NOVEL FLOW CYTOMETER-BASED PLATFORMS FOR DIRECTED EVOLUTION

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Dipl. Ing. (FH)
Bioingenieurwesen

Christian Pitzler
aus Dormagen

Berichter: Universitätprofessor Dr. Ulrich Schwaneberg
Universitätprofessor Dr. Andrij Pich


Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek online verfügbar
Acknowledgements

First of all I want to thank Prof. Dr. Ulrich Schwaneberg for giving me the opportunity to work on my PhD-thesis in his research group and for his effort to pursue my scientific and personal development. I also would like to thank my committee members Prof. Dr. Andrij Pich, Prof. Dr. Lothar Elling and Prof. Dr. Alan Slusarenko for being part of my PhD disputation.

The open and friendly atmosphere in the Schwaneberg group made me have a great time during my PhD studies. I want to thank especially both of my supervisors Ronny and Ljubica for their support, discussions and motivation. In addition I thank the subgroup leaders Joelle, Marco, Felix and Leilei as well as my research students Alina and Michael for their motivation and contribution to research projects.

Special thanks go to my collaboration partners Nina and Feng for their support and successful projects. I also want thank my research students for their motivation and contribution to research projects.

I want to thank my cooking group fellows Julia, Nina, Erik, Christian, Marcus and Anne for the lunches with lots of yummy taste experiences.

My family and friends are acknowledged for supporting me during my studies and in my whole life. Last but not least I thank Lena for her endless love and support in any situation.
Abstract

Novel Flow Cytometer-based Platforms for Directed Evolution

Modern biotechnology has successfully launched synthesis routes that compete with traditional chemical processes for production of bulk chemicals and pharmaceuticals. In order to alter naturally occurring enzymes to perform under industrial conditions, directed evolution has become the predominant tool. Directed evolution mimics natural evolution through iterative rounds of mutagenesis and screening, yielding enzyme variants with i.e. improved activity, stability or tolerance towards organic solvents. The main bottleneck in directed evolution is to screen through the generated sequence space that can easily exceed $10^8$ clones in standard epPCR based approaches. A common directed evolution campaign lasts up to one year, whereby the screening for improved variants is the most time consuming part of the whole project. Novel screening formats that minimize sample size and increase throughput greatly reduce the cost and time burden of directed evolution campaigns. However, so far each of these formats lack a broad application potential, like being applicable to more than one specific enzymatic reaction or being limited to display technologies. Here, two ultra-high throughput screening platforms with broad utility were developed that use fluorescence activated cell sorting in order to isolate improved variants.

The first developed flow cytometer-based technology for high throughput screening is based on a coupled reaction of a phytase from *Yersinia mollaretii* (YmPh) and a glucose oxidase, converting glucose-6-phosphate to glucono-δ-lactone and hydrogen peroxide. Fenton’s reaction produces hydroxyl radicals, acting as initiator of poly(ethyleneglycol)-acrylate-based polymerization incorporating a fluorescent monomer. As a consequence, a fluorescent hydrogel is formed around *E. coli* cells expressing active YmPh. Validation of the method was performed by screening model libraries with defined ratios of active and inactive YmPh with flow cytometry, resulting in a 5-fold enrichment of the active population. Further screening of in total $1.8 \times 10^7$ events of an YmPh epPCR library yielded an enrichment of the active population from 40 % close to 90 % and a variant with 97 U/mg higher specific activity.

The application of different glucose derivatives as substrates pursued the advancement of the technology into a general high-throughput screening toolbox for directed evolution of hydrolases. Libraries of lipolytic (*Bacillus subtilis* lipase A, BSLA) and cellulolytic enzymes (CelA2) were screened and yielded improved variants (BSLA: 1.3-fold increase in $k_{cat}$, CelA2: 1.7-fold increase in $k_{cat}$) after one round of directed evolution. Particularly noteworthy is the usage of a natural substrate (cellobiose) in a high throughput screening format enabling identification of cellulase variants with increased activity.

The second flow cytometer-based technology developed during this work includes *in vitro* compartmentalization (IVC) of a cellulase DNA library inside polymersomes and subsequent sorting on flow cytometer. This technology represents the first ultra-high throughput screening platform for directed evolution that makes use of the beneficial properties of polymersomes. The strategy of IVC in polymersomes and subsequent flow cytometer sorting was validated by screening $3.9 \times 10^7$ events of a celA2 library expressed *in vitro*. Remarkably, analysis of 400 single variants in microtiter plate before and after flow cytometer sorting yielded an impressive three times enrichment of the active population in one round of sorting.

The developed technologies for high throughput flow cytometer screening allow due to their reliable, straightforward and fast implementation the realization of one round of directed evolution in less than one week. Consequently, these methods can solve demanding and challenging problems in directed evolution like revealing structure-function relationships and performing many rounds of directed evolution with high mutational load.
# Table of Contents

Abstract ................................................................................................................................................................. 3

1 General Introduction .............................................................................................................................................. 1

1.1 Biotechnology and protein engineering ........................................................................................................ 1

1.2 Gene diversity generation .................................................................................................................................. 2

1.3 High throughput screening (HTS) .................................................................................................................. 4

1.3.1 Hydrogels ..................................................................................................................................................... 6

1.3.2 Compartmentalization technologies ......................................................................................................... 7

1.3.3 Compartmentalization with polymersomes ............................................................................................. 10

1.4 Enzymes .......................................................................................................................................................... 12

1.4.1 Phytases ...................................................................................................................................................... 12

1.4.2 Lipases .......................................................................................................................................................... 13

1.4.3 Cellulases ................................................................................................................................................... 14

2 Goal of this work .................................................................................................................................................. 16

3 A Fluorescent Hydrogel-Based Flow Cytometry HTS Platform for Phytases ........................................... 18

3.1 Declaration ....................................................................................................................................................... 18

3.2 Pilot experiment and objective ...................................................................................................................... 18

3.3 Introduction ....................................................................................................................................................... 19

3.4 Results .............................................................................................................................................................. 20

3.4.1 Optimization of coupled YmPh/GOx reaction in microtiter plate format .............................................. 21

3.4.2 Detection of YmPh/GOx initiated hydrogel polymerization ..................................................................... 22

3.4.3 Flow cytometry analysis of YmPh cells with a defined active to inactive ratio ..................................... 24

3.4.4 Generation and screening of a YmPh mutant library ............................................................................. 25

3.4.5 Characterization of wild type YmPh and variants M1 and M2 ................................................................. 27

3.4.6 Structural model of YmPh and M1 ............................................................................................................ 29

3.5 Discussion ......................................................................................................................................................... 31

3.6 Materials and methods .................................................................................................................................. 34

3.6.1 Plasmids and strains ................................................................................................................................... 34
# Table of Contents

3.6.2 Protein expression and cell labeling ........................................................................... 34  
3.6.3 2,2’-azino-bis(3-ethylbenzothioazioline-6-sulphonic acid) (ABTS) screening system in 96 well microtiter plates ....................................................................................... 34  
3.6.4 Flow cytometry analyses and cell sorting ....................................................................... 35  
3.6.1 Confocal Microscopy ..................................................................................................... 35  
3.6.2 Scanning force microscopy (SFM) ................................................................................ 35  
3.6.3 Diversity generation and YmPh library construction ...................................................... 35  
3.6.4 Protein expression in 96-well microtiter plates ............................................................. 36  
3.6.5 Screening in 96-well microtiter plates ......................................................................... 37  
3.6.6 Expression and purification of YmPh and variant M1 .................................................. 37  
3.6.7 Homology model generation with YASARA .................................................................. 38  

4 Toolbox for Detecting Hydrolytic Activity by Flow Cytometry-Based HTS .................. 39  
4.1 Declaration ....................................................................................................................... 39  
4.2 Objective ........................................................................................................................ 39  
4.3 Introduction ...................................................................................................................... 40  
4.4 Results .............................................................................................................................. 41  
4.4.1 Detection of Fur-Shell labelled cells with confocal microscopy ................................. 42  
4.4.2 Flow cytometer analyses of BSLA and CelA2 libraries ............................................... 43  
4.4.3 Characterization of improved variants ....................................................................... 45  
4.4.4 Determination of kinetic parameters ($K_M$, $k_{cat}$) .................................................... 47  
4.4.5 Homology model of BSLA and CelA2 ...................................................................... 48  
4.5 Discussion ....................................................................................................................... 51  

4.6 Materials and Methods .................................................................................................... 53  
4.6.1 Cloning into expression vector .................................................................................. 53  
4.6.2 Diversity generation by random mutagenesis ............................................................ 53  
4.6.3 Hydrogel Formation .................................................................................................... 55  
4.6.4 Flow cytometry analyses and cell sorting ................................................................... 55  
4.6.5 Protein expression in 96-well microtiter plates (MTP) ............................................... 55  
4.6.6 Screening systems ....................................................................................................... 56  
4.6.7 Purification .................................................................................................................. 56
<table>
<thead>
<tr>
<th>5</th>
<th>Flow Cytometry-Based HTS Platform for Cell Free Expression in Polymersomes</th>
<th>58</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Objective</td>
<td>58</td>
</tr>
<tr>
<td>5.2</td>
<td>Introduction</td>
<td>58</td>
</tr>
<tr>
<td>5.3</td>
<td>Results</td>
<td>60</td>
</tr>
<tr>
<td>5.3.1</td>
<td><em>In vitro</em> expression of CelA2</td>
<td>60</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Dynamic light scattering (DLS)</td>
<td>61</td>
</tr>
<tr>
<td>5.3.3</td>
<td><em>In vitro</em> expression in polymersomes and flow cytometer analysis</td>
<td>62</td>
</tr>
<tr>
<td>5.4</td>
<td>Discussion</td>
<td>69</td>
</tr>
<tr>
<td>5.5</td>
<td>Materials and methods</td>
<td>73</td>
</tr>
<tr>
<td>5.5.1</td>
<td>Preparation of lyophilizes polymersomes</td>
<td>73</td>
</tr>
<tr>
<td>5.5.2</td>
<td>Encapsulation of the transcription-translation machinery</td>
<td>73</td>
</tr>
<tr>
<td>5.5.3</td>
<td>Activity determination in 96 well MTP format</td>
<td>74</td>
</tr>
<tr>
<td>5.5.4</td>
<td>FACS sorting, DNA recovery, and transformation</td>
<td>74</td>
</tr>
<tr>
<td>5.5.5</td>
<td>Dynamic light scattering (DLS)</td>
<td>75</td>
</tr>
<tr>
<td>6</td>
<td>Conclusions</td>
<td>76</td>
</tr>
</tbody>
</table>

Publications | 78 |
Abbreviations | 79 |
Declaration | 80 |
References | 81 |
Acknowledgements | 89 |
Curriculum Vitae | 90 |
1 General Introduction

1.1 Biotechnology and protein engineering

In nature, evolution persists through millions of years as a accumulation of adaptations gradually producing new properties and species. The process of evolution through adaption and selection was firstly proposed by Charles Darwin [1]. In contrary to natural evolution, alteration of living organisms by humans through domestication and artificial selection can be traced back since the Neolithic. Examples for alteration of organisms by mankind are the transformation of corn from the small grass teosinte or the evolution of dairy cows [2]. However, usage of organisms by mankind was not restricted to metabionta. Yeast and bacteria have been used centuries ago for manufacturing bread, yoghurt, beer, or wine and their evolution yielded e.g. highly tolerant ethanol-resistant strains.

Natural evolution has deloped an immense diversity of catalytically active molecules that can be used individually or together by combining natural functional units. Enzymes, however, have evolved in their natural habitats and conditions, exhibiting their optimum functionality and stability to well-defined chemical environments and temperature ranges.

A new era of biotechnology, where novel enzymes can be designed and altered for a broad range of applications (e.g. bioplastic [3], biofuels [4] pharmaceuticals [5]) arose with the development of novel methods originating from diverse sciences such as chemical engineering, bioprocess engineering, and bioinformatics. The dawn of modern biotechnology in the early 1970s was accomplished introducing exogenous genetic material by transformation into a bacterium [6]. Since then, incredible advances have been made in finding ways to manipulate organisms, genes, genomes, and in elucidating their biomolecular functions.

Modern biotechnology enabled the application of enzymes in a wide range of industrial processes and products by providing technologies for manipulating enzyme properties towards non-natural conditions. By doing so, biotechnology may contribute to solve major problems of modern society as to lower the greenhouse effect and to find alternatives to a petrochemical-based economy. Spiegelman et al. introduced the concept of “evolution in a test tube” in 1967 [7]. In this artificial form of evolution called “directed evolution” genes are evolved towards new or altered functions, by applying selection pressures on mutant gene libraries expressed in heterologous systems such as Escherichia coli. Directed evolution of enzymes in vitro became a powerful technology for adapting enzyme properties to obtain catalysts that perform specific functions under defined conditions. Despite researcher’s ability to alter or synthesize genetic information, in-depth knowledge on how to design genes accomplishing a desired function in a correctly folded protein de novo is still missing. Most
protein structures *per se* are so complex that it is nearly impossible to foresee how altering their structure will affect their function. Hence, the approach of trial-and-error in directed evolution campaigns is nowadays still the fastest way to adapt protein variants to desirable traits and functions. The procedure of directed evolution comprises three steps (Figure 1):

- **Step I: Gene diversity generation:** Chemical and enzymatic methods, *in vitro* and *in vivo* approaches [8]
- **Step II: Screening and selection:** See chapter 1.3
- **Step III: Isolation:** Isolating the gene encoding for the improved enzyme variant for kinetic characterization or another round of gene diversity generation and screening for further improvement of the enzyme.

Directed evolution is performed in iterative cycles, until the enzyme is altered to the desired property. In principle, this general setup of a directed evolution experiment might be useful for optimizing any enzyme property, which can be screened in an economically feasible way, even if the molecular basis of that property is not known [9].

![Figure 1: Principle of a directed evolution experiment comprising three steps: I. Gene diversity generation II. Screening for improved variants III. Isolation of gene encoding for improved protein variant. Rounds of directed evolution can be iteratively repeated, until the enzyme yields the desired property or activity improvement [10].](image)

### 1.2 Gene diversity generation

Success stories in the field of directed enzyme evolution comprise many applications, starting from biotransformation for bioremediation, replacement of organo-metallic catalysts in chemical synthesis, biosensors to vaccines and therapeutic proteins [11-15]. Furthermore,
directed evolution has played a key role in elucidating protein structure-function relationship [16].

A number of methods are nowadays available to generate gene diversity in vitro. In general, they can be divided into rational (e.g. site-specific), semi-rational (e.g. recombination of sequences PTRec, or saturation of multiple positions with OmniChange), and random mutagenesis (e.g. epPCR, or SeSaM) approaches. [17-21]

Deciding how to generate gene diversity depends on the particular application. The most commonly used random mutagenesis techniques are PCR-derived methods under low fidelity conditions. A robust and simple directed evolution technique is to generate variants with random mutations by error-prone PCR (epPCR). Taq DNA polymerase from Thermus aquaticus that lacks a 3’→5’ proofreading exonuclease activity is commonly used in epPCR experiments. It has an intrinsic high error rate of $8.0 \times 10^6$ mutations per bp per duplication [22]. This gives around one substitution per 125000 bp per PCR cycle which is obviously insufficient for randomizing genes that seldom exceed 5000 bp in length. This technique utilizes various approaches to increase the error rate of DNA polymerases, such as addition of MnCl$_2$, unbalanced nucleotide concentrations, increased extension times, or increased concentrations of MgCl$_2$ [23]. In addition, there are alternative polymerases having higher error rates like Mutazyme or engineered Pfu polymerases, which can be used in epPCR [24, 25]. Besides this, epPCR methods lead to a transition-biased mutation spectra due to the redundancy of the genetic code. Furthermore, the lack of consecutive mutations reduces the sequence space to 40 % of the possible protein sequence space [26]. Nevertheless, epPCR is still the leading method for diversity generation in directed evolution due to its technical simplicity and robustness. Additionally, average epPCR mutation rates can be adjusted through careful manipulation of amplification reaction conditions (PCR cycles, dNTPs concentration, MnCl$_2$ concentration). epPCR techniques reach their performance limit when it comes to exploiting the whole available protein’s sequence space; concretely this means the capacity to replace every amino acid by their 19 equivalents. Recently, the limitations of epPCR were shown and discussed in directed evolution experiments of a small enzyme (lipase from Bacillus subtilis (BSLA), 181 amino acids) [27]. BSLA resistance towards an ionic liquid was studied by screening a BSLA library, where every position was saturated and overall 292 beneficial substitutions at 104 different amino acid positions were identified. In a parallel epPCR experiment, only 18 beneficial positions were identified, proving that 80 % beneficial amino acid positions are not targeted in a traditional directed evolution experiment [28]. It was proposed that higher quality libraries can be generated with the Sequence Saturation Mutagenesis method (SeSaM) [21]. Here, usual limitations of epPCR-derived methods caused by polymerase bias (transitions favored over transversions), is tackled by using universal bases, that allow also mutations of subsequent nucleotides. Furthermore, the
OmniChange method represents an advancement on focused semi-rational saturation mutagenesis of single positions, usually performed with the QuikChange method [20]. In OmniChange, multiple positions (up to five) in a protein sequence are simultaneously saturated, leading to the identification of cooperative effects of mutated amino acids that are not identified with common methods.

1.3 High throughput screening (HTS)

A major challenge of directed evolution is to cover the generated combinatorial complexity of protein sequence space by screening. In order to explore the complete protein sequence space it is necessary to analyze millions of genetic variants for identification and isolation of the improved one. For a small protein of 200 amino acids, the theoretical protein sequence space contains $20^{200} \approx 10^{260}$ sequences. Approximately screening of $10^{260}$ unique protein sequences are beyond the capacity of any screening technology. The largest protein libraries screened by now are reported for display technologies of mRNA-protein fusions and have had around $10^{15}$ variants [29]. Due to technical limitations of random mutagenesis methods and redundancy of the genetic code, an average amino acid substitution of 3.15–7.40 per amino acid position can be achieved with up-to-date methods [30]. Ultimately, this leads to a generated sequence space between $10^{100}$ and $10^{173}$ possible mutants. During a common directed evolution experiment using polymerase chain reaction (PCR), usually $10^7$ to $10^{12}$ amplicons (synthesized DNA molecules) are generated. In order to cover a meaningful fraction of this sequence space, it is necessary to develop novel ultra-high throughput screening technologies.

A crucial factor in a directed evolution campaign is the selection of the most appropriate screening system to monitor the desired enzyme property. The assay conditions of the screening have to mimic the final application condition as close as possible, such as the choice of substrate, physical and chemical characteristics. Most screening systems have to be customized for the enzyme of choice and to be sensitive over the substrate concentrations expected at the final application [31]. Additionally, a decision on the configuration of the screening format (low, medium or high throughput) has to be made on the basis of the library size. Typically, variants generated in a directed enzyme evolution campaign are screened in agar plates or microtiter plates, limiting the screening to $5 \times 10^5$ variants by using robotic automation, which is many orders of magnitude lower than the typical size of a random mutagenesis library (around $10^{10}$ variants for PCR-based libraries). In general, lower throughput screenings can be applied to monitor a wide range of activities but are cost intensive, since each clone is tested separately [32]. On the other hand, higher throughput enzyme screenings have a simpler composition since assay costs are reduced through miniaturization of the assay. Up-to-date ultra-high-throughput screens, such as microfluidic systems and flow cytometer-based assays carry required volumes to extremes,
and include millions of assay points in a sample of less than one milliliter. Novel flow cytometer as well as microfluidic-based systems have an analysis rate (throughput) of up to $10^7$ events per hour and emerged as suitable screening formats for large libraries [33-35] (Figure 2). Due to their outstanding throughput, they increase success probabilities in directed evolution experiments and play a pivotal role in the question whether new enzymes can efficiently be isolated from metagenome libraries [36].

![Figure 2: Overview of screening methods used in directed evolution campaigns.](image)

Flow cytometer-based screening systems require fluorescent-based product detection to discriminate among enzyme variants. The linkage between genotype and phenotype is the basis of selection-based evolution. In nature, the linkage between genotype and phenotype is maintained by encapsulation of the inner components of the cell and separation by cell membranes. One technology that maintains the linkage between genotype and phenotype is the display technology. Several display strategies have been reported for screening through libraries with flow cytometers, including yeast, bacterial or phage display [37-39]. However, expanding display technologies to select for properties besides protein-protein interaction or
joining two molecules with *e.g.* sortases seems to be virtually impossible, because the linkage between genotype and phenotype is not maintained [40]. Two alternative technologies to display screens are introduced in the following chapters promising a more general applicability. One alternative is to deploy compartmentalization technology that has the potential to be applicable for all enzymes that convert a substrate into a fluorescent product (1.3.2). The other alternative is the formation of a fluorescent hydrogel that links active enzymes to the particular genetic information (1.3.1).

### 1.3.1 Hydrogels

Hydrogels are water-swollen, cross-linked networks that have great potential for use in biotechnology applications. Their characteristics *e.g.* softness, biocompatibility and ability for rapid diffusion of molecules make them useful for drug delivery, cell culture, wound healing and sensing applications [41-44]. Fluorescent-labeled hydrogels are a reliable fluorescent-based screening technology in 8-well strip plate format for antigen detection [45]. Fluorescent monomers are thereby copolymerized during hydrogel formation as a result of a glucose oxidase (GOx) catalyzed conversion of β-D-glucose into δ-gluconolactone. The formed hydrogen peroxide generates hydroxyl radicals in presence of ferrous ions (Fe²⁺) and water. Subsequently, the hydroxyl radicals act in presence of poly(ethyleneglycol) (PEG)-acrylate as initiators for synthesis of PEG-based hydrogels [46]. Interestingly, the radical-initiated polymerization is efficiently catalyzed at ambient temperature and pressure by GOx and unaffected by dissolved oxygen [47]. PEG-based hydrogels are highly permeable, biocompatible and have been employed to entrap bacteria through thiol-ene click reactions or TEMED/potassium persulfate; and alternatively ‘DNA-damaging’ UV light exposure was used for radical generation [48-53]. In addition, fibroblasts cells (NIH3T3) have been shown to be highly viable after entrapment in PEG-based hydrogels generated by GOx mediated polymerization [46].

Fluorescent hydrogels that surround cells would be an attractive novel compartment technology for flow cytometry-based screening systems. Prerequisite is a formation of a fluorescent polymer shell that can be linked to the activity of expressed enzyme variants. So far, labelling cells which express active enzymes with fluorescent hydrogels in high throughput screening systems has never been tested. A screening system based on this approach provides independence of fluorogenic substrates that are often not available for any given reaction or difficult to synthesize. Furthermore, omitting fluorogenic molecules in enzyme assays allows the usage of natural substrates, a criterion which is highly desired for directed evolution campaigns carried out for industrial enzymes.
1.3.2 Compartmentalization technologies

In order to mimic compartmentalization akin to a cell membrane in a laboratory, usually water-in-oil-in-water emulsions (double emulsions) are used. The isolated environment that keeps this linkage is essential when substrates or fluorescent products are freely diffusing in and out of the cell, when using cell lysates or if enzymes are secreted into the media. A substrate is incubated inside the isolated environment of a compartment together with the expressed enzyme. The enzyme converts the substrate into a fluorescent product which indicates the functional activity of a particular genetic variant. A library of thiolactonases in single bacterial cells was encapsulated by Aharoni et al. in water-in-oil-in-water double emulsions, thus maintaining a linkage between the gene, the encoded enzyme variant, and the fluorescent product [54]. Another approach used cell lysates containing a model library of arylsulfatases encapsulated in double emulsions with microfluidic systems. With this method rare events such as 1 positive out or 100000 negative could be isolated showing a potential applicability for libraries with high mutational load [55].

These methods share the common drawback that variant libraries are expressed in vivo, requiring to be cloned into a suitable plasmid and to be transformed with the expression host. The presence of “nicked” DNA when transformed often leads to poor transformation efficiency in the expression host causing a loss of diversity during transformation.

In vitro protein expression is a technique that enables researchers to rapidly produce small amounts of functional proteins. Two main components are needed to accomplish in vitro protein production: the DNA template (linear operon or plasmid) encoding the target protein and a reaction solution containing the necessary transcriptional and translational molecular machinery. Besides these, additives supporting the transcription and translation machinery are used to express proteins, including RNA polymerases for mRNA transcription, ribosomes for polypeptide translation, tRNA and amino acids, enzymatic cofactors, energy sources, and cellular components essential for proper protein folding. Translation is usually carried out with E. coli cell extract and more recently, systems based on extracts from insect cells, mammalian cells and human cells have been developed that implement post translational modifications.

In vitro expression systems offer an attractive alternative not only eliminating steps of cloning, transformation, culture growth and induction, but also preventing a tremendous loss of diversity from \(\sim 10^{10}\) variants to \(\sim 10^{5}\) variants that comes along with the transformation step in traditional cloning strategies. Cell-free transcription–translation synthesis helps to avoid protein aggregation into inclusion bodies, simplifies protein purification, offers easy control over reaction conditions and the ability to supplement the reaction mixture [56, 57]. Furthermore, in vitro protein expression provides much greater flexibility and significant time-savings, such as the ability to express toxic proteins, using substrates that are impermeable
for the cell wall, and the expression of a linear gene to a functional enzyme in one day. The latter example was shown for Old Yellow Enzyme, indicating that the enzyme activity for linear DNA templates are only slightly lower, yielding 80% to 90% synthesis relative to circular DNA and comparable expression levels to heterologous expression on *E. coli* [58]. By using *in vitro* expression for high throughput screening, the gene encoding the enzyme of interest is entrapped together with a substrate and the cell extract within a compartment enabling the link between genotype and phenotype. Upon transcription and translation, active enzymes convert the substrate into a fluorescent product and the fluorescent droplet can be detected by a flow cytometer [59]. A schematic representation in Figure 3 shows the principle of cell-free expression in the compartmentalized structure of a water-in-oil-in-water double emulsion.

**Figure 3:** The double emulsion provides a closed environment for directed evolution of *in vitro* expressed enzymes. The surfactant stabilized oil phase mimics a cell wall and thereby retains encapsulated machinery for cell-free protein synthesis encapsulated. A double emulsion contains an *in vitro* expression kit, a molecule of mutant DNA, and the substrate of the fluorogenic reaction. Upon transcription and translation, active enzymes convert the substrate into a fluorescent product. This idea to produce enzymes *in vitro* and use emulsions for compartmentalization was first developed by Tawfik *et al.*, in 1998 and in 2003 proven to be valid for a directed phosphotriesterase evolution experiment using microbeads [60, 61].

A summary of recently published flow cytometer and microfluidic-based technologies is shown in Table 1.
Table 1: Compilation of flow cytometry-based and microfluidics-based high-throughput screening methods for directed enzyme evolution, including developed enzymes, obtained method, throughput and references (Ref.).

<table>
<thead>
<tr>
<th>Enzymes</th>
<th>Method</th>
<th>Throughput</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>in vitro compartmentalization</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β-galactosidase</td>
<td>extrusion to form double emulsions</td>
<td>20,000 events/s</td>
<td>[62]</td>
</tr>
<tr>
<td><strong>whole cell compartmentalization</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>glucose oxidase</td>
<td>stirring to form double emulsions</td>
<td>100 events/s</td>
<td>[63]</td>
</tr>
<tr>
<td><strong>whole cells</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>esterase</td>
<td>extrusion to form double emulsions</td>
<td>8,000 events/s</td>
<td>[64]</td>
</tr>
<tr>
<td>thiolactonases</td>
<td>homogenizing to form double emulsions</td>
<td>4,000 events/s</td>
<td>[65]</td>
</tr>
<tr>
<td>β-glucosidase</td>
<td>homogenizing to form double emulsions</td>
<td>~1,000,000 sorted events</td>
<td>[66]</td>
</tr>
<tr>
<td>phosphatase</td>
<td>internalized probe</td>
<td>2800 events/s</td>
<td>[67]</td>
</tr>
<tr>
<td>glycosyltransferases</td>
<td>internalized probe</td>
<td>4000 events/s</td>
<td>[59]</td>
</tr>
<tr>
<td>orthogonal nucleoside analog kinase</td>
<td>internalized probe</td>
<td>&lt;2000 events/s</td>
<td>[68]</td>
</tr>
<tr>
<td>protease</td>
<td>fluorescent cell-surface labeling</td>
<td>50,000,000 events screened</td>
<td>[69]</td>
</tr>
<tr>
<td>enantioselective esterase</td>
<td>fluorescent cell-surface labeling</td>
<td>15,000-30,000 events/s</td>
<td>[70]</td>
</tr>
<tr>
<td>glycosyltransferase</td>
<td>internalized probe</td>
<td>&gt;10,000,000 sorted events</td>
<td>[31]</td>
</tr>
<tr>
<td><strong>microfluidics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α-amylase</td>
<td>on-chip sorting of cells</td>
<td>323 droplets/s</td>
<td>[71]</td>
</tr>
<tr>
<td>horseradish peroxidase</td>
<td>on-chip sorting of cells</td>
<td>2,000 droplets/s</td>
<td>[72]</td>
</tr>
</tbody>
</table>

In essence, these reports conclude that the current technology allows sampling through a protein sequence space that can be generated by peptides as short as six to seven amino acids. Most proteins are significantly larger, and for exploring novel directed evolution strategies with high mutational load, screening systems with throughputs higher than $10^7$ are needed. The development of an *in vitro* expression-based ultra-high throughput screening system for flow cytometry that detects enzymatic reactions during fluorescent product formation allows sampling through unprecedented sequence space. Through maintenance of the genetic diversity in large libraries (~$10^{10}$ variants) expressed in cell-free systems, valuable insights into mechanisms of protein evolution are provided [73]. The possibility to screen large libraries in high throughput allows deciphering enzyme functions for example reveal synergistic or additive interactions between non-adjacent amino acids. The concept of understanding structure-function relationships by recombination and saturation multiple
amino acid positions is proposed by Cheng et al. to be part of next generation directed evolution campaigns (directed evolution 2.0) [28].

1.3.3 Compartmentalization with polymersomes

Block copolymers are macromolecules consisting of two or more different hydrophobic and hydrophilic homopolymers that have unique physico-chemical properties. Block copolymers have molecular weights up to several orders of magnitude higher than phospholipids. The hydrophobic part of block copolymers has been constructed of poly(ethyl ethylene), poly(butadiene), poly(dimethylsiloxane), poly(styrene) poly(lactide), poly(ε-caprolactone) and poly(trimethylene carbonate) whereas (acrylic acid), poly(L-glutamic acid) and poly(ethyleneglycol) have been frequently selected as water-soluble blocks [74-80].

As a consequence of their bilayer-like structure, block copolymers are able to self-assemble in aqueous solution into highly entangled membranes that, assemble into their entropically most stable state as hollow vesicles, called polymersomes [81]. Miscellaneous block copolymers with varying properties can be produced by applying blocks with different functionalities, molecular weights, compositions and molecular architectures. In return, properties of polymersomes including stability, fluidity, as well as structural features are greatly influenced by characteristics of the homopolymers [81]. Depending on the method, polymersomes may be formed with relatively high control on their size distribution, ranging from tens of nanometers to tens of micrometers [82]. Typically, polymersomes are formed by techniques such as film rehydration, electroformation, phase transfer, or ultrasonication. Polymer hydration techniques are based on the hydration of amphiphilic block copolymers to induce self-assembly into polymersomes. Polymersomes generated with these methods usually show a large size distribution. In order to generate monodisperse polymersomes, a templated assembly from double emulsions formed via microfluidic approaches was described, resulting in a size between 50 and 100 µm [83]. Polymersomes can encapsulate hydrophilic, hydrophobic and amphiphilic molecules like any other vesicular structure [79, 84, 85]. Most methods for polymersome formation are not compatible with the mild conditions required for encapsulation of biological compounds.

A suitable triblock copolymer for forming polymer vesicles in mild conditions is poly(2-methyloxazoline-b-dimethylsiloxane-b-2-methyloxazoline) (PMOXA-PDMS-PMOXA), due to its ability to self-assemble into polymersomes in aqueous solutions. Vesicles formed of the triblock copolymer were used for incorporation of aquaporins into the membrane, for membrane insertion of an in vitro expressed G-protein coupled receptor, and to induce membrane permeability by a photoreaction [86-88]. The molecular structure of PMOXA-PDMS-PMOXA is shown in Figure 4.
In comparison to liposomes, built up by phospholipids, polymersome membranes usually are multiple times thicker, due to the longer homopolymers chains of the block copolymers. Hence the membrane provides polymersomes a couple of advantages compared to liposomes, making them more suitable for encapsulation. Polymersome membranes allow encapsulation of both hydrophobic and hydrophilic compounds [89]. Polymersomes have almost an order of magnitude higher mechanical response than liposomes, and their permeability towards water is more than one order of magnitude lower compared to liposomes [74]. Moreover, phospholipids are prone to oxidation, but polymersomes may be assembled from relatively stable compounds and can have significantly longer shelf lives [90]. Applications of polymeric vesicles comprise systems for drug delivery, nanoreactors or for cell-free expression of enzymes [87, 90, 91]. One example of compartmentalized gene expression inside polymersomes was given by Nallani et al. The authors designed a method to encapsulate an in vitro expressed lipase into vesicles composed of the diblock copolymer (polystyrene-polyisocyanopeptide) and sorted by flow cytometry the fluorescent population according to the fluorescent signal generated after substrate conversion [92]. Diffusion in and out of the polymersomes could be selectively stopped by the polyelectrolyte poly-L-lysine, allowing the fluorescent product to be entrapped. This approach showed the general compatibility of the in vitro enzyme production machinery towards polymersomes, in which a single vesicle included thousands of DNA molecules.
To our knowledge, encapsulation and expression of a library of single gene variants in polymersomes for the purpose of a directed evolution experiment has not yet been reported in literature.
1.4 Enzymes

Enzymes are present in every living organism and perform reactions of anabolism and catabolism.

With more than 2000 fully sequenced genomes [93] and several thousand enzymes being isolated and characterized a distinct classification was necessary to handle the large and growing amount of available data. Enzymes are grouped into six classes (EC number = enzyme commission number) according to the specifically catalyzed reaction:

- **EC 1: Oxidoreductases**: Catalyze oxidation/reduction reactions
- **EC 2: Transferases**: Transfer a functional group
- **EC 3: Hydrolases**: Catalyze the hydrolysis of various bond types
- **EC 4: Lyases**: Cleave various bonds by means other than hydrolysis and oxidation
- **EC 5: Isomerases**: Catalyze isomerization changes within a single molecule
- **EC 6: Ligases**: Catalyze the joining of two molecules

A large compilation of reported and characterized enzymes can be found in the ‘BRENDA’ database (http://www.brenda-enzymes.org) [94]. Since this work focused on the three enzymes phytases, lipases and cellulases a more detailed presentation of those can be found in the following chapters.

1.4.1 Phytases

Phytases (EC-number: 3.1.3.8) catalyze the hydrolysis of phytic acid into inorganic phosphate and myo-inositol phosphate derivatives (Figure 7). Four structurally different classes of phytases have been reported and comprise (1) histidine acid phosphatase (HAP), (2) β-propeller phytase, (3) cysteine phosphatase, and (4) purple acid phosphatase (Lei et al. 2007). Catalytically, most of phytases belong to the family of histidine acid phosphatases that is characterized by a conserved active site hepta-peptide motif R-H-G-X-R-X-P and the catalytically active dipeptide HD. This group of enzymes catalyzes the phytic acid hydrolysis in two steps: a nucleophilic attack from the histidine in the active site of the enzyme to the scissile phosphoester bond of phytic acid and protonation of the leaving group by the aspartic acid residue of the HD motif (37). The large active-site region allows HAPs to accommodate various kind of phosphate esters, such as pNPP, AMP, ATP, fructose 1,6-bisphosphate, and glucose 6-phosphate. Members of the HAP family have attracted industrial interest because of their ability to efficiently hydrolyze phytate (>100 U/mg) (38).
Figure 5: Phytase catalyzes hydrolysis of phytate into myo-inositol and inorganic phosphate.

At physiological conditions phytic acid exists as phytate that chelates approximately 12-20 % calcium and 1-2 % iron and zinc. Phytases are widely used in food and feed industries for improving digestibility and assimilation of nutrients of foods and feeds by reducing the antinutritional effects of phytic acid (36). Although phytate serves as the major source of energy and phosphorus for seed germination, the bound phosphorus is poorly available to simple-stomached animals. Thus, inorganic phosphorus, a non-renewable and expensive mineral, is supplemented in diets for swine, poultry, and fish to meet their nutrient requirement. Meanwhile, the unutilized phytate phosphorus from plant feeds is excreted, becoming an environmental pollutant in areas of intensive animal agriculture. Excessive phosphorus in soil runs off to lakes and the sea, causing eutrophication and stimulating growth of aquatic organisms that may produce neurotoxins injurious to humans (39).

1.4.2 Lipases

Lipases (EC number: 3.1.1.3) are a class of hydrolases with the biological function to catalyze the hydrolysis of triacylglycerol to glycerol and free fatty acids (Figure 6). In contrast to esterases, lipases generally show activation at oil-water interfaces and their substrates are relatively long-chained acylglycerols (>C_12) [95]. Due to their high specificity and wide range of reactions, they play an important role in chemical, food, and pharmaceutical industries. Lipases are ubiquitous and found in animals, plants, fungi, and bacteria. Microbial lipases are of special interest because of their high activity, selectivity and stability towards extreme conditions (temperature, pH, organic solvents). Lipases derived from the gram-positive bacteria Bacillus subtilis differ significantly from commonly known lipases. One of these enzymes is Bacillus subtilis Lipase A (BSLA), which consists of only 181 amino acids (19.3 kDa) and is the smallest lipase known. It exhibits an optimum activity at pH of 10 and shows highest activity on glycerol esters with medium length (C_8) fatty acid chains. Because of its small size, it has been an interesting candidate for mutagenesis experiments [96]. BSLA does not possess a lid-like structure, which normally covers the active site of lipases in...
the absence of oil-water interfaces, and therefore is not dependent on interfacial activation. Additionally, alanine replaces the first glycine residue in the G-X-S-X-G pentapeptide, which is highly conserved in the active site of other lipases.

![Diagram of lipolytic hydrolysis](image)

**Figure 6:** Scheme of lipolytic hydrolysis of triacylglycerol into glycerol and carboxylic acids.

### 1.4.3 Cellulases

Cellulases (EC-number: 3.2.1.4) are involved into the hydrolysis of cellulose, whereas the complete digestion refers to a whole network of hydrolytic enzymes including endoglucanase, exoglucanase and β-glucosidases. Endoglucanases hydrolyze accessible intramolecular β-1,4-glycosidic bonds of cellulase chains randomly to produce new chain ends. Exoglucanases cleave cellulose chains at the reducing or non-reducing ends to release soluble cellobiose or glucose and β-glucosidases hydrolyze cellobiose into glucose (Figure 7) [97]. Cellulose is a linear condensation polymer consisting of D-anhydroglucopyranose joined together by β-1,4-glycosidic bonds with a degree of polymerization from 100 to 20,000 [98]. Coupling of adjacent cellulose chains and sheets of cellulose by hydrogen bonds and van der Waal’s forces results in a parallel alignment and a crystalline structure with straight, stable fibers.

Cellulases are important industrial enzymes, applied in pulp and paper, textile, laundry, biofuel production, food and feed industry, brewing, and agriculture. [99]
Figure 7: Scheme of enzymatic cellulose hydrolysis. Endos stands for endoglucanase, exosR for exoglucanase cleaving at reducing ends, exosNR for exoglucanase cleaving at non-reducing ends and β-Gase for β-Glucosidase. Figure was taken from Zhang et al., 2006.
2 Goal of this work

Within the framework of the German Federal Ministry of Education and Research (BMBF) this work was financed under the program “Basis technologien für die nächste Generation biotechnologischer Verfahren” (FKZ 031A165) supporting the development of novel production routes for biotechnology. The goal of this thesis was to generate novel high throughput screening technologies for directed evolution experiments carried out in vivo and in vitro that are robust in use and have a broad application spectrum. In order to fill the gap between generated sequence space and limited screening capacity, high throughput screening enables the sampling of mutant libraries of a size that would be inaccessible by microtiter plate-based assays on either cost or time considerations. Particularly the time effort to screen through libraries is reduced from several months to about one week and the volumes of reagents needed for screening are million-fold less than that required for microtiter plate screening, resulting in significant savings in cost and availability of reagents [100]. These technologies allow the isolation of active variants from large mutant libraries containing a high number of inactive clones. Novel HTS-technologies allow researchers to pursue novel mutagenesis strategies with high mutational loads e.g. the saturation of more than five positions at once and thereby contribute to a better understanding of structure-function relationships and in addition engineering of enzymes exhibiting improved properties after only one round of directed evolution. To date only nine research articles with focus on directed evolution campaigns employing flow cytometer-based high throughput technologies are published (Table 1).

The development of an ‘in vivo technology’ for whole cells, called Fur-Shell, comprises a high throughput screening platform for directed enzyme evolution that by utilizing free radical polymerization labels cells expressing active enzymes with fluorescent hydrogel. For validation of the method, directed evolution of phytase, and as a further development towards a toolbox for hydrolytic enzymes for a cellulase and a lipase will be performed. Within the Fur-Shell technology, the E. coli cell membrane is used as a “naturally confined barrier” and its surface is used as a template for hydrogel production. Compared to existing in vivo flow cytometer screenings, this approach achieves versatile application for hydrolases and oxidases overcoming drawbacks as leakage of the substrate/product due to cell membrane permeability. Additionally, the novel technology platform uses non-fluorescent glucose derivatives as substrates overcoming the need to synthesize broad varieties of fluorogenic substrates.

The development of an ‘in vitro technology’ comprises an ultra-high throughput screening platform using cell-free expressed enzymes encapsulated in polymersomes. Validation of the
method will be performed through a directed evolution campaign of CelA2 and determination of enrichment factors as a benchmark for identification of improved enzyme variants. Cell-free expression systems advance directed evolution campaigns, because they can be used for the expression of toxic, proteolytically sensitive or unstable proteins. In order to maintain the genotype-phenotype linkage, not more than one gene encoding for a single enzyme variant should be present per polymersome. Upon active enzyme production and subsequent substrate conversion fluorescent product labels polymersomes which will be sorted by flow cytometry in ultra-high throughput. Compartamentalization technology based on polymersomes provides an advanced encapsulation system compared to already reported systems using double emulsion and liposomes. Polymersomes have almost an order of magnitude higher mechanical response than liposomes, and their permeability towards water is at least ten times reduced compared to liposomes [73]. In addition, amphiphilic properties of polymersomes enable usage of broad range of substrates having hydrophilic as well as hydrophobic moieties. The use of polymersomes for encapsulation will advance the in vitro compartamentalization (IVC) technology, which is currently limited to double emulsions and liposomes, to a more robust and broader applicable platform in directed enzyme evolution.
3 A Fluorescent Hydrogel-Based Flow Cytometry HTS Platform for Phytases

3.1 Declaration

Parts of this chapter have been published in the journal Chemistry & Biology and are adapted to this thesis with permission of Elsevier


3.2 Pilot experiment and objective

Johnson et al. reported formation of fluorescent hydrogel layers through enzyme induced polymerization based on glucose oxidase (GOx) and glucose with subsequent production of hydroxyl radicals in Fenton’s reaction [101]. On this basis, an experiment was performed within this thesis using the Saccharomyces cerevisiae strain (ngd29)/pYES2 secreting active GOx in order to investigate if in vivo produced enzyme can induce hydrogel formation. The addition of 20 mM β-D-glucose and a monomer solution (poly(ethyleneglycol)-diacrylate and N-vinylpyrrolidone) to GOx expressing yeast cells led to a time dependent formation of hydrogel around the cells secreting active enzyme indicated by formation of an amorphous structure around the cells (Figure 8).

![Figure 8: Microscope image of yeast cells (EZ12) taken 0 h, 8 h and 24 h after polymerization initiation. Fluorescence image shows yeast cells after 24 h stained with nile red for better identification of cells inside the hydrogel. In the negative control where cells do not secrete GOx, no formation of hydrogel could be observed after 24 h.](image)

Based on data obtained in this pilot experiment the aim of this part was to develop a flow cytometer-based screening platform that selects for active enzyme variants upon fluorescent hydrogel formation in a coupled enzyme reaction of phytase and GOx.
3.3 Introduction

A major constrain for practical applications of phytases is an insufficient activity at the present pH values inside the stomach of monogastric animals which is around pH 3.5. A couple of directed evolution campaigns have been performed to render phytases more suitable for industrial processes [102-104]. Solving crystal structures of phyA (Aspergillus niger), appA (Escherichia coli) or YkPhy (Yersinia kristensenii) has enabled researchers to identify structure-function relationships of microbial phytases [105-107]. A multi-site saturation mutagenesis (OmniChange) of YmPh (Yersinia mollaretii) yielded an improved phytase at low pH values (3-fold improved stability at pH 2.8) and 32 % improved residual activity (58°C for 20 min) [108].

A proposed reaction for the conversion of glucose-6-phosphate by a phytase, leading towards a polymerization through free radicals is shown in Figure 9. In the first step, YmPh catalyzes the hydrolysis of glucose-6-phosphate into β-D glucose and phosphate (Figure 9a).

In the second step, GOx oxidizes β-D-glucose to δ-gluconolactone and generates thereby hydrogen peroxide. Ferrous ions (Fe^{2+}) and hydrogen peroxide produce free hydroxyl radicals (Fenton’s reaction) (Figure 9b), which initiate hydrogel formation by polymerizing PEG-acrylate monomers [109]. During the polymerization reaction, fluorescent Polyfluor 570 acrylate is co-polymerized into the hydrogel which surrounds E. coli cells that contain YmPh in active form and resembles a fur-like structure (Fur-Shell). Fluorescent Polyfluor 570 has an excitation maxima at \( \lambda_{ex} \) 548 nm and emission maxima at \( \lambda_{em} \) 570 nm (Figure 9c). The incorporation of the fluorescent monomers into the hydrogel enables a flow cytometry based cell sorting of the E. coli population expressing active enzymes.
Figure 9: Coupled phytase (YmPh)/glucose oxidase (GOx) reaction cascade leading to a fluorescent hydrogel shell (PEG-Polyfluor 570). (a) Glucose-6-phosphate is hydrolyzed by YmPh, releasing β-D-glucose and inorganic phosphate. (b) Produced β-D-glucose is converted by GOx to δ-glucosonolactone and thereby generates hydrogen peroxide. Hydrogen peroxide produces in presence of Fe²⁺ hydroxyl radicals (Fenton reaction). (c) In the presence of hydroxyl radicals, poly(ethylene glycol)-diacylate, Polyfluor 570 and 1-Vinyl-2-pyrrolidinone monomers polymerize and generate due to incorporated Polyfluor 570 a fluorescent hydrogel around E. coli cells that express active YmPh variants. (Figure was reprinted with permission of Elsevier)

3.4 Results

The results section is divided into three parts in which the principle of developed Fur-Shell technology platform is introduced, optimized and validated in a directed YmPh evolution experiment. In the first section, the principle of the coupled YmPh/GOx reaction system is described. The second section comprises an optimization of hydrogen peroxide formation in microtiter plate format and analyzes hydrogel formation using confocal microscopy and scanning force microscopy (SFM) measurements. The third section describes a validation of the developed screening platform based on coupled enzyme reaction system (YmPh/GOx) and fluorescent hydrogel formation by screening an YmPh-mutant library in a flow cytometry setup and analyzing sorted populations in microtiter plate format.
3.4.1 Optimization of coupled YmPh/GOx reaction in microtiter plate format

Upon conversion of glucose-6-phosphate into β-D-glucose and inorganic phosphate, oxidation of produced β-D-glucose is mediated by GOx and hydrogen peroxide is produced. Subsequently, through Fenton reaction hydrogen peroxide is converted into highly reactive hydroxyl radical species which are initiating a hydrogel formation. The amount of hydroxyl radicals formed is directly proportional to the amount of produced hydrogen peroxide during oxidation of β-D-glucose. Optimization of polymerization time in regard of fluorescent hydrogel formation was performed using already reported 4-MUP detection system. Upon induction with auto-induction media *E. coli* BL21 lacI Q1/pALXtreme YmPh and *E. coli* BL21 lacI Q1/pALXtreme empty vector were three times washed and subsequently 4-MUP assay was performed every two hours. *E. coli* cells harboring an empty vector showed in the first 12 hours a basal activity on the substrate 4-MUP which decreases by increase in incubation time. *E. coli* cells harboring the YmPh showed a strong increase in fluorescence per sec in the first 4 h of reaction time (Figure 10). Further incubation revealed a constant level of fluorescence over the analyzed time period of 18 hours. Taking into account the background reaction coming from *E. coli* cells harboring an empty vector the highest increase of fluorescence signal per second was observed after 16 h and was chosen as optimal time for further experiments.

![Figure 10: Identification of the best time point for hydrogel formation.](image)

The ABTS assay (3.6.3) was used for optimization of coupled YmPh/GOx reaction in 96-well microtiter plate format as well as for determining the produced hydrogen peroxide amount in the coupled YmPh/GOx reaction. Figure 11a shows increase in absorbance coming from oxidation of ABTS by produced hydrogen peroxide. The differences between YmPh (+) and empty vector containing cells (YmPh (-)) under non-optimized reaction conditions are minor,
therefore optimization of reaction conditions was required. In order to reduce the background signal different reaction parameters were varied i.e. substrate concentration (1 - 500 mM glucose-6-phosphate), concentration of GOx (0.4 - 3.0 µM GOx type VII from *Aspergillus niger* (K<sub>M</sub> of 33-110 mM for β-D-glucose [110]) or 0.4 - 3.0 µM GOx type II from *Aspergillus niger* (K<sub>M</sub> of 33-110 mM for β-D-glucose [110])), cell density (0.5 and 2; OD<sub>600</sub>), and pH (4.5 to 7.0). In the final reactions following concentrations were used: glucose-6-phosphate (20 mM), GOx type II from *Aspergillus niger* (1.55 µM), *E. coli* cells in 250 mM sodium acetate buffer, OD<sub>600</sub>: 1pH (5.5). Under these conditions a 6-fold signal to noise ratio after 16 minutes of reaction with ABTS was observed, indicating a sufficient difference among the absorbance signal coming from *E. coli* cells expressing active YmPh wild type and cells harboring an empty vector (Figure 11b).

**Figure 11**: Optimization of hydrogen peroxide production via the coupled reaction system YmPh/GOx. *E. coli* BL21(DE3) lac<sup>Q1</sup>pALXtreme-5b YmPh cells (+) are displayed as filled circles, *E. coli* BL21(DE3) lac<sup>Q1</sup>/empty pALXtreme-5b YmPh (-) cells are displayed as empty circles. Graphs show the increasing concentration of hydrogen peroxide using the ABTS assay: (a) before optimization and (b) after optimization. Final reaction conditions for ABTS assay were determined as follows: cell suspension in sodium acetate buffer (140 µl, pH 5.5, 250 mM) with an optical density of 1 (OD<sub>600</sub>), glucose-6-phosphate (20 mM), GOx Type II (1.58 µM), Horseradish peroxidase (0.025 U/ml), ABTS (2.5 mM). (Figure was reprinted with permission of Elsevier)

### 3.4.2 Detection of YmPh/GOx initiated hydrogel polymerization

Hydrogel polymerizations on the surface of *E. coli* cells were initiated with the addition of 24.8 % PEG<sub>575</sub>-DA, 2.5 % 1-vinyl-2-pyrrolidone, 0.005 % Polyfluor 570, and 25 µM Fe<sub>2</sub>SO<sub>4</sub>*7H<sub>2</sub>O in a total reaction volume of 5 ml. After 16 min incubation *E. coli* cells were washed and polymerization was visualized by confocal microscopy (Leica TCS SP8, Leica Microsystems, Wetzlar, Germany). Figure 12 shows overlays of transmission and fluorescence images in order to illustrate labeling efficiency with copolymerized Polyfluor 570 of YmPh (-) cells (Figure 12a) and YmPh (+) cells (Figure 12b). Comparison of both Figures demonstrates a striking difference in fluorescence polymer shell formation within YmPh (+) and YmPh (-) cells. In case of the YmPh(+), nearly all cells are showing fluorescence,
whereas the negative cells only show a slight background signal which results most likely from a genomic *E. coli* phytase [106]. A 3.7-fold higher fluorescence intensity on average was determined in YmPh (+) cells compared to YmPh (-) cells by grey level quantification using ImageJ software.

![Figure 12: Overlay of fluorescence and transmission images which are recorded by confocal microscope. (a) *E. coli* BL21(DE3) lacI^Q1^ / empty pALXtreme-5b cells with no YmPh activity. (b) *E. coli* BL21(DE3) lacI^Q1^ / pALXtreme-5b cells expressing active YmPh. Both cell types are incubated under optimized polymerization conditions for 16 min. (Figure was reprinted with permission of Elsevier)](image)

In order to investigate morphological changes on the surfaces of bacterial cells scanning force microscopy (SFM) has been reported to act as a reliable method [111]. Therefore, SFM measurements of YmPh (-) and YmPh (+) *E. coli* cells were performed to confirm and to analyze polymer shell formation around YmPh (+) cells. Treated YmPh (-) and YmPh (+) samples were washed five times with demineralized water prior to SFM measurements to avoid formation of salt crystals from PBS during the drying process. A SFM phase image of a single dried *E. coli* cell (negative control) is shown in Figure 13a. The *E. coli* cell retains its shape with a relatively smooth surface after drying. An SFM phase image of a single, dried YmPh (-) cell that was incubated with 20 mM glucose-6-phosphate, 1.58 mM GOx, 24.8 % PEG575-DA, and 2.5 % 1-vinyl-2-pyrrolidone for 16 min (background reaction) is shown in Figure 13b. The *E. coli* cell shows the formation of a wrinkled structure on the surface that on one hand may occur due to surface instabilities during shrinkage of the cell upon drying and on the other hand may also point to the formation of a small polymer brush. The latter is in accordance to the results obtained by confocal microscopy (faint fluorescent signal, Figure 12a). A YmPh (+) cell (Figure 13c) under identical incubation conditions displays pronounced and significantly more heterogeneous features in the polymer shell surrounding the *E. coli* cell compared to the background reactions shown in Figure 13b. In addition, the phase images and the height profiles (cross-section in insets in Figure 13) show a polymer shell
stretching across the silicon wafer and surrounding the *E. coli* cells, again confirming the successful polymerization around the cells. Height profile and phase image in Figure 13c clearly indicating the corona formed by the polymer shell (black circles) explain why YmPh (+) cells show a 3.7-fold enhanced fluorescence in confocal microscope analysis (Figure 12).

Figure 13: SFM height (small) and phase images (large) (z-range: 700 nm, 35-40°; image size 3 x 3 µm) with respective cross sections. Positions of cross section are marked in height images with white bars. All samples were washed using demineralized water instead of PBS to avoid salt crystals. (a) untreated *E. coli* BL21 (DE3) lacI<sup>Q1</sup> cell. (b) YmPh (-) cells, incubated with 20 mM glucose-6-phosphate, 1.58 µM glucose oxidase (GOx), 25 µM Fe<sub>2</sub>SO<sub>4</sub>*7H<sub>2</sub>O, 24.8 % w/w PEG<sub>575</sub>-DA, 2.5 % w/w 1-vinyl-2-pyrrolidone for 16 min. (c) YmPh (+) cells, treated with 20 mM glucose-6-phosphate, 1.58 µM glucose oxidase (GOx), 25 µM Fe<sub>2</sub>SO<sub>4</sub>*7H<sub>2</sub>O, 24.8 % w/w PEG<sub>575</sub>-DA, 2.5 % w/w 1-vinyl-2-pyrrolidone for 16 min. (Figure was reprinted with permission of Elsevier)

### 3.4.3 Flow cytometry analysis of YmPh cells with a defined active to inactive ratio

*E. coli* cells containing YmPh (+) and *E. coli* cells containing YmPh (-) and model libraries containing different ratios of *E. coli* YmPh (+) and *E. coli* YmPh (-) in a ratio of 1:9 and 3:7 (YmPh (+) vs. YmPh (-)) were analyzed for determining parameters to efficiently sort by flow cytometer. Analysis and sorting was performed with an event rate of 5000 cells/s. Figure 14 shows that an increasing amount of cells expressing active YmPh leads to a clear shift towards higher fluorescence. In case of 0 % YmPh (+) only 2 % of the population is located inside the gate (P1) which was used for sorting active variants (Figure 14a). 10 % YmPh (+) result in 15 % of positive events inside the gate (P1) (Figure 14b), whereas for 30 % YmPh (+), 55 % of the population was within the sorting gate (P1; Figure 14c). In sample containing 100 % YmPh (+), 81 % of the population was inside the gate (P1) (Figure 14d). Thus, P1 was selected as an optimal gate for sorting on flow cytometer.
Figure 14: Flow cytometry analysis of YmPh-expressing cells mixed in different active/inactive ratios. (a) 0 % YmPh (+) cells, 100 % YmPh (-) cells. (b) 10 % YmPh (+) cells, 90 % YmPh (-) cells. (c) 30 % YmPh (+) cells, 70 % YmPh (-) cells. (d) 100 % YmPh (+) cells, 0 % YmPh (-) cells. The results are plotted as forward scatter (FSC) versus fluorescence signal ($\lambda_{ex}$ 561 nm and $\lambda_{em}$ 585 nm). Density of the population is represented as red > yellow > green > blue. Lines indicate gate P1 for sorting criterion of the fluorescent population. (Figure was reprinted with permission of Elsevier)

Sorted cell populations were subsequently grown on LB_Amp plates, picked and transferred into two 96-well microtiter plates for expression. YmPh activities were determined using the reported 4-MUP detection system [108, 112]. Enrichment in the active population after sorting proves the general applicability of the Fur-Shell technology for use as pre-screening system for large libraries with high mutational load, as well as the correct implementation of the gating parameters. A five-times enrichment of the active population (50 % YmPh (+) cells) was achieved, starting from a model library consisting of 10 % YmPh (+) cells. The model library harboring 30 % YmPh (+) population was 2.9 times enriched, corresponding to a 89 % YmPh (+) population after sorting.

3.4.4 Generation and screening of a YmPh mutant library

An epPCR library of YmPh was generated using 0.2 mM MnCl$_2$ and activity determination of 200 random clones with the 4-MUP assay in 96-well microtiter plates yielded a fraction of
40% active variants. Mutation frequency of the library was determined as 4.8 mutations/gene by sequencing five randomly selected mutants. Clones were considered as "active" when the increase in fluorescence per second was at least three times higher than the increase in fluorescence per second of empty vector control cells (YmPh (-)).

The random mutant library (epPCR; 0.2 mM MnCl$_2$) of YmPh in *E. coli* BL21(DE3) lacI$^Q$ was analyzed by flow cytometry (Figure 15a). The right dot plot in Figure 15a shows the active population fraction of 37% which is in very good agreement to the 4-MUP activity detection system revealing a 40% active population and within the expected range of the flow cytometry results shown in Figure 14c (30% active population). The 'upper' 10% of the YmPh population were sorted and the sorting efficiency is confirmed by flow cytometer re-analysis (Figure 15b) which shows that >90% of the sorted population are within the sorting gate (P1).

A comparison in activity of YmPh variants using 96-well microtiter plate based 4-MUP assay is shown in Figure 15c (epPCR library before sorting) and Figure 15d (flow cytometer enriched epPCR library). The YmPh population was analyzed in 96-well plates to determine the activity ratio in before and after sorting. After epPCR library sorting, total activity was increased 2.2-fold starting from 40% before sorting to 89% active variants after sorting.

**Figure 15: Flow cytometer analysis of an YmPh mutant library (epPCR) employing the 4-MUP activity detection system.** (a) Density plot of YmPh mutant library showing the initial population forward scatter (FSC) versus side scatter (SSC) and FSC versus fluorescence intensity (561/585 nm) (0.2 mM MnCl$_2$ epPCR) (density = red > yellow > green > blue). (b) Re-analysis of the sorted YmPh mutant library comprising 10% of the most fluorescent population. The phytase activity of a sample of the population was determined using the 4-MUP detection system in 96-well microtiter plate format. (c) YmPh mutant library before sorting (epPCR; 40% active)
and (d) sorted YmPh mutant library (89 % active). Activity of wild type YmPh is highlighted as black bar (Figure was reprinted with permission of Elsevier)

In total, ~1000 YmPh variants within the sorting gate P1 were collected and plated onto LB<sub>Amp</sub> agar plates, grown and transferred to 96-well microtiter plates. YmPh variant expression was subsequently carried out in auto-induction medium and YmPh activity was determined using the 4-MUP activity detection system. Six YmPh variants with significantly increased 4-MUP activity (between 1.3- and 3.0-fold) were selected for rescreening (Figure 16). After rescreening, two clones – M1 and M2 showing increased activity towards 4-MUP compared to YmPh wild type were purified and characterized.

![Figure 16: Rescreening of 6 variants showing increased activity compared to YmPh wild type using the 4-MUP detection system. The reported values are the average of six measurements and error bars represent the standard deviation from the mean (n = 6).](image)

### 3.4.5 Characterization of wild type YmPh and variants M1 and M2

The YmPh variants with the highest 4-MUP activity were sequenced (M1: Gly7Cys, Ala166Thr, Gly335Ser, Gly353Asp, His433Arg; M2: Asp85Gly, Asn88Ser), purified, kinetically characterized and compared to the YmPh wild type in terms of 4-MUP and phytic acid conversion. The pH profile of M1 was slightly broadened compared to YmPh wild type, whereas the pH profile of M2 remained more or less the same. Interestingly, YmPh, M1 and M2 showed a pH optimum at pH 3.9 for 4-MUP and a pH optimum at pH 4.5 for phytic acid (Figure 17a, b).
Figure 17: pH profile of YmPh and mutants M1 and M2. Phytase activity was measured at different pH values ranging from 3.5 to 5.5 with two different substrates, (a): and 4-MUP and (b): phytic acid.

In previous studies by Shivange et al. in 2011 a tetrameric assembly of YmPh enzymes was detected with an acrylamide gel performed under natural conditions (tetramer, ∼188 kDa; monomer, ∼47 kDa, [113]). Monomeric enzymes display Michaelis-Menten kinetics, whereas multimeric enzymes show often allosteric interactions characterized by sigmoidal kinetics, described by a Hill coefficient greater than 1 [114]. Enzyme kinetics of YmPh, M1, and M2 were determined at pH 4.5 using 4-MUP and sodium phytic acid suggesting that the product formation is driven by an allosteric interaction between YmPh monomers (Figure 18a, b). The curves for 4-MUP and phytic acid follow a sigmoidal shape due to a slower substrate conversion at low substrate concentrations.

Figure 18: Initial activities were plotted against different substrate concentrations and fitted with Hill equation. (a): shows YmPh and mutants M1 and M2 kinetics with substrate 4-MUP. (b): shows YmPh and mutants M1 and M2 kinetics with substrate the substrate phytic acid.

M1 showed a 31 % increased catalytic activity ($k_{cat} = 1293 \text{ s}^{-1}$) for the artificial substrate 4-MUP (applied selection pressure) compared to YmPh wild type (Table 2) and a slightly (5 %) increased activity for the natural substrate phytic acid ($k_{cat} = 5540 \text{ s}^{-1}$; see Table 3). Compared to other directed evolution campaigns, the increase in specific activity of around 30 % for 4-MUP seems to be little, but can be attributed to the already high initial activity of YmPh wildtype of 315 U mg$^{-1}$. 
Table 2: Kinetic parameters determined with 4-MUP for YmPh wild type and variants M1 and M2. Values in the brackets represent standard errors determined in triplicate measurements.

<table>
<thead>
<tr>
<th></th>
<th>$k_{cat} (s^{-1})$</th>
<th>Specific activity [U mg$^{-1}$]</th>
<th>Hill coefficient</th>
<th>$k_{half} [\mu M]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>YmPh</td>
<td>998 (± 18)</td>
<td>315 (± 6)</td>
<td>2.08 (± 0.12)</td>
<td>1414 (± 48)</td>
</tr>
<tr>
<td>M1</td>
<td>1293 (± 61)</td>
<td>412 (± 19)</td>
<td>2.26 (± 0.19)</td>
<td>1005 (± 96)</td>
</tr>
<tr>
<td>M2</td>
<td>1065 (± 22)</td>
<td>339 (± 7)</td>
<td>2.26 (± 0.17)</td>
<td>1247 (± 48)</td>
</tr>
</tbody>
</table>

Table 3: Kinetic parameters determined with phytic acid for YmPh wild type and variant M1. Values in the brackets represent standard errors determined in triplicate measurements.

<table>
<thead>
<tr>
<th></th>
<th>$k_{cat} (s^{-1})$</th>
<th>Specific activity [U mg$^{-1}$]</th>
<th>Hill coefficient</th>
<th>$k_{half} [\mu M]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>YmPh</td>
<td>5267 (± 206)</td>
<td>1681 (± 66)</td>
<td>1.77 (± 0.16)</td>
<td>150 (±12)</td>
</tr>
<tr>
<td>M1</td>
<td>5541 (± 157)</td>
<td>1769 (± 66)</td>
<td>1.92 (± 0.16)</td>
<td>130 (±8)</td>
</tr>
<tr>
<td>M2</td>
<td>5355 (± 62)</td>
<td>1766 (± 33)</td>
<td>1.84 (± 0.16)</td>
<td>147 (± 11)</td>
</tr>
</tbody>
</table>

3.4.6 Structural model of YmPh and M1

A homology model of YmPh was generated through identification of possible targets by running 3 PSI-BLAST iterations to extract a position specific scoring matrix (PSSM) from UniRef90, and then searching the PDB for a match (i.e. hits with an E-value below the homology modeling cutoff 0.5). Finally, YASARA combined the best parts of the 25 models to obtain a hybrid model, in order to increase the accuracy beyond each of the contributors. In Figure 19 YmPh wildtype and mutant M1 are illustrated. Amino acid were substituted using FoldX algorythm. The identified substitutions in variant M1 are not located close to the active center, but on surface exposed regions of the enzyme (Ala166Thr, His433Arg).

![Figure 19: A structural homology model of YmPh and M1 generated with YASARA software.](image)

The amino acid substitutions in YmPh and M1 are shown in space-filling model. Active site loop (S42-T47) is highlighted in red color and catalytically important residues involved in substrate binding (R37, R41, E241, and D327) are...
shown in green color in the active site that is depicted by molecular surface. a): Front view on YmPh enzyme. b): Back view on the YmPh including amino acids that are substituted in M1. c): back view of mutant M1. M1 forms one additional hydrogen bond (indicated by arrows) between Arg433 and Pro430 compared to the wildtype enzyme.
3.5 Discussion

The developed technology platform for flow cytometer-based sorting achieves a throughput of 1.8 x 10^7 events per hour, leading to approximately 10 million screened events in a single experiment. Fur-Shell uses a new principle for fluorescence sorting of E. coli cells that express active enzyme variants. The detection principle employs a coupled enzymatic reaction in which the active variants generate β-D-glucose which is subsequently converted to δ-gluconolactone and thereby producing hydrogen peroxide which initiates hydrogel formation (PEG-diacylate polymerization) on the E. coli surface with co-polymerized fluorescent Polyfluor 570. In Figure 20 the workflow of the flow cytometer-based sorting of E. coli cells of the herein developed method called “Fur-Shell” is illustrated.

**Figure 20: Flow cytometer based sorting principle of the Fur-Shell screening technology employing a YmPh/GOx coupled reaction.** Cells expressing active YmPh release β-D-glucose and generate in the coupled GOx reaction hydrogen peroxide. Hydrogen peroxide initiates polymerization through Fenton’s reaction and finally builds up a fluorescent Polyfluor 570 containing PEG-hydrogel around the E. coli cells expressing active YmPh variants. Sorting of fluorescent E. coli cells is performed at λ_ex 561 nm and λ_em 585 nm. (Figure was reprinted with permission of Elsevier)

In order to determine the optimal time point for fluorescent hydrogel formation, the highest ratio in activity for E. coli BL21 lacIQ1/pALXtreme 5b cells expressing YmPh and harboring
an empty vector was identified (Figure 10). A basal 4-MUP activity was detected for cells harboring an empty vector, indicating the presence of additional phosphatases or phytases in *E. coli*. Indeed, a quite active phytase possessing the highest specific activity of any phytase characterized is present in the genome of *E. coli* [106]. Regardless of the background activity, the excellent signal to noise ratio of cultures with and without YmPh allows efficient separation of non-active variants from active ones (see Figure 14) and yield an enrichment of for instance a population comprising 10 % active clones five times. Validation of the novel screening method was performed in a directed YmPh evolution experiment with a subsequent rescreening of 6 variants in 96-well microtiter plates employing 4-MUP and phytate screening systems [112, 113]. The YmPh library which was generated by epPCR was enriched from 40 % active variants to 89 % active variants after one round of sorting ($10^7$ variants). In the microtiter-plate-based rescreen, two variants M1 and M2 were identified and kinetically characterized. The best variant M1 exhibits a 97 U/mg higher specific activity for screening substrate 4-MUP and interestingly also the natural substrate phytic acid was converted faster (increase by 88 U/mg) at equal conditions. The latter represents a remarkable increase in specific activities in a single round of evolution [115]. Additionally the pH range in which the enzyme is active was broadened for the two substrates (Figure 17), indicating a higher stability at both lower (around pH 3.7) and higher pH values (above pH 5.3). This higher stability can be attributed to the amino acid substitutions (Ala166Thr, His433Arg) due to their surface exposure. In a FoldX interaction comparison between dimers of different orientation between wildtype and M1 a 1.5 kcal/mol lower energy for M1 was determined. In a recent publication Wang et al. identified position Gln349 in *Escherichia coli* phytase appA being responsible for a higher thermostability [116]. Position 349 in appA is located in a structural alignment with YmPh (Align objects with MUSTANG in YASARA) in close proximity to position Gly335 in YmPh. Although thermal stability of YmPh mutant M1 was not tested, it was expected that position Gly335 is beneficial in promoting the enzymatic properties of M1.

A crucial factor in the implementation of the screening platform is the recovery of cells after flow cytometry sorting by streaking on agar plates. Only cells surviving the treatment with reactive monomers and hydrogen peroxide can be recovered. Although improved variants were isolated, this approach could eventually hinder the recovery of the most active variants due to higher production of hydrogen peroxide or a more distinct hydrogel shell, compared to cells with lower activities.

Despite intensive research efforts only a few screening systems have successfully been used in directed evolution experiments in the last five years [63, 117-119]. Main challenges in compartmentalization systems are to achieve the stability of for instance double emulsions and to prevent diffusion of products into the surrounding media. So far compartments were
made of polymersomes, liposomes or water in oil in water emulsions [120-122]. All three technologies share a common strategy, as the generation of a diffusion barrier to separate the inner reaction, either by di- or triblock copolymers, lipids or surfactant-stabilized, water-immiscible oil phases. These systems rely on the monodispersity of the formed reaction containers in order to maintain reproducibility. The controlled formation, incubation and analysis of such reaction containers is technically challenging and time consuming in optimization especially when minimizing product diffusion.

The Fur-Shell technology omits the use of confinement of cells in liposomes or polymersomes and copolymerizes a fluorophore on the surface of *E. coli* in a reaction time of \(~1\) h (including preparation and washing steps) and its formation was supported by SFM images showing a significant difference in density of the polymer shell on the outer membrane of *E. coli* cells which harbor active YmPh variants. The incorporation of small molecules (Polyfluor 570) into the hydrogel structure ensures a stable fluorescence signal without leakage [101].

Flow cytometry-based screening systems are based on single cells and exhibit often significant differences in catalytic performance depending on the metabolic states of cells [123]. We therefore see the main application of the Fur-Shell technology, as all other whole cell-based flow cytometer screening system, as prescreening systems which enables researchers to sort out in their directed evolution experiments the most active variants from huge populations and to explore for instance novel mutagenesis strategies with high mutational loads. In addition, the enrichment in the active population reduces the overall screening effort and increases the chances to find improved variants in case the screening does not cover the generated diversity.
3.6 Materials and methods

Chemicals were of analytical grade or higher quality and purchased from Carl Roth GmbH (Karlsruhe, Germany), Sigma-Aldrich (Hamburg, Germany) and AppliChem (Darmstadt, Germany). Enzymes were purchased from New England Biolabs (Beverly, MA, USA). Oligonucleotides were purchased from Eurofins MWG Operon (Ebersberg, Germany) in salt-free form. Methacryloxyethylthiocarbamoyl-rhodamine B (Polyfluor 570) was purchased from Polysciences Inc. (Warrington, PA, USA).

3.6.1 Plasmids and strains

The Yersinia mollaretii phytase (YmPh) gene (Genebank accession number JF911533.1) was subcloned into pALXtreme-5b [124], resulting in the construct named pALXtreme-5b YmPh. Plasmid constructs and libraries were transformed into E. coli XL10 ultra-competent cells (Agilent Technologies, Palo Alto, CA, USA). YmPh expression and production was carried out in E. coli BL21 (DE3) lacIq cells.

3.6.2 Protein expression and cell labeling

E. coli BL21 (DE3) lacIq cells containing pALXtreme-5b YmPh or a mutant library were cultivated in ZYM-505 non-inducing media (10 ml, 8 h, 37 °C, 250 rpm). The main culture (50 ml ZYM-5052 auto-inducing medium) was inoculated with 500 µl of a preculture and cells were grown (16 h, 37 °C, 250 rpm) [125]. Subsequently, cells were harvested by centrifugation and washed three times with sodium acetate buffer (250 mM sodium acetate, 1 mM calcium chloride, 100 µl Tween-20, pH 5.5) and diluted to a cell density of 1 (OD600). Polymerisation reactions were initiated by mixing cell suspensions in sodium acetate buffer (2.4 ml, 250 mM), glucose-6-phosphate (0.5 ml, 20 mM), GOx Type II (5 µl, 1.58 µM), PEG575-DA (24.8 % w/w), 1-vinyl-2-pyrrolidone (2.5 % w/w), Polyfluor 570 (5 µl, 0.005 % w/w) and Fe2SO4*7H2O (5 µl, 25 µM). Polymerization reaction proceeded for 16 min and was terminated by centrifuging the cell suspension (18,350 × g, 10 min, Eppendorf Centrifuge 5810 R) and three times washing with phosphate buffer (PBS, 12mM total phosphate, pH 7.4) to ensure removal of monomers.

3.6.3 2,2'-azino-bis(3-ethylbenzothioazioline-6-sulphonic acid) (ABTS) screening system in 96-well microtiter plates

Optimization of reaction conditions was carried out with the ABTS assay in 96-well microtiter plates to follow hydrogen peroxide production. 140 µl E. coli BL21-Gold (DE3) cells were adjusted to an optical density of 2 (OD600) in sodium acetate buffer (pH 4-7, 250 mM). Addition of glucose-6-phosphate (1-500 mM), GOx Type II (0.4-3 µM), Horseradish peroxidase (0.5 U/ml), ABTS (50 mM) led to an increase in absorbance at 414 nm. The total
volume of the assay was 200 µl. ABTS absorbance at $\lambda_{\text{abs}}$ 405 nm was measured at room temperature using a Tecan Infinite M1000 microtiter plate reader (Tecan Group AG, Männedorf, Switzerland).

3.6.4 Flow cytometry analyses and cell sorting

Cell analyses and sorting was performed using a BD Influx cell sorter (BD Biosciences, San Jose, CA). The BD Influx flow cytometer was equipped with a 100 µm nozzle and the sheath fluid phosphate buffered saline (PBS) (1.05 mM KH$_2$PO$_4$, 3 mM Na$_2$HPO$_4$, 155 mM NaCl) was used as sheath fluid. A sample of Polyfluor 570 labeled cells was diluted 10 times in PBS and analyzed according to forward and side scatter as well as to its fluorescence intensity ($\lambda_{\text{ex}}$ 561 nm/$\lambda_{\text{em}}$ 585 nm). Cells were sorted at a rate of 5000 events/s allowing a throughput of 1.8 $10^7$ events/h. Sorted cells were collected in PBS and aliquots of 100 µl were plated onto LB$_{amp}$ plates.

3.6.1 Confocal Microscopy

Confocal microscopy (Leica TCS SP8, Leica Microsystems) was performed using 63x oil immersion objective. A continuous wave laser (DPSS, 20 mW: 561 nm) was used for excitation and a highly sensitive prism spectral detector filter was used for emission (570 nm). Gain was set according to a minimal signal for the strains harboring an empty vector to 530.

3.6.2 Scanning force microscopy (SFM)

Untreated cells (negative control) or labeled YmPh (+) and YmPh (-) cell samples were washed 5 times with demineralized water to remove buffer and avoid formation of salt crystals. 1 µl of each cell sample was dried on silicon wafers (Sigma Aldrich) and surface images were recorded by SFM, using an intermittent mode (Bruker Dimension Icon with OTESPA tips with a spring constant of 12-103 Nm$^{-1}$ and a resonant frequency of 278-357 kHz).

3.6.3 Diversity generation and YmPh library construction

A mutant library of Yersinia mollaretii phytase (YmPh) was generated by epPCR. In detail, MnCl$_2$ (0.2 mM) was added to a PCR containing phosphorothioated primers (40 pmol), taq-polymerase (5 U), dNTP-mix (0.2 mM each), the DNA template (30 ng/µl) in TE buffer (total volume of 50 µl) (Table 4). PCR was carried out as follows: initial denaturation 120 s at 94 °C, 30 cycles of: [denaturation 30 s at 94 °C, annealing 30 s at 57 °C, elongation 90 s at 68 °C], followed by a final elongation step, 10 min at 68 °C.
Table 4: Primers for epPCR (small letters indicates phosphorothioate nucleotides)

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence 5’-3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLIC_YmPh_FW</td>
<td>ctttaagaaggaGATATACATATGCGATTAACTG</td>
</tr>
<tr>
<td>PLIC_YmPh_RV</td>
<td>ggctttgttagcAGCCGGATC</td>
</tr>
</tbody>
</table>

The amplification of the vector backbone was performed similarly as for the insert by using the following primers (40 pmol) and adjusting the elongation time to 120s at 68°C (Table 5).

Table 5: Primers for epPCR (small letters indicates phosphorothioate nucleotides)

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence 5’-3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLIC_pALX_FW</td>
<td>gctaaacaagccCGAAAGGAAGCT</td>
</tr>
<tr>
<td>PLIC_pALX_RV</td>
<td>tccttcttaagTTAAACAAAATTATTCTAGAGGGG</td>
</tr>
</tbody>
</table>

PCR products were purified using NuceloSpin® Gel and PCR Clean-up kit according to the manufacturer’s protocol (Macherey-Nagel GmbH & Co. KG, Düren, Germany) and eluted in demineralized water. Respective PCR constructs were subsequently assembled by PLICing as previously reported [124] and transformed into *E. coli* XL10 ultra-competent cells (Agilent Technologies, Palo Alto, CA, USA). Transformation efficiency was determined to be 50000 CFU/µg DNA and transformation was carried out two times in order to further increase the total number of cells harboring a *YmPh* mutant. Subsequently, the transformed YmPh-library was grown in a liquid LB<sub>amp</sub> culture and plasmid DNA was harvested through plasmid preparation, by using NuceloSpin<sup>®</sup> Plasmid kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany), according to the manufacturer’s protocol. The isolated YmPh-Plasmid mutant library was transformed into competent *E. coli* BL21 lacI<sup>D1</sup> cells for YmPh expression.

### 3.6.4 Protein expression in 96-well microtiter plates

*E. coli* colonies grown on LB<sub>amp</sub> plates after flow cytometry-sorting were transferred using toothpicks into 96-well microtiter plates (Flat bottom, polystyrene plates; Corning GmbH, Kaiserslautern, Germany), containing a non-inducing ZYM-505 media (150 µl) supplemented with ampicillin (100 µg/ml). After overnight cultivation in a microtiter plate shaker (16 h, Multitron II, Infors GmbH, Einsbach, Germany; 37 °C, 900 rpm, 70 % relative humidity), each plate was replicated into a second one containing ZYM-5052 auto-induction medium (150 µl; supplemented with 100 µg/ml ampicillin). The first set of 96-well microtiter plates was stored at −80 °C after addition of glycerol (100 µl; 50 %) and the second set of plates was incubated (Multitron II, Infors GmbH, Einsbach, Germany; 37 °C, 900 rpm, 70 % relative humidity; 10 h) for auto-induced expression of YmPh variants and subsequent screening.
3.6.5 Screening in 96-well microtiter plates

96-well microtiter plates containing sorted cells were centrifuged (3220 g, 4 °C, 20 min, Eppendorf Centrifuge 5810 R, Hamburg, Germany) and cell pellets were stored at -80 °C overnight. Cell pellets were thawed and lysed by the addition of lysozyme (100 μl, 2 mg/ml, 50 mM Tris-HCl, pH 7.4) followed by incubation (37 °C, 1 h). Clear supernatant was obtained by centrifugation (4 °C, 3,220 g, 20 min) and an aliquot from each well (60 μl) was transferred into a corresponding one in a 96-well microtiter plate containing sodium acetate buffer (140 μl, 250 mM sodium acetate; 1 mM CaCl₂, 100 μl Tween-20, pH 5.5). Subsequently, the diluted cell lysates (50 μl) were transferred into black 96-well microtiter plates (Flat bottom, polystyrene plates; Greiner bio-one, Frickenhausen, Germany). Activity determination in 96-well microtiter plate format was subsequently performed by supplementing 4-MUP in a substrate solution (50 μl per well; 1 mM 4-MUP, 250 mM sodium acetate, pH 5.5, 1 mM CaCl₂, 100 μl Tween-20). The YmPh activity was determined as increase in the relative fluorescence at the excitation/emission maxima (360 nm/465 nm) using a Tecan Infinite M1000 microtiter plate reader (Tecan Group AG, Männedorf, Switzerland). Three YmPh wild type and three empty vector controls were included in each microtiter plate. Relative activity for each clone was defined by calculating the increase of fluorescence over time of each variant (slope) compared to the average slope (in triplicate) of wild type.

3.6.6 Expression and purification of YmPh and variant M1

YmPh purification was carried out at room temperature using an ÄKTAprime plus system (GE Healthcare Europe GmbH, Munich, Germany) and a preparative liquid chromatography column (internal diameter 15 mm, length 125 mm, maximum pressure 70 bars; YMC Europe GmbH, Dinslaken, Germany). A pre-culture of cells expressing YmPh or M1 were grown (10 ml ZYM-505 in 50 ml Erlenmeyer flasks; 37 °C, 8 h; 250 rpm; 100 μg/mL ampicillin) and used for inoculation (1:100 dilution) of a production culture (200 ml ZYM-5052 in 2 l Erlenmeyer flask; 37 °C, 16 h; 250 rpm; 100 μg/mL ampicillin). Respective cultures of E. coli cells expressing wild type YmPh or variant M1 were harvested by centrifugation (2820 g, 15 min, 4 °C; Sorvall RC-6 Plus, Thermo Scientific) and resuspended in the running buffer (20 mL, 25 mM sodium acetate, pH 5.5, 0.1 mM CaCl₂, 10 μl Tween-20). The resuspended cell culture was pre-homogenized (50 ml; on ice) by sonication (Sonics Vibra-Cell, Sonics & Materials Newton, USA; 30 s; at 40 % amplitude with 30 cycles). Cell lysis was carried out using a high pressure homogenizer (~1,300 bar, 3 cycles; EmulsiFlex C3, Avestin Europe GmbH, Mannheim, Germany). Cell debris was pelleted by centrifugation (18,350 × g, 60 min, 4 °C, Eppendorf Centrifuge 5810 R) and the supernatant was further cleared by filtration through a low-protein-binding filter (0.45 μm; Minisart RC 25 disposable syringe filter;
Sartorius, Hamburg, Germany). Wild type YmPh or M1 variant were purified by cation exchange chromatography, pre-equilibrated with three column volumes of running buffer. The filtered and cleared cell lysates (10 ml) were loaded onto a Toyoperl SP-650C (Tosoh Bioscience GmbH, Stuttgart, Germany) cation-exchange column (bed volume 15 × 100 mm). YmPh or M1 variant were eluted with a NaCl gradient (1 M; 0 % to 60 % in 100 ml, flow rate 3.5 ml/min) in the running buffer. The YmPh protein fractions obtained by ion-exchange chromatography were desalted twice by dialysis using a Spectra/Por dialysis membrane (Spectrum Lab Inc., CA, USA) against sodium acetate buffer (1000 ml, 250 mM sodium acetate, pH 5.5, 1 mM CaCl₂, 100 µl Tween-20) and stored at 4 °C. After the final purification step, the concentration of YmPh and M1 were determined using the Thermo Scientific Pierce BCA protein assay kit (Rockford, IL, USA). The protein content of purified YmPh and M1 samples reached >85 % and kinetic characterization were performed using these samples as previously described [113].

For determination of pH profile and optimum, purified YmPh and M1 samples were diluted to 42 ng/ml and activity determination was carried out using sodium phytate (1 mM) or 4-MUP (1 mM) as substrates.

3.6.7 Homology model generation with YASARA

Since the target sequence of the YmPh wildtype enzyme was the only available information, possible templates were identified by running 3 PSI-BLAST iterations to extract a position specific scoring matrix (PSSM) from UniRef90, and then searching the PDB for a match (i.e. hits with an E-value below the homology modeling cutoff 0.5). Finally, YASARA tried to combine the best parts of the 25 models to obtain a hybrid model, aiming to increase the accuracy beyond each of the contributors.
4 Toolbox for Detecting Hydrolytic Activity by Flow Cytometry-Based HTS

4.1 Declaration

Parts of this chapter have been published in the journal Chemical Communications and are adapted to this thesis with permission of The Royal Society of Chemistry


*shared co-authorship

4.2 Objective

The aim of this part was to adapt and expand the in part II developed Fur-Shell screening platform to other enzymes converting glucose derivatives into glucose. In detail, it has been tested if the lipase *Bacillus subtilis* lipase A (BSLA), the \( p \)-nitrobenzyl esterase from *Bacillus licheniformis* (\( p \)NBEBL) and the cellulase CelA2 are possible target enzymes for the conversion of \( \beta \)-D-glucose-pentaacetate and cellobiose into \( \beta \)-D-glucose. The implementation of additional enzymes expands the Fur-Shell technology into a universal screening platform for hydrolytic enzymes.

An overview about the projected reaction including substrates and enzymes is shown in Figure 21.

![Figure 21: Reaction overview with different enzymes (BSLA, CelA2, pNBEBL) leading towards fluorescent hydrogel formation around *E. coli* cells.](image)

The experiment of the esterase \( p \)NPBEGL is in detail described in the thesis of Nina Lülsdorf. A: glucose derivative is converted by the respective enzyme into \( \beta \)-D-glucose.
Supplemented glucose oxidase (GOx) converts glucose into glucono-δ-lactone and hydrogen peroxide. B: Through Fenton reaction radicals are formed, initiating free radical polymerization of monomers around *E. coli* cells that express active enzyme.

### 4.3 Introduction

In chapter 3, a proof of concept for a novel screening principle named Fur-Shell which is based on fluorescent hydrogel formation around *E. coli* cells with an YmPh phytase as an example was presented. Significantly, the Fur-Shell technology overcomes technical limitations in flow cytometry-based screening systems in terms of compartmentalization and leakage of a fluorogenic substrate and/or fluorescent product. In this chapter, an advancement of the Fur-Shell technology into a general high-throughput screening toolbox for directed evolution of hydrolases is presented. Validation of the proposed system is performed by screening three hydrolases: a *p*-nitrobenzyl esterase from *Bacillus licheniformis* (pNBEBL) [126], a *Bacillus subtilis* lipase A (BSLA) [127], and a cellulase (CelA2) isolated from a metagenome library by Streit *et al.* [128]. Validation was performed for each of the three hydrolases through a single round of directed evolution by screening an epPCR random mutagenesis library. All three hydrolases address enzymes which are of significant synthetic and/or industrial importance (*e.g.* lipases are used in laundry detergents, in the synthesis of pharmaceuticals, and in food processing [129]; cellulases are applied in *e.g.* depolymerisation of cellulose and in food industry [130-132]).

The work for this project was carried out in cooperation with Nina Lülsdorf; the detailed results from pNBEBL are reported in her thesis. Figure 22 shows the principle of the Fur-Shell technology in three steps. In **Step 1** the library is generated by epPCR, insert and vector fragment are cloned by PLICing and the mutant library is subsequently transformed and expressed in *E. coli* BL21-Gold (DE3) cells [124]. **Step 2** comprises the Fur-Shell technology, which relies on the conversion of a substrate (glucose-pentaacetate for lipase; cellobiose for cellulase,) into β-D-glucose. Subsequently, in a glucose oxidase coupled reaction hydrogen peroxide and glucono-δ-lactone are produced. Through Fenton reaction, radical species from hydrogen peroxide are generated, initiating a PEG-based co-polymerization of the fluorescent Polyfluor 570 monomer [133]. *E. coli* cells expressing active enzyme variants are surrounded by a fluorescent hydrogel shell and can in **Step 3** be analyzed and sorted by flow cytometer at rates of around 5000 events per second. The sorted *E. coli* cells expressing active hydrolase variants are plated on agar plates and afterwards transferred for screening in microtiter plates (MTPs) (**Step 4a**). Alternatively as reported here in **Step 4b**, performance parameters such as low cell survival rate of *E. coli* cells (<8 %) were optionally addressed by introducing a plasmid isolation step of sorted cells to rescue mutated genes that encoded improved hydrolase variants in non-viable *E. coli* cells [133].
Figure 22: Flow cytometer-based sorting principle in four steps of the Fur-Shell toolbox for hydrolases. 

**Step 1:** Library generation by epPCR and subsequent cloning by PLICing. **Step 2:** *E. coli* BL21-Gold (DE3) cells producing enzyme variants are incubated with substrate, glucose oxidase and fluorescent labelled monomers (Polyfluor 570). Fluorescent hydrogel is formed around cells expressing active enzyme variants (+) which allows in **Step 3** analysis and sorting (enrichment) by flow cytometer at a rate of about 5000 events per second. Finally the *E. coli* BL21-Gold (DE3) cells are either plated on LB-agar plates (**Step 4a**), or used for plasmid isolation and transformation (**Step 4b**), in both cases followed by transferring clones into MTPs. 

Isolated plasmids are subsequently transformed into competent *E. coli* BL21-Gold (DE3) cells and most beneficial variants were identified after screening of hydrolase clones in 96-well MTP format.

**4.4 Results**

Enzyme initiated hydrogel formation was shown in chapter 3.4.2 for a phytase (YmPh) in combination with a glucose oxidase. In this chapter fluorescent hydrogel formation for *E. coli*
BL21 expressing CelA2 and BSLA is confirmed through confocal microscope images as well as through flow cytometer.

### 4.4.1 Detection of Fur-Shell labelled cells with confocal microscopy

Confocal microscopy images showing an overlay of transmission and fluorescence were recorded, pointing out the difference in the fluorescent signal of *E. coli* BL21-Gold (DE3) cells producing either active hydrolases (pET22b(+) -BSLA, or -CelA2) or cells harboring empty pET22b(+) (Figure 23A: lipase BSLA; C: cellulase CelA2 ). For *E. coli* BL21-Gold (DE3) cells expressing active hydrolase variants (Figure 23B: lipase BSLA; D: cellulase CelA2) a strong fluorescent signal was detected which confirms hydrogel formation through incorporation of Polyfluor 570. *E. coli* BL21-Gold (DE3) cells harboring empty pET22b(+) (Figure 23 A, C) showed little to no fluorescence.

![Overlay of fluorescence and transmission images which were recorded by confocal microscope analysis.](image)

**Figure 23:** Overlay of fluorescence and transmission images which were recorded by confocal microscope analysis. A/C: shows the empty vector signal of *E. coli* BL21-Gold (DE3) pET22b(+) incubated with the substrate glucose pentaacetate or cellobiose. B: *E. coli* BL21-Gold (DE3) pET22b(+)-BSLA were incubated with the substrate glucose pentaacetate. D: *E. coli* BL21-Gold (DE3) pET22b(+)-CelA2 was incubated with the substrate cellobiose.
4.4.2 Flow cytometer analyses of BSLA and CelA2 libraries

Flow cytometry analyses of Fur-Shell labelled *E. coli* BL21-Gold (DE3) cells confirm the visual impression of recorded confocal microscopy images (Figure 24). Comparison of Figure 24 (A-C, B-D) depicts a significant difference in fluorescence intensity among *E. coli* cells expressing active hydrolases and those harboring an empty vector. The fluorescent intensity of *E. coli* cells expressing wildtype hydrolases was around 20-fold higher compared to *E. coli* cells harboring an empty vector (Figure 24 C: lipase BSLA; D: cellulase CelA2). One epPCR random mutagenesis library per hydrolase gene (lipase *bsla* and cellulase *cela2*) was generated using 0.2 mM MnCl₂ with an average mutation frequency of 13.2 (*bsla*, 0.2 mM MnCl₂) and 4.2 (*cela2*, 0.2 mM MnCl₂) mutations per kb. The populations of the two mutant libraries were analyzed by flow cytometer and showed a reduced fluorescent signal compared to wildtype due to the presence of inactive hydrolase variants (Figure 24 E: lipase BSLA, F: cellulase CelA2). The sorting gate P1 was set to minimize background (0.9%, see Figure 24 A, B) from *E. coli* BL21-Gold (DE3) cells harboring an empty vector (Y-axis). Additionally, P1 was adjusted using calibration beads with the forward scatter laser (X-axis) to contain only single cells being smaller than 3 µm in size in order to exclude associated cells [134].

![Figure 24: Flow cytometry analysis of Fur-Shell labelled *E. coli* cells.](image)

Density plots were used to indicate the fluorescent signals generated through *E. coli* cells coated with fluorescent hydrogel shells (recorded by forward scatter (FSC) versus fluorescence signal (λ<sub>ex</sub> 561 nm / λ<sub>em</sub> 585 nm)). Gate (P1) was set to sort all events with a size <3 µm in order to minimize sorting of associated cells [134]. *E. coli* BL21-Gold (DE3) cells were incubated with glucose pentaacetate (BSLA) or cellobiose (CelA2) as substrate. A/B: *E. coli* BL21-Gold (DE3) pET22b(+)
cells were used as a negative control to determine the threshold for fluorescence. C/D: *E. coli* BL21-Gold (DE3) pET22b(+)–BSLA and -CelA2 wildtype cells were used as a positive control. Random mutagenesis libraries were generated with epPCR: E/F: 0.2 mM MnCl₂ for BSLA, and 0.2 mM MnCl₂ for CelA2.

*E. coli* BL21-Gold (DE3) cells expressing active hydrolase variants were sorted and collected (Figure 22, Step 3). DNA of 5*10⁵ sorted *E. coli* cells was recovered by using option 4b (Figure 22, Step 4b). This option includes a plasmid isolation step to maintain the genetic diversity, because cells exhibiting higher activity than others might become more damaged due to a thicker hydrogel shell or a higher local hydrogen peroxide concentration. Therefore, isolated plasmids were re-transformed for expression into *E. coli* BL21-Gold (DE3) cells and plated on LB-agar. Variants of the two hydrolases (in total 2160 clones) were transferred into 96-well MTPs for subsequent screening. In case of CelA2 the substrate 4-MUC that chemically is similar to cellobiose was employed, due to an established screening procedure in 96-well MTP format [135]; and in case of the BSLA *p*-nitrophenyl acetate (*p*NPA) [136] was used as substrate because of a good sensitivity of the *p*NPA assay.

Enrichment factors were determined by activity measurements and calculated by taking ratio of active to inactive populations (active variants after sorting divided by active variants before flow cytometry sorting). Enrichment of the active population above 20 to 30 % (better 50 %) is required to minimize screening efforts in MTP formats. In detail, after one round of sorting CelA2 library was enriched from 3.5 % active population to 38 % (11.7-fold enrichment); BSLA library was enriched 1.5-fold from 29 % to 44 % active population (Table 6). A decrease reaction time for hydrogel formation from 10 min to 2 min, as well as reduction of the monomer concentration from 24.8 % w/w to 12.4 % w/w (PEG₅₇₅-DA) and 2.5 % w/w to 1.25 % (1-vinyl-2-pyrrolidone) did not yield a further improvement in enriching the library. In all cases a sufficiently high population of active hydrolase variants was obtained for 96-well MTP screening.

Table 6 Enrichment of the random mutagenesis libraries were calculated by using the respective MTP screening assay.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Before sorting [%]</th>
<th>After sorting [%]</th>
<th>Enrichment [fold]</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSLA</td>
<td>29</td>
<td>44</td>
<td>1.5</td>
</tr>
<tr>
<td>CelA2</td>
<td>3.5</td>
<td>38</td>
<td>11.7</td>
</tr>
</tbody>
</table>

Remarkably, improved variants were found in MTP format by screening of 900 BSLA variants with *p*NPA and 900 variants with 4-MUC. Two variants with significantly increased activities were chosen for rescreening and the reported values are the average of six measurements and error bars represent the standard deviation from the mean value (Figure 25).
4.4.3 Characterization of improved variants

An fast protein liquid chromatography (FPLC) protocol using ion exchange chromatography was established for purification of BSLA. The best two lipase A variants and BSLA wildtype were overexpressed in *E. coli* BL21-Gold (DE3) in shake flasks and purified by Äkta (cation exchange chromatography, Toyopearl SP-650C column (Tosoh Bioscience)). Adsorbed enzyme was eluted with 1 M NaCl in HEPES buffer (10 mM, pH 7.4). Collected fractions of eluent were assayed for BSLA activity and further analyzed by SDS PAGE to confirm the correct size of the target protein (Figure 26). A major peak eluting at NaCl concentrations between 5 % and 15 % showed highest activities for pNPB. SDS PAGE analysis confirmed that the first fractions within this peak can be assigned to lipase A, by exhibiting an intense band at approximately 19 kDa, and only one faint second band at approximately 21 kDa. The four peak fractions with highest activities for lipase A were pooled for subsequent determination of protein concentration and purity estimation (Table 7). Purity of all lipase A samples determined by automated electrophoresis was higher than 92 %.
Figure 26: Äkta profile of BSLA variant L1. Within the elution peak around 150 ml, lipolytic activity was detected. The blue bars indicate the samples that were showing the highest activity in the pNPB screening system. These fractions were pooled for further characterization. SDS-PAGE on the right shows the purity of the eluted fractions 5-9 with one major band around 19 kDa matching to the size of BSLA.

In order to perform a kinetic characterization of the two best CelA2 variants, another purification protocol was established using Äkta (FPLC) with a HiPrep 16/60 Sephacryl S-400 HR (GE Healthcare Europe GmbH) gel filtration column. The size exclusion method was chosen to semi purify CelA2, because the protein of choice can be isolated in a fast and simple way. Purification of CelA2 using affinity purification would have required a cloning step due to pET22b(+) plasmid lacking tags for affinity purification. Another option of purification using ion exchange chromatography remained not successful in preliminary experiments (internal communication with Dr. Christian Lehmann). In Figure 26 the chromatogram of CelA2 variant C1 is shown exemplary for all cellulase purifications. Fractions 14 – 18 exhibited 4-MUC activity, indicating the cellulase being present in these samples. A SDS gel loaded with a sample of the fractions exhibiting 4-MUC activity confirmed the presence of two strong bands at the expected size of 69 kDa. Samples of fractions 16 and 17 were pooled and used for further experiments.
Figure 27: FPLC profile of CelA2 variant C1. Within the elution peak around 90 ml, cellulolytic activity was detected. The blue bars indicate the samples that were showing the highest activity in the 4-MUC screening system. These fractions were pooled for further characterization. SDS-PAGE on the right shows the purity of the eluted fractions 14-18 with two major bands around 69 kDa matching to the size of CelA2.

In order to determine the total protein concentration of each purified sample a BCA assay kit (Life Technologies) was used in a 96-well microtiter plate. Purity of each variant was determined with automated electrophoresis system (Experion, Bio-Rad). The specific concentration of each enzyme variant is shown in (Table 7).

Table 7: Summary of specific enzyme concentrations for BSLA and CelA2. Specific enzyme concentrations were determined according to the concentration (BCA assay) and the purity (Experion automated electrophoresis) of the respective enzyme.

<table>
<thead>
<tr>
<th>Variant</th>
<th>Specific enzyme concentration [µg/ml]</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSLA WT</td>
<td>71.01</td>
</tr>
<tr>
<td>BSLA L1</td>
<td>98.15</td>
</tr>
<tr>
<td>BSLA L2</td>
<td>154.95</td>
</tr>
<tr>
<td>CelA2 WT</td>
<td>29.95</td>
</tr>
<tr>
<td>CelA2 C1</td>
<td>198.77</td>
</tr>
<tr>
<td>CelA2 C2</td>
<td>27.69</td>
</tr>
</tbody>
</table>

Samples of BSLA and CelA2 were diluted to the same protein concentration (71.8 µg/ml and 10 µg/ml) and used for the determination of enzyme kinetics.

4.4.4 Determination of kinetic parameters ($K_M$, $k_{cat}$)

Figure 28 presents Michaelis Menten kinetics for the two purified enzyme variants as well as the corresponding wildtypes (BSLA left, CelA2 right), determined with the respective substrates. A summary of the kinetic data obtained for the two enzymes are summarized in Table 8. The BSLA variants L1 (Arg57Gln, Ser163Gly) and L2 (Tyr139Asp) were kinetically characterized with the colorimetric substrate $p$NPA at 22°C and both showed 1.3-fold
increase in $k_{cat}$ compared to the wildtype BSLA. CelA2 was kinetically characterized with the fluorogenic substrate 4-MUC at 30°C. The cellulase variants C1 (Val37Ala, Glu275Gly, Glu398Val) and C2 (Asn135Ser) both showed an increase in $k_{cat}$ of around 70%.

**Table 8: Characterization of BSLA and CelA2 regarding $K_m$, $k_{cat}$ and specific activities in U/mg.** BSLA was kinetically characterized with the colorimetric substrate $p$NPA (at 22°C, pH 6.5). CelA2 was kinetically characterized with the fluorogenic substrate 4-MUC (at 30°C; pH 7.2). Units are defined as: 1 U of BSLA catalyzes the conversion of 1 µmol $p$NP per minute, 1 U of CelA2 catalyzes the conversion of 1 µmol 4-MU per minute.

<table>
<thead>
<tr>
<th>Enzyme (substitution)</th>
<th>$K_m$ [µM]</th>
<th>$k_{cat}$ [1/s]</th>
<th>U/mg [µmol/min*mg]</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSLA wildtype</td>
<td>1259±92</td>
<td>3.4±0.1</td>
<td>10.7±0.3</td>
</tr>
<tr>
<td>L1 (Arg57Gln, Ser163Gly)</td>
<td>1807±243</td>
<td>4.2±0.2</td>
<td>13.1±0.7</td>
</tr>
<tr>
<td>L2 (Tyr139Asp)</td>
<td>1893±241</td>
<td>4.3±0.2</td>
<td>13.4±0.7</td>
</tr>
<tr>
<td>CelA2 wildtype</td>
<td>170±8</td>
<td>0.15±0.02</td>
<td>0.14±0.02</td>
</tr>
<tr>
<td>C1 (Val37Ala, Glu275Gly, Glu398Val)</td>
<td>189±9</td>
<td>0.26±0.03</td>
<td>0.22±0.03</td>
</tr>
<tr>
<td>C2 (Asn135Ser)</td>
<td>172±9</td>
<td>0.26±0.03</td>
<td>0.23±0.03</td>
</tr>
</tbody>
</table>

**4.4.5 Homology model of BSLA and CelA2**

The crystal structure of BSLA was loaded into YASARA and is illustrated in Figure 29A. For BSLA wildtype enzyme the catalytic triad Ser77, Asp133 and His156 is indicated in green and active site residues are indicated in red. Amino acids from the variants L1 and L2 were substituted with the FoldX plugin and are indicated in blue (Figure 29B and C). The amino
acid substitutions of L1 are both located in \( \alpha \)-helices (Figure 29C). Amino acid substitution Ser163Gly in L1 is in close proximity to the active site residue Tyr161 and is reported to have an effect on thermostability of BSLA [137]. An effect of position 57 in BSLA has not been reported yet. Amino acid substitution Tyr139Asp of L2 is located on the surface in loop of BSLA and the side chain is turned to the outside of the molecule (Figure 29B). An effect of position Tyr139 in variant L2 has been reported for higher activity towards DMSO and this position is a direct neighbor of the active site residue Lys140 [138].

![Figure 29: Yasara models of energy minimized, repaired BSLA and its variants. A: BSLA wildtype enzyme, catalytic triad Ser77, Asp133 and His156 is indicated in green and active site residues are indicated in red. Amino acids from the variants L1 (B) and L2 (C) were substituted with FoldX plugin and are indicated in blue. Variant C1 (Val37Ala, Glu275Gly, Glu398Val) shows a 1.9-fold increase in \( k_{\text{cat}} \) compared to wildtype CelA2. Since a crystal structure of CelA2 is currently not available, a homology model was generated using the “homology modeling” routine in the YASARA software package [139]. In the homology model, CelA2 shares 41 % sequence identity to a Glycosyl Hydrolase Family 9 (GH9) cellobiosidase from \textit{Clostridium cellulovorans}. The homology model of CelA2 is illustrated in Figure 30 for the mutants C1 and C2. Amino acids from the variants C1 and C2 were substituted with the FoldX plugin. Position Val37 is not covered by the homology model. Position Glu275 is located on the surface between two \( \beta \)-sheets, position 398 is located in the middle of a \( \alpha \)-helix, and position 37 was not covered by the homology model which was generated on the basis of a glycosyl hydrolase family 9 (GH9) cellobiosidase from \textit{Clostridium cellulovorans} (PBD ID: 3RX7). None of three substituted positions in CelA2 variant C1 have been reported yet. CelA2 variant C2 shows a 1.9-fold increase in \( k_{\text{cat}} \) compared to wildtype CelA2. Amino acid substitution of C2 at position Glu398 is located at
the molecule surface and oriented to the inside of the molecule. Compared to wildtype CelA2 M1 shows a reduced net charge due to substitution of two negatively charged residues to non-charged residues (Glu275Gly, Glu398Val). Wildtype CelA2 sequence shows 12 % sequence identity with a cellulase isolated from Clostridium thermocellum (CelA), where the substitution Asp278Asn showed 25 % more specific activity compared to wildtype [140]. Although the sequence identity between CelA and CelA2 is low, the increased activity in variant C1 might be related to the substitution at position 275 (in close proximity to position 278 in CelA). An influence of position 135 in variant C2 is not reported in literature yet.

**Figure 30:** Homology model of energy minimized, repaired CelA2 C1 (A) and C2 (b) variants generated with the YASARA software package. Substituted positions are highlighted. Homology model is based on PBD ID 3RX7.
4.5 Discussion

In this chapter the Fur-Shell method was expanded to lipases, cellulases and esterases. Fur-Shell formation was validated for every hydrolase via confocal microscopy and flow cytometer analysis and significant difference in fluorescent signal was observed for cells expressing active hydrolases compared to cells expressing inactive hydrolases.

For fluorescent hydrogel formation a signal sequence, directing the expressed enzyme to the periplasm, seemed to be crucial. The hydrolases BSLA and CelA2, were cloned into pET22b(+) vector harboring a pelB sequence for periplasmic expression [141]. In pET28a(-), a vector lacking the pelB leader sequence, no sufficient signal for differentiating negative from positive cells could be detected. This leads to the assumption, that the whole reaction of hydrogel formation takes place outside or in the periplasm of the cell.

Compared to standard directed enzyme evolution campaigns (low mutagenesis frequency, 1200-15000 clones screened per round; often 1.5-2.5-fold improved activity) the obtained improvements were comparable to one round of evolution in case of BSLA and CelA2 [142]. The number of clones screened in 96-well MTPs (on average ~720) is significantly lower compared to standard directed evolution experiments. Therefore, the principal application of the Fur-Shell technology is a pre-screening system used to isolate cells expressing active enzymes from large cell populations (>10^7 cells) and thereby to minimize screening efforts in a cost-effective manner.

The three hydrolase examples show that the Fur-Shell screening principle has been advanced into a general platform for directed hydrolase evolution by reporting first validated protocols for a lipase, a cellulase, and an esterase. The developed protocols are easy in use and time-efficient when compared to flow cytometry based screening systems in directed evolution. Compared to emulsion systems challenges such as leakage of fluorogenic substrates and/or fluorescent products and crosstalk of double emulsions systems are solved. Additionally, the problem of synthesizing fluorogenic substrates, which must meet the condition to become fluorescent upon cleavage, is solved by using substrates that are glucose derivatives.

The principle of fluorescent hydrogel formation around whole cells can likely be expanded to alternative fluorescent hydrogels (Fur-Shells) as well as enzyme classes and has from our point of view the potential to establish flow cytometry as standard screening format in directed enzyme evolution.

We have developed a new polymerization-based reaction for flow cytometer screening of hydrolases. To the best of our knowledge, no previously reported flow cytometer screening system is able to generate a fluorescent signal upon conversion of its natural substrates. Cellobiose is naturally cleaved by β-glucosidases and CelA2 wildtype shows particularly
cellobiose activity. All FACS screens are followed up by more accurate secondary screening in MTP, so the development of an assay that provides a distinct signal by FACS as well as in microtiter plates is advantageous.
4.6 Materials and Methods

4.6.1 Cloning into expression vector

*E. coli* strains DH5α and BL21-Gold (DE3) were purchased from Agilent Technologies (Santa Clara, USA).

*B. subtilis* lipase A (*BSLA*)

*BSLA* (GenBank: NP_388152.1), as well as the expression vector pET22b(+) were double digested with *Nde*I and *Xho*I, the purified fragments were ligated with T4 DNA ligase (5 U). The resulted plasmid was transformed with *E. coli* BL21-Gold (DE3) for enzyme expression.

*Cellulase A2* (*CelA2*)

The gene *CelA2* (GenBank: JF826524.1; [135]) was ordered as a synthetic gene with optimized codon usage for *E. coli*.

4.6.2 Diversity generation by random mutagenesis

Three sets of mutant libraries were generated by epPCR using pET22b(+)-pNBEBL pET22b(+)-*BSLA*, and pET28b(+)-*CelA2* as template. The standard PCR was carried out using *Taq* polymerase (2.5 U), dNTP mix (0.2 mM each), DNA template (30 ng/µl) supplemented with different MnCl₂ concentrations (0.0, 0.1, 0.2 and 0.3 mM) (Table 9). Primers (0.4 pmol) were used according to the different templates (Table 10).

Table 9: Error prone PCR (epPCR) program for insert *BSLA* and *CelA2*.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>94</td>
<td>60</td>
<td>1x</td>
</tr>
<tr>
<td>Denaturation</td>
<td>94</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td></td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>BSLA</td>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CelA2</td>
<td>60</td>
<td></td>
<td>25x</td>
</tr>
<tr>
<td>Elongation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSLA</td>
<td>72</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>CelA2</td>
<td>72</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>Final elongation</td>
<td>72</td>
<td>300</td>
<td>1x</td>
</tr>
</tbody>
</table>
Table 10: Primers for epPCR (small letters indicate phosphorothioate nucleotides)

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence 5'-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>F_BSLA</td>
<td>cttgctcagcggagctctcatta</td>
</tr>
<tr>
<td>R_BSLA</td>
<td>cccgcgatggccatggatatc</td>
</tr>
<tr>
<td>F_CelA2</td>
<td>cttgctcagcggagctcgaattcttataa</td>
</tr>
<tr>
<td>R_CelA2</td>
<td>cccgcgatggccatggagcagacacctatgaaaaag</td>
</tr>
</tbody>
</table>

Amplification of the vector backbone was performed by using pET22b(+) as template to subclone all genes into the vector (pET22b(+)) containing a pelB leader sequence. PCR was carried out using PfuS polymerase (Table 11) and respective primers for each construct (Table 12). Resulting PCR products were digested with DpnI (20 U) overnight at 37°C, and subsequently DpnI was inactivated (80°C; 20 min). The PCR products were purified using QIAaquik PCR Purification Kit. Cloning was performed by PLICing [124] and hybridized DNA fragments were transformed with competent E. coli BL21-Gold (DE3).

Table 11: PCR program for vector backbone amplification

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>98</td>
<td>60</td>
<td>1x</td>
</tr>
<tr>
<td>Denaturation</td>
<td>98</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSLA</td>
<td>60</td>
<td>30</td>
<td>25x</td>
</tr>
<tr>
<td>CelA2</td>
<td>60</td>
<td>180</td>
<td></td>
</tr>
<tr>
<td>Elongation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSLA</td>
<td>72</td>
<td>180</td>
<td></td>
</tr>
<tr>
<td>CelA2</td>
<td></td>
<td>180</td>
<td></td>
</tr>
<tr>
<td>Final elongation</td>
<td>72</td>
<td>300</td>
<td>1x</td>
</tr>
</tbody>
</table>

Table 12: Primers for vector backbone amplification (small letters indicates phosphorothioate nucleotides).

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence 5'-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>F_pET22b(+)-BSLA</td>
<td>ggcacgtcggccgtggcaacgcagcgacgcagccagccaggcag</td>
</tr>
<tr>
<td>R_pET22b(+)-BSLA</td>
<td>tccgctcggcagcttgccgccgcactcgagc</td>
</tr>
<tr>
<td>F_pET22b(+)-CelA2</td>
<td>cgacagtgtgtatcatcactcat</td>
</tr>
<tr>
<td>R_pET22b(+)-CelA2</td>
<td>tccgctcggcagcttgccgccg</td>
</tr>
</tbody>
</table>
4.6.3 Hydrogel Formation

For generating hydrogel around *E. coli* BL21-Gold (DE3) cells expressing active enzyme, pET22b(+) plasmids containing a *pelB* leader sequence were obtained. Application of a *pelB* leader sequence leads to a periplasmic expression of the encoded genes. A 10 ml preculture of *E. coli* BL21-Gold (DE3) cells producing either active hydrolases (pET22b(+)-BSLA, or -CelA2) or cells harboring pET22b(+) was grown overnight in 500 ml erlenmeyer flasks (LB media supplemented with 100 µg ampicillin /ml, 37°C, 250 rpm). 1 % of the preculture were used to inoculate a 20 ml production culture in a 500 ml erlenmeyer flask (LB media supplemented with 100 µg Ampicillin/ml, 37°C, 250 rpm). Upon induction with 0.1 mM isopropyl thio-β-D-galactoside (IPTG) at an optical density (OD) of $\lambda_{600 \text{ nm}} = 0.6$, enzyme was expressed (4 h, 30°C, 250 rpm). Cells were harvested by centrifugation (10 min, 4°C, 3220 g, Eppendorf centrifuge 5810 R) and cell pellet was washed two times with 10 ml PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na$_2$HPO$_4$, 2 mM KH$_2$PO$_4$, pH 6.5). Cell density was adjusted to an OD$_{600}$ of 2 using PBS. Hydrogel formation was initiated by using 960 µl cell suspension, 24.8 % Poly(ethylene glycol)$_{575}$ diacrylate (PEG$_{575}$-DA), 2.5 % 1-vinyl-2-pyrrolidone, 20 µl (0.5 % stock concentration) Polyfluor 570, 2 µl (25 µM stock concentration) Fe$_2$SO$_4$·7H$_2$O dissolved in ddH$_2$O, 200 µl (200 mM stock concentration) respective substrate dissolved in PBS (cellobiose) or ddH$_2$O (glucose pentaacetate), 2 µl GOx type II dissolved in PBS (1.58 mM stock concentration) in a total reaction volume of 1.5 ml. Reaction was stopped by three times washing (1 min, 4°C, 3220 g) with 1 ml PBS. Prepared samples were used for confocal microscopy and flow cytometry analyses.

4.6.4 Flow cytometry analyses and cell sorting

Flow cytometry was carried out as described previously [133]. Sorted cells were pelleted by centrifugation (1 min, 11000 rpm). Plasmid was isolated using plasmid isolation kit (Macherey Nagel) and transformed with competent *E. coli* Bl21-Gold (DE3) and plated on LB-agar plates supplemented with 100 mg ampicillin/l.

4.6.5 Protein expression in 96-well microtiter plates (MTP)

Colonies were transferred into 96-well flat bottom microtiter plates (MTP) containing 150 µL LB-medium supplemented with ampicillin (100 µg/ml). After overnight cultivation in a MTP shaker (37°C, 900 rpm, 70 % humidity), 150 µl main culture (LB-medium supplemented with 100 µg/ml ampicillin) were inoculated with 10 µl pre-culture (v-bottom MTP. The plates were cultured for 2.5 h (37°C, 900 rpm, 70 % humidity) until OD$_{600}$ reached 0.6 and protein expression was induced by addition of IPTG (0.1 mM). After expression (4 h, 30°C, 900 rpm, 70 % humidity), plates were centrifuged (15 min, 4°C, 4000 g). Medium supernatant was removed and cell pellets were stored overnight at -20°C. Cell disruption was performed by
resuspending the cell pellet in 150 μL PBS-buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 2 mM KH₂PO₄, pH 7.4) supplemented with lysozyme (1 mg/ml). After incubation (1 h, 37°C, 900 rpm, 70 % humidity), plates were centrifuged (15 min, 4°C, 4000 g). Supernatants of lysates were subsequently used in respective 96-well MTP screening systems.

4.6.6 Screening systems

**Spectrophotometric assay for determining lipolytic activity of BSLA**

The activity of BSLA was determined by using p-nitrophenyl acetate (pNPA). pNPA was dissolved in acetonitrile (stock concentration 100 mM) and subsequently diluted in triethanolamin buffer (TEA, 50 mM, pH 7.4) (stock concentration 1 mM). Reaction was initiated inside 96-well MTPs by mixing 90 μl of pNPA-TEA solution with 10 μl cell lysate containing BSLA. The release of p-nitrophenolate was monitored at 410 nm in a MTP reader (Tecan Sunrise).

**Fluorometric assay for determining cellulolytic activity of CelA2**

Activity of CelA2 was determined by using 4-methylumbelliferyl-β-D-cellobioside (4-MUC, [143, 144]) as described previously [135].

4.6.7 Purification

**Enzyme production and purification**

*E. coli* BL21-Gold (DE3) cells containing the plasmid pET22b(+)-BSLA, pET22(b)-CelA2 or a corresponding variant were grown in shake flasks overnight (10 ml LB media supplemented with 100 μg/ml ampicillin, 37°C, 250 rpm) and used for inoculation of a production culture (100 ml LB media supplemented with 100 μg/ml ampicillin, 37°C, 250 rpm). Enzyme expression was induced by the addition of IPTG (0.1 mM final concentration) when the production culture reached an optical density (OD₆₀₀) of 0.6 and further cultivation was continued (30°C, 250 rpm, 4 h). *E. coli* cells were harvested by centrifugation (4°C, 10000 rpm, 20 min, Sorvall RC-6 Plus, Thermo Scientific) and frozen at -20°C overnight. Cell pellets were resuspended in 20 ml HEPES buffer (10 mM, pH 7.4). After cell disruption by sonification (3 times 30 s with 30 s cooling intervals, EmulsiFlex C3; Avestin Europe GmbH) on ice, cell debris was removed by centrifugation (4°C, 10000 rpm, 20 min) and supernatant was filtered (0.45 μm; Minisart RC 25 disposable syringe filter; Sartorius) and used for purification by cation exchange chromatography (BSLA) or gel filtration (CelA2) with a protein purification system (ÄKTAprime plus, GE Healthcare Europe GmbH Life Sciences):

- 10 ml of BLSA-samples were loaded on a Toyopearl SP-650C column (Tosoh Bioscience) and eluted through an increasing gradient (0 % to 20 %) of sodium chloride solution (1 M).
1.2 ml of CelA2 samples were loaded on a HiPrep 16/60 Sephacryl S-400 HR (GE Healthcare Europe GmbH GE Life Sciences) gel filtration column. Peak fractions were collected and analyzed for enzyme activity and by SDS PAGE. Fractions showing the highest enzyme activity were pooled and stored at 4°C until further use.

**Determination of protein quantity and purity**

Total protein concentration was determined using a BCA Protein Assay Kit (Pierce Chemicals) according to the manufacturer's protocol. Experion Automated Electrophoresis System (Bio-Rad) and the Pro260 Analysis Kit were used for purity determination by following the manufacturer's instructions.

**Determination of enzyme kinetics for BSLA**

Purified enzyme solutions were diluted to a concentration of 71.8 µg/ml. For determination of kinetic parameters, 140 µl of pNPA solution in different concentrations (0 mM – 10 mM, in 1 % acetonitrile, 100 mM potassium phosphate buffer, pH 6.5) was added to 10 µl sample containing purified lipase A in 96-well MTP. The increase of absorption at 410 nm was recorded for 5 min at room temperature using a Tecan Infinite M1000 MTP reader (Tecan Group AG). The maximum linear rate was used to determine enzyme activity. Product formation rates were calculated using a standard curve obtained by measuring absorbance values of a p-nitrophenylate dilution series. The determination of enzyme kinetics was performed with GraphPad Prism 6 software using a Michaelis-Menten kinetic derivation.

**Determination of enzyme kinetics for CelA2**

Purified enzyme solutions were diluted to a concentration of 10 µg/ml. For determination of kinetic parameters, 20 µl 4-MUC substrate solutions of different concentrations (0.0 µM – 1 µM, 40 µl 250 mM potassium phosphate buffer, pH 7.2) was added to 40 µl sample containing purified CelA2 in 96-well MTP. The increase in fluorescence at 450 nm was recorded for 20 min at 30°C using a Tecan Infinite M1000 MTP reader (Tecan Group AG). Product formation rates were calculated using a standard curve obtained by measuring fluorescence values of a 4-methylumbelliferone dilution series. The determination of enzyme kinetics was performed with GraphPad Prism 6 software using a Michaelis-Menten kinetic derivation.
5 Flow Cytometry-Based HTS Platform for Cell Free Expression in Polymersomes

5.1 Objective
The goal of this part is to design a platform for flow cytometer-based ultra-high throughput screening by using in vitro expression of gene libraries encoding for enzyme of interest within polymersomes acting as cell like compartments.

In a first step, optimization of polymer formation will be performed in order to generate polymer vesicles having properties that complement a directed evolution experiment (i.e., no diffusive loss of entrapped components, stability at the wide range of reaction conditions). Next step will involve the encapsulation and expression of a cellulase (celA2) gene inside the polymersomes with subsequent cell free cellulase expression. Active enzyme variants able to convert a fluorogenic substrate label a polymersome compartment fluorescent. Polymersomes will be analyzed and sorted by fluorescence activated cell sorting based on a qualitative discrimination between fluorescent and non-fluorescent populations. In a next step, genes encoding for active enzymes will be recovered and transformed into E. coli for expression. In the final step, quantitative characterization of 2000 variants will be performed in a microtiter plate assay. In order to validate the developed screening platform a directed evolution experiment using in vitro cellulase library synthesis will be performed within polymersomes and screened by flow cytometry with the aim to generate a proof of concept for future directed evolution campaigns carried out in polymersomes in ultra-high throughput.

5.2 Introduction
In many ways the increasing effectiveness of directed evolution has been driven by advancements that are based on new concepts, rather than new technologies. Novel concepts include neutral fitness networks or various advancements related to library quality, such as OmniChange or SeSaM, [20, 21, 145, 146]. Regarding novel technologies, so far, only one enzymatic activity has been reengineered using the strategy of isolating active variants of in vitro expressed enzymes encapsulated in double emulsions by flow cytometer. Mastrobattista and colleagues screened a mutant library of Ebg, an E. coli protein of unknown function, but known to be evolvable for β-galactosidase activity [147].

The main challenge of in vitro compartmentalization as a general flow cytometer-based ultra-high throughput screening platform for directed evolution is the inherent low sensitivity due to low levels of in vitro expressed protein. For enzyme engineering that is performed by IVC, the DNA amount must be adjusted to in average one molecule of DNA per compartment in order to maintain the linkage between one genotype to one phenotype.
Inspired by *in vitro* compartmentalization inside double emulsions performed by Mastrobattista *et al.* a decade ago, we aim in this chapter to adopt the idea of compartmentalizing an *in vitro* expressed mutant library and sort out the active population by flow cytometry. But unlike in the reported approach, encapsulation is performed in polymeric membranes. This strategy is regarded to be more promising than compartmentalization in surfactant-based double emulsions, as polymersomes are able to encapsulate hydrophilic, hydrophobic and amphiphilic molecules such as any other vesicular structure, but their thick membrane provides them with superior stability. By using *in vitro* expression for high throughput screening, the gene encoding the enzyme of interest is entrapped together with a substrate and the cell extract within a compartment that enables the link between genotype and phenotype. Upon transcription and translation, active enzymes convert the substrate into a fluorescent product and the fluorescent droplet can be detected by a flow cytometer [59]. In Figure 31 the strategy of synthesizing an enzyme from a single DNA molecule *in vitro* inside polymersomes is depicted. After *in vitro* expression from a DNA template active enzyme converts the co-encapsulated substrate and generated fluorescent polymersomes can be sorted by flow cytometer in ultra-high throughput.

![Figure 31: Polymersome providing a closed environment for directed evolution of *in vitro* expressed enzymes.](image)

A polymersome contains an *in vitro* expression kit, a plasmid of mutant DNA, and the substrate of the fluorogenic reaction. Upon transcription and translation, active enzymes convert the substrate into a fluorescent product.

Notably, this technology design provides the first opportunity to engineer a cellulase through directed evolution offering an ultra-high throughput screening. In addition, due to the substitution of double emulsions by polymersomes, its beneficial chemical composition widens the possible application more hydrophobic substrates and thereby expands the technology to more than one enzyme class.
5.3 Results

5.3.1 *In vitro* expression of CelA2

CelA2 production in test tubes was carried out using two different kit-based approaches. Both systems use *E. coli* lysate for cell-free protein production. In order to confirm the substrate conversion and to determine the optimal temperature for *in vitro* expression, temperatures of 30°C, 33.5°C, and 37°C were tested with a kit purchased from Roche and a kit purchased from RiNa GmbH according to the respective protocols. Upon reaction initiation through incubation at the particular temperature, resorufin-β-D-cellobioside was added in order to follow enzyme production and to determine the background reaction, resulting from the cell lysate [148]. In Figure 32, the fluorescent signal from conversion of resorufin-β-D-cellobioside to resorufin is shown after 4 h reaction time. Fluorescent signal obtained after enzyme synthesis using Roche kit was overall lower compared to fluorescent signals obtained with RiNa kit. The poor overall performance from Roche kit can be attributed to the absence of nuclease inhibitors (their addition still leads to a lower total production). The optimal temperature for cell free CelA2 synthesis was identified to be 37 °C (31.015 RFU) with a ~5-times lower background reaction (6.554 RFU) using the RiNa kit. Therefore, for all further experiments CelA2 was expressed at 37°C. The Roche kit was excluded in further approaches as the RiNa kit yields the highest fluorescent signal upon conversion of CelA2.

![Figure 32: Endpoint measurement of the fluorescent signal generated through the conversion of resorufin-β-D-cellobioside into resorufin by *in vitro* produced CelA2. The cell-free CelA2 (-) indicates the application of DNA leading to inactive CelA2, whereas (+) indicates the application of DNA leading to active CelA2. Standard deviation is determined from triplicate measurements. Reactions were carried out with two different kits at 30°C, indicated in light grey, at 33.5°C indicated in grey, at 37°C indicated in dark grey.](image-url)

To determine the incubation time for gathering maximal CelA2 activity resorufin-β-D-cellobioside *in vitro* reaction mixture containing celA2 gene and substrate was sampled by
measuring fluorescence signal at different time points (0 h, 1 h, 2 h, 3 h, and 4 h). A doubling in activity can be observed after 1 h, 2 h, and 3 h of expression indicating a constant CelA2 production in the first three hours (Figure 33). The fluorescent signal obtained after 4 h of incubation was comparable to fluorescent signal after 3 h of incubation, indicating a finishing of enzyme synthesis, through the depletion of compounds necessary for in vitro CelA2 production. Ultimately in all further experiments, CelA2 was expressed for 4 h.

Figure 33: Increase in fluorescent signal over time of in vitro expressed cellulase CelA2 sampled at 0 h, 1 h, 2 h, 3 h, and 4 h. Increase in activity was followed through conversion of resorufin-β-D-cellobioside at λ<sub>ex</sub> 571 nm and λ<sub>em</sub> 585 nm.

5.3.2 Dynamic light scattering (DLS)

Polymersomes were generated through overnight stirring of 10 mg triblock copolymer PMOXA-PDMS-PMOXA with intermediate sonication steps, followed by lyophilization to remove internal and external water phase. Lyophilized polymersomes were frozen for storage at -20°C and rehydrated with aqueous solution according to method reported for liposomes [149].

Dynamic light scattering (DLS) analysis of 1.5 mg PMOXA-PDMS-PMOXA polymersomes rehydrated in 1 ml ddH<sub>2</sub>O is shown in Figure 34. Three peak fractions with size-distributions in diameter ranging from 30 to 80 nm, from 100 to 700 nm and from 4000 to 6500 nm confirmed the assembly of triblock copolymers into higher structures. According to intensity values, the main fraction of polymersomes has a size between 200 and 400 nm.
Figure 34: Dynamic light scattering analysis of a PMOXA-PDMS-PMOXA polymersome sample. The sample was analyzed in triplicates and the diameter of the recorded polymersomes is plotted against the measured intensity.

5.3.3 In vitro expression in polymersomes and flow cytometer analysis

Frozen polymersomes were recovered upon rehydration using in vitro expression reaction on ice for 30 min. Afterwards, 1 µl of protease (1 mU/mg) was added to each reaction to degrade non-encapsulated proteins omitting in vitro expression outside the polymersomes. CelA2 was expressed inside polymersomes in batches of 11.5 µl, mixed with 0.5 mg lyophilized PMOXA-PDMS-PMOXA vesicles. The usage of small batches of in vitro expression reaction that can be pooled, in case more sample is needed, compensated for the high cost of the in vitro expression kits. The samples were diluted prior to FACS analysis 30 times, so especially for optimization experiments and for negative and positive controls the usage of one batch was sufficient.

Crucial for the success of in vitro ultra-high throughput screening is the amount of DNA molecules per compartment. More than one DNA molecule impairs the chance to enrich the active variants of a library caused by an ambiguous assignment of one fluorescent signal to more than one DNA molecule (genotype-phenotype linkage). The process of loading DNA into compartments is assumed to happen purely random; the distribution is dictated by Poisson statistics

\[
P(x = k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad (1)
\]

in which P(x = k) is the probability to have k DNA molecules per polymersome, and \( \lambda \) is the mean number of DNA molecules per vesicle. This number was determined by calculating the number of balls of a diameter of 3 µm that can be generated with a volume of 11.5 µl.
According to Poisson distribution shown in Table 13 the number of polymersomes (3 µm) which include more than one DNA molecule is shown for different DNA concentrations.

Table 13: \( \lambda \): Mean number of DNA molecules per polymersome of a size of 3 µm and the probability \( P \) of more than one DNA molecule in a polymersome according to Poisson distribution.

<table>
<thead>
<tr>
<th>DNA concentration</th>
<th>( \lambda )</th>
<th>( P = 1 )</th>
<th>( P &gt; 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ng:</td>
<td>0.24</td>
<td>0.188</td>
<td>0.024</td>
</tr>
<tr>
<td>3 ng:</td>
<td>0.72</td>
<td>0.351</td>
<td>0.162</td>
</tr>
<tr>
<td>5 ng:</td>
<td>1.20</td>
<td>0.361</td>
<td>0.337</td>
</tr>
<tr>
<td>10 ng:</td>
<td>2.39</td>
<td>0.218</td>
<td>0.689</td>
</tr>
<tr>
<td>50 ng:</td>
<td>11.96</td>
<td>0.001</td>
<td>0.999</td>
</tr>
</tbody>
</table>

In vitro expression inside polymersomes was evaluated with different DNA concentrations (1 ng, 5 ng, 10 ng, and 50 ng) in order to determine a correlation of the fluorescent signal as well as the minimal DNA concentration sufficient for background signal discrimination via flow cytometer. The substrate resorufin-\( \beta \)-D-celllobioside was applied to follow the in vitro synthesis of active CelA2 in polymersomes. Gate P1 was set in order to exclude events larger than 3 µm and the signal caused by the background reaction (negative control, respective amount of pIX3.0RMT7/CelA2 Glu580Gln DNA yielding inactive CelA2 upon expression). Figure 35 shows a strong correlation of the fluorescent signal to the amount of deployed DNA. The load of 1 ng DNA into polymersomes led to a population of 2.2 % in gate P1. A further increase to 7.1 %, 13 % and 35.5 % was achieved by loading polymersomes with 5 ng, 10 ng and 50 ng DNA. In conclusion, there is a positive correlation between the DNA amounts loaded with the polymersomes and the intensity of the fluorescent signal. In addition, it was shown that the lowest amount necessary for fluorescent signal detection of CelA2 catalyzed resorufin-\( \beta \)-D-celllobioside conversion via flow cytometry was 1 ng of plasmid DNA.
Figure 35: Density plot of a flow cytometer analysis where forward scatter is plotted against fluorescence intensity. The experiment was carried out with expression of different amounts of DNA in polymersomes. Control reactions performed with the same amount of DNA leading to inactive CelA2 can be seen in the framed boxes. Polymersomes contain 1 ng, 5 ng, 10 ng, or 50 ng of pIX3.0RMT7/CelA2His288Phe DNA or the respective amount of pIX3.0RMT7/CelA2Glu580Gln leading to inactive CelA2. Sorting gate P1 was used to analyze polymersomes loaded with different amounts of DNA.

Based on these promising results a model library, consisting of defined ratios of DNA coding for active and inactive CelA2 was generated (Figure 36). The total DNA concentration of 3 ng ensured that only around 16 % of the polymersomes contained more than one DNA molecule. The genotype-phenotype link was assessed by mixing a model library in a ratio of 1:4 DNA coding for active to inactive CelA2 (25 % active CelA2). The library contained 0.75 ng pIX3.0RMT7/CelA2His288Phe and 2.25 ng pIX3.0RMT7/CelA2Glu580Gln DNA. In correlation to the previous results, a slight increase of fluorescent signal for the model library was recorded compared to the negative control. Although the amount of population in P1 was quite low, the power of flow cytometry to isolate rare events was used. In total, 39.000.000 events were screened and 110.000 events were sorted according to gate P1.
Part III: Flow Cytometry-Based HTS Platform for Cell Free Expression in Polymersomes

Figure 36: Density scatter plot of FACS analysis and sorting. A model library generated with 0.75 ng pIX3.0RMT7/CelA2 His288Phe DNA and 2.25 ng pIX3.0RMT7/CelA2Glu580Gln DNA (right) was sorted according to gate P1. Negative (left) and positive controls (middle) were analyzed with 3 ng of pIX3.0RMT7/CelA2Glu580Gln DNA, and pIX3.0RMT7/CelA2His288Phe DNA, respectively.

Plasmids were isolated and transformed with competent E. coli DH5α cells, due to their high transformation efficiency. More than 250 colonies grew on selective agar plates, providing a sufficient number to draw statistical conclusions on the enrichment through sorting the fluorescent population by flow cytometry. Subsequently, plasmids were again isolated from colonies on agar plates and transformed with the expression host E. coli BL21. In addition, the 1:4 DNA (25% active CelA2) was transformed into E.coli BL21 cell and plated on azo-carboxymethylcellulose plates to assess the activity reaction in the library before sorting. Cellulase activity on azo-carboxymethylcellulose plates was indicated through halo formation.

Analysis of 200 clones before and after screening revealed an enrichment of the active population of 1.67 times from 17.85% active variants to 29.86% active variants (Figure 37). A discrepancy in the numbers between active CelA2 variants before sorting (17.8%) and expected activity ratio of 25% can be explained with inaccurate determination of the DNA concentrations with NanoDrop. Nonetheless, the result suggests that enrichment in the active population can be achieved through flow cytometry and subsequent DNA recovery.
Figure 37: Azo-carboxymethylcellulose plates with expressed *E. coli* BL21 pIX3.0RMT7/CelA2 His288Phe DNA or pIX3.0RMT7/CelA2E580Q colonies. The plates show halo formation of 112 clones in case active enzyme is produced. Plate on the left represents the amount of clones expressing active cellulase before flow cytometer sorting (20 clones show halo formation). Plate on the right represents the amount of clones expressing active cellulase after flow cytometer sorting (36 clones show halo formation). Crosses should guide the eye to formed halos.

Furthermore, this strategy was applied to a ‘real’ CelA2 library generated with error-prone PCR. Accordingly, an epPCR library was generated using 0.1 mM MnCl$_2$ leading to a mutation frequency of 2.5 per 1000 bp and 11.7 % clones coding for active CelA2. Flow cytometer analysis was performed using 5 ng DNA library, because several attempts with 3 ng DNA yielded not enough clones for a statistical evaluation of the enrichment. This DNA amount still provides a distribution of 67 % polymersomes with one or less DNA molecules per polymersome. In order to validate the enrichment of active CelA2 population using this flow cytometer-based strategy, in vitro expressed CelA2 epPCR library within polymersomes was analyzed and sorted. The analysis of the library revealed a subpopulation of 4.5 % within gate P1, thus being less fluorescent that the population of expressed wildtype (5.6 % in P1). More than 30.000.000 events were analyzed and the 2-3 % most fluorescent population was collected within gate P2 (Figure 38).
Figure 38: Density scatter plot of FACS analysis and sorting of a CelA2 epPCR library generated with 0.1 mM MnCl$_2$. Positive (top) with 5 ng of pIX3.0RMT7/CelA2His288Phe DNA, negative controls (middle) with 5 ng of pIX3.0RMT7/CelA2Glu580Gln DNA, and an epPCR library with 0.1 mM MnCl$_2$ with pIX3.0RMT7/CelA2His288Phe DNA (bottom) as template were analyzed.

The plasmid DNA was isolated from the collected droplets and transformed as reported for the model library first into competent *E. coli* DH5α cells and afterwards into competent *E. coli* BL21Gold (DE3) cells on selective LB agar plates. Single colonies were transferred and expressed inside 96-well MTP and the quality of the flow cytometer sorting was assessed through comparing the amount of active population before and after sorting. In total 180 clones were subjected to activity measurement, this time performed with 4-MUC screening system (Figure 39). Screening revealed 11.7 % clones coding for active enzyme before sorting and 35.0 % clones coding for active enzyme after sorting. Thus, one round of sorting results in a three times enrichment of the active population.
Figure 39: The CelA2 activity was determined using the 4-MUP detection system in 96-well microtiter plate format. CelA2 mutant library before sorting (left, 11.7 % active) and sorted CelA2 mutant library (right, 35 % active). Activity of wild type CelA2 is highlighted as green bar. Threshold for clones regarded as being active was set to 0.5 RFU s^{-1} increase in activity per second (blue line).
5.4 Discussion

In this chapter a novel technology for high throughput screening of CelA2 libraries inside polymersomes in combination with a flow cytometer screening was developed. Compartmentalization of genes in vitro has previously been applied to formats as double emulsions [62], liposomes [121], and microbeads [146] in which the selection and isolation of a gene is linked to a fluorescent signal. In comparison to liposomes, built up by phospholipids, polymersome membranes usually are multiple times thicker, due to the higher molecular weight of the block copolymers. Hence the membrane provides polymersomes a couple of advantages compared to liposomes, making them more suitable for encapsulation.

Here, a compartmentalization technology is developed in polymersomes based on encapsulation of gene library encoding for enzyme variants, fluorogenic substrate and lysate for cell free expression. The principle relies on production of enzyme variants and subsequent conversion of fluorogenic substrates in case active enzyme variants are present. In the first phase, CelA2 production was followed through fluorescent substrate conversion in order to estimate the time necessary for reaching the highest detectable fluorescent product signal. For in vitro expression of CelA2 a doubling in the fluorescent signal generated through conversion of resorufin-β-D-cellobioside was measured over the first 3 hours of enzyme production at 37°C. The final cease in CelA2 synthesis after 4 hours of in vitro synthesis was most likely induced through depletion of compounds necessary for in vitro enzyme production. In E. coli cells, in vitro transcription–translation is coupled with two units performing at the same speed - a RNA polymerase with a speed of 60nt per second and ribosomes translating the mRNA with a speed of 20 amino acids per second [142]. In cell-free systems this reaction is uncoupled, due to the use of bacteriophage T7 RNA polymerase instead of E.coli RNA polymerase, which is five to eight times faster than E.coli RNA polymerase [142]. One way to slow down the T7 RNA polymerase might be to use lower temperatures for protein synthesis. An improvement in the coupling of the transcription and translation is expected, if the reduced temperature slows down the rate of translation to a lesser extent than the rate of transcription. However, in a case of CelA2 production, effect of uncoupled transcription and translation did not lead to higher fluorescent signal indicating increased CelA2 amounts at temperatures lower than 37 °C (33.5 °C, 30 °C).

Commonly applied double emulsions in IVC have as a main drawback a leakage of essential hydrophobic components like fluorescent products into the oil phase yielding in a prominent cross talk between adjacent compartments [150]. Instead, polymeric vesicles consist of a dense hydrophilic brush with relatively long hydrophilic polymers on the surface as well as on the inside of polymersomes and therefore may increase their stability and decrease the flux in- and outside. Additionally, the presence of a hydrophilic brush on the surface reduces the protein adsorption onto the polymersomes [151]. Here, polymersome formation was carried
out using PMOXA-PDMS-PMOXA triblock copolymers, which were described to form vesicular structures in aqueous solutions [78]. The formation of higher molecular weight structures of diblock copolymers was confirmed by dynamic light scattering revealing three fractions according to size distribution (Figure 34). Only one fraction consisting of a diameter between 100 nm and 700 nm was analyzed by flow cytometer as the smallest fraction from 30 to 70 nm was too small to be detectable and the largest fraction from 4000 to 6500 nm was excluded through applied gating criteria (Figure 35). The size of the main fraction with range between 200 nm - 700 nm is in good agreement to previously reported data regarding the size of polymersomes formed with PMOXA-PDMS-PMOXA triblock copolymers [78].

In order to achieve a correlation between a gene and its corresponding enzyme, it is necessary to lower DNA amount to enable encapsulation of one gene per compartment. CelA2 synthesis inside polymersomes was carried using different DNA concentrations (concentration range from 1- 50 ng) in order to find minimal necessary amount of DNA for detection of the fluorescent signal by flow cytometer (Figure 35). A linear correlation between DNA amount and increase in fluorescent signal in subpopulation gated in P1 could be observed for DNA concentrations between 1 ng and 10 ng (Figure 40).

![Figure 40: Relation between fluorescent subpopulation in gate P1 and the amount of DNA polymersomes are loaded with.](image)

At higher DNA concentrations (> 10 ng) the correlation between fluorescent signal detected by flow cytometer and the DNA concentration was lost. This can be explained by a statistically higher number of polymersomes being loaded with DNA molecules. Furthermore, this result indicates that at low DNA concentrations there is a strong dependency of DNA concentrations and transcription speed, whereas at higher DNA concentrations, the protein synthesis seems to be more dependent on the speed of the translation reaction. At higher DNA concentrations (> 40 ng) the reaction reached plateau indicating saturation of \textit{in vitro} transcription-translation mixture by DNA. Nevertheless, the detection threshold of the flow
cytometer for IVC in polymersomes was 1 ng of encapsulated DNA – a concentration at which less than one fourth of the total number of polymersomes is loaded. Non-compartmentalized protein expression was hindered by the addition of protease in order to prevent unspecific fluorescence signal formation due to conversion of the substrate outside of the polymersomes by residual produced enzyme. In order to prove the genotype-phenotype linkage a model library consisting of 25 % DNA coding for active CelA2 was expressed inside polymersomes and 110,000 out of 39,000,000 analyzed events were sorted by flow cytometer (Figure 36). After transformation into E. coli BL21Gold (DE3), analysis of in total 400 clones before (200 clones) and after sorting (200 clones) revealed an enrichment of the active population of 1.67-times from 17.85 % active variants to 29.86 % (Figure 37). The achieved enrichment of the CelA2 model library showed that polymersomes are an excellent system for encapsulation and retention of all essential reaction compounds needed to synthesize enzyme variants. In order to demonstrate the applicability of the screening based polymersome technology for directed evolution, a 0.1 mM MnCl$_2$ epPCR mutant library of CelA2 was generated, sorted and transformed the same way as the model library. In total 180 clones were subjected to activity measurement (Figure 39). Analysis of the clones in 96 well MTP revealed 11.7 % clones coding for active enzyme before sorting and 35.0 % clones coding for active enzyme after sorting. This threefold enrichment of the active population exceeds the enrichment reported for an IVC experiment carried out in double emulsions with a 2-times enrichment of a 1:4 diluted ebg model library [62]. The excellent enrichment achieved with IVC in polymersomes contributes to a further improvement of directed evolution and high throughput technology by reducing the time for iterative cycles of directed evolution (from 2 years to 4-6 months), as expression of the protein and further sorting for improved variants can be performed in one day. In addition, the presented platform offers directed evolution of toxic enzymes for cell based hosts.

The herein developed flow cytometer screening system of in vitro compartmentalization inside polymersomes effectively solves a decade old problem that drastically limited broad application of IVC for enzyme engineering. In this study it was shown that the use of polymersomes enables the selection for enzymes exhibiting a $k_{cat}$ value as low as 0.509 s$^{-1}$ [152]. Compared with in vivo flow cytometer screening, this strategy allows the use of cell impermeable substrates and paves the way to express linear template DNA in vitro, thereby omitting the steps of cloning and transformation and the associated loss of diversity. The application of the developed technique allows researchers to screen with ultra-high throughput through unprecedented sequence (~$10^{10}$ variants) space in a time efficient and cost-effective way. The drastically increased screening capacity offers identification of significant improved enzyme catalysts after only one round of directed evolution which could only be achieved in two to three rounds of traditional directed evolutions campaigns [62].
This approach allows analysis of structure-function relationships based on cooperative effects *e.g.* for synergistic or additive interaction between mutated amino acids having a large distance to each other [73] and contributes to better molecular understanding.
5.5 Materials and methods

5.5.1 Preparation of lyophilizes polymersomes
The Triblock copolymer Poly(2-methylloxazoline\textsubscript{14}-b-dimethylsiloxane\textsubscript{65}-b-2-methylloxazoline\textsubscript{14}) (PMOXA-PDMS-PMOXA) was purchased from Polymer Source Inc.. 10 mg PMOXA-PDMS-PMOXA were diluted in 1 ml Milli-Q water and stirred overnight in a glass beaker. The turbid polymer suspension was exposed two times to shear stress (sonication 3 x 30 s with 30 s cooling intervals; amplitude 30 (VCX 130, Sonics & Materials Inc.)) with in between stirring for 1 h. Polymersome formation was finalized with 10 freeze thaw cycles of 3 min incubation in liquid nitrogen followed by 3 min incubation at 60 °C. Finally, 40 µl aliquots were made in PCR tubes and flash frozen in liquid nitrogen. Polymersomes were lyophilized overnight (Alpha 1-2 LD plus, Christ) and empty vesicles were stored at -20 °C.

5.5.2 Encapsulation of the transcription-translation machinery
The components of the transcription-translation reaction were purchased from RiNa (EasyXpress \textit{E. coli} Kit) thawed and mixed on ice. The assembly of the components is shown in Table 14 and a master mix was prepared in case more reactions were conducted.

<table>
<thead>
<tr>
<th>Components</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reaction buffer</td>
<td>4 µl</td>
</tr>
<tr>
<td>DNA</td>
<td>1 µl</td>
</tr>
<tr>
<td>Resorufin-(\beta)-D-cellobioside [2 mM]</td>
<td>2.64 µl</td>
</tr>
<tr>
<td>Chaperones 1:5 diluted</td>
<td>0.4 µl</td>
</tr>
<tr>
<td>Cell extract</td>
<td>3.5 µl</td>
</tr>
<tr>
<td>total</td>
<td>11.5 µl</td>
</tr>
</tbody>
</table>

The vector for cell-free expression pIX3.0 was received from RiNa. The vector pIX3.0RM7/CelA2His288Phe (Figure 41) was constructed and kindly provided by Dr. Ronny Martinez. The vector includes the cellulase CelA2, a \(\beta\)-lactamase for ampicillin resistance, a T7 promoter, a ribosomal binding site, and a T7 terminator.
Figure 41: Scheme of the constructed vector pIX3.0RMT7/CelA2His288Phe. The vector construct was used for in vitro expression of CelA2 under the control of a T7 Promoter.

An aliquot of lyophilized vesicles were hydrated with 11.5 μl of the assembled in vitro expression reaction. The mixture was briefly vortexed for 30 s until the vesicles were resuspended. Subsequently, the reaction was incubated on ice for 30 min to allow the polymersomes to swell. Protease from Bacillus licheniformis (Sigma P4860) was diluted 1:1000 to 2.4mU/g with Milli-Q water and 1 μl was added to the reaction to degrade non-encapsulated proteins. In vitro expression of CelA2 was performed in a PCR cycler (Eppendorf) at 37 °C for 4 h.

5.5.3 Activity determination in 384-well MTP format

Reaction was assembled in 384-well MTP as follows: 5 μl of in vitro expressed enzyme (see 5.5.2) at the respective timepoints was added to 20 μl KPi buffer (0.2 M, pH 7.2) and 20 μl resorufin-β-D-cellobioside (0.05 mM in Milli-Q water). For in vitro expression 65 ng circular plasmid DNA pIX3.0RMT7/CelA2His288Phe coding for active CelA2 or 65 ng circular plasmid DNA pIX3.0RMT7/CelA2Glu580Gln coding for inactive CelA2 were used in a total reaction volume of 25 μl. Reaction was incubated at 37 °C. Increase in fluorescent signal was measured over 15 min using Tecan infinite MTP reader at λem571/λex585 nm.

5.5.4 FACS sorting, DNA recovery, and transformation

Flow cytometry was carried out as described previously [133]. For DNA recovery, sorted polymersomes were heated to 85 °C for 5 min and 1μl Triton X100 was added per ml polymersome suspension. Plasmid was isolated using PCR purification kit (Macherey Nagel) and transformed with competent E. coli DH5α cells using heat shock to recover the plasmids [153]. Cell suspension was plated after recovery on selective LB_Amp agar plates and incubated overnight at 37 °C. Formed colonies on each plate were suspended with addition of 500 μl LB broth and gentle treatment with a drigalski spatula. Plasmid isolation of pelleted cells was performed with plasmid isolation kit (Macherey Nagel). For CelA2 expression
plasmid were transformed with competent *E. coli* Bl21Gold (DE3) cells and plated on selective LB agar plates supplemented with 100 mg ampicillin/l. The fluorometric assay for determining cellulolytic activity of CelA2 was performed by using 4-methylumbelliferyl-β-D-cellobioside (4-MUC, [143, 144]) as described previously [135].

5.5.5 **Dynamic light scattering (DLS)**

Dynamic light scattering experiments were performed with ZETASIZER Nano ZS (Malvern Instruments GmbH). Samples (100 µL) were diluted with Milli-Q water in a spherical glass cuvette. Measurements were performed at a scattering angle of 173 °, using a Helium-Neon gas laser (λ=633 nm) at 25 °C. The mean scattering intensity and the radius (in logarithmic scale) obtained from the cumulant fits is shown in the respective graph.
6 Conclusions

A major challenge of protein engineering campaigns is to screen through a whole generated protein sequence space. In order to cover this sequence space, it is often necessary to quickly analyze millions of genetic variants. In order to fill the gap between generated sequence space and limited screening capacity, novel high throughput screening technologies have to be developed, facilitating high screening throughput and its broad applicability. Novel flow cytometer- as well as microfluidic-based systems have a throughput of up to $10^7$ events per hour and emerged as suitable screening formats for large libraries [33-35]. Due to their outstanding throughput, they increase success probabilities in enzyme engineering campaigns and play a pivotal role in the question whether new enzymes can efficiently be isolated from metagenome libraries [36].

In this work two novel technologies based on flow cytometry that provide high throughput and a broad application potential were developed. One technology, named Fur-Shell, introduces the novel principle of labeling active clones with a fluorescent hydrogel. An impressive throughput of $1.8 \times 10^7$ events per hour was achieved, enabling to cover for example the screening of a simultaneous site saturation library consisting of six positions (64,000,000 variants) in one day. Importantly this strategy overcomes common drawbacks attributed to flow cytometry with whole cells like retention of fluorescent or fluorogenic compounds induced through membrane permeability. The problem of synthesizing not yet available fluorogenic substrates, which meet the condition to become fluorescent upon cleavage, is avoided by using glucose-derived substrates. A broad applicability of the Fur-Shell technology was demonstrated in screening four different hydrolases, yielding improvements for every single hydrolase after only one round of directed evolution. It is out of question to successfully expand the screening system to the class of oxidases and even the usage of transferases or oxygenases seems reasonable in case a suitable substrate for an enzyme generating hydrogen peroxide is applied. Of special interest might be also the application of the Fur-Shell technology in organic solvents or high salt applications. PEG is soluble in many organic solvents, including aliphatic ketones and alcohols, chloroform, glycol ethers, esters, and aromatic hydrocarbons and also glucose oxidases find applications in biosensors under non-aqueous conditions [154].

The second technology developed during this work includes cell-free production of a cellulase DNA library inside polymersomes with subsequent flow cytometer screening to isolate the active variants. Polymersomes have received great attention as extremely interesting systems for various applications in the past decade. This is represented in an increasing number of reports dealing with polymersomes in the past years (50 % of total publications on polymersomes were published in the past 3 years). Polymersomes are
intended to provide a barrier for hydrophilic, hydrophobic and amphiphilic molecules like any other vesicular structure, but their large membrane diameter and superior stability are unique and beneficial for protein engineering applications.

A method was developed to express small amounts (11 µl) of CelA2 inside polymersomes as in vitro expression kits are inordinately expensive, and thus this advancement also serves to reduce the cost of screening, as for control reactions low quantities were required.

The strategy of IVC in polymersomes and subsequent flow cytometer sorting was validated by screening 39,000,000 events of a CelA2 library. Remarkably, microtiter plate analysis using the 4-MUC screening system revealed an impressive three times enrichment of the active population in one round of sorting [62].

In conclusion, two novel technologies for high throughput flow cytometer screening were developed that allow due to their reliable and fast implementation to perform one round of directed evolution within less than one week. The reduced time effort allows directed evolution to become implemented as a routine tool in enzyme engineering campaigns for industrial purposes. Besides, the developed HTS methods can solve demanding and challenging problems in computational biology like revealing structure-function relationships through performing many rounds of directed evolution with high mutational load.
Publications

Journals

December 2014  
**Pitzler C.** Wirtz G., Vojcic L., Hiltl S., Böker A., Martinez R., and Schwaneberg U.; A Fluorescent Hydrogel-Based Flow Cytometry High-Throughput Screening Platform for Hydrolytic Enzymes; Chemistry & Biology (2014); 21(12) 1733-1742

January 2015  
Vojcic L.\(^\$\), **Pitzler C.**\(^\$\), Körfer G. \(^\$\), Jakob F. \(^\$\), Martinez R., Maurer K.-H., Schwaneberg U.; Advances in protease engineering for laundry detergents; New Biotechnology (2015); only online version available

February 2015  

April 2015  
Lülsdorf N.\(^\$\), **Pitzler C.**\(^\$\), Biggel M., Martinez R., Vojcic L., and Schwaneberg U.; A flow cytometer based whole cell screening toolbox for directed hydrolase evolution through fluorescent hydrogels; Chem. Commun., (2015); 41(51), 8679-8682

\(^\$\) shared first authorship

Patents


Contribution to DWI annual report 2014

May 2015  
**Pitzler C.**; DWI Leibniz-Institut für Interaktive Materialien; Annual Report 2014; A fluorescent hydrogel-based flow cytometry high-throughput screening platform for hydrolytic enzymes

Conferences

July 2014  
16\(^{th}\) European Congress on Biotechnology
Talk: “Lessons on directed evolution of hydrolases”
Poster: “A Fluorescent Hydrogel-based Flow Cytometry Screening Platform for Hydrolytic Enzymes”
Abbreviations

4-MUC 4-Methylumbelliferyl-β-D-cellobioside
4-MUP 4-Methylumbelliferyl-phosphate
Amp Ampicillin
BSA Bovine serum albumin
BSLA \textit{Bacillus subtilis} lipase A
Da Dalton
DNA Deoxyribonucleic acid
dNTP Deoxynucleotide triphosphate
\textit{E. coli} \textit{Escherichia coli}
epPCR error-prone polymerase chain reaction
FACS Fluorescent activated cell sorting
FPLC Fast protein liquid chromatograph
GOx Glucose oxidase
HTS High throughput screening
IVC \textit{In vitro} compartmentalization
LB Luria-bertani medium
MTP Microtiter plate
OD Optical density
\textit{pNPA} \textit{para}-nitrophenyl acetate
\textit{pNBEBL} \textit{para}-nitrobenzyl esterase from \textit{Bacillus licheniformis}
PCR Polymerase chain reaction
PDB Protein data bank
PEGDA Poly(ethylene glycol) diacrylate
PLICing Phosphorothioate-based ligase independent gene cloning
Ref References
RCB Resorufin-β-D-cellobioside
RFU Relative fluorescent units
RNA Ribonucleic acid
SDM Site-directed mutagenesis
SeSaM Sequence saturation mutagenesis
SFM Scanning force microscopy
TEA Triethanolamine
YASARA Yet another scientific artificial reality application
YmPh \textit{Yersinia mollaretii} phytase
Declaration

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Cologne 09th of July 2015

--------------------------------------------
Christian Pitzler
References


References


References


References


Acknowledgements
Curriculum Vitae

__________________________
Christian Pitzler
Luxemburger Straße 20
50674 Cologne
+49 177 - 5525640
Christian.pitzler@gmx.de

Date & place of birth
27. August 1981 in Dormagen, Germany

Nationality
German

__________________________
Academic education

Seit 11/2011
RWTH Aachen University, Biotechnology department, lab of Prof. Ulrich Schwaneberg
Title: “Novel flow cytometer-based platforms for directed evolution”

Bioengineer at Henkel Ag & Co KgaA, Düsseldorf.

04/2008 - 09/2009
Diploma thesis and practical term at Henkel Ag & Co KgaA, Düsseldorf.
Title: “Konstruktion und fermentative Charakterisierung von Bacillus licheniformis Protease Produktionsstämmen”

03/2004 - 09/2009
Diploma study in bioengineering, Fachhochschule Aachen
Degree: Diplom Ingenieur (FH)

09/2002 - 02/2004
Study in bioinformatics, University of Saarbrücken

__________________________
Civilian service

09/2001 – 06/2002
Arbeiter Samariter Bund, Cologne

__________________________
Education

08/1992 - 06/2001
Leibniz Gymnasium Dormagen
Degree: Allgemeine Hochschulreife

09/1988 - 07/1992
Katholische Grundschule Köln Worringen