laboratory of Prof. Donald Hilvert (ETH, Zurich), whom I had the opportunity to work with on the retro-aldolase project. His high motivation had a crucial impact on the progress of the project and I hope for a continuation of our fruitful collaboration.

I wish to acknowledge Dr. Linas Mazutis for being a model PhD student for me and for the help provided for the protease screening project. Within the same project I am indebted to people from the protein engineering department of Novozymes (Denmark) for the very important experience I earned during our cooperation.

My thanks are extended to Isabelle Borsenberger for being always receptive and very efficient in solving any bureaucratic issues a non-European Union citizen like me had to struggle with. Her help left me much more time for science.

Many thanks for technical support, scientific and casual discussions go to all former and current LBC members with whom I daily enjoyed working closely. Also, special thanks to the helpers who read and corrected parts of the manuscript: Michaël, Rebecca, Clément, Elodie.

It is an honor for me that jury members: Prof. Uwe Bornscheuer, Prof. Nicolas Winssinger and Prof. Patrick Tabeling took time out of their busy schedules and agreed to participate in the evaluation of this doctoral work.

I wish to thank my old friends from Cauşeni and Москва as well as new friends I met in Strasbourg for making my student life delightful.

Last but not less important than anyone else here, I would like to thank my family, my parents and my brothers. Although they were far away from me, I could always feel their deep support for everything I do.

Abstract

In the course of billions of years, natural evolution has resulted in emergency of complex organisms from simple molecules. The understanding of the principles driving this fascinating process in Nature has allowed improvement of biomolecules through directed evolution in the laboratory. One of the main conditions for biological evolution was the compartmentalization of replicating entities. This process of compartmentalization should have taken place early in the emergence of life and yield in the development of actual cells. In analogy to cells, man-made micrometer-sized droplets can assure the genotype-phenotype linkage by compartmentalization in the laboratory conditions. In this regard, droplet-based microfluidics is a powerful tool for high-throughput generation and various manipulations of monodisperse cell-like droplets populations allowing *in vitro* selections of specific phenotypes.

The research performed within this doctoral thesis focuses on development and application of droplet-based microfluidic systems for directed evolution of enzymes. Each microfluidic platform is tailored to biocatalysts particularities and the properties to be evolved. For the improvement of a *de novo* designed enzyme retro-aldolase, it was possible to integrate several droplet-based microfluidic modules functionally with each other on a single chip, to create a novel complex device, allowing the screening of improved enzyme variants in one step. The screening of industrial proteases required the development of microfluidic platform with a droplet fusion step, resulting in a precise selection of protease activity and tolerance to detergent. Using the platforms above it was possible to run 1-3 high-throughput directed evolution cycles after which improved clones were isolated. Finally, a universal completely *in vitro* microfluidic system was developed for fundamental molecular evolution studies on *Streptomyces griseus* aminopeptidase.

Table of contents

RESUME DE THESE EN FRANÇAIS **1**

INTRODUCTION **7**

I. Enzyme engineering **7**

1. General introduction 7
2. *De novo* design 9
3. Computer-assisted enzyme redesign 10
4. Directed evolution 11

II. Droplet-based microfluidics **21**

1. Numbers in droplet fluidics 21
2. Droplet generation 23
3. Microfluidic droplet manipulations 31
4. Platforms for biological applications 47

THESIS SUBJECT **55**

I. Materials and methods (general) **57**

1. Microfluidic device fabrication 57
2. Microfluidic device operation 58
3. Fluorinated oils and surfactants 59
4. Optical setup and detection 60

DIRECTED EVOLUTION OF DE NOVO DESIGNED RETRO-ALDOLASE **63**

I. Materials and methods **66**

1. Synthesis of retro-aldolase fluorogenic substrate 66
2. 96-multiwell plate screening assay 67
3. Protein production and purification 68
4. Kinetic analysis in cuvettes 69
5. Library generation and cloning 69
6. Growth of E coli for microfluidic assay 70
7. Plasmid recovery 70
8. The exchange of retro-aldolase reaction product between droplets 70
9. Kinetics in droplets 72
10. Microfluidic assay with off-chip incubation 73

II. Results and discussions **75**

1. Fluorogenic substrates for retro-aldolase assay in droplets 75
2. Retro-aldolase activity and kinetics in droplets 82
3. Validation of full RA directed evolution cycle 86

4. Development of integrated microfluidic device 89
5. Retro-aldolase library screening 93
6. Conclusions and perspectives 99

HIGH-THROUGHPUT SCREENING OF PROTEASES UNDER DEFINED CONDITIONS 101

I. Materials and methods 104

1. Cell inoculation before encapsulation 104
2. Cell encapsulation and growth in droplets 104
3. Fusion with assay components 105
4. FADS for improved activities 106
5. DNA recovery 107
6. DNA transformation in cells 109

II. Results and discussion 110

1. Optimization of *B. subtilis* growth 110
2. Protease activity in droplets 112
3. Passive fusion with assay components 114
4. Protease kinetics 118
5. Model experiment in final screening conditions 120
6. Protease library screening 122
7. Conclusions and perspectives 127

COMPLETELY *IN VITRO* MICROFLUIDIC PLATFORM FOR DIRECTED EVOLUTION OF *STREPTOMYCES GRISEUS* AMINOPEPTIDASE 129

I. Materials and methods 133

1. Substrates 133
2. SGAP activity assay in bulk 133
3. In vitro transcription-translation in bulk 133

II. Results and discussions 134

1. Fluorogenic substrates for SGAP activities 134
2. Optimal IVTT kit for SGAP expression 135
3. Protein expression in droplets 137
4. DNA amplification in droplets 140
5. How many fusion steps? 143
6. Model experiment 147
7. Conclusions and perspectives 153

GENERAL DISCUSSION 154

BIBLIOGRAPHY 162

List of abbreviations

3D three–dimensional

AC alternating current

APD avalanche photodiode detector

BSA bovine serum albumin

Ca Capillary number

CCD charged-coupled device

CMC critical micelle concentration

CV coefficient of variance

Da Dalton

DAQ data acquisition

DC direct current

DEP dielectrophoresis

DMSO dimethyl sulfoxide

DNA deoxyribonucleic acid

E. coli Escherichia coli

EDTA ethylenediaminetetraacetic acid

FACS fluorescent activated cell sorter

FADS dielectrophoresis-based fluorescence activated droplet sorter

FFJ flow focusing junction

GFP green fluorescent protein

HRCA hyperbranched rolling circle amplification

IPTG isopropyl-β-D-1-thiogalactopyranoside

ITO indium tin oxide

IVC *in vitro* compartmentalization

IVTT *in vitro* transcription/translation

k_{cat} catalytic constant

K_M Michaelis-Menten constant

LB Luria Broth

MALDI matrix–assisted laser desorption/ionization

MD molecular dynamics

MS mass spectrometry

OD optical density

PAA poly(acrylic acid)

PBS phosphate buffered saline

PCR polymerase chain reaction

PDMS poly(dimethylsiloxane)

Pe Péclet number

PEG poly(ethylene)glycol

PET poly(ethylene)

PMT photomultiplier tube

PFPE perfluorinated polyether

RA retro-aldolase

RCA rolling circle amplification

Re Reynolds number

RFU relative fluorescence unit

RNA ribonucleic acid

RNase endoribonuclease

rpm rotation per minute

SAM self–assembled monolayer

SGAP *Streptomyces griseus* aminopeptidase

SELEX systematic evolution of ligands by exponential enrichment

tRNA transfer RNA

UV ultraviolet

QM/MM quantum mechanics/molecular mechanics

w/o/w water-in-oil-in-water

wt wildtype

Résumé de thèse en français

L'ingénierie des protéines fonctionnelles est un processus d'amélioration des propriétés physiques ou catalytiques d'enzyme au travers d'approches rationnelles et d'évolution dirigée. Actuellement, le design rationnel, l'évolution dirigée, aussi bien que la combinaison des deux méthodes sont largement utilisés pour optimiser et modifier les caractéristiques désirées des protéines. Malgré le progrès de la modélisation moléculaire des protéines, les méthodes de prédiction restent aléatoires et un grand nombre de variantes restent à tester. De ce fait, le développement et l'utilisation d'un système de criblage d'activité de protéines à très haut débit, comme la microfluidique en gouttes, est indispensable. Le principe de cette technique repose l'utilisation de puces PDMS/verre dans lesquelles sont gravés des réseaux de canaux microscopiques pour la manipulation de liquides. Il est ainsi possible de créer et manipuler de façon rapide et précise des microgouttes de phase aqueuse dispersées dans une phase d'huile fluorée et stabilisées par ajout d'un tensio-actif. Chaque goutte, de quelques picolitres à quelques nanolitres, est un microréacteur indépendant permettant la compartimentation *in vitro* de l'information génétique (l'ADN) et phénotypique (protéines) de manière analogue aux cellules, *in vivo*. Il est possible de fusionner les gouttes avec d'autres gouttes pour en modifier le contenu, de les diviser, de les incuber en lignes de délai ou encore de les trier. Cette flexibilité d'action sur les gouttes et le faible volume de ces dernières, permet la réalisation d'opérations complexes tout en augmentant la sensibilité de système de détection. La microfluidique en gouttes a déjà démontré son utilité pour l'évolution dirigée de protéines telles que la péroxydase [1] et d'identifier des variants très efficace dont l'activité est uniquement limitée par la diffusion du substrat. Pendant ce criblage 10^8 de variants ont été analysés en 10 h avec une réduction de coût d'un million fois en comparaison des criblages standard en microplaques.

En dépit des quelques exemples d'applications réussies en évolution dirigée, le développement et l'utilisation des plateformes microfluidique n'est pas trivial et leur conception est influencée par les particularités des protéines étudiées ainsi que du type d'évolutions recherché. Cette thèse de doctorat présente trois projets d'évolution dirigée de protéines en trois approches différentes avec expression d'enzyme *in vitro* et *in vivo*.

Les plateformes microfluidiques ont été développées et validées pour chaque projet. De plus, plusieurs banques de variants ont été criblées avec, dans certains cas, isolement de molécules 5-10 fois que le clone parental.

L'évolution dirigée de retro-aldolase

Les enzymes naturelles sont des catalyseurs écologiques très efficaces. Cependant, beaucoup de réactions synthétiques importantes manquent de biocatalyseurs. Le design d'enzymes possédant de nouvelles activités catalytiques a un grand intérêt pratique [2] avec des applications potentielles en biotechnologie, biomédecine et pour les processus industriels. De plus, la création *de novo* de protéines fonctionnelle peut approfondir la compréhension du mécanisme de biocatalyse et d'évolution moléculaire d'enzymes.

Ce projet de recherche, dirigé en coopération avec l'Ecole Polytechnique de Zurich (ETH Zurich), a comme but l'évolution d'enzyme retroaldolase (RA) artificiellement modelé. Comme le criblage en microplaque n'a permit d'améliorer que de dix fois l'activité enzymatique en huit étapes de mutation/sélection, il a été décidé d'utiliser des systèmes de criblages microfluidique à très haut débit.

Une série d'aldols et d'aldéhydes complémentaires a été synthétisée afin de palier le manque de substrat fluorogène permettant le suivi de l'activité rétroaldolase en gouttes. Les constantes catalytiques et la rétention des substrats dans les gouttes ont été testées. Le couple substrat/produit fluorescent optimal a été utilisé pour suivre l'activité retro-aldolase en émulsion. La vitesse de clivage du substrat similaire en goutte et en microplaque, a démontré l'absence d'effet significatif de l'interface des gouttes sur l'activité enzymatique. Pour simplifier les manipulations microfluidiques un système d'expression de protéine dans le cytoplasme cellulaire a été choisit. Les étapes microfuidiques ont été limitées à une co-encapsulation (d'une part les cellules dans le milieu de culture, d'autre part le substrat avec la solution de lyse), suivi par une incubation de l'émulsion 1 heure hors puce, réinjection, analyse et tri de gouttes les plus fluorescentes.

Enfin, pour démontrée que le système est applicable au criblage de banques de retro-aldolase, trois souches bactériennes produisant des variantes de RA ayant différents niveaux d'activité ont été mélangés. Le variant le plus actif a pu être récupéré

et isolé à l'aide du système microfluidique développé. La diversité génétique a été générée à partir du gène initial par mutagenèse aléatoire et par mutagenèse à saturation, mais également en utilisant des méthodes de recombinaison *in vitro* par *DNA shuffling*. Après chaque étape de mutation/sélection une augmentation de l'activité totale des banques a été observé. Afin de différencier finement les variants les plus actifs il a était nécessaire de passer d'une incubation en dehors de la puce à une incubation sur puce. Pour ce faire, une puce microfluidique intégrée unique permettant la production de gouttes, leur incubation puis leur tri a été développée. L'intégration d'une ligne de délai entre le générateur de gouttes et le trieur permet l'incubation des gouttes pour un laps de temps précis sur la puce avant leur sélection. De plus, il est possible d'ajuster la taille de cette ligne de délai afin d'obtenir le temps d'incubation désiré. Dans notre cas nous avons utilisé des lignes de délai permettant une incubation de 45, 30, 14 et 7 minutes.

Après trois cycles de sélection et une caractérisation sur micro-plaque, nous avons pu isoler des clones ayants des activités 5-10 fois supérieure.

Il est prévu de réaliser plusieurs cycles de sélection/mutation additionnels, suivis d'une caractérisation des mutants les plus performants.

Criblage à très haut débit de protéases

Les protéases comme les lipases, amylases, et cellulases sont largement utilisées dans l'industrie des détergents, surtout pour la production de lessives biologiques. Ces enzymes décomposent rapidement ou libèrent les souillures qui nécessitent normalement des conditions drastiques pour être nettoyées.

L'objectif de ce projet est d'évolué des variants plus actif de protéases dans des conditions définies (détergent, pH, etc). Ce criblage de protéases à très haut débit s'inscrit dans une collaboration avec un partenaire industriel.

L'évolution dirigée *in vivo* utilisant des souches bactériennes pour la production de protéines simplifie les manipulations microfluidiques, parce que l'amplification d'ADN et la production des enzymes encodées sont réalisées naturellement par la cellule. Les étapes proposées pour l'évolution de protéases sont comme suit: la banque de gènes transformée dans une souche de *Bacillus subtilis*, puis chaque variant est

encapsulé en gouttes de 3 picolitres dans des conditions où chaque goutte ne contient au plus qu'une seule bactérie. Pendant l'incubation les bactéries se divisent et secrètent des protéases dans la goutte. Les gouttes sont fusionnées passivement avec des gouttes contenant un substrat fluorogène, un détergent et une solution tampon au pH établi. Après un court intervalle de temps nécessaire à la conversion du substrat l'émulsion est analysée et triée. A partir de l'émulsion collectée, l'ADN correspondant aux protéases les plus actives est récupéré et utilisé pour de nouvelles étapes de sélections et mutations.

La fusion passive est une technique microfluidique, qui permet la coalescence sélective de deux gouttelettes, l'une (stabilisée avec un tensioactif) provenant d'une émulsion réinjectée et l'autre (partiellement stabilisée) produite sur la puce. La présence de détergent était nécessaire au test enzymatique. Ce détergent ayant un effet stabilisateur, la fusion passive n'était pas réalisable dans les conditions standard [3]. Après modification de la géométrie de canaux microfluidiques, réduction de la quantité de tensioactif et changement de type de l'huile, la fusion passive a pu être adaptée pour l'évolution de ces protéases.

Pour la mise au point définitive de toutes les étapes de criblage, une sélection modèle a pu être réalisée avec succès, démontrant la possibilité d'enrichir spécifiquement le variant actif, d'un mélange actif/inactif. Quelques banques de gènes fournies par le partenaire industriel ont été criblées. Dans le processus de sélection le top 1-3% des clones les plus actifs a été récupéré. Pour chaque banque le processus de sélection a été répété 2 à 3 fois jusqu'à ce qu'une population plus active soit bien définie sur les histogrammes de distribution d'activité. Une deuxième génération d'évolution a été obtenue par mutation de l'ADN codant pour les protéases les plus actives de la première génération. Le résultat final devrait être présenté si le partenaire industriel en donne l'autorisation.

Evolution dirigée *in vitro* d'aminopeptidase *Streptomyces griseus*

L'aminopeptidase de *Streptomyces griseus* (SGAP) est une protéine fonctionnelle, multi-spécifique qui possède plusieurs activités enzymatiques et peut être considérée comme un intermédiaire évolutif de protéines portant des activités distinctes. De plus, sa petite taille et sa thermostabilité, font de SGAP un modèle attractif pour l'étude de l'évolution moléculaire en microfluidiques de gouttes. L'activité enzymatique principale

de SGAP est le clivage d'acides aminés N-terminaux de peptides. D'autre part, SGAP possède une activité de phosphodiesterase utilisant un mécanisme distinct de celui de l'activité aminopeptidique. Enfin, lorsque les ions zinc du site actif sont substitués par des ions cuivre, SGAP présente une catécholoxidase.

Une méthodologie ainsi qu'une plateforme de microfluidique en gouttes (création, fusion et tri de gouttes) ont été développées en vue de l'évolution dirigée de SGAP. Un système complètement *in vitro* a tout d'abord été employé pour amplifier les gènes de chaque variants d'une banque individualisés en microgouttes. Celles-ci ont ensuite été fusionnées par électro-coalescence avec des gouttes contenant les éléments nécessaires à la transcription-traduction *in vitro* permettant l'expression de populations clonale. Parmi la grande variété des kits commerciaux permettant de réaliser l'expression de protéines par transcription et traduction *in vitro* (IVTT), seul deux ne présentaient pas d'activité aminopeptidase bruit de fond. Les composants de ces systèmes IVTT reconstituent les enzymes et les facteurs de transcription et traduction, purifiées séparément [4]. Comme SGAP, dans sa conformation secondaire, forme un pont disulfure, le puresystem S-S a été choisit car est facilite la formation de ces liaison.

D'autre part, la synthèse et la mise au point des substrats fluorogènes adaptés au dosage d'activité enzymatique en gouttelettes ont été réalisées. Ainsi, après la pico-injection de substrat et une courte incubation, les gouttelettes ont été triées en fonction de l'activité enzymatique du variant exprimé dans la goutte. Les étapes de fusion et de pico-injection sont obligatoires du fait de l'incompatibilité des processus d'amplification de l'ADN avec l'expression d'enzyme. De même, la transcription-traduction doit être séparée du dosage enzymatique. Toutes les étapes sont validées avec une expérience modèle à partir d'un mélange de gènes codants pour la SGAP active ou inactive. La plateforme développée pour l'évolution dirigée de SGAP peut facilement être adaptée au criblage *in vitro* d'autres enzymes en modifiant uniquement le substrat.

Introduction

I. Enzyme engineering

1. General introduction

Enzymes are extremely proficient catalysts, which naturally evolved to catalyse a wide spectrum of chemical reactions. At the same time, enzymes also display a high degree of regio- and stereoselectivity, which results in limited substrate scope. These functional proteins govern almost all chemical transformations in living organisms. Because of their high catalytic proficiency and mild reaction conditions, enzymes are broadly used as industrial, medicinal and laboratory tools in the pharmaceutical [5], detergent [6], biofuel [7] and food industries [8], as well as diagnostics [9] and molecular biology research. The study of protein evolution pathways and specialization may address fundamental questions related to evolution of life.

Despite enzyme's natural proficiency there is still a huge unexplored potential to enhance or modify their characteristics. Protein engineering approaches are aiming to change, improve or even acquire new desired enzyme properties, thus protein design it is possible to augment the native catalytic function, thermostability or to increase the protein expression levels. Furthermore, the enzymes can be adapted to new reaction conditions or tailored to secondary or new activities (Figure 1).

The strategies for designing or enhancing enzyme properties in protein engineering are:

- *De novo* design
- Rational redesign
- Directed evolution
- All possible combinations among all three general methods

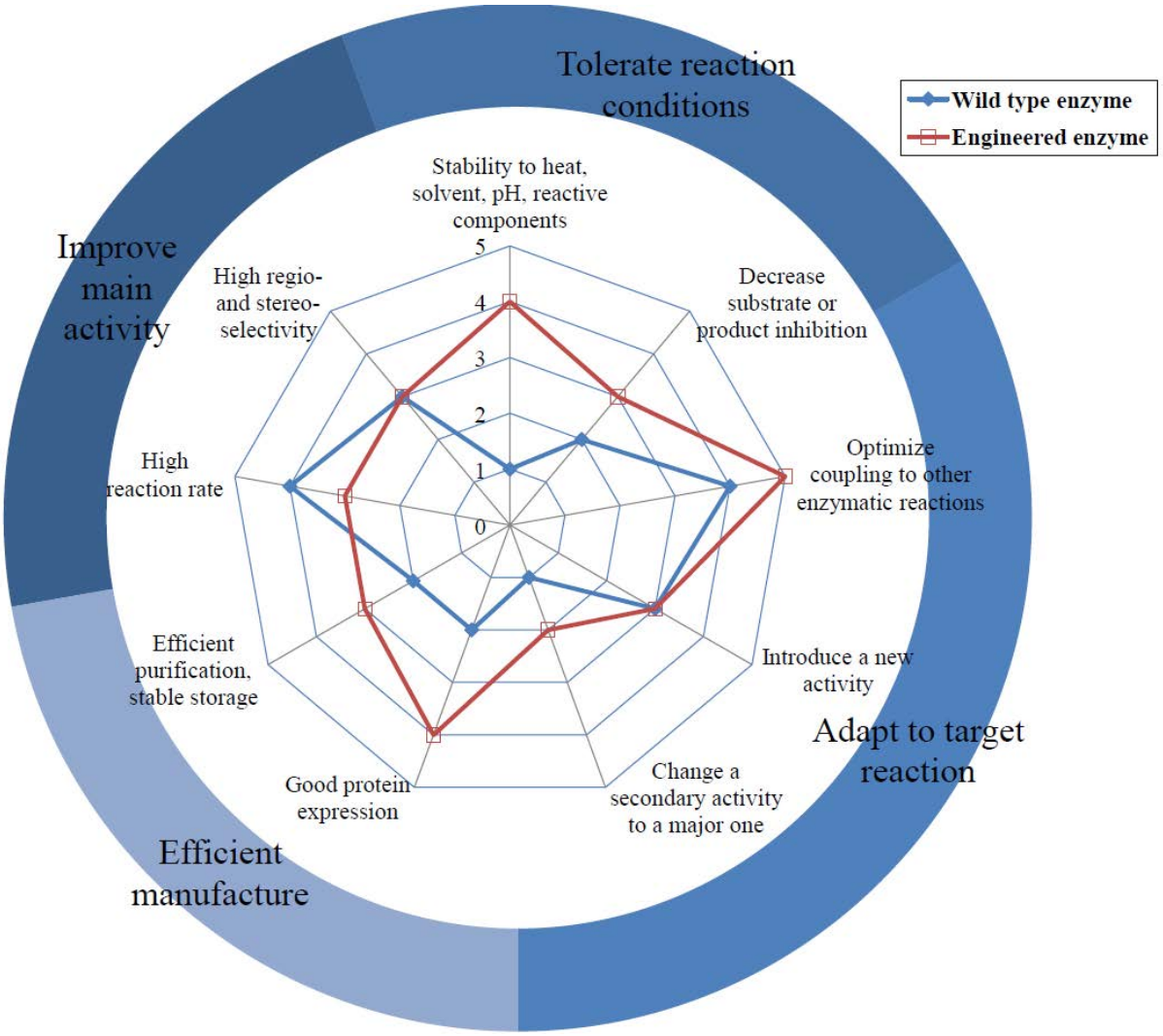

Figure 1. Protein engineering criteria. *From 0 to 5 the levels of property improvement. Adapted from [10].*

2. *De novo* design

The design of the effective biological catalysts for any desired chemical reaction is one of the most promising advances in biotechnology. *De novo* protein design is a highly challenging approach, offering broad possibilities for new protein structures [11]. This strategy aims to design and to construct a synthetic protein using the information on the 3D structures of natural protein accumulates and folding rules. The output of such computational modeling is an amino acid sequence which folds into an optimized protein capable to catalyse a desired reaction. The protocol for *de novo* enzyme design (Figure 2) has been suggested by authors [12] and optimized in [13].

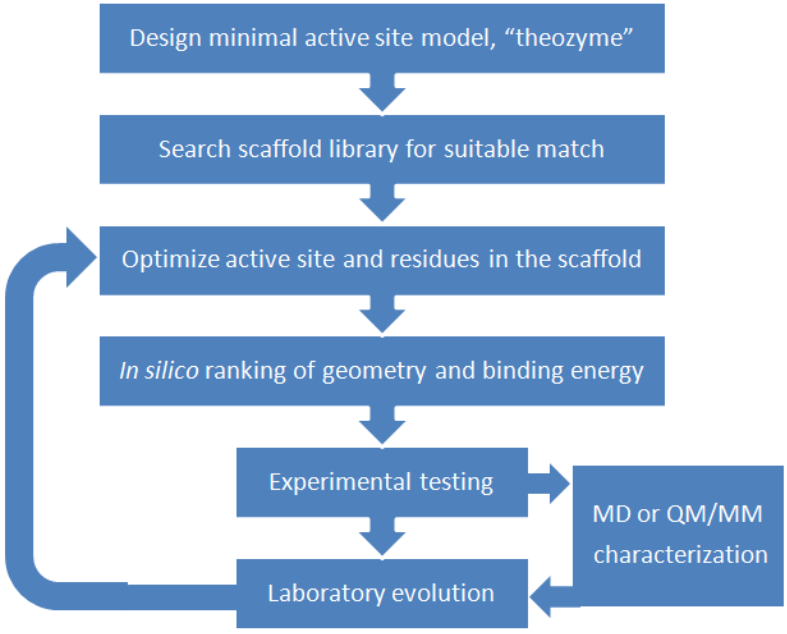

Figure 2. Protocol for *de novo* enzyme design. *Adapted from [13].*

A number of successful enzymes designed from scratch have been monitored in the literature recently: Kemp-eliminase [14], retro-aldolase [15], Diels-Alderase [16]. As any rational design method, *de novo* design requires in-depth knowledge of enzyme's (usually unavailable) catalytic mechanism, complicating the prediction of synthetic enzyme behavior. As a result, the efficiencies of *de novo* engineered biocatalysts are far too low to give a direct outcome and can only be improved to the levels of natural enzymes by consecutive laboratory evolution.

3. Computer-assisted enzyme redesign

Since nature offers such a large choice of versatile and efficient functional proteins, many existing templates can be reprogrammed to fit required properties using computational design algorithms. The variety of computational methods span from quantum mechanical cluster models containing a smaller subset of catalytic residues, QM/MM calculation taking into account the whole system and MD simulations that analyze structural changes to search for the origins of activity in designed systems, to semi-empirical approaches like EVB based on valence bond theory [17].

The majority of rational design strategies are employing primitive electrostatic models based on protein crystal structure. Theoretical enzymes are constructed by computing the optimal geometry for transition-state stabilization by functional groups. Since backbone perturbations are often neglected, computational simulations may result in many false-positive protein structures.

Along with classic computational methods, semi-rational approaches are applied as well. Sequence-based enzyme redesign [18], REAP (reconstructing evolutionary adaptive path) [19] and ProSAR (protein sequence activity relationship) [20] offer promising predictions. Here multiple sequence alignments and phylogenic analyses are helping to identify functional hotspots estimating local protein variability and leading to desired improvements. The estimations made by semi-rational techniques allow predicting a narrow genetic diversity enriched in variants displaying desired catalytic properties, also known as “small but smart libraries”, which facilitates laboratory evolution.

Although computational design methods are still lacking in precision, the growing knowledge of the enzymes’ mechanisms of action, together with the expansion of protein structure databases, seem very bright for the development of protein engineering via rational design.

4. Directed evolution

Laboratory evolution, also known as directed evolution, brought the engineering of "the third wave of biocatalysis" [21, 22]. Basically, directed evolution mimics Darwinian evolution in laboratory conditions at an accelerated rate. In the broadest sense, the directed evolution approach consists of two steps: generation of diversity and selection of desired phenotypes. Since the predictions based on protein structure-function relationships are still inaccurate, the success of the evolutionary strategy depends on fine tuning between the creation of an optimal genetic library and a suitable selection method.

a. Generation of diversity

The effectiveness of directed protein evolution depends on the quality and the size of the genetic library, defined as a mixture of DNA molecules which represent variants of the original gene. The genetic diversity should be large enough to increase the chances of finding a desirable property. At the same time, the increase in library size makes the selection more challenging [23, 24]. In this regard, the design of protein knowledge-based libraries is highly rewarding [25, 26]. In the literature there is a large spectrum of library generation methods which can be grouped into three main categories: random mutagenesis, focused mutagenesis and recombination. The combination of different mutation strategies renders evolution more efficient, moreover, when coupled with rational design approaches [27, 28].

Random mutagenesis [29] is commonly used when protein structure-function knowledge is lacking or to mimic natural evolution pathways. Even when the enzyme structure is available, random generation of diversity is applied to reveal favorable mutations in unexpected positions outside of the active site [30]. Such mutations influence protein thermostability and tolerance of reaction conditions, by inducing conformational changes in the enzyme backbone. This method introduces random point nucleotide changes in the DNA sequence: (i) transitions, which consist of substitution of a purine (pyrimidine) nucleotide with another purine (pyrimidine); (ii) transversions, where a purine (pyrimidine) is replaced by a pyrimidine (purine); (iii) deletions, where

one or more nucleotides are deleted from the sequence; (iv) insertions, where one or several nucleotides are incorporated into the gene; and (v) inversions, the 180° rotation of dsDNA segments of two base pairs or longer. These changes can be achieved by treating the DNA with mutagens [31], by passing the cloned genes through mutator strains [32], by error-prone PCR (epPCR) [33] or by error-prone rolling circle amplification (epRCA) [34].

Despite numerous advantages over other library generation techniques, epPCR only introduces a limited number of amino-acid substitutions, because this reaction biases the distribution of mutation type in favor of transitions (A>G, T>C). Multiple substitutions within a single codon are extremely rare [35], therefore, not every amino acid is accessible. To reduce the redundancy, several strategies have been developed like: doped libraries [36], where subsets of amino acids are generated that reveal only certain combinations of amino acids in a protein sequence; oligonucleotide-directed randomizations [37], where different codon systems (NNN, NNB, NNS, NNK, MAX) are employed. For the same reason classic random methods are usually combined with DNA-shuffling [38], which mimics natural gene recombination. This method is based on the mixing and concatenation of genetic material from a number of parent sequences. Recombination mutagenesis has proven to concentrate accumulated positive mutations over the rounds of directed evolution [39].

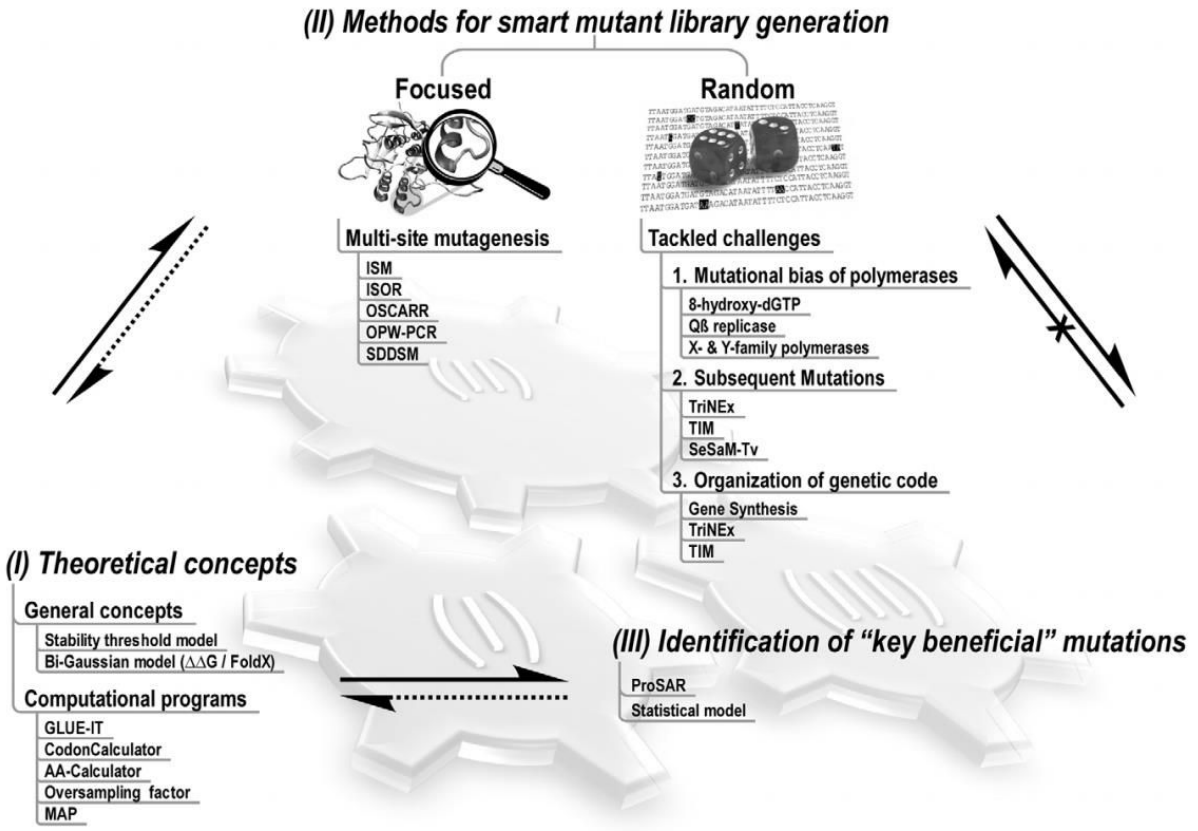

Figure 3. Methods for smart library generation. *Reprinted from [24].*

In contrast to random methods, which are arbitrarily distributing mutations, targeted mutagenesis focuses on single hotspots or selected regions of a protein found to be responsible for a desired function (Figure 3). Focused mutagenesis methods allow one to reduce the sizes of libraries simultaneously, enriching the content in positive hits. A large number of concepts for designing targeted libraries, often including computational mediated methods, has been reported and reviewed in the literature [24, 40-42]. One of the most recent examples of successful of semi-rational design is the engineering of P450 mono-oxygenase by targeting a small number of residues in the active site [43]. A systematic iterative saturation mutagenesis [44] enabled the increase in substrate specificity of Bayer-Villiger mono-oxygenase.

Since every protein behaves differently, there is no generalized strategy of diversity generation and selection in directed evolution. Many developed advances remain as a proof of principle and fail when applied to other proteins. In this context, natural evolution remains still unbeaten, "turning many protein engineers into evolutionists" [45]. The advancement in enzyme engineering depends not only on

technological breakthroughs, but also on a deep understanding of how these biomolecules evolved in nature. The reconstruction of mutational trajectories has shown the presence of dead ends in the mechanisms of protein divergence [46]. To explain this observation it is necessary to examine the effects of protein mutations [47]. A single mutation could be deleterious, neutral or advantageous for a certain protein property. However, a pleiotropic mutation favoring the gain in one activity might be accompanied by a loss in another property. This effect is also known as mutation trade-off, which usually occurs between: new function and stability, new and existing function, or new function and expression. In reality the effects of mutations are even more complex, implying the interactions between two or among several mutations. The notion of epistasis is introduced when an effect of a given mutation depends on the presence of other mutation [48, 49]. For example, a new-function mutation cannot be fixed unless another stabilizing, compensatory mutation is present, or, a compensatory mutation can be neutral on its own but beneficial in combination with a destabilizing mutation [50]. In laboratory conditions, the problem of revealing the cooperative mutations is partially overcome by DNA recombination [51]. In this regard, selecting more mutants every round increases the chances of finding the very best variant, rather than selecting only the active one [52]. Another delicate tool for protein engineering, which helps to bypass the problem of dead ends in protein evolutionary divergence, is neutral drift selection [53]. During first evolution rounds only neutral selection is applied allowing the accumulation of mutations without altering protein function or stability. The resulting gene pool contains polymorphic non adaptive variations enhancing protein evolvability and tolerance to dead ends. Afterwards, a more stringent selection pressure is applied to improve the desired property [54-57].

b. Selection and screening

The methods for screening and selection are far less developed than the strategies for diversity generation, making selection, the limiting step in directed protein evolution. While the approaches for diversification are quite general and could be applied for any gene, the selection or screening must be adapted to the enzyme and to the desired protein property that needs to be improved, at the same time following the

first rule of directed evolution "you get what you select for". A suitable selection method for enzyme evolution must fit several criteria:

- specific to the desired property
- sensitive to low enzyme activity
- robust, reproducible, simple in set-up
- high-throughput assay
- possibility to increase selection pressure

The selection approach must be highly specific to the evolvable property and capable to sense even low amounts of improvements. Obtained results should be reproducible and the method robust and easy to handle. High-throughput capacity enables the analysis of large mutant libraries, increasing the chances of finding the best variants. A simple calculation show that, for even a small protein of 100 amino acids there are 20^{100} ($\sim 10^{130}$) possible sequences, or more than the number of atoms in the universe. Thus, the development of high-throughput methods allowing the screening of billions of variants a day is very important in finding the best protein sequence-space. Another desirable characteristic is the possibility to vary the selection gate to narrow the search within every round. Before choosing a selection method it is necessary to clearly define the objectives and the size of libraries.

Since proteins cannot replicate themselves, the selection of a certain protein property should simultaneously select the gene encoding it [58]. Thus, genotype and phenotype must be linked physically in some way [59]. Cells naturally assure the compartmentalization by a system of membranes (Figure 5.a). *In vivo* systems usually use the cell apparatus and select for new activities related to the survival of the cell or production/cleavage of a fluorophore or chromophore. *In vivo* system can be applied for selection of variants that tolerate restrictive growing conditions, when the gene essential for cell survival is substituted by a complementing gene construct. One prominent example is the evolution of beta-lactamases [60], where the elevated concentration of the antibiotic acts as a restrictive condition. The selections based on production of chromophores have the drawback of low sensitivity allowing the selection only of very active variants. *In vivo* selection methods express the phenotype in different

ways reducing the reproducibility of the selection system. The use of cell lysates can overcome this problem but the throughput is quite low.

In contrast, fluorescent activated cell sorter (FACS) can perform high-throughput screening of cells. Using this method one can sort 10^8 cells a day and the method has been applied to protein evolution through cell-surface display [61-63]. Often FACS is used for selection of binding activity while the selection of biocatalysts is complicated by the difficulties of linking the DNA and the product of the enzyme reaction. There are several examples of FACS mediated selections for protein evolution: where the fluorescent product stays in cells while the substrate is washed out [64]; where the product is captured on the surface of the cell [65, 66]; or where the co-encapsulation is assured by double emulsions [67]. Nevertheless, the use of FACS for directed protein evolution is limited due to the general disadvantages imposed by the *in vivo* genotype-phenotype linkage: the inability to select under conditions different from the cellular environment, a low dynamic range, cells opposing selection pressure, low transformation efficiency and problems with selection of toxic proteins.

The major drawback of *in vivo* selection methods is the fact that only metabolically relevant reactions can be targeted, however *in vitro* systems can overcome this shortcoming. These methods enable the screening of large libraries ($>10^9$) in stringent control of selection conditions. There are several *in vitro* high-throughput screening methods based on physical genotype-phenotype linkage (Figure 5. b): phage display [69, 70], although the expression is done by cells; non-covalent linkages as ribosome display [71-73] and CIS display [74]; covalent linkage as mRNA display [75, 76]. Many of these techniques are based on cell-free expression systems [77], often commercially available extracts from *E. coli*, wheat germ or yeast. Some *in vitro* transcription/ translation (IVTT) systems have been developed using the PURE approach based on reconstitution from separately purified components [78]. In all display techniques, selection is made *in vitro* by affinity purification on the immobilized ligand. Therefore, the physical linkage has been mainly applied to selections for binding activity or single catalytic turnover activity [79], although there are exceptions [80] which have resulted in the selection of antibodies with multiple catalytic turnovers.

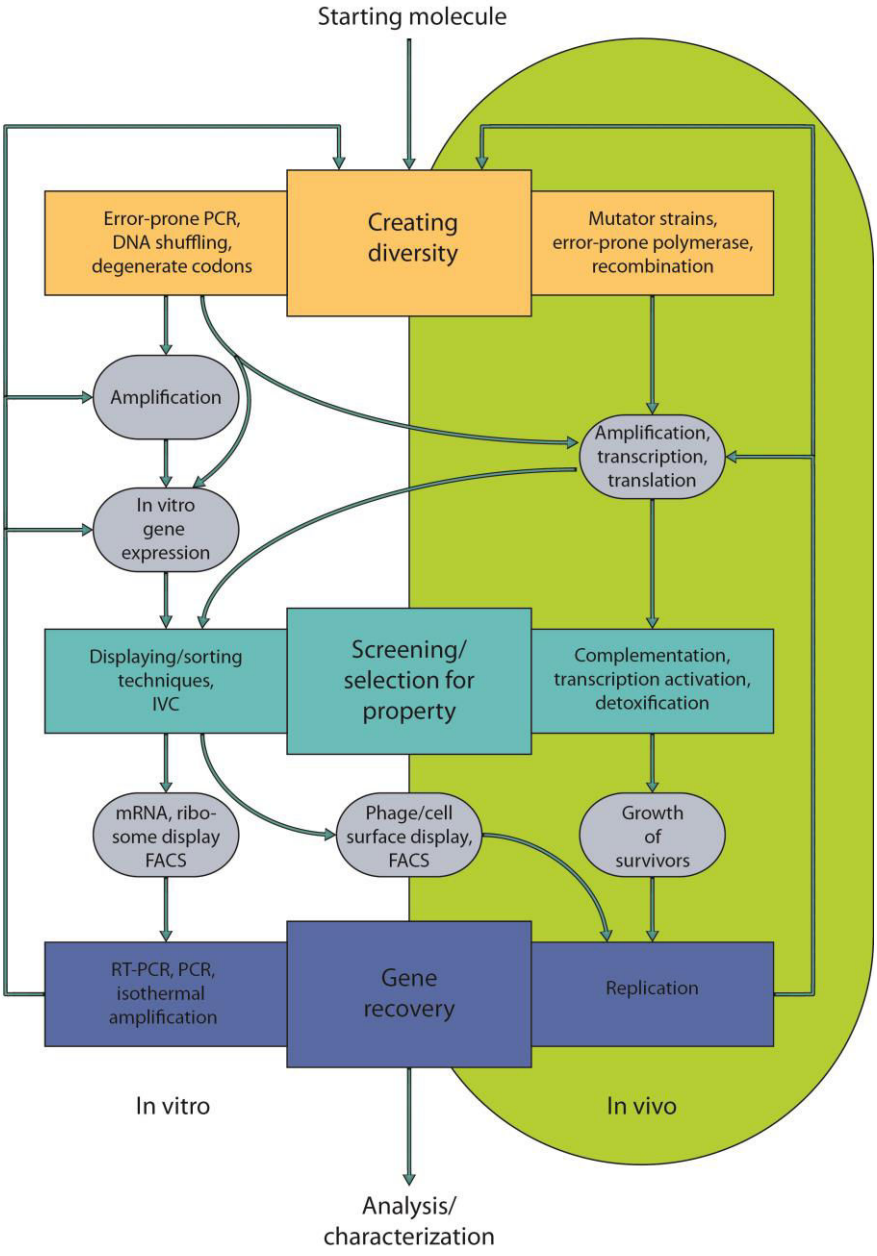

Figure 4. Principal strategies and techniques of directed evolution. *In center, Darwinian evolution principles; right part, in vivo pathways; left part, in vitro pathways. Arrows indicate the possible directed evolution routes. Adapted from [68].*

Despite a number of powerful properties, the methods based on physical genotype-phenotype linkage are limited for selection of multiple-turnovers biocatalysts, because of difficulties related to association of the protein encoding gene and the products of the enzymatic reaction [69]. *In vitro* compartmentalization (IVC) mimics natural cell compartments by generating micro-sized droplets (Figure 5. c), where the

genes together with expression and assay components are co-encapsulated to generate artificial reactors for selection of only desired properties [70]. IVC has already been successfully used to select for DNA-modifying enzymes: DNA methyltransferases [71], DNA polymerases [72, 73], restrictases [74]; or enzymes that produce substrates for the polymerase chain reaction [75].

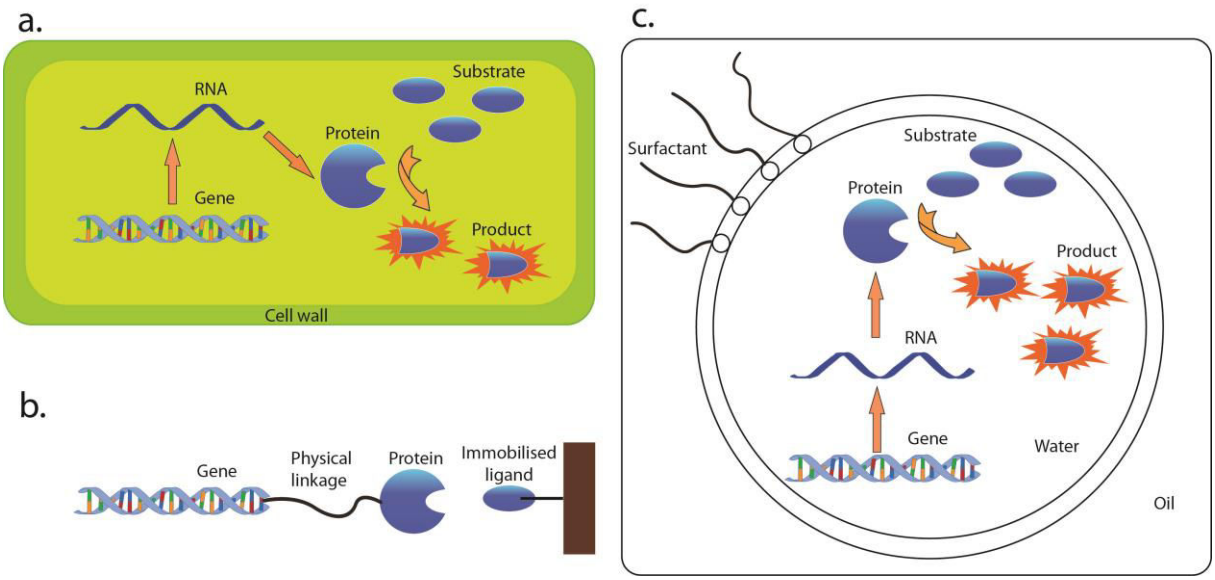

Figure 5. Genotype-phenotype linage.

*Different ways to link genes and the properties of proteins they encode. (**a**) Genotype-phenotype linkage in vivo. Cells assure the compartmentalization of genes and the signal of their expression. (**b**) Physical genotype-phenotype linkage. The gene and the product are physically linked as in phage display, ribosome display and SELEX. (**c**) Genotype-phenotype linkage by in vitro compartmentalization. The gene, the protein and the product are isolated in the same droplet compartment.*

If the droplets are used in an one-step procedure as in the above examples, selections are limited to the conditions imposed by the *in vitro* expression. To increase the spectra of conditions for IVC, protein synthesis must be uncoupled from the selection. Several strategies have been developed to keep the protein connected to DNA after expression in droplets, moving the selection into a multiple turnover scenario [76]. This includes: STABLE display, which uses the biotin-streptavidin link [77]; and microbeads display, where the DNA and expressed protein are bound to the same bead [78]. Analogically, the use of double water-in-oil-in-water emulsions coupled with FACS provides an efficient and high-throughput enzyme selection [67, 79]. Many more applications of bulk-emulsions are summarized in the review [80].

Nonetheless, IVC has weaknesses such as polydispersity of droplets, as well as limited manipulation of droplet size and content. For example, a polydisperse emulsion in a directed protein evolution experiment means that an identical gene in a different sized droplet can have a different phenotype, making it difficult to efficiently select for small differences in enzymatic activity. Also, to reduce the frequency of droplets containing more than one gene, the gene concentration must be greatly decreased, pushing down the throughput of the method. Manipulations of droplets after production might help to tune the selection for precise improvement of desired enzyme activity. Although it is possible to deliver hydrophobic substrates through the oil phase [78, 81] and water soluble components can be delivered through nanodroplets, or swollen micelles [82], and the pH can be altered, for example by the acid delivery method [78], none of these techniques is universal and highly controllable. In the end, it is impossible to measure reaction kinetics in individual droplets in bulk emulsions systems: the screens or selections are based on an end-point assay. These limitations can potentially be overcome using droplet-based microfluidic systems, in which highly monodisperse droplets can be generated at kHz frequencies, incubated, fused, split, sorted and manipulated in controllable and sophisticated ways.

II. Droplet-based microfluidics

Droplet-based microfluidics enables the manipulation of discrete fluid volumes in the form of microdroplets that provide numerous benefits for conducting biological and chemical assays. Among these benefits are the large reduction in the quantity of reagents and enhancing the speed of assays by reducing the volumes over which processes such as heating, diffusion, and convective mixing occur. Once the droplets are generated, carefully designed droplet operations allow for the multiplexing of a large number of droplets to enable high-throughput complex biological screening and selection.

1. Numbers in droplet fluidics

To have a better understanding of droplet manipulations it is worth regarding physical phenomena and dimensionless numbers behind them. Although the physics of fluidics starts with Navier-Stokes equations, for simplicity reasons only solved forms of it will be mentioned.

Common liquids like water belong to the Newtonian category where the viscosity depends only on concentration and temperature. For Newtonian liquids the Reynolds number characterizes the relative importance of inertial and viscous forces

$$Re = \frac{VR}{\nu} = \frac{\rho VR}{\mu}$$

where V is the average fluid velocity, R is a length characteristic of the geometry, μ is dynamic viscosity and ρ is density. The tendency of a flowing liquid phase to develop turbulence is determined by Re (laminar flow Re < 2000, turbulent flow Re > 2000). In microfluidics, the Reynolds number corresponds to a laminar flow regime (having values < 1, although some systems for cell handling reach *Re* = 10), thus, viscous stresses and pressure gradients dominate, while inertial and gravity forces become negligible.

The capillary number is used in droplet-based microfluidics to compare viscous forces to surface tension forces

$$Ca = \frac{\mu V}{\gamma} \sim \frac{w}{d}$$

where γ is the interfacial tension, w is cross-sectional dimension of the channel and d is diameter of the droplet. In other words, Ca is the ratio of the viscous forces to the capillary forces. Because droplet production exploits a shear-rupturing mechanism, the size of a generated droplet can be linked to the capillary number (see next section).

The Weber number (Bond number) is used to characterize the disruption of an interface under the action of strong inertial forces (gravity forces), representing the ratio of inertial forces to surface tension forces

$$We = \frac{\rho V^2 R}{\gamma}$$

$$Bo = \frac{\Delta \rho g R^2}{\gamma}$$

The surface tension maintains the droplet interface convex. And, if the inertia force or gravity force increases, the interface becomes concave, leading to droplet disruption.

The convective motions in a droplet are determined by the Marangoni effect. Even if the droplets do not move, internal motions are frequent. Convections are mainly caused by interfacial forces, especially by a gradient of interfacial tension, a gradient of interfacial temperature or a gradient of concentration on the interface. To generalize, the Marangoni number determines the strength of convective motions

$$Ma = \frac{\Delta \gamma R}{\rho \nu \alpha}$$

In laminar flows diffusion is far less involved in mixing comparing to convection. The number describing the relative rate of convection to diffusion in a flow is Péclet number

$$Pe = \frac{Ul}{D}$$

where U is average axial flow speed, l is the cross-section of the channel and D is the diffusion coefficient.

2. Droplet generation

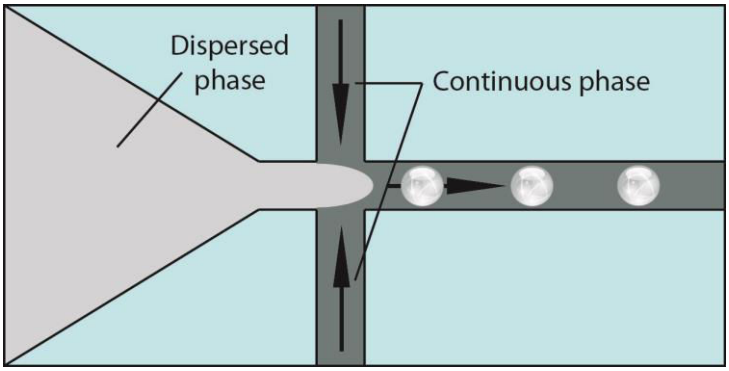

Figure 6. Droplet production in the flow-focusing junction.
Aqueous phase is dispersed into continuous oil phase.

All operations in droplet-based microfluidics start from micrometer-sized water-in-oil (w/o) droplets formation. Microfluidic droplet generators are dispersing the aqueous stream into continuous carrier oil phase (Figure 6). The coefficient of variance (CV), which represents the ratio between the standard deviation to the mean diameter of the droplet, is less than 1%, resulting in a high level of monodispersity of generated emulsion. Depending on the carrier oil and the drop size, droplets can be generated at frequencies varying from 0.001-10 kHz. The fastest droplet formation has been reported by [83] using fluorocarbon oil, reaching 30 000 droplets per second. For that reason, most of the work in the field of droplet-based microfluidics is focused around droplets in fluorocarbon oils [84, 85], which are less viscous and allow very high throughput operations in comparison with systems based on mineral or silicon oil. Moreover, fluorocarbon oil is hydrophobic and lipophobic, thus it possess low solubility for biochemical and chemical compounds soluble in water and prevents the molecular diffusion between drops. Comparing to hydrocarbon oils, fluorocarbon oils result in less swelling of PDMS used for microfluidic device preparation [86]. Finally, fluorocarbon oils have a good solubility for gases mandatory for cell viability in droplets [87-89].

Droplet formation is a complex and dynamic process that can be regarded as interplay between shear stress tending to draw the fluids along the channel and surface forces acting towards the droplet formation to minimize the surface tension [90]. The

size and the frequency of droplet generation depend on a variety of parameters: capillary number *Ca*; flow rate, viscosity and content of dispersed phase; viscosity and flow rate of the carrier oil; surfactant concentration; temperature; geometry and material of the microfluidic device. The influence of gravity and inertial forces on droplet generation can be neglected due to the low *Bo* and *We* numbers [91].

By now, a general unified theory of droplet generation does not exist, and all explanations are specific to particular cases related to the device configuration.

a. Drop production regimes and inlet geometries

Since fluorinated oils have a high interfacial tension with water, droplet production requires a high shear stress provided by the geometry of the device. A detailed understanding of parameters affecting the droplet production using different channel configurations is crucial in designing microfluidic devices for all droplet operations. In microfluidics there are two basic geometries of drop makers (Figure 8), each having a distinctive inlet channel configuration: T-junction (TJ) [90, 92, 93] and flow-focusing junction (FF) [94-97]. TJ-drop makers use two inlets and FF-drop maker use three inlets. In FF the two inlets for continuous flows connect at the junction with the dispersed flow inlet, being forced to pass together through the outlet channel. In all cases, the continuous phase applies shear force (viscous drag) on the dispersed phase while an interfacial force opposes to the elongation of the liquid neck. During this process the neck gets thinner and leads to droplet formation from the aqueous phase. A very thorough analysis of droplets formation using TJ configuration has been done by Garstecki *et al.* [92]. The proposed model fit well with experimental results for a wide range of flow rates and for different viscosities of the continuous phase and was confirmed by a series of independent experiments and numerical simulations [92, 98, 99].

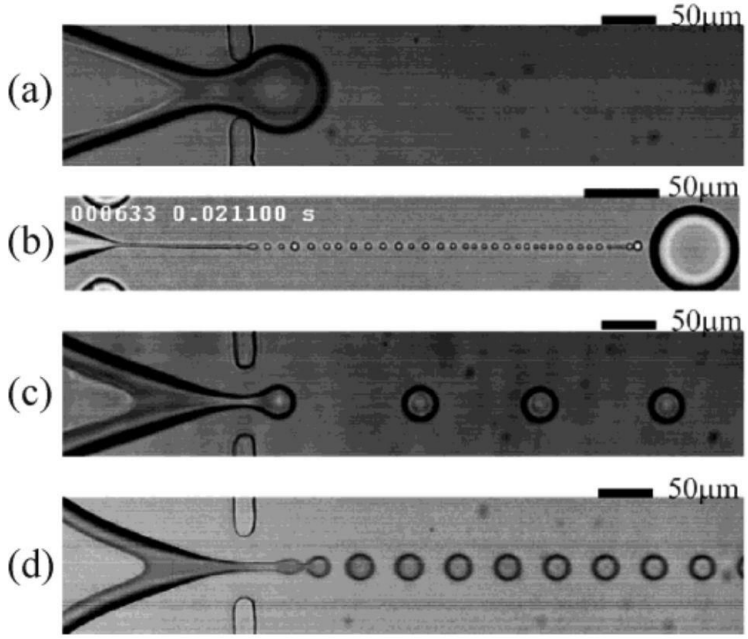

Figure 7. Different droplet breakup processes: *(a) squeezing; (b) tip-streaming; (c) dripping and (d) jetting. Reprinted from [100].*

Theoretical work, also based on the TJ and FF studies has identified three distinct mechanisms of drop formation (Figure 7): squeezing, dripping, and jetting [100-102]. The squeezing mechanism proceeds as the emerging dispersed phase obstructs the flow of the continuous phase, causing its pressure to rise; this rise in pressure, in turn, allows the continuous phase to squeeze on the dispersed phase, pinching off a drop in the process. When the width of the main channel is much greater than the width of the inlet channel (w/win > 4) droplets do not fill the cross-section of the main channel quickly enough before they pinch-off, and thus, the squeezing regime is replaced by shear stresses or the dripping regime. The dripping mechanism occurs when shear stresses overcome the interfacial tension, and drop breakup is caused by the shearing of the dispersed phase by the continuous phase. This type of droplet formation can be described predominantly by the capillary number. The transition from dripping to jetting occurs at smaller *Ca* values when the viscosity ratio of the liquids increases. This mechanism is characterized by the formation of long threads in the dispersed phase, which are broken due to the Plateau–Rayleigh instability. The jetting regime appears at very high flow rates or with low interfacial tension (i.e. high *Ca* numbers). Jetting is

generally irregular, leading to a more polydisperse size distribution. The squeezing mechanism is distinguished from the other two by the existence of large pressure fluctuations in the continuous phase. Guillot [103] demonstrated that by having a constant flow rate for the continuous phase while increasing the flow rate of the dispersed phase it is possible to pass through all three regimes of droplet break-up. So far, the dripping droplet production regime is more common because of monodispersity and the high frequency of generated droplets, although, the jetting regime has been well exploited for double-emulsion production [104] and selective encapsulation of single cells [105].

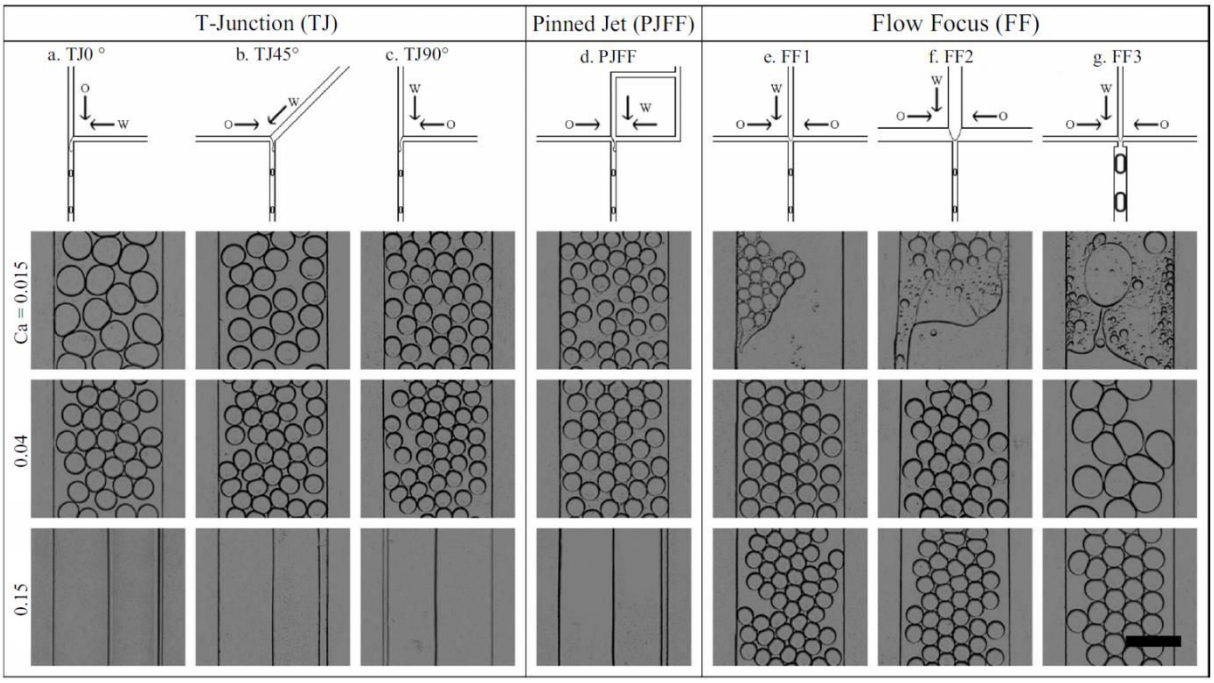

Figure 8. Impact of channel geometry on droplet production. *Schematics of drop makers with different inlet channel geometries (top row). Example images of drops formed by each device for different capillary numbers (lower rows). Reprinted from [106].*

Experimental testing of drop makers with different inlet configurations has revealed that the regimes at which droplets are monodisperse differ [106]. Here, authors have introduced several different geometries including pinned-jet flow focusing geometry (PJFF). Representative images for drops formed in all devices are shown for different capillary numbers (Figure 8). FF drop makers are producing monodisperse drops at high and moderate capillary numbers, while TJ and PJFF do so at moderate and low capillary numbers. Also, for the same conditions the droplets formed by TJ0° and

TJ45° are larger than those formed with TJ90° and PJFF, suggesting that water angle influences the droplet generation at low capillary numbers. To quantify the drop properties as a function of the flow-rate ratio (continuous phase/dispersed phase) authors [106] fixed the Ca = 0.015 for TJ and Ca=0.04 for FF. At low flow-rate ratios drop makers were producing large plug drops separated by short segments of oil. As the ratio was increased the droplets became smaller for all geometries. The further increase reduced the drop size only slightly, while most of continuous phase contributed only to droplet spacing.

b. Double emulsions

Double emulsions, which represent smaller droplets of one fluid suspended in a larger droplet of a second, immiscible fluid, are important technologies for biological screenings and selections (Figure 9). Since FACS machines operate only with aqueous solutions, the oil continuous phase should be changed to water. Double water-in-oil-in-water (w/o/w) emulsions have been already employed for high-throughput evolution experiments [78, 79]. Monodisperse double emulsion can be generated using standard TJ [107], FF [108] and PJFF [109] junctions on microfluidic device, as well as capillary devices [104]. In these axisymmetric flow-focusing glass systems, the dispersed phase is surrounded symmetrically by a continuous phase, which allows constant droplet production for extended times.

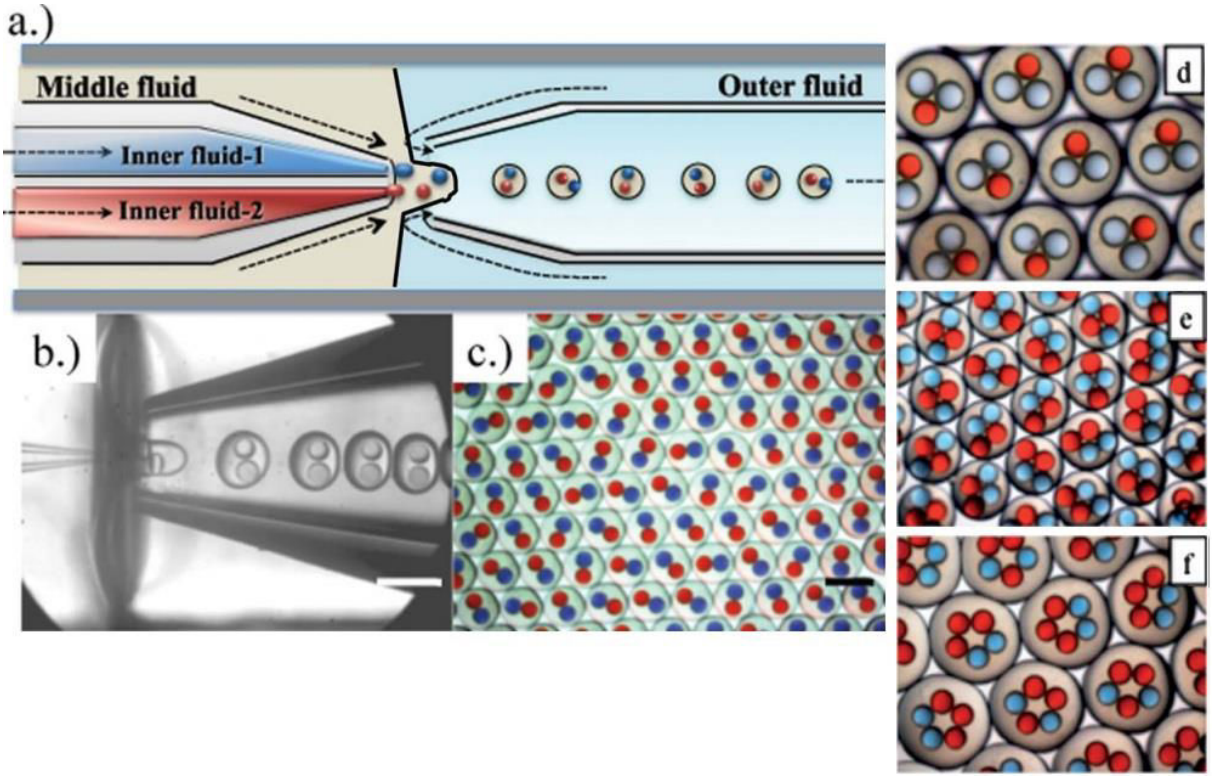

Figure 9. Multi-component double emulsions. *(a) Schematic of a microfluidic capillary device for preparation of multiple component double emulsions using a single-step emulsification. (b and c) Optical microscopy images showing double emulsion generation and monodisperse double emulsions with two different inner drops, one red and one blue. (d,e,f) Combinations of monodisperse double emulsions with two different types of inner drops. Reprinted from [110].*

Generation of multiple component double emulsions has a large potential for encapsulation of different types of drops for later release of incompatible actives or for using multi-component drops as isolated microreactors loaded with different reagents [110].

c. Droplet stability, biocompatibility and compartmentalization

A list of reports found in the literature point out that the optimal droplet size for high-throughput biological screenings and selections is in the range of 10-50 μm, which corresponds to the volume of 0.5-65 pL [105, 111, 112]. Decreasing the droplet size may increase the throughput, but production of femtoliter droplets is more difficult to achieve and precise droplet manipulation is still challenging [113]. Moreover, as the size

of the droplets decreases, the influence of droplet interfacial forces becomes more significant, this could perturb biological processes inside droplets.

Once droplets are generated they need to be stable and confined to assure proper compartmentalization, also the droplet environment needs to be biocompatible to allow biological processes to occur. Enhancing droplet stability can be achieved by the use of surfactants, which are amphiphilic molecules composed of a hydrophilic "head" and a hydrophobic "tail". During drop formation, surfactant supplied with the continuous phase is redistributed onto the interface of the droplet with the hydrophilic part facing the center of the droplet and the fluorophilic part facing the fluorocarbon oil. Surfactants reduce the surface tension facilitating production and manipulation of droplets, as well as preventing the droplets from coalescing [83, 114]. When stored as an emulsion, droplets come into direct contact with one another, resulting in stable confined coexistence for an extended period of time (even years). But in some cases, uncontrolled droplet merging can happen [115]. Reduction of monodispersity occurs when the contact layers between droplets break or when the dispersed phase diffuses into the continuous phase. Since fluorinated oils have a low solubility for non-fluorinated compounds, the Ostwald ripening is a slow process. Thus, the integrity of emulsions is mostly affected by the presence of the surfactant. If the concentration of surfactant is below the critical micellar concentration (CMC), the droplets are not fully stabilized. In contrast, when the concentration is above CMC, the surface of the droplets is fully saturated with surfactant molecules, making droplets resistant against uncontrolled coalescence. In the work [116], authors studied the kinetic aspects of drop stabilization by surfactants. Apparently, the dynamics of surfactant self-assembling at the droplet interface depend not only on surfactant concentration but also on pre-incubation time before the droplets come into contact with each other. These observations led to the elaboration of an efficient and selective one-to-one droplet fusion technique [117].

In addition to stability, surfactants must provide a biologically inert interior surface for the water drops. The most common functional groups used to prevent adsorptions of biomolecules on droplet interface are polyethylene glycols (PEG). Fluorosurfactants containing PEG groups have been proven to be efficient and highly compatible with enzymatic and cell-based assays [83, 118]. In addition to PEG-based

surfactants, morpholinophospate, pyridinium, carbohydrate and Jeffamine polyetheramine head-groups were proven to be biocompatible [89, 119, 120].

Another consideration in selecting oil/surfactant systems is the uncontrolled exchange of compounds between droplets. The transport of small molecules from the aqueous to the oil phase [121, 122], as well as between droplets of different solute concentrations has been demonstrated to occur [122]. To prevent the molecular exchange between droplets, a leaky compound can be modified by adding hydrophilic groups to it.

d. Droplet generation from parallel streaming

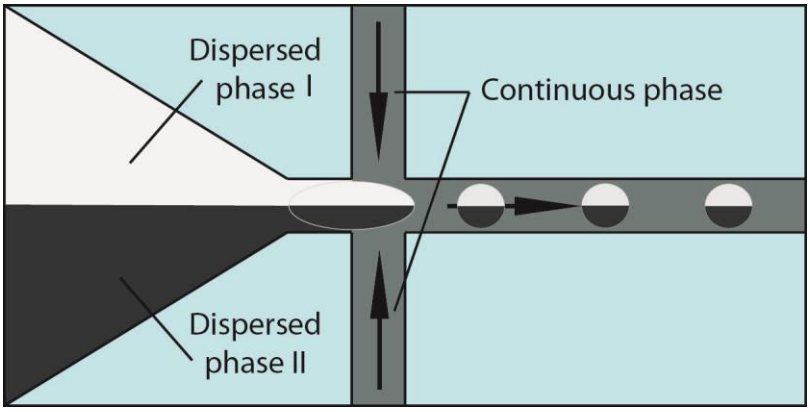

Figure 10. Coencapsulation of two streams.

A significant advantage of microfluidic systems is the ability to vary the content of the droplets by simply changing the flow rate of liquid streams prior to droplet formation. This option provides a powerful tool to control the composition of each droplet, which is very important for many reactions and applications [87, 123]. Since in microfluidic systems $10 < Pe < 10^5$, two streams of miscible fluids injected into a microchannel flow side-by-side mix only by diffusion [124]. The distance that molecules diffuse between these adjacent streams depends on the molecular weight of the molecules, the viscosity of the liquids, the distance from the channel junction and the velocity of the fluids. If the contact time between the flows is short enough, the content of the streams stays separated before encapsulation and the mixing of components takes place within droplets.

3. Microfluidic droplet manipulations

a. Mixing within droplets

The mixing of reagents in droplets is an essential process determined by Marangoni convective motions, and it depends on several aspects: droplet size; composition and ratios of liquids to be mixed; channel size and configuration; speed of the droplet inside the channel. It has been shown that droplets moving in straight channel generate a steady, recirculation flow in each hemisphere [125-128]. The flow mixes the fluid only within the two hemispheres of the droplet but not across the centreline. Authors [128] studied the mixing dynamics inside microfluidic system generated droplets using two-photon fluorescence lifetime imaging (Figure 12).

If the two phases are generated from a FF device, mixing in droplets occurs only by diffusion (Figure 11. a), which is often too slow for some applications like kinetic analysis and high-throughput screening. But, if a droplet is moving through a winding channel it sequentially undergoes different internal Marangoni circulation changes, moving the centreline up and down (Figure 11. c). Also in the winding channels droplets are reoriented, rotating the centreline by a certain angle which further enables convective mixture (Figure 11. d). The angle in the winding channels also plays a crucial role for the mixing efficiency as shown in the picture. The authors [127] have proven that the 45° and 90° bend configurations are more efficient for intradroplet mixing than 135°. The theory of mixing in droplets using this technique is further developed in [129, 130], and the comparison of experiment and simulation has been done in [131].

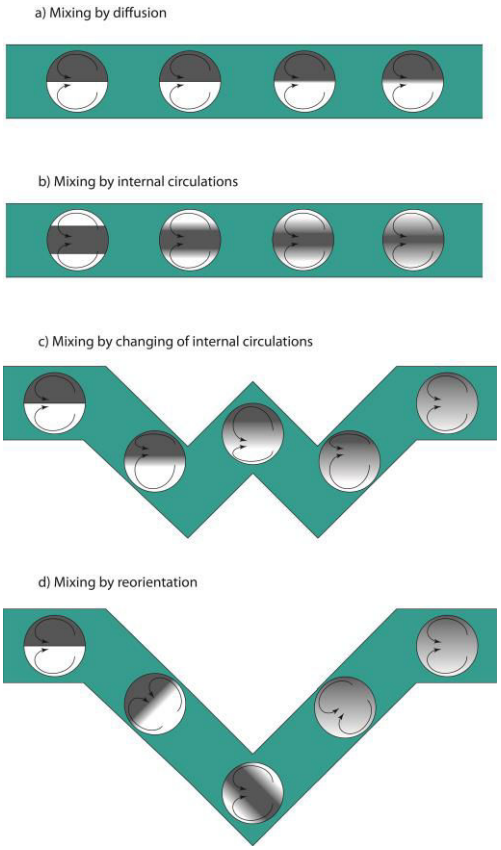

Figure 11. Mixing within droplets. *Two solutions (black and white) are co-emulsified using parallel streaming. The arrows indicate the internal flows. (**a**) Mixing in a straight channel, when droplets are generated from streams next to each other. (**b**) Mixing in a straight channel, when droplets are formed using internal and external streams. (**c**) Mixing in a serpentine channel, when droplets are generated from streams next to each other. (**d**) Mixing by Baker's transformation in winding channel.*

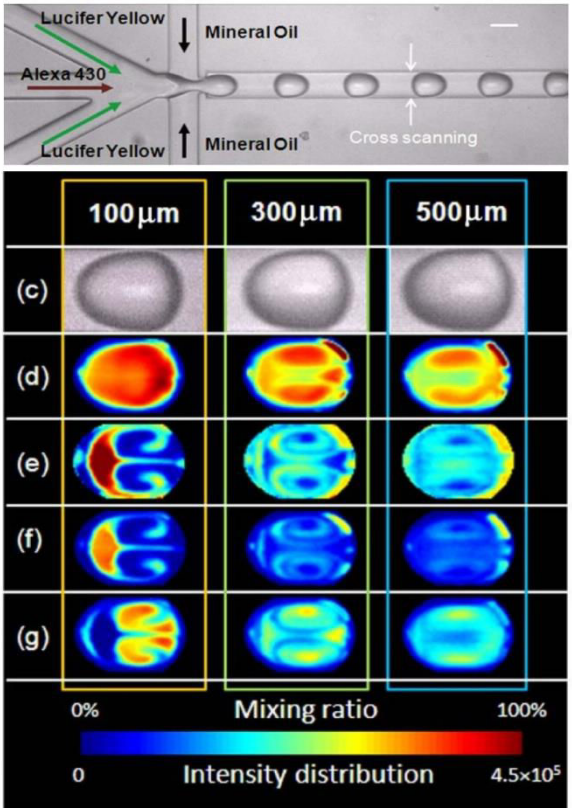

Figure 12. Visualization of mixing dynamics in droplets. *The mixing of two streams containing each a specific fluorophore is monitored at different distances from the junction using two-photon lifetime imaging microscopy. Reprinted from [131].*

To improve mixing of crowded solutions, several research groups have added bumps onto the edges of the winding channel design [132, 133]. This feature further enhances the asymmetry of shear forces on either side of the droplet as it traverses the winding channels, which aids mixing.

The advantages of rapid mixing in droplet microfluidics systems have motivated new methods for directly observing the kinetics of reactions with millisecond

resolution[134], and will allow higher throughput testing for many different types of applications.

b. Droplet splitting

The ability to split droplets is a necessary operation for the execution of assays and the production of sample replicates. For example, fission is used to divide plug into many smaller volume droplets, which allows individual droplets to be paired and mixed with different reagents [135], and also provide a smaller volume container, which allows for rapid mixing and reduces the reaction time in the droplet [136].

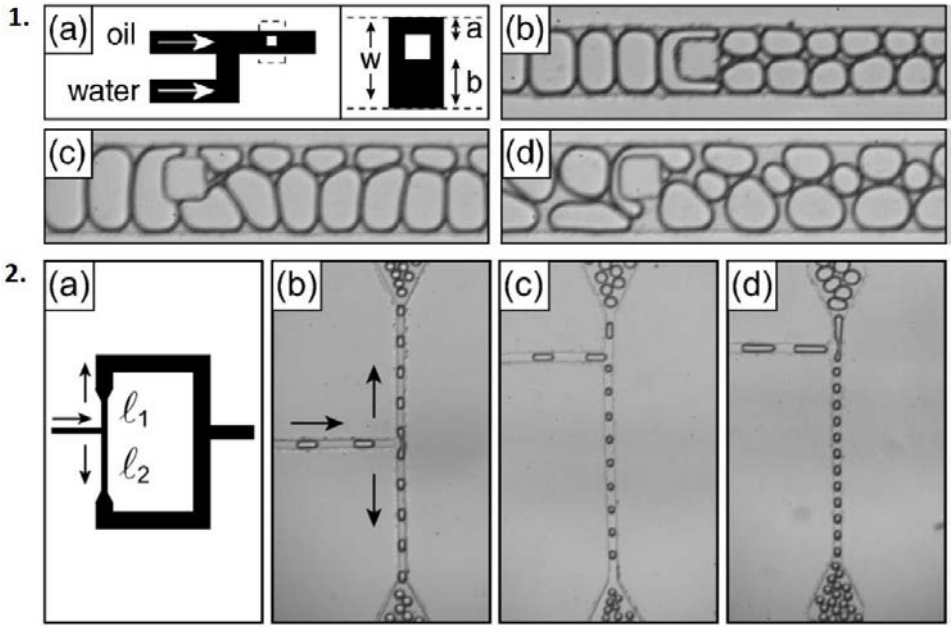

Figure 13. Geometry mediated splitting. 1. *Obstruction passive droplet breakup. The splitting ratio is determined by the position of the pillar.* **2.** *Passive breakup at T junction. The size of daughter droplets is conditioned by the ratio of daughter channel lengths. Reprinted from [137].*

The most common splitting method has been developed by Link *et al.* [137], where the fission is mediated by geometry of the channels (Figure 13). Here, a simple bifurcation of the main microfluidic channel is introduced in order to split droplets. Geometrical fission depends on the relative size of the droplet and the width of the channel. If the length of the droplet in the microchannel is greater than the

circumference on the edge of the droplet, splitting always occurs. But smaller droplets do not break if they are moving too slow, and the flow rate needs to be increased to achieve splitting. The droplet splits evenly if the resistances of the two daughter channels, downstream of the bifurcation, have the same fluidic resistance. The volume ratio of daughter droplets can be changed by variation of the resistance ratio in side-arms. Another geometry-mediated fission employs a large post near the middle of a microfluidic channel to induce droplet fission. By adjusting the position of the post, the ratio of sizes of daughter droplets can be changed. A more detailed analysis of the physical processes involved in geometrical break-up is described in the paper [138].

Since fission product volumes are limited by the fixed geometry, several methods employing electrical fields [139, 140], heat [141] and lasers [142] have been developed to achieve more control over the droplet fission process. In addition, such methods have the ability to divide only droplets possessing a desired property. These additional functionalities come at the cost of a more complex device, but may be necessary for some applications.

c. Fusion

Droplet fusion is mandatory for droplet manipulation in many biological applications, where new reagents need to be added after droplet formation. There are two main steps that allow fusion to occur: firstly, the droplets to be fused should come into close contact for a minimal amount of time; secondly, the destabilization of droplets interface must be induced [143]. We can discriminate passive and active fusion mechanisms by the method which induces this instability.

Passive droplet fusion

Passive fusions do not require active control or an electric field, and they occur by channel geometry mediation or physico-chemical induction. The earliest mechanisms for passive fusion used a section of expanded channel, also known as “expansion volume” fusion mechanisms. In these methods continuous phase is drained from in between droplets allowing the interfaces of adjacent droplets to interact, and further to merge, due to the minute disturbances in surface tension. Either a gradually tapering channel [144], a wider section of a channel [145] or even a 3D channel configuration

[146] may be employed as an expansion volume. By controlling the flow rates and the geometry of the expansion volume consecutive droplets can be fused [147]. The expansion volume mechanisms require careful control of droplets size, spacing and frequency. Authors [148] improved the geometry mediated mechanism by overcoming the problem of inter-droplet spacing. Here, as a droplet enters the expansion chamber, it is squeezed between two sets of pillars (Figure 14. a). The increase in surface tension holds the droplet, while the continuous phase flows around it. The droplet is held until the next droplet comes into contact; fuses; and the hydrodynamic pressure pushes the fused droplet downstream. Analogically, to trap droplets prior to fusion, a membrane valve was used [149].

Another fusion mechanism which does not require droplet synchronization relies on a hydrophilic patch inside the microchannel [151, 152]. The hydrophilic droplet is stopped over the hydrophilic patch and afterward, successive droplets gather behind it. When the viscous drag forces overcome surface tension, droplets are released and fused. Although this mechanism requires no special geometry, the possibility for contamination between droplets exists, due to the interaction of droplets with the hydrophilic patch in the channel. However, the droplet fusion mechanisms described above are usually carried out in the absence of surfactant, which is undesirable when further droplet manipulations are foreseen. Alternatively, droplets stabilized by surfactant are less prone to reactant exchange between them. Also, such droplets are resistant to uncontrolled coalescence which facilitates droplet storage and reinjection for further droplet operations.

Several passive fusion techniques allowing the selective coalescence of surfactant-stabilized droplets have been developed. In the work [150], droplet fusion was induced when the laser was directed to the interface between two droplets (Figure 14. b). This fusion is not suitable for biological applications due to the degradation effects of laser on molecules like DNA and proteins. Other passive coalescence in the presence of surfactant is enabled by constrictions in a microfluidic channel. Bremond *et al*. [153] proved that coalescence occurs not during the droplet impact, but during the separation phase, in which the surfactant interface is destabilized. The exploitation of surfactant stabilization effects led to the development of a high-throughput and highly

controllable passive fusion technique [117]. This mechanism (Figure 14. c) allows selective one-to-one fusion between a surfactant-stabilized droplet and a newly created droplet lacking in surfactant on the interface. Just after the fusion act, droplets are additionally stabilized by providing surfactant in the carrier oil. The resulting droplets are stable and ready to be involved in other microfluidic manipulations. This passive fusion technique does not require lasers, electrical fields, special channel treatment or precise droplet synchronization; and allows highly efficient (> 99% one-to-one fused drops) and high-throughput (> 1kHz fusion frequency) droplet coalescence. A further development of the mechanism permitted selective two-to-one and even three-to-one droplet fusions [154].

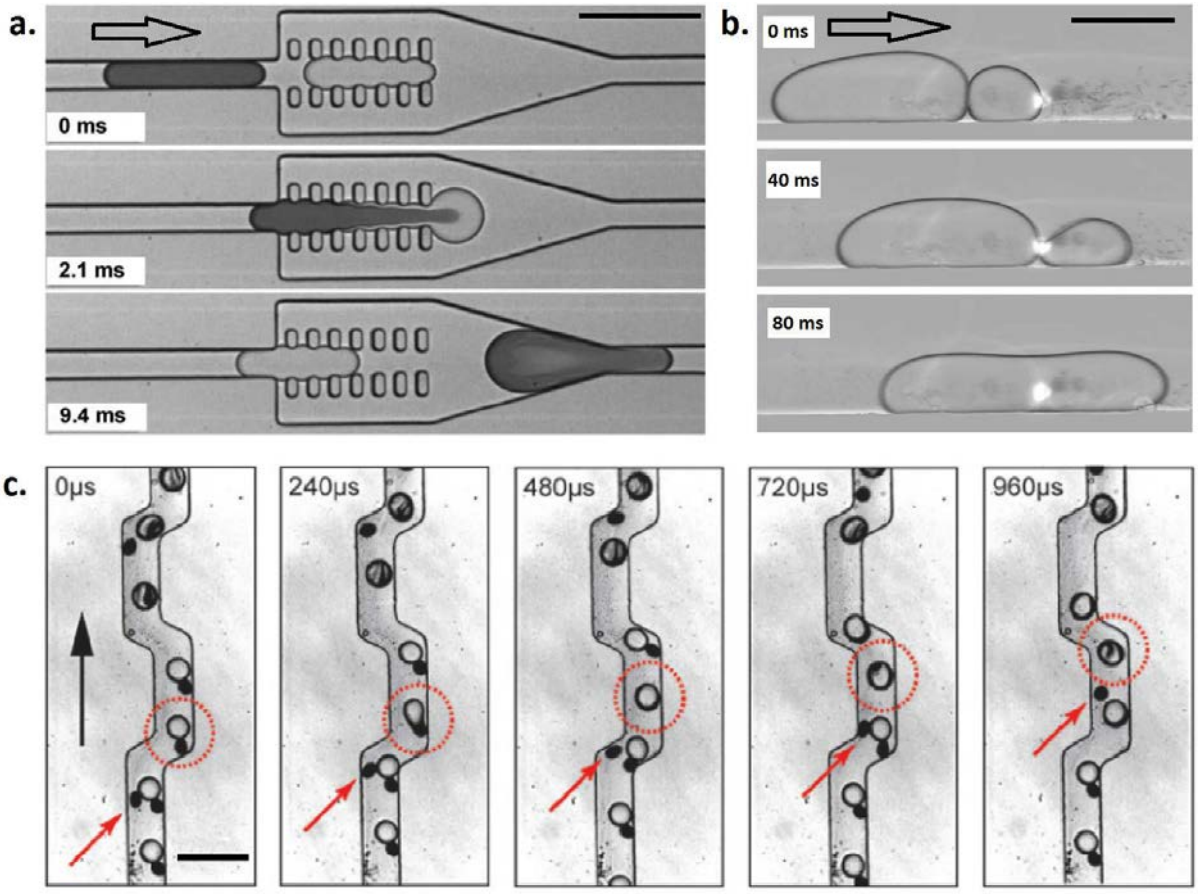

Figure 14. Passive droplet fusion techniques. *a)Pillar-induced droplet merging [148]. b) Laser-induced coalescence [150]. c) High-throughput droplet fusion exploiting surfactant stabilization effect [117]. The bars are representing 100 μm.*

Active droplet fusion

In contrast to passive fusion methods, active methods emply active controls such as electro-coalescence, dielectrophoresis and optical tweezers. The requirement of electrodes and the equipment necessary for active fusion control make these methods more complex. The advantage of these fusion mechanisms is that the employment of electricity can hasten the development surface instabilities initiating fusion more quickly. Also, in the majority of cases the fusion efficiency does not depend on the droplet contents making the active fusion reliable and robust.

There are a number of groups that elaborated fusion methods where the position and the size of electrodes differ. Several fusion techniques are not applicable to biological screenings, due to the low-throughput and high voltage. Examples are devices that employ; a platinum wire inside the channel [155]; an electrode chamber that slows and orientates droplets to be fused [156] or an electrode surface for slowing and trapping drops for fusion [157]. To address this problem authors [158] demonstrated a mechanism where only 1V of direct current (DC) is enough to fuse drops. This feature has been successfully applied to on-chip static droplet assays [159]. An alternative approach imposes an opposite electrical charge on two droplet populations produced on the same chip, which fuse together when they come into contact [139]. However, it is difficult to assure the stability of newly formed droplets, also the reinjection of a charged emulsion is hardly possible. Electrowetting is another method used for droplets manipulation and active fusion in particular [160]. Here, using computer software, the electrodes can be actuated in a programmed order inducing the movement of drops, and fusion when droplets are next to each other.

The most attractive active fusion technique for biological screenings employs alternating current (AC) for electro-coalescence of paired droplets [161, 162]. When two drops (Figure 15. a), with conductivity higher than the continuous phase, are exposed to an electrical field, their induced dipoles align leading to attraction force between them. Once droplets are in contact, the strong dipolar force rapidly drains the surfactant film between two interfaces causing droplet coalescence. The frequency of fused drops in these cases may reach from 300 Hz to 3 kHz.

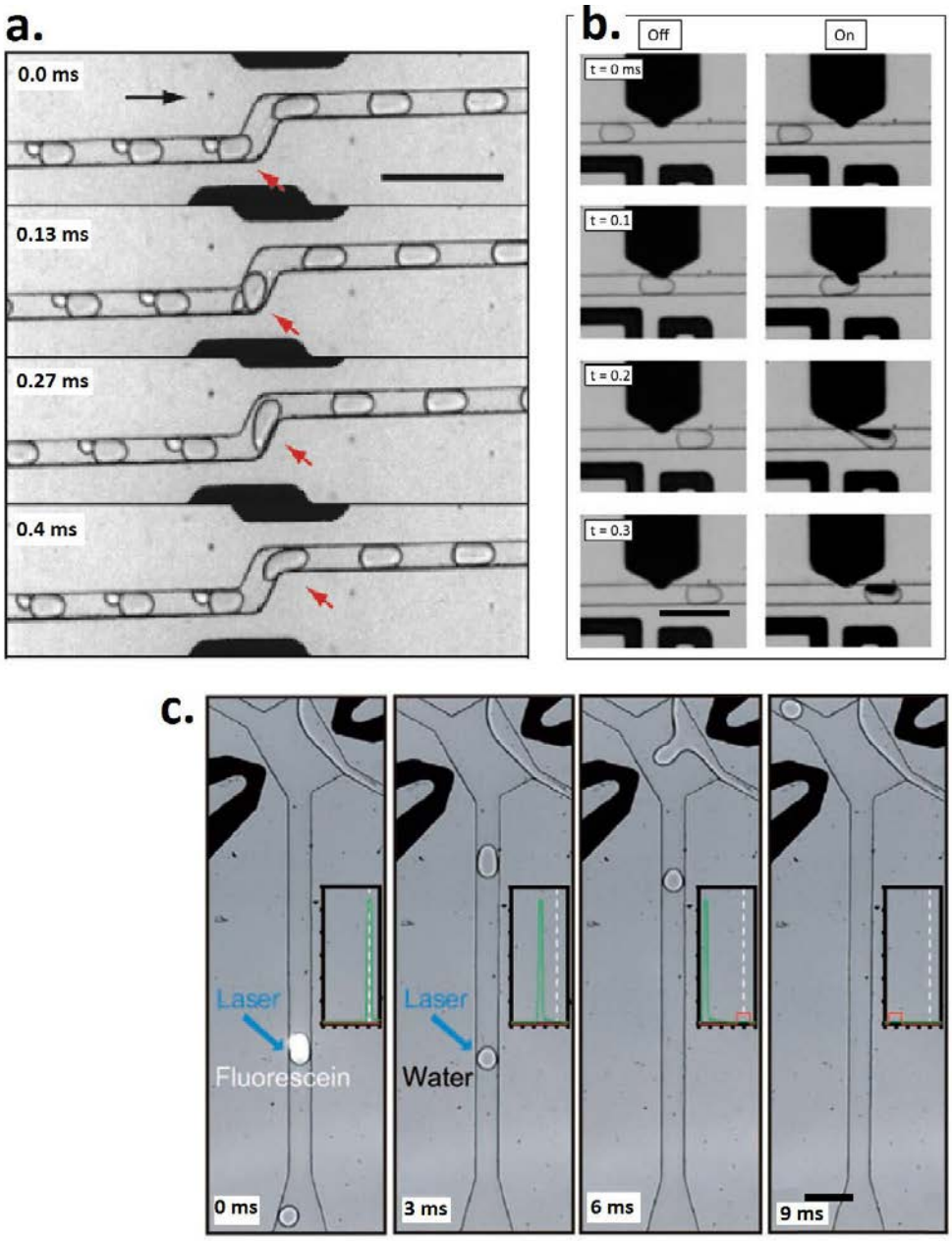

Figure 15. Electro-coalescence-based high-throughput droplet operations.

a) *Electro-coalescence of droplets. [163].* ***b)*** *Injection of picoliter volumes into passing droplets [164].*

c) *Selective fusion of droplets to continuous flow [165].*

Adding reagents into passing droplets

Aqueous solution can be dispensed into droplets not only through fusion with other droplet populations, but also directly from solution. In this strategy (Figure 15.b), the amount of reagent is merged into a passing droplet from a microchannel crossing the main channel [166]. The disadvantage of this technique is that the possibility of contamination is higher, since all passing droplets have direct contact with the second reagent stream. This problem may be overcome by making the injection channel smaller [167]. In this case, the *Pe* number increases, meaning that the addition of reagent is mainly by convection and not by diffusion effects.

In the case of surfactant-stabilized droplets, as expected, the addition of reagent is facilitated by electric fields [164]. Authors proved that from a certain AC voltage (30V) it is possible to inject picoliter volumes of reagent into a droplet. The amount of fluid injected into passing droplets may be tuned by adjusting the pressure in the dispensing channel as well as the velocity of the passing droplets. Also, it is possible to add successively a series of fluids to the droplets, while passing through the main channel assembled with a several picoinjectors. Moreover, this mechanism operates at high frequencies (10 000 Hz) which further increases the potential for high-throughput biological applications.

Droplet fusion to continuous flow

For some applications it is necessary to transfer the droplet content into a continuous flow for further analysis or manipulations. For this reason, in the work [165] has been described a method, where selected droplets from a stream of oil are electrically induced to merge with a parallel aqueous stream (Figure 15. c). If coupled with a detection system, this method could be used for sorting applications allowing selective droplet fusion to aqueous flow.

d. Incubation

A large variety of biological applications require droplet incubations from minutes to hours or even days. For example, enzymatic kinetics need to be monitored from seconds to hours depending on activity; and cells are growing from hours to days. Thus, large spectra of engineering solutions have been developed for on-chip droplet incubation. For a short period of time (~s) droplets can be incubated in a shallow and narrow microfluidic channel (Figure 16. a) where droplets remain in a single-file [134].

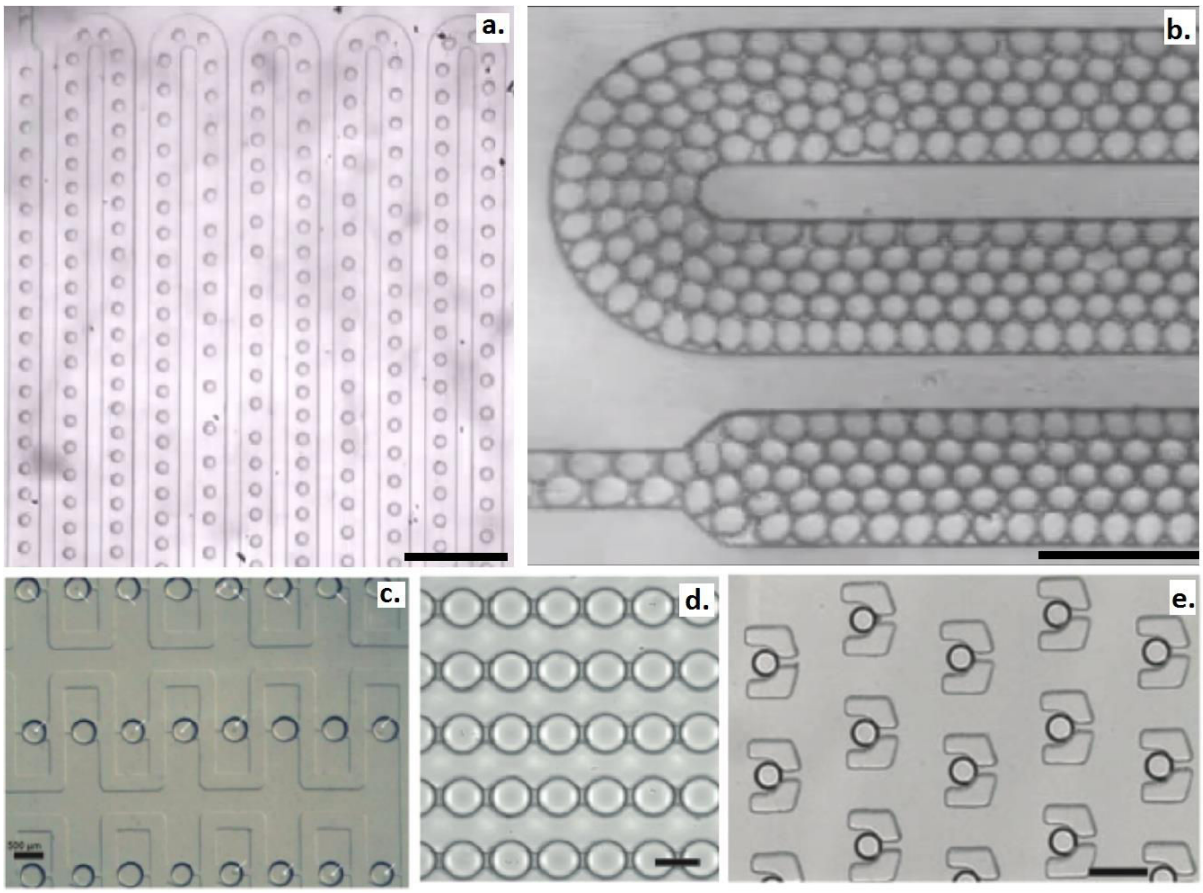

Figure 16. Droplet incubation strategies. ***a)*** *Single-file droplet incubation in delay-line.* ***b)*** *Close-packed droplet incubation in delay-line.* ***c)*** *Droplet trap array employed for C. elegance growth monitoring [168].* ***d)*** *Dropspots device [169].* ***e)*** *Droplet trapping array used for cell-based assays [170]. Scale bars 500 µm.*

However, this mechanism is not suitable for longer time scales, due to the limited increase in the length of the channel and the related pressure drop. Using deep and wide delay-lines it is possible to reduce back-pressure at the same time achieving longer

incubation time. However, this causes the problem of dispersion in the incubation time, related to the unequal speeds with which droplets pass through the delay-line. Authors [171] have studied the dispersion of droplet incubation time in microfluidic channel, emphasizing three main regimes (Figure 17. a): low droplet densities, where the droplets remain in the fastest streamlines moving with almost the same speed; medium droplet densities, where droplets pushed to the walls experience slower speed than droplets in the middle; high droplet densities, here droplets are densely packed in crystal-like structure and move as one block. Moreover, by introducing constrictions with the size of one to two droplets (Figure 17. b), the incubation dispersion is reduced for all regimes, due to repetitive droplet shuffling. Deep delay-lines with constrictions are used for 1-60 min incubation times.

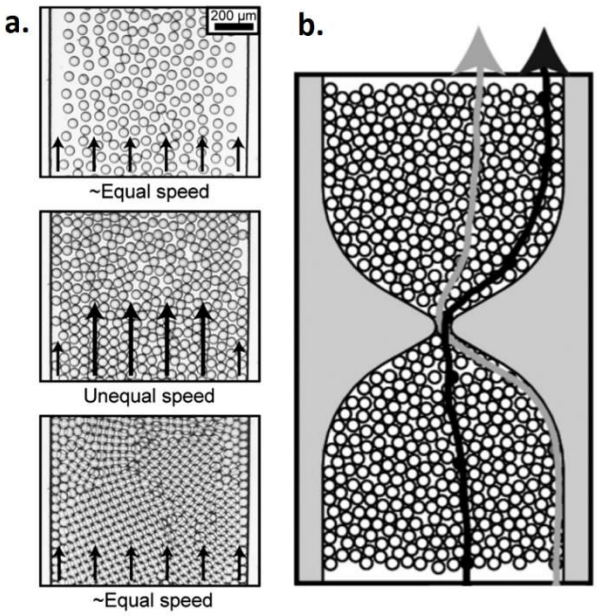

Figure 17. Reducing dispersion in delay-lines. ***a)*** *Dispersion of droplets at different droplet densities.* ***b)*** *Introduction of constrictions reduces dispersion [171].*

In cases where longer incubation time is needed, on-chip reservoirs [172, 173] or static droplet arrays [168-170, 174-176] are employed (Figure 16. c-e). Static droplet traps are used to monitor the enzymatic activity or organism behavior in a single droplet. However, the number of stored droplets usually does not exceed several thousand.

In the end, if a microfluidic experiment requires the incubation of large numbers of droplets for a long period of time (>1h), droplets can be stored off-chip in the form of

emulsion. After a certain time period, the emulsion can be re-injected on-chip for further droplet manipulation.

e. Sorting

High-throughput selection of droplets containing desired variants is the bottleneck for biological applications of droplet-based microfluidics. The simplest sorting techniques rely on device geometry that allows droplet sorting by size. In some cases, it is important to remove small, non-fused droplets or satellite-droplets formed during different microfluidic operations. Thus, this type of hydrodynamic sorting is needed to avoid contamination. Moreover, there are systems where droplet size is directly connected to the droplet content. For example, in some cases, generated droplets which contain cells are larger than the empty ones (Figure 18. a), which allows geometrical sorting of only occupied droplets [105]; or, selective droplet-shrinking due to the cells growth (Figure 18. b), which makes the separation of droplets occupied by live cell colonies possible [177]. Hydrodynamic sorting exploits the property that larger droplets are diverted to a stream with a higher shear. Using this mechanism (Figure 18. d), polydisperse emulsions have been successfully sorted in up to five different-sized droplet populations [178, 179].

A label-free, passive droplet sorting technique using tensiophoresis has been reported recently [180]. Since the interfacial tension of a droplet directly depends on the droplet composition, it may be used to separate droplets with different protein concentration, enzymatic activities or pH-value (Figure 18. c).

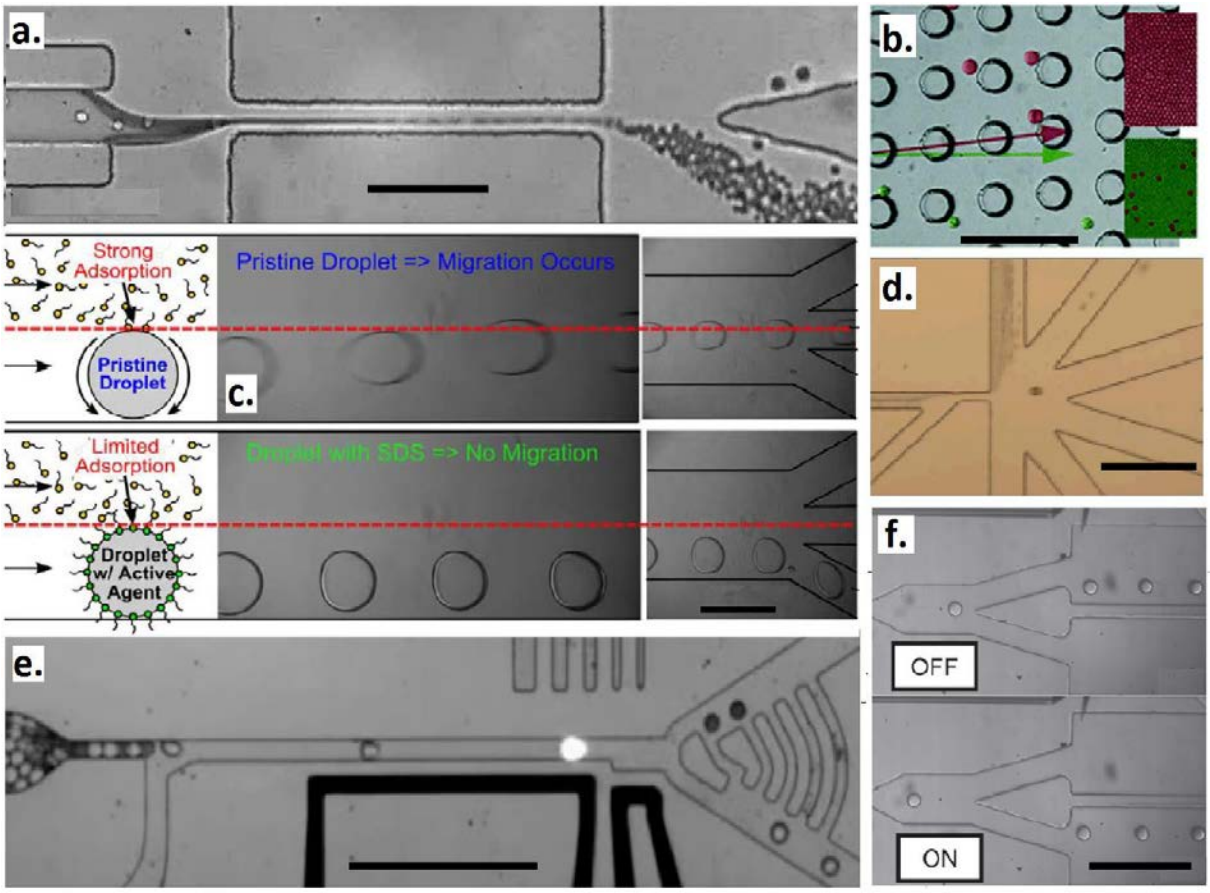

Figure 18. Droplet sorting operations. ***a)*** *Hydrodynamic self-sorting of occupied droplets directly after cell encapsulations [105].* ***b)*** *Droplet size based separation by deterministic lateral displacement [177].* ***c)*** *Droplet sorting by chemical composition using tensiophoresis [180].* ***d)*** *Size-dependent sorting of reinjected emulsion into six droplet subpopulations [178] .* ***e)*** *Fluorescence-activated droplet sorting [181].* ***f)*** *Sorting based on surface acoustic waves [182]. Scale bars – 500 μm.*

As in many other droplet manipulations, dielectrophoresis (DEP) is commonly used in sorting of droplets, cells or other particles. The application of nonuniform electric fields exerts a force on droplets that may be used to direct droplets into one of several outlets of the device. Since the dielectric permittivity of an aqueous droplet is much higher than that of the oil; a positive dielectrophoretic force can pull the droplet towards the high electric field region. The first experiment to show the application of DEP for sorting in droplet-based microfluidics has been performed by the group [183]. In the absence of the electric field, droplets flow into the waste channel with lower hydrodynamic resistance. When the electrodes are actuated, droplets are redirected into the collection channel. In the same manner that fluorescent cells can be sorted by FACS,

droplet sorting can be triggered on fluorescence using dielectrophoresis. This mechanism, named fluorescence-activated droplet sorting (FADS), detects passing fluorescent droplets and applies an AC field to redirect desired droplets to an alternate outlet channel. A device by Baret *et al*. was able to achieve a sorting speed of 2.000 droplets per second (Figure 18. e), with a false sorting rate of 1 in 10.000 droplets [181].

Other possibilities to sort droplets involve: magnetic sorting, when encapsulation of magnetic particles is used to deflect droplets into different channels with a magnetic field [184]; optical sorting, where by positioning the laser on either side of the post structure in the middle of the channel, droplets can be sorted into one of two daughter channels [185]; surface acoustic wave (SAW) sorting (Figure 18. f), where a piezoelectric material creates SAWs that induce acoustic streaming to move droplets into the collection channel [182].

f. Detection

Detection methods in droplet-based microfluidics are essential for the identification of chemical species that are being analyzed. Alongside mass spectrometry, electrophoresis, Raman spectroscopy and electrochemical methods, fluorescence detection is the most commonly used in microfluidics. An optical system consists of light-emitting sources, detection systems and a microchip. Due to the transparency of the microfluidic PDMS-glass chip, all droplet manipulations are available for optical monitoring. The majority of reported studies use a reverse microscope coupled with a video camera. A high-speed camera is preferable when the control of high-frequency droplet operations is required.

The choice of light emitting sources and optical sensing system are strongly related to the type of information that needs to be collected from the droplets. Wide field fluorescence microscopy requires only a fluorescence microscope equipped with a sensitive camera. Charged-coupled device (CCD) and complementary metal–oxide–semiconductor (CMOS) image sensors are used for capturing the fluorescent images digitally, and are often employed when the fluorescence of a chip area needs to be visualized. Wide-field fluorescence microscopy has been used to follow enzyme kinetics in continuously moving droplets [134]. Since the droplet frequencies are higher than the frame rate of the cameras, reported kinetics are obtained from the average fluorescence

intensity of several hundred droplets passing through the detection area during the exposure time. To monitor the contents of individual droplets by fluorescence imaging the flow of droplets needs to be stopped. A large number of microfluidic devices have been coupled with fluorescence imaging to interrogate thousands of individual compartments following the time course of long reactions [121, 170].

Laser-induced fluorescence spectroscopy is employed when high-throughput detection is needed. Moreover, this technique allows one to integrate further handling steps for each individual reactor (for example sorting). The confocal laser induced fluorescence detection system is able to resolve >100 kHz fluorescence events. Moreover, this system makes it possible to register the intensity and number of fluorescent particles (like cells [186]) in droplets. Besides lasers, fluorescence can be induced with other light sources such as light emitting diodes (LEDs) or Hg-lamps. While the extra sensitive detection is achieved by photomultiplier tubes (PMTs) or avalanche photodiode detectors (APDs). Signal output from optic sensors is collected by data acquisition card processing in computer software, like LabView. The extracted information contains data on droplet fluorescence, size and spacing.

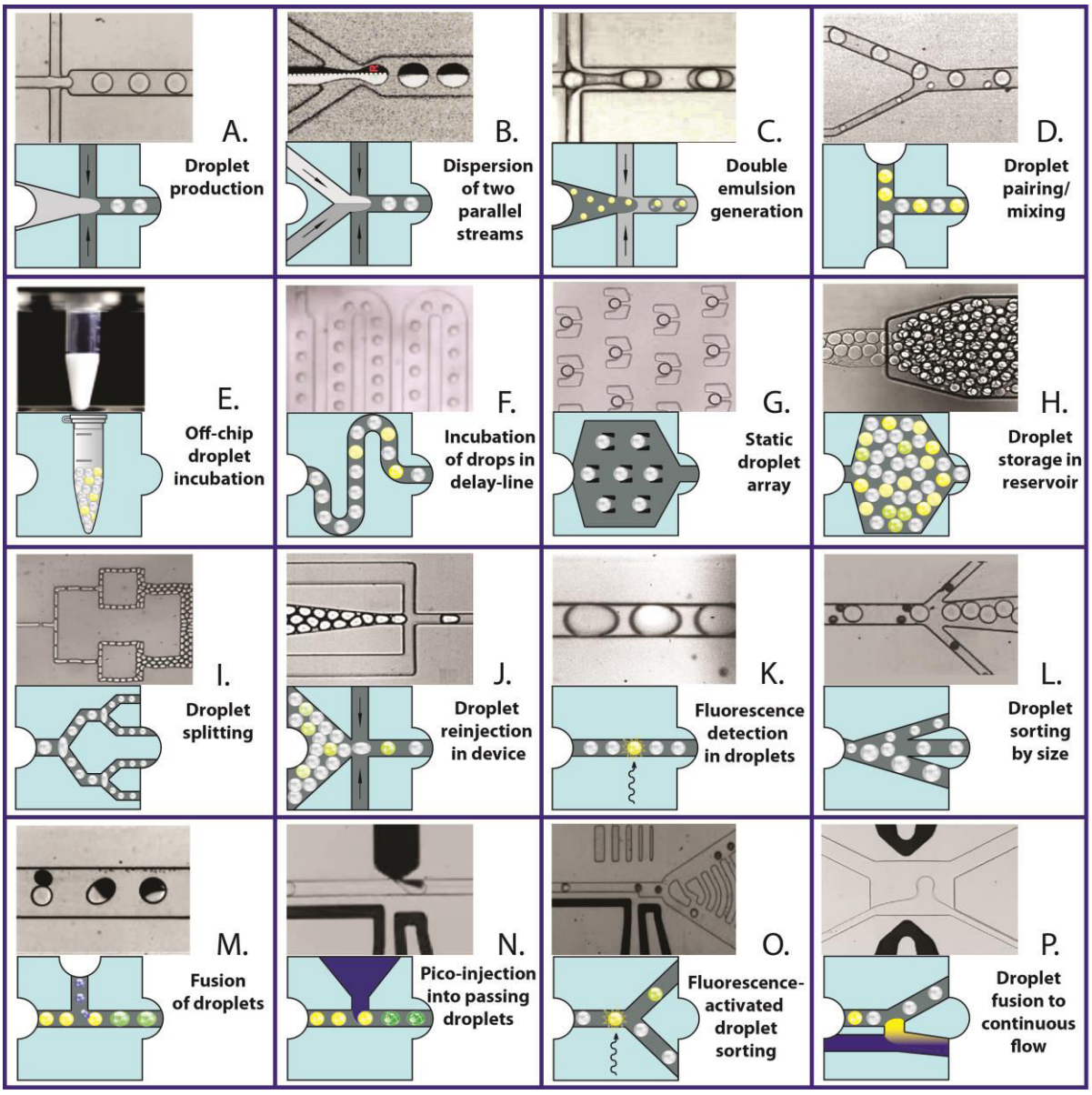

Figure 19. Microfluidic operations and corresponding jigsaw pieces. *The modules for manipulation of microfluidic droplets can be integrated to suit the requirements of specific biological experiments.*

4. Platforms for biological applications

This section focuses on high-throughput applications of droplet-based microfluidics for directed evolution of proteins, as well as, other biological assays.

a. Encapsulation

All droplet-based evolution experiments start with encapsulation of cells or DNA molecules. Simple dilution prior to emulsification is an effective means of ensuring that single cells or biomolecules are compartmentalized in droplets. Inevitably, the encapsulation leads to a Poisson distribution which results in empty droplets, droplets with one cell (biomolecule), droplets with two cells (biomolecules), etc. Poisson distribution is characterized by the equation $P(X = k) = \frac{e^{-\lambda}}{k!}\lambda^k$, where P(X = k) is the probability to have k biomolecules/particles per droplet, and λ is a mean number of biomolecules/particles per droplet. In order to avoid multiple encapsulations, it is necessary to operate with more dilute solutions. As a result, the presence of large number of empty droplets reduces the throughput of the system. Since cells can be regarded as particles, a number of groups elaborated mechanisms to avoid Poisson distribution limitations for cell encapsulation. Chabert and Viovy demonstrated single-cell compartmentalization during jet break-up; however the droplet formation was relatively slow (0.1 kHz) [105]. Several works reported single-cell compartmentalization by relying on cells to self-organize prior to droplet formation. Self-organization occurs when a high density suspension of cells is forced to travel rapidly through a high aspect-ratio microchannel, for which the cell diameter is a large fraction of the channel's narrow dimension [187]; or, when it passes through a curved microchannel which is capable of cell ordering based on Dean forces [188]. The use of cell-ordering mechanisms may increase the system throughput, although accurate design of a microchip is required.

b. Cell-based systems

Cell expression machinery is often employed for protein evolution. Despite the well known drawbacks, *in vivo* systems are still widely used when suitable selection is available. Fast growth, easy handling and low cost are not the only advantages, cells are also performing gene amplification and protein expression simultaneously, which simplifies droplet manipulations. The use of gas permeable perfluorocarbon oils allows the cells to be kept alive for several days, and cell cultures in droplets have comparable viability to cultures in flasks [89, 189]. The development of cell-based assays in microfluidic droplets led to significant progress in the application of droplet-based microfluidic platforms for protein evolution.

The simplest microfluidic setups employ cells with periplasmatic expression or secretion or surface expression of proteins. In these cases, enzymes are easily accessible to the substrate without cell lysis, which makes possible the recovery of live cells after the assay. For example, a cell-based assay for alkaline phosphatase has been performed in picoliter droplets (Figure 20. b) using *E. coli* over-expressing the enzyme in periplasm [190]. The catalytic turnover of the substrate in droplets was measured in less than a minute interval, showing similar kinetics compared to the bulk experiment. An integrated, multilayer microfluidic device has been developed for monitoring enzyme kinetics and protein expression in individual cells compartmentalized in droplets [191]. Cells are coencapsulated with substrate and IPTG and stored on a static droplet array for 20 h (Figure 20. f). Bacteria express both a red fluorescent protein (mRFP1) and the enzyme, alkaline phosphatase (AP), from a biscistronic construct. By measuring the fluorescence intensity of both the mRFP1 inside the cells and a fluorescent product formed as a result of the enzymatic activity outside the cells, gene expression and enzymatic activity can be simultaneously and continuously monitored. In the case of cytoplasmatic expression of proteins it is possible to use membrane permeable substrates [89]; otherwise cells need to be lysed.

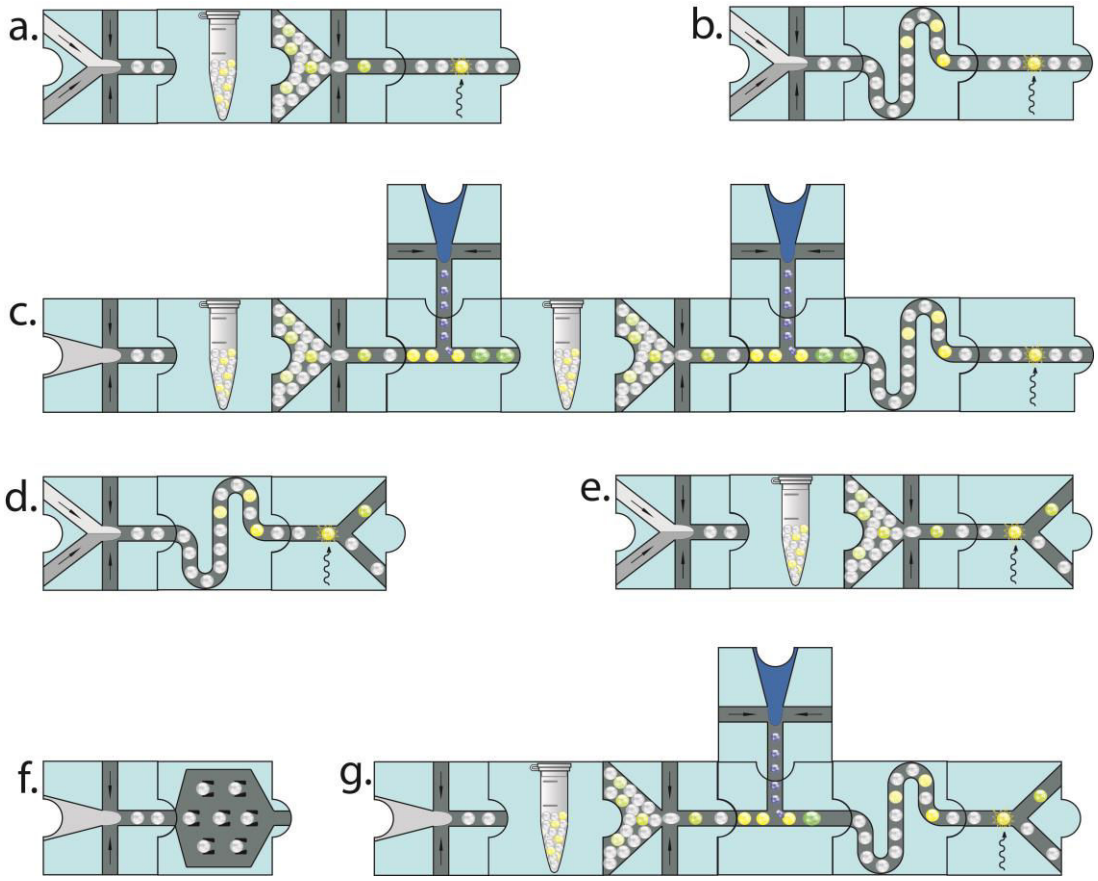

Figure 20. Microfluidic platforms for cell-based assays. ***a)*** *Encapsulation and screening of mammalian cells and multicellular organisms. Cells are co-encapsulated with media, incubated off-chip. Droplets are re-injected in different time intervals to monitor cell viability [89].* ***b)*** *Cell-based enzyme assays. Droplet generation occurs by confluence of three aqueous streams: substrate, buffer and cells. After formation, droplets pass through a delay-line where different kinetic points are taken [190].* ***c)*** *Toxicity screening. Several dilutions of a cytotoxin were compartmentalized with coding fluorophores. After re-injection the library was fused with pre-encapsulated cells and incubated off-chip. Re-injection and fusion with encapsulated cell viability dye was followed by on-chip incubation in a delay line and fluorescence analysis of the viability assay and identification of the drugs using the coding fluorophores [192].* ***d)*** *Directed evolution of HRP. A yeast display library was compartmentalized with its substrate, incubated in a delay line for 5 min followed by enzyme activity sorting [57].* ***e)*** *Directed evolution of PAS. Cell library is co-encapsulated with lytic agent and substrate. After 1 hour off-chip incubation droplets are re-injected for sorting of most active variants [193].* ***f)*** *Simultaneous detection of gene expression and enzymatic activity in cells using droplet array incubation [191].* ***g)*** *Single-cell hybridoma screening. After a 6 h off-chip incubation period, droplets hosting cells are fused with droplets containing the fluorogenic substrate and subsequently incubated for 30 min in a delay line. The final sorting module allows specific collection of droplets with low fluorescence intensity [194].*

A general method for a single-cell-based assay in microdroplets is based on cell compartmentalization followed by lysis to access the over-expressed enzyme. In the work [195], selected single cells expressing β-galactosidase were mixed in droplets with the fluorogenic substrate. After photolysis, the cell contents are released into the droplet, allowing the reaction to occur. RainDance Technologies reported a droplet-based microfluidic platform (Figure 20. c), using cell-based assays for a cytotoxicity screen of single mammalian cells using an optically encoded library [192].

Recently, two rounds of directed evolution of an arylsulfatase from *Pseudomonas aeruginosa* (PAS) have been performed using a cell lysate assay in microfluidic droplets [193]. Cells encoding an enzyme library were co-encapsulated with the assay mixture; and, after 1h off-chip incubation the emulsion was sorted for the improved variants (Figure 20. e). In two rounds of mutagenesis the wt-clone was improved by 6-fold in activity and 6-fold in expression. Due to the low level of microfluidic module integration, protein expression was suppressed to improve the resolution in the second round of evolution.

Directed evolution of yeast-displayed horse radish peroxidase (HRP) has been performed using a single microfluidic chip [57]. After the second generation was possible to isolate variants, whose efficiencies approached the maximum rate allowed by diffusion-limited encounters (k_{cat}/K_M = 2.5*10^7 M^{-1}s^{-1}). Because the starting wild-type HRP already had a high catalytic activity, a microfluidic module for incubation has been easily integrated on the same chip (Figure 20. d). Five minutes delay-line, for the substrate conversion, did not create back-pressure issues for the droplet generation or droplet sorting. Within this work, authors made a time and cost comparison of microfluidic emulsions to the other screening technologies, resulting in 10^6 times less expenses for droplet-based microfluidic experiments.

A high level of integration of microfluidic modules has been recently achieved in the work of El Debs *et al.* for screening of single-cell hybridomas using droplet-based microfluidics [194]. Here on a single microchip droplets were reinjected, fused, incubated and sorted (Figure 20. g). Due to the large size of the droplets the throughput achieved was only 5*10^4 cells per hour (15 Hz). Nevertheless, this microfluidic platform is, to date, the fastest and most efficient system to sort hybridoma cells for the release of antibodies inhibiting a clinically relevant drug.

c. Amplification of DNA

Amplification of DNA in droplets has many advantages related to the small volume of samples and high-throughput format. Encapsulation of a single gene per droplet avoids competition, like favored amplification of shorter DNA fragments over longer ones; and unwanted recombination between multiple homologous DNA fragments [196]. Moreover, emulsion PCR increases the throughput of genome sequencing and analysis methods. For example, DNA amplification in droplets has been already employed for development of next-generation sequencing methods [197, 198], identifying rare mutations [199-201] or high-throughput screening of transcription-factor targets [202]. Initially, emulsion PCRs (ePCRs) have been thermocycled in bulk [196]. Further, microfluidic setups for on-chip PCR were performed in large droplets (μL-nL) [203-205]. However, only the use of picoliter droplets allowed for high-throughput PCR on-chip. DNA amplification is done by thermocycling the whole chip [206, 207], or running the droplets through different temperature zones created on the chip [111, 208]. The latest examples allow for continuous, high-throughput PCR by reducing the time of temperature exchange inside the droplets. Continuous, on-chip PCR is also circumvents the issues related to poor droplet stability at high temperatures that can impede further droplet manipulations. However, efficient PCR amplification of fragments longer than 500 bp (which is the case of most protein encoding genes) is difficult to achieve using on-chip methods. Trying to extend PCR amplification in droplets to larger DNA fragments, authors [209] used an optimized carrier oil and surfactant mixture to keep the compact emulsion stable at high temperatures. In this setup, picoliter droplets containing template are generated on-chip, collected and thermocycled off-chip. Since the resulting emulsion consists of monodisperse droplets, further microfluidic operations become possible.

Thermocycling can be entirely avoided using isothermal amplification [210], however the problem of nonspecific amplification may appear.

d. In vitro transcription/translation systems

The advantages and perspectives of IVTT expression have been already discussed in the section "Selection and screening".

The first *in vitro* expression of green fluorescent protein (GFP) using droplet-based microfluidics has been reported by Dittrich *et al.* [211, 212]. This system has been further developed by the group [172], who fabricated a reservoir device to store up to one million subpicoliter droplets (Figure 21. a). On the same chip, single copies of DNA templates (encoding GFP) were emulsified together with an IVTT mixture; droplets were stored in the sealed reservoir allowing GFP expression; and, at selected time points, droplet fluorescence intensity was monitored in the outlet channel. The model droplet-based microfluidic studies on IVT expression of GFP enabled further development of *in vitro* enzyme screenings.

Since the expression of a single *lacZ* gene was reported to be challenging, authors [210] developed a more advanced microfluidic platform for high-throughput (up to 8 kHz) analysis of β-galactosidase (Figure 21. b). Here single *lacZ* genes were isothermically amplified in droplets using a hyperbranched rolling circle amplification (HRCA). The IVTT/substrate mixture was added through an electro-coalescence mechanism, where 15 pL droplets containing expression mix were fused with 2 pL droplets containing pre-amplified genes. After two hours incubation for β-galactosidase expression, enzymatic activity was monitored in a reinjection device. By adding a new sorting step (Figure 21. c), this high-throughput analysis platform has been adjusted into a similar platform for selection of potential β-galactosidase improved variants [213]. The isothermal amplification has been substituted by emulsion PCR, which required additional droplet size sorting, due the uncontrolled coalescence during thermocycling. A model experiment starting from 1 active to 100 inactive *lacZ* genes resulted in over 500-fold enrichment of active β-galactosidase variant.

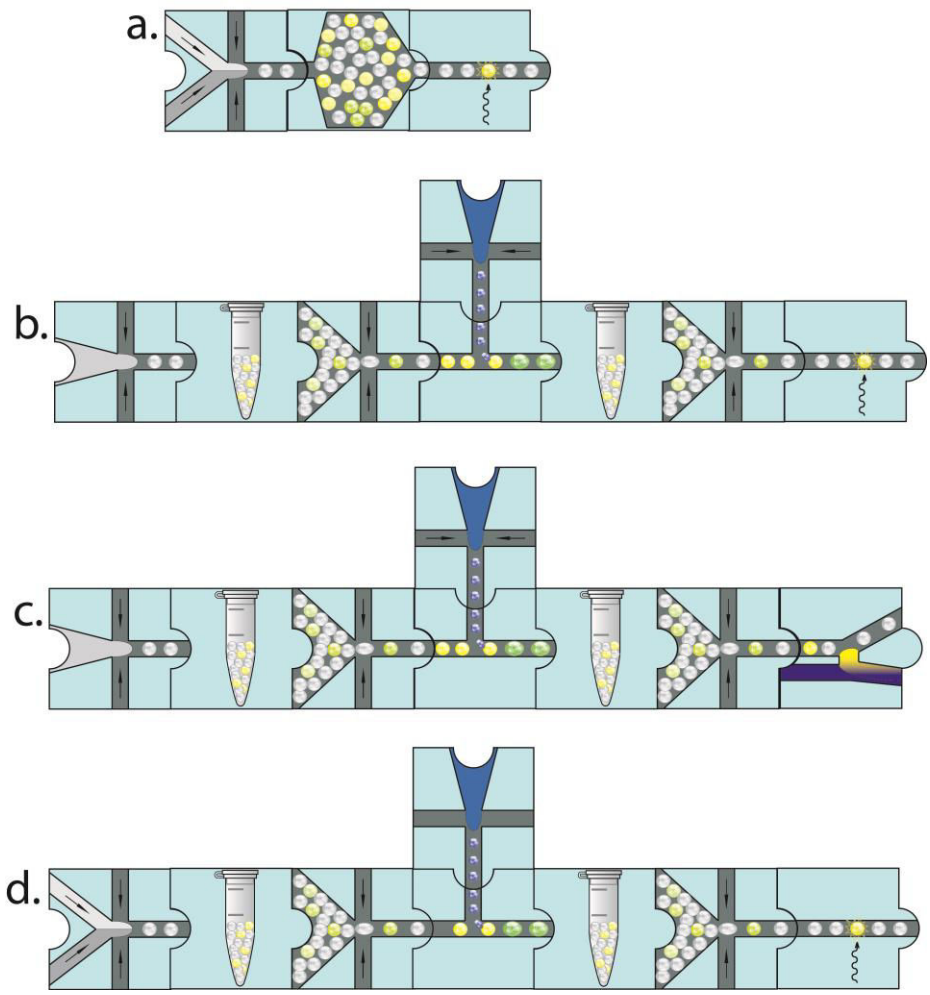

Figure 21. Microfluidic platforms for in vitro protein evolution. ***a)*** *Monitoring in vitro expression of GFP. The GFP encoding genes are co-encapsulated with IVTT mixture. Produced droplets are stored in a reservoir on-chip. Green fluorescence is monitored at the outlet of the device after a certain incubation time [172].* ***b)*** *Single DNA molecule isothermal amplification and analysis. Droplets containing HRCA mixture and β-galactosidase genes are incubated off-chip for amplification to occur. Further, droplets are fused with IVTT/substrate mixture and incubated for protein expression and substrate conversion. Successively, the fluorescent readout is performed [210].* ***c)*** *In vitro screening of β-galactosidase. The same as* ***b),*** *except the final step, where the detection has been substituted by selection of the more active variant by fusion with a continuous flow [213].* ***d)*** *Kinetic analysis of in vitro translated laccase. The coupled in vitro transcription and translation (IVTT) of genes is followed, after droplet fusion, by an enzymatic assay of the translated CotA protein (laccase) [214].*

In vitro experiments described above are assuming the compatibility of gene expression and enzymatic assay conditions, which is not the case for many enzymes. In this regard, multistep microfluidic droplet processing has been reported for the *in vitro* translated laccase [214]. The assay mixture has been added after the protein expression followed by kinetic analysis (Figure 21. d). The systems [172, 214] are built for protein expression from a single gene, while β-galactosidase experiments showed the limitations of this approach. The [210, 213] platforms need to be modified as well, when the assay reaction inhibits protein expression.

Although there are several examples of droplet-based microfluidic platforms for high-throughput screening of *in vitro* translated proteins, no real evolutionary experiments have been performed yet. Moreover, none of these models can be considered as universal *in vitro* enzyme selection technology.

Thesis subject

The number of successful applications of droplet-based microfluidics for biological assays registers a continuous growth in recent years. This trend is explained by the high-throughput format, the ability to miniaturize assays minimizing screening costs and the possibility to vary the selection conditions offered by droplet-based microfluidic technique. Despite a series of microfluidic platforms for directed evolution reported in literature, so far only two were employed for real evolution experiments.

This work describes the development of high-throughput droplet microfluidic platforms fine-tuned for protein of interest and their employment in directed evolution experiments. When not available, fluorogenic assay for monitoring desired enzyme activity (-ies) in droplets was developed. In the course of retro-aldolase evolution a novel fluorogenic substrate was designed, tested and successfully applied for libraries screening. Moreover, the *in vivo* expression in cytoplasm simplified the microfluidic operations which allowed the successive integration of microfluidic modules on the same chip. After a couple of evolution rounds the initial retro-aldolase variant was significantly improved. In other project, to meet industrial requirements a high-throughput screening platform for protease evolution in detergent has been assembled and validated. Two evolution rounds showed the accumulation of a certain pool of beneficial amino acid mutations over the selection/mutation rounds.

Finally, the research described in this work highlighted that *in vitro* expression systems are sensitive to the amount of supplied DNA and reaction conditions. This observation led to the development of a multistep completely *in vitro* microfluidic platform. Here starting from a single gene amplified in droplet, to which the *in vitro* expression mixture and the assay components are added successively, to allow first the protein production and after the enzymatic assay.

I. Materials and methods (general)

1. Microfluidic device fabrication

The large variety of microfluidic designs used in this work were created using software AutoCAD for computer-aided design from Autodesk, Inc. The corresponding photolithography masks were printed by Selba S. A. The microfluidic devices were fabricated by patterning channels and electrodes in poly(dimethylsiloxane) (PDMS) using conventional soft lithography methods [215]. Depending on desired layer depth, a certain photoresist SU8 (MicroChem Corp.) was spin-coated onto silicon wafers (Siltronix), patterned by UV exposure (MJB3 contact mask aligner; SUSS MicroTec) through a photolithography mask and subsequent developed (SU-8 developer; MicroChem Corp.). When a multilayer device was needed, several soft lithography rounds were performed starting from the thinner layer.

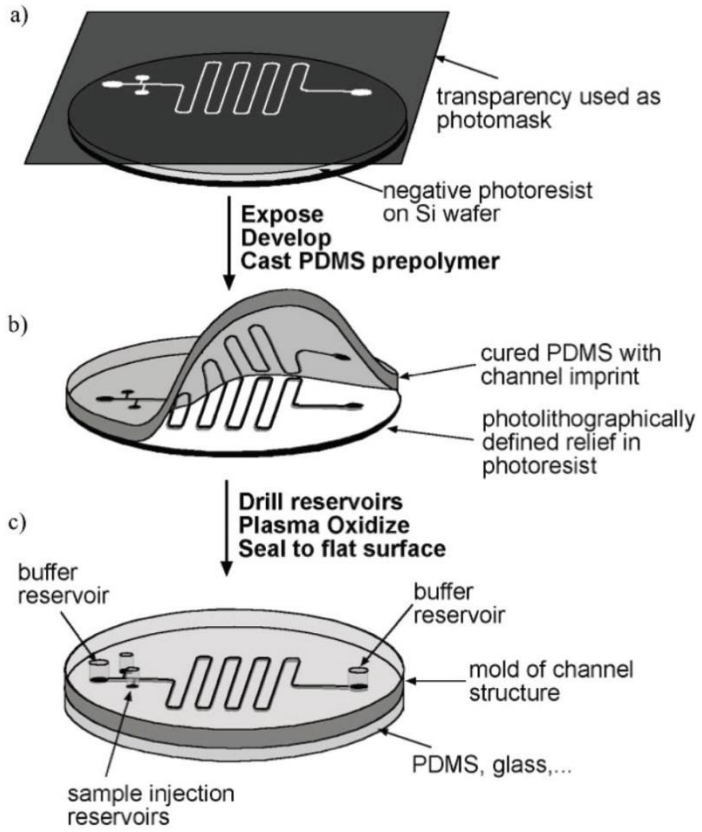

Figure 22. Schematic of fabrication of PDMS microfluidic devices using soft lithography. *Reprinted from [163].*

Next, a 10:1 (w/w) mixture of Sylgard 184 silicone elastomer and curing agent (Dow Corning Corp), degassed under vacuum, was poured onto the silicon wafer and cured at 65°C for 2 hours. Afterwards, the structured PDMS layer was peeled from the master and input/output ports were punched out of the PDMS with a 0.75 mm-diameter (or 0.50 mm for integrated module) Uni-Core biopsy punch. The structured side of the PDMS piece was brought into conformal contact with a glass slide (1 mm thickness) by exposing both parts to oxygen plasma (PlasmaPrep 2 plasma oven; GaLa Instrumente GmbH) and pressing them together. For devices with electrodes, indium tin oxide (ITO) coated glass slides were used (Delta Technologies). Prior to use, microfluidic channels were coated with a commercial hydrophobic surface coating agent (perfluorododecyl-1H, 1H, 2H, 2H-triethoxy-silane, SigmaAldrich) and subsequently flushed with N_2. Electrodes used for sorting, electro-coalescence and pico-injection experiments were fabricated in close proximity to the microfluidic channels as described previously [216]. The patterned electrode channels were filled with metal: the device was heated to 85°C and 51In/32.5Bi/16.5Sn low-temperature solder (Indium Corporation) was melted inside the electrode channels. Electric connections with the solder electrodes were made with short pieces of electric wire (Radiospares). Electrodes were used in combination with an Agilent A33220A function generator connected to a TREK Mode 623B amplifier.

2. Microfluidic device operation

In typical experiment each of the phases were injected into the PDMS channels via PTFE tubing (0.56/1.07 mm, 0.36/0.87 mm Fisher Bioblock) connected to 1 mL syringes (Omnifix-F®, BRAUN) and 0.6 x 25 mm (0.4 x 16 mm for integrated module) Neolus needles (TerumoCorporation). The flow rates of liquids and oils were controlled by syringe pumps (PHD 22/2000, Harvard Apparatus). The flow rates for aqueous phases were in the range of 10-200 μL/hr, for the carrier oil 40-300 μL/hr and for reinjected emulsion 20-100 μL/hr.

3. Fluorinated oils and surfactants

The fluorinated oils used in this work were HFE7500, HFE7100 and FC40 from 3M (Figure 23). The majority of experiments were performed using HFE7500 fluorinated oil. HFE7100 was used for the microchannel coating procedure and FC 40 for passive fusion experiments.

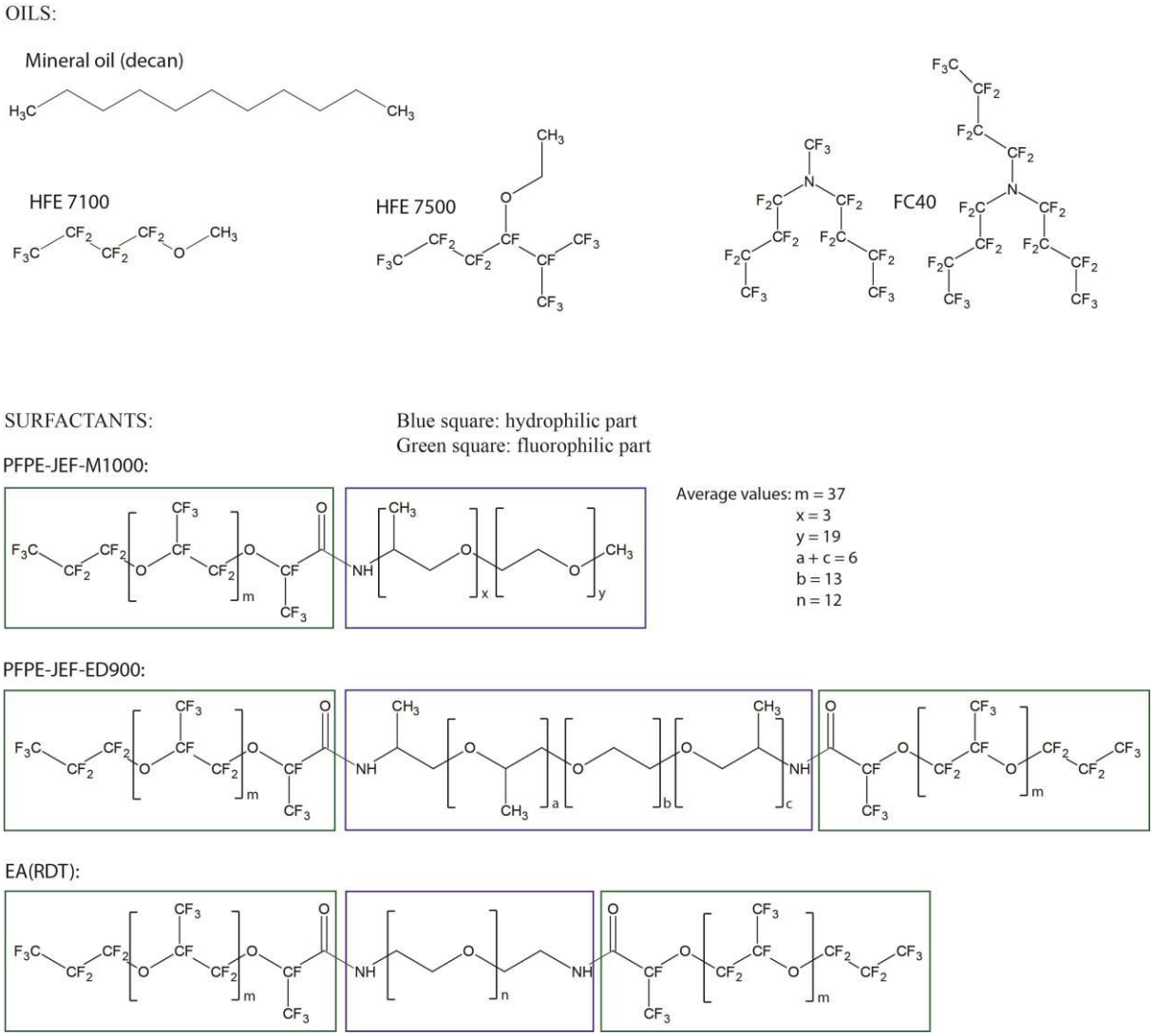

Figure 23. Fluorinated oils and surfactants. Oils: *HFE 7100 - Methoxy-nonafluorobutane; HFE 7500 - 2-trifluoromethyl-3-ethoxydodecofluorohexane; FC 40 –a mixture of perfluoro-tri-N-butylamine and perfluoro-di-N-butylmethylamine;* **Surfactants:** *PFPE-JEF-M1000 - One poly(perfluoropropylene)glycol tail combined with a single Jeffamine M1000 head group; PFPE-JEF-ED900 – Two poly(perfluoropropylene)glycol tails coupled with a single Jeffamine ED900 head group; EA - Two poly(perfluoropropylene)glycol tails coupled with a poly(ethylene)glycol (PEG) head group.*

EA – surfactant obtained from RainDance Technologies - was used for experiments at initial stages, later being substituted by PFPE-JEF-ED900 and PFPE-JEF-M1000, both synthesized by Estelle Mayot in our lab (Figure 23). The switch from EA to PFPE-JEF-ED900 did not cause a significant difference in emulsion stability or biocompatibility. PFPE-JEF-M1000 was used when emulsion incubation at high temperatures was required.

4. Optical setup and detection

a. Setup I

Light from a 488 nm, 25 mW solid-state laser ('LAS1'; Newport-Spectraphysics), attenuated 1000 times with a neutral density filter 3OD ('NDF'; Thorlabs GmbH), was combined with light from a 532 nm, 25 mW solid-state laser ('LAS2'; Newport-Spectraphysics) using a single-edge dichroic beamsplitter ('DBS1'; LM01-503-25 from Semrock Inc.). The combined laser beams were transmitted through a multi-edge dichroic beam splitter ('DBS2'; Di-T488/532/638-25x36x5.0; Semrock Inc.) to the side camera port of an Axiovert 200 inverted microscope (Carl Zeiss SAS). Inside the microscope the laser light passed through a beam splitter ('BS') and was reflected up into the objective ('OBJ'; LD Plan Neofluar 40×/0.6 from Carl Zeiss SAS) by a conventional mirror ('M'). The combined laser beams were focused to a ~20 μm-diameter spot inside a channel in the microfluidic chip ('CHIP') where it excited droplets one at a time as they flowed past. The fluorescent emission from each droplet passed back along the path of the laser beam, and was reflected by the dichroic beam splitter ('DBS2') to the PMTs (both H5784-20; Hamamatsu Photonics KK). Two notch filters ('NF1' and 'NF2'; NF01-488-25 and NF01-532U-25, respectively; Semrock Inc.) eliminated potentially damaging reflections from the lasers and the light was split between the two PMTs by a single-edge dichroic beamsplitter ('DBS3'; FF562-Di02-25x36; Semrock Inc.). Light emitted by fluorescein or DNA-bound RDT-D1 intercalating dye was measured by one PMT ('PMT1') after filtering through a bandpass filter ('F2'; FF02-510/20-25 from Semrock Inc.). The remaining PMT ('PMT2') measured the light emitted by Texas Red or Resorufin after filtering through another bandpass filter ('F3'; FF01-617/73-25; Semrock Inc.). Light from the microscope's halogen lamp ('LAMP')

illuminated the channels and droplets, allowing the droplets to be observed by a Phantom v4.2 high-speed digital camera ('CAM'; Vision Research) when fluorescence measurements were being taken.

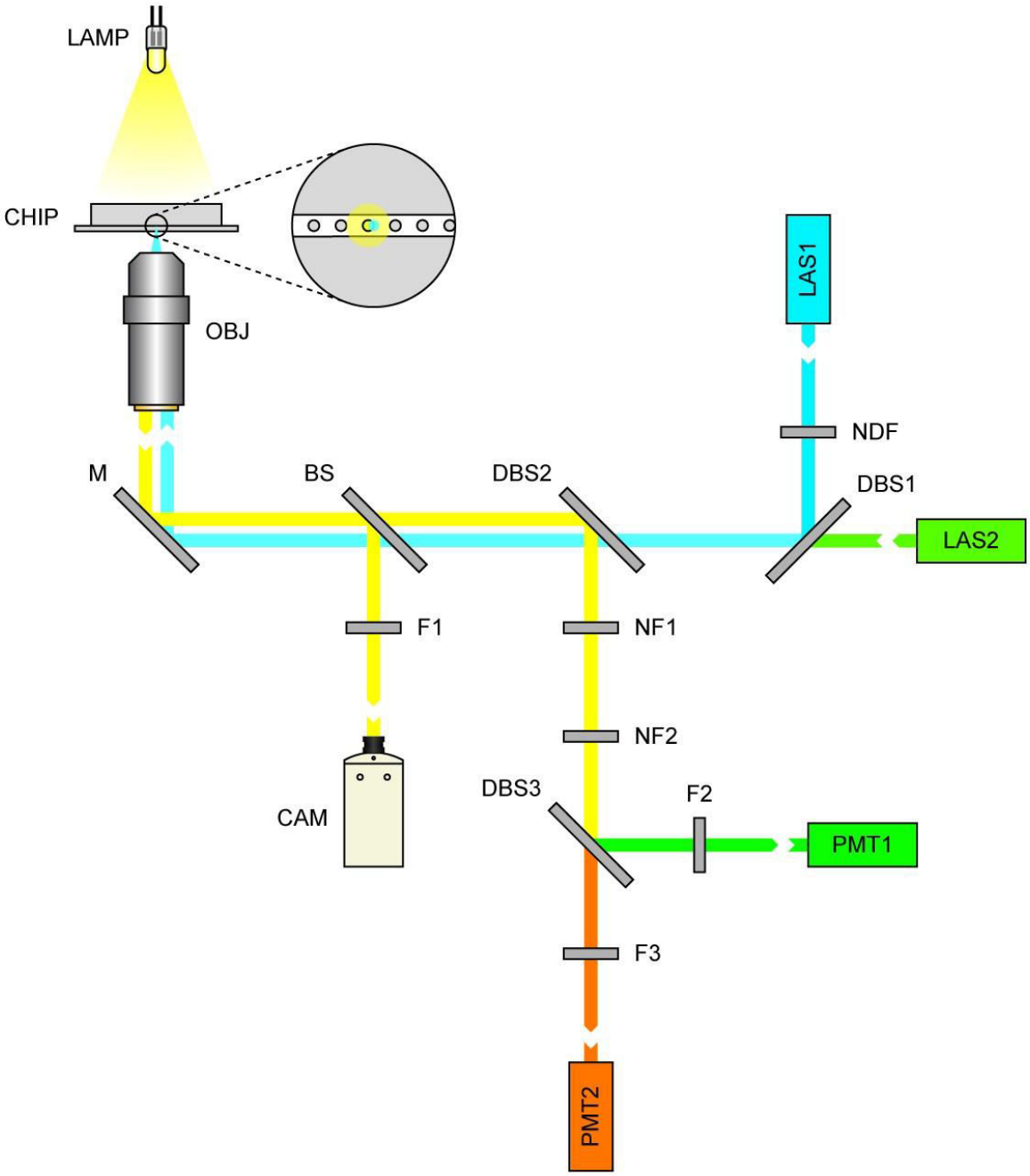

Figure 24. Schematic representation of the optical setup I.

b. Setup II

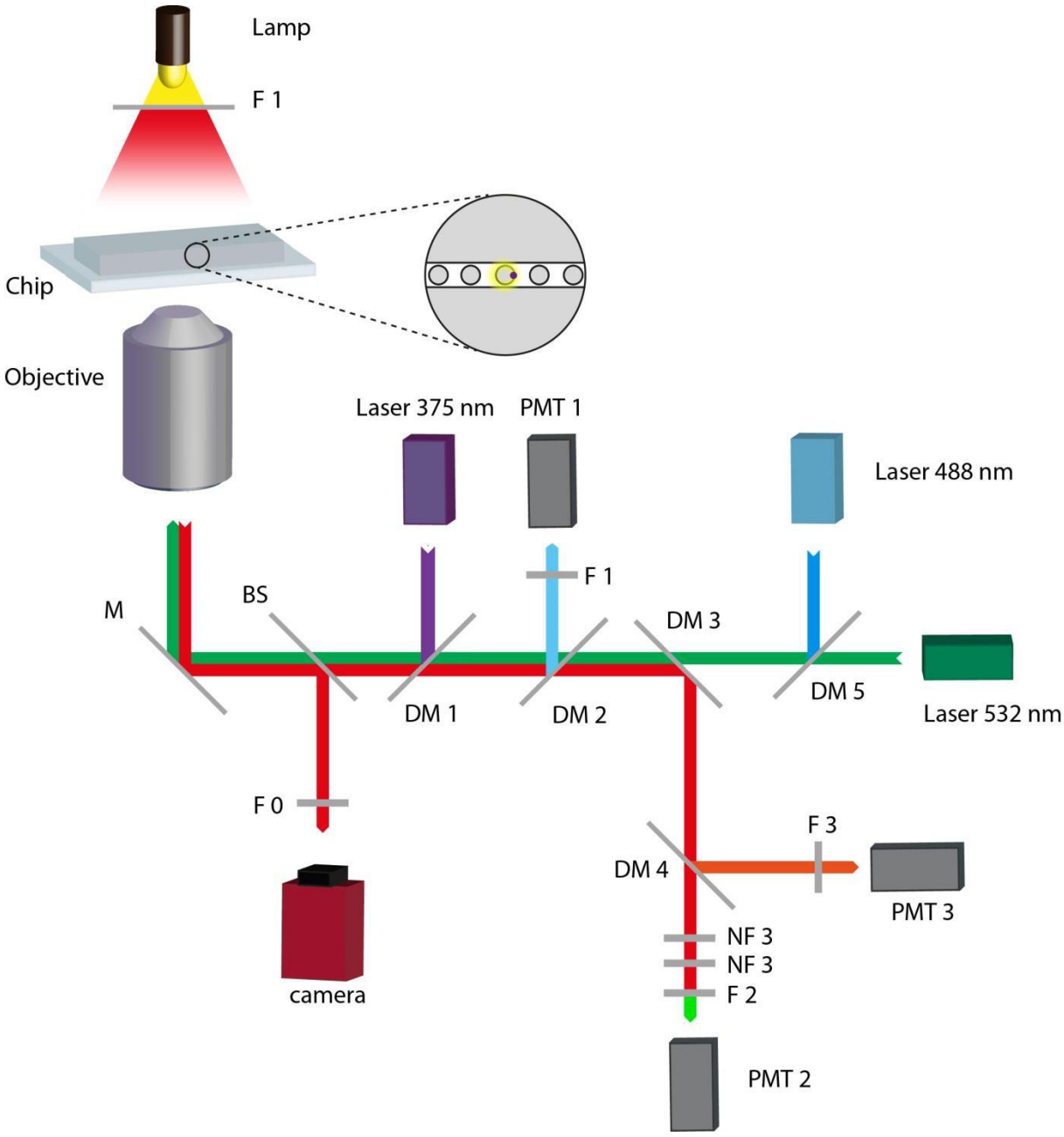

Figure 25. Schematic representation of the optical setup II. *BS: FF665-Di02-25x36, DM1: Di01-R405-25x36, DM2: FF482-Di01-25x36, DM3: Di01-T405/488/532/647-13x15x0.5, DM4: FF562-Di02-25x36, DM5: FF509-Di01-25x36, F0: FF01-736/128-25, F1: FF01-447/55-25, F2: FF02-525/40-25, F3: FF01-607/70-25, NF3: NF01-532U-25.*

Directed evolution of de novo designed retro-aldolase

De novo design has been proven to generate new biocatalysts, but the catalytic activity of such enzymes is not nearly comparable to natural ones [2, 15, 217]. Thus, up to date, all virtually created enzymes need further optimization through site-directed mutagenesis and laboratory evolution.

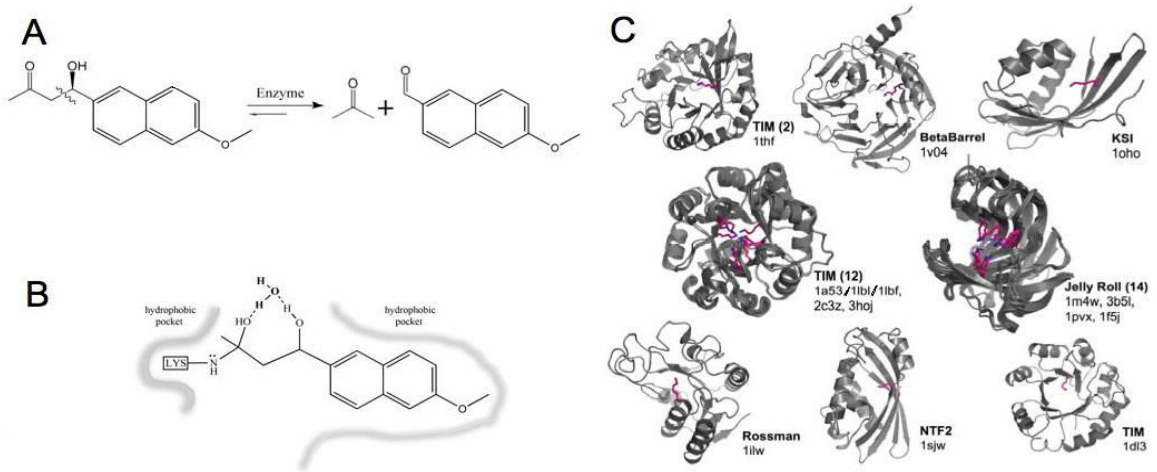

Figure 26. A. *The retro-aldol reaction of methodol.* ***B.*** *Scheme of the active site.* ***C.*** *Different scaffold backbones and 30 different catalytic lysine positions utilized by the active design. Reprinted from [218].*

Jiang *et al.* [15] developed a general computational method for constructing enzyme active sites for multistep reactions, called *Rosetta*. This methodology has been employed to design a novel retro-aldolase, capable to break carbon-carbon bond in a nonnatural substrate: methodol (Figure 26. A). *RosettaMatch* have been used to find sites in known scaffolds to imitate the desired catalytic motif. Next, the optimization of surrounding side chains for favorable interactions with the substrate/transition state model was performed by the *Rosetta design* algorithm. The experimental testing of 72 predicted variants resulted in 32 designs with detectable retro-aldolase activity. In a more recent work [218], the same protein structure prediction program was optimized to yield 33 active retro-aldolases from 42 predictions. All 65 active designs represent 14 different protein backbone scaffolds. Althoff *et al.* [218] decided to focus on three

distinct scaffolds represented by the four most promising variants: the jelly roll (RA60), the NTF2 fold (RA110), and the TIM barrel (RA45 and RA95) (Figure 26. C).

Table 1. Enzymatic activity of four retro-aldolase designs before and after optimization. *The catalytic antibody 38C2 was included for comparison. Data from [218] and [219].*

	K_M(μM)	k_{cat} (min^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)	Fold increase in k_{cat}/K_M
RA60*design*	510	0.0093	0.27	-
RA60.2*improved*	660	0.070	1.8	7
RA110*design*	1600	0.005	0.048	-
RA110.4*improved*	278	0.070	4.2	88
RA110.4-3*evolved*	76	0.240	53	1100
RA110.4-6*evolved*	69	0.230	55	1100
RA45*design*	800	0.0017	0.036	-
RA45.2*improved*	439	0.0139	0.54	15
RA45.2-8*evolved*	54	0.100	32	900
RA45.2-10*evolved*	80	0.230	47	1300
RA95*design*	540	0.0020	0.053	-
RA95.5*improved*	220	0.190	14.5	270
RA95.5-5*evolved*	230	4.1	270	5100
RA95.5-8*evolved*	**210**	**10.2**	**810**	**15300**
catAB 38C2	24	0.71	490	-

The k_{cat}/K_M of initially selected variants ranged from 0.03-0.3 $M^{-1}s^{-1}$; thus further experimental active site optimization has been carried out. Individual amino acid substitutions have been generated near or at the active site, as well as at the some surface positions believed to be responsible for solubility or expression in the same protein family. Positive mutations have been combined and tested for additive effects. The most active improved designs in each of the scaffolds were characterized with Michaelis-Menten kinetics. The cycles of site-directed mutagenesis and combination improved catalytic efficiencies from a minimum 7-fold for RA60.2 up to 270-fold for RA95.5 (Table 1). Nevertheless, these variants displayed the efficiency constants at least

20 times lower than k_{cat}/K_M of the catalytic antibody 38C2, one of the most active artificial aldolase.

In order to increase the activity even further, several rounds of directed evolution were performed on the most successful designs from each scaffold. The genes encoding these enzymes were diversified by error-prone PCR and DNA shuffling. With every cycle, 800 clones were screened in microtiter plates and 1% of the best variants were pooled out for the next evolution step. The first rounds of directed evolution resulted in a significant increase in catalytic efficiencies: 13-fold for RA110.4*improved/* RA110.4-3*evolved*, 60-fold for RA45.2*improved/*RA45.2-8*evolved* and 20-fold for RA95.5*improved/*RA95.5-5*evolved*. But further attempts to improve retro-aldolase variants faced the limitations of the medium throughput screening system. Variants RA45.2-8 and RA95.5-5 showed less than 1.5-fold improvement per round, while RA110.4-3 did not evolve at all during three mutations cycles (Table 1).

Because of the limited size of the screened library and the small number of selected clones, the number of evolutionary pathways of proteins becomes increasingly restricted with every round, leaving a vast protein functional sequence space unexplored. Within an actual collaborative project (with Richard Obexer and Prof. Donald Hilvert from ETH Zurich), we are aiming to develop and apply an ultra-high throughput droplet-based microfluidic screening platform for *de novo*-designed retro-aldolase evolution. Such a system will allow the screening of up to 10^7 clones in a day with significantly reduced reagent amounts, increasing the chances to evolve artificial enzymes to the level of natural ones.

I. Materials and methods

1. Synthesis of retro-aldolase fluorogenic substrate

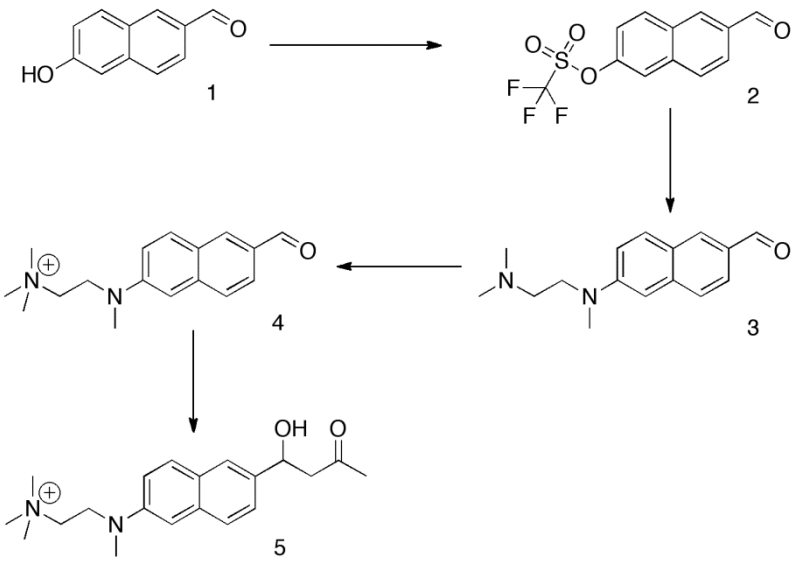

Figure 27. Route of chemical synthesis of fluorogenic substrate:
2-[(6-formyl-2-naphthyl)(methyl)amino]-N,N,N-trimethylethanaminium.

Synthesis of compound **2**:

500mg 6-hydroxy-2-naphthaldehyde (Sigma-Aldrich) and 700µL of dry pyridine were dissolved in 50 mL dry THF and cooled to -20°C. Slowly, 540µL of freshly distilled triflic anhydride were added. After 1h, the reaction mixture was brought to RT and stirred over night. The reaction was quenched by the addition of water and extracted 3 times with dichloromethane. The organic phase was washed with diluted HCl, water, and brine and dried over sodium sulfate. After evaporation of the solvent under vacuum, the residual oil was purified by silica gel chromatography (2.5:1 hexane ethylacetate). 880mg (yield: 99%) of a pale yellow solid were obtained. 1H NMR (300MHz, CDCl3): ∂ = 10.17 (s, 1H), 8.37 (s, 1H), 7.97 (m, 2H), 7.95 (d, 1H), 7.49 (dd, 2H).

Synthesis of compound **3**:

In a flame dried flask, 70mg **2**, 60µL trimethylenediamine, 115mg cesium carbonate, 32mg BINAP and 23mg Pd2dba3 were dissolved in 3mL dry toluene. The mixture was heated to 100°C for 3h. The reaction was quenched by addition of water and extracted 3 times with dichloromethane. The organic phase was washed with water

and brine and dried over sodium sulfate. The solvent was evaporated in vacuum and the residual dark oil was purified by silica gel chromatography (2:8:0.1 hexane:ethylacetate:triethylamine). 30mg (yield: 50%) of a dark yellow oil were obtained. 1H NMR (300MHz, CDCl3): ∂ = 9.99 (s, 1H), 8.13 (s, 1H), 7.8 (d, 2H), 7.63 (d, 1H), 7.16 (dd, 1H), 7.86 (d, 1H), 3.63 (t, 2H), 3.12 (s, 3H), 2.68 (t, 2H), 2.37 (s, 6H).

Synthesis of compound **4**:

30mg of **3** were dissolved in 5mL dry dichloromethane and 6µL methyliodide were added. After 12h, the reaction mixture was extracted 3 times with water. Water was removed by lyophilization and a yellow solid was obtained in nearly quantitative yields. 1H NMR (300MHz, CD3OD): ∂ = 9.94 (s, 1H), 8.24 (s, 1H), 7.95 (d, 1H), 7.76 (s, 2H), 7.34 (dd, 1H), 7.10 (d, 1H), 4.05 (t, 2H), 3.64 (t, 2H), 3.31 (s, 9H), 3.18 (s, 3H).

Synthesis of compound **5**:

20mg of **5** were dissolved in 700µL water, 300µL acetone and 10µL pyrrolidine. The reaction was stirred at 4°C for approx. two days. Reaction progress was monitored by analytical HPLC. After completion, the reaction mixture was directly injected onto a Nucleosil 300-7 C8 column. The eluents were water and acetonitrile, containing 1M of triethylamionium acetate. Fractions, containing the product were pooled and lyophilized. 8mg (yield: 50%) of an orange solid were obtained. 1H NMR (300MHz, CD3CN): ∂ = 7.74 (m, 3H), 7.39 (dd, 1H), 7.05 (dd, 1H), 6.78 (d, 1H), 5.17 (m, 1H), 3.87 (t, 2H), 3.52 (t, 2H), 3.16 (s, 9H), 3.05 (s, 3H), 2.84 (m, 2H), 2.14 (s, 3H).

2. 96-multiwell plate screening assay

PBS: 50mM sodium phosphate, 160mM NaCl, pH 7.5

High salt PBS: 50mM sodium phosphate, 300mM NaCl, pH 7.5

Supplemented M9 medium: 60mM Na_2HPO_4, 20mM KH_2PO_4, 8mM NaCl, 18mM NH_4Cl, 0.1mM $CaCl_2$, 1mM $MgSO_4$, 0.4% Glucose, 5µg/mL thiamine, 0.08% Yeast ForMedium Complete Supplement Mixture, pH 7.5)

LB-Kan: LB supplemented with 30µg/mL kanamycin sulphate

Precultures were prepared by transferring single colonies to 96-U-well plates (Nunc) containing 150µL LB-Kan (LB supplemented with 30µg/mL kanamycin sulphate) and incubated overnight at 37°C and 250rpm. 2mL 96-deep-well-plates (Eppendorf) filled with 1mL LB-Kan were inoculated with 20µL preculture and incubated for 2.5h at 37°C and 250rpm. Protein production was induced by addition of IPTG to a final concentration of 0.5mM. After 5h, cells were harvested by centrifugation, the medium was discarded and the pellets were frozen at -20°C. Cell pellets were resuspended in 300µL of PBS buffer (50mM sodium phosphate, 160mM NaCl, titrated with NaOH to pH 7.5) containing 1mg/mL egg white lysozyme and 5µg/mL DNAseI and incubated at 25°C for 1h. One cycle of freeze-thaw was performed and hereafter the volume was adjusted to 1mL by addition of 600µL PBS. Cell debris was removed by centrifugation, and 195µL of the cleared lysate were transferred to a black 96-well fluorescence plate (Nunc) containing 5µL of the substrate solution (8mM (±)-methodol in acetonitrile). Fluorescence was recorded (330nm excitation and 452nm emission wavelength) over time in a fluorescence platereader (Varioskan, Thermo Fisher).

3. Protein production and purification

100mL LB-Kan was inoculated with 200µL fresh overnight culture and incubated at 37°C and 250rpm until an OD_{600} value of 0.4 was reached. The culture was then induced with IPTG to a final concentration of 0.5mM. After 5h, cells were harvested by centrifugation and stored at -20°C. Cells were resuspended in 10mL high salt PBS buffer containing 1mg/mL lysozyme, 5µg/mL DNAseI and 10µg/mL RNAseA and incubated for 1h at 25°C. Cells were sonicated and the lysate was cleared by centrifugation. The soluble fraction was loaded onto a Ni-NTA column with a bed volume of 1mL (Qiagen). The column was washed thereafter with multiple column volumes of high salt PBS containing 20mM and 40mM imidazole. The protein was eluted with PBS buffer containing 250mM imidazole and dialyzed to PBS. Protein concentration was determined by measuring absorbance at 280nm and dividing by the calculated extinction coefficient (ExPASy ProtParam tool, http://web.expasy.org/protparam).

4. Kinetic analysis in cuvettes

Michaelis-Menten kinetics were measured using 1mL UV quartz cuvettes in a UV-VIS spectrometer (Perkin Elmer). The precise substrate concentration was measured at 330nm ($\varepsilon_{methodol,\ 330nm}$ $1{,}420 M^{-1}s^{-1}$) prior to enzyme addition. The final enzyme concentration was 0.1µM in PBS buffer containing 2.7% acetonitrile. The substrate concentration ranged from 0 to 500 µM. At higher concentrations, the substrate precipitated. Reaction progress was monitored at 310 and 350nm ($\varepsilon_{6\text{-methoxy-2-naphthaldehyde},\ 310nm}$ $16{,}100 M^{-1}s^{-1}$, $\varepsilon_{6\text{-methoxy-2-naphthaldehyde},\ 350nm}$ $5{,}970 M^{-1}s^{-1}$). The initial slopes were corrected for the background reaction without enzyme. The values were fitted according to the Michaelis-Menten equation ($v_0/[E] = k_{cat}[S]/(K_m + [S])$) to give the k_{cat} and K_m values.

5. Library generation and cloning

Diversification of the parent enzyme, or the parent mixture was achieved either by random mutagenesis or cassette mutagenesis.

Random mutagenesis was performed using the GeneMorphII kit (Agilent). In the first step a linear template was generated by PCR amplification. Next, the template was gel purified and subjected to epPCR according to the supplier's manual. The epPCR fragement and the acceptor vector pET29b(+) (Novagen) were digested with NdeI and XhoI (NEB) and ligated with T4 ligase (NEB). The ligation mixture was purified with the DNA Clean and Concentrator-5 Kit (Zymo Research) and desalted on VIVACON 500 (30,000 MWCO HY, SartoriusStedim Biotech) spin columns prior to transformation into electro competent XL1-Blue *E. coli* cells (Agilent).

Cassette mutagenesis was achieved using a degenerate primer and a non-degenerate primer, sharing an overlapping region. In the first step two small fragments were generated directly from the plasmid using outer primers and the library primers. The fragments were mixed in an equimolar ratio and assembled by PCR. The subsequent steps were performed as described above.

DNA shuffling was performed in analogy to the method published by Stemmer [38].

6. Growth of *E. coli* for microfluidic assay

Isolated and desalted library plasmids were transformed into electro competent BL21-Gold(DE3) cells (Agilent). The recovered cells were incubated for 10h at 37°C and 250rpm in LB-Kan. A fresh LB-Kan culture was inoculated with the transformed cells to an OD_{600} value of 0.1. After 1.5h, protein production was induced by addition of IPTG to a final concentration of 0.5mM. After 5h, cells were harvested by centrifugation and washed with supplemented M9 medium (60mM Na_2HPO_4, 20mM KH_2PO_4, 8mM NaCl, 18mM NH_4Cl, 0.1mM $CaCl_2$, 1mM $MgSO_4$, 0.4% Glucose, 5µg/mL thiamine, 0.08% Yeast ForMedium Complete Supplement Mixture, pH 7.5). The cells were filtered through a 5µM syringe filter, and diluted to the desired OD_{600} value. For one cell per one 20 pL-droplet encapsulation, an OD_{600} of 0.25 at the droplet nozzle is required.

7. Plasmid recovery

The sorted droplets were demulsified by addition of 1H,1H,2H,2H-Perfluoro-1-octanol, and the aqueous phase was diluted by addition of 200 µL lysis buffer (10mM Tris-HCl (pH 8), 10mM EDTA, 100mM NaCl, 1% Triton X-100, 10mg/mL proteinase K) to protect plasmid from degradation by nucleases. The aqueous phase was isolated and incubated for 2h at RT prior to DNA isolation using the DNA Clean and Concentrator -5 Kit (Zymo Research) according to the manufacturer's manual. The eluted DNA was directly amplified with Phusion Hot Start High-Fidelity DNA Polymerase (Finnzymes) and primers RiO13 and RiO14. The gel-purified fragment was then amplified and recloned as described above.

8. The exchange of retro-aldolase reaction product between droplets

The retention time of synthesized naphthaldehydes was determined by preparing solutions of 50 µM and 500 µM in water or in 100 mM phosphate-buffered saline (PBS; pH 7.00). Each aqueous solution was infused at 100 µL/hr into a dual droplet maker (Figure 28, A). Fluorinated oil HFE-7500 containing 2 % (w/w) EA surfactant is infused at 500 µL/hr to generate 40 pL droplets by a flow-focusing nozzle (Figure 28, A. ii). At the same time, the initial population distribution was monitored right after mixing on dual droplet maker device (Figure 28, A. i. green spot). Fluorescence intensities were

registered by excitation with 375 nm UV-laser and detection by green photomultiplier at 525±40 nm (Figure 25). The collected emulsion was stored off-chip in a glass capillary for several hours. In different time intervals, a small sample of the droplet mixture was analyzed using a reinjection module (Figure 28, B). The emulsion was reinjected into an analysis device at 20 µL/hr and the spacing was done by surfactant-free fluorinated oil at 100 µL/hr. Droplets fluorescence is monitored with the same detection system while passing through the microfluidic module (Figure 28, B. i. green spot).

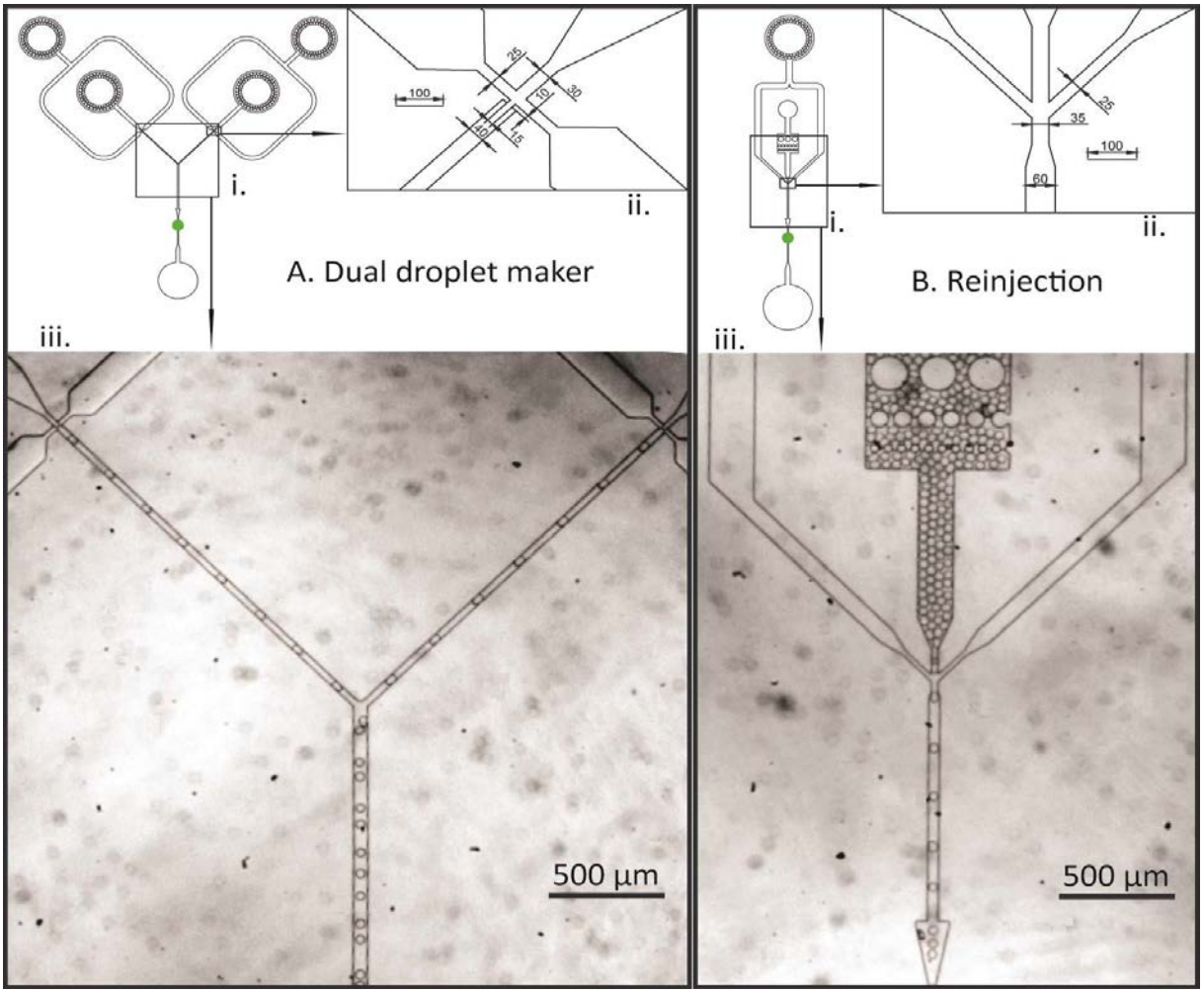

Figure 28. Microfluidic modules used for leakage tests.. *(A) Dual droplet maker module. (B) Reinjection module. (i) Microfluidic designs. Measurement points where the laser was focused are indicated by green spots. (ii) Nozzle parameters in µm. Depths of the channels in both devices are 25 µm. (iii) Pictures of droplet formation and mixing (A) and droplet reinjection (B). Dimensions are given in µm.*

9. Kinetics in droplets

Aqueous phase I (*E. coli* suspension to yield 10% droplet occupancy after droplet generation, or *E. coli* lysate corresponding to one cell per droplet occupancy) and aqueous phase II with substrate and assay components (as in "Microfluidic assay with off-chip incubation") were injected at 15 µL/hr into the microfluidic device for short-time kinetics (Figure 29). The flow-rate of the oil phase with 2% surfactant was maintained at 30 µL/hr to form ~ 18 pL droplets (~500 Hz). Once generated, droplets passed through a delay-line with multiple measurement points for fluorescence detection.

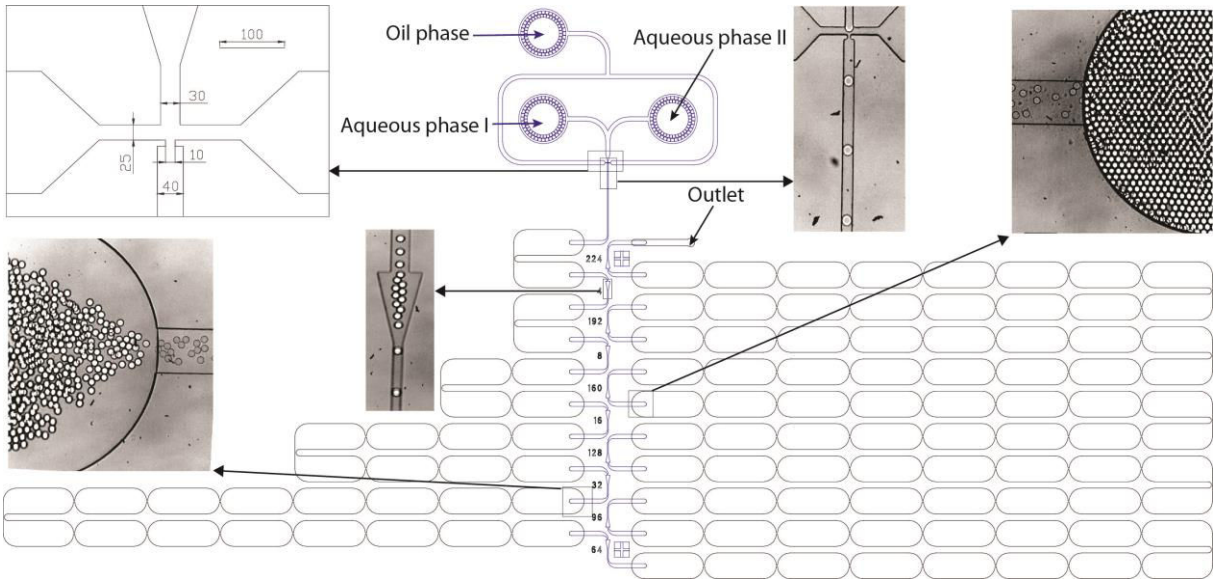

Figure 29. Microfluidic device for short time kinetics. *Droplet production module and measurement-points are 20 µm-deep, layer in blue. Delay-line is 75 µm deep, in black. The dimensions of all device features are indicated in µm.*

10. Microfluidic assay with off-chip incubation

A syringe containing aqueous phase I - solution *E. coli* cells, expressing the enzyme library, and a substrate syringe with aqueous phase II (500μM ammonium substrate, 20mM EDTA, 4mg/mL polymyxin b, 1mg/mL egg white lysozyme, 5μM ammonium aldehyde, 30% Percoll in supplemented M9 medium) were connected to the PDMS chip for parallel stream coencapsulation (Figure 30. A). The water phases were run at 100 μL/hr each, while the oil phase containing 2% surfactant (EA or PFPE-JEF-ED900) was injected at 300 μL/hr to generate 25 pL-sized droplets (~2.2 kHz). The resulting monodisperse emulsion was collected on ice in a 1-mL syringe containing surfactant-free oil. After incubation at 37°C for a certain period of time (from 30min to 1h depending on the case), emulsion was reinjected at 20 μL/hr which corresponds to ~ 400 Hz at the sorting junction (surfactant-free oil at 400 μL/hr). The sorting parameters were: F = 30 kHz, ts_{ort} = 0.8 ms, U_{sort} = 1.3 kV_{p-p}. Sorting was triggered according to the desired fluorescence threshold (Figure 30. B).

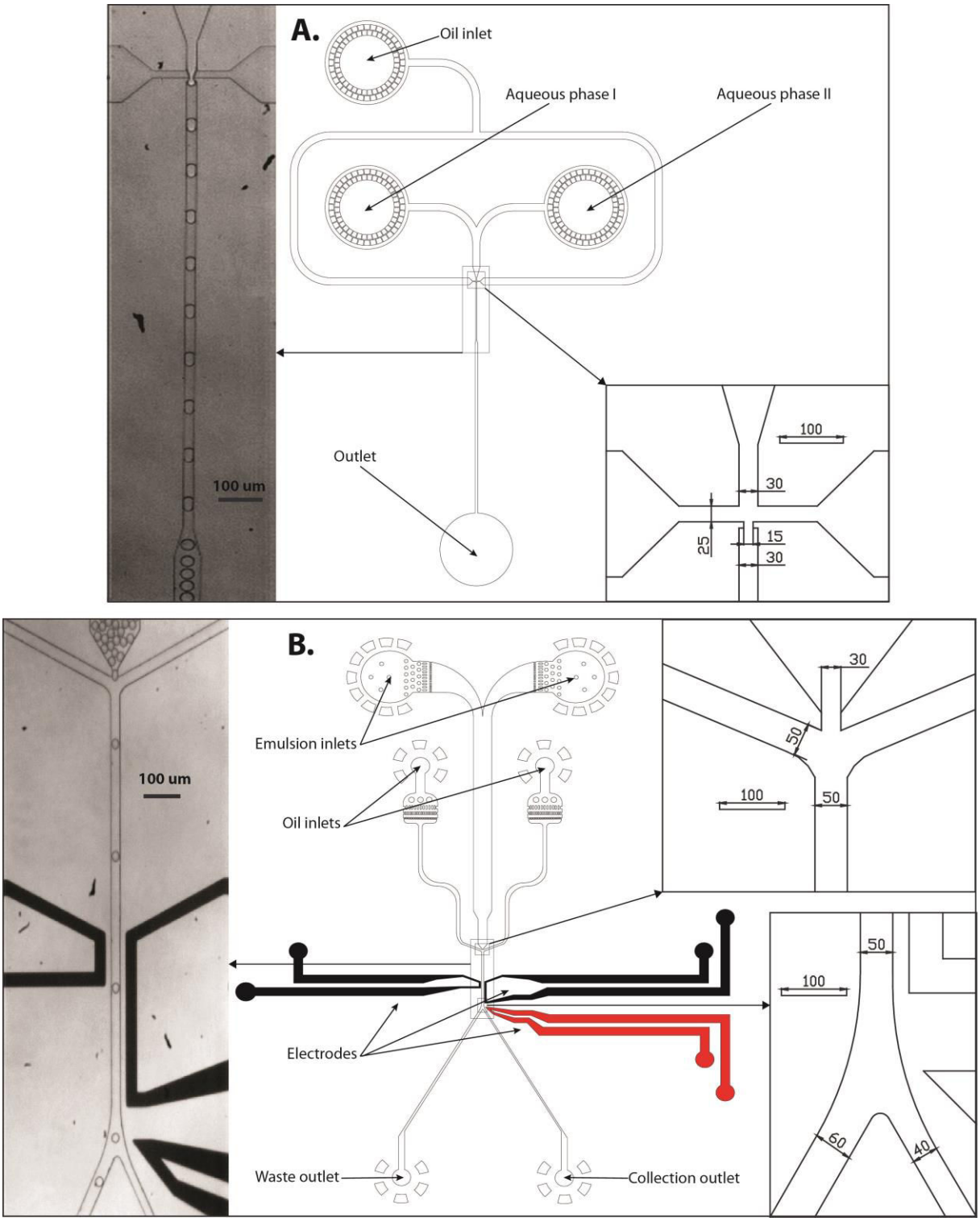

Figure 30. A. Microfluidic device for coencapsulation of parallel aqueous streams. *Depth 20 µm.* **B. Microfluidic device for FADS of reinjected emulsion**. *Depth 25 µm. The dimensions of all device features are indicated in µm.*

II. Results and discussion

1. Fluorogenic substrates for retro-aldolase assay in droplets

The ideal substrate to use for the evolution of an enzyme activity is the substrate specific to actual activity. However, it is usually not possible to develop a working high throughput assay based on the actual substrate. With this regard, it can be assumed that the closer a substrate analog mimics the actual substrate, the more likely it is to improve an appropriate biocatalyst. The substrate and the corresponding product must fit several criteria to be suitable for droplet-based screening of retro-aldolase variants:

- ✓ High sensitivity and specificity to the enzyme
- ✓ Appropriate physicochemical properties
  - ➢ Stability
  - ➢ Good solubility in water
- ✓ Availability; commercial or from synthesis
- ✓ Suitability for microfluidic screening
  - ➢ Compatibility with microfluidic fluorescent detection system
  - ➢ Behavior in droplet environment

The substrate conversion must specifically correspond to activity of the enzyme acting upon this substrate. At the same time, the substrate should be chemically stable during storage and possess good solubility in water to achieve the desired concentration in assay solution. Droplet-based microfluidics does not require large amounts of substrate; nevertheless it is important to have constant supply of it, commercially or by synthesis.

All the conditions described above are applied to almost every substrate used for the screening of an enzyme activity. The implementation of a microfluidic platform imposes more restrictions on substrate properties. The detection setups of microfluidic stations allow the monitoring of fluorescence in droplets, thereby, it would be more convenient to have a fluorescent product with the excitation maximum close to wavelengths of lasers installed on stations (UV: 375 nm; Blue: 488 nm; Green: 532 nm),

and emission maxima corresponding to the spectral intervals of optical filters on PMT's (Blue: 447/55 nm; Green: 525/40 nm; Orange: 607/70 nm).

In droplet-based microfluidics, droplets are regarded as micro-reaction vessels in which separate reactions can be run. To assure the proper compartmentalization, the exchange of content between droplets should be avoided. In particular, for enzyme activity screening, it is important to limit the exchange of fluorescent product.

Miniaturization increases the influence of the interface forces on the physico-chemical processes in the droplets. The adsorption of the substrate to the droplet walls could reduce the amount available for enzyme cleavage. At the same time, if the product sticks to the interface, this can modify recorded signal from the droplet.

The standard fluorescent assay for the retro-aldol reaction was based on the conversion of commercially available methodol **S1** (Figure 31) into the corresponding naphthaldehyde **P1** (λ_{exc} = 330 nm, λ_{em} = 452 nm). The excitation maximum of fluorophore **P1** is quite far from the wavelength of the 375nm-UV laser, making methodol unsuitable for the microfluidic assay. By contrast, the exchange of oxygen by nitrogen in position 6 of naphthaldehyde, shifts the excitation maximum towards the emission of the UV-laser area λ_{exc} = 364 nm and the fluorescence can be recorded with green PMT λ_{em} = 531 nm. The aldol sensor system **S2/P2** has already been successfully applied for the antibody catalysis assay [220].

The fluorogenic assay using droplet-based microfluidics necessitates the fluorescent product remaining confined in droplets until the assay read-out. The encapsulation of **P2** in 20 pL droplets using 2% EA surfactant in HFE 7500 (hydrocarbon and fluorocarbon fragments) and FC 40 (only fluorocarbon chains) has shown no signal in drops. Eventually, all of the fluorophore disperses from the droplet phase into the oil phase in several minutes. The effect can be explained by the hydrophobicity of **P2** [121]. The partition coefficient predicted for **P2** using ACD/Laboratories software is logD = 2.98, showing the hydrophobic character of the fluorescent product. For instance, in the same conditions, another commonly used fluorophore - resorufin, leaks from droplets in less than 2 hours. The partition coefficient for **P2** is greater than logD = 2.63 of resorufin, which makes **P2** more hydrophobic, and as a consequence, it is leaking from the droplets even faster.

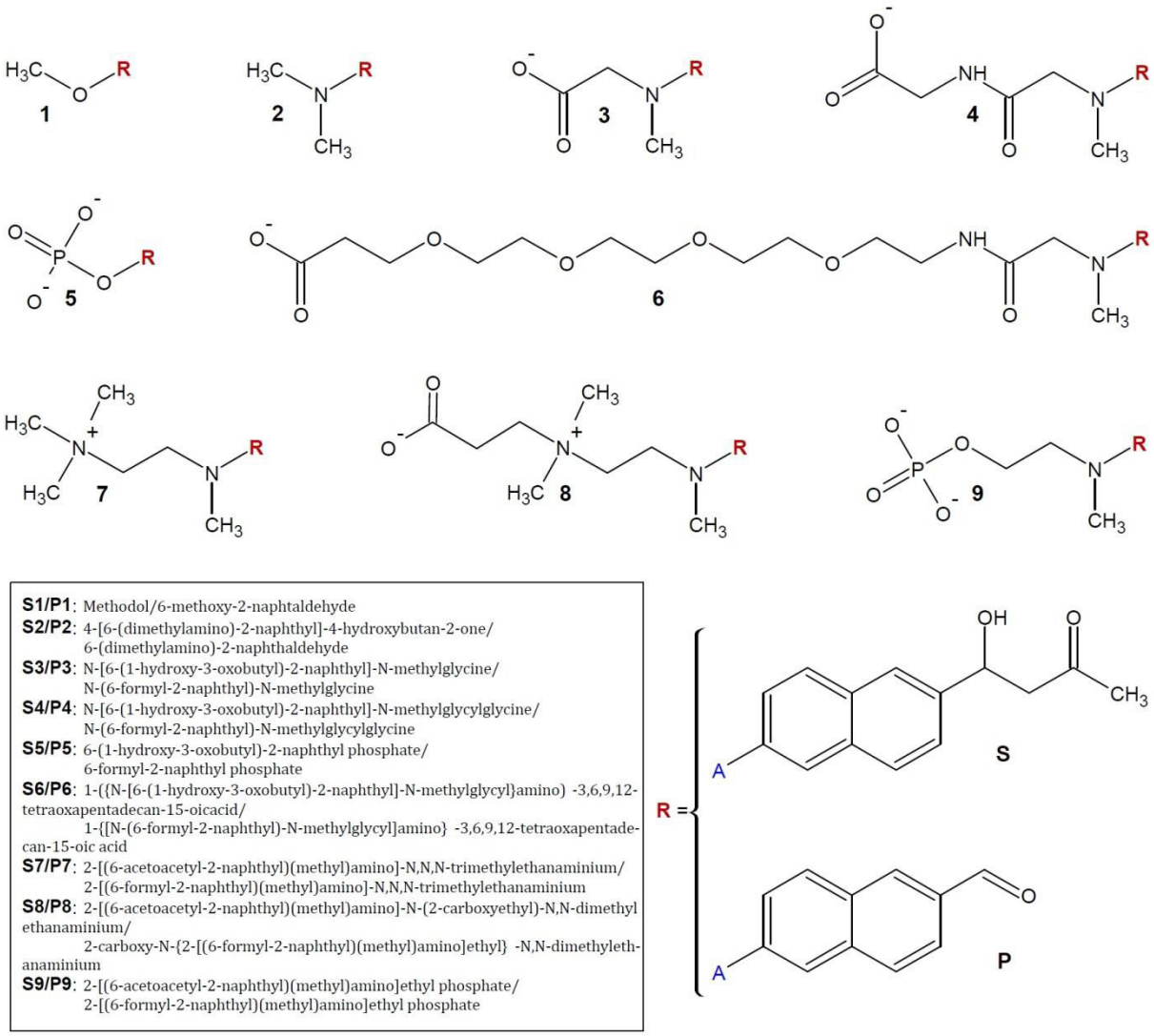

Figure 31. List of potential substrates and corresponding products for retro-aldolase assay.

To improve the retention time of the fluorophore in droplets, the chemical structure needs to be modified by attaching hydrophilic group. An identical issue was solved in the work of Woronoff *et al.* [122], switching hydrophobic aminocoumarin-based substrates to hydrophilic ones by attaching a sulfonic group. To limit the exchange between droplets, a series of substrates and complementary fluorescent products (Figure 31) have been synthesized by Richard Obexer and supplied for leakage tests.

The microfluidic workflow used for retention time characterization is represented as a combination of puzzle pieces (Figure 32), each piece reflecting a microfluidic operation (Figure 19).

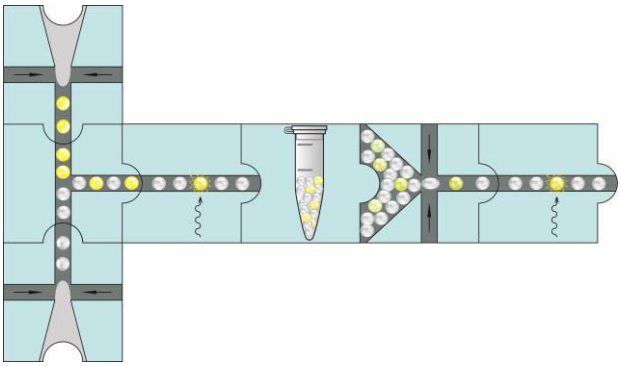

Figure 32. Microfluidic workflow for retention time characterization. *Two droplet populations with different concentration of naphthaldehyde are generated, mixed, analyzed for detection of initial fluorescent distribution, collected and stored off-chip, reinjected and analyzed after several hours.*

In order to be able to screen for enzymatic activity, the reaction product, in our case the naphthaldehyde, must stay confined in droplets at least several hours. The droplet population distribution was analyzed up to 8-10 h.

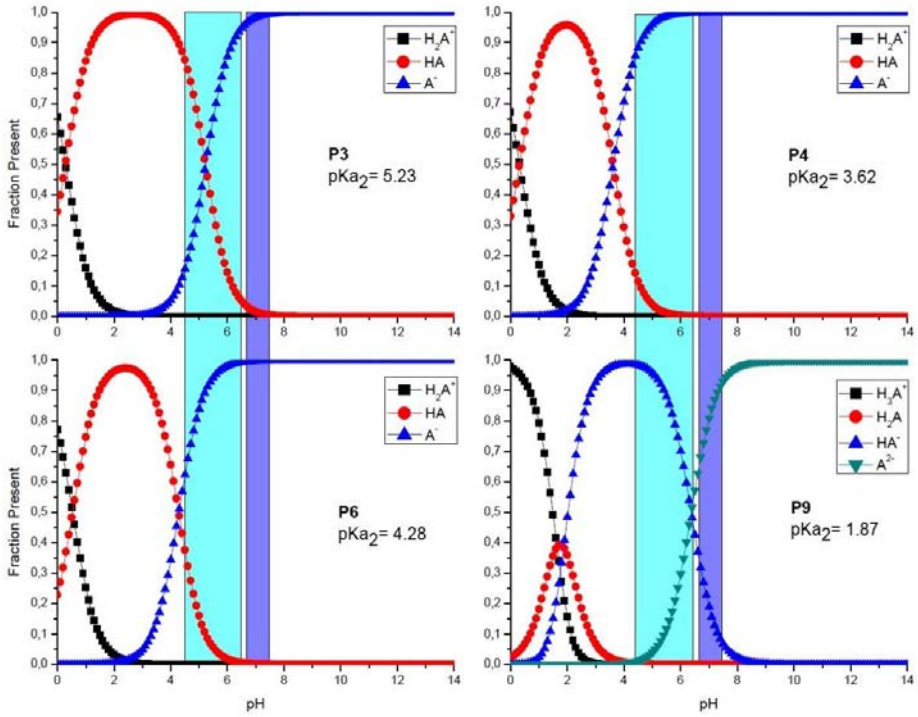

Figure 33. pH-distribution diagrams for aldehydes P3 ,P4, P6 and P9. *The diagrams are representing the protonation states of the aldehydes, depending on the pH. The curves were calculated using ACDlabs software.*

Leakage experiments (Figure 34) with aqueous solutions of compounds **P3**, **P4** and **P6,** whose solubility was increased by adding carboxylic groups, have shown a dependence of retention time on the pKa of the carboxylic group. The aldehydes **P4** and

P6 display pKa_2 lower than pH = 5 and show better retention in droplets than **P3**, with pKa_2 = 5.2. From the pH-distribution diagrams (Figure 33) it can be stated, that at the pH of water, a lot of **P3** is in HA non-ionic form, while **P4** and **P6** are predominantly represented as deprotonated species. Thus, the non-ionic form of the aldehyde is responsible for diffusing into the carrier oil [221] towards droplets with lower concentration.

In this regard, an aldehyde which is in its ionic form in all pH-ranges might serve as a universal fluorescent marker for retro-aldolase activity. However, retention time experiments of tertiary ammonium compound **P7** in distilled water droplets showed leakage within fewer than eight hours. The experiments on **P7** indicate that leakage can take place not only through diffusion but also via the micellar transport [222, 223]. When the same experiment on **P7** was performed in buffer, the two droplet populations held their initial positions, even after 8 hours of incubation. It seems that the ionic force created by the presence of buffer salts limits micelle formation from the droplet interface.

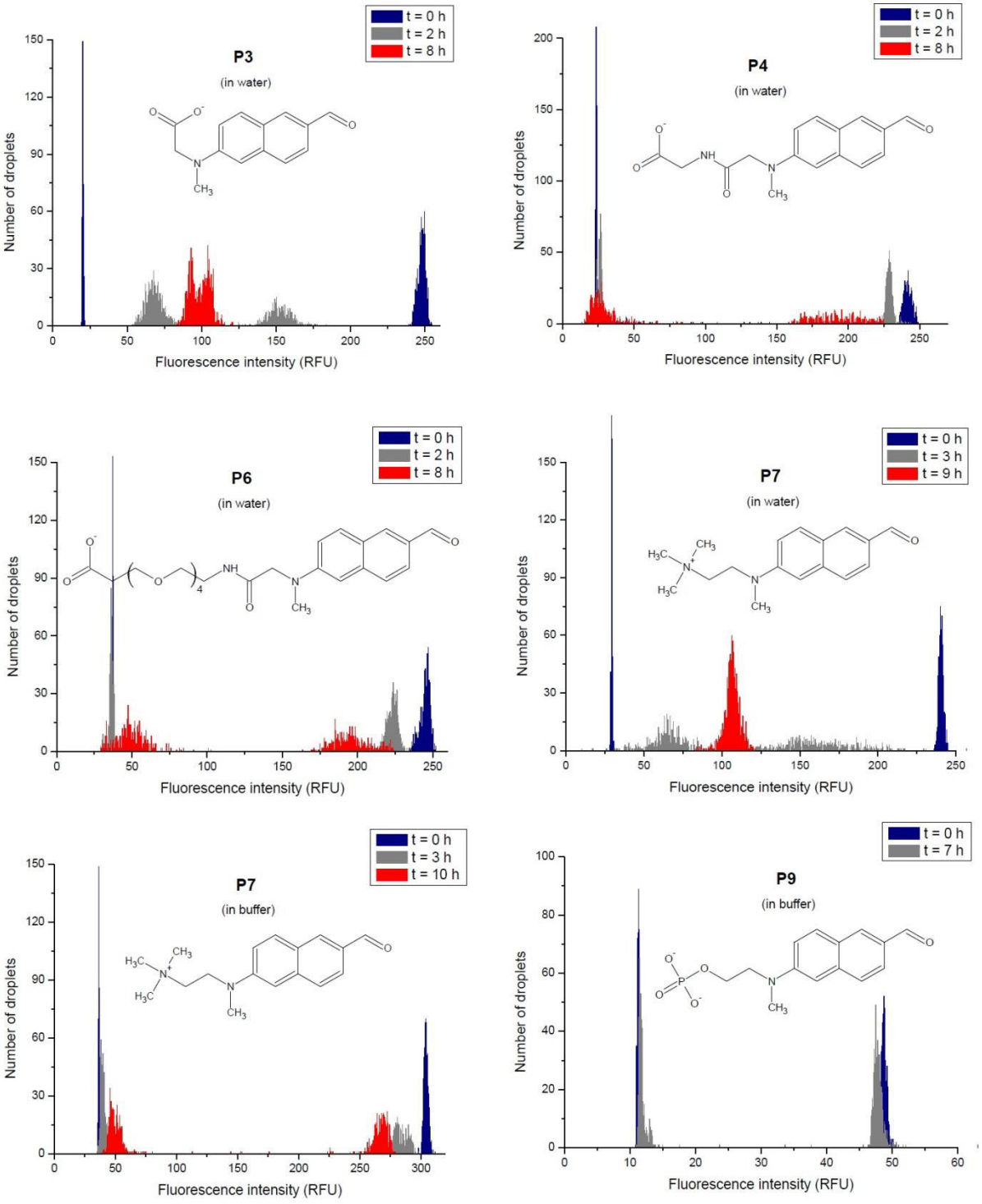

Figure 34. Histograms for retention time characterization. *Two populations of droplets containing 50 μM and 500 μM of certain naphthaldehyde in water or in buffer were mixed together and reinjected in different time intervals to monitor their fluorescence.*

Table 2. Comparison of substrate/product pairs for retro-aldolase screening suitability.

	Catalytical Caracteristics	Availability Aldol	Availability Aldehyde	Solubility	Microfluidic Detection	Retention Time	Suitability for Screening
S1 P1	Ok: kcat = 0.18 1/s Km = 400 µM	Synthesis Yield: 40%	Commercially	Low Requires co-solvent	No: λ_{exc} << UV laser	Not tested	No
S2 P2	Ok: kcat = 0.05 1/s Km = 100 µM	Synthesis Yield: 10%	Synthesis Yield: 40%	Low Requires co-solvent	Ok: λ_{exc} ~ UV laser	Goes into the oil phase	No
S3 P3	N/A	N/A	Synthesis Yield: 5%	Good	Ok: λ_{exc} ~ UV laser	$t_{1/2}$ ~ 2 h in water, probably, does not leak in buffer	No
S4 P4	N/A	N/A	Synthesis Yield: Traces	Good	Ok: λ_{exc} ~ UV laser	Does not leak $t_{1/2}$ > 8h	No
S5 P5	N/A	Synthesis Yield: 10%	Synthesis Yield: 30%	Good	No: λ_{exc} << UV laser	Not tested	No
S6 P6	N/A	N/A	Synthesis Yield: Traces	Good	Ok: λ_{exc} ~ UV laser	Does not leak $t_{1/2}$ > 8h	No
S7 P7	Ok: kcat = 0.07 1/s Km = 220 µM	Synthesis Yield: 10%	Synthesis Yield: 20%	Good	Ok: λ_{exc} ~ UV laser	$t_{1/2}$ < 3 h in water, does not leak in buffer $t_{1/2}$ > 10h	Ok
S8 P8	N/A	N/A	Synthesis Yield: 10%	Good	Ok: λ_{exc} ~ UV laser	Not tested	N/A
S9 P9	Low catalytic activity	Synthesis Yield: 10%	Synthesis Yield: 20%	Good	Ok: λ_{exc} ~ UV laser	Does not leak $t_{1/2}$ > 7h	No

In Table 2, all aldol/aldehyde (S/P) pairs are compared for choosing the optimal retro-aldolase substrate for assay in droplets. Highlighted in red are the conditions, which are impeding a certain S/P pair to be suitable for the screening. As has been mentioned before, the standard methodol (S1/P1) assay is not compatible with available optic setups on stations, nor is the S5/P5 pair. The hydrophobic P2 is partitions almost entirely into the fluorinated oil phase eliminating the pair S2/P2 from RA screening in droplets. Despite their good retention time, the carboxylic-based aldols (S3, S4, S6) were difficult to synthesize and purify. The phosphate-based substrate S9 has shown a low activity with several RA clones, explained by a negatively charged group which prevents the substrate from entering the enzyme's active site. However, S9/P9 could be used for the evolution of a retro-aldolase that tolerates negatively charged substrates.

In conclusion, the substrate/product pair that fits all the criteria stated in the beginning is the ammonium salt-based S7/P7 pair.

2. Retro-aldolase activity and kinetics in droplets

Once the optimal substrate had been established, retro-aldolase activity in droplets was probed thereafter on a microfluidic platform. First, purified enzyme and ammonium substrate **S7** were coencapsulated in picoliter-sized droplets. After incubation, the emulsion turned fluorescent, proving that the substrate was converted inside the droplets.

To simplify protein evolution experiments and microfluidic operations, a cell-based expression system was chosen. Therefore, *E.coli* BL21-Gold strain expressing retro-aldolase in the cytoplasm was chosen for protein production, since this expression system has been previously established and optimized for high protein yields. High protein concentration is required to get a clear signal over noise within a reasonable time lasting in 0.5-2 h interval. Hence other expression systems, including periplasmatic export, failed.

Efficient analysis and resolution of clones during the directed evolution rounds depends on the precision of the assay readout. For that reason, time courses of RA reactions were measured using the integrated module described in: "Materials and methods. Kinetics in droplets". RA-95.5-8 was expressed in *E. coli* grown in liquid

culture. In one experiment, the cells were diluted to the concentration corresponding to one cell per one droplet with average occupancy (λ = 1). Cells were then lysed prior encapsulation in monodisperse droplets together with substrate and assay buffer. In the microfluidic device for kinetic monitoring (Figure 29), droplets passed through the delay-line with multiple constriction points, where individual droplets were analyzed by laser-induced fluorescence. Plotting the fluorescence intensity over the time (Figure 35) reveals a reaction slope decrease after 25 minutes, at only 25 % substrate conversion. This observation can be rationalized by the fact that the enzyme preferentially reacts with (R)-methodol [Richard Obexer, personal communication], and hence the 'real' substrate concentration is only 50%, since racemic substrate was used in this experiment.

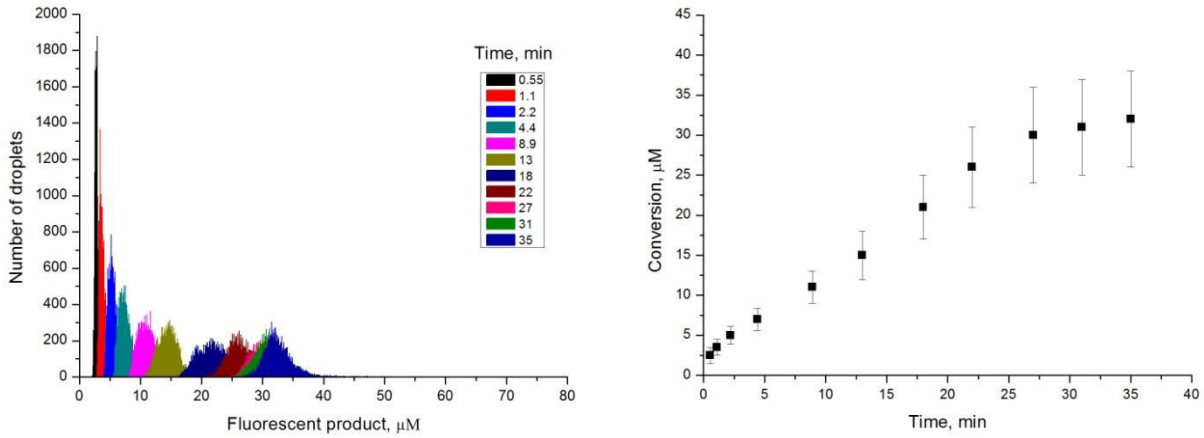

Figure 35. Retro-aldolase kinetics of cell lysate expressing RA95.5-8 (lysis before encapsulation).

In another similar experiment, instead of cell lysate, intact, single cells were diluted to λ = 0.1 and co-emulsified with lysis agents and substrate. According to a Poisson distribution, low occupancy reduces the probability to have multiple cells in one drop. In this case, 90 % of droplets will be empty, 9 % will contain one single cell, and 1% multiple co-encapsulations. The same device was used to follow the time course of retro-aldol activity of single cells (Figure 36). From the droplet fluorescence distribution at each kinetic point, it is possible to monitor over time the gradual differentiation of the droplet population containing cells from the empty droplet population. Since the

distribution of activity within the same cell population is widely spread (Figure 36. B), the kinetic plotting on a graph is not possible. In order to see if the reaction reached the plateau, the emulsion was collected off-chip after passing through the kinetic delay-line, and reinjected after 18 hours. At this time-point, the cell-containing population of droplets was completely separated from the empty droplets. The substrate conversion achieved at this point was 60 %, much higher compared to 25 % after 25 min. This proves that the retro-aldol reaction continues after half an hour, but at slower rates.

The time-dependent assay of cells lysed before encapsulation (Figure 36. A) reflects the average kinetics of the whole population of *E. coli* expressing the enzyme. The variation in droplet fluorescence is mainly due to the optical set-up and retention time dispersion between droplets. In contrast, the experiment, in which the cells are lysed within the droplet, makes it possible to monitor the retro-aldolase activity for the whole population at the single cell level (Figure 36. B). These experiments demonstrate how differently single cells can behave. The scattering caused by the optical set-up or the chip are hence negligible, compared with the error caused by the cells.

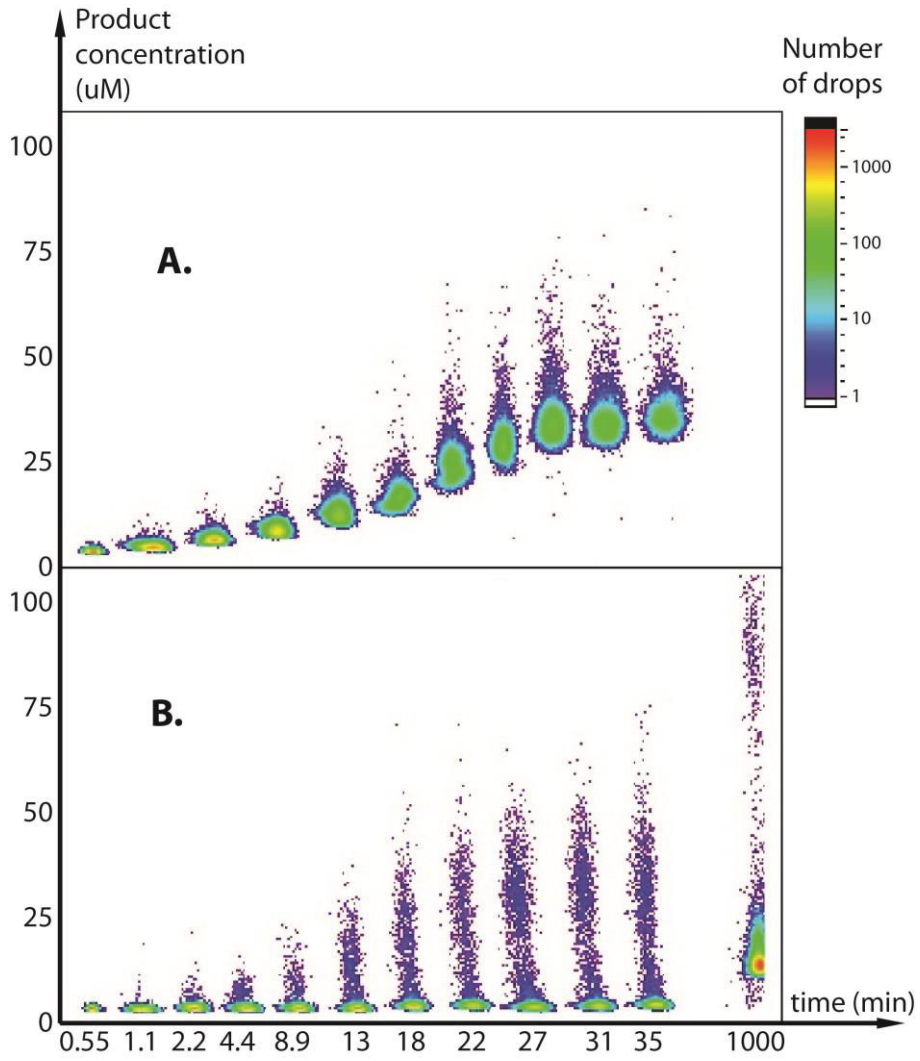

Figure 36. Kinetics of RA assay of cell lysates in droplets. A. *The cells were diluted to correspond to λ = 1 for average occupancy in droplets, lysed and only after, coencapsulated with assay mixture.* **B.** *The cells were diluted to correspond to λ = 0.1 for average occupancy in droplets, coencapsulated with assay mixture and lysis agents, so the lysis takes place in droplets.*

In conclusion, the retro-aldol kinetic reactions in droplets have shown a wide dispersion of activity within the single population of cells expressing the same variant. Nevertheless, the separation of the active cell-containing droplet population starts at 20 min. The optimal incubation time for initial rounds of RA library screening is lasting in the interval of hours, the same range as of off-chip emulsion incubation.

3. Validation of full RA directed evolution cycle

To demonstrate the feasibility of the microfluidic retro-aldolase screening platform three RA95.5 mutants with different activities were chosen: RA95.5-8 (k_{cat}/K_m = 48600 M^{-1} min^{-1}), RA95.5-5 (k_{cat}/K_m = 17800 M^{-1} min^{-1}) and RA95.5-5k/o (k_{cat}/K_m = 5.1 M^{-1} min^{-1}). The proteins were expressed in cells separately cultured in growth media. The cells were diluted to achieve $\lambda \sim 0.3$ and mixed 1:1:1 prior to encapsulation with substrate and lysis mixture (Figure 30. A), as described in "Material and methods. Microfluidic assay with off-chip incubation". The droplets were collected off chip in a syringe filled with oil and stored for 3h at 37 °C for reaction to occur. Following incubation, droplets were reinjected into sorting device for fluorescence-activated droplet sorting (Figure 30. B).

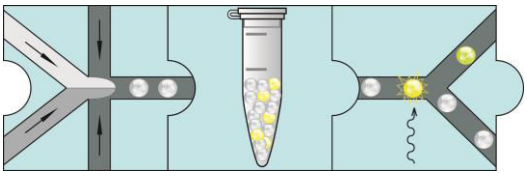

Figure 37. Microfluidic workflow for RA screening with off-chip incubation.

The distribution of droplet fluorescence indicates the overlapping of populations containing different RA variants; nevertheless it is possible to distinguish the peaks corresponding to each clone (Figure 38). The sorting gate was set for the top 5 % most active fraction aiming to enrich for the RA95.5-8 clone.

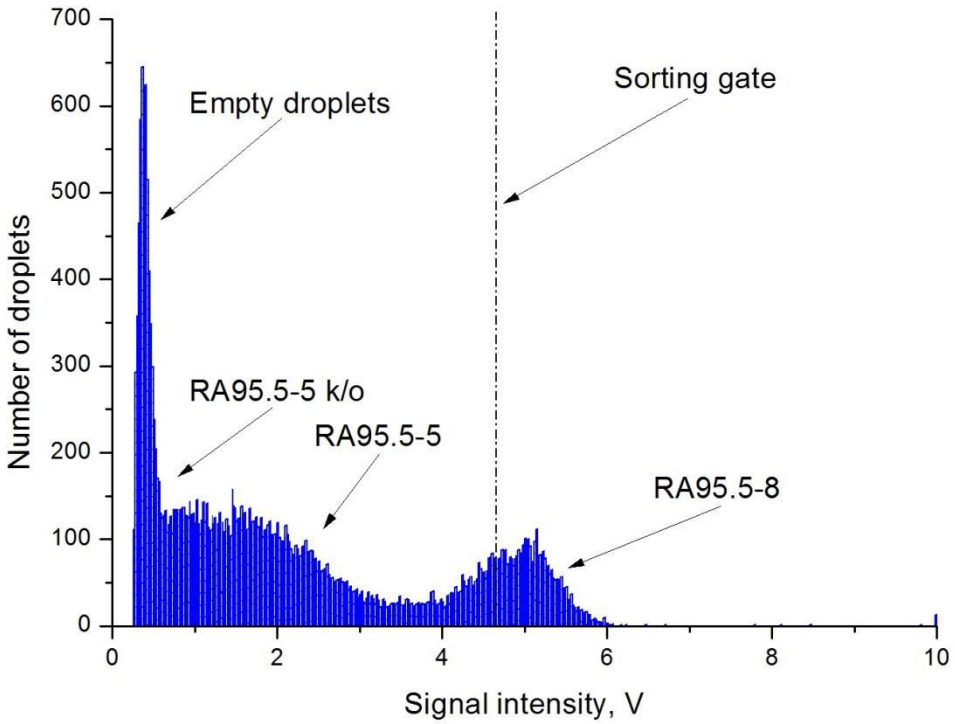

Figure 38. Histogram of retro-aldolase screening of 1:1:1 mixture of clones RA95.5-8 : RA95.5-5 : RA95.5-5k/o. *Droplet population corresponding to cells expressing the most active variant RA955-8 can be mpnitored, although is not well discriminated from the less active variant RA95.5-5 The sorting gate was set to specifically enrich for the most active RA95.5-8 clone.*

After several hours of screening, the collected emulsion was broken and DNA was isolated. The sorted DNA was subjected to restriction digestion along with the corresponding controls (RA95.5-8, RA95.5-5, RA95.5-5k/o). To distinguish all three sequences two enzymes were required: ApoI distinguishes between RA95.5-8 and the other two sequences, whereas BsrI identifies RA95.5-5k/o. For separation and identification of the digested fragments agarose gel electrophoresis was performed.

Figure 39. The restriction digestion pattern of chosen mutants.

The restriction digests indicate that the sorted DNA sample mainly contains RA95.5-8 (Figure 40). The content of the inactive RA95.5-5k/o variant cannot be exactly determined since the 249 bp band of RA95.5-5k/o overlaps with the "partial digest band" of RA95.5-8 and RA95.5-5 (207 bp + 42 bp). The ApoI digestion pattern, however, shows that only traces of RA95.5-5k/o and RA95.5-5 are present in the sorted DNA.

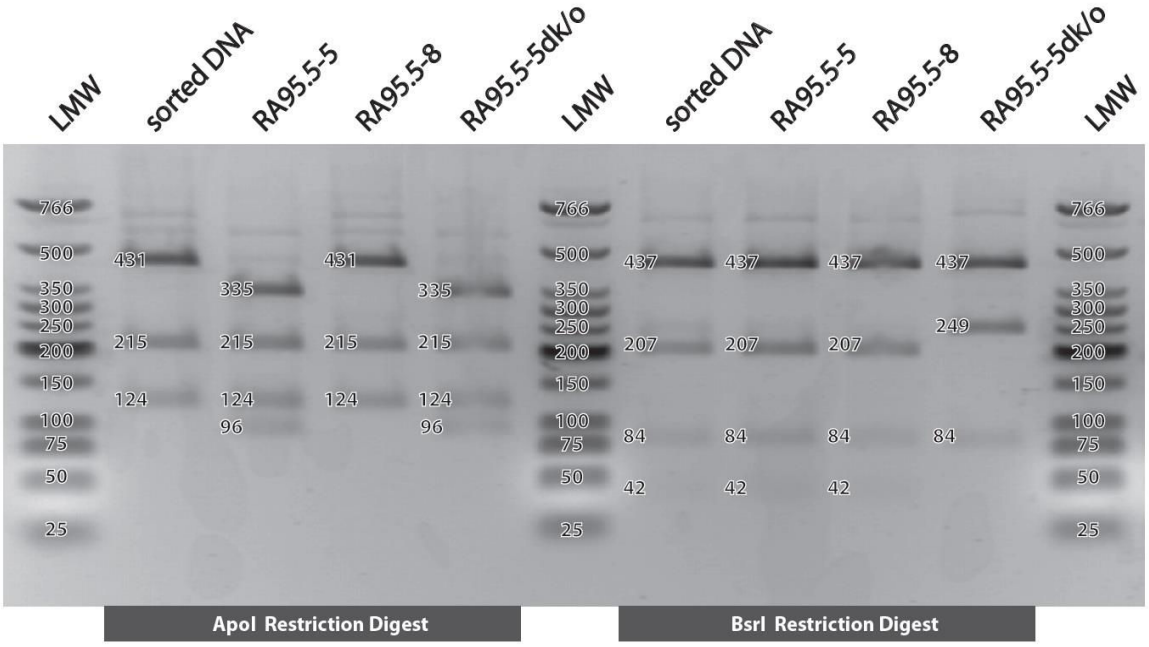

Figure 40. Agarose gel of digested DNA. *Unlabeled bands represent the whole gene and partial digests.*

In conclusion, we were able to collect almost exclusively the most active mutant and thus strongly enriched this population. Given the rather small difference in catalytic activity between RA95.5-5 and RA95.5-8 (3-fold) and rather broad distribution of catalytic activity within one population (shown as well in the previous kinetic experiment) the system seems to be feasible for directed evolution cycles of RAs.

4. Development of integrated microfluidic device

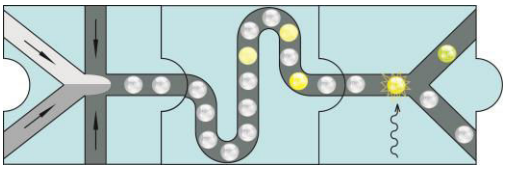

Figure 41. Microfluidic workflow for RA screening with on-chip incubation.

The microfluidic platform used for the first round of RA directed evolution consists of off-chip droplet incubation for 1-3 hours. This platform is suitable for the screening of variants whose catalytic activity is not high enough to reach the maximum substrate conversion in the given time interval. However, in the course of directed evolution experiments, the catalytic properties of clones will be eventually improved. In this case, to preserve the discrimination of library variants in the initial reaction period, it is necessary to shorten the emulsion incubation interval. Off-chip emulsion collection, storage and re-injection induce dispersion in the time which droplets spend in the tubing and syringe. Thus, the platform described above is not applicable for shorter incubation times which are needed for the screening of highly-active enzyme variants. This problem can be overcome by using an integrated microfluidic device consisting of three modules: a co-flow droplet generation module, a 20-45 min delay-line and a sorting module.

Since sorting conditions are not affected by the presence of the delay-line, the FADS module was kept as close as possible to the device from the original microfluidic platform. The delay-line was designed according to the work of Frenz *et al.* [171]. Multiple constrictions along the delay-line were introduced to reduce the Taylor dispersion of droplets within the channels.

The most challenging part of integrated device development is the droplet production module. The backpressure from the delay-line significantly influences the frequency of droplet formation, droplet size and spacing between droplets as well as flow stability. In the platform applied for the first round of selection, a co-flow device consisting of a flow-focusing cross (FF) was used. Such a configuration allows high droplets frequency generation (>1kHz) applying high flow rates of carrier oil (300 µL/hr) and aqueous phases (100 µL/hr each). These characteristics are not compatible with integrated device requirements. Since the initial FADS device operates up to 600 Hz

to avoid sorting errors, the droplet maker should not exceed this limit. At the same time high total flow rate causes an increase in pressure leading to device delamination or fluid leakage through the connections of inlets. Dropping down the total flow rate solved these issues.

At low capillary numbers (low flow rates) FF configuration leads to unstable polydisperse drop formation (Figure 8). By contrast, T-junction (TJ) and pinned-jet-flow focusing (PJFF) form monodisperse droplets under these conditions [106]. PJFF configuration can be adjusted for co-flow droplet generation. In addition to stable droplet formation at low capillary numbers, PJFF-like co-flow configuration allows the aqueous phases to come into contact just before droplet generation. This way the reaction starts only in droplets resulting in reliable compartmentalization.

Taking into consideration the reasoning from above, a three-layer integrated device was designed and experimentally optimized. The first, 18 μm-deep layer consists of a PJFF-like co-flow droplet maker; the second - 25 μm-deep layer is the sorting module; the third - 75 μm-deep layer is the delay-line having the overall volume aprox. 33 μL.

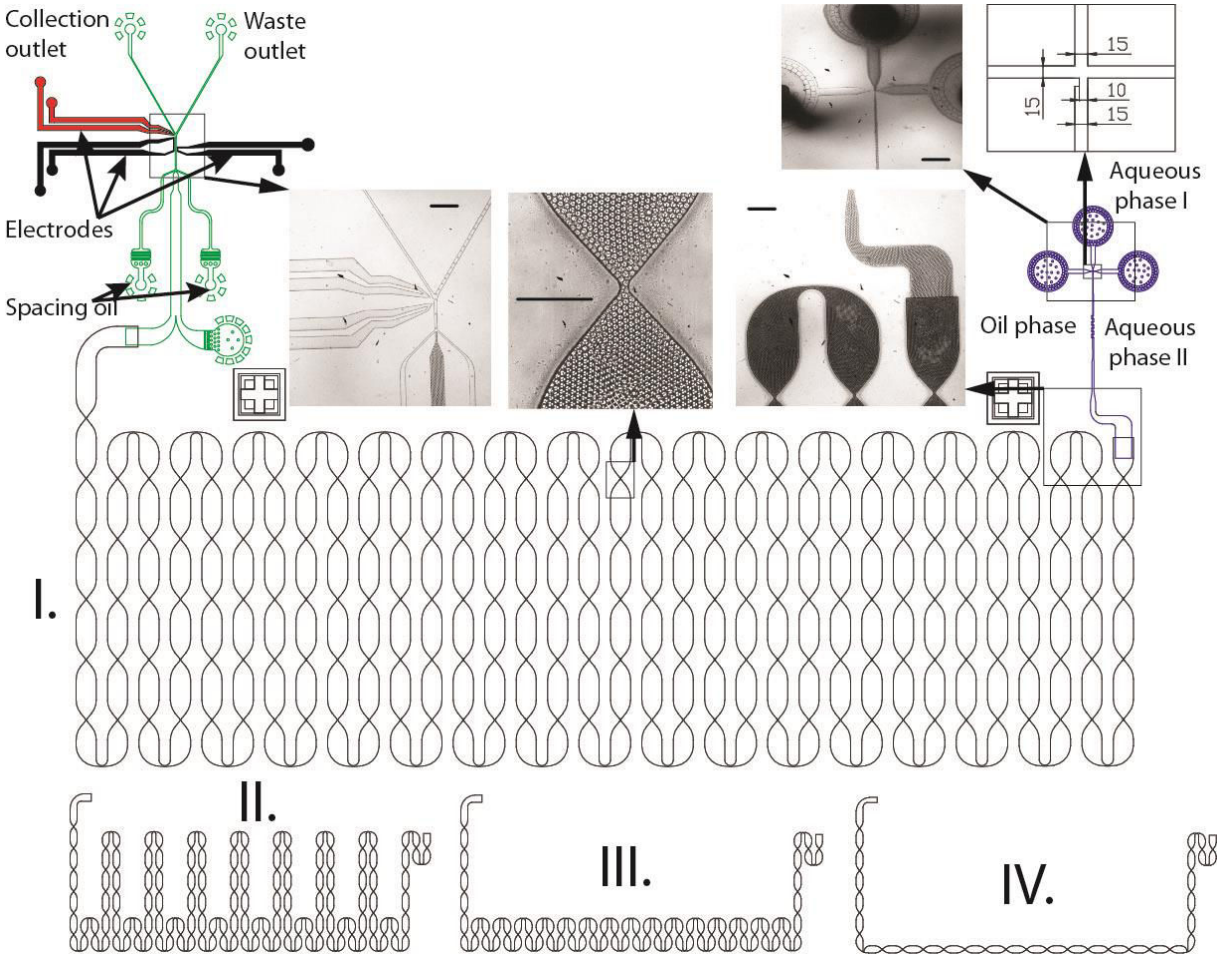

Figure 42. Integrated device overview. *Blue layer (18 µm-deep) consists of PJFF-like co-flow droplet maker; green layer (25 µm-deep) - the sorting module; black layer (75 µm-deep) is the delay-line. Fabrication of devices where the delay-line* ***I*** *is substituted with* ***II***, ***III*** *or* ***IV*** *allows the reduction of incubation time from 45 minutes to 30, 16 or 7 minutes under the same flow-rate conditions: Aqueous phase I = Aqueous phase II = Oil phase = 15 µL/hr.*

Since the overall flow-rate of dispersed and continuous phase injected into the delay-lines is 45 µL/hr, the time for each droplet from the generation nozzle to the sorting module becomes: 33 µL /45 µL/hr = 45 minutes. Thus, using this microfluidic device it is possible to bridge the gap between off-chip incubation (< 50 min) and on-chip droplets incubation (> 10 min), prior high-throughput sorting.

Under the conditions of oil deficiency, when the flow-rate of the dispersed phase is higher than the flow-rate of continuous phase, newly formed droplets are moving in one layer close to each other. When entering the deeper delay-line, droplets are repacking in a three-layer crystal-like highest density arrangement, which allows the

emulsion to flow as a jammed pack. In this case droplets are moving forward to the sorting area with a constant velocity. The integration of smaller delay-lines makes the microfluidic chip even easier to operate, since the back-pressure drops together with the decrease of the length of the line.

The experimental proof of integrated module efficiency was obtained during sorting of the same library. Two rounds of selection were run in parallel with off-chip incubation and using the on-chip technique. In the case of the integrated module (Figure 43. B), the population corresponding to occupied droplets was shifted in a more active area and had a more narrow distribution, while after off-chip rounds of selection the population distribution was quite spread (Figure 43. A).

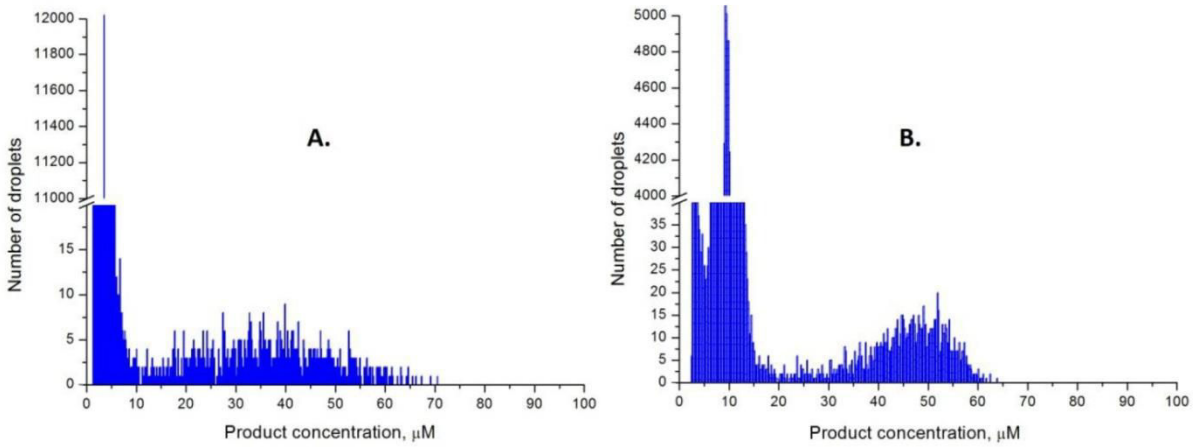

Figure 43. Distribution histograms obtained after two selection rounds of RA95.5-b4220 library. *A: off-chip incubation. B: using integrated chip.*

To summarize, a series of integrated chips have been designed and optimized for the screening of RA variants possessing high activities. On a single chip, cells can be encapsulated with assay components, incubated for 45, 30, 16 or 7 minutes and sorted at 500 Hz frequencies for the improved activity.

5. Retro-aldolase library screening

Since among other clones, RA95.5-8 displays higher catalytic parameters and better evolvability, it has been decided to use this retro-aldolase variant as a starting point for directed evolution in droplets.

In the course of this work, the evolved retro-aldolase RA95.5-8 was further improved over another three generations by directed evolution. The main strategy is outlined in Figure 44, and was based on employing (sorted) populations of active mutants for the generation of new libraries. The idea behind this approach was to mimic natural evolution more closely and to maintain diversity, in order to converge classical directed evolution with neutral drift experiments.

To generate genetic diversity, a large variety of mutation strategies were employed: error-prone PCR, cassette mutagenesis, DNA shuffling and back-shuffling (Figure 44). Each DNA library was transformed into cells and encapsulated ($\lambda \sim 0.05$-1) with substrate and lysis agents using an off-chip microfluidic platform or the integrated chip. Due to the unavailability of the integrated device, the first screening rounds were done employing the off-chip incubation method. The sorting gate was set to select the top 0.1-1% of the most active droplet population from each library. Two to three consecutive rounds of selection were performed before mutants were analysed in a plate assay and prior to introducing new mutations. To streamline and speed up this very linear process, several libraries were evolved in parallel.

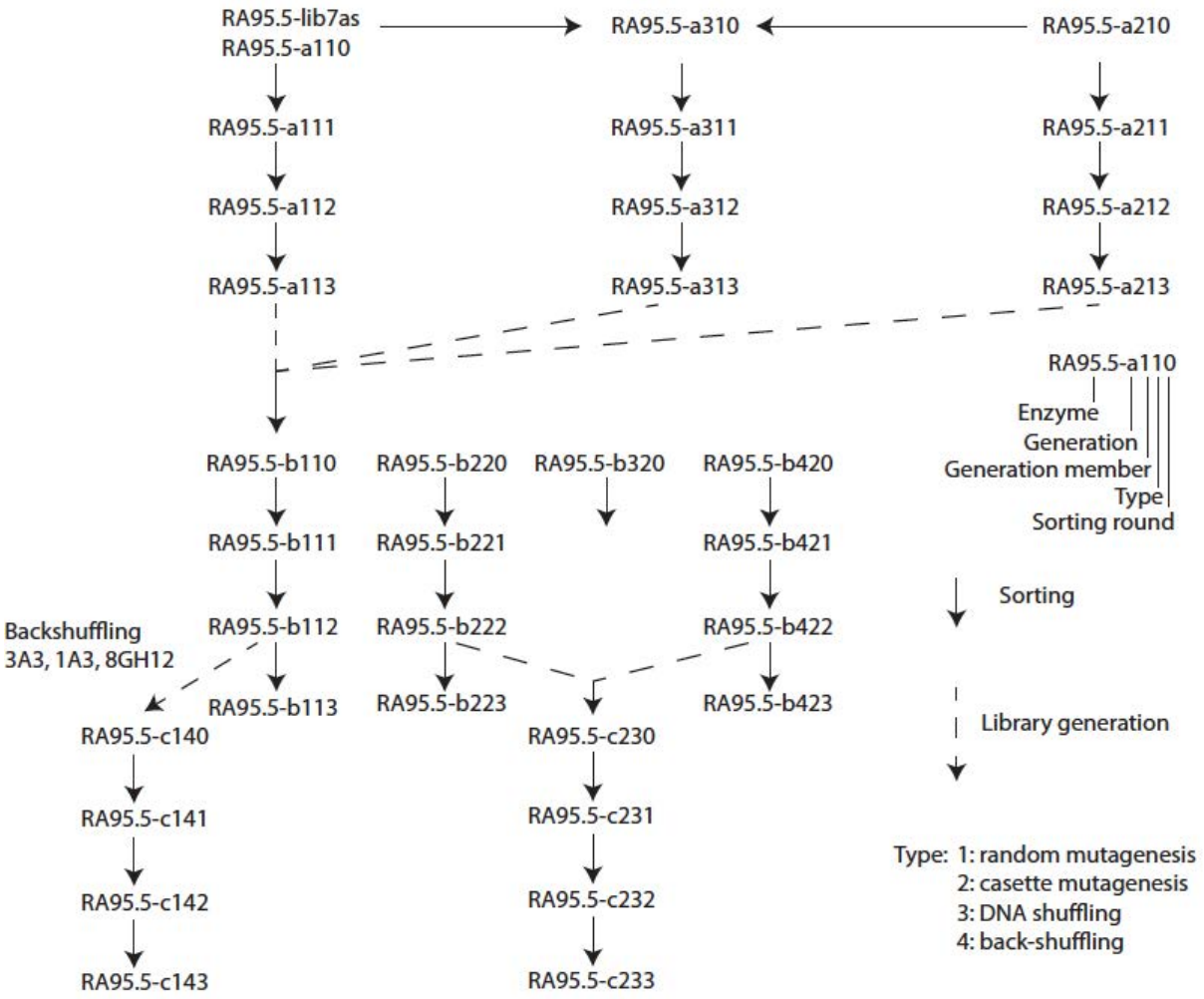

Figure 44. Directed evolution rounds of RA using droplet-based microfluidics.

In the first generation, two random mutatgenesis libraries were analysed. RA95.5-a110 is identical to the library RA95.5-lib8a(s) that was previously analysed by plate assay, to isolate the RA95.5-8 mutant. The rationale behind rescreening this library using the microfluidic assay was for proof of principle. It was hence expected, that clones with the same or higher activity than RA95.5-8 should be identified. Crude lysate activity measurements in 96-well plates revealed that active mutants were successfully enriched over three rounds of sorting. Two mutants, RA95.5-a113 BA1 and RA95.5-a113 BA3, showed a 1.8 and 1.9 fold activity increase over RA95.5-8 in the crude lysate assay. In parallel, a random mutagenesis library based on RA95.5-8 and a 1:1 mixture of both libraries (RA95.5-a310) was sorted and analysed. As was the case for the RA95.5-a110 library, enrichment was observable; however no significantly improved clone was isolated from these libraries.

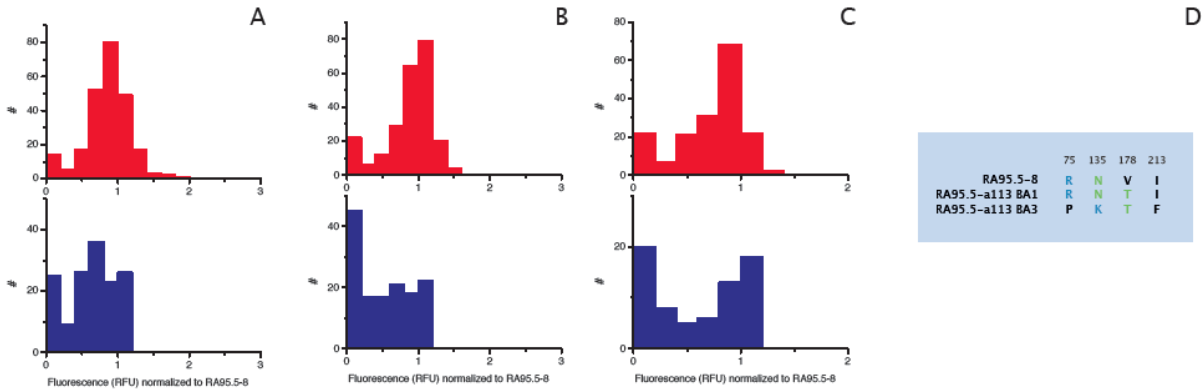

Figure 45. Histograms of 96-well plate assay. *Indicated in blue is the unsorted library, and shown in red is the activity distribution after three rounds of sorting. A: RA95.5-a110 library, B: RA95.5-a310 library, C: RA95.5-a210 library. D: Alignment of new mutations identified in the best clones from this generation.*

Thus, it was assumed, that the activities of those clones were all in a similar range, which is basically smaller than the activity scattering caused intrinsically by using single cells. This would hamper discrimination of clones, and makes sorting a statistical process. To circumvent this problem, it was attempted to increase the mutational load and therefore create a stronger discrimination between active and less active clones.

Therefore, a semi-rational approach to systematically identify beneficial mutations was applied. For this purpose, libraries in some loop regions were extensively mutagenized by cassette mutagenesis. In addition to the loop libraries RA95.5-b220 (residues 56-60) and RA95.5-b420 (residues 183-186 and 210) (Figure 47. B), which are located on the most flexible loops, another library RA95.5-b320 (85- 89) was created on a very rigid loop, which however is in very close proximity to the catalytic lysine. The outcome of these libraries showed that in the case of library RA95.5-b420 and RA95.5-b320, these positions were very sensitive to mutations. In the case of library RA95.5-b320, the frequency of droplets containing active mutants was negligible, and hence it was decided to omit this library for sorting. Library RA95.5-b420 was sorted three times and rescreened in a 96-multiwell plate to identify the best mutants. Sequencing however showed that no mutations were acquired in the loop region 183-186, and only K210 was mutated to L and F. This mutation yielded a two-fold increased activity in the crude lysate assay. Concerning library RA95.5-b220, the frequency of active mutants was the highest in comparison to libraries RA95.5-b320 and RA95.5-b420 (Figure 47), which

indicates that this loop is very tolerant to mutations. With regard to the structure, this can be rationalized, since this loop is most distant to the substrate with regard to the other two loop libraries.

Plate screening of the sorted mutants resulted in a number of mutants, showing wild type like activity levels and only a few clones with increased activity. Sequencing of a number of mutants showed that no consensus sequence was however observable and that many new loop sequences were accepted. Although most of these libraries did not yield improved variants, except RA95.5-b420, they shed light on the catalyst itself. The results of these libraries indicate that the loop regions 85-89 and 183-186 were already adapted to support catalysis. Although this strategy was not very successful for the highly evolved RA95.5 variant, it might prove useful for the evolution of non-optimized computational designs.

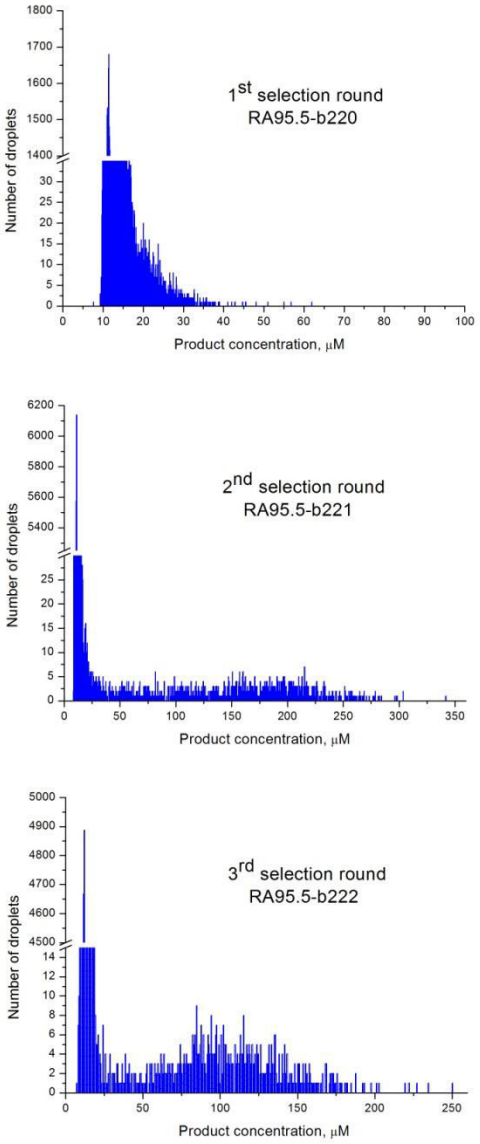

Figure 46. Active population enrichment over evolution rounds. *Not normalized.*

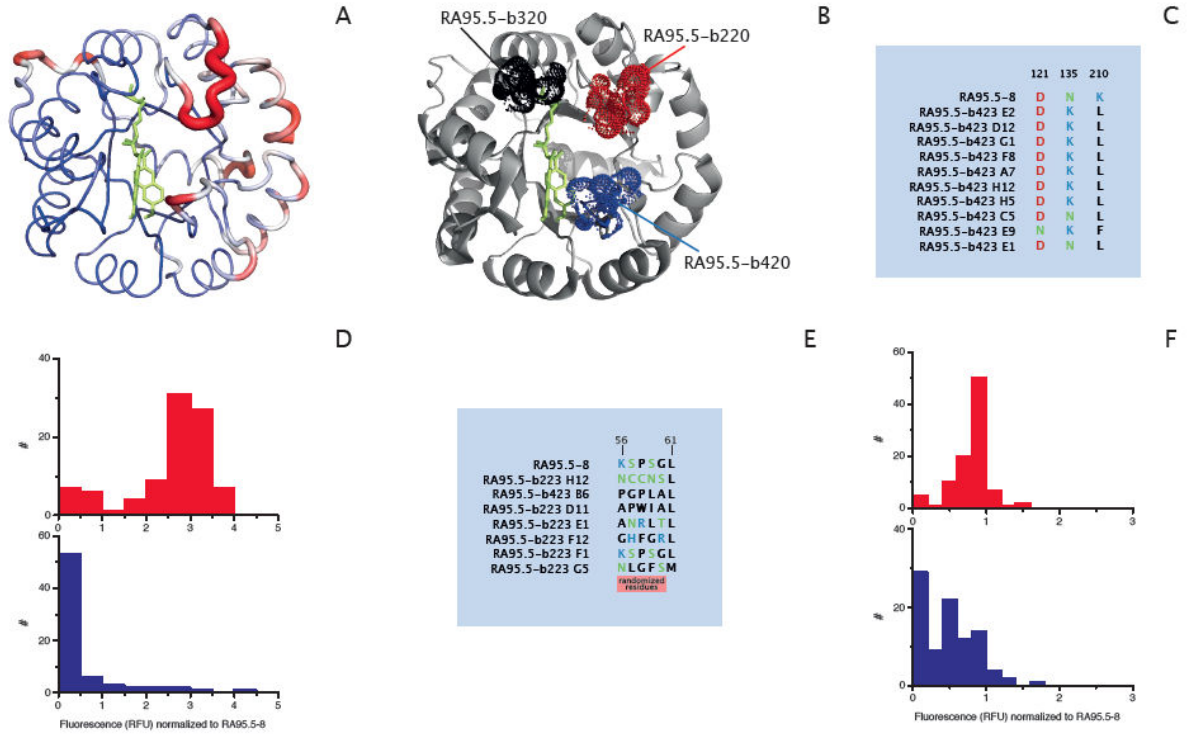

Figure 47. Mutations in loop regions. *A: B factor representation of the crystal structure, red and thick regions correspond to high B values. B: Location of the loop libraries in the protein scaffold. C: Alignment of isolated mutants from the RA95.5-b423 library. D: Histogram of plate assay of RA95.5-b420 library. Shown in blue is the activity distribution before sorting and shown in red is the distribution after three rounds of sorting. E: Alignment of mutants isolated from the third round of screening from the RA95.5-b220 library. F: Histogram of plate assay of RA95.5-b220 library. Shown in blue is the population before sorting; shown in red is the population after three rounds of sorting.*

Additionally, the two loop libraries RA95.5-b220 and RA95.5-b420 were shuffled after the second round of sorting to give library RA95.5-c230. After sorting multiple times, the diversity at positions 56-60 was completely lost in the best mutants. Only clones bearing mutations at position K210 were isolated. Additionally two clones, identical to K210M, which did not appear in the previous library RA95.5-b420, were isolated. This methionine mutant RA95.5-c233 CD4 was further characterized (Figure 47) and revealed a k_{cat}/K_M value of 2670 $M^{-1}s^{-1}$, making this enzyme three times more proficient than the starting enzyme RA95.5-8.

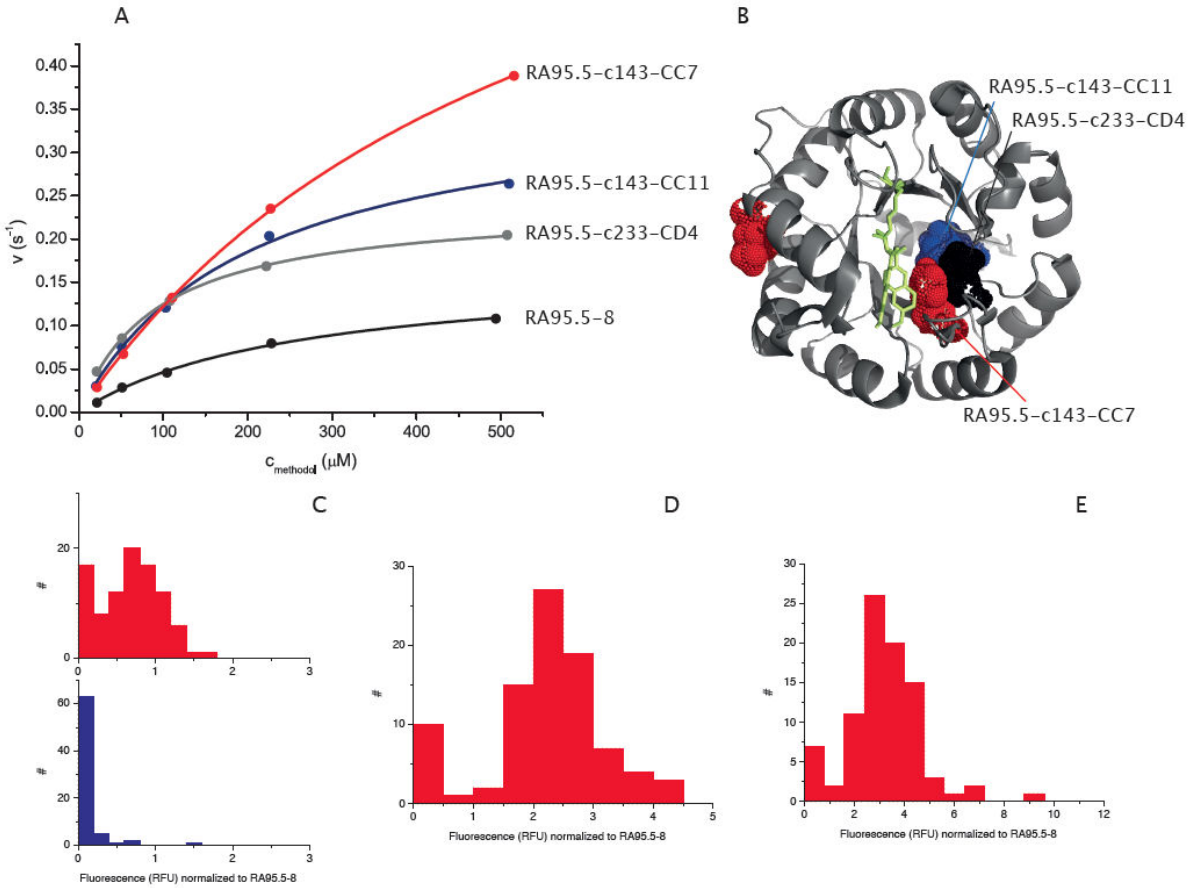

Figure 48. Improved variants after 2-3 rounds of mutagenesis. *A. Michaelis Menten kinetics of selected mutants. B: Location of the mutations of the characterized mutants. C: Histogram of plate assay of RA95.5-b110 library. Shown in blue is the activity distribution before sorting and shown in red is the distribution after three rounds of sorting. D: Histogram of the rescreen of sorted library RA95.5-c233. E: Histogram of the rescreen of sorted library RA95.5-c143.*

In parallel to the loop libraries, a random mutagenesis library, based on the sorted variants of the previous rounds was created. The average amino acid mutation frequency was set to an approximate value of 10. The library was sorted three times and analysed, revealing that the average activity was smaller than the activity of RA95.5-8. Sequencing of the best mutants showed that the amino acid mutation frequency was decreased to around 2.5 mutations per protein. It can be stated that proteins with a low mutation load had a selective advantage. Thus, to come by this problem and to isolate highly beneficial mutations, back shuffling was performed after the second round of sorting. The population was shuffled with mutants RA95.5-8, RA95.5-a113 BA1 and RA95.5-a113 BA3 in an overall 1:3 ratio. Sorting of this library gave rise to mutants with improved catalytic properties. The isolated and characterized mutants each bore one

mutation (RA95.5-c143 CC11: L231M and RA95.5-c143 CC7 being an exception with one surface mutation L17Q and an active site mutation F180Y) located directly in the active site. Intriguingly, these mutations including the K210M mutation of clone RA95.5-c233 CD4 cluster in same part of the active site and are all in van der Waals contact, as suggested by the crystal structure of RA95.5-5 (Figure 47).

Although the kinetics of the mutants could not be satisfyingly determined, since the K_M exceeded the solubility of the substrate, clear improvement in k_{cat} is observable (Figure 48). Additionally, it has to be mentioned that for the plate screening and kinetic characterization, methodol was used as a substrate due to its availability. Single point measurements of various mutants however, suggest that the activity for methodol differs from the activity for the actual ammonium substrate, used in the microfluidic experiments by the factor of two or more [Richard Obexer, personal communication]. Considering that RA95.5-8 has been already evolved for over 8 rounds, the 5-10 increase in activity of crude lysate over two rounds of mutagenesis can be considered rather an encouraging result.

6. Conclusions and perspectives

In the course of this work we developed a high-throughput droplet-based microfluidic platform for *de novo* designed retro-aldolase screening. Since there was no suitable fluorogenic substrate, we designed, synthesised and tested for droplet leakage a series of UV-fluorescent naphthaldehydes. The substrate/product pair which fit the best criteria was chosen for the retro-aldolase assay in droplets. To avoid multiple droplet manipulation we decided to choose cells, as protein factories, expressing enzyme in the cytoplasm. Retro-aldolase kinetics of single bacteria in droplets revealed a wide variance in activity within the single population of cells expressing the same enzyme variant. Nevertheless we were able to enrich the most active variant during the model experiment.

Moreover, we created an integrated microfluidic device allowing droplet production, incubation in a delay-line, kinetic measurements and sorting triggered on fluorescence by dielectrophoresis, all performed on a single device. This integrated chip

reduces the droplet time-dispersion during the off-chip incubation allowing better discrimination between different variants.

The directed evolution experiment aimed to improve the RA95.5-8 variant, isolated after 8 rounds of evolution using microtiter plate screening. After two-three rounds of evolution employing the microfluidic platform we were able to isolate variants with a 5-10-fold increase in activity of the crude lysate.

We are planning to continue the evolution rounds towards the further improvement of retro-aldolase activity. In parallel, we are characterizing the isolated clones and supplying this information to the computational design team[a], which will be used to increase the efficiency of *de novo* computational design algorithms.

[a] The Baker Laboratory, Seattle, WA, USA

High-throughput screening of proteases under defined conditions

The detergent industry is the largest single market for enzymes, constituting 25-30% of total enzyme sales (about 1.2 billion € in 2011). The use of enzymes in detergent mixtures can often eliminate the need for high temperatures and extreme pH values, while at the same time offering reduced environmental impact and increased washing efficiency.

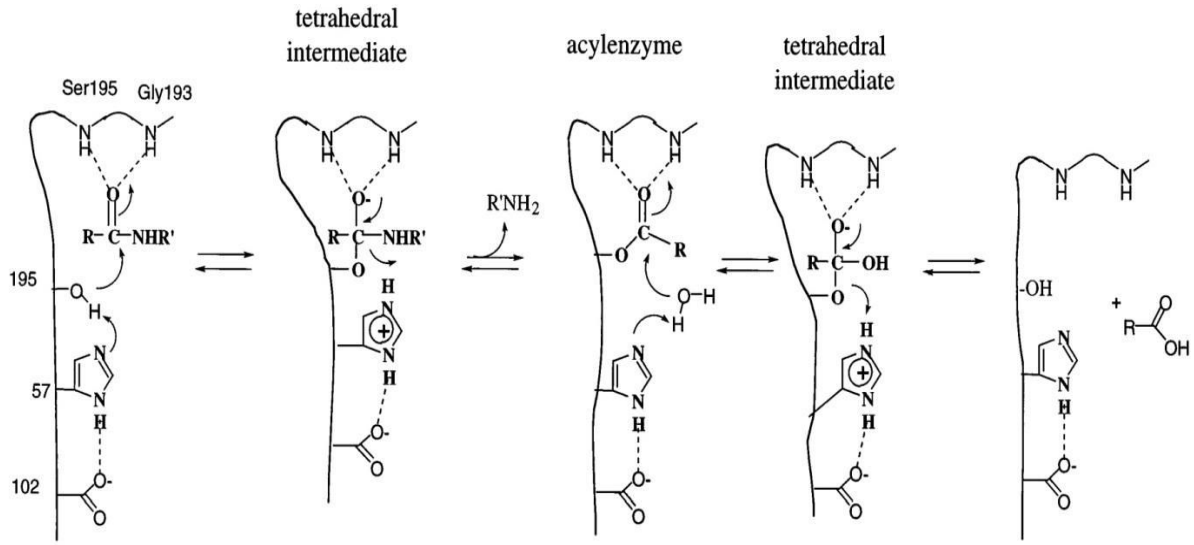

Figure 49. General mechanism for serine proteases *(catalytic triad Asp-His-Ser)* **[224].**

Alongside other enzymes like amylases, lipases and cellulases, proteases are often used in detergents for effective removal of organic stains from clothing and other materials. All the proteolytic enzymes employed in the detergent industry are non-specific serine endoproteases, giving preferential cleavage on the carboxyl side of hydrophobic amino acid residues, which gives the ability to hydrolyse most peptide links. They convert their substrates into small, readily soluble fragments which can be removed easily from fabrics. Only serine protease like subtilisins may be used in

detergent formulations: thiol proteases (papain) are easily oxidised by bleaching agents, and metalloproteases (thermolysin) can lose their metal cofactors due to complex formation with water softening agents or hydroxyl ions. Another important point is the production of subtilisins as extracellular enzymes, which greatly facilitates enzyme purification and employment in other downstream processing steps.

In addition to having the non-specific protease activity, proteolytic detergent enzymes must be stable during storage and tolerate reaction conditions. For example, granulated enzymes, when released into washing media, must withstand anionic and non-ionic detergents, soaps, oxidants, other less-aggressive compounds and, sometimes, high temperatures (up to 60°C).

Due to high commercial interest, the engineering of detergent proteases has been carried out since 1984, involving rational and random approaches [225], but most of the progress has been published in patents and not in the scientific literature. One recent available example is the directed evolution of subtilisin 309 towards the modification of substrate specificity using phage display [226]. However, this high-throughput screening system was successfully applied only for improving the binding specificity, but not for catalytic activity or stability in reaction environment.

Our current collaborative project between Laboratoire de Biochimie and Novozymes aims at developing a droplet-based microfluidic system for the high-throughput screening of proteases under conditions that closely simulate the final operating conditions for commercial proteases (pH, salt, high water-hardness, detergent, temperature). To keep the selection system as close as possible to the industrial entourage, was decided to use *B. subtilis* strain for protein expression. Since the conditions of screening for activity are largely incompatible with the viability of bacteria, protease secretion must be completely uncoupled from the enzymatic assay. This allows us to select proteases under any desired conditions.

For this reason, it is necessary to create a microfluidic platform in which droplets containing single *B. subtilis* cells will be incubated to allow cell growth and protease secretion, and then be fused one-to-one with droplets containing a fluorogenic protease assay and reagents to modify the conditions of the assay to simulate the operating conditions for commercial enzymes. The steps for our work plan are: (a) optimization of *B. subtilis* growth and protein secretion in droplets, (b) discrimination of *B. subtilis* producing proteases with different activities, and (c) sorting *B. subtilis* producing an active protease, from *B. subtilis* not producing any significant protease activity (d) performing several cycle iterations of directed evolution of the protease using the most preferred assay conditions.

I. Materials and methods

1. Cell inoculation before encapsulation

All received bacterial samples in glycerol stock were thawed on blue ice, aliquoted by 50 µl into small PCR-tubes, refrozen and kept at -80 °C. Prior to analysis, 50 µl sample from the glycerol stock was thawed on wet-ice. 5 µl of the sample was inoculated with 3 mL of growth media (Luria Broth-Media 87.5%, Terrific Broth-Media 12.5%; 100 mM Tris-HCl [pH 8.0]; 6 µg/mL chloramphenicol) in a 14 mL round-bottom falcon tube. The tube was incubated at 37 °C, 60° angle of tube in the shaker, 200 rpm, for 4-8 hours.

2. Cell encapsulation and growth in droplets

From the inoculated cell solution, 1 mL of culture was harvested with a centrifuge (6000 rpm for 2 min). Supernatant was removed and cells were gently resuspended and washed twice with 1 mL of fresh media. The final pellet was resuspended in 1 mL of media and filtered through a Syringe-Driven Filter Unit (Millex-SV 5.00 µm, Low Protein Binding Durapore® (PVDF), 25 mm). The OD_{600} was adjusted to 0.06 (for occupancy λ=0.3). The resulting cell solution was kept on ice. The cold cell suspension was transferred into a 1-mL syringe (Omnifix-F®, BRAUN). The syringe cooled with ice, was fixed in horizontal position on a pump, and connected to the aqueous phase inlet of the microfluidic device (Figure 50) through a needle and a short fragment of PTFE tubing. In the conditions when flow-rate of the cell solution was 200 µL/hr and flow-rate of oil phase 300 µL/hr (HFE7500 with 2% EA), 4 pL droplets were generated at 14 kHz frequency. The outlet of the device was connected through short PTFE tubing and a needle, to an open (without piston) 5-mL syringe containing 1 mL of surfactant-free HFE7500 oil. The 5-mL syringe was fixed vertically, higher than the level of chip, so that the produced emulsion was collected on top of oil layer inside the syringe. Since the collected droplets experienced direct contact with air, a small cotton plug saturated with water was inserted from the top part, which helped to maintain a humid air atmosphere above the emulsion preventing the evaporation of water from the droplets.

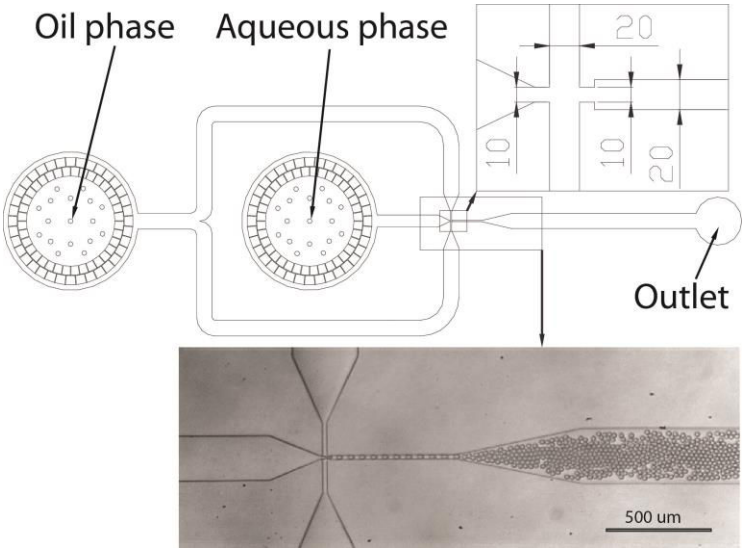

Figure 50. Microfluidic device for 0.5-4 pL-sized droplet generation. *Depth 15 µm. The dimensions of all device features are indicated in µm.*

The droplets were collected for 30 min, and a 5-mL syringe containing the emulsion was incubated at 37 °C for 8-10 h.

3. Fusion with assay components

After incubation the emulsion containing droplets with *B. subtilis* and secreted protein was transferred from the 5-mL syringe into a 1-mL syringe with a piston. Another 1-mL syringe was filled with assay solution:

- 3.9 mg/mL Model L detergent;
- 15°dH: 0.315 mg/mL $CaCl_2 \cdot 2H_20$, 0.109 mg/mL $MgCl_2 \cdot 6H_20$, 0.337 mg/mL $NaHCO_3$, 100 mM Tris [pH = 7.8]
- The solution above was filtered through 5 µm syringe filter before adding:
- Bodipy TR-X Casein from Molecular Probes - 0.05 mg/mL.

Other oil containing syringes were prepared and injected into the fusion device with the following flow-rates:

Phase	Composition	Flow-rate (µL/hr)
Aqueous	Assay solution	80-110
Drop formation oil	HFE7500 surfactant-free	150-200
Emulsion	Droplets with cells	25-40
Reinjection oil	HFE7500 surfactant-free	100-150
2xStabilization oil	HFE7500 2% surfactant	120-180

Droplets were collected for 1h into a glass capillary on ice. Afterward, the emulsion was incubated at 37 °C for 40 min.

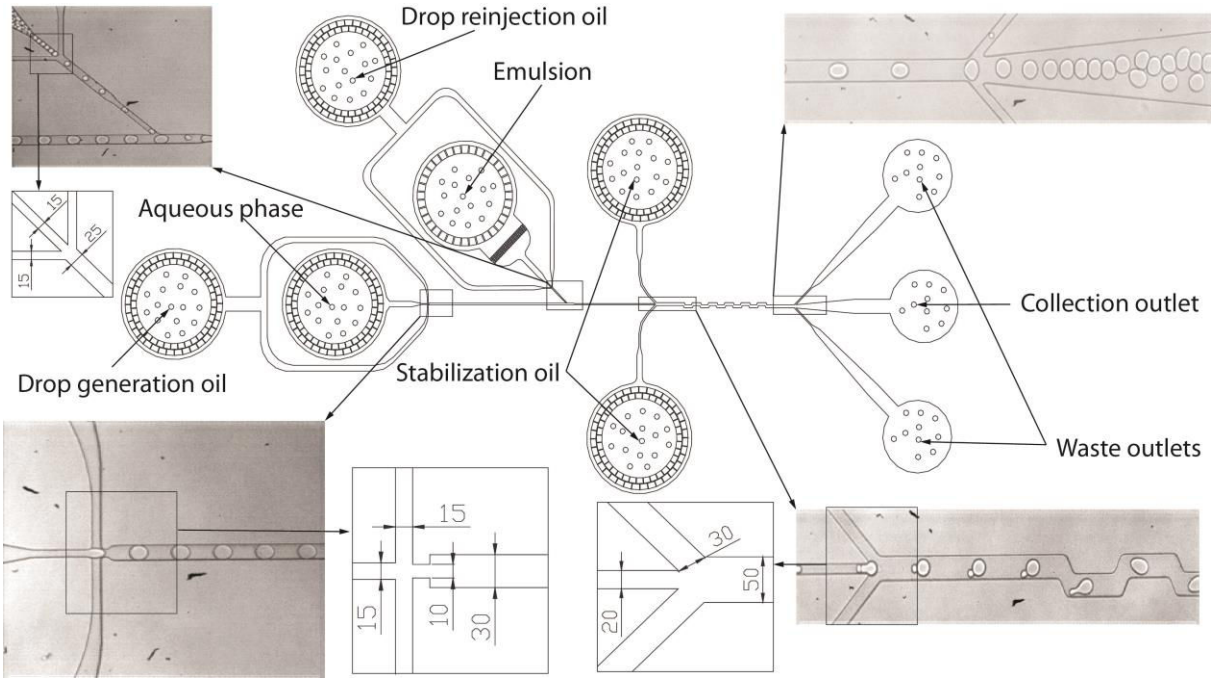

Figure 51. Optimized microfluidic device for passive fusion. *Depth 20 µm. The dimensions of all device features are indicated in µm.*

4. FADS for improved activities

After incubation the emulsion was reinjected from ice into the sorting device through one of the emulsion inlets (Figure 52). 25 pL-sized "control droplets" prepared beforehand (50 µM Fluorescein, 0.1 µg/mL TexasRed) were reinjected through the other emulsion inlet. The flow-rates of emulsion phases were at 20 µL/hr each, while the flow-

rates of surfactant-free oil for reinjection were at 300 µL/hr each. The droplet fluorescence was monitored using the set-up (Figure 24) and recorded for 5 minutes. Afterward, the flow rate of "control droplets" was stopped and the flow-rate of incubated emulsion was decreased to 15 µL/hr, reinjection oil flow-rates were as well decreased to 200 µL/hr each. The sorting switch was turned on and positive droplets are deflected using AC (F = 30 kHz, ts_{ort} = 0.8 ms, U_{sort} = 1.3 kV_{p-p}). The reinjected library was sorted at 300 Hz for 4 hours, with sorting gate set to collect 1% of the most active drops.

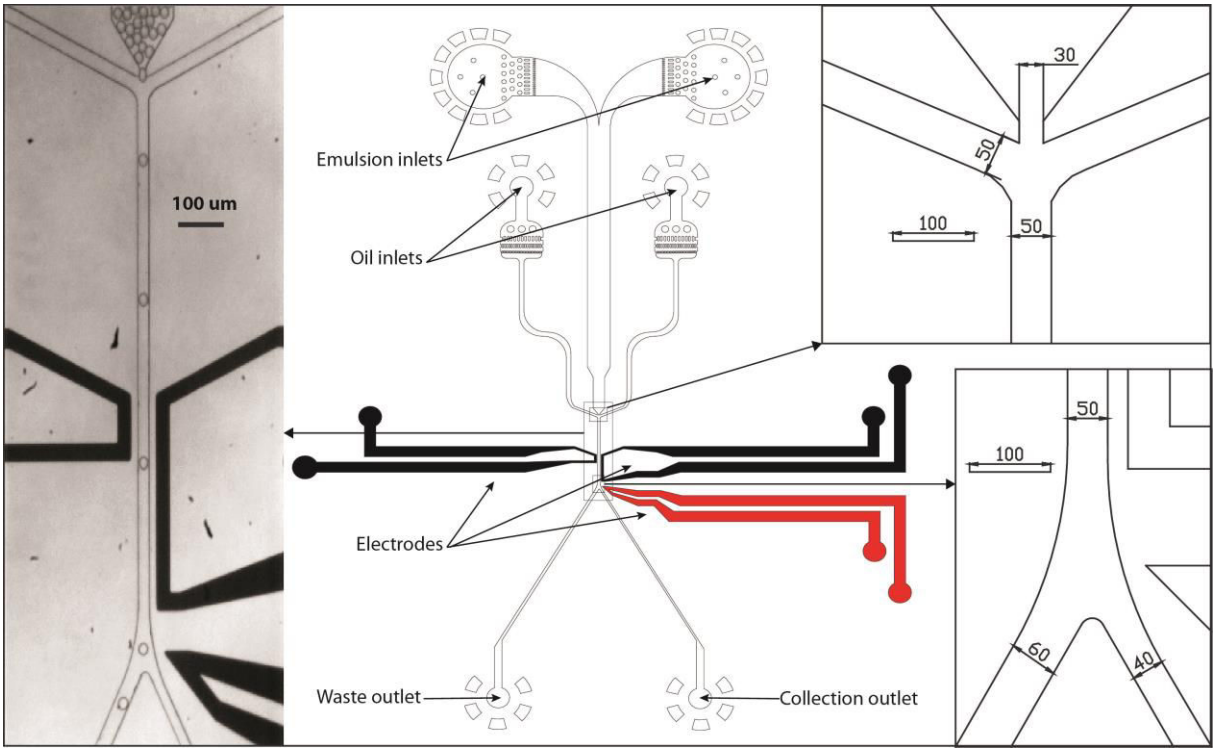

Figure 52. Microfluidic device for fluorescence-activated droplet sorting. *Depth 25 µm. The dimensions of all device features are indicated in µm.*

5. DNA recovery

Sorted droplets were collected in a tubing loop that connected the positive outlet of the device with a 2-mL eppendorf tube. To avoid inspiration of the unsorted emulsion, the tubing was cut right next to the outlet of the microfluidic device. If the sorted emulsion was visible by eye, we carefully removed the other end of the tubing from the eppendorf tube and collected the emulsion in a small PCR tube. Otherwise we

flushed all the oil from the tubing to the 2-mL eppendorf. We washed collection tubing several times. To the w/o emulsion was added 1/10 volume of *1H,1H,2H,2H*-perfluoro-1-octanol and 50 µL of water (Molecular Probes). We mixed by vortexing several times. We centrifuged at 1000 rpm to have the water and oil phases separated. We took 50 µL of water phase and follow the protocol: Gram-Positive Bacteria Genomic DNA Purification Protocol using *GeneJET Genomic DNA Purification Kit* (Fermentas), (Modifications of protocol are in bold):

- To **50 µL** water phase add 180 µl of the appropriate enzyme solution (20 mg/mL lysozyme or 200 µg/mL lysostaphin; 20 mM Tris·HCl, pH 8.0; 2 mM EDTA; 1.2% Triton).
- Incubate for at least 30 min at 37°C.
- Add 200 µl of Lysis Solution and 20 µl of Proteinase K. Mix thoroughly by vortexing or pipetting to obtain a uniform suspension.
- Incubate the sample at 56°C while vortexing occasionally or use a shaking water bath, rocking platform or thermomixer until the cells are completely lysed (~30 min).
- Add 20 µl of RNase A Solution, mix by vortexing and incubate the mixture for 10 min at room temperature.
- Add **300 µL** of **70%** ethanol and mix by pipetting or vortexing.
- Transfer the prepared lysate to a GeneJET™ Genomic DNA Purification Column inserted in a collection tube. Centrifuge the column for 1 min at **8000 rpm**. Discard the collection tube containing the flow-through solution. Place the GeneJET™ Genomic DNA Purification Column into a new 2 mL collection tube (included).
- Add 500 µL of Wash Buffer I (with ethanol added). Centrifuge for 1 min at **8000 rpm**. Discard the flow-through and place the purification column back into the collection tube.
- Add 500 µL of Wash Buffer II (with ethanol added) to the GeneJET™ Genomic DNA Purification Column. Centrifuge for 3 min at maximum speed **13 400 rpm**. *Optional*. If residual solution is seen in the purification column, empty the collection tube and re-spin the column for 1 min. at maximum speed. Discard the collection tube containing the flow-through solution and transfer the GeneJET™ Genomic DNA Purification Column to a sterile 1.5 mL microcentrifuge tube (not included). Place the QIAamp Mini spin column in a new 2 mL collection tube (not provided) and discard the old collection tube with the filtrate. Centrifuge at full speed for 1 min.
- Add 100 µL of Elution Buffer to the center of the GeneJET™ Genomic DNA Purification Column membrane to elute genomic DNA. Incubate for **5 min** at room temperature and centrifuge for 1 min at **8000 rpm**.

6. DNA transformation in cells

The DNA encoding the sorted protease variants were recloned by Novozymes; briefly, the DNA was amplified by PCR and cloned in *Bacillus subtilis* using standard but proprietary protocols. This resulted in a recloned pool of *Bacillus subtilis* clones encoding the enriched population of genes ready for a new round of sorting and also for DNA sequencing of individual clones.

II. Results and discussion

1. Optimization of *B. subtilis* growth

During the visual monitoring of *B. Subtilis* growth inside droplets, we observed at least three different morphologies (rods, filaments and spheres). These morphology distributions were characteristic for all tested protease secreting strains. Moreover, each cell morphology showed different growth rates: the rod-type grew fastest and formed the largest colonies. Spheres grew slowest and produced only few colonies. They also moved differently, the rod-type being the most active, swimming very energetically.

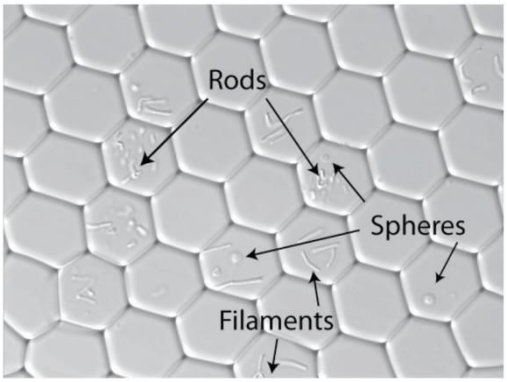

Figure 53. Different cell morphologies in droplets.

Initial experiments showed variations between individual experiments in terms of cell density per droplet, as well as protease activity. We noticed that cell survival is critically dependent on oxygen availability. If a cell-containing emulsion is incubated in a closed system (capillary, syringe, tube), cells grow slowly and show pronounced lyses. Therefore, different strategies to grow cells inside droplets were tested.

The GFP-POS strain was emulsified into 10 pL droplets together with the fluorogenic protease substrate Casein TR-X (the hydrolysed substrate has excitation/emission maxima of approximately 589 nm and 617 nm, respectively). The cell density was adjusted to achieve 30% single occupancy and less than 5 % double occupancy. The collected emulsion was split into four syringes, each filled with 200 µL carrier oil (bottom layer). Three syringes contained a 100 µL top layer: LB-Media, decan, and mineral oil; while the fourth one had

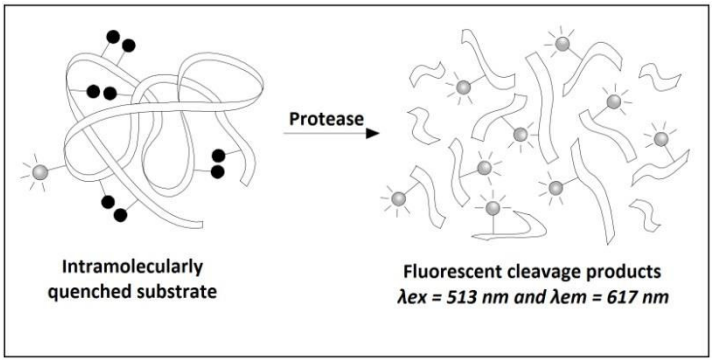

Figure 54. Cleavage of BODIPY TR-X substrate by proteases.

only a water-saturated atmosphere above. Emulsions were incubated for 24 hours and placed on glass slides for analysis.

According to the observed results (Figure 55), the syringe containing no top layer was most suitable for cell survival and growth. In the presence of mineral oil, cell growth was often accompanied by cell lysis. In the syringe with decane, cell lysis events were rarer but still present. Finally, in the presence of LB-media as the top layer, significant droplet coalescence occurred.

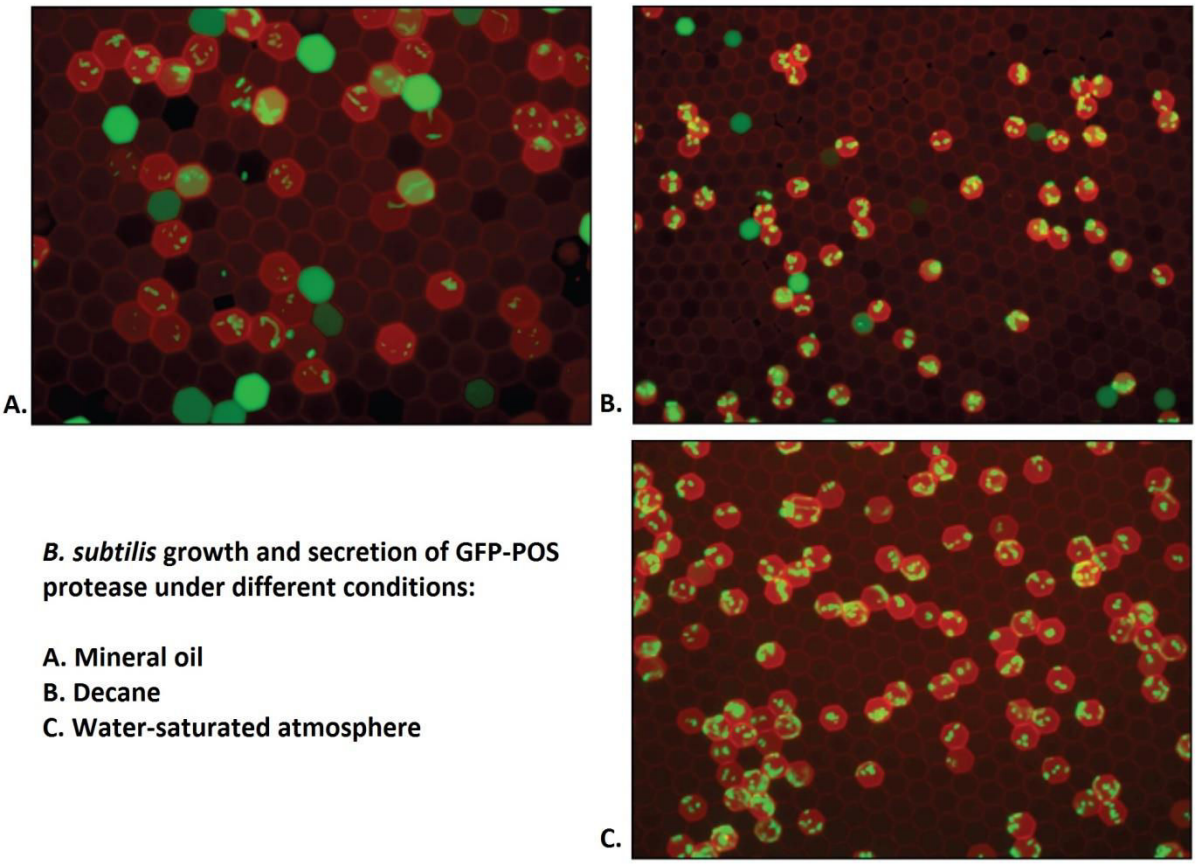

Figure 55. Confocal microscopy images of emulsions after 24 hours incubation under different conditions. *Droplet size 10 pL.*

In the experiment where cells were incubated for extended times using the water-saturated atmosphere method, *B. subtilis* showed viability even after 48 hours. Thus, in all further experiments this type of incubation has been employed.

Another important question related to bacterial growth is whether the bacterial strains expressing different protease variants grow at the same rate or whether the secreted protease interferes with cell growth. For that reason, several *B. subtilis* strains

POS (1x), Sol843 (0.01x-activity of wt-variant - POS), Sol842 (0x) and LBei307 (2x) were inoculated in the same conditions in a transparent microtiter plate. The cell density has been monitored over 9 hours on a plate reader. Monitoring of optical density increase over the time did not reveal any dependence of growth rate and secreted protease variant for tested strains (Figure 56). In this manner, the protease library will be screened according to the activity and expression level of the clone, and not to bacterial growth.

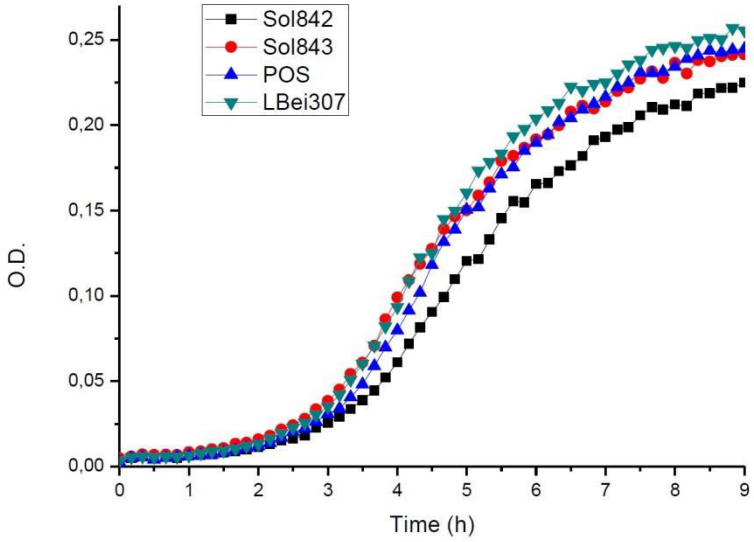

Figure 56. Overnight growth of bacteria secreting different protease variants. *POS (1x), Sol843 (0.01x), Sol842 (0x) and LBei307 (2x).*

2. Protease activity in droplets

For a microfluidic platform to be feasible for directed evolution, the ability to exclusively select improved variants is necessary. The more precise is the screening, the more likely is the enrichment in desired enzyme quality. For that reason, Novozymes supplied several *B. subtilis* strains expressing protease with different activities ranging from 0 to 250% of wild-type clone activity (POS).

To see if it is possible to discriminate all the clones by activity using the microfluidic platform, the following experiment has been carried out. Five separate emulsions containing five different strains, (POS 1x, Neg 0x, LBei 306 1.5x, LBei 307 2x, LBei 310 0.5x), were emulsified into 3 pL droplets at $\lambda \sim 0.3$ (on average <30% of droplets contain single cells) and labelled with different concentrations of green

fluorescent marker. Droplets were then incubated in an air-thermostat overnight, reinjected and passively fused [117] with 12 pL droplets produced on-chip and containing 0.1mg/mL substrate dissolved in 0.1M Tris-HCl [pH 8.0], 1mM $CaCl_2$ solution. Fused droplets were then collected and incubated at 30°C for 2 hours followed by reinjection and fluorescence detection.

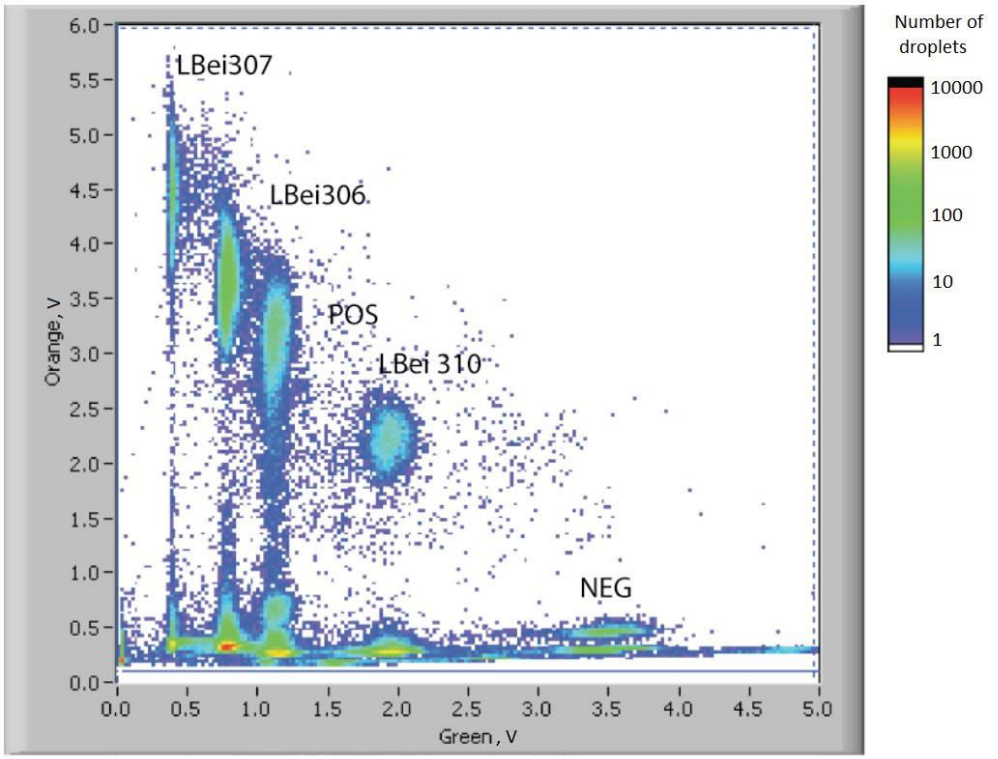

Figure 57. Comparison of protease activity among five strains. *Droplets containing different strains were labelled with a fluorescein marker, which corresponds to the Green signal. Protease activity corresponds to the Orange signal.*

From Figure 57 we can conclude that individual populations corresponding to different protease activities are distinguishable. Nonetheless, the activities of the LBei307 and LBei306 strains were only 1.3x and 1.1x, which is lower than the microtiter plates results. This might be explained by the fact that the activity was screened after 2 hours and in that time the most active variants probably reached the region outside of linear kinetics. Thus, in order to be able to distinguish clones with higher activity, the incubation time after fusion with assay components needs to be optimized.

All attempts to recover viable cells after the activity assay failed. Eventually, the reaction solution (0.1M Tris-HCl [pH 8.0], 1mM $CaCl_2$) in this experiment might cause osmotic cell lysis or cells to not stand for long time in protease solution. Moreover, the assay conditions will get even harsher when the activity is screened in detergent-containing assay solution. So, another important conclusion which needs to be drawn from the experiment above is that the recovery of genetic material must be done by DNA amplification, because cells do not survive the assay.

3. Passive fusion with assay components

The experiments described in the section "Protease activity in droplets" were carried out without detergent. The passive fusion of cell-containing droplets with assay detergent solution was impossible under standard conditions [117]. This outcome can be explained when considering the passive fusion mechanism. This selective one-to-one droplet fusion method exploits transient states in the build-up of surfactant molecules at the droplet interface.

Here (Figure 58), reinjected surfactant-stabilized droplets are brought together and fused with newly created droplet that are not fully stabilized by surfactant. The newly formed interface of the fused droplets is partly stabilized; thus additional droplet stabilization is assured by supplying further surfactant dissolved in the carrier oil. When detergent solution is used to generate droplets for fusion, a newly formed drop is supplementary stabilized by the components of the detergent, thus when coming in contact with reinjected droplet, the fusion act does not occur. Looking closer at standard enzyme detergent composition (Table 3), it is possible to identify ionic surfactant molecules like sodium alkane sulphonate and sodium alkane carboxylates. These surface active substances constitute almost 1/3 of detergent mass, and presumably are the main reason for premature droplet stabilization.

Table 3. Compositions of an enzyme detergent [227].

Constituent	Composition (%)
Sodium tripolyphosphate (water softener, loosens dirt)	38.0
Sodium alkane sulphonate (surfactant)	25.0
Sodium perborate tetrahydrate (oxidising agent)	25.0
Soap (sodium alkane carboxylates)	3.0
Sodium sulphate (filler, water softener)	2.5
Sodium carboxymethyl cellulose (dirt-suspending agent)	1.6
Sodium metasilicate (binder, loosens dirt)	1.0
Bacillus protease (3% active)	0.8
Fluorescent brighteners	0.3
Foam-controlling agents	Trace
Perfume	Trace
Water	to 100%

To overcome the issues related to passive fusion with detergent solution the standard protocol needs to be modified. Different variations in flow-rates of injected phases lead only to the decrease of throughput with no significantly improved outcome. Thus, the surfactant concentration in the oil for droplet generation and droplet reinjection was decreased from 2% to 0.5% of EA in FC40 to achieve one-to-one fusion in the area **B** (Figure 58). However when fused droplets come into contact with each other in the collection outlet, uncontrolled coalescence occurs. The increase in surfactant concentration of stabilization oil to 5% EA, improved droplet stability after fusion; nevertheless about 20% of fused droplets still coalesced.

Since the variations in fluorinated surfactant concentration gave only modest results, it was decided to introduce modifications into the fusion device itself. The surfactant-mediated passive fusion theory stipulates that the stabilization of droplet against coalescence depends on the building-up of surfactant on the droplet interface. This process occurs in a millisecond time scale after droplet formation and depends on surfactant concentration, fluorinated-oil type and on the composition of aqueous phase. Probably, the surfactant interface of the detergent droplet is already built-up before coming into contact with the reinjected droplet. In an attempt to reduce the time between on-chip droplet generation and droplet fusion the length of the channel in the **A.** region (Figure 58) was reduced. This did not yield any improvements. This means

that the high concentration of ionic surfactants in the detergent solution facilitates a very fast droplet interface formation.

The increase in the length of pairing region **B.** (Figure 58) from 1500 µm to 2500 µm gave a fusion improvement from 90 % to 95%, while the further length increase to 3500 µm reduced the fusion efficiency. Eventually, the longer contact time between unstabilized droplet and reinjected droplet allows a closer interaction of interfaces contributing to more efficient droplet merging. However, when paired droplets are spending too much time in contact with each other, the surfactant contained in the spacing oil used for reinjection builds on the surface of the on-chip produced droplet resulting in stabilization before the fusion act. And finally, the increase in the length of **C**-region (Figure 58) resulted in only improved stability of the fused droplet at the collection, as expected.

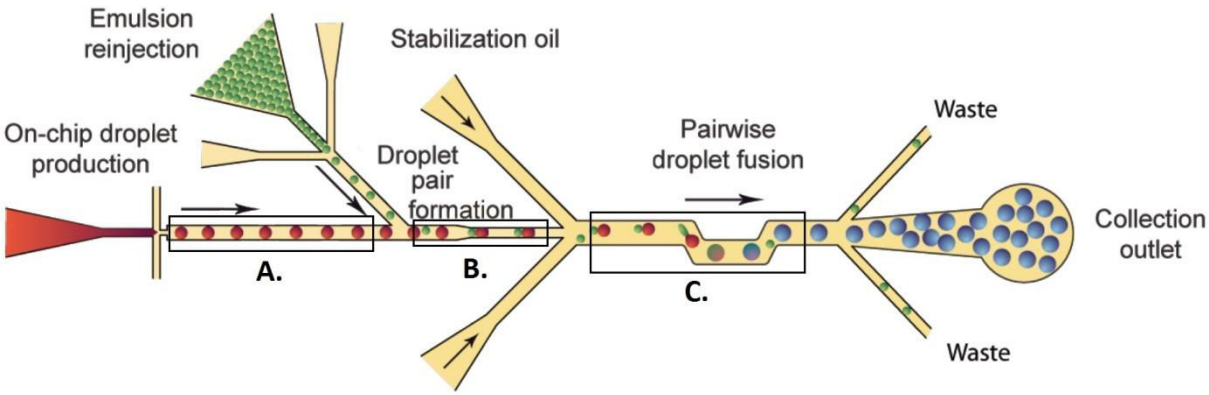

Figure 58. Passive fusion technique. *Regions on the microfluidic device:* ***A.*** *Unstabilized droplets before pairing;* ***B.*** *Pairing region;* ***C.*** *Zig-zag channel for fusion and stabilization.*

All efforts for passive fusion with detergent droplets in FC40 oil described above allowed selective efficient 95% one-to-one fusion with no coalescence at the collection outlet. But still, when the fused emulsion after incubation at 37°C was reinjected for sorting, 5-10% of droplets coalesced. This is related to well known problems of droplet stability in FC40 oil, observed in other unpublished experiments as well.

Since during directed evolution experiments we are aiming to select the top 1-5% most active protease variants, uncontrolled droplet merging can contaminate the positive hits, dropping down significantly sorting efficiency. In this regard, it has been

decided to test HFE7500 oil as a continuous phase, in which no successful surfactant-mediated passive fusion experiments have been reported before.

Passive fusion in HFE7500 was not possible since the surfactant distribution on the droplet interface occurs faster in HFE7500 than in FC40 and formed droplets are more stable [Linas Mazutis, personal communication]. This effect was seen when the passive fusion experiment was carried out involving the same surfactant concentration as in FC40 conditions. Generated on-chip droplets were stabilized before fusion using only 0.5 % surfactant concentration in HFE7500.

The further decrease in surfactant concentration showed the possibility of detergent-containing droplet generation employing surfactant-free HFE7500. This effect can be explained by droplet stabilization only by surfactants contained in the detergent. Since HFE7500 represents a mixture of fluorinated and partially fluorinated hydrocarbons (Figure 23), it possesses fluorophilic and hydrophobic properties, which is not characteristic for completely fluorinated FC40. Thus, a detergent containing droplet is more stable in HFE7500 due to the build-up on the droplet interface of alkane carboxylates and alkane sulphonates. The hydrophobic alkane tail of the detergent surfactant has a higher affinity for partially hydrophobic HFE7500 than for the entirely fluorophilic FC40, allowing alkane-based surfactants to be distributed on the interface (hydrophilic head inside the drop, hydrophobic tail in the oil phase) thus stabilizing the drop in HFE7500. But, detergent stabilized droplets fuse when they come into contact with each other. In this manner, alkane-based ionic surfactants do not fully protect droplets from coalescence, which is exactly what passive fusion needs.

In the passive fusion device (Figure 58) detergent droplets were generated using surfactant-free HFE7500 and efficiently fused one-to-one with reinjected droplets. Full droplet confinement was achieved by stabilization with 2% fluorinated surfactant in HFE7500. The collected emulsion of fused droplets was stored at 37˚C, and reinjected for monitoring the droplet stability. The emulsion represented monodisperse droplets even after 2 days of incubation, which is more than enough for protease directed evolution experiments.

In conclusion, the adding of assay components to the droplets containing bacteria and secreted protease was impossible using standard conditions of passive fusion (the

same for electro-coalesence and pico-injection) due to the additional interface stabilization effects of detergent constituents. Selective one-to-one passive fusion (98% one-to-one fused droplets, < 0.5 % two-to-one) at high-throughput level (1.8 kHz) has been achieved using HFE7500 oil as the dispersed phase, resulting in high stability of droplets after fusion. All droplet operations for the entire microfluidic workflow of the protease screening project were done in this manner in HFE7500.

4. Protease kinetics

Since the pure wild-type enzyme was unavailable, all the following kinetics are based on cell suspensions.

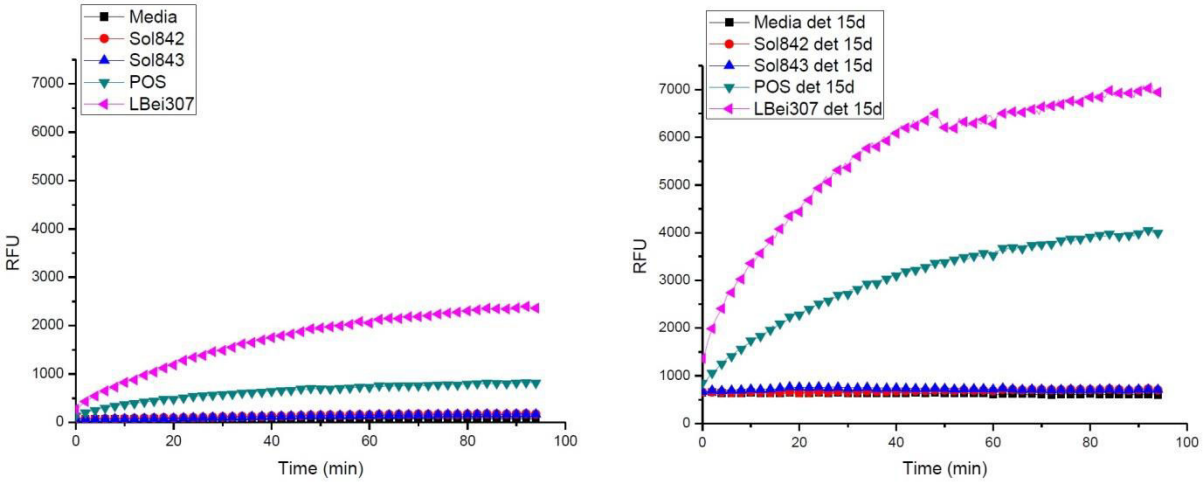

Figure 59. Protease assay with and without detergent. *80 µL of assay mixture (**left**) (0.05 mg/mL Bodipy TR-X; 100 mM Tris; pH = 8) or assay/detergent mixture (**right**) (0.05 mg/mL Bodipy TR-X; 15°d; 3.8 g/l - Model L detergent; pH = 7.8) was mixed with 20 µL cell culture inoculated for 10h. The kinetic curves were registered on spectrophotometer (λexc=589 nm, λexc=617 nm). Protease strains: POS (1x), Sol843 (0.01x), Sol842 (0x) and LBei307 (2x).*

From the figure above (Figure 59) it can be concluded that, the protease activity in the detergent assay is 3-4 times higher than in buffer, which might be related to the high concentration of alkaline earth cations (Ca^{2+} and Mg^{2+}) in hard water. These ions can additionally activate the protease in the detergent assay mixture. Also, it is possible that the detergent solubilizes better the protease and its substrates. The 2-fold difference in activities between LBei307 (2x) and POS (1x) is maintained when switching to the detergent solution, but this might not be the case for other clones.

Another observation reveals the presence of background fluorescence in the mixture of hard water and detergent. The effect might be caused by changes in casein-based substrate structure resulting in loss of quenching efficiency.

In order to establish the optimal incubation time for the protease activity assay in droplets, the following experiment has been run. Two strains POS (1x) and LBei307 (2x) were encapsulated and incubated overnight, fused afterwards with substrate/detergent containing droplets and reinjected into an analyzing device in different time intervals.

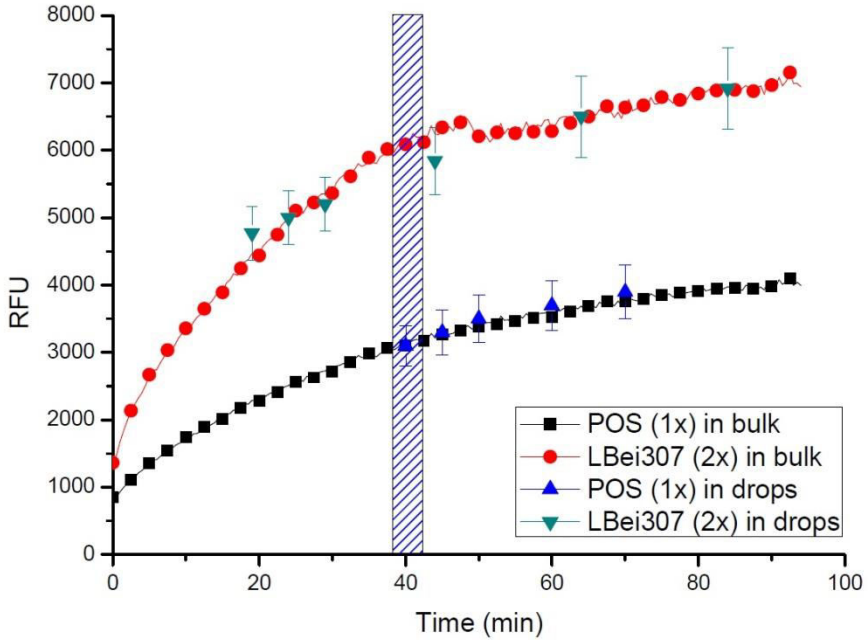

Figure 60. The protease kinetics in droplets versus bulk.

The kinetic points obtained for the activity of protease strains in droplets were plotted against the same data from the bulk experiment (Figure 60). The activity of the two variants in droplets matches the kinetics in microtiter plates. To avoid reaching the plateau, the droplets should be incubated no longer than 40 min after fusion.

5. Model experiment in final screening conditions

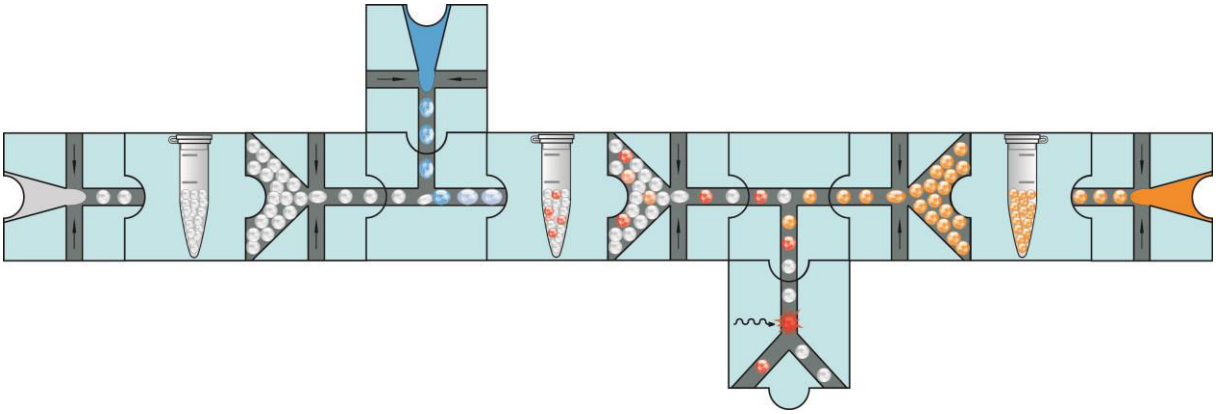

Figure 61. Microfluidic workflow for protease screening. *From left, cells secreting enzyme are compartmentalized in droplets, incubated for 10h, fused with assay mixture containing detergent, incubated for 40 min, and reinjected into the sorting device together with "control droplets".*

Since the optimal fusion and incubation conditions were found, the validation of the full evolution cycle can be performed. For that reason two strains Sol843 (0.01x) and POS (1x) were grown overnight in falcon tubes at 37°C. The next day, cultures were washed twice with media and filtered through 5µm filters. The solutions were diluted to achieve 30% average occupancy in droplets and mixed in equal volumes. Afterwards, cells were encapsulated in 4pL droplets and inoculated for 10h. The emulsion containing grown cells was fused with substrate/detergent 20 pL droplets. The resulted emulsion has been collected on ice, incubated for 40 min at room temperature and reinjected from the capillary kept on ice, into a sorting device, alongside 25 pL "control droplets" (50 µM Fluorescein, 0.1 µg/mL TexasRed). These "control droplets" will be used for normalization of protease libraries in comparison to the wild-type POS (1x) clone.

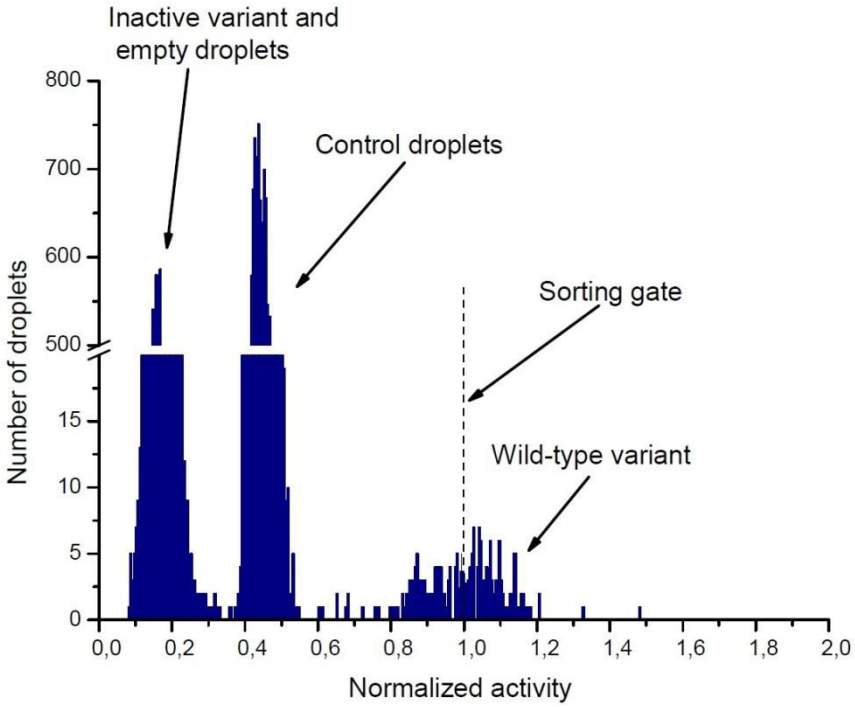

Figure 62. Model experiment for protease screening. *The histogram is representing droplet population distribution during the model sorting.*

From the fluorescence distribution diagram (Figure 62) we can distinguish three populations corresponding to: inactive variant Sol843 (0.01x) and empty droplets (between 0.1-0.3 fluorescence units); control droplets (0.4-0.55); wild-type POS (1x) (0.8-1.2). The sorting threshold was set to specifically enrich for the wild-type variant, 5% of total droplets. After 30 min sorting, the genetic material was recovered from 10^4 collected droplets.

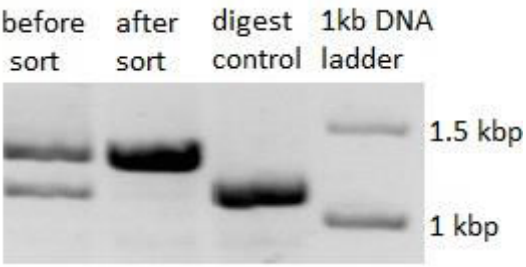

Figure 63. Agarose gel of digested DNA. *Before sorting: two bands correspond to digested Sol843 variant and undigested POS clone; after sort: the enrichment in heavy DNA band is observable.*

Since the inactive variant has a specific restriction site, the digestion with XmaI was used to discriminate it from the wild-type variant. The enrichment was calculated by the ratio of normalized band intensities and constituted 130-fold active variant over the inactive one. This value exceeds the theoretical enrichment allowed by Poisson-

distribution. For 1:1 ratio Sol843(0.01x):POS(1x) and 30% average droplet occupancy after initial cell encapsulation we have: 74% empty droplets, 11% of droplets occupied by a single Sol843 secreting bacteria, 11% of droplets occupied by a single POS secreting bacteria, 3.3 % of droplets occupied by two cells (1.67% - one POS and one Sol843; 0.83% - two POS; 0.83% - two Sol843). If we neglect the droplets occupied by three or more cells and assuming the equal and constant growth rates of cells in droplets with single and double starting occupancy, for the maximum theoretical enrichment we have η_{max} = (11% + 2*0.83% + 1.67%) / 1.67% = 8.6. Such a large deviation from the theoretical value can be explained by the false assumption that the protease activity from a droplet containing a single POS bacterium from the start is the same as from a droplet containing one POS and one Sol843 bacterium right after encapsulation. Eventually, the rates of bacterial growth and protease secretion decrease with each cell division cycle, as a result of nutrient depletion and metabolite accumulation. Due to the competition in growth between the POS strain and the Sol843 strain in the same droplet the number of POS protease molecules can be lower than in droplets where only POS strain grew from one cell. Moreover, the registered activity from the same enzyme amount can be reduced over time owed to autoproteolysis. This explains why it is not possible to see 1% droplet population corresponding to double POS initial encapsulation. Thus, when selecting only the right part of the peak of the most active droplet population, we are avoiding contamination with droplets containing both strains, which are probably hidden in the left shoulder of the same peak. This observation can have a beneficial effect on the library screening.

6. Protease library screening

The RDT8 protease library (10^6 variants) supplied by our industrial partner was generated by error-prone mutation of the wild-type variant (3-4 mutations per gene). During the first round of selection we monitored, as expected, a large distribution of activity (Figure 64). The top most active 1% of the droplet population (this corresponds to 3% of the most active variants of the library) was sorted, recovered and sent for recloning. The second screening (10^4-10^5 library size) showed an enriched population with the maximum in the wt area. Nevertheless the shape of the active population is quite spread covering the variants from 50% to 130% of POS activity. This wide

distribution can be the result of double encapsulations of cells in droplets during the emulsion generation or the activity loss resulting from of the transformation. During the 2nd screening the most active clones were reselected again and retransformed in cells.

The histogram from the 3rd screen (10^3-10^4 library size) represented a more narrow active population but in the same wt area (70%-130%). The top hits were reselected, recovered and sent to Novozymes. A part of the gene pool was error-prone mutated and further recloned; another part was transformed in the cells with no changes. To get a better discrimination of active clones, it was decided to reduce the incubation time after fusion from 40 to 30 min, which is the minimum off-chip time interval. The 4th screen (10^2-10^3 library size) of the initial library in the new assay incubation interval showed a shifted peak of the droplet population containing active clones (Figure 64). The peak has an asymmetric shape with a steeper left shoulder and a smoother right shoulder. This interesting population shape has been monitored in other evolution experiments as well [193], and is a result of enrichment in more active clones in the process of selection rounds together with the decrease in library size.

The sorting of the library generated from error-prone PCR on the hits selected after 3rd round revealed a flipped back activity as intended.

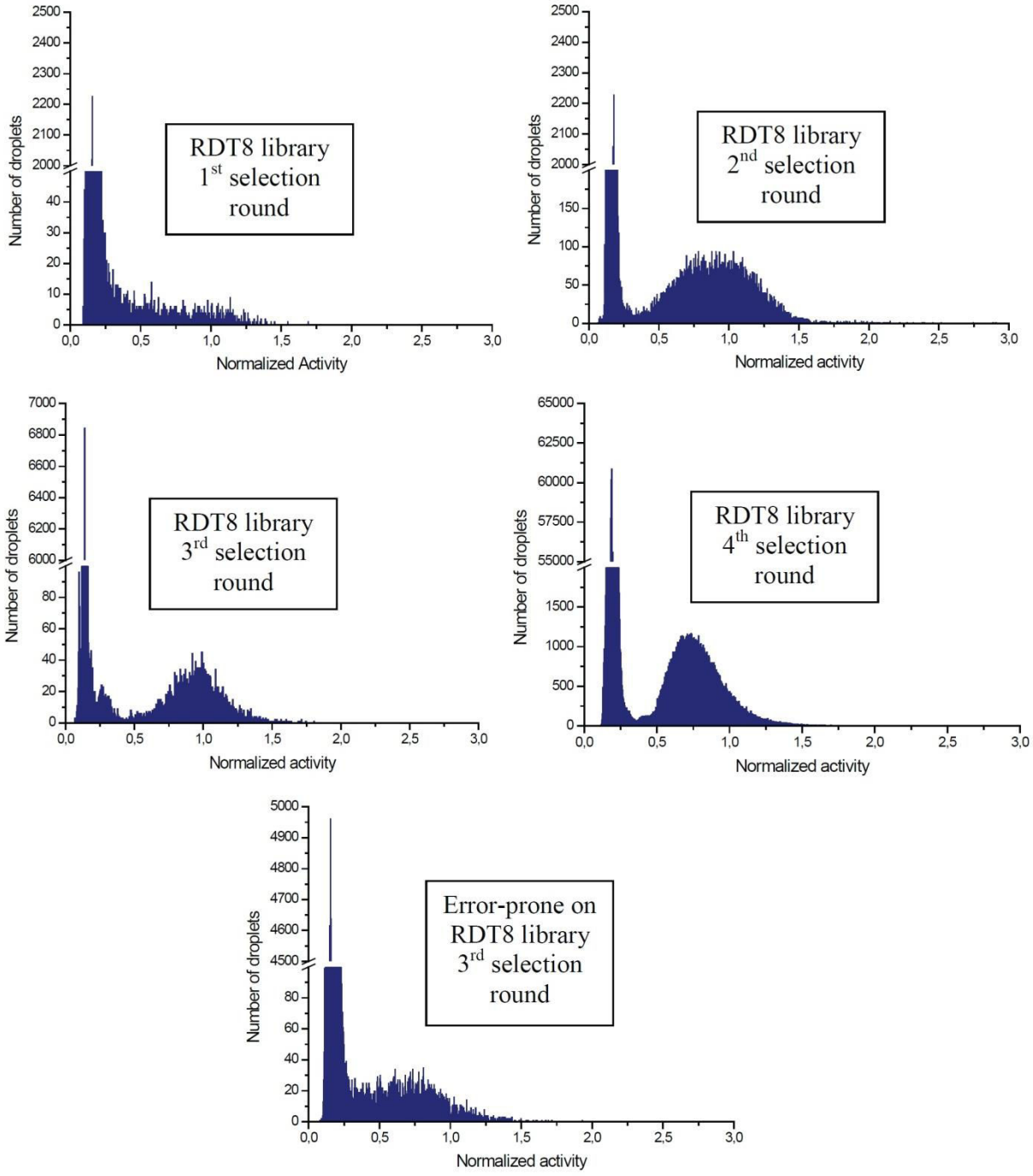

Figure 64. Error-prone PCR library profiles over the selection/mutation cycles. *The fluorescence intensities were normalized in comparison to the POSWT variant. The last two histograms were obtained from the screening after 30 min incubation, but were normalized by 40 min incubation of POSWT clone, which caused their shifts.*

Looking at the evolution of population distributions over the selection rounds it can be stated that the developed microfluidic platform can enrich a well defined active population only in two selection rounds. Although further selection may yield a smaller library, which is reach in active hits.

To monitor the quality and the composition of screened variants, 46-80 clones from the selected gene pools were sequenced (Figure 65). Although such a small sampling does not cover the whole number of selected clones, it may still give valuable information on the accumulation of positive mutations. In the same figure the decrease in wt-fraction can be easily observed. After the first library sorting only - 16.3% of all clones represented the POS-variant (data set 1), while after the second round of mutagenesis (data set 4) the variant was present only as - 4,3%. The premature drop of the wt-fraction in data set 2 (down to 2.1%) can be explained by low number of screened samples, only 48. While the sequencing of 96 clones after the 3rd selection showed - 6.3% wt (data set 3). Considering the presence of wt (4.3%) in data set 4 obtained from the error-prone PCR on the gene pool from the 3rd selection round, and the absence of wt after the sorting of the same library (data set 5), it is possible to assume that the majority of selected clones are better expressed or are more active than the wild-type variant.

In some cases a point mutation in an amino acid sequence is found to prevail only once in the sequencing data sets and not seen in others, like T, X, Y, Z. Probably, these mutations do not significantly influence the expression, stability or protease activity and are favoured by codon bias. The point mutation H is present in all data sets and its percentage increases until the 4th sequencing and decreases in the 5th. Such a distribution can be caused by many reasons and it is difficult to speculate without knowing the nature of the amino acid substitution. The most promising mutations seem to be A, B, C, D. All of them accumulate till the data set 4, where the error-prone PCR decreases their percentage. However, the point substitutions A-D are even more frequent after first sorting of second mutant generation (data set 5). This means that the presence of A-D mutations has a beneficial impact on protease activity in the detergent environment.

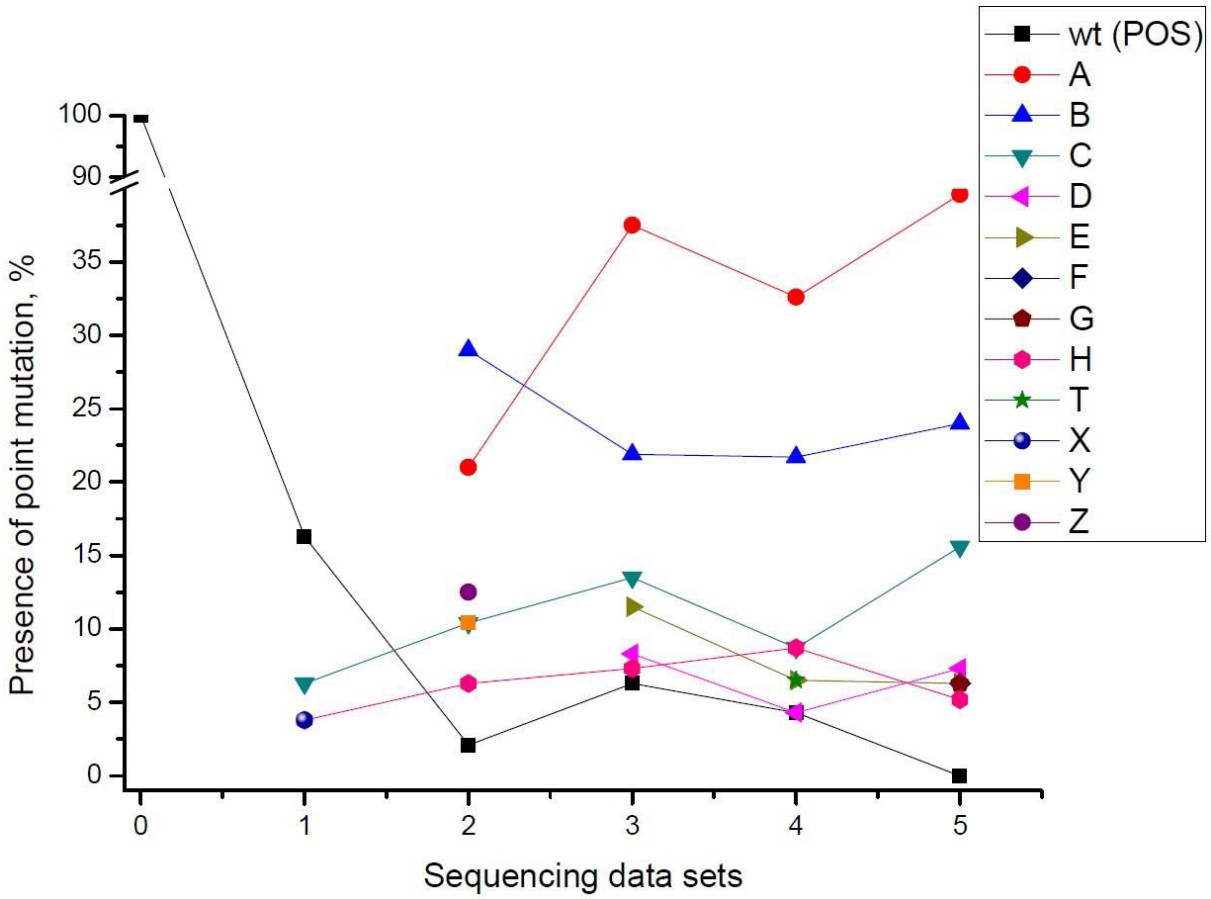

Figure 65. Sequencing data sets for protease screening rounds. *Data sets are representing the frequencies of certain point mutations in amino acid sequence of selected gene pool, wt(POS) – wild type variant (no mutation). Data set 1: 80 clones picked from the gene pool after the 1st sorting round; data set 2: 48 clones picked after 2nd sorting round; data set 3: 96 clones picked after 3rd selection round; data set 4: 46 clones picked right after error-prone PCR on 3rd selection round, no selection (second mutant generation); data set 5: 46 clones picked after sorting of error-prone PCR on 3rd selection round (1st sorting of second mutant generation).*

A more correct interpretation of the monitored frequency and effect of point mutations on protease activity, stability or expression can be performed after knowing the amino acid substitutions behind them and whether these mutations are accumulate on the same sequence, or on different sequences. The evolution results will be more complete after testing and comparing the individual mutants with the starting wild-type clone.

7. Conclusions and perspectives

Within the actual work a high-throughput microfluidic platform was developed and applied for screening of protease variants in specified assay conditions, mimicking closely the conditions of real enzyme application. In the near future, iterative evolution cycles will be continued towards further protease improvement.

In perspective, we are considering the development of an integrated module to allow the droplet fusion, followed by on-chip medium-term incubation (10-30 min), and followed by sorting on the same chip. Such an integrated chip will allow more precise discrimination between highly active variants. However, at this stage in directed evolution process, the enrichment in active clones is still possible using the microfluidic platform with off-chip incubation.

Another strong point of the developed platform is the flexibility for modification of assay conditions. The droplet-based microfluidic setup can be easily adapted to a new type and concentration of detergent; pH values; as well as, the screening temperature, which can be adjusted by incubating the droplets at lower or higher temperatures. Since the enzyme detergent market is continuously evolving, the availability of such a flexible and high-throughput screening system can quickly generate biocatalysts adapted to new customer preferences and technological breakthroughs.

Cooperation with the industrial sector allows the expansion of droplet-based microfluidics applications for protein evolution beyond academic examples. The screening of commercially useful enzymes can provide an additional proof for the efficiency of microfluidic platforms for directed evolution.

Completely *in vitro* microfluidic platform for directed evolution of *Streptomyces griseus* aminopeptidase

The high evolutionary interest in *Streptomyces griseus* aminopeptidase (SGAP) is related to its wide palette of displayed activities. SGAP is a promiscuous enzyme having detectable aminopeptidase, phosphodiesterase and cathehol oxidase activities. This, together with its small size (284 amino acids), heat stability, makes it an attractive model enzyme for fundamental studies of molecular evolution using droplet-based microfluidics.

Wild-type SGAP (Figure 66) is a zinc metalloprotein, containing two zinc (Zn^{2+}) ions in its active site. At the same time, SGAP is a metallo-activated protein, being stabilized and activated in the presence of calcium (Ca^{2+})[228]. It has been demonstrated that as aminopeptidase substrates SGAP prefer large, hydrophobic residues (Leu, Met, Pha) as the first or second amino acid. The phosphodiesterase activity is 800-fold lower than leucine aminopeptidase one, nevertheless, SGAP is only 10 times slower than the fastest phosphodiesterases with the same substrate. The secondary, phosphodiesterase activity is displayed in the same di-zinc configuration as for protease activity [229], while for

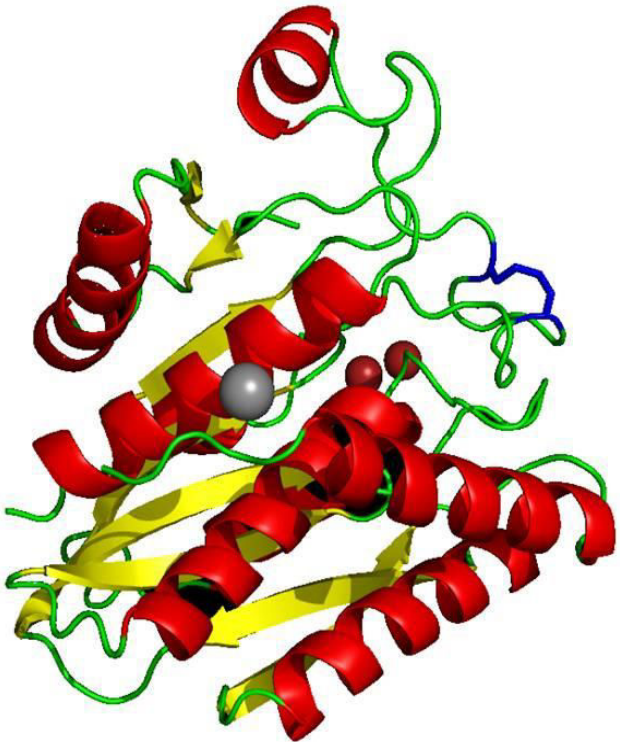

Figure 66. *Streptomyces griseus* aminopeptidase structure. *Structure of SGAP at 1.58 Å resolution. α-helices are indicated in red, β-sheets are indicated in yellow and the Cys245-Cys250 disulfide bridge is blue. The two zinc ions are represented by dark red spheres and the calcium ion is depicted as a dark gray.*

catechol oxidase activity SGAP requires the substitution of zinc by copper in the active site [230].

Table 4. Kinetic parameters of SGAP activities.

Activity	Substrate	$k_{cat}(s^{-1})$	$K_m(mM)$	$k_{cat}/K_m(s^{-1}\ mM^{-1})$
Aminopeptidase	Leu-pNA	441	0.55	802
	Phe-pNA	22	0.47	46
Phosphodiesterase	BNPP	0.45	4.5	0.1
Catechol oxidase	DTC	1.45	0.44	3.3

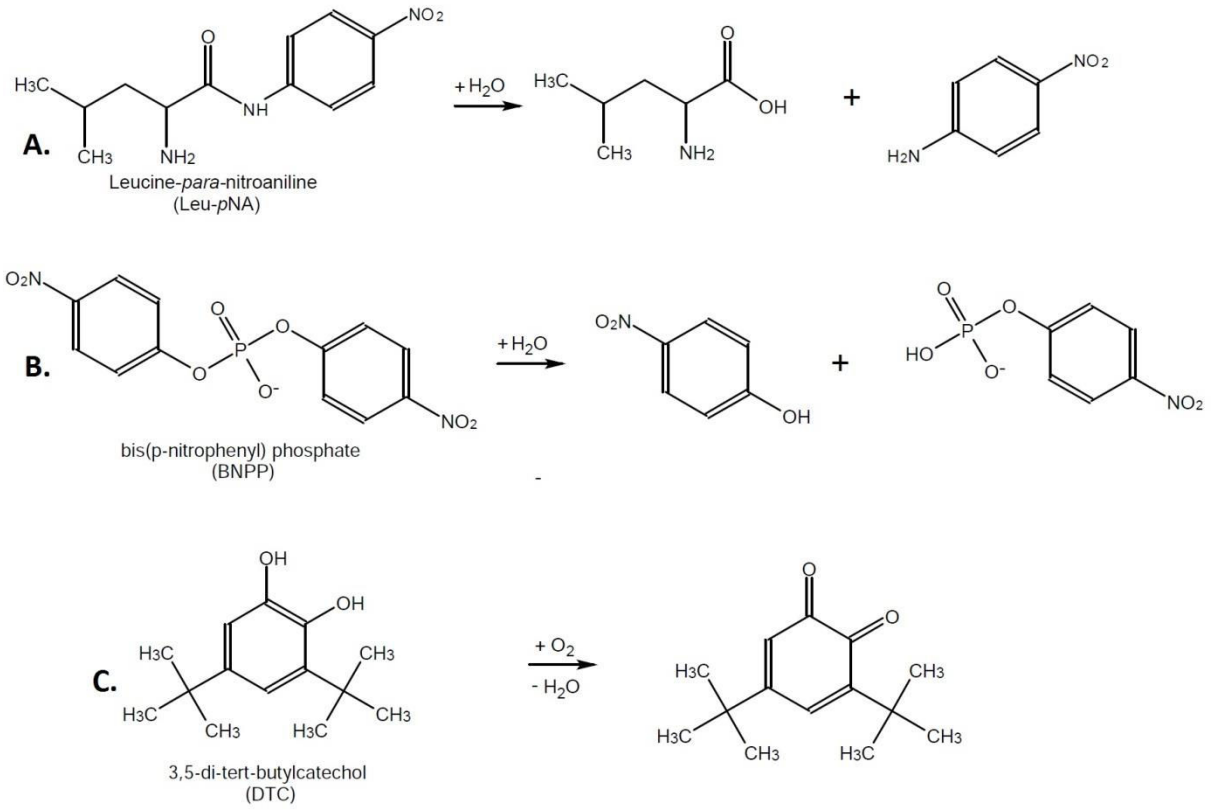

Figure 67. SGAP enzymatic activities. A. Aminopeptidase. B. Phosphodiesterase. C. Catechol oxidase.

When we look at the large meaning of the term "multiple activities", there are several types of promiscuity and SGAP cover most of them [231]. The broad substrate specificity for the peptidase activity means that SGAP possess cross-reactivity[a]; the phosphoesterase activity could be interpreted as multi-specificity[b] and moonlighting[c] [232]; and the catechol oxidase activity could be interpreted as promiscuity[d].

The long-term goal of this project is to address several outstanding questions in evolution biology through directed evolution of SGAP. For example:

- *How does genetic drift function?* For this purpose it is necessary to have a duplicated gene in the system and select for two activities simultaneously. Then, it would be possible to find out whether the parental genes evolve in a different direction for each function thus separating into specialists or whether there would be a different outcome;

- *What if the evolution is performed by counter selection?* Whether the sequence and function would differ between an evolved protein that is theoretically allowed to retain its original function, and one that is specifically chosen to lose the original function.

- *What is the influence of specific generalist selection?* If two reactions are at the same time, it is also possible to specifically select a protein that possesses both activities. It would be interesting to see whether this protein is more prone at evolving a new third activity.

---

[a] Cross-reactivity is when a protein has overlapping functions for ligands resembling the original ligand. They can also be substrate analogues.

[b] Multi-specificity is when a protein has a comparable function, but for a distinctly different ligand. This may or not be based on different residues than those involved in the original function.

[c] Moonlighting is when a protein has an additional function that is not based on its original active site. Often this is structural or regulatory, rather than functional.

[d] Promiscuity (poly-reactivity) is when a protein has multiple functions at a single active site. The mechanism of action is often different between the functions.

- *Is evolvability evolvable? [233]* The potential of an entity to evolve is called – evolvability. The question – "Why some proteins evolve faster than others?" is a contentious issue in evolution biology. For example, it seems difficult to assess whether evolvability can be evolved or selected. To prove it experimentally it is necessary to go back and forth between two activities in the scope of directed evolution. Hypothetically, several rounds of selection could be performed for phosphodiesterase activity, followed by several rounds of selection for aminopeptidase function. Going back and forth between aminopeptidase activity and phosphodiesterase activity probably will "train" the protein to be more evolvable. To prove it, the evolvability of "trained" enzyme could be compared with the wild-type clone towards the cathechol oxidaxe activity. Or that the time to switch between the different activities reduces with "training".

More advanced high-throughput screens and higher quality sequence libraries will make the searches easier and will enable evolution to answer these complex questions. Thus, in this project we are intending to build a completely *in vitro* droplet-based microfluidic platform for SGAP evolution.

I. Materials and methods

1. Substrates

Bis-(l-leucinyl)-rhodamine 110, Leu_2Rho, (Promega, Charbonniere, France) was dissolved to a concentration of 100 mM in DMSO and aliquoted in PCR-tubes.

Leucine p-nitroanilide, Leu-pNA, (Sigma-Aldrich)

Resorufin p-nitrophenylphosphate was kindly donated by Dr. Kerstin Blank (Nijmegen University).

2. SGAP activity assay in bulk

SGAP activity measurements off-chip were performed in microtiter plates on the spectrophotometer using the assay solution containing: 2 mM $CaCl_2$, 200 µM $ZnCl_2$, 1mM Leu-pNA (or 10 µM Leu_2Rho) and 40 mM Tricine (pH = 7.8). Absorption wavelength 400 nm.

3. In vitro transcription-translation in bulk

The following in vitro transcription-translation kits were tested using standard protocols:

- *Novagen*, EcoPro™ T7 System
- *Roche*, Rapid Translation System RTS 100, *E. Coli* HY
- *Invitrogen*, Expressway™ Cell-Free *E.coli* Expression System
- *Qiagen,* EasyXpress™ Linear Template Kit
- *Promega, E.coli* T7 S30 Extract System for Linear Templates
- *Cosmobio*, PureSystem™ Classic II
- *Cosmobio*, PureSystem™ SS II
- *New England BioLabs*, PURExpress™ In Vitro Protein Synthesis Kit

II. Results and discussion

1. Fluorogenic substrates for SGAP activities

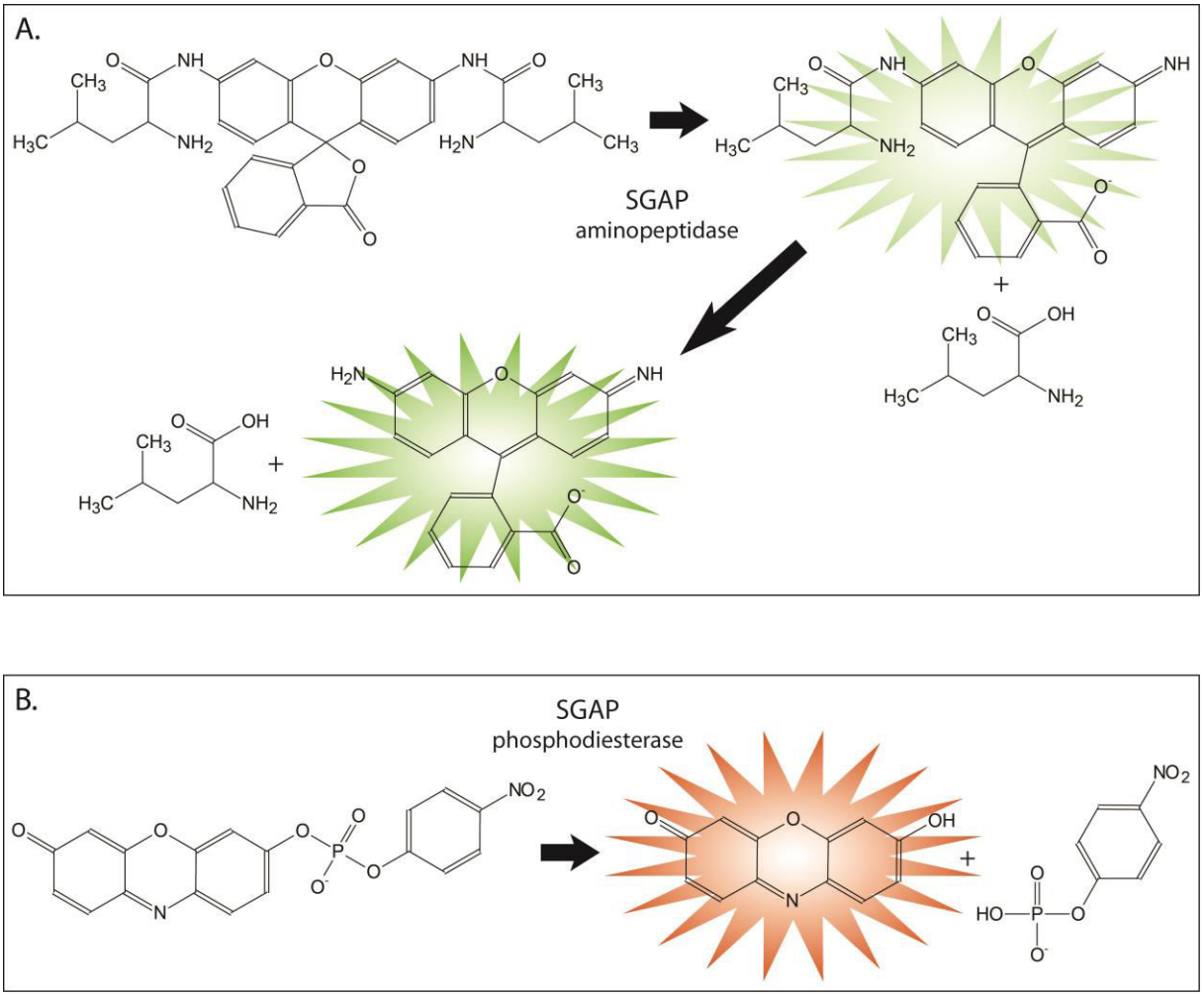

Figure 68. Fluorogenic substrates for monitoring SGAP activities in droplets. A. Aminopeptidase activity. *Bis-(l-leucinyl)-rhodamine 110 (Leu$_2$Rho) is first cleaved to monosubstituted Leu-Rho110 and further, cleaved to Rho110. Both Leu-Rho110 and Rho110 are fluorescent in green (λex = 488 nm and λem = 525 nm).* **B. Phosphodiesterase activity.** *The cleavage of resorufin para-nitrophenylphosphate liberates fluorescent resorufin (λex = 532 nm and λem = 585 nm).*

Fluorogenic substrate available for detecting the SGAP leucine-aminopeptidase activity in droplets is bis-(L-leucinyl)-rhodamine 110 (Leu$_2$Rho). The hydrolysis of this compound liberates rhodamine 110 (Figure 68. A), a fluorescent dye (λex = 488 nm and λem = 525 nm). Since, Leu$_2$Rho is a double substituted substrate; recorded fluorescent

signal in the course of aminopeptidase cleavage corresponds to the course of two reactions: hydrolysis of Leu_2Rho to Leu-Rho and hydrolysis of Leu-Rho to Rho. Resulted green fluorescence represents the sum of fluorescence of Rho and Leu-Rho, the latter is also green-fluorescent, having the quantum yield 4-5 times lower than Rho makes the analysis more difficult. However, the apparent Michaelis constants of SGAP activity for Leu_2Rho cleavage was measured in the work [234] : Km = 60 μM.

To monitor the phosphodiesterase activity we are using resorufin-p-nitrophenylphosphate, the cleavage of which gives red fluorescent dye resorufin (λex = 532 nm and λem = 585 nm). The kinetic constants for this substrate are not available.

It is thus, in principle, possible to measure simultaneously two enzymatic activities for the same enzyme in a droplet.

2. Optimal IVTT kit for SGAP expression

Before starting microfluidic experiment it is necessary to find suitable conditions for the *in vitro* expression of SGAP. For that reason a series of commercially available *in vitro* transcription/translation (IVTT) kits were first evaluated in bulk. Each IVTT sample (50 μL) was prepared according to the manufacturers' protocols and was supplemented with 0.5 μg of linear DNA of wild type SGAP or inactive SGAP-E131A. The DNA coding for these genes was introduced in a plasmid harbouring all the sequences required for efficient expression of proteins *in vitro* (start codon (ATG); stop codon (TAG, TGA or TAA); T7 promoter sequence at upstream from the coding sequence; ribosomal binding site (RBS; SD sequence) at approximately 10 bp upstream from start codon; additional sequence of 6 bp or more at downstream from stop codon). After 2 hours of incubation at 37°C, 10 μL of each sample were transferred into microtiter plate well containing 90 μL (1.1x) of assay reagents (1mM Leu-pNA, 2 mM $CaCl_2$, 200 μM $ZnCl_2$ and 40 mM Tricine). Apparent SGAP concentration was estimated from aminopeptidase activity calculated from initial slopes of plotted kinetic curves. The expression of completely inactive SGAPE131A was needed as a control. Results presented in Figure 69 showed that more than a half of IVTT kits display a significant background aminopeptidase activity. These expression systems are not suitable for SGAP directed evolution since they are not reflecting the exact activity of encoded gene. Thus, in

counter selection experiments towards improvement of phosphodiesterase activity, it will be impossible to monitor the loss of aminopeptidase activity. The background observed with kits from *Novagen*, *Roche*, *Invitrogen* and *Qiagen* can be easily understood if one conserved that these systems are based on bacterial extracts or lysates. In contrast, the IVTT systems from *Cosmobio* and *NEB* are reconstituted from the purified components necessary for *E. coli* translation [4]: 3 initiation factors (IF1, IF2, IF3), 3 elongation factors (EF-G,EF-Tu, EF-Ts), 3 release factors (RF1, RF2, RF3), ribosome recycling factor, 20 aminoacyl-tRNA synthetases, methionyl-tRNA formyltransferase and T7 RNA polymerase, *E. coli* 70S ribosome, amino acids, NTPs, *E.coli* tRNA and energy recycling system.

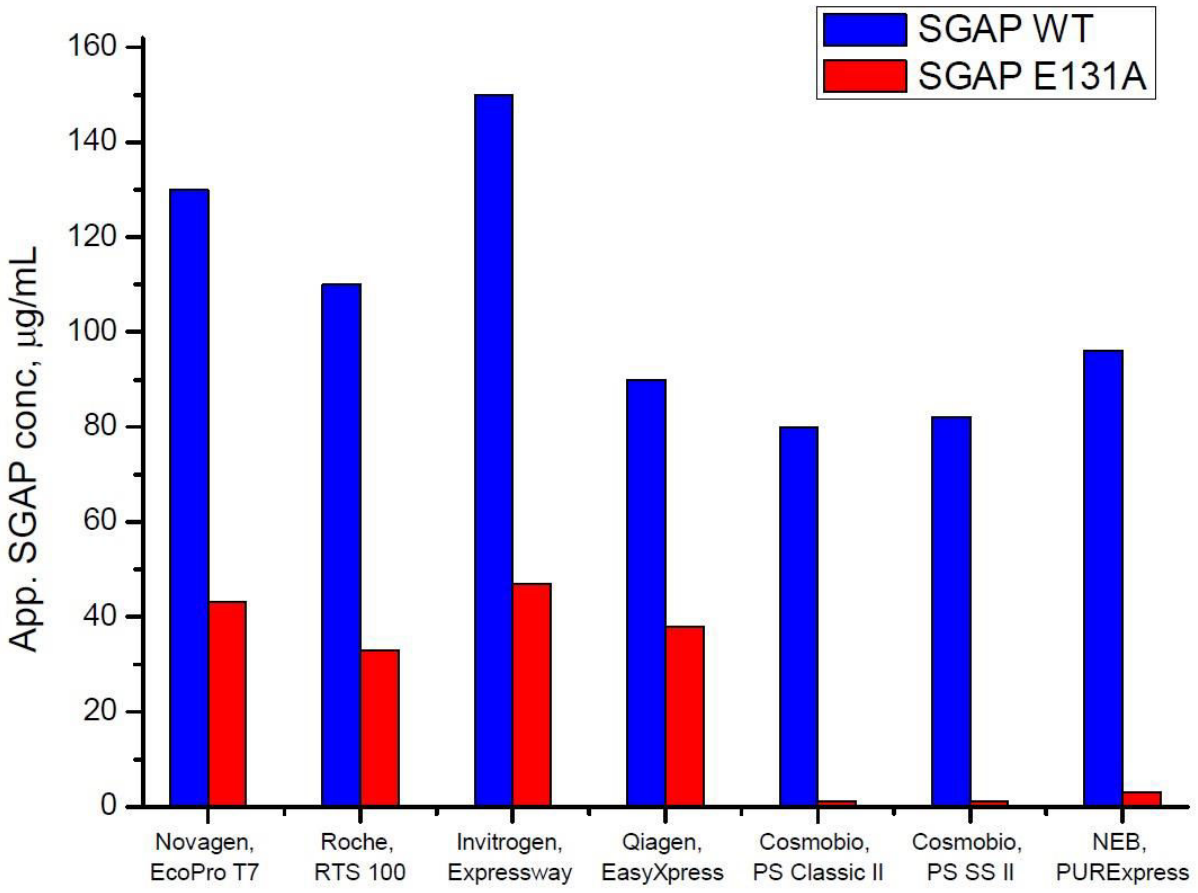

Figure 69. Expression of SGAP using *in vitro* transcription/translation kits. *The bars are representing the apparent SGAP concentration after 2 hours incubation at 37°C. In blue – the aminopeptidase activity of expressed wild type SGAP; in red – the aminopeptidase activity of expressed SGAP E131A (inactive variant), representing the background activity of the kit.*

Since SGAP contains a disulfide bond, special IVTT kits facilitating sulfide bridge formation were tested as well: Puresystem S-S II (*Cosmobio*) and Purexpress (*NEB*) supplied with Disulfide Bond Enhancer. Nonetheless, the activity and stability difference between SGAP expressed by standard kits and oxidative versions was hardly noticeable. Based on these observations, all further experiments on *in vitro* SGAP expression were carried out employing Puresystem Classic II or Puresystem S-S II from *Cosmobio*.

The studies on expression kinetics revealed as expected a 40 min lag-phase and a maximum protein concentration after 4 hours of incubation. When DNA template and IVTT mixture are incubated longer a decrease in SGAP activity was monitored. For example after 8 hours of incubation the activity of expressed SGAP was only 90% of the same mixture after 4 hours of incubation.

In conclusion, Puresystem (*Cosmobio*) has proven to be an efficient *in vitro* expression system for SGAP directed evolution, possessing no detectable background aminopeptidase or phosphodiesterase activity. The optimal incubation time for protein production lasts between 4-6 hours at 37°C.

3. Protein expression in droplets

Having established the optimal condition for IVTT reaction in bulk, the SGAP expression was evaluated in droplets. Pre-cooled IVTT mixture and SGAPWT DNA template were coencapsulated using a co-flow droplet maker (Figure 70). IVTT droplets were produced at rate of 5 kHz using HFE 7500 oil containing 2% surfactant. The volume of IVTT droplets was 20 pL with around $1.5*10^5$ copies of DNA per droplet.

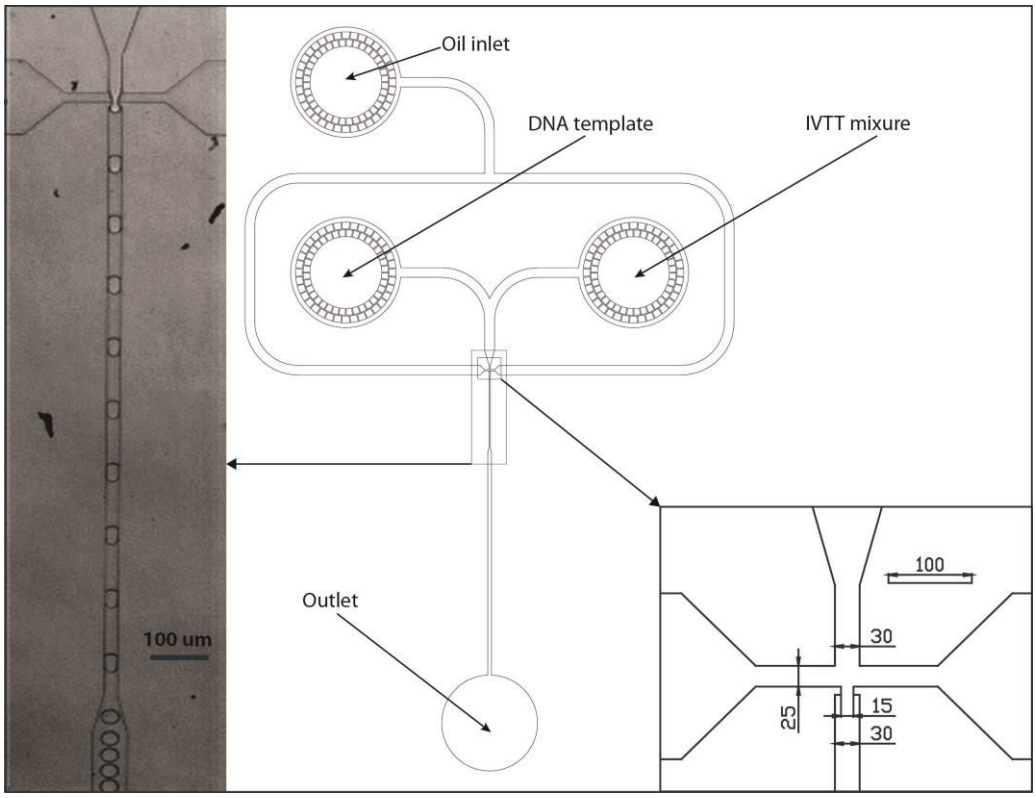

Figure 70. Microfluidic device for coencapuslation of two streams. *Depth 20 µm. The dimensions of all device features are indicated in µm.*

An emulsion, containing approximately 10^7 droplets was collected off-chip on ice, since the expression at 4°C does not occur. The emulsion was transferred to 37°C and incubated for 4 hours into an air thermostat to initiate the reaction and allow the production of enough protein. After incubation the emulsion was broken and the resulted solution was assayed for SGAP activity. Nonetheless, no aminopeptidase activity was detected. This means, that the protein was not produced in droplets. It was assumed that some of the components of IVTT were inactivated prior encapsulation, or as a result of adsorption on the droplets interface some of IVTT enzymes or cofactors were unavailable for SGAP expression. A quick test, in which the DNA solution and IVTT components were kept 1 h on ice before mixing in bulk showed significant amounts of expressed SGAP, proving wrong the first assumption.

To prevent the adhesion of biomolecules to droplets interface it is necessary to introduce surfactant-like molecules which will equally distribute on the internal droplet surface limiting the adsorption of precious enzymes. Therefore, BSA (bovine serum albumin) in one case and non-ionic surfactant Pluronic F-127 in other case, were separately emulsified together with IVTT mixture and DNA template. Analysis of the

aminopeptidase activity after incubation revealed that SGAP was expressed in droplets containing PEG-based copolymer Pluronic F-127 and BSA (Figure 71). Moreover, in the droplets where Pluronic was added to prevent adsorption, the production of SGAP was essentially identical to that obtained in microtiter plate. The effect of BSA on droplet surface is probably the same, but BSA has been proven to inhibit *in vitro* expression. The decrease in produced enzyme in droplets correlates with the amount produced in bulk in presence of BSA (data not shown).

In this manner, in the presence of only 0.4% Pluronic F-127 surfactant it is possible to create a favourable droplet interface for reliable *in vitro* expression of SGAP.

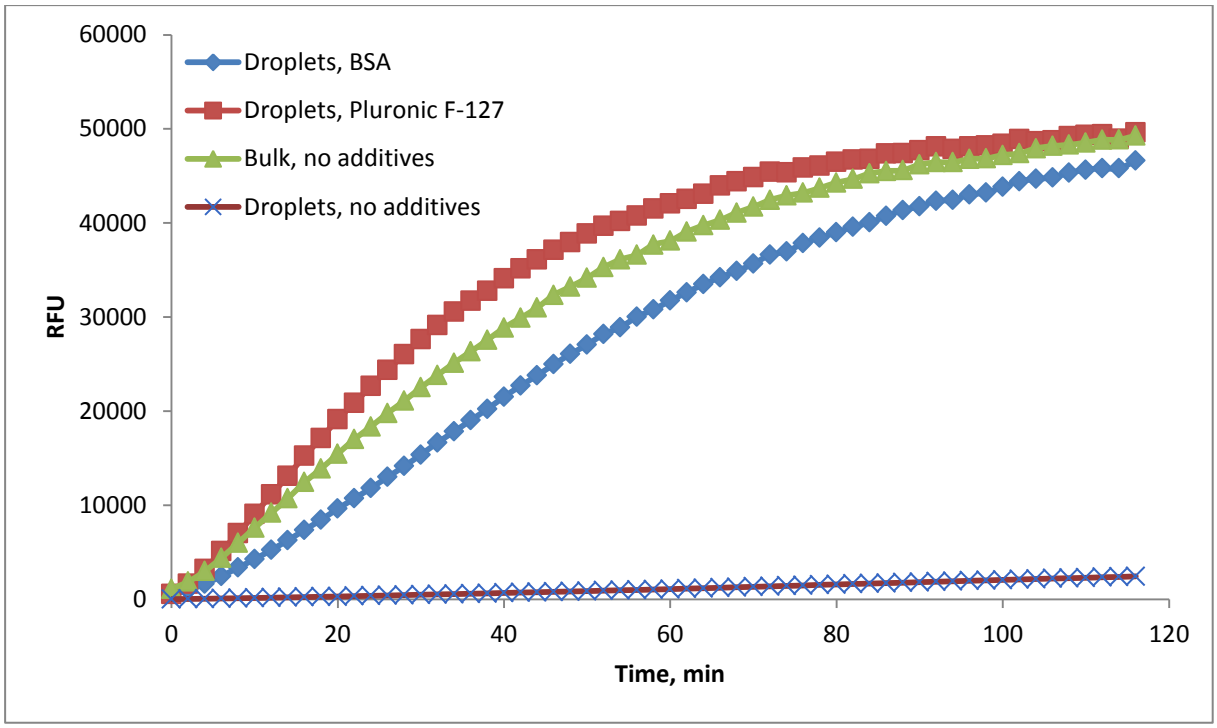

Figure 71. The activity of SGAP expressed in droplets. *IVTT reaction was performed in bulk; and in droplets in the presence of Pluronic F-127, BSA and without additives. 10 µL of IVTT mixture was mixed with 90 µL of assay mixture (2 mM $CaCl_2$, 200 µM $ZnCl_2$, 10 µM Leu_2Rho and 40 mM Tricine). The fluorescence was excited with λ = 488 nm and recorded at λ = 532 nm.*

4. DNA amplification in droplets

For directed evolution experiment, the SGAP library represented by linear 1.2 kb DNA fragments will be encapsulated in droplets following Poisson distribution to achieve one gene per droplet compartmentalization. Preliminary experiments on *in vitro* expression starting from a diluted solution of DNA template (corresponding to average of 1 DNA molecule per for 2 pL droplets), showed no detectable aminopeptidase signal. Thus, the expression of SGAP in droplets starting from a single gene is not possible, and the protein production must be accompanied or preceded by DNA amplification. Moreover, amplification will enhance the gene recovery efficiency after the sorting of improved variants.

As already discussed in the introduction part, the two available strategies for large DNA fragments (>500 bp) amplification are involving off-chip droplet incubation. The first involves the thermocycling of emulsion containing PCR mixture and the genes. But, the incubation at high temperatures can cause uncontrolled coalescence. In this regard, the use of other available fluorinated surfactant PFPE-JEF-M1000 is increasing the droplet stability at temperatures above 60°C [Michaël Ryckelynck, personal communication]. The second option is based on isothermal amplification using HRCA (hyperbranched rolling circle amplification), which has been already employed for building droplet-based microfluidic platform [210]. The last strategy does not need any special thermo stabilizing surfactant, but the gene can get contaminated due to the unspecific amplification of DNA impurities. Since the HRCA method amplify any circular DN molecule, special precautions must be taken to minimize contamination. To see which strategy is more suitable for single-copy SGAP amplification in droplets, both of them were tested.

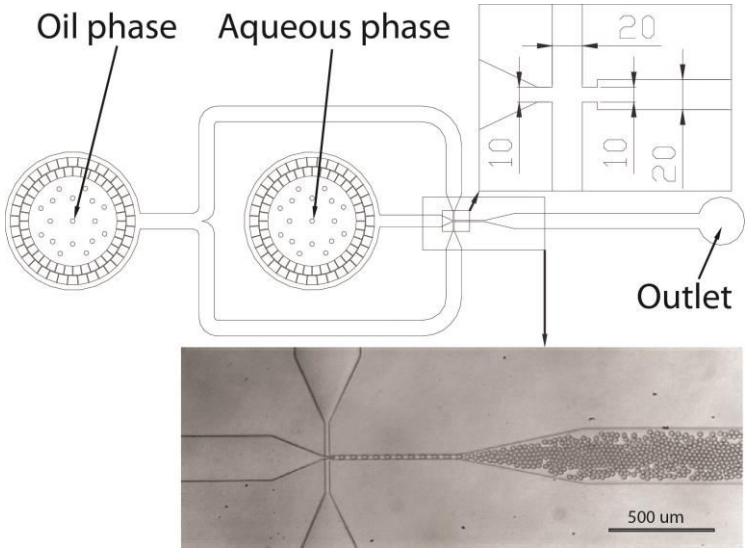

Figure 72. Microfluidic device for production of 2pL-sized droplets. *Depth 15 µm. The dimensions of all device features are indicated in µm.*

PCR mixture precooled at 4°C (containing 1 µM PIVB4 primer, 1 µM LMB10 primer, 300 µM dNTPs mixture, 0.1% Pluronic F-68, DreamTaq polymerase, buffer and linear DNA template) was encapsulated in 2 pL droplets using 2.5 % PFPE-JEF-M1000 surfactant in HFE7500 fluorinated oil. The emulsion was generated at 14 kHz frequency and collected for 30 minutes in a 500 µL PCR tube with a PDMS plug. After thermocycling the emulsion was broken and analyzed on an agarose gel.

The comparison of PCR in droplets and in bulk revealed a lower amount of amplification product obtained in emulsion at λ = 1. However, one should keep in mind that the molecules distribute in droplets following a Poisson distribution. Therefore when λ is 1, the actual expected droplet occupancy is only around 60%, therefore the total amount of DNA recovered from the emulsion is expected to be 60% of that of the same reaction performed in bulk or with all the droplets occupied (Figure 73).

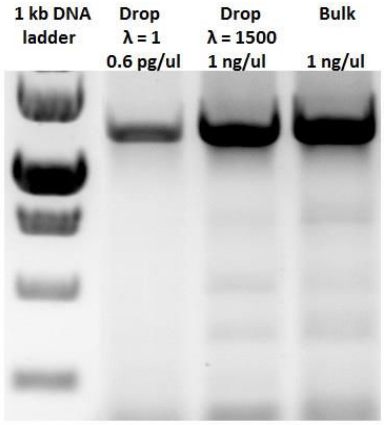

Figure 73. Gel analysis of emulsified and bulk PCR reactions. *Amplifications of linear DNA template encoding SGAPWT. After thermocycling PCR mixtures were analyzed on 1% agarose gel stained with ethidium bromide. The GeneRuler 1kb DNA Ladder (Thermo Scientific) was used as standard and the size of the bands are 250, 500, 750, 1000, 1500 bp. The DNA concentration used for each PCR is indicated on top of each band. For the PCR in emulsion, the average number of genes per droplet is described by λ number.*

HRCA system was tested using Illustra GenomiPhi V2 DNA amplification kit (*GE Healthcare*). According to manufacturer's protocol the sample buffer and DNA template (supplied with 1 ng/µL yeast RNA) were mixed and denaturated for 3 minutes at 95°C and cooled to 4°C. The Phi29 polymerase and reaction buffer were cooled to 4°C as well and mixed with denaturated DNA sample. Amplification mixture was encapsulated employing the same microfluidic system for droplet generation as in emulsion PCR experiment (Figure 72). Droplets were collected in a glass capillary and incubated for 4 hours at 30°C. The emulsion was then broken with, the aqueous phase recovered, the polymerase inactivated, the HRCA products purified by ethanol precipitation and monomerized with NotI restrictase. The linearized DNA was loaded on 1 % agarose gel.

The obtained results suggest that rolling circle amplification is highly efficient, generating 10^4-10^5 DNA copies from a single gene. In droplet experiments, it was possible to monitor amplification products from 5% average droplet occupancy. However, as the DNA template concentration goes down the products of non specific amplification are more pronounced (or not?), nevertheless the contamination is not significant.

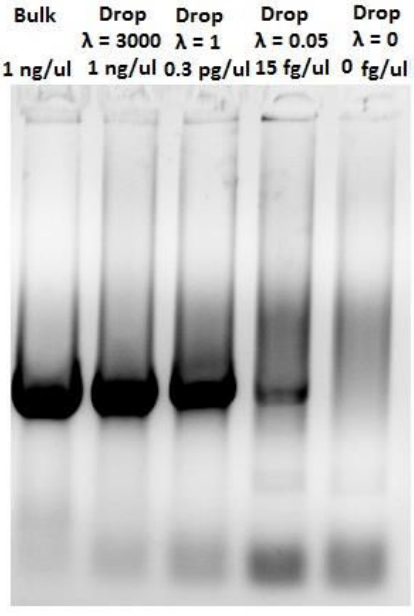

Figure 74. Gel analysis of emulsified and bulk HRCA reactions. *Amplifications of linear DNA template encoding SGAPWT. After HRCA incubation, DNA purification and monomerisation mixtures were analyzed on 1% agarose gel stained with ethidium bromide. The DNA concentration used for each isothermal amplification is indicated on top of each band. For the PCR in emulsion, the average number of genes per droplet is described by λ number. The ladder not shown.*

Comparing the two strategies for DNA amplification in droplets, the HRCA requires more precautious in DNA handling to avoid contaminants, but unlike PCR-amplification, produces significantly more DNA product and there is no droplets coalescence caused by thermocycling (2-5% of thermocycled emulsion were coalesced after PCR amplification). Considering the reasons above the amplification in

droplets using hyperbrached rolling circle amplification was chosen for SGAP directed evolution experiments.

5. How many fusion steps?

Droplet based-microfluidics allows a high-throughput manipulation of emulsions, like adding picoliter-sized aliquots to already formed droplet. Although, there are many described fusion methods their integration is not trivial and requires the tuning of microfluidic device configuration, flow-rates etc... Also, during the iterative directed evolution cycles additional droplet manipulation steps will make the screening of improved variants more complex and will challenge the dispersity of the emulsion.

Ideally, an *in vitro* screening experiment staring from a single gene consists of several consecutive steps:

- DNA amplification
- *In vitro* protein expression
- Activity assay
- Sorting of improved variants

In terms of microfluidic operations, to the droplets with amplified DNA, it is necessary to add IVTT mixture by first fusion, and assay components by second fusion, followed by sorting. Such a complex microfluidic platform can be difficult to operate and every fusion step introduces a certain degree of droplets polydispersity and cross-contamination. In published examples of microfluidic platforms with *in vitro* expression, only one fusion step was required. In the work [214], CotA gene amplification step was omitted due to the possibility to express detectable quantities of enzyme from a single gene, when in the work [210] the IVTT mixture and assay components were added simultaneously. In this regard, it is worth trying to simplify microfluidic platform for SGAP evolution as well, by coupling together two or more stages of the screening.

a. Coupling gene amplification and protein expression

Since the amplification requires isothermal conditions (30°C) close to the *in vitro* expression temperature (37°C), it was decided to couple amplification and expression in a single step. The IVTT kit contains more than 30 enzymes and slight variations in pH and buffer composition can drastically affect the expression efficiency. For this reason, it is rational to suppose that is easier to adapt one enzyme, Phi29 polymerase, to the conditions of reactions of other 30 expression enzymes, than the reverse. The comparison of the buffers revealed the close values of pH (pH = 7.3 for IVTT mixture and pH = 7.5 for phi29 buffer) and a series of compounds in almost similar concentrations (DTT, Mg^{2+}, NH_4^+). But, when single stranded DNA template (0.1 ng/µL), hexamers, dNTPs and phi29 polymerase were added to IVTT mixture, no amplification or expression product was detected at 37°C or 30°C. Variations in concentration of different constituents of IVTT buffer lead only to modest results, 2-3 amplifications per gene. Probably, Phi29 polymerase is inactivated or inhibited by some of IVTT buffer components (like creatine phosphate, spermidine, putrescein, etc), or other uknown reasons.

In conclusion, the amplification and SGAP gene expression need to be performed consecutively.

b. Coupling protein expression and activity assay

The other option for minimizing microfluidic manipulations consists in adding the assay components together with expression mixture. To test their compatibility we performed following experiment. A series of samples IVTT mixture supplied with 10 ng/µL of SGAPWT DNA were prepared. Further, to different tubes we added separately one of the assay buffer components ($CaCl_2$, $ZnCl_2$, Tricine and DMSO, supplied to increase substrate solubility). After 2 hours incubation at 37 °C, 5 µL aliquots were taken from each sample to estimate the efficiency of SGAP expression. Except for the beneficial effect of 2 mM $CaCl_2$, all other additives inhibited *in vitro* protein production (Figure 75). Although the presence of tricine is not mandatory since it's added only for pH regulation in assay buffer. DMSO can be avoided as well using other co-solvents less harmful for IVTT reaction. The biggest issue represents the significant inhibition of *in vitro* expression by Zn^{2+}, which obligatory for SGAP aminopeptidase and phosphodiesterase

activity. This effect has been previously monitored in the work of Alirezaei *et al.* [235], where zinc inhibited protein synthesis in cultured brain cells from the cerebral cortex of embryonic mice. Apparently, zinc ions reversibly inhibit the elongation and the initiation steps of the protein translation process.

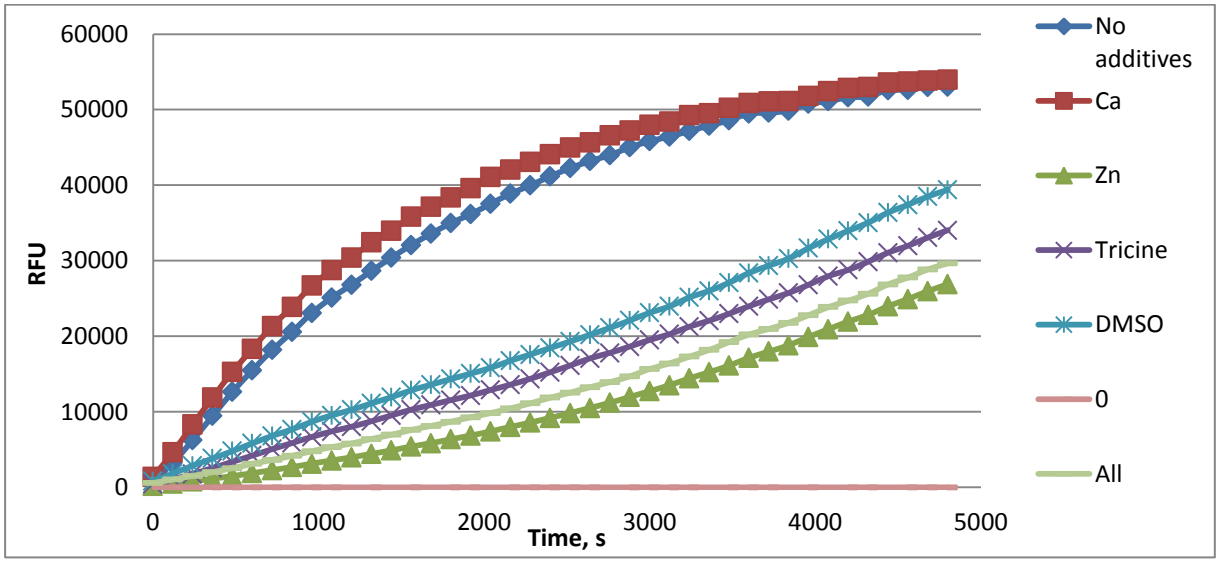

Figure 75. The influence of assay reagents on protein expression. *Several IVTT reactions were performed in bulk each in the presence of one of the components: 2 mM $CaCl_2$, 200 µM $ZnCl_2$, 40 mM Tricine, DMSO, all of them together and without additives. Then 10 µL of IVTT mixture was mixed with 90 µL of assay mixture (2 mM $CaCl_2$, 200 µM $ZnCl_2$, 10 µM Leu_2Rho and 40 mM Tricine). The fluorescence was excited with λ = 488 nm and recorded at λ = 532 nm.*

The optimization of zinc concentration revealed, that at 10 µM $ZnCl_2$ (instead of 200 µM) it is possible to achieve 80% of protein expression with only 25% loss in SGAP aminopeptidase activities, making an acceptable compromise to start with.

c. Time limitations of assay caused by product leakage

The coupling of SGAP expression with assay means that droplets should be incubated for 2-4 hours to allow sufficient protein production. In parallel, fluorogenic substrates will be converted by constantly produced enzyme. Thus, the fluorescent products must stay confined in droplets during all incubation time.

However, previous studies shown that Rho110 - the product of aminopeptidase reaction, as well as resorufin – the product of phosphodiesterase reaction are prone to

exchange between droplets (Figure 68). The retention times in droplets for these substrates were calculated before (Figure 76). Although, in these experiments the composition of either buffer or fluorinated oil (FC-40 in case of resorufin) are not the same as in our experiment, rhodamine and resorufin are likely to leak in a time interval of 2-4 hours for our conditions. Thus, the *in vitro* protein expression must be uncoupled from activity assay procedure.

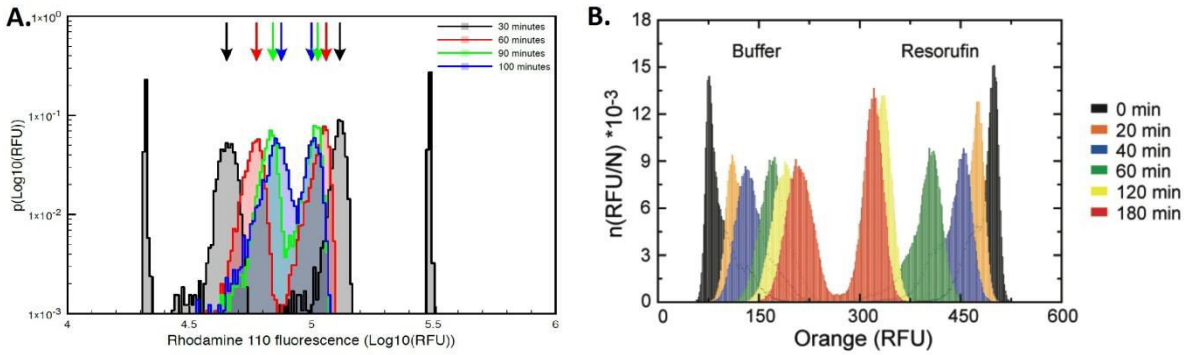

Figure 76. Retention time in droplet for rhodamine110 and resorufin. A. *The exchange rate of rhodamine 110. Droplets populations contained 10 μM Rho110 or 100 μM Rho110 in 40 mM HEPES pH 8.5 [234].* ***B****. Resorufin exchange between the droplets created with FC-40 oil [163].*

In conclusion, all tries to simplify the microfluidic workflow for SGAP evolution were unsuccessful due to the high sensitivity of amplification and expression systems to the changes in reaction media, as well as limited choice of fluorogenic substrates. In this case the microfluidic platform should contain following steps: generation of droplets containing single gene, amplification in droplets, fusion with expression components, IVTT in droplets, fusion with assay components, droplet sorting triggered on fluorescence.

6. Model experiment

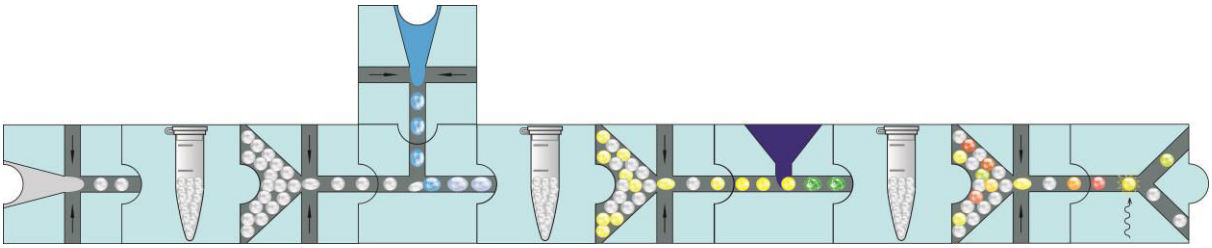

Figure 77. The microfluidic workflow for completly in vitro SGAP screening. *DNA solution with HRCA components are coencapsulated in 2 pL droplets to achieve single gene per droplet; collected emulsion is incubated at 30°C for amplification to occur; further, the droplets are reinjected into electro-coalescence device for fusion with 25 pL droplets containing in vitro transcription/translation mixture; resulted emulsion is incubated at 37°C for protein expression; after, 4pL of assay mixture is infused into the droplets with expressed protein; short incubation is followed by fluorescence analysis of both SGAP activities and sorting of the desired fraction.*

In order to assemble all parts of the developed microfluidic platform it is necessary to integrate successively all screening steps. For the model selection experiment we equally premixed DNA encoding active and inactive variants of enzyme SGAPWT : SGAPE131A – 1 : 1. DNA sample was diluted to achieve 30% average droplet occupancy (λ = 0.3) and coencapsulation with amplification components was done following previously described procedure for HRCA in droplets. Collected droplets were incubated at 30°C for 6 hours. After incubation, the emulsion consisting of droplets with amplified fragments was fused with droplets containing IVTT mixture. For the comparison reasons with previous experiments (give again the ref here to avoid misunderstanding), electro-coalescence has been chosen as a fusion technique (Figure 78). Droplet pairing and fusion on the electro-coalescence device has been optimized and tested in some previous fusion experiments performed in the laboratory.

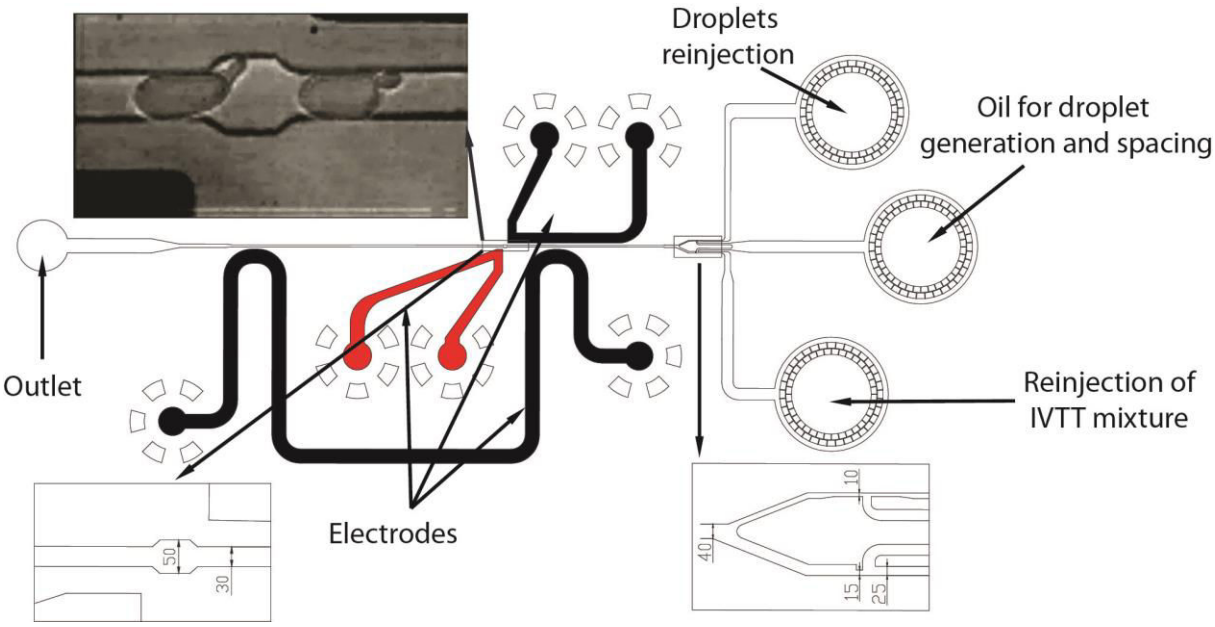

Figure 78. Droplet fusion based on electro-coalescence. *The depth of the channels was 20 μm The sizes are given in μm.*

Bulk experiments revealed that when HRCA buffer was added in a small ratio to IVTT mixture, no significant influence on the efficiency of *in vitro* transcription-translation is observed. Therefore, we choose to dilute 2pL DNA containing droplets in 25 pL of IVTT reagents.

2pL droplets with amplified DNA were reinjected from capillary kept on ice into the droplet fusion device at 20-40 μL/hr. While, droplets with IVTT mixture were generated directly on the droplet fusion device. To do so, the solution with expression mixture was infused, also from ice, at 150-200 μL/hr flow rates. The HFE7500 oil supplied with 2% surfactant was injected in the middle inlet (Figure 78) at 250-300 μL/hr. The oil flow was spitted in two on the device: one part was spacing the reinjected droplets with DNA fragments; the other was used to generate 25 pL IVTT-containing droplets. The flow-rates were adjusted to synchronize reinjected droplets and newly formed ones. Pair wised droplets, one 2 pL droplet with HRCA products and the other 25 pL droplet with expression components, were fused together when passing through the region with close proximity to the active electrode (in red on Figure 78) where an AC field of 80-120 V (current frequency 30 kHz) was applied. The interface between droplets in contact is shortly destabilized and passing droplet-pairs are fusing. The resulted droplet is stable since the droplet interface is quickly (μs intervals) rebuilt. The

red electrode is constantly on, allowing the fusion of each passing droplet pair, approximately 1.5 kHz. This way, the amplified genes are mixed with expression elements, preserving the compartmentalization. The collected emulsion after fusion was incubated at 37°C for 4 hours.

To add the assay reagents to the droplets containing already expressed enzyme we have chosen the pico-injection technique. The other fusion methods are not so efficient for adding small volumes (1-8 pL) into already formed droplets. The droplet size should be maintained as small as possible because, the increase in droplet volume is always accompanied by loss in sorting throughput.

Pico-injection of assay reagents into the droplets containing expressed protein has been performed using the design from the work [164], which was further optimized in our laboratory by Faith Coldren (Figure 79). Droplets with expressed SGAP variants were injected into the microfluidic chip at 50 µL/hr and spaced with 200 µL/hr HFE7500 supplied with 2% surfactant. The pico-injector with assay mixture is maintained at 40 kPa, and to the red electrode (Figure 79) was applied 100 V AC (current frequency 40 kHz) to allow 4 pL assay mixture injection into each droplet. The throughput of the pico-injection constituted around 1 kHz. The concentration of assay constituents in the resulted 31 pL droplet were: 2 mM $CaCl_2$, 200 µM $ZnCl_2$, 40 mM Tricine, 100 µM resorufin-P-para-nitrophenyl, 20 µM Leu_2Rho. Collected emulsion in capillary placed on ice was further incubated for 1 hour at 37°C.

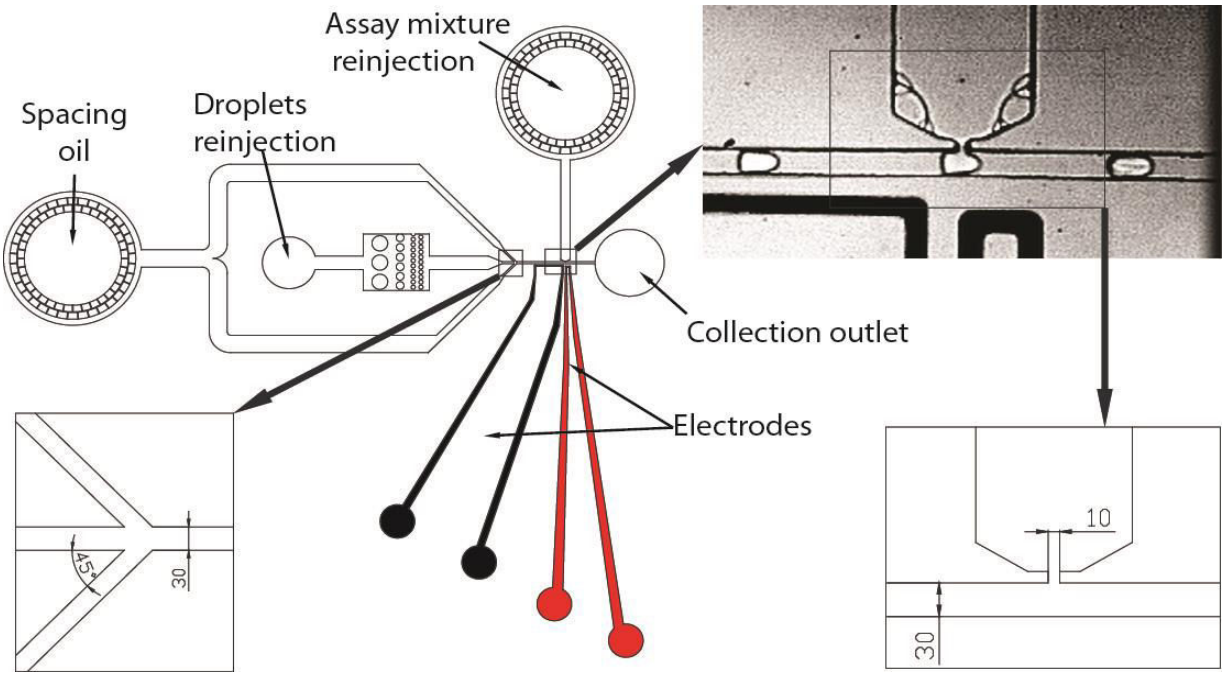

Figure 79. The pico-injection of small volumes into passing droplets. *The depth of the channels was 20 μm The sizes are given in μm.*

The analysis of droplets fluorescence and sorting of SGAPWT variant was performed using the standard droplet sorter employed in other projects as well. To be capable to register both activities the droplets fluorescence was monitored with two lasers/PMT systems at the same time (Figure 25).

The data analysis of 2D-plot histograms revealed the presence of droplet population containing the SGAPWT variant (Figure 80). Plotting the droplet size over the green fluorescence, which corresponds to aminopeptidase activity Figure 80. A, shows two distinct populations: one, corresponding to empty droplets and droplets with amplified inactive variant (SGAPE131A); the other, in more fluorescent region, corresponds to wild type clone of SGAP. The percentage of droplets with high green fluorescence constitutes almost 6% of overall droplets number. This data does not correlate with theoretical occupancy: for 15% average occupancy, SGAPWT should be present in 12% of all droplets. This fact can be explained by DNA loss before or during the encapsulation.

To observe the phosphodiesterase activity, we plotted the droplets size towards the orange fluorescence (Figure 80. B). In this case, there is no definitive separation of SGAPWT droplet population as in the green channel; nevertheless it is possible to monitor the presence of certain fraction of droplets with more intense orange

fluorescence. And finally, the plotting of green fluorescence over the orange fluorescence brings out the presence of two distinct populations (Figure 80. C.): one, in the region with low green and low orange corresponding to empty droplets and high green-low orange corresponding to SGAPWT.

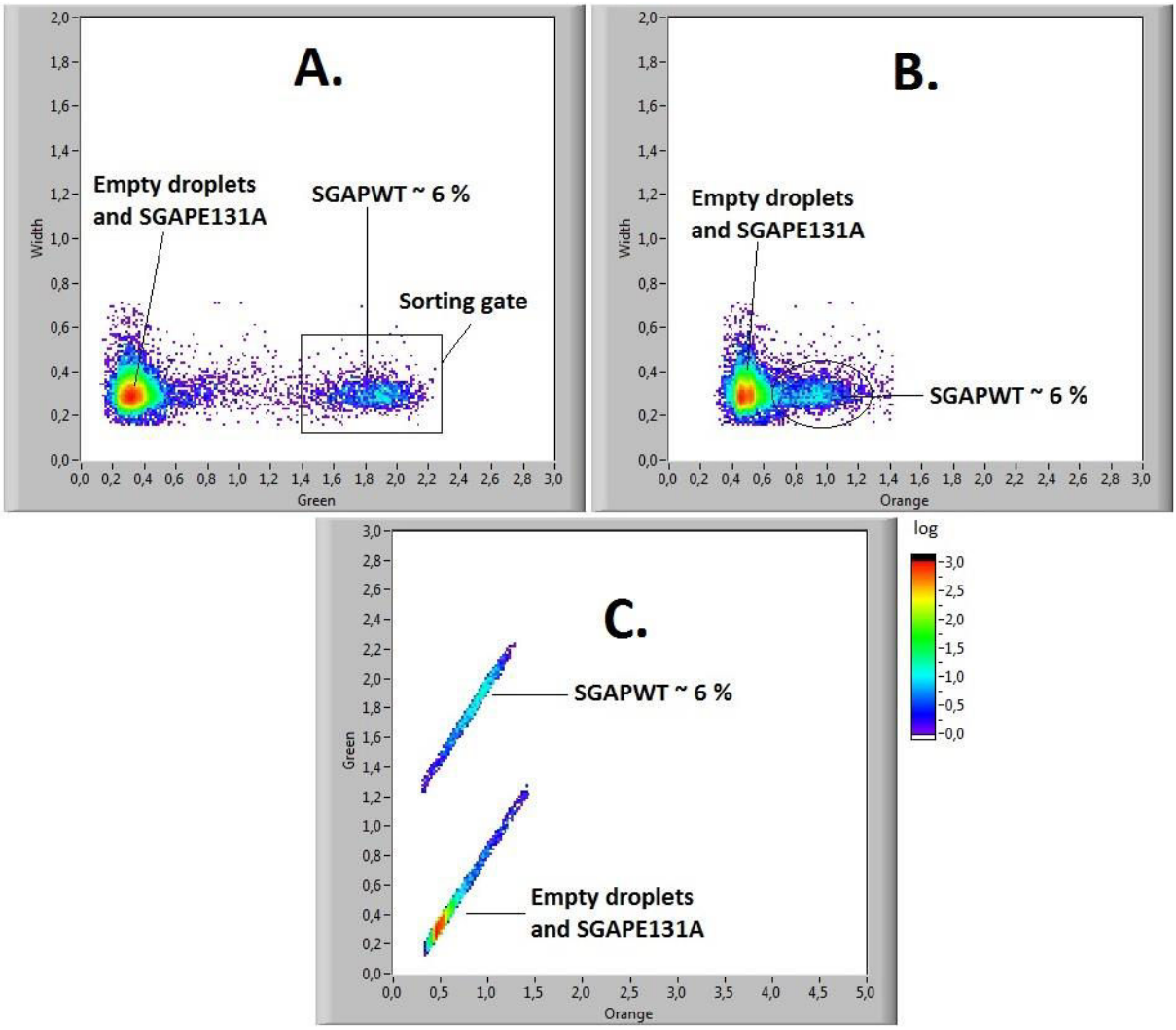

Figure 80. 2D plot histograms recorded during the model experiment. A. *Droplet size is plottet versus Green signal(V) correspondingto aminopeptidase activity. It is possible to monitor a well defined population of droplets with high Green fluoresence, about 6%.* **B.** *Droplet size plottet versus Orange signal(V) corresponding to phosphodiesterase acivity. The active droplet population is overlaping with empty droplets and droplets containing inactive variant.* **C.** *Green signal (V) versus Orange signal (V). Two populations are well separeted.*

Since the SGAP kinetics were not analysed in droplets, it is difficult to say whether the recorded aminopeptidase activity (or phosphodiesterase) is on the initial kinetic linear slope or whether it is a result of full substrate conversion. In addition, it is necessary to take into account the leakage of fluorophores. In the work of Chen *et al.* [236], it has been described the competing dynamics of the generation of fluorescent molecules in positive drops, against its leakage into the negative drops, where such molecules are absent. The authors are supplying a kinetic model to estimate the optimal incubation time to achieve maximum signal-to-noise ratio, knowing the leakage rates of fluorophore and the kinetic parameters of the enzymatic reaction. However, in our system we are monitoring two SGAP activities, both based on leaky fluorescent products. Thus, even if we had the data for SGAP kinetics in droplets and fluorophores leakage, the optimal incubation time calculated for one activity can be far from optimal incubation of the second activity. The decrease in the resolution of SGAP variants is another reason to avoid leaky droplets assay.

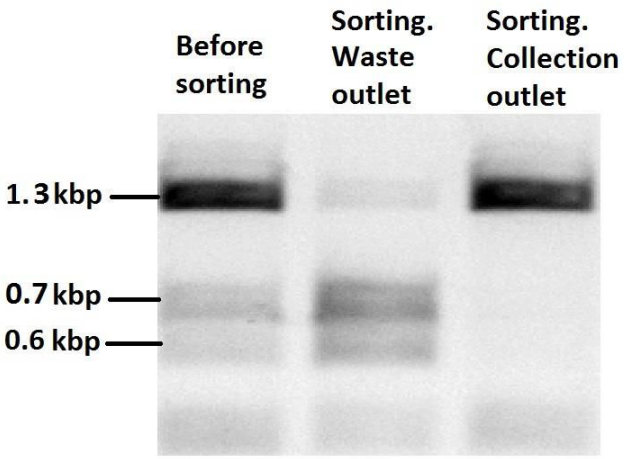

Figure 81. Agarose gel of digested DNA. *Two bands at 0.6 kbp and at 0.5 kbp correspond to digested SGAPE131A. The 1.1 kbp band corresponds to undigested SGAPWT. After sort: the enrichment in heavy DNA band is observable.*

The sorting gate was set to enrich for the SGAPWT variant. After 2 hours of sorting, collected droplets were broken, the content DNA recovered by PCR. To distinguish the SGAPWT and SGAPE131A, the DNA mixture was digested with SacII restrictase resulting in 8-9 fold enrichment of the active clone.

In conclusion, the experiment proved the possibility to screen SGAP clones for two activities using developed completely *in vitro* microfluidic platform.

7. Conclusions and perspectives

Within this project we developed and validated an entirely *in vitro* microfluidic platform for library variants screening of *Streptomyces griseus* aminopeptidase. In the future this platform will be employed for directed evolution of SGAP. This high-throughput droplet-based microfluidic platform will be combined with next generation sequencing for genotype-phenotype mapping of large population of variants 10^6 generated from the initial wild-type clone. Thus, for every evolution round the DNA sequence of a pool of variants will linked to level of aminopeptidase and phosphodiesterase activity. This will allow deep understanding of evolutionary pathways of an enzyme and to answer fundamental evolutionary questions, stated above.

For better discrimination between enzyme variants and more reliable assay, the developed platform needs some improvement. In order to be able to monitor precisely the enzyme's activities in droplets, it is necessary to replace the fluorogenic substrates used in the existing platform by monosubstituted fluorogenic substrates, whose cleavage release fluorophores resistant to leakage between droplets. Another concern is related to the different expression levels between the variants within the same library. The existing system does not discriminate the level of enzyme activity from the protein expression, thus cannot distinguish if the increase in substrate conversion rate is due the improvement in activity or in expression.

In principle, the developed platform can be applied for *in vitro* directed evolution of any enzyme. To readapt the platform for another enzyme directed evolution it is necessary to substitute only the components of assay mixture with a new suitable substrate and preferred buffer.

Since the platform is developed for completely *in vitro* screening, it allows overcoming the issues related to enzyme expression by cells. A more detailed comparison of droplet-based microfluidic platforms with expression *in vivo* and *in vitro* is to be found in the section: "General discussion".

General discussion

This work describes the development of three different high-throughput (10^6-10^7 variants a day) droplet-based microfluidic platforms for directed evolution of enzymes. Each microfluidic platform is tailored to the requirements of the particular enzyme to be investigated and designed to specifically improve the desired biocatalytical properties. For example, to monitor the activity in droplets of *de novo* designed retro-aldolase we synthesized and tested a series of new fluorogenic substrates; or, for creating a specific environment for protease screening we adapted a droplet fusion technique for adding the detergent.

Development of the microfluidic platform is always a delicate interplay between both microfluidic operations from one part and expression systems with assay conditions from the other. Since directed protein evolution is an iterative process consisting of many repetitive rounds of mutagenesis and screening, the easier are the platforms in handling, the more efficient and faster the evolution rounds can be carried out. When the enzyme library was displayed *in vivo* (cytoplasmic expression with cell lysis after compartmentalization - retro-aldolase evolution project; or secretory expression in droplets during cell growth - protease screening project) the microfluidic set-up for screening was simpler, since cells replicate the DNA and produce large amounts of protein. For example, in the retro-aldolase evolution project the microfluidic workflow consisted of only three microfluidic steps (cell encapsulation in droplets, incubation and droplet sorting), which made possible the integration of all three steps on a single microfluidic chip, minimizing droplet manipulation. For this reason, platforms built with *in vivo* enzyme expression advanced further in directed evolution of the corresponding biocatalyst.

When the enzyme library was generated using *in vitro* expression, the set-up was more complicated. In the project SGAP evolution, because of the incompatibility of DNA amplification with *in vitro* expression, as well as incompatibility of the expression with assay conditions, the compromise was achieved by the integration of additional microfluidic steps. Droplet-based microfluidics offers a wide palette of droplet

manipulation techniques, which can be adapted and integrated to meet the particularities of a biological entity and screening conditions. One of the most important tool in microfluidics allows new reagents to be added to pre-formed droplets at defined times, to start, modify, or terminate a reaction. For the *in vitro* SGAP platform, after amplification of single DNA molecule in droplet, IVTT components were added by an electro-coalescence mechanism, while the assay mixture was pico-injected into the droplets containing the expressed protein.

Table 5. Comparison of microfluidic operations for controlled addition of volumes to droplets.

Technique	Device fabrication	Device operation	Throughput max. reported/ actual work	Added volume to droplet in this work	Sensitivity to droplet content
Passive surfactant-mediated fusion [117]	Simple PDMS-glass device	Requires reinjection of stabil. oil	1.55 kHz /1.8 kHz	20 pL were added to 4 pL-sized drops	Moderate. Adjustment of surfactant concentration
Droplet electro-coalescence[a] [214]	Requires ITO glass and electrodes	Requires electric field	1.7 kHz /1.5 kHz	25 pL were added to 2 pL-sized drops	Low. Adjustment of electric field
Pico-injection into passing droplets[b] [164]	Requires ITO glass and electrodes	Requires electric field	10 kHz /1 kHz	4 pL were added to 27 pL-sized drops	Low. Adjustment of electric field

In this work several different microfluidic techniques were used for the same purpose: to add reactants volumes to pre-existing droplets. The fact that passive surfactant-mediated fusion, droplet electro-coalescence and pico-injection into passing

[a] Electro-coalescence of reinjected droplets with droplets produced on the same chip.
[b] Here we are regarding initial pico-injection mechanism and not "electrode-free" version.

droplets, were all employed in the same work allows their practical comparison, moreover when only few examples of their integration in microfluidic platforms is recorded in literature. All three were operated at 1-2 kHz frequencies and the choise of one technique as opposed to another stemmed from device fabrication/operation, sensitivity to droplet content and the ratio of volumes that need to added to pre-formed droplets. The fabrication and operation of electro-coalescence and pico-injection devices is complicated by electrode integration and the need for electric field generators, nevertheless they are less sensitive to the droplets content, while the components of droplets involved in passive fusion can significantly influence fusion efficiency. This effect was monitored during the development of protease screening platform, where the presence of detergent prevented fusion under normal conditions. The advantage of the pico-injection technique over the other two are from the fact that a large range of volumes that can be added to the droplets using the same device. Adjusting the pressure of injected solution, reinjected droplet frequency and the field it is possible to add to a 20 pL droplet volumes ranging from 0.5 pL to 20 pL. This property of pico-injection was successfully exploited in SGAP evolution project to add only 4 pL of assay mixture to the 27 pL droplets with expressed protein. In the future, the integration of these fusion modules on the same chip with delay-line (<30 min) and sorter is very promising for screening of variants based on initial rates. In conclusion, all fusion techniques are used efficient if integrated correctly, complementing each other, rather than competing. he robustness and capacity of droplet-based microfluidic platform depends on the right choice of the fusion system.

To compare the microfluidic platforms for directed evolution developed in this work with previously published high-throughput screening assays it is worth regarding coefficient of variation (CV) of the signal from a droplet population of a single enzyme variant [193]. The signal CV represents the standard deviation/mean fluorescence. A low variation coefficient (close to the ideal limited by microfluidic set-up) corresponds to high confidence in each data point and will allow sensitive detection of small activity differences. However, it is important that fluorescent measurement corresponds to initial rates. Measuring after the initial phase where most of all substrate is converted would likely result in a low CV, making impossible to discriminate different variants.

Table 6. Signal variation coefficients comparison.

System	Coefficient of Variation (CV)
In vitro expression in double emulsion droplets / FACS [79]	1.0
Cytoplasmic expression / FACS [64]	0.9
Bacterial-displayed protein / FACS [237]	0.8
PAS Cytoplasmatic expression/FADS [193]	0.3
Retro-aldolase project Cytoplasmatic expression/FADS	0.35
Protease screening project Secretory expression in droplets/fusion with assay components/FADS	0.12
***In vitro* evolution of SGAP project** HRCA from single gene in drops/fusion with IVTT/ fusion with assay components/FADS	*0.2*[a]
Droplets with dye The minimal CV allowed by microfluidic set-up	0.05

From the Table 6 the advantage of FADS ("Sorting") over the FACS-based screening systems in terms of resolution is obvious. The comparison among the microfluidic platforms based on FADS, suggests that the systems where the assays were based on the activity from single cells had higher variance. In the retro-aldolase project as well as in the example of *PAS* evolution [193], single cells expressing the same variant are lysed just after compartmentalization. Thus, the dispersion in recorded signal comes from different phenotypic behaviour of cells within the same population (different levels of expression of the same enzyme). In contrast, in protease screening project, the CV is 3-times lower. Here, single cell encapsulation followed by bacterial growth in droplets and followed by fusion with assay components, allowed the monitoring of the average activity of cell population bearing the same enzyme variant. Protease screening system

[a] The reaction might no longer be in the linear phase of the substrate conversion and actual CV can be lower than the real one.

with a CV of 0.12, gives significantly better signal quality than current state-of-the-art technologies. Accordingly, the addition of a fusion step can increase the variants resolution of an *in vivo* expressed library.

Screening and selection based on the *in vitro* expression can open new horizons in fundamental directed protein evolution studies. Such systems should allow precise screening of desired property, since they are much simpler, genetically and biochemically, than *in vivo* systems. They will allow rapid iteration of cycles of mutation and selection where no time consuming transformation steps are required. Thus, making possible the investigate protein evolution pathways independently of the cell influence. Furthermore, *in vitro* systems permit expression of toxic proteins and incorporation of non-natural amino acids in protein sequence. The completely *in vitro* droplet-based microfluidic platform developed for SGAP directed evolution possesses all the aforementioned advantages. The fact that, an *in vitro* library selection in microfluidic droplets is still not on record, will make further directed evolution experiments ground-breaking.

Supplementary Information

Sequence of SGAP vector

>LMB2-10E

GA TGG CGC CCA ACA GTC CCC CGG CCA CGG GGC CTG CCA CCA T < 42

AC CCA CGC CGA AAC AAG CGC TCA TGA GCC CGA AGT GGC GAG C < 84

>LMB2-11E

CC GAT CTT CCC CAT CGG TGA TGT CGG CGA TAT AGG CGC CAG C < 126

AA CCG CAC CTG TGG CGC CGG TGA TGC CGG CCA CGA TGC GTC C < 168

GG CGT AGA GGA TCG AGA TCT CGA TCC CGC GAA ATT AAT ACG A < 210

>T7\promoter

CT CAC TAT AGG GAG ACC ACA ACG GTT TCC CTC TAG AAA TAA T < 252

>Enhancer >T7\RBS sgap_wt

TT TGT TTA ACT TTA AGA AGG AGA TAT ACC ATG GCG CCG GAC A < 294

M A P D I

TC CCG CTG GCG AAC GTC AAG GCC CAC CTC ACG CAG CTC TCG A < 336

P L A N V K A H L T Q L S T

CG ATC GCC GCG AAC AAC GGC GGC AAC CGC GCC CAC GGC CGC C < 378

I A A N N G G N R A H G R P

CC GGC TAC AAG GCG TCC GTC GAC TAC GTG AAG GCC AAG CTC G < 420

G Y K A S V D Y V K A K L D

AC GCG GCC GGA TAC ACC ACC ACG CTC CAG CAG TTC ACC TCG G < 462

A A G Y T T T L Q Q F T S G

GC GGG GCC ACC GGC TAC AAC CTG ATA GCC GAC TGG CCC GGC G < 504

G A T G Y N L I A D W P G G

GC GAC CCC AAC AAG GTC CTG ATG GCC GGG GCC CAC CTC GAC T < 546

D P N K V L M A G A H L D S

CG GTC TCC TCC GGC GCC GGG ATC AAC GAC AAC GGC TCC GGC T < 588

V S S G A G I N D N G S G S

CG GCC GCC GTG CTG GAG ACC GCG CTC GCC GTC TCC CGC GCC G < 630

A A V L E T A L A V S R A G

GG TAC CAG CCC GAC AAG CAC CTG CGG TTC GCC TGG TGG GGC G < 672
Y Q P D K H L R F A W W G A

CG GAG GAG CTG GGC CTG ATC GGC TCG AAG TTC TAC GTT AAC A < 714
E E L G L I G S K F Y V N N

AC CTG CCG TCC GCC GAC CGC TCC AAG CTC GCC GGA TAT CTC A < 756
L P S A D R S K L A G Y L N

AC TTC GAC ATG ATC GGC TCG CCC AAC CCC GGT TAC TTC GTC T < 798
F D M I G S P N P G Y F V Y

AC GAC GAC GAC CCG GTC ATC GAG AAG ACC TTC AAG GAC TAC T < 840
D D D P V I E K T F K D Y F

TC GCC GGC CTG AAC GTC CCG ACC GAG ATC GAG ACC GAG GGC G < 882
A G L N V P T E I E T E G D

AC GGC CGC TCC GAC CAC GCC CCG TTC AAG AAC GTC GGC GTC C < 924
G R S D H A P F K N V G V P

CC GTC GGC GGA CTC TTC ACC GGC GCC GGC TAC ACC AAG TCC G < 966
V G G L F T G A G Y T K S A

CC GCC CAG GCG CAG AAG TGG GGC GGG ACG GCC GGG CAA GCC T < 1008
A Q A Q K W G G T A G Q A F

TC GAC CGC TGC TAC CAC TCC TCG TGC GAC AGC CTG AGC AAC A < 1050
D R C Y H S S C D S L S N I

TC AAC GAC ACC GCC CTC GAC CGC AAC AGC GAT GCA GCA GCA C < 1092
N D T A L D R N S D A A A H

AT GCC ATT TGG ACC CTG AGC AGC GGC ACC GGC GAA CCA CCA A < 1134
A I W T L S S G T G E P P T

CC GCG GCC GCA TGT TGC CCG GGC TGC TGT TAA TGA CTC GAG C < 1176

GA GCT CTG CAG CCC GGG ATC CGG TAA GAT CCG GCT GCT AAC A < 1218

AA GCC CGA AAG GAA GCT GAG TTG GCT GCT GCC ACC GCT GAG C < 1260

>T7\terminator

AA TAA CTA GCA TAA CCC CTT GGG GCC TCT AAA CGG GTC TTG A < 1302

GG GGT TTT TTG CTG AAA GGA GGA ACT ATA TCC GGA TAT CCA C < 1344

>PIVB-7

AG GAC GGG TGT GGT CGC CAT GAT CGC GTA GTC GAT AGT GGC T < 1386

>PIVB-4

CC AAG TAG CGA AGC GAG CAG GAC TGG GCG GCG GCC AAA < 1424

Features :

T7\terminator : [1273 : 1310]

T7\RBS : [269 : 274]

Enhancer_1 : [258 : 266]

T7\promoter : [203 : 219]

sgap_wt : [285 : 1136]

Primers

LMB2-11E : [81 : 101]

LMB2-10E : [1 : 18]

PIVB-7 : [1340 : 1373]

PIVB-4 : [1408 : 1424]

Bibliography

1. Agresti, J.J., et al., *Ultrahigh-throughput screening in drop-based microfluidics for directed evolution.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(9): p. 4004-9.
2. Rothlisberger, D., et al., *Kemp elimination catalysts by computational enzyme design.* Nature, 2008. **453**(7192): p. 190-195.
3. Mazutis, L., J.-C. Baret, and A.D. Griffiths, *A fast and efficient microfluidic system for highly selective one-to-one droplet fusion.* Lab on a Chip, 2009. **9**(18): p. 2665-2672.
4. Shimizu, Y., et al., *Cell-free translation reconstituted with purified components.* Nat Biotech, 2001. **19**(8): p. 751-755.
5. Rasor, J.P. and E. Voss, *Enzyme-catalyzed processes in pharmaceutical industry.* Applied Catalysis A: General, 2001. **221**(1–2): p. 145-158.
6. Kirk, O., T.V. Borchert, and C.C. Fuglsang, *Industrial enzyme applications.* Current opinion in biotechnology, 2002. **13**(4): p. 345-351.
7. Yeoman, C.J., et al., *Chapter 1 - Thermostable Enzymes as Biocatalysts in the Biofuel Industry*, in *Advances in Applied Microbiology*, I.L. Allen, S. Sima, and M.G. Geoffrey, Editors. 2010, Academic Press. p. 1-55.
8. Demarche, P., et al., *Harnessing the power of enzymes for environmental stewardship.* Biotechnology Advances, 2012. **30**(5): p. 933-953.
9. Ispas, C.R., G. Crivat, and S. Andreescu, *Review: Recent Developments in Enzyme-Based Biosensors for Biomedical Analysis.* Analytical Letters, 2012. **45**(2-3): p. 168-186.
10. Bornscheuer, U.T., *Rational Design vs Directed Evolution*, in *Summer School on Protein Engineering*. 2010: Greifswald, Germany.
11. Baker, D., *An exciting but challenging road ahead for computational enzyme design.* Protein Science, 2010. **19**(10): p. 1817-1819.
12. Sterner, R., R. Merkl, and F.M. Raushel, *Computational Design of Enzymes.* Chemistry & Biology, 2008. **15**(5): p. 421-423.
13. Barrozo, A., et al., *Computational Protein Engineering: Bridging the Gap between Rational Design and Laboratory Evolution.* International Journal of Molecular Sciences, 2012. **13**(10): p. 12428-12460.
14. Rothlisberger, D., et al., *Kemp elimination catalysts by computational enzyme design.* Nature, 2008. **453**(7192): p. 190-5.
15. Jiang, L., et al., *De novo computational design of retro-aldol enzymes.* Science, 2008. **319**(5868): p. 1387-91.
16. Siegel, J.B., et al., *Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction.* Science, 2010. **329**(5989): p. 309-13.
17. Kamerlin, S.C.L. and A. Warshel, *The empirical valence bond model: theory and applications.* Wiley Interdisciplinary Reviews-Computational Molecular Science, 2011. **1**(1): p. 30-45.

18. Lutz, S., *Beyond directed evolution--semi-rational protein engineering and design.* Current opinion in biotechnology, 2010. **21**(6): p. 734-43.
19. Cole, M.F. and E.A. Gaucher, *Exploiting models of molecular evolution to efficiently direct protein engineering.* Journal of molecular evolution, 2011. **72**(2): p. 193-203.
20. Fox, R.J., et al., *Improving catalytic function by ProSAR-driven enzyme evolution.* Nat Biotechnol, 2007. **25**(3): p. 338-44.
21. Bornscheuer, U.T., et al., *Engineering the third wave of biocatalysis.* Nature, 2012. **485**(7397): p. 185-94.
22. Turner, N.J., *Directed evolution drives the next generation of biocatalysts.* Nat Chem Biol, 2009. **5**(8): p. 567-73.
23. Reetz, M.T., D. Kahakeaw, and R. Lohmer, *Addressing the numbers problem in directed evolution.* Chembiochem, 2008. **9**(11): p. 1797-804.
24. Shivange, A.V., et al., *Advances in generating functional diversity for directed protein evolution.* Current Opinion in Chemical Biology, 2009. **13**(1): p. 19-25.
25. Bornscheuer, U.T., et al., *Optimizing lipases and related enzymes for efficient application.* Trends Biotechnol, 2002. **20**(10): p. 433-7.
26. Yoshikuni, Y., T.E. Ferrin, and J.D. Keasling, *Designed divergent evolution of enzyme function.* Nature, 2006. **440**(7087): p. 1078-82.
27. Shivange, A.V., et al., *Advances in generating functional diversity for directed protein evolution.* Curr Opin Chem Biol, 2009. **13**(1): p. 19-25.
28. Patrick, W.M. and A.E. Firth, *Strategies and computational tools for improving randomized protein libraries.* Biomol Eng, 2005. **22**(4): p. 105-12.
29. Labrou, N.E., *Random mutagenesis methods for in vitro directed enzyme evolution.* Curr Protein Pept Sci, 2010. **11**(1): p. 91-100.
30. Morley, K.L. and R.J. Kazlauskas, *Improving enzyme properties: when are closer mutations better?* Trends Biotechnol, 2005. **23**(5): p. 231-7.
31. Kadonaga, J.T. and J.R. Knowles, *A simple and efficient method for chemical mutagenesis of DNA.* Nucleic Acids Res, 1985. **13**(5): p. 1733-45.
32. Greener, A., M. Callahan, and B. Jerpseth, *An efficient random mutagenesis technique using an E. coli mutator strain.* Methods Mol Biol, 1996. **57**: p. 375-85.
33. McCullum, E.O., et al., *Random mutagenesis by error-prone PCR.* Methods Mol Biol, 2010. **634**: p. 103-9.
34. Fujii, R., M. Kitaoka, and K. Hayashi, *One-step random mutagenesis by error-prone rolling circle amplification.* Nucleic Acids Res, 2004. **32**(19): p. e145.
35. Wong, T.S., et al., *A statistical analysis of random mutagenesis methods used for directed protein evolution.* J Mol Biol, 2006. **355**(4): p. 858-71.
36. Tomandl, D., A. Schober, and A. Schwienhorst, *Optimizing doped libraries by using genetic algorithms.* Journal of Computer-Aided Molecular Design, 1997. **11**(1): p. 29-38.
37. Hughes, M.D., et al., *Removing the Redundancy From Randomised Gene Libraries.* Journal of Molecular Biology, 2003. **331**(5): p. 973-979.
38. Stemmer, W.P., *DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution.* Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(22): p. 10747-51.
39. Crameri, A., et al., *Improved green fluorescent protein by molecular evolution using DNA shuffling.* Nat Biotechnol, 1996. **14**(3): p. 315-9.
40. Tang, L., et al., *Construction of "small-intelligent" focused mutagenesis libraries using well-designed combinatorial degenerate primers.* Biotechniques, 2012. **52**(3): p. 149-58.

41. Siloto, R.M.P. and R.J. Weselake, *Site saturation mutagenesis: Methods and applications in protein engineering.* Biocatalysis and Agricultural Biotechnology, 2012. **1**(3): p. 181-189.
42. Jochens, H. and U.T. Bornscheuer, *Natural diversity to guide focused directed evolution.* Chembiochem, 2010. **11**(13): p. 1861-6.
43. Hoffmann, G., et al., *Changing the substrate specificity of P450cam towards diphenylmethane by semi-rational enzyme engineering.* Protein Eng Des Sel, 2011. **24**(5): p. 439-46.
44. Reetz, M.T., et al., *Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: rigorous comparison with traditional methods.* J Am Chem Soc, 2010. **132**(26): p. 9144-52.
45. Peisajovich, S.G. and D.S. Tawfik, *Protein engineers turned evolutionists.* Nat Methods, 2007. **4**(12): p. 991-4.
46. Weinreich, D.M., R.A. Watson, and L. Chao, *Perspective: Sign epistasis and genetic constraint on evolutionary trajectories.* Evolution, 2005. **59**(6): p. 1165-74.
47. Camps, M., et al., *Genetic constraints on protein evolution.* Crit Rev Biochem Mol Biol, 2007. **42**(5): p. 313-26.
48. Poelwijk, F.J., et al., *Empirical fitness landscapes reveal accessible evolutionary paths.* Nature, 2007. **445**(7126): p. 383-6.
49. Breen, M.S., et al., *Epistasis as the primary factor in molecular evolution.* Nature, 2012. **advance online publication**.
50. Bloom, J.D., et al., *Protein stability promotes evolvability.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(15): p. 5869-74.
51. Coco, W.M., et al., *DNA shuffling method for generating highly recombined genes and evolved enzymes.* Nat Biotechnol, 2001. **19**(4): p. 354-9.
52. Tracewell, C.A. and F.H. Arnold, *Directed enzyme evolution: climbing fitness peaks one amino acid at a time.* Curr Opin Chem Biol, 2009. **13**(1): p. 3-9.
53. Bloom, J.D., et al., *Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution.* Biol Direct, 2007. **2**: p. 17.
54. Bershtein, S., K. Goldin, and D.S. Tawfik, *Intense neutral drifts yield robust and evolvable consensus proteins.* J Mol Biol, 2008. **379**(5): p. 1029-44.
55. Gupta, R.D. and D.S. Tawfik, *Directed enzyme evolution via small and effective neutral drift libraries.* Nat Methods, 2008. **5**(11): p. 939-42.
56. Aharoni, A., et al., *The 'evolvability' of promiscuous protein functions.* Nat Genet, 2005. **37**(1): p. 73-6.
57. Agresti, J.J., et al., *Ultrahigh-throughput screening in drop-based microfluidics for directed evolution.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(9): p. 4004-9.
58. Griffiths, A.D. and D.S. Tawfik, *Man-made enzymes--from design to in vitro compartmentalisation.* Current opinion in biotechnology, 2000. **11**(4): p. 338-53.
59. Leemhuis, H., et al., *New genotype-phenotype linkages for directed evolution of functional proteins.* Curr Opin Struct Biol, 2005. **15**(4): p. 472-8.
60. Salverda, M.L.M., J.A.G.M. De Visser, and M. Barlow, *Natural evolution of TEM-1 β-lactamase: experimental reconstruction and clinical relevance.* FEMS Microbiology Reviews, 2010. **34**(6): p. 1015-1036.
61. Yoo, T.H., et al., *Directed evolution of highly selective proteases by using a novel FACS-based screen that capitalizes on the p53 regulator MDM2.* Chembiochem, 2012. **13**(5): p. 649-53.

62. Bershtein, S. and D.S. Tawfik, *Advances in laboratory evolution of enzymes.* Curr Opin Chem Biol, 2008. **12**(2): p. 151-8.
63. Yang, G. and S.G. Withers, *Ultrahigh-throughput FACS-based screening for directed enzyme evolution.* Chembiochem, 2009. **10**(17): p. 2704-15.
64. Aharoni, A., et al., *High-throughput screening methodology for the directed evolution of glycosyltransferases.* Nat Methods, 2006. **3**(8): p. 609-14.
65. Antipov, E., et al., *Highly L and D enantioselective variants of horseradish peroxidase discovered by an ultrahigh-throughput selection method.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(46): p. 17694-9.
66. Georgiou, G., et al., *Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines.* Nat Biotechnol, 1997. **15**(1): p. 29-34.
67. Aharoni, A., et al., *High-throughput screening of enzyme libraries: Thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments.* Chemistry & Biology, 2005. **12**(12): p. 1281-1289.
68. Jackel, C., P. Kast, and D. Hilvert, *Protein design by directed evolution.* Annu Rev Biophys, 2008. **37**: p. 153-73.
69. Aharoni, A., A.D. Griffiths, and D.S. Tawfik, *High-throughput screens and selections of enzyme-encoding genes.* Curr Opin Chem Biol, 2005. **9**(2): p. 210-6.
70. Tawfik, D.S. and A.D. Griffiths, *Man-made cell-like compartments for molecular evolution.* Nat Biotechnol, 1998. **16**(7): p. 652-6.
71. Cohen, H.M., D.S. Tawfik, and A.D. Griffiths, *Altering the sequence specificity of HaeIII methyltransferase by directed evolution using in vitro compartmentalization.* Protein Eng Des Sel, 2004. **17**(1): p. 3-11.
72. Ghadessy, F.J., J.L. Ong, and P. Holliger, *Directed evolution of polymerase function by compartmentalized self-replication.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(8): p. 4552-7.
73. Ghadessy, F.J., et al., *Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution.* Nat Biotechnol, 2004. **22**(6): p. 755-9.
74. Doi, N., et al., *In vitro selection of restriction endonucleases by in vitro compartmentalization.* Nucleic Acids Res, 2004. **32**(12): p. e95.
75. Dower, W.J. and L.C. Mattheakis, *In vitro selection as a powerful tool for the applied evolution of proteins and peptides.* Curr Opin Chem Biol, 2002. **6**(3): p. 390-8.
76. Kelly, B.T., et al., *Miniaturizing chemistry and biology in microdroplets.* Chem Commun (Camb), 2007(18): p. 1773-88.
77. Yonezawa, M., et al., *DNA display of biologically active proteins for in vitro protein selection.* J Biochem, 2004. **135**(3): p. 285-8.
78. Griffiths, A.D. and D.S. Tawfik, *Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization.* EMBO J, 2003. **22**(1): p. 24-35.
79. Mastrobattista, E., et al., *High-throughput screening of enzyme libraries: in vitro evolution of a beta-galactosidase by fluorescence-activated sorting of double emulsions.* Chem Biol, 2005. **12**(12): p. 1291-300.
80. Griffiths, A.D. and D.S. Tawfik, *Miniaturising the laboratory in emulsion droplets.* Trends Biotechnol, 2006. **24**(9): p. 395-402.
81. Agresti, J.J., et al., *Selection of ribozymes that catalyse multiple-turnover Diels-Alder cycloadditions by using in vitro compartmentalization.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(45): p. 16170-5.

82. Bernath, K., S. Magdassi, and D.S. Tawfik, *Directed evolution of protein inhibitors of DNA-nucleases by in vitro compartmentalization (IVC) and nano-droplet delivery.* J Mol Biol, 2005. **345**(5): p. 1015-26.
83. Holtze, C., et al., *Biocompatible surfactants for water-in-fluorocarbon emulsions.* Lab Chip, 2008. **8**(10): p. 1632-9.
84. Hudlicky, M., *Chemistry of organic fluorine compounds : a laboratory manual with comprehensive literature coverage*. 2nd (rev. ed. 1976, New York ; London: Ellis Horwood. xiv, 903 p, 4 p of plates.
85. Studer, A., et al., *Fluorous synthesis: a fluorous-phase strategy for improving separation efficiency in organic synthesis.* Science, 1997. **275**(5301): p. 823-6.
86. Lee, J.N., C. Park, and G.M. Whitesides, *Solvent compatibility of poly(dimethylsiloxane)-based microfluidic devices.* Anal Chem, 2003. **75**(23): p. 6544-54.
87. Song, H., D.L. Chen, and R.F. Ismagilov, *Reactions in droplets in microfluidic channels.* Angew Chem Int Ed Engl, 2006. **45**(44): p. 7336-56.
88. Lowe, K.C., M.R. Davey, and J.B. Power, *Perfluorochemicals: their applications and benefits to cell culture.* Trends Biotechnol, 1998. **16**(6): p. 272-7.
89. Clausell-Tormos, J., et al., *Droplet-based microfluidic platforms for the encapsulation and screening of Mammalian cells and multicellular organisms.* Chem Biol, 2008. **15**(5): p. 427-37.
90. Joanicot, M. and A. Ajdari, *Applied physics. Droplet control for microfluidics.* Science, 2005. **309**(5736): p. 887-8.
91. Gunther, A. and K.F. Jensen, *Multiphase microfluidics: from flow characteristics to chemical and materials synthesis.* Lab Chip, 2006. **6**(12): p. 1487-503.
92. Garstecki, P., et al., *Formation of droplets and bubbles in a microfluidic T-junction-scaling and mechanism of break-up.* Lab Chip, 2006. **6**(3): p. 437-46.
93. Garstecki, P., H.A. Stone, and G.M. Whitesides, *Mechanism for flow-rate controlled breakup in confined geometries: a route to monodisperse emulsions.* Phys Rev Lett, 2005. **94**(16): p. 164501.
94. Thorsen, T., et al., *Dynamic pattern formation in a vesicle-generating microfluidic device.* Phys Rev Lett, 2001. **86**(18): p. 4163-6.
95. Zheng, B., J.D. Tice, and R.F. Ismagilov, *Formation of droplets of alternating composition in microfluidic channels and applications to indexing of concentrations in droplet-based assays.* Anal Chem, 2004. **76**(17): p. 4977-82.
96. Tice, J.D., A.D. Lyon, and R.F. Ismagilov, *Effects of viscosity on droplet formation and mixing in microfluidic channels.* Anal Chim Acta, 2004. **507**(1): p. 73-77.
97. Romero, P.A. and A.R. Abate, *Flow focusing geometry generates droplets through a plug and squeeze mechanism.* Lab Chip, 2012.
98. van der Graaf, S., et al., *Lattice Boltzmann simulations of droplet formation in a T-shaped microchannel.* Langmuir, 2006. **22**(9): p. 4144-52.
99. Christopher, G.F., et al., *Experimental observations of the squeezing-to-dripping transition in T-shaped microfluidic junctions.* Phys Rev E Stat Nonlin Soft Matter Phys, 2008. **78**(3 Pt 2): p. 036317.
100. Lee, W., L.M. Walker, and S.L. Anna, *Role of geometry and fluid properties in droplet and thread formation processes in planar flow focusing.* Physics of Fluids, 2009. **21**(3).
101. De Menech, M., et al., *Transition from squeezing to dripping in a microfluidic T-shaped junction.* Journal of Fluid Mechanics, 2008. **595**: p. 141-161.

102. Gu, H., M.H. Duits, and F. Mugele, *Droplets formation and merging in two-phase flow microfluidics.* Int J Mol Sci, 2011. **12**(4): p. 2572-97.
103. Guillot, P. and A. Colin, *Stability of parallel flows in a microchannel after a T junction.* Phys Rev E Stat Nonlin Soft Matter Phys, 2005. **72**(6 Pt 2): p. 066301.
104. Utada, A.S., et al., *Monodisperse double emulsions generated from a microcapillary device.* Science, 2005. **308**(5721): p. 537-41.
105. Chabert, M. and J.L. Viovy, *Microfluidic high-throughput encapsulation and hydrodynamic self-sorting of single cells.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(9): p. 3191-6.
106. Abate, A.R., et al., *Impact of inlet channel geometry on microfluidic drop formation.* Phys Rev E Stat Nonlin Soft Matter Phys, 2009. **80**(2 Pt 2): p. 026310.
107. Okushima, S., et al., *Controlled production of monodisperse double emulsions by two-step droplet breakup in microfluidic devices.* Langmuir, 2004. **20**(23): p. 9905-8.
108. Xu, Q., et al., *Preparation of monodisperse biodegradable polymer microparticles using a microfluidic flow-focusing device for controlled drug delivery.* Small, 2009. **5**(13): p. 1575-81.
109. Abate, A.R. and D.A. Weitz, *High-order multiple emulsions formed in poly(dimethylsiloxane) microfluidics.* Small, 2009. **5**(18): p. 2030-2.
110. Adams, L.L.A., et al., *Single step emulsification for the generation of multi-component double emulsions.* Soft Matter, 2012. **8**(41): p. 10719-10724.
111. Kiss, M.M., et al., *High-Throughput Quantitative Polymerase Chain Reaction in Picoliter Droplets.* Analytical Chemistry, 2008. **80**(23): p. 8975-8981.
112. Srisa-Art, M., et al., *Monitoring of real-time streptavidin-biotin binding kinetics using droplet microfluidics.* Analytical Chemistry, 2008. **80**(18): p. 7063-7067.
113. Chiu, D.T. and R.M. Lorenz, *Chemistry and Biology in Femtoliter and Picoliter Volume Droplets.* Accounts of Chemical Research, 2009. **42**(5): p. 649-658.
114. Baret, J.C., *Surfactants in droplet-based microfluidics.* Lab Chip, 2012. **12**(3): p. 422-33.
115. Bibette, J., et al., *Stability criteria for emulsions.* Phys Rev Lett, 1992. **69**(16): p. 2439-2442.
116. Baret, J.C., et al., *Kinetic aspects of emulsion stabilization by surfactants: a microfluidic analysis.* Langmuir, 2009. **25**(11): p. 6088-93.
117. Mazutis, L., J.C. Baret, and A.D. Griffiths, *A fast and efficient microfluidic system for highly selective one-to-one droplet fusion.* Lab Chip, 2009. **9**(18): p. 2665-72.
118. Roach, L.S., H. Song, and R.F. Ismagilov, *Controlling nonspecific protein adsorption in a plug-based microfluidic system by controlling interfacial chemistry using fluorous-phase surfactants.* Anal Chem, 2005. **77**(3): p. 785-96.
119. Courrier, H.M., et al., *Evaluation of cytotoxicity of new semi-fluorinated amphiphiles derived from dimorpholinophosphate.* Biomaterials, 2003. **24**(4): p. 689-96.
120. Li, X., et al., *Hydrophobic tail length, degree of fluorination and headgroup stereochemistry are determinants of the biocompatibility of (fluorinated) carbohydrate surfactants.* Colloids Surf B Biointerfaces, 2009. **73**(1): p. 65-74.
121. Courtois, F., et al., *Controlling the retention of small molecules in emulsion microdroplets for use in cell-based assays.* Anal Chem, 2009. **81**(8): p. 3008-16.
122. Woronoff, G., et al., *New generation of amino coumarin methyl sulfonate-based fluorogenic substrates for amidase assays in droplet-based microfluidic applications.* Anal Chem, 2011. **83**(8): p. 2852-7.
123. Huebner, A., et al., *Microdroplets: a sea of applications?* Lab Chip, 2008. **8**(8): p. 1244-54.

124. Stroock, A.D., et al., *Patterning electro-osmotic flow with patterned surface charge.* Phys Rev Lett, 2000. **84**(15): p. 3314-7.
125. Handique, K., et al., *On-chip thermopneumatic pressure for discrete drop pumping.* Anal Chem, 2001. **73**(8): p. 1831-8.
126. Song, H., et al., *Experimental test of scaling of mixing by chaotic advection in droplets moving through microfluidic channels.* Appl Phys Lett, 2003. **83**(12): p. 4664-4666.
127. Sarrazin, F., et al., *Mixing characterization inside microdroplets engineered on a microcoalescer.* Chemical Engineering Science, 2007. **62**(4): p. 1042-1048.
128. Zeng, Y., et al., *Quantitative imaging of mixing dynamics in microfluidic droplets using two-photon fluorescence lifetime imaging.* Optics Letters, 2011. **36**(12): p. 2236-2238.
129. Song, H., et al., *Experimental test of scaling of mixing by chaotic advection in droplets moving through microfluidic channels.* Applied Physics Letters, 2003. **83**(22): p. 4664-4666.
130. Bringer, M.R., et al., *Microfluidic systems for chemical kinetics that rely on chaotic mixing in droplets.* Philos Transact A Math Phys Eng Sci, 2004. **362**(1818): p. 1087-104.
131. Jiang, L.G., et al., *Visualizing millisecond chaotic mixing dynamics in microdroplets: A direct comparison of experiment and simulation.* Biomicrofluidics, 2012. **6**(1).
132. Liau, A., et al., *Mixing crowded biological solutions in milliseconds.* Analytical Chemistry, 2005. **77**(23): p. 7618-7625.
133. Pompano, R.R., H.W. Li, and R.F. Ismagilov, *Rate of mixing controls rate and outcome of autocatalytic processes: Theory and microfluidic experiments with chemical reactions and blood coagulation.* Biophysical Journal, 2008. **95**(3): p. 1531-1543.
134. Song, H. and R.F. Ismagilov, *Millisecond kinetics on a microfluidic chip using nanoliters of reagents.* Journal of the American Chemical Society, 2003. **125**(47): p. 14613-14619.
135. Clausell-Tormos, J., A.D. Griffiths, and C.A. Merten, *An automated two-phase microfluidic system for kinetic analyses and the screening of compound libraries.* Lab on a Chip, 2010. **10**(10): p. 1302-1307.
136. Abate, A.R. and D.A. Weitz, *Faster multiple emulsification with drop splitting.* Lab Chip, 2011. **11**(11): p. 1911-5.
137. Link, D.R., et al., *Geometrically mediated breakup of drops in microfluidic devices.* Physical Review Letters, 2004. **92**(5).
138. Ménétrier-Deremble, L. and P. Tabeling, *Droplet breakup in microfluidic junctions of arbitrary angles.* Physical Review E, 2006. **74**(3): p. 035303.
139. Link, D.R., et al., *Electric control of droplets in microfluidic devices.* Angewandte Chemie-International Edition, 2006. **45**(16): p. 2556-2560.
140. Fan, S.K., T.H. Hsieh, and D.Y. Lin, *General digital microfluidic platform manipulating dielectric and conductive droplets by dielectrophoresis and electrowetting.* Lab on a Chip, 2009. **9**(9): p. 1236-1242.
141. Darhuber, A.A., J.P. Valentino, and S.M. Troian, *Planar digital nanoliter dispensing system based on thermocapillary actuation.* Lab on a Chip, 2010. **10**(8): p. 1061-1071.
142. Baroud, C.N., et al., *Thermocapillary valve for droplet production and sorting.* Physical Review E, 2007. **75**(4).
143. Bibette, J., F.L. Calderon, and P. Poulin, *Emulsions: basic principles.* Reports on Progress in Physics, 1999. **62**(6): p. 969-1033.

144. Hung, L.H., et al., *Alternating droplet generation and controlled dynamic droplet fusion in microfluidic device for CdS nanoparticle synthesis.* Lab on a Chip, 2006. **6**(2): p. 174-178.
145. Tan, Y.C., et al., *Design of microfluidic channel geometries for the control of droplet volume, chemical concentration, and sorting.* Lab on a Chip, 2004. **4**(4): p. 292-298.
146. Um, E. and J.K. Park, *A microfluidic abacus channel for controlling the addition of droplets.* Lab on a Chip, 2009. **9**(2): p. 207-212.
147. Sivasamy, J., et al., *Reliable addition of reagents into microfluidic droplets.* Microfluidics and Nanofluidics, 2010. **8**(3): p. 409-416.
148. Niu, X., et al., *Pillar-induced droplet merging in microfluidic circuits.* Lab on a Chip, 2008. **8**(11): p. 1837-1841.
149. Lin, B.C. and Y.C. Su, *On-demand liquid-in-liquid droplet metering and fusion utilizing pneumatically actuated membrane valves.* Journal of Micromechanics and Microengineering, 2008. **18**(11).
150. Baroud, C.N., M.R. de Saint Vincent, and J.P. Delville, *An optical toolbox for total control of droplet microfluidics.* Lab on a Chip, 2007. **7**(8): p. 1029-1033.
151. Fidalgo, L.M., C. Abell, and W.T.S. Huck, *Surface-induced droplet fusion in microfluidic devices.* Lab on a Chip, 2007. **7**(8): p. 984-986.
152. Liu, Y. and R.F. Ismagilov, *Dynamics of Coalescence of Plugs with a Hydrophilic Wetting Layer Induced by Flow in a Microfluidic Chemistrode.* Langmuir, 2009. **25**(5): p. 2854-2859.
153. Bremond, N., A.R. Thiam, and J. Bibette, *Decompressing emulsion droplets favors coalescence.* Physical Review Letters, 2008. **100**(2).
154. Mazutis, L. and A.D. Griffiths, *Selective droplet coalescence using microfluidic systems.* Lab on a Chip, 2012. **12**(10): p. 1800-1806.
155. Wang, W., C. Yang, and C.M. Li, *Efficient On-Demand Compound Droplet Formation: From Microfluidics to Microdroplets as Miniaturized Laboratories.* Small, 2009. **5**(10): p. 1149-1152.
156. Tan, W.H. and S. Takeuchi, *Timing controllable electrofusion device for aqueous droplet-based microreactors.* Lab on a Chip, 2006. **6**(6): p. 757-763.
157. Zagnoni, M. and J.M. Cooper, *On-chip electrocoalescence of microdroplets as a function of voltage, frequency and droplet size.* Lab on a Chip, 2009. **9**(18): p. 2652-2658.
158. Priest, C., S. Herminghaus, and R. Seemann, *Controlled electrocoalescence in microfluidics: Targeting a single lamella.* Applied Physics Letters, 2006. **89**(13).
159. Wang, W., C. Yang, and C.M. Li, *On-demand microfluidic droplet trapping and fusion for on-chip static droplet assays.* Lab on a Chip, 2009. **9**(11): p. 1504-1506.
160. Schwartz, J.A., J.V. Vykoukal, and P.R.C. Gascoyne, *Droplet-based chemistry on a programmable micro-chip.* Lab on a Chip, 2004. **4**(1): p. 11-17.
161. Ahn, K., et al., *Electrocoalescence of drops synchronized by size-dependent flow in microfluidic channels.* Applied Physics Letters, 2006. **88**(26).
162. Niu, X.Z., et al., *Electro-Coalescence of Digitally Controlled Droplets.* Analytical Chemistry, 2009. **81**(17): p. 7321-7325.
163. Mazutis, L., *Droplet-based microfluidics for protein evolution*. 2009, Strasbourg University.
164. Abate, A.R., et al., *High-throughput injection with microfluidics using picoinjectors.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(45): p. 19163-19166.

165. Fidalgo, L.M., et al., *From microdroplets to microfluidics: Selective emulsion separation in microfluidic devices.* Angewandte Chemie-International Edition, 2008. **47**(11): p. 2042-2045.
166. Shestopalov, I., J.D. Tice, and R.F. Ismagilov, *Multi-step synthesis of nanoparticles performed on millisecond time scale in a microfluidic droplet-based system.* Lab on a Chip, 2004. **4**(4): p. 316-321.
167. Li, L., J.Q. Boedicker, and R.F. Ismagilov, *Using a multijunction microfluidic device to inject substrate into an array of preformed plugs without cross-contamination: Comparing theory and experiments.* Analytical Chemistry, 2007. **79**(7): p. 2756-2761.
168. Shi, W.W., et al., *Droplet-based microfluidic system for individual Caenorhabditis elegans assay.* Lab on a Chip, 2008. **8**(9): p. 1432-1435.
169. Schmitz, C.H.J., et al., *Dropspots: a picoliter array in a microfluidic device.* Lab on a Chip, 2009. **9**(1): p. 44-49.
170. Huebner, A., et al., *Static microdroplet arrays: a microfluidic device for droplet trapping, incubation and release for enzymatic and cell-based assays.* Lab on a Chip, 2009. **9**(5): p. 692-698.
171. Frenz, L., et al., *Reliable microfluidic on-chip incubation of droplets in delay-lines.* Lab on a Chip, 2009. **9**(10): p. 1344-1348.
172. Courtois, F., et al., *An integrated device for monitoring time-dependent in vitro expression from single genes in picolitre droplets.* Chembiochem, 2008. **9**(3): p. 439-446.
173. Koster, S., et al., *Drop-based microfluidic devices for encapsulation of single cells.* Lab on a Chip, 2008. **8**(7): p. 1110-1115.
174. Tan, W.H. and S. Takeuchi, *A trap-and-release integrated microfluidic system for dynamic microarray applications.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(4): p. 1146-1151.
175. Marcoux, P.R., et al., *Micro-confinement of bacteria into w/o emulsion droplets for rapid detection and enumeration.* Colloids and Surfaces a-Physicochemical and Engineering Aspects, 2011. **377**(1-3): p. 54-62.
176. Dewan, A., et al., *Growth kinetics of microalgae in microfluidic static droplet arrays.* Biotechnology and Bioengineering, 2012. **109**(12): p. 2987-2996.
177. Joensson, H.N., M. Uhlen, and H.A. Svahn, *Droplet size based separation by deterministic lateral displacement-separating droplets by cell-induced shrinking.* Lab on a Chip, 2011. **11**(7): p. 1305-1310.
178. Maenaka, H., et al., *Continuous and size-dependent sorting of emulsion droplets using hydrodynamics in pinched microchannels.* Langmuir, 2008. **24**(8): p. 4405-4410.
179. Tan, Y.C., Y.L. Ho, and A.P. Lee, *Microfluidic sorting of droplets by size.* Microfluidics and Nanofluidics, 2008. **4**(4): p. 343-348.
180. Kurup, G.K., Basu A. S., *Passive, label-free droplet sorting by chemical composition using tensiophoresis*, in *Microtas*. 2012.
181. Baret, J.C., et al., *Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity.* Lab on a Chip, 2009. **9**(13): p. 1850-1858.
182. Franke, T., et al., *Surface acoustic wave (SAW) directed droplet flow in microfluidics for PDMS devices.* Lab on a Chip, 2009. **9**(18): p. 2625-2627.
183. Ahn, K., et al., *Dielectrophoretic manipulation of drops for high-speed microfluidic sorting devices.* Applied Physics Letters, 2006. **88**(2).

184. Zhang, K., et al., *On-chip manipulation of continuous picoliter-volume superparamagnetic droplets using a magnetic force.* Lab on a Chip, 2009. **9**(20): p. 2992-2999.
185. Cordero, M.L., et al., *Thermocapillary manipulation of droplets using holographic beam shaping: Microfluidic pin ball.* Applied Physics Letters, 2008. **93**(3).
186. Huebner, A., et al., *Quantitative detection of protein expression in single cells using droplet microfluidics.* Chem Commun (Camb), 2007(12): p. 1218-20.
187. Edd, J.F., et al., *Controlled encapsulation of single-cells into monodisperse picolitre drops.* Lab Chip, 2008. **8**(8): p. 1262-4.
188. Kemna, E.W.M., et al., *High-yield cell ordering and deterministic cell-in-droplet encapsulation using Dean flow in a curved microchannel.* Lab on a Chip, 2012. **12**(16): p. 2881-2887.
189. Martin, K., et al., *Generation of larger numbers of separated microbial populations by cultivation in segmented-flow microdevices.* Lab Chip, 2003. **3**(3): p. 202-7.
190. Huebner, A., et al., *Development of quantitative cell-based enzyme assays in microdroplets.* Anal Chem, 2008. **80**(10): p. 3890-6.
191. Shim, J.U., et al., *Simultaneous determination of gene expression and enzymatic activity in individual bacterial cells in microdroplet compartments.* J Am Chem Soc, 2009. **131**(42): p. 15251-6.
192. Brouzes, E., et al., *Droplet microfluidic technology for single-cell high-throughput screening.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(34): p. 14195-200.
193. Kintses, B., et al., *Picoliter Cell Lysate Assays in Microfluidic Droplet Compartments for Directed Enzyme Evolution.* Chemistry & Biology, 2012. **19**(8): p. 1001-1009.
194. Debs, B.E., et al., *Functional single-cell hybridoma screening using droplet-based microfluidics.* Proceedings of the National Academy of Sciences, 2012. **109**(29): p. 11570-11575.
195. He, M., et al., *Selective encapsulation of single cells and subcellular organelles into picoliter- and femtoliter-volume droplets.* Anal Chem, 2005. **77**(6): p. 1539-44.
196. Williams, R., et al., *Amplification of complex gene libraries by emulsion PCR.* Nature Methods, 2006. **3**(7): p. 545-550.
197. Mardis, E.R., *Next-generation DNA sequencing methods.* Annual Review of Genomics and Human Genetics, 2008. **9**: p. 387-402.
198. Shendure, J. and H.L. Ji, *Next-generation DNA sequencing.* Nature Biotechnology, 2008. **26**(10): p. 1135-1145.
199. Li, M., et al., *BEAMing up for detection and quantification of rare sequence variants.* Nature Methods, 2006. **3**(2): p. 95-97.
200. Pekin, D., et al., *Quantitative and sensitive detection of rare mutations using droplet-based microfluidics.* Lab Chip, 2011. **11**(13): p. 2156-66.
201. Kumaresan, P., et al., *High-throughput single copy DNA amplification and cell analysis in engineered nanoliter droplets.* Analytical Chemistry, 2008. **80**(10): p. 3522-3529.
202. Kojima, T., et al., *PCR amplification from single DNA molecules on magnetic beads in emulsion: application for high-throughput screening of transcription factor targets.* Nucleic Acids Research, 2005. **33**(17).
203. Chabert, M., K.D. Dorfman, and J.L. Viovy, *Using droplet microfluidics to design a continuous flow PCR system.* Houille Blanche-Revue Internationale De L Eau, 2006(6): p. 51-56.

204. Mohr, S., et al., *Numerical and experimental study of a droplet-based PCR chip.* Microfluidics and Nanofluidics, 2007. **3**(5): p. 611-621.
205. Curcio, M. and J. Roeraade, *Continuous segmented-flow polymerase chain reaction for high-throughput miniaturized DNA amplification.* Analytical Chemistry, 2003. **75**(1): p. 1-7.
206. Beer, N.R., et al., *On-chip single-copy real-time reverse-transcription PCR in isolated picoliter droplets.* Analytical Chemistry, 2008. **80**(6): p. 1854-1858.
207. Beer, N.R., et al., *On-chip, real-time, single-copy polymerase chain reaction in picoliter droplets.* Analytical Chemistry, 2007. **79**(22): p. 8471-8475.
208. Schaerli, Y., et al., *Continuous-Flow Polymerase Chain Reaction of Single-Copy DNA in Microfluidic Microdroplets.* Analytical Chemistry, 2009. **81**(1): p. 302-306.
209. Tewhey, R., et al., *Microdroplet-based PCR enrichment for large-scale targeted sequencing.* Nature Biotechnology, 2009. **27**(11): p. 1025-U94.
210. Mazutis, L., et al., *Droplet-Based Microfluidic Systems for High-Throughput Single DNA Molecule Isothermal Amplification and Analysis.* Analytical Chemistry, 2009. **81**(12): p. 4813-4821.
211. Dittrich, P.S., M. Jahnz, and P. Schwille, *A new embedded process for compartmentalized cell-free protein expression and on-line detection in microfluidic devices.* Chembiochem, 2005. **6**(5): p. 811-+.
212. Dittrich, P.S., M. Jahnz, and P. Schwille, *A novel approach for miniaturized in vitro protein expression in microfluidic channels.* Micro Total Analysis Systems 2004, Vol 1, 2005(296): p. 52-54.
213. Fallah-Araghi, A., et al., *A completely in vitro ultrahigh-throughput droplet-based microfluidic screening system for protein engineering and directed evolution.* Lab on a Chip, 2012. **12**(5): p. 882-891.
214. Mazutis, L., et al., *Multi-step microfluidic droplet processing: kinetic analysis of an in vitro translated enzyme.* Lab Chip, 2009. **9**(20): p. 2902-8.
215. Xia, Y.N. and G.M. Whitesides, *Soft lithography.* Annual Review of Materials Science, 1998. **28**: p. 153-184.
216. Siegel, A.C., et al., *Cofabrication of electromagnets and microfluldic systems in poly(dimethylsiloxane).* Angewandte Chemie-International Edition, 2006. **45**(41): p. 6877-6882.
217. Kaplan, J. and W.F. DeGrado, *De novo design of catalytic proteins.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(32): p. 11566-11570.
218. Althoff, E.A., et al., *Robust design and optimization of retroaldol enzymes.* Protein Science, 2012. **21**(5): p. 717-726.
219. Giger, L., *Directed evolution as a versatile tool to investigate natural and artificial aldolases*. 2011, ETH: Zurich.
220. List, B., C.F. Barbas, and R.A. Lerner, *Aldol sensors for the rapid generation of tunable fluorescence by antibody catalysis.* Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(26): p. 15351-15355.
221. Hai, M.T. and S. Magdassi, *Investigation on the release of fluorescent markers from w/o/w emulsions by fluorescence-activated cell sorter (vol 96, pg 393, 2004).* Journal of Controlled Release, 2004. **97**(2): p. 383-383.
222. Wen, L. and K.D. Papadopoulos, *Visualization of water transport in W1/O/W2 emulsions.* Colloids and Surfaces A: Physicochemical and Engineering Aspects, 2000. **174**(1–2): p. 159-167.

223. Wen, L.X. and K.D. Papadopoulos, *Effects of surfactants on water transport in W-1/O/W-2 emulsions.* Langmuir, 2000. **16**(20): p. 7612-7617.
224. Hedstrom, L., *Serine protease mechanism and specificity.* Chem Rev, 2002. **102**(12): p. 4501-24.
225. Maurer, K.H., *Detergent proteases.* Current opinion in biotechnology, 2004. **15**(4): p. 330-334.
226. Legendre, D., et al., *Display of active subtilisin 309 on phage: Analysis of parameters influencing the selection of subtilisin variants with changed substrate specificity from libraries using phosphonylating inhibitors.* Journal of Molecular Biology, 2000. **296**(1): p. 87-102.
227. Chaplin, M.F., *Enzyme Technology*. 1990.
228. Spungin, A. and S. Blumberg, *Streptomyces-Griseus Aminopeptidase Is a Calcium-Activated Zinc Metalloprotein - Purification and Properties of the Enzyme.* European Journal of Biochemistry, 1989. **183**(2): p. 471-477.
229. Park, H.I. and L.J. Ming, *A 10(10) Rate Enhancement of Phosphodiester Hydrolysis by a Dinuclear Aminopeptidase-Transition-State Analogues as Substrates?* Angew Chem Int Ed Engl, 1999. **38**(19): p. 2914-2916.
230. da Silva, G.F. and L.J. Ming, *Catechol oxidase activity of di-Cu2+-substituted aminopeptidase from Streptomyces griseus.* J Am Chem Soc, 2005. **127**(47): p. 16380-1.
231. James, L.C. and D.S. Tawfik, *Conformational diversity and protein evolution--a 60-year-old hypothesis revisited.* Trends Biochem Sci, 2003. **28**(7): p. 361-8.
232. Ercan, A., H.I. Park, and L.J. Ming, *A "moonlighting" dizinc aminopeptidase from Streptomyces griseus: mechanisms for peptide hydrolysis and the 4 x 10(10)-fold acceleration of the alternative phosphodiester hydrolysis.* Biochemistry, 2006. **45**(46): p. 13779-93.
233. Pigliucci, M., *Opinion - Is evolvability evolvable?* Nature Reviews Genetics, 2008. **9**(1): p. 75-82.
234. van Ditmarsch, D., *Selection of Streptomyces griseus aminopeptidases by droplet microuidics*. 2009.
235. Alirezaei, M., et al., *Zinc-induced inhibition of protein synthesis and reduction of connexin-43 expression and intercellular communication in mouse cortical astrocytes.* European Journal of Neuroscience, 2002. **16**(6): p. 1037-1044.
236. Chen, Y., A. Wijaya Gani, and S.K.Y. Tang, *Characterization of sensitivity and specificity in leaky droplet-based assays.* Lab on a Chip, 2012. **12**(23): p. 5093-5103.
237. Varadarajan, N., et al., *Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(19): p. 6855-6860.