Development of a flow cytometer-based *in vitro* compartmentalization screening platform for directed protein evolution

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

vorgelegt von

Master of Science (M. Sc.)
Biologie

Georgette Körfer (geb. Wirtz)

aus Stolberg (Rhld.)

Berichter: Universitätsprofessor Dr. Ulrich Schwaneberg
Universitätsprofessor Dr. Lothar Elling

Tag der mündlichen Prüfung: 26. Juli 2016

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek online verfügbar.

Table of contents

Table of contents

Abbreviations and acronyms

4-MUC	4-Methylumbelliferyl-β-D-cellobioside
4-MUP	4-Methylumbelliferyl-phosphate
Amp	Ampicillin
bp	Base pair
BSA	Bovine serum albumin
CelA2(WT)	CelA2 cellulase (wildtype)
Da	Dalton
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
E. coli	*Escherichia coli*
EF	Elongation factor
epPCR	Error-prone polymerase chain reaction
EV	Empty vector
FDC	Fluorescein-di-β-D-cellobioside
HTS	High throughput screening
IVC	*In vitro* compartmentalization
LB	Luria-Bertani medium
MTP	Microtiter plate
PCR	Polymerase chain reaction
PDB	Protein data bank
PLICing	Phosphorothioate-based ligase independent gene cloning
RFU	Relative fluorescent units
SDM	Site-Directed Mutagenesis
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SSM	Site-Saturation Mutagenesis
SeSaM	Sequence saturation mutagenesis
TB	Terrific broth
uHTS	Ultrahigh throughput screening
YASARA	Yet another scientific artificial reality application
YmPh(WT)	*Yersinia mollaretii* phytase (wildtype)

Abstract

Directed evolution is a powerful algorithm for tailoring industrially important biocatalysts. Methodological and conceptual advancements in directed evolution in the past decades enabled engineering of enzymes towards unnatural substrates with catalytic efficiencies beyond natural enzymes. Despite all progress in the field of directed evolution the exploration of generated sequence space and acceleration of iterative rounds of mutagenesis and screening is limited by throughput of commonly employed screening technologies. Development of ultrahigh throughput screening (uHTS) represents the main challenge in today's state-of-the-art directed evolution and ideally, in contrast to medium screening formats, enables rapid analysis of up to 10^7 events per hour using flow cytometry. The possibility to screen millions of variants within one day using uHTS offers the opportunity to generate libraries of high mutational load to successfully identify and investigate structure-function relationships or combinatorial effects within directed evolution campaigns. The latter is up to now unobtainable employing standard screening formats due to cost and time limitations of directed evolution campaigns. Most of reported uHTS screening systems employ *in vivo* screening where the bacterial host is used as carrier of enzyme variants. The main limitation of the cell-based systems is the low transformation efficiency resulting in reduced gene diversity. In this study we developed an uHTS platform in which cells were replaced by cell-free transcription-translation machinery encapsulated within artificial cell-like compartments (water-in-oil-in-water (w/o/w) emulsions). Each emulsion compartment encapsulates a single gene copy of a highly diverse gene library (10^8). Upon cell-free transcription and translation of genes, encoding active enzyme variants, compartments are labelled by conversion of fluorogenic substrate into fluorescent product and are analyzed by flow cytometry. *In vitro* compartmentalization (IVC) enables genotype-phenotype linkage and isolation of genes encoding for active enzyme variants from an excess of genes encoding inactive enzymes. The uHTS screening platform was developed for cellulase (CelA2) using *in vitro* production of CelA2 enzyme within (w/o/w) emulsions and fluorescein-di-β-D-cellobioside (FDC)-fluorescein substrate-product pair (λ_{ex} 494 nm/λ_{em} 516 nm). The developed uHTS cell-free compartmentalization platform was validated by screening of a random epPCR cellulase mutagenesis library and yielded an improved cellulase variant within only two weeks. The identified variant CelA2-H288F-M1 (N273D/N468S) showed 5-fold reduced K_M value and 13.3-fold increased specific activity (220.60 U/mg) compared to cellulase wild type (16.57 U/mg). The uHTS *in vitro* based platform was employed to study differences between three focused mutagenesis approaches (iterative site-saturation mutagenesis, multiple site-saturation mutagenesis, and simultaneous saturation mutagenesis of four amino acid positions *i.e.* OmniChange) in order to explore combinatorial effects. As an outcome the best performing variant contained only two amino acid substitutions (F288 and Q524) and showed 8-fold improvement in specific activity compared to parent.

In order to expand the uHTS platform on other enzymes, in this study *in vitro* based screening assay for *Yersinia mollaretii* phytase wildtype (YmPhWT) was optimized. Due to discrepancies between pH optimum for cell-free expression machinery (pH 7) and acidic pH optimum of YmPhWT, optimization of YmPhWT through a directed evolution campaign was performed. The goal of the campaign was to broaden the activity of the phytase towards neutral pH. Phytase libraries were generated *via* epPCR and subsequent simultaneous site-saturation mutagenesis based on identified beneficial positions in the epPCR library screening. The variants YmPh-M10 and -M16, showing 7-fold improved specific activity at pH 6.6 towards the model substrate 4-methylumbelliferylphosphate (4-MUP) compared to YmPhWT, were isolated.

The flow cytometer-based uHTS-IVC technology platform drastically reduces time requirements for one round of directed evolution from several months to only one day and enables an efficient screening of mutant libraries (10^{10}) with high diversity covering a significant portion of the generated sequence space in a time efficient manner. Consequently, the flow cytometer-based uHTS-IVC technology platform offers an efficient solution to tackle present challenges of directed evolution, like the performance of multiple, iterative rounds of diversity generation and screening of mutant libraries with unlimited diversity, in order to successfully identify beneficial variants with altered selectivity, specificity or activity. In addition, the flow cytometer-based uHTS-IVC technology platform has an impressive potential to study demanding/challenging scientific questions *e.g.* the exploration of combinatorial effects and answering fundamental questions on structure-function relationships within biocatalysts. Additionally, the uHTS technology platform can - from our point of view - be adapted to other enzyme classes since it uses a widely applicable fluorescence sorting of cell-free expressed active enzyme variants within (w/o/w) emulsions compartments.

1. General introduction

1.1. Directed evolution and gene diversity generation

Directed evolution is a powerful tool in protein engineering to tailor enzyme properties for the requirements of industrial or therapeutic applications (*e.g.* in chemical, pharmaceutical and food industry) and to elucidate structure-function relationships of proteins (Roccatano *et al.*, 2006; Wong *et al.*, 2006). The concept of directed evolution was introduced in 1967 by Spiegelman *et al.,* performing "evolution in a test tube" (Mills *et al.*, 1967), mimicking the process of Darwinian evolution with a few differences: in contrast to the natural Darwinian evolution, directed evolution requires an outside intelligence, not blind chance, and does not create brand new species (macroevolution), but only improvements within a species (microevolution), during a process that does not take millions of years, but happens rapidly (Stemmer *et al.*, 2003). Since then, directed evolution has been successfully applied in various fields of industry ranging from biotransformation for chemical synthesis (Jaeger *et al.*, 2004), biosensors (Doi *et al.*, 1999), bioremediation (Furukawa, 2003), and protein structure function relationships (Lee *et al.*, 2000), to medical and pharmaceutical applications in form of vaccines (Locher *et al.*, 2004) or therapeutic proteins (Roccatano *et al.*, 2006). Up to date, directed evolution matured to the point where it plays a key role in industrial biocatalyst development. In order to overcome the limitations of natural enzymes, such as low catalysis of non-natural substrates, low stability (towards temperature, pH, and oxidative agents), poor activity in non-aqueous media, requirements for expensive cofactors in protein engineering various strategies have been established to tailor enzyme properties for the requirements of industrial applications (Farinas *et al.*, 2001; Wong *et al.*, 2006). Upon generation of biocatalysts with novel or improved functions and properties, directed evolution can be either used for improvement of a specific, already exhibited enzyme property (*i.e.* activity, thermal stability, enantioselectivity) or for the combination of properties that are usually not found together in natural enzymes (Matsuura *et al.*, 2006).

In contrast to rational design, directed evolution can be applied to newly discovered enzymes with unknown crystal structures or mechanistic information, *e.g.* key amino acid positions, as it does not require an understanding of structure-function relationships of the target protein for its optimization (Wong *et al.*, 2006). A typical directed evolution experiment comprises three major steps: (I) starting from diversity generation, where mutations are randomly introduced in the gene of interest, to transformation and expression

of the mutated gene library in an appropriate expression host, (II) screening of the generated mutant library for improved variants using a suitable screening system mimicking conditions comparable with the final application conditions, and (III) isolation of the gene encoding for the improved protein variant, which is then usually used for iterative rounds of diversity generation and screening until a desired property or level of improvement is achieved, as shown in Figure 1 (Güven *et al.*, 2010; Wong *et al.*, 2006).

Figure 1: Principle of a directed evolution experiment. The figure is taken from Güven *et al.* (Güven *et al.*, 2010).

One of the key challenges for performance of a successful directed evolution experiment is the generation of mutant libraries of sufficient size with high diversity (Wong *et al.*, 2006). Nowadays, the large number of available chemical- or molecular-biologically-based methods for mutant library generation enables a fast and simple generation of highly diverse gene libraries (Wong *et al.*, 2006). For generation of random mutant libraries in directed evolution experiments the most commonly used method is **error prone PCR (epPCR)**. Due to incorrect incorporation of nucleotides, random mutations are introduced during PCR, by using *Taq* DNA polymerase from *Thermus aquaticus* which lacks a 3'-5' proofreading exonuclease activity and has an intrinsic high error rate of 8.0×10^{-6} mutations/bp/duplication (Cirino *et al.*, 2003). This results in around 1 substitution per 125,000 bp per PCR cycle which is insufficient for randomizing genes that seldom exceed 5,000 bp in length. In order to increase the error-rate, the fidelity of the DNA polymerase can be reduced by addition of $MnCl_2$, increasing the concentration of $MgCl_2$, supplementing unbalanced dNTP

concentrations or by increasing the number of amplification cycles (Cadwell *et al.*, 1992; Cirino *et al.*, 2003). Besides the advantages like robustness and simplicity of epPCR-based methods, typical disadvantages are the limited ability to generate sufficient diversity, due to the biased mutational spectra of the DNA polymerase, the redundancy of the genetic code and the lack of subsequent mutations due to low mutation frequencies and high numbers of transitions which are leading to so called "cold or hot spots" in a protein sequence (Wong *et al.*, 2004). In the last decade, the advanced random mutagenesis method **Sequence Saturation Mutagenesis (SeSaM)** was developed, which overcomes limitations caused by the polymerase bias. SeSaM is able to saturate every single nucleotide position of the target sequence with all four standard nucleotides and introduces the subsequent nucleotide mutations, so it can be used to generate transversions with a higher probability than epPCR (Wong *et al.*, 2004). Prominent recombination-based methods in directed evolution are **DNA shuffling** (Stemmer, 1994), **Staggered Extension Process (StEP)** (Zhao *et al.*, 1998) and **Phosphorothioate-based DNA Recombination (PTRec)** (Marienhagen *et al.*, 2012), which combine and accumulate beneficial mutations from different gene variants. Prominent techniques for focused diversity generation are **Site-Directed Mutagenesis (SDM)** or **Site-Saturation Mutagenesis (SSM)**, allowing the quick introduction of point mutations, with a mutagenesis efficiency of 70-90 %, into specific DNA-sequences using a pair of complementary mutagenesis primers for the amplification of a complete plasmid in one single PCR (Wang *et al.*, 1999). SDM and SSM can be used when structure and function of the enzyme is at least partly explored and it becomes reasonable to analyze amino acid exchanges at certain positions. This method was expanded using **Iterative Saturation Mutagenesis (ISM)** which employs subsequent cycles of SSM and screening (Reetz *et al.*, 2007). A recent advancement of these focused mutagenesis methods was achieved using multiple site-saturation with **OmniChange** method which enables a simultaneous site-saturation of up to five codons, allowing engineering of "localizable" properties (e.g. activity, selectivity) and the discovery of cooperative effects between amino acid substitutions that are improbable to identify using single codon saturation methods (Dennig *et al.*, 2011). The OmniChange method enables within only four steps the generation of diverse mutant libraries in a robust, reliable and simple manner, without DNA modifying enzymes like ligases or endonucleases, using a DNA sequence independent multi DNA

fragment assembly, based on chemical cleavage of phosphorothioated nucleotides by ethanol-iodine (Dennig *et al.*, 2011).

The method of choice for generation of the mutant library in a directed evolution campaign strongly depends on the enzyme property which should be optimized or re-engineered as well as on the screening throughput. Error-prone based mutagenesis methods, which are leading mainly to similar amino acid substitutions, could be beneficial for tailoring of enzyme properties like enantioselectivity (Arnold, 2001; Reetz *et al.*, 1999), whereas engineering of thermostability, requiring many mutations leading to chemically different amino acid substitutions, might be achieved using SeSaM method. All in all, nowadays there are plenty diversity generation methods available (OmniChange, SeSaM, etc.) to generate huge and highly diverse mutant libraries ($\sim 10^{15}$ mutants) covering and investigating the sequencing space of the gene of interest in order to tailor biocatalysts with improved and novel properties for industrial or pharmaceutical applications (Reetz *et al.*, 2008). In summary, these novel methods for diversity generation enable the generation of mutant libraries of high mutational load requiring potent screening formats to sample through the generated diversity.

1.2. Screening systems

As a huge variety of methods for the generation of highly diverse mutant libraries is already available (Ruff *et al.*, 2013), currently the biggest bottleneck of directed evolution campaigns is the ability to screen a significant portion of the generated sequence space, as for a protein of only 100 amino acids the theoretical sequence space comprises 20^{100} ($\sim 1.27 \times 10^{130}$) different sequences. The first requirement for designing a screening system, is its complementarity to the final application conditions (*i.e.* substrate, temperature, environmental stress). The established screening system has to mimic the application conditions as close as possible, in order to apply a correct selection pressure (Schmidt-Dannert *et al.*, 1999). Besides identification of improved variants for the desired property according to the settled rule "you get what you screen for" (Arnold *et al.*, 1999), the developed screening system has to offer scanning through a statistically meaningful fraction of the generated sequence space in a cost and time efficient manner (Wong *et al.*, 2006). The screening system has to be individually developed for each enzyme, should be optimized for medium to high throughput screening and sensitive over the final application range of

the substrate concentration, enabling the identification of desired hits in the generated mutant library (Aharoni *et al.*, 2005b).

The most common screening systems are solid phase, such as agar plates (Table 1), filter papers or membranes, offering high throughput of 10^4-10^6 clones/week, but having less detail level, as they only allow differentiation between active and inactive variants and are therefore widely used as prescreening methods (Lin *et al.*, 2002). Microtiter plate (MTP) based screening systems (Table 1) offer medium throughput (10^3-10^5 clones/week) and are mainly based on colorimetric and fluorimetric detection systems (Cohen *et al.*, 2001). MTP-based systems are in contrast to solid phase based screening methods expensive and labor intensive, but more accurate than solid phase screening methods (Olsen *et al.*, 2000). Flow cytometer-based screening technologies and microfluidic devices, offering <u>**ultrahigh throughput screening (uHTS)**</u> capabilities (10^7 events/hour), are becoming the most promising screening format in directed evolution experiments as they are able to screen a significant fraction of the generated sequence space in a cost-efficient and non-laborious manner (Aharoni *et al.*, 2005b; Farinas, 2006).

Table 1: Comparison of screening and selection technologies (Leemhuis *et al.*, 2009).

Strategy	Library size	Advantage	Disadvantage
Selection	$\sim 10^9$	Yields desirable variants only	Only possible if activity gives growth advantage
Agar plate screen	$\sim 10^5$	Simple to operate	Limited dynamic range
Microtiter plate screen	$\sim 10^4$	All analytical methods possible, excellent dynamic range	Relatively low screening capacity
Cell-in-droplet screen	$\sim 10^9$	Large libraries	Fluorescence detection and DNA modifying enzymes only
Cell as microreactor	$\sim 10^9$	Large libraries	Fluorescence detection only
Cell surface display	$\sim 10^9$	Large libraries	Fluorescence detection only
In vitro compartmentalization	$\sim 10^{10}$	No cloning steps, large libraries	Fluorescence detection and DNA modifying enzymes only

1.3. Ultrahigh throughput screening (uHTS)

In order to identify improved enzyme variants with the desired properties (increased pH-stability, activity or thermostability) for industrial applications it is necessary to develop efficient screening systems offering high or ultrahigh throughput for significant coverage of

variants in highly diverse gene libraries in form of a pre-screening and enrichment of active variants from a vast pool of inactive variants for subsequent detailed analysis.

1.4. Compartmentalization technologies

The microdroplets or compartments can be generated either by dispersion in water-in-oil (w/o) emulsions (Tawfik *et al.*, 1998) or in **water-in-oil-in-water (w/o/w) emulsions** that can be sorted by flow cytometer (Bernath *et al.*, 2004a). In addition, stable micro-compartments can be generated employing block copolymers, based on polyethylene glycol and biodegradable polyesters or polycarbonates, forming hollow vesicles, *e.g.* polymersomes, and entrapping the gene or cell of interest together with fluorogenic substrate within the polymer vesicle (Discher *et al.*, 2002; Meng *et al.*, 2005). Another method for generation of micro-compartments is the preparation of liposomes using freeze-dried empty liposomes for entrapment of genes or cells (Sunami *et al.*, 2006). Hydrogels, an alternative method applicable for many hydrolases, generate a fluorescent shell around cells producing active protein variants (Pitzler *et al.*, 2014). Additionally, micro-compartments can be generated, analyzed and sorted on-chip using microfluidic devices (Brouzes *et al.*, 2009; Kintses *et al.*, 2010).

In general, as compartmentalization can be used to entrap whole cells or cell-free mutant libraries, two types of compartmentalization can be distinguished: *in vivo* and *in vitro* compartmentalization. In the case of *in vivo* **compartmentalization** a gene library is cloned into expression host cells. The single cells are entrapped, *e.g.* in (w/o/w) emulsions, and sorted by flow cytometer (Aharoni *et al.*, 2005a). As shown in Figure 2, active cells, expressing the desired protein convert the substrate inside of the emulsion droplet into the fluorescent product. The fluorescent microdroplet with the active cell is detected and sorted directly into microtiter plates or collected and plated on agar plates for further analysis (Aharoni *et al.*, 2005a).

Figure 2: Compartmentalization of single cells in (w/o/w) emulsions and sorting with flow cytometer. (1.) A gene library is cloned and transformed into host cells resulting in a cell mutant library. (2.) Cells are expressed for production of the enzyme of interest. (3.) The mutant library is encapsulated in (w/o) emulsions and subsequently (4.) in (w/o/w) emulsions. (5.) Active cells, expressing and secreting the desired enzyme, convert the substrate inside of the emulsion droplet into the fluorescent product. (6.) The fluorescent microdroplet with the active cell is detected and sorted directly into microtiter plates or collected and plated on agar plates for further analysis.

In case of *in vitro compartmentalization (IVC)* single genes are cell-free transcribed and translated within a micro-compartment, *e.g.* a (w/o/w) emulsion (Figure 3). Especially for cell-free protein production, most commonly water-in-oil-in-water (w/o/w) emulsions, shortly called double emulsions, are employed which consists of an inner aqueous phase, surrounded by an oil phase (primary (w/o) emulsion) and an additional water phase. The gene encoding the enzyme of interest is entrapped together with the substrate inside the (w/o/w) emulsion compartment. Upon the protein expression, the substrate is converted by the *in vitro* produced enzyme into the fluorescent product and the fluorescent microdroplet can be detected by the optical and electronics system of the flow cytometer for identification of active enzymes for further enrichment (Figure 3), (Griffiths *et al.*, 2003; Mastrobattista *et al.*, 2005). Conversion of the primary emulsion into (w/o/w) emulsion, mimicking the lipid bi-layer of the cell-membrane, makes it possible to sort the droplets by flow cytometry (Bernath *et al.*, 2004a; Bernath *et al.*, 2004b). Using the IVC-based approach, the limitations of cell-based systems, *e.g.* low transformation efficiency of the expression hosts or low protein yields upon *in vivo* expression of membrane or toxic proteins, are addressed since the genetic material is directly put into the compartment, making this technique advantageous and interesting for research and industry (Blanusa, 2009).

Figure 3: *In vitro* compartmentalization (IVC) of a gene library in (w/o/w) emulsions and sorting with flow cytometer. (1.) A gene library is encapsulated in (w/o) emulsions and subsequently (2.) in (w/o/w) emulsions. Genes encoding for active enzyme variants are cell-free transcribed and translated within the *in vitro* expression mixture and produce the desired enzyme. Active enzyme variants convert the substrate inside of the emulsion droplet into the fluorescent product. (3.) The fluorescent microdroplet with the gene encoding for the active enzyme variant is detected and sorted with the flow cytometer. (4.) The sorted DNA encoding for active enzyme variants are recovered from the emulsions for the next round of diversity generation and sorting.

The generation of (w/o/w) emulsions can be performed by different techniques consisting of two emulsification steps. An elaborate and complex technique is the generation of (w/o/w) emulsions using microfluidic devices (Okushima *et al.*, 2004; Shah *et al.*, 2008), (Figure 4). More simple and rapid methods for generation of (w/o/w) emulsions are emulsification by ultra turrax (Hai *et al.*, 2004), by stirring, or by membrane extruder (Miller *et al.*, 2006), (Figure 4).

Different detergent and oil systems have been tested while the most promising ones for compartmentalization in (w/o/w) emulsions seem to be: Tween 80, Span 80 in light mineral oil (LMO)/Tween 20 in PBS (Hai *et al.*, 2004); ABIL EM 90 in LMO/CMC, triton-X102 in PBS (Aharoni *et al.*, 2005a) and decane, cholesterol and Span 60/Tween 20 in PBS (Miller *et al.*, 2006).

Figure 4: Preparation of (w/o) and (w/o/w) emulsion compartments, entrapping gene or cell mutants, by different techniques (*i.e.* stirring, extrusion, homogenizing or microfluidics). The (w/o/w) emulsions are prepared in two steps, first a high-shear emulsification step with lipophilic surfactants for generation of the (w/o) emulsion, and second a low shear emulsification step with hydrophilic surfactants for generation of the (w/o/w) emulsion. The (w/o) and (w/o/w) emulsion pictures were taken by transmission microcopy and the scale bars represent a length of 20 μm.

1.5. Cell-free protein expression

For the *in vitro* expression of proteins it is necessary to perform transcription and translation of a DNA molecule into mRNA molecules and the subsequent translation of the mRNA molecules into polypeptides in a cell-free environment, e.g. within a laboratory mixture containing ribosomes, polymerases, chaperones etc. (Figure 5). The first successful attempts to translate an exogenous mRNA molecule in a cell-free environment were performed by Nirenberg and Matthaei in 1961 (Nirenberg *et al.*, 1961). Six years later, Zubay and his team managed to couple transcription and translation and developed in this way an *in vitro* system for translation of DNA molecules (DeVries *et al.*, 1967). Nowadays *in vitro* translation is an important technique to analyze the process of mRNA translation into functional proteins, especially in the field of functional and structural proteomics (Kigawa *et al.*, 1999).

1.6. Enzymes

1.6.1. Cellulases

Cellulases (EC-number: 3.2.1.4) catalyze the hydrolysis of β-1,4-glycosidic bonds in cellulose and cellulose compounds (polymers build-up of anhydro-β-1,4-glucose units connected by β-1,4-D-glycosidic bonds). There are three different types of cellulases (Percival Zhang *et al.*, 2006): (1) endoglucanases (EC-number: 3.2.1.4), hydrolyzing internal β-1,4-glycosidic bonds, (2) exoglucanases (EC-number: 3.2.1.91), hydrolyzing β-1,4-glycosidic bonds at the non-reducing end and (3) β-glucosidases (EC-number: 3.2.1.21), cleaving cellobiose and oligosaccharides to glucose. Figure 6 shows the hydrolysis of cellulose by the cellulolytic system of *Trichoderma reesei* in form of a simplified scheme. At internal amorphous sites in the cellulose polysaccharide chain endoglucanases cleave the β-1,4-D-glycosidic bonds to oligosaccharides of different lengths and chain ends (reducing or non-reducing ends) (Figure 6). In contrast to endoglucanases, exoglucanases are able to cleave microcrystalline cellulose, releasing cellobiose as reaction product, acting either from the reducing end (CBHI) or from the non-reducing end (CBHII) (Teeri, 1997). The soluble cellobiose can be converted subsequently to glucose by β-glucosidase. As for complete cellulose conversion the accessibility of the cellulase to its substrate is main limitation, depending on physical properties like polymerization, crystallinity and accessibility, determined by size and porosity of the cellulose, most cellulases have different domain structures, *e.g.* the cellulose binding domain (Chandra *et al.*, 2007). The cellulose binding domain (CBD) is connected by a flexible linker to the catalytic domain and enables the cellulase to recognize the solid surface of the cellulose and facilitates the movement across a single chain of decrystallized cellulose fiber, which is immediately moving into the active site tunnel of the catalytic domain. It was reported that the binding efficiency of the cellulase on the cellulose surface directly correlates with the efficiency of crystalline cellulose degradation (Esteghlalian *et al.*, 2001; Linder *et al.*, 1995) and a removal of the CBD results in a drastic decrease (80 %) in cellulase activity on solid substrates (Gilkes *et al.*, 1988).

Figure 6: Schematic representation of the hydrolysis of amorphous and microcrystalline cellulose with the cellulolytic system of *T. reesei*. The solid circles represent non-reducing ends and the open circles represent reducing ends. The figure is modified from Lynd *et al.* (Lynd *et al.*, 2002).

Cellulases can be isolated from fungi, bacteria and protozoa and play an important role in coffee production, biofuel production, laundry, textile, paper and detergent industry. In order to meet the increasing global energy demands it is necessary to find alternatives to fossil fuels, like the production of tailor made biofuels. For the production of liquid biofuels an efficient conversion of lignocellulosic material into ethanol is necessary which can be performed by cellulases, which are hydrolyzing cellulose polymers in biomass into monomeric sugars that can be easily fermented to ethanol (Kumar *et al.*, 2009). The cellulase used for this project is a **CelA2 cellulase** (celA2, GenBank: JF826524.1), consisting of 604 amino acids and having a molecular weight of around 69 kDa. CelA2 was isolated from a metagenomic library from a biogas plant sample (Ilmberger *et al.*, 2012) and was cloned into *E. coli* BL21-Gold(DE3) strain. The variant **CelA2-H288F** revealed an outstanding performance in ionic liquids (IL) and sodium chloride and is therefore an excellent candidate to be applied in the cellulose hydrolysis process in the generation of biofuels (Lehmann *et al.*, 2012).

1.6.2. *Phytases*

Phytases, (EC-number: 3.1.3.8), also called *myo*-inositol hexakis phosphohydrolases, are mainly found in plants, fungi, microorganisms, and animals and belong to the enzyme group of phosphomonoester hydrolases, which catalyze the hydrolytic cleavage of phytates (*myo*-inositol-1,2,3,4,5,6 hexakis dihydrogen phosphate) into inositol, phosphates, divalent cations and proteins (Yao *et al.*, 2012), as shown in Figure 7.

Figure 7: The hydrolysis of phytate by phytase into inositol, phosphate and other divalent elements. The cleavage of phosphate groups by phytase results in the release of metals, metal-binding enzymes and proteins. Many phytases however cleave only five phosphates. The figure is taken from Yao *et al.* (Yao *et al.*, 2012).

Phosphorous and inositol are stored in plants or seeds (*i.e.* cereals, oil seeds or legumes) in form of phytate which cannot be hydrolyzed in the gut of monogastric animals, like dogs, birds and humans due to low level of phytase activity (Bitar *et al.*, 1972). This leads to phosphorous deficiency which requires supplementation of phosphorous in form of expensive feed additives (Konietzny, 2004). Additionally, phytates are strong chelating agents as they have the ability to fix divalent cations (*e.g.* Ca^{2+}, Zn^{2+}, Fe^{2+}) and proteins in stable complexes making them unavailable for the organism (Rao *et al.*, 2009). Phytases catalyze the step-wise hydrolysis of phosphomonoester bonds by release of inorganic phosphate, enabling monogastric animals to assimilate phosphorus. Therefore, phytases are attractive food supplements for the treatment and prevention of iron and mineral deficiency caused by phytate in humans and animals as they significantly improve the food and feed nutritional quality (Sandberg *et al.*, 1996). Beneficial uses of phytase comprise phosphate-mobilizing feed additives (*i.e.* swine, poultry or fish), development of transgenic organisms, and soil amendment for intensive crop production (Shivange *et al.*, 2012; Yao *et al.*, 2012). Furthermore, unabsorbed phosphates in form of phytate are usually excreted and released to the soil. The phytate is subsequently converted by soil-microorganisms which leads to phosphate run-off and environmental pollution (Konietzny, 2004). In this case,

treatment of agricultural soils with phytases can significantly decrease the environmental pollution by 30-50 % caused by excretion of phosphorous in form of phytate by monogastric animals (Konietzny, 2004; Shivange *et al.*, 2012; Yao *et al.*, 2012). In addition, the usage of cost-intensive phosphorous fertilizers in agriculture can be decreased upon usage of phytases in transgenic plants or for soil amendment on the fields to increase the crop production by improved phosphate assimilation from phytate in plants and animals (Hong *et al.*, 2008). Phytases can either be divided into acid and alkaline phytases dependent on their pH optima or they can be grouped according to their catalytic mechanism into three classes (1) histidine acid phosphatase (HAPs), (2) ß-propeller phytase, and (3) cysteine phytase or purple acid phosphatase (Rao *et al.*, 2009; Tye *et al.*, 2002; Yao *et al.*, 2012). Many phytases isolated from fungi, bacteria and plants belong to the HAPs class of enzymes (Wyss *et al.*, 1999) and share the same conserved active site amino acid motif (RHGXRXP) and the catalytic dipeptide (HD) (Shivange *et al.*, 2012; Van Etten *et al.*, 1991). The enzyme of choice for this project is a highly active tetrameric phytase, which belongs to the group of acidic phosphatases, with a size of 43 kDa (188 kDa in native page electrophoresis) and a specific activity of 1073 U/mg which was isolated from *Yersinia mollaretii* (Ym-phytase) and cloned into *E.coli* BL21-Gold(DE3)lacI[Q1] strain (Shivange *et al.*, 2012).

1.6.3. *Fluorescence-based assays for detection of phytase and cellulase activity*

As the proposed ultrahigh throughput screening (uHTS) system is based on flow cytometry, for development of this screening platform it is necessary to find suitable fluorescent assays to detect phytase and cellulase activity with the available BD Influx flow-cytometer-based sorting device in our laboratory. This flow cytometer, in our case, includes three excitation lasers, a UV laser for excitation at 355 nm, a blue laser for excitation at 488 nm and a yellow-green laser for excitation at 651 nm, as well as emission detection filters for 450/50 nm, 530/40 nm, 585/29 nm, 612/20 nm, 670/30 nm, 673/30 nm, 710/50 nm, 750/LP.

Several available fluorescent assays, *e.g.* 4-MUC activity (Lehmann *et al.*, 2012) assay or 4-MUP activity assay (Shivange *et al.*, 2012), adapted for MTP screening have to be tested and optimized for *in vitro* compartmentalization systems. The main requirements for fluorescent assays employed in a flow cytometer screening platform are: (1) dye diffusion (*e.g.* fluorescent substrate/product of the enzymatic reaction should stay inside the inner water phase of the emulsion droplet), which can be prevented by charged moieties into

Since 4-MU is not charged, it is possible that it might diffuse into the oil phase. In this case several alternative substrates that cannot diffuse out of the emulsion can be used: 1.) 4-methylumbelliferyl phosphate-6-carboxylic acid (4-MUPCA). In this reaction the phytase converts the substrate to the blue fluorescent product 4-methylumbelliferone-6-carboxylic acid (4-MUCA) (λ_{ex} 360 nm; λ_{em} 465 nm), by cleavage of the phosphate group (Figure 9).

Figure 9: Conversion of 4-methylumbelliferyl-phosphate-6-carboxylic-acid (4-MUPCA) substrate into the fluorescent product 4-methylumbelliferone-6-carboxylic-acid (4-MUCA) by phytase.

2.) Fluorescein di-phosphate triammonium salt (FDP) is another alternative substrate, which should not likely diffuse into the oil phase. In this reaction the phytase converts the FDP to the green fluorescent fluorescein (λ_{ex} 490nm; λ_{em} 514nm), by cleavage of the two phosphate groups (Figure 10).

Figure 10: Conversion of fluorescein di-phosphate triammonium salt (FDP) substrate into the fluorescent product fluorescein by phytase.

1.6.3.2. Fluorescence-based assays for detection of cellulase activity

For the cellulase, 4-methylumbelliferyl-β-D-cellobioside (4-MUC) is used as fluorogenic substrate for fluorescence based microtiterplate and flow cytometric assays (Lehmann *et al.*, 2012). The conversion of the blue fluorescent product 4-methylumbelliferone is detected by increase of fluorescence (ex. 360 nm, em. 465 nm), see Figure 11.

Figure 11: Conversion of 4-methylumbelliferyl-β-D-cellobioside (4-MUC) substrate into the fluorescent product 4-methylumbelliferone (4-MU) by cellulase.

Since 4-MU is not charged, it is possible that it might diffuse into the oil phase. In this case fluorogenic substrate fluorescein-di-β-D-cellobioside (FDC) can be used as alternative substrate for flow cytometer-based screening within (w/o/w) emulsion compartments, which should not likely diffuse into the oil phase. In this reaction, cellulase converts the FDC substrate to the green fluorescent product fluorescein (λ_{ex} 494 nm/λ_{em} 516 nm), by cleavage of the β-1,4-D-glycosodic bonds (Figure 12).

Figure 12: Conversion of fluorescein di-β-D-cellobioside (FDC) substrate into the fluorescent product fluorescein by cellulase.

2. Goal of this work

The main goal of this project was the development of a novel and routine-ready flow cytometer-based ultrahigh throughput *in vitro* compartmentalization (uHTS-IVC) screening platform for directed protein evolution (Figure 13). The project was financed by the German Federal Ministry of Education and Research (BMBF) within the framework of *"Basistechnologien für die nächste Generation biotechnologischer Verfahren"* (FKZ 031A165), which is supporting the development of novel, robust technology platforms for industrial applications in the field of biotechnology.

Screening throughput is one of the key criteria for successful directed enzyme evolution and discovery of novel enzymes. Medium to high throughput screening systems based on microtiter plates (MTP) or agar plates usually offer analysis of 10^4-10^5 variants, which is still significantly lower compared to the size (10^8-10^9) and potential complexity of random mutagenesis libraries (Agresti *et al.*, 2010; Wong *et al.*, 2006). As nowadays the development of rapid screening systems for efficient coverage of the generated sequencing space is the main limitation of directed evolution campaigns, ultrahigh throughput enzyme screening (uHTS) technologies offering a throughput of up to 10^7 events per hour are a key tool for today's state-of-the-art directed evolution experiments, as they enable rapid analysis of millions of variants (Griffiths *et al.*, 2006; Mayr *et al.*, 2009; Miller *et al.*, 2006). Ultrahigh throughput of 10^7 events per hour enables efficient coverage of the generated sequence space in highly diverse mutant libraries (Dennig *et al.*, 2011; Reetz *et al.*, 2010; Shivange *et al.*, 2009) and allows successful re-engineering of biocatalysts with high industrial application potential (*e.g.* for chemical, pharmaceutical, food industry) through iterative rounds of directed evolution in a time and cost efficient manner (Prodanovic *et al.*, 2011).

A recent highlight in directed enzyme evolution was the development of an alternative ultrahigh throughput screening platform employing *in vitro* enzyme library production within water-in-oil-in-water (w/o/w) emulsion compartments (*in vitro* compartmentalization) and subsequent analysis by flow cytometry (Bernath *et al.*, 2005). *In vitro* compartmentalization (IVC) technology enables the miniaturization of reaction volumes to microliter scale, by production of 10^{10} reaction compartments per milliliter of reaction with diameters ranging from 0.5-10 µm, resulting in a dramatic reduction of consumable costs, workload and assay time, compared to standard medium to high throughput screening systems (Agresti *et al.*, 2010; Bernath *et al.*, 2005; Mastrobattista *et al.*, 2005; Taly *et al.*, 2007). Upon *in vitro*

transcription-translation of the gene library into enzyme variants, active variants convert fluorogenic substrate into fluorescent product thus labelling water-in-oil (w/o) single emulsions. In order to enable subsequent ultrahigh throughput analysis and sorting by flow cytometer in an aqueous environment, (w/o) single emulsions are dispersed in an external water phase resulting in water-in-oil-in-water (w/o/w) emulsions (Bernath *et al.*, 2005; Ruff *et al.*, 2012; Taly *et al.*, 2007). IVC technology is completely independent of host cells, enabling the production of toxic and membrane bound proteins (Merk *et al.*, 2015; Worst *et al.*, 2015). Additionally, cell-free technology overcomes the challenge of diversity loss in directed evolution experiments due to low transformation efficiency of expression hosts (*i.e. E.coli* or *Bacillus*) since the DNA material is directly added in the inner aqueous phase of the compartment (Tu *et al.*, 2011). Several criteria must be fulfilled in order to achieve successful qualitative discrimination between active and inactive enzyme variants, by using *in vitro* enzyme compartmentalization-based flow cytometry screening platform: (1) reduced diffusion rates of fluorogenic substrate and/or fluorescent-based product from (w/o/w) emulsion compartments, (2) reduced diffusion rates of cell-free reaction components from (w/o/w) emulsion compartments, (3) monodispersity of (w/o/w) emulsion, and (4) stability of (w/o/w) emulsion compartments.

In order to establish the proposed flow cytometry-based uHTS-IVC screening platform and to make it applicable in directed evolution campaigns it is necessary, in the first step, to develop a flow cytometer-based ultrahigh throughput screening combined with *in vitro* compartmentalization (IVC) technology. In the second step, one or two industrial important enzymes (*i.e.* phytase and/or cellulase) have to be optimized for desired properties, *e.g.* activity, by application of the established screening platform as a proof of concept. The strategy to develop and validate the flow cytometer-based uHTS-IVC technology platform comprises the following steps: (1) development of an efficient protocol for emulsion generation, (2) selection of the fluorescent assays suitable for flow cytometric application in emulsions, (3) optimization of *in vitro* expression for enzymes (phytase and/or cellulase) in emulsion droplets, (4) sorting of populations with high mutational load and low numbers of active variants and (5) performing the directed evolution experiment for phytase and/or cellulase as a proof of principle for validation the novel developed technology platform. In total up to now, only three success stories for directed evolution of β-galactosidase (Mastrobattista *et al.*, 2005), [FeFe] Hydrogenase (Stapleton *et al.*, 2010) and

phosphotriesterase (Griffiths *et al.*, 2003), employing IVC-based flow cytometry in (w/o/w) emulsions have been reported.

In this thesis, for the first time the development of an uHTS technology platform-based on IVC (uHTS-IVC) in (w/o/w) emulsion for directed cellulase (CelA2-H288F) evolution will be presented. The cell-free transcribed and translated cellulase mutant libraries will be encapsulated together with fluorescein-di-β-D-cellobioside (FDC) as substrate within (w/o/w) emulsion compartments. Translated active cellulase variants catalyze the cleavage of the β-1,4-D-glycosodic bonds and the resulting fluorescein will be detected by fluorescence measurements (λ_{ex} 494 nm/λ_{em} 516 nm) using flow cytometry to discriminate between active and inactive cellulase variants. The validation of the developed flow cytometry-based uHTS-IVC platform will be performed in form of a CelA2-directed evolution campaign by screening an error-prone PCR (epPCR) cellulase mutant library using linear DNA template. In addition, three different types of focused mutagenesis (*e.g.* OmniChange) on positions in close proximity to the active site of cellulase will be screened using uHTS-IVC platform, by using *in vitro* transcription-translation of plasmid DNA, and combinatorial effects will be investigated.

Additionally, the highly active *Yersinia mollaretii* phytase will be evolved for higher activity at neutral pH (*i.e.* pH 6.6), in order to broaden the application potential of this acid phosphatase in various industries (*e.g.* feed industry, agriculture) and to make it applicable in cell-free expression systems employed in the flow cytometry-based uHTS-IVC screening platform. Therefore, the pH profile of the phytase has to be expanded to neutral pH which will be realized through a directed evolution campaign using epPCR as random mutagenesis method for diversity generation, followed by focused mutagenesis using SSM of the identified beneficial positions from the random approach.

Compared to existing *in vivo* and *in vitro* flow cytometer-based screening systems, the novel flow cytometry-based uHTS-IVC screening platform inaugurates applications for various enzymes, especially hydrolases, overcoming the limitations of previous flow cytometer-based screening systems, *i.e.* leakage of substrate/product due to cell membrane permeability, it uses completely cell-free expression systems instead of frequently used lysed cells in the compartments, no additional microbeads entrapped together with the protein in the compartment, no complicated cell surface display technology, and it is completely independent of host cells fostering the production of toxic or membrane bound

proteins, which are difficult to express *in vivo*. The application of the simple and rapid flow cytometry-based uHTS-IVC screening platform for directed cellulase evolution reveals revolutionary advances in regard to screening capabilities matching to requirements of novel methods for diversity generation in today's directed evolution campaigns. The quick and simple encapsulation of DNA using optimized extrusion technique, as well as DNA recovery after sorting, and the good compatibility between cell-free extracts and emulsion components (*i.e.* surfactants, oils) for generation of robust and stable emulsion compartments, that are able to withstand the shear forces upon flow cytometer-based analysis and sorting, demonstrate the broad applicability of the routine-ready flow cytometry-based uHTS-IVC technology platform.

Consequently, the development of the novel flow cytometry-based uHTS-IVC technology platform will be a major breakthrough for the directed evolution of proteins, as it overcomes the current limitations of today's state-of-the-art directed evolution such as insufficient throughput capacities and diversity losses due to limited transformation efficiencies of the host cells. The uHTS-IVC technology will offer significant advancements regarding reduction of time, screening efforts and expenses facilitating the performance of iterative rounds of mutagenesis and screening. The powerful uHTS-IVC technology platform will enable studying cooperative effects of amino acid side chains by screening advanced mutant libraries of high mutational loads with up to five simultaneous site-saturations (*i.e.* OmniChange). The latter will increase the possibility to identify beneficial variants showing altered activity, specificity or selectivity. The development of a robust uHTS-IVC technology screening platform will increase success probabilities to identify beneficial variants from diverse mutant libraries containing high numbers of inactive variants, which is unfeasible using standard screening systems due to time and cost limitation of directed evolution campaigns. Additionally, the use of IVC technology will facilitate the performance of iterative rounds of mutagenesis and screening and enable a rapid engineering of proteins exhibiting improved properties in a cost and time efficient manner. In addition, the employment of the flow cytometry-based uHTS-IVC technology platform will open up new exciting opportunities to study challenging scientific questions like the exploration of non-understood enzyme properties (*e.g.* organic solvent, ionic liquid resistance, pH stability).

3.3. Introduction

Screening throughput is a key criteria for successful directed enzyme evolution and discovery of novel enzymes. Medium to high throughput screening systems based on microtiter plates (MTP) or agar plates usually offer analysis of 10^4-10^5 variants, which is still significantly lower compared to the size and potential complexity of random mutagenesis libraries (10^8-10^9) (Agresti *et al.*, 2010; Wong *et al.*, 2006). On the other hand, ultrahigh throughput enzyme screening (uHTS) technologies enable a throughput of up to 10^7 events per hour and are a key tool for today's state-of-the-art directed evolution experiments (Griffiths *et al.*, 2006; Mayr *et al.*, 2009; Miller *et al.*, 2006). Ultrahigh throughput enables an efficient coverage of the generated sequence space in highly diverse mutant libraries (Dennig *et al.*, 2011; Reetz *et al.*, 2010; Shivange *et al.*, 2009) and allows successful re-engineering of biocatalysts with high industrial application potential (*e.g.* for chemical, pharmaceutical, food industry) through iterative rounds of directed evolution in a time and cost efficient manner (Prodanovic *et al.*, 2011). A recent highlight in directed enzyme evolution was the development of an alternative ultrahigh throughput screening platform employing *in vitro* enzyme library production within water-in-oil-in-water (w/o/w) emulsion compartments and subsequent analysis by flow cytometry (Bernath *et al.*, 2005). Water-in-oil (w/o) single emulsions are generated using extrusion or homogenizing and contain an inner aqueous phase comprising gene mutant library, cell-free transcription-translation reaction mixture, and the fluorogenic substrate (Bernath *et al.*, 2005). The inert oil phase of the emulsion compartment mimics the bacterial cell membrane by encapsulation of ideally one DNA molecule per compartment, enabling genotype-phenotype linkage (Bernath *et al.*, 2004b). Upon *in vitro* transcription-translation of the gene library into enzyme variants, active variants convert the fluorogenic substrate into the fluorescent product thus labelling water-in-oil (w/o) single emulsions. In order to enable subsequent analysis and sorting by flow cytometer in an aqueous environment, (w/o) single emulsions are dispersed in an external water phase resulting in water-in-oil-in-water (w/o/w) emulsions (Bernath *et al.*, 2005; Ruff *et al.*, 2012; Taly *et al.*, 2007). *In vitro* compartmentalization (IVC) technology enables the miniaturization of reaction volumes by production of 10^{10} reaction compartments per milliliter of reaction with diameters ranging from 0.5-10 µm, resulting in a dramatic reduction of consumable costs, workload and assay time (Bernath *et al.*, 2005; Mastrobattista *et al.*, 2005; Taly *et al.*, 2007). IVC technology is completely independent of

host cells, enabling the production of toxic and membrane bound proteins (Merk *et al.*, 2015; Worst *et al.*, 2015). Additionally, cell-free technology overcomes the challenge of diversity loss in directed evolution experiments due to low transformation efficiency of expression hosts (*i.e. E.coli* or *Bacillus*) since the DNA material is directly added in the inner aqueous phase of the compartment (Tu *et al.*, 2011). Several criteria must be fulfilled in order to achieve a successful discrimination between enzyme variants using the flow cytometry-based screening platform in combination with *in vitro* enzyme compartmentalization: (1) reduced diffusion rates of fluorogenic substrate and/or fluorescent-based product from (w/o/w) emulsion compartments, (2) reduced diffusion rates of cell-free reaction components from (w/o/w) emulsion compartments, (3) monodispersity of (w/o/w) emulsion, and (4) stability of (w/o/w) emulsion compartments.

Up to now, only a few success stories for directed evolution of β-galactosidase (Mastrobattista *et al.*, 2005), [FeFe] Hydrogenase (Stapleton *et al.*, 2010) and phosphotriesterase (Griffiths *et al.*, 2003) employing IVC-based flow cytometry have been reported.

In this project, for the first time the development of an ultrahigh throughput *in vitro* compartmentalization (uHTS-IVC) technology platform using (w/o/w) emulsions for directed cellulase (CelA2-H288F) evolution was reported. The cell-free transcribed and translated cellulase mutant library was encapsulated together with fluorescein-di-β-D-cellobioside (FDC) as substrate within (w/o/w) emulsion compartments. Translated active cellulase variants catalyze the cleavage of the β-1,4-D-glycosodic bonds and the resulting fluorescein can be detected by fluorescence measurements (λ_{ex} 494 nm/λ_{em} 516 nm) using flow cytometry to discriminate between active and inactive cellulase variants. The developed flow cytometry-based uHTS-IVC platform was validated by screening an error-prone PCR (epPCR) cellulase mutant library.

3.3.1. *Principle of the flow-cytometer-based uHTS-IVC platform*

The **flow cytometer-based <u>u</u>ltra<u>h</u>igh <u>t</u>hroughput <u>s</u>creening platform for directed protein evolution using <u>*in*</u> <u>*v*</u>*itro* <u>c</u>ompartmentalization (uHTS-IVC platform)** significantly reduces the time necessary for tailoring and optimization of enzymes attractive for industrial application and decreases the labor intensity (*i.e.* picking of thousands of single clones into microtiter plates). The principle of the uHTS-IVC platform through seven subsequent steps is summarized in Figure 13. Within approximately six hours a gene diversity library with high mutational load containing >10^8 mutants can be generated using standardized epPCR protocol (Figure 13, Step 1) and encapsulated in (w/o/w) emulsion droplets together with fluorogenic substrate and *in vitro* extract mixture for cell-free enzyme production (Figure 13, Steps 2-3). The compartmentalization in emulsion droplets enables genotype-phenotype linkage since it ensures that the gene, the enzyme it encodes and the generated fluorescent product remain within the same compartment (Figure 13, Step 3). The resulting fluorescent product labels (w/o/w) emulsions containing an active enzyme variant and enables its analysis by flow cytometry. The flow cytometry screening platform offers screening of 10^7 events per hour and sorting of 5,000 events per second by qualitative differentiation between fluorescent and non-fluorescent events (Figure 13, Step 4). The genes encoding for active enzyme variants in the sorted sample can be isolated and amplified by PCR in approximately three hours (Figure 13, Step 5). Amplified PCR products can be used as a template for further iterative rounds of directed evolution. After several rounds of directed evolution the genes encoding for improved enzyme variants are cloned and transformed into the expression host (Figure 13, Step 6). Expressed active enzyme variants can be further characterized in MTP format (*e.g.* for improved activity, thermal and organic solvent resistance) (Figure 13, Step 7) using already established spectrophotometric or fluorimetric detection systems. The ultrahigh throughput screening platform offers rapid screening of enormous high number of enzyme variants (>10^8) and enables enzyme engineering through 4-6 iterative rounds of directed evolution in about two months which is a significant technological advancement compared to traditional directed evolution campaigns (up to 1-2 years).
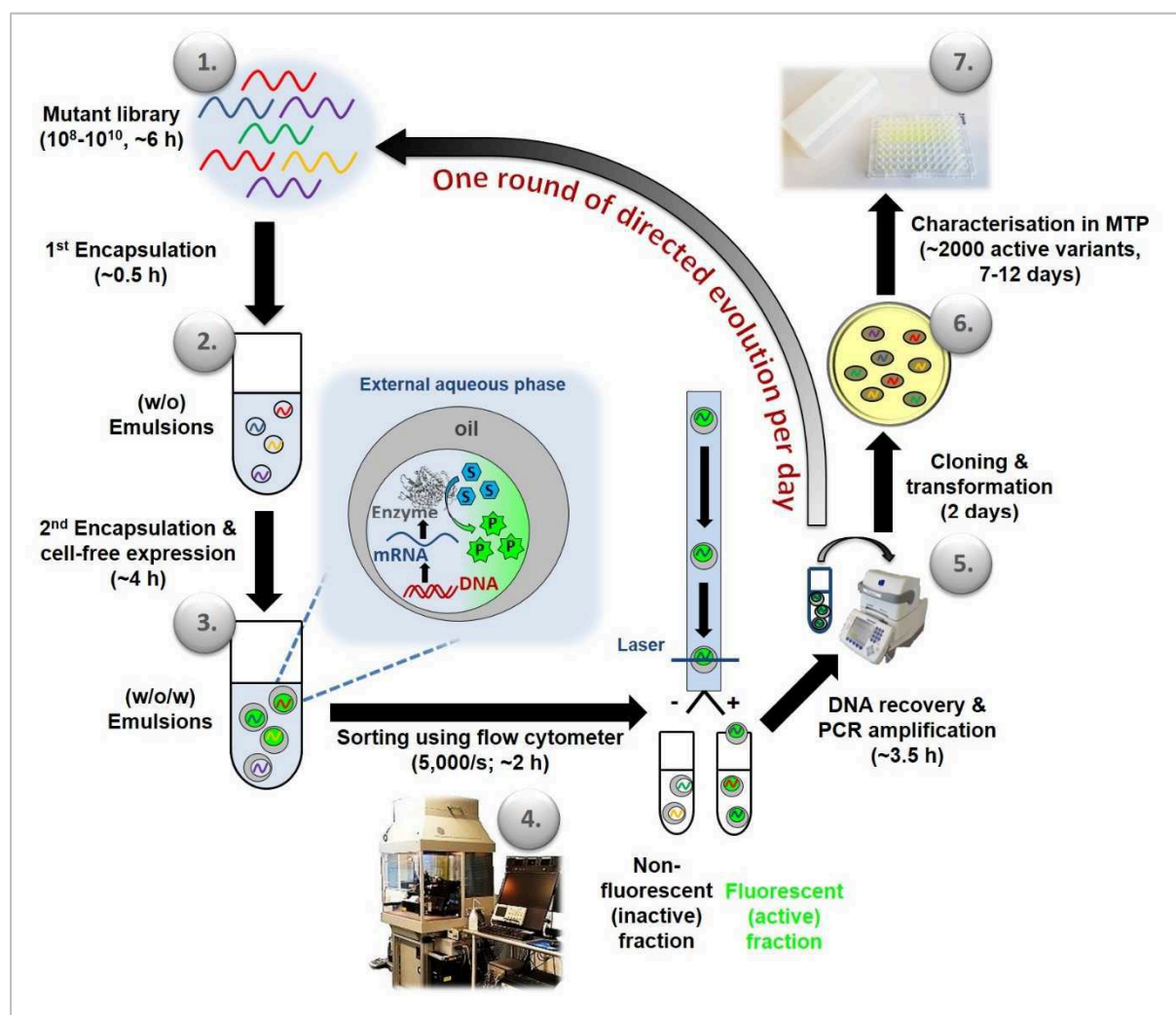
Figure 13: Principle of flow cytometer-based uHTS-IVC screening platform comprising seven steps: (1.) Mutant library generation using a linear DNA template (approx. 6 h), (2.) entrapment of mutant library in (w/o) single emulsions within 0.5 h, (3.) cell-free expression of mutant library and generation of (w/o/w) emulsions within 4 h, (4.) sorting of active enzyme variants within (w/o/w) emulsions using flow cytometer within 2 h, and (5.) DNA recovery from (w/o/w) emulsions and PCR gene amplification in 3.5 h. A whole round of directed evolution using flow cytometer-based uHTS-IVC platform (diversity generation, screening by flow cytometry, amplification) can be completed within 16 h. (6.) Cloning and transformation into expression host (2 days), and (7.) screening of up to 2,000 beneficial clones in MTP format and characterization of only a few beneficial enzyme variants (7-12 days).

3.4. Results

The result section is divided into three parts in order to introduce the novel uHTS-IVC-based flow cytometry screening technology platform. In the first section, the principle of the developed screening technology platform is described (3.3.1); subsequently in the second result section optimization steps for cell-free production of two model enzymes (CelA2-H288F cellulase variant and YmPhWT phytase) in (w/o/w) emulsions and subsequent analysis using flow cytometry are described. In detail, optimization steps include: a) optimization of cell-free phytase and cellulase production in tubes and selection of the model enzyme for establishment of the uHTS-IVC platform (3.4.1-3.4.4), b) optimization of (w/o/w) emulsion generation (3.4.5), c) optimization of fluorescence-based assays for enzyme detection and selection of the model substrate for establishment of the uHTS-IVC platform (3.4.6-3.4.7), d) optimization of enzyme production within (w/o/w) emulsion, screening and sorting by flow cytometer (3.4.8), e) analysis and sorting of enzyme model libraries with defined active-to-inactive DNA ratios with the flow cytometry (3.4.9). Finally, in the third result section the developed screening platform for directed protein evolution was validated by flow cytometry based screening of a cellulase epPCR gene library produced within (w/o/w) emulsions using cell-free extracts (3.4.10-3.4.11). Identified improved cellulase variants were purified and kinetically characterized (3.4.12-3.4.14).

3.4.1. Optimization of cell-free phytase production

3.4.1.1. Addition of chaperones

Cell-free YmPhWT synthesis was performed using *E.coli* protein synthesis kit (P1105-05, RiNA GmbH, Berlin, Germany) like stated in the protocol of the kit (EasyXpress Protein Synthesis Handbook 04/2008, QIAGEN GmbH, Hilden, Germany). For better folding of the protein a chaperone mix (DnaK, DnaJ, GroE, SVC100; RiNA GmbH, Berlin, Germany; 2.8 µl/50 µl *in vitro* reaction volume) was added additionally to the *in vitro* reaction mixture. For comparison, the *in vitro* reaction mixture containing plasmid DNA (pIX3.0RMT7-YmPhWT, 240 ng) was incubated (30°C) with and without chaperone mix (DnaK, DnaJ, GroE, SVC100, RiNA GmbH). Samples were analyzed after different incubation times (1.5, 6, and 22 h) using 4-MUP activity assay for 384-well plate and by SDS-PAGE. As negative control for 4-MUP activity assay the elongation factor (EF), supplied within the *in vitro* expression kit, was used. The sample containing the EF is expected to show no activity towards 4-MUP substrate, but

Figure 15: SDS-PAGE analysis of i*n vitro* expression of pIX3.0RMT7-YmPhWT and the elongation factor (**EF**, positive control; 32 kDa) with and without additional chaperone mix after 22 h incubation at 30°C and 37°C. **PL** represents the SDS-PAGE protein ladder (PageRuler™ Prestained Protein ladder, 10-170 kDA, #26616, Thermo Fisher Scientific, Waltham, MA, USA).

As cell-free phytase production in general seems to be very inefficient, an additional experiment was performed to confirm that the phytase enzyme itself has no inhibitory effect on the *in vitro* enzyme production. Figure 16 shows cell-free expression of the EF as well as samples containing only the *in vitro* expression mixture without template DNA, serving as negative control. The samples were incubated with and without addition of purified YmPhWT (0.725 µg/ml) to the *in vitro* expression mixture prior incubation (30°C, 0-1.5 h). The EF was successfully cell-free produced (Figure 16) so it can be concluded that no inhibitory effect of the phytase enzyme on the *in vitro* protein synthesis can be observed, even in high concentrations of purified YmPhWT (0.725 µg/ml). Samples analyzed after 4 h showed the same result like depicted in Figure 16 (data not shown).
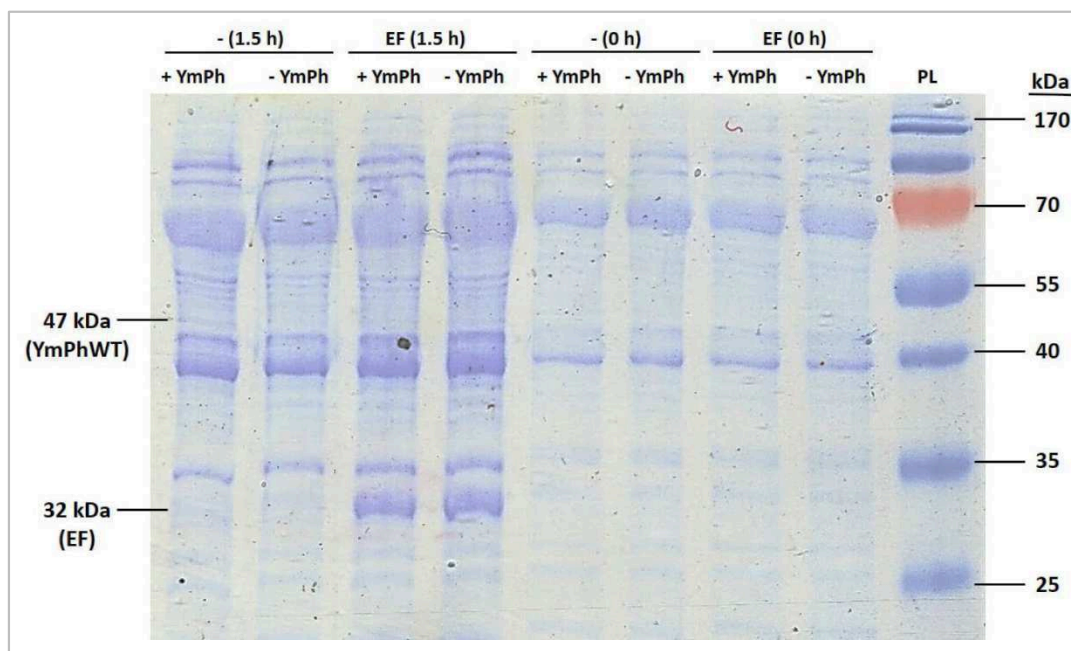
Figure 16: SDS-PAGE analysis of *in vitro* expression of the elongation factor (**EF**) as positive control for cell-free protein synthesis and the negative control (**-**) without template DNA after incubation (30°C, 0-1.5 h). **PL** represents the SDS-PAGE protein ladder (Thermo Fisher Scientific). The samples were incubated with (**+ YmPh**) and without (**- YmPh**) addition of purified YmPhWT (0.725 µg/ml) to the *in vitro* expression mixture prior incubation (30°C, 0-1.5 h).

3.4.1.2. *Optimization of incubation temperature, incubation time and amount of template DNA*

In order to determine the optimal *in vitro* expression conditions different incubation temperatures (15°C, 20°C, 25°C, 30°C) and amounts of template DNA (120-800 ng pIX3.0RMT7-YmPhWT) were tested. Samples were analyzed using 4-MUP activity assay in 384-well plates after different incubation times (1.5-24 h). Highest YmPhWT protein concentrations were obtained after incubation at 20°C for 24 h using 240 ng DNA (pIX3.0RMT7-YmPhWT) as template for the cell-free protein expression (Figure 17; Figure A1, appendix).
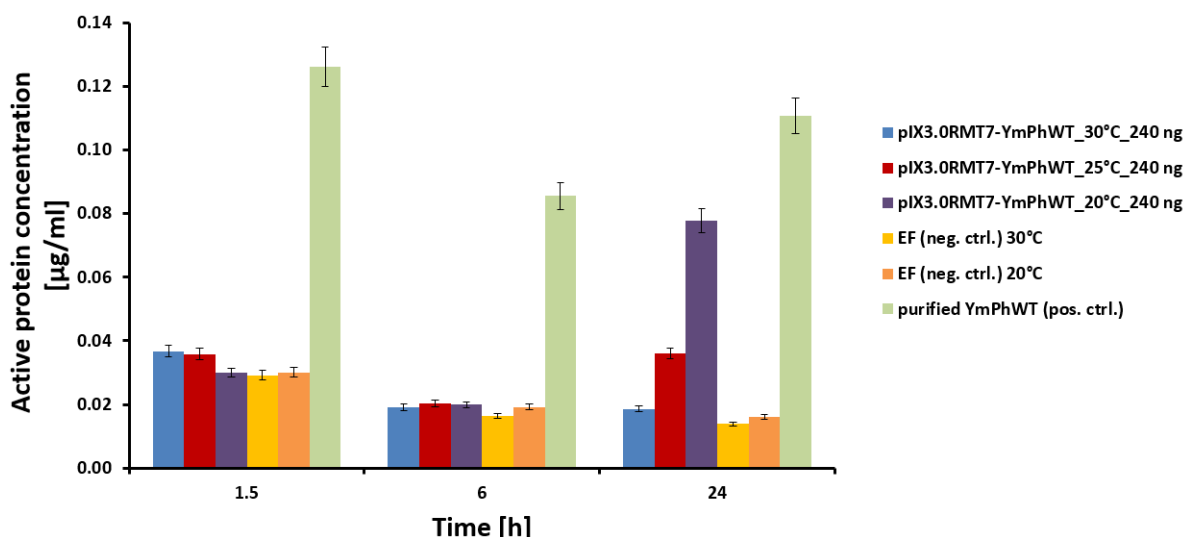
Figure 17: *In vitro* expression of pIX3.0RMT7-YmPhWT after incubation (20-30°C) with 240 ng of template DNA. Samples were analyzed with 384-well-based 4-MUP activity assay after 1.5 h, 6 h and 24 h. Calculated active protein concentrations [µg/ml] are shown on the y-axis and the x-axis represents the time in [h]. As experiments were performed only once a standard deviation of 5 % was added.

Variation of the DNA amount (120-800 ng pIX3.0RMT7-YmPhWT) showed the optimal template DNA amount is 240 ng, as an increase of template DNA >240 ng is not resulting in higher protein yields (Figure 18). Finally, optimal cell-free YmPhWT synthesis conditions are incubation at 20°C for 24 h using 240 ng DNA (pIX3.0RMT7-YmPhWT).
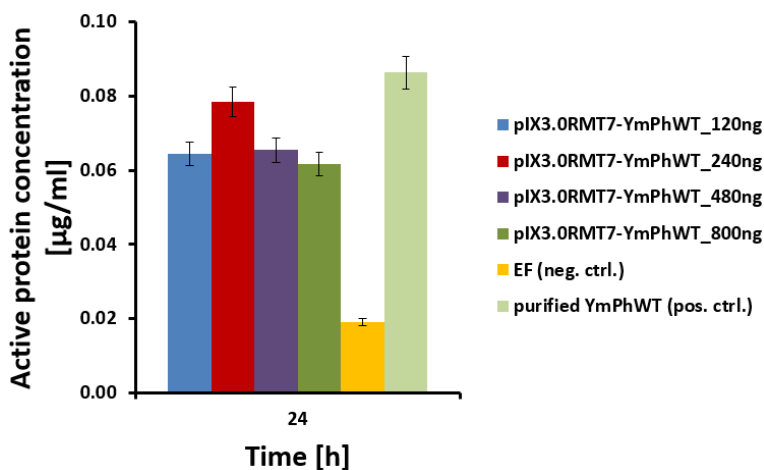
Figure 18: *In vitro* expression of YmPhWT after incubation (24 h, 20°C) with varied amounts of template DNA (120-800 ng pIX3.0RMT7-YmPhWT). Samples were analyzed with 384-well-based 4-MUP activity assay after 24 h. The calculated active protein concentrations [µg/ml] are shown on the y-axis and the x-axis represents the time in [h]. As experiments were performed only once a standard deviation of 5 % was added.

3.4.1.3. Selection of the in vitro expression vector

In order to achieve optimal enzyme yields upon cell-free YmPhWT production two different *in vitro* expression vectors (pNEB, pIXRMT7) were compared. Pre-tests were performed with *in vivo* expression vectors pET-22b(+) and pAIxtreme-5b in comparison to the *in vitro*

expression vector pIX3.0RMT7. The construct pIX3.0RMT7-YmPhWT showed slightly better performance after incubation for 5 h at 30°C (not optimized conditions) compared to the constructs pET-22b(+)-YmPhWT and pAlxtreme-5b-YmPhWT, which showed similar performance like the EF (negative control), (Figure 19).
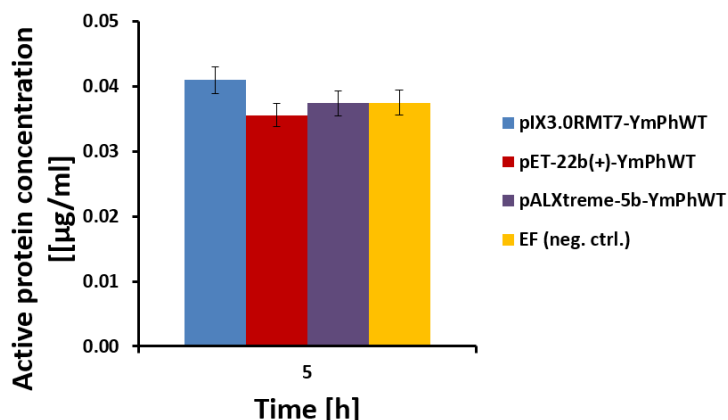
Figure 19: *In vitro* expression of pET-22b(+)-, pALXtreme-5b- and pIX3.0RMT7-YmPhWT after incubation (5 h, 30°C, 240 ng template DNA). Samples were analyzed with 384-well-based 4-MUP activity assay after 5 h. The calculated active protein concentrations [µg/ml] are shown on the y-axis and the x-axis represents the time in [h]. As experiments were performed only once a standard deviation of 5 % was added.

In order to select the optimal vector for *in vitro* phytase expression pNEB-YmPhWT and pIX3.0RMT7-YmPhWT constructs (240 ng DNA) were incubated (20°C, 24 h). Samples were analyzed using 4-MUP activity assay in 384-well plate. Highest YmPhWT protein concentrations were obtained for the construct pIX3.0RMT7-YmPhWT, so the vector pIX3.0RMT7 was selected as the vector of choice for all further cell-free experiments (Figure 20).
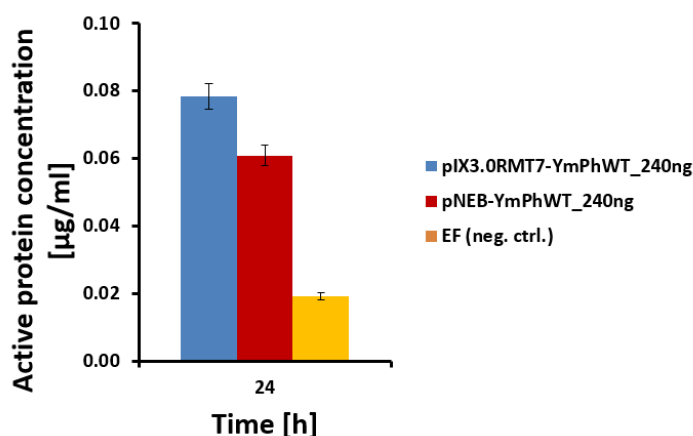
Figure 20: *In vitro* expression of pNEB- and pIX3.0RMT7-YmPhWT after incubation (20°C, 24 h, 240 ng template DNA). Samples were analyzed with 384-well-based 4-MUP activity assay after 24 h. Calculated active protein concentrations [µg/ml] are shown on the y-axis and the x-axis represents the time in [h]. As experiments were performed only once a standard deviation of 5 % was added.

Abbreviations and acronyms

4-MUC	4-Methylumbelliferyl-β-D-cellobioside
4-MUP	4-Methylumbelliferyl-phosphate
Amp	Ampicillin
bp	Base pair
BSA	Bovine serum albumin
CelA2(WT)	CelA2 cellulase (wildtype)
Da	Dalton
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
E. coli	*Escherichia coli*
EF	Elongation factor
epPCR	Error-prone polymerase chain reaction
EV	Empty vector
FDC	Fluorescein-di-β-D-cellobioside
HTS	High throughput screening
IVC	*In vitro* compartmentalization
LB	Luria-Bertani medium
MTP	Microtiter plate
PCR	Polymerase chain reaction
PDB	Protein data bank
PLICing	Phosphorothioate-based ligase independent gene cloning
RFU	Relative fluorescent units
SDM	Site-Directed Mutagenesis
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SSM	Site-Saturation Mutagenesis
SeSaM	Sequence saturation mutagenesis
TB	Terrific broth
uHTS	Ultrahigh throughput screening
YASARA	Yet another scientific artificial reality application
YmPh(WT)	*Yersinia mollaretii* phytase (wildtype)

Figure 22: SDS-PAGE analysis of *in vitro* expressed CelA2-H288F using pIX3.0RMT7- and pNEB-CelA2-H288F construct and the EF after incubation (20°C, 23 h). PL represents the SDS-PAGE protein ladder (Thermo Fisher Scientific).

3.4.2.2. Expression of CelA2-H288F from circular and linear DNA templates

In vitro cellulase expression using CelA2-H288F linear template (LT) DNA was performed in the same conditions which were optimal for circular DNA (20°C, 240 ng, Figure 21) and compared to the circular pIXRMT7-CelA2-H288F construct. Samples were analyzed after incubation (0-23 h) using optimized 4-MUC activity assay in 384-well plate. Highest CelA2-H288F protein concentrations were obtained after 6 h incubation for the circular construct pIX3.0RMT7-CelA2-H288F as well as for the linear CelA2-H288F template DNA (Figure 23). Already after 3 h incubation (20°C) a significant amount of CelA2-H288F enzyme was produced, indicating that *in vitro* production of cellulase from a linear CelA2-H288F DNA template is more efficient than observed for YmPhWT using circular template DNA (Figure 20, 23).
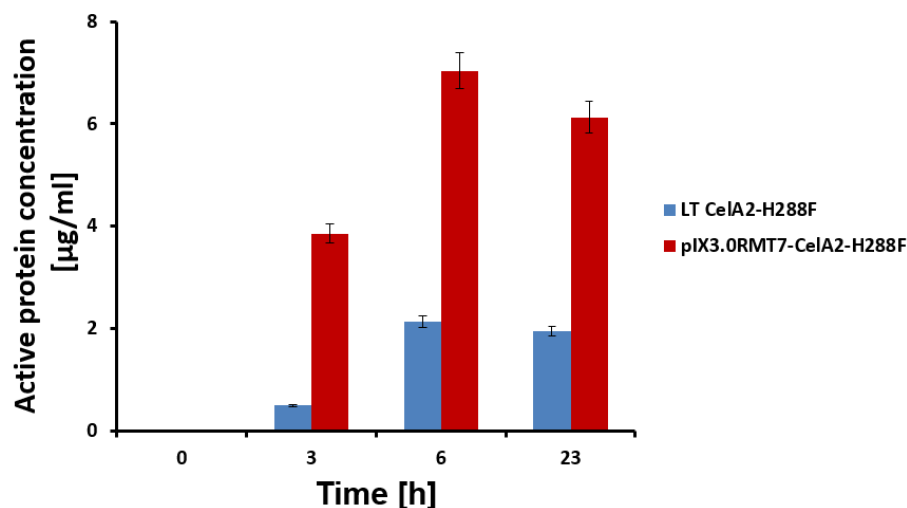
Figure 23: *In vitro* expression of CelA2-H288F using the pIX3.0RMT7-CelA2-H288F construct and CelA2-H288F linear template (LT) DNA after incubation for 0-23 hours (20°C, 240 ng template DNA). Samples were analyzed with 384-well-based 4-MUC activity assay after 0-23 h. Calculated active protein concentrations [µg/ml] are shown on the y-axis and the x-axis represents the time in [h]. As experiments were performed only once a standard deviation of 5 % was added.

The SDS-PAGE analysis showed no CelA2-H288F band at the expected size of 69 kDa, either due to the presence of the proteins in cell-free extract having the same size as CelA2 or the obtained CelA2-H288F protein concentration was too low to be visualized on the SDS gel (Figure 24).
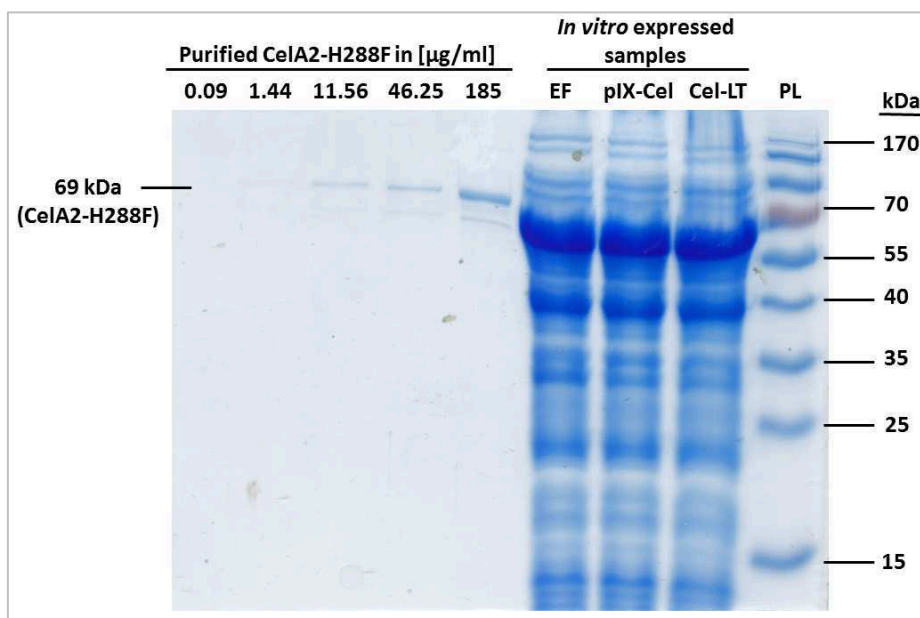
Figure 24: SDS-PAGE analysis of *in vitro* expressed CelA2-H288F using the circular pIX3.0RMT7-CelA2-H288F construct and CelA2-H288F linear template DNA after incubation (20°C, 23 h). PL represents the SDS-PAGE protein ladder (Thermo Fisher Scientific).

3.4.2.3. Optimization of incubation temperature and incubation time for cell-free expression of circular pIX3.0RMT7-CelA2-H288F

In order to achieve the maximum enzyme yields using cell-free cellulase expression, the incubation temperature (25°C-37°C) and time (1 h-3 h) were varied to determine the optimal *in vitro* expression conditions. As shown in Figure 25 the optimal incubation temperature for cell-free production of cellulase from circular DNA template (pIX3.0RMT7-CelA2-H288F, 240 ng) is 30°C, yielding ~22 µg/ml active cellulase. After 3 h of incubation (30°C) the highest active protein concentration was obtained (Figure 25). The negative control, harboring a mutation at position E580Q, resulting in a complete loss of cellulase activity, shows no conversion of 4-MUC substrate at any time point and was used as negative control for all later cell-free experiments (Figure 25).
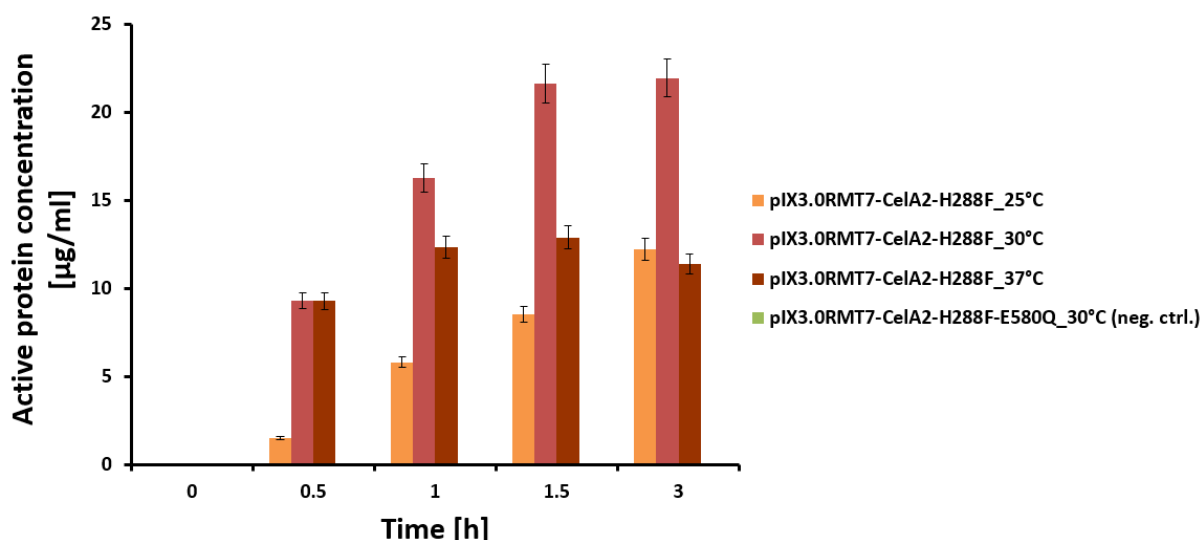
Figure 25: Temperature profiles for cell-free CelA2-H288F production using circular template DNA after incubation for 0-3 hours (20°C, 240 ng pIX3.0RMT7-CelA2-H288F template DNA). Samples were analyzed with 384-well-based 4-MUC activity assay after 0-3 h. Calculated active protein concentrations [µg/ml] are shown on the y-axis and the x-axis represents the time in [h]. As experiments were performed only once a standard deviation of 5 % was added.

3.4.2.4. Optimization of incubation temperature, incubation time and amount of template DNA for cell-free expression of linear CelA2-H288F

The optimization of the amount of CelA2-H288F LT DNA for cell-free cellulase synthesis after incubation (20°C, 0-24 h) is illustrated in Figure 26. Increasing LT DNA concentrations (120-480 ng) result in an increase of the active protein concentration (Figure 26). Incubation of samples longer than 7 h results in no further increase in active protein concentration (Figure 26). The active protein concentration for the sample using 480 ng CelA2-H288F LT DNA is comparable to the sample using 240 ng circular template DNA

(pIX3.0RMT7-CelA2-H288F), indicating that cell-free cellulase production from circular template DNA is more efficient than from LT DNA (Figure 25, 26).
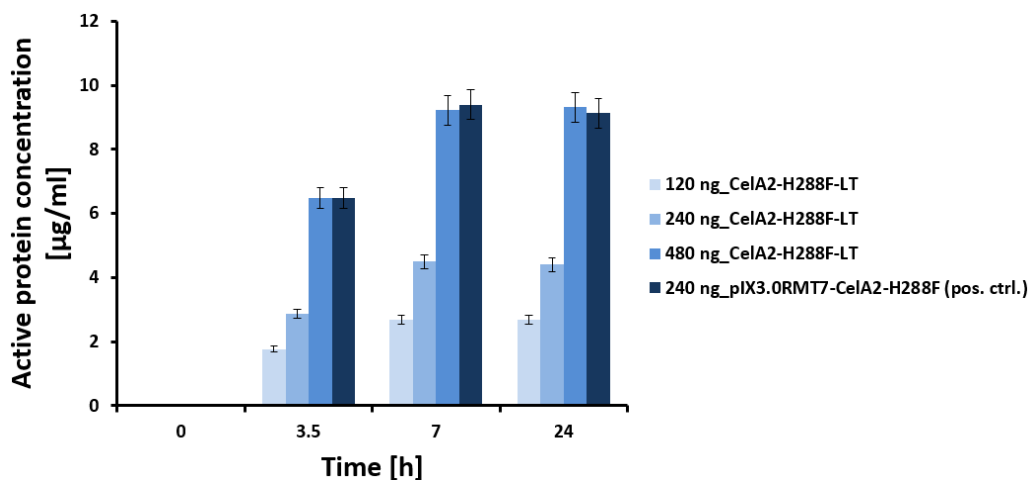
Figure 26: Optimization of incubation time and CelA2-H288F LT DNA amount for cell-free cellulase synthesis after incubation (20°C, 0-24 h). The construct pIX3.0RMT7-CelA2-H288F (240 ng) was used as positive control. Samples were analyzed with 384-well-based 4-MUC activity assay after 0-24 h. Calculated active protein concentrations [µg/ml] are shown on the y-axis and the x-axis represents the time in [h]. As experiments were performed only once a standard deviation of 5 % was added.

In order to further increase the enzyme yield using cell-free cellulase expression, the incubation temperature (20°C-37°C) was varied to determine the optimal *in vitro* expression temperature. As shown in Figure 27, the optimal incubation temperature for cell-free production of cellulase from LT DNA (CelA2-H288F, 240 ng) is 25°C. After 4 h incubation (25°C) maximum active protein yield (~15 µg/ml) was reached for *in vitro* produced CelA2-H288F from LT DNA (Figure 27), so in all later experiments with cell-free cellulase production from LT DNA samples were incubated for 4 h at 25°C using 240 ng LT DNA.
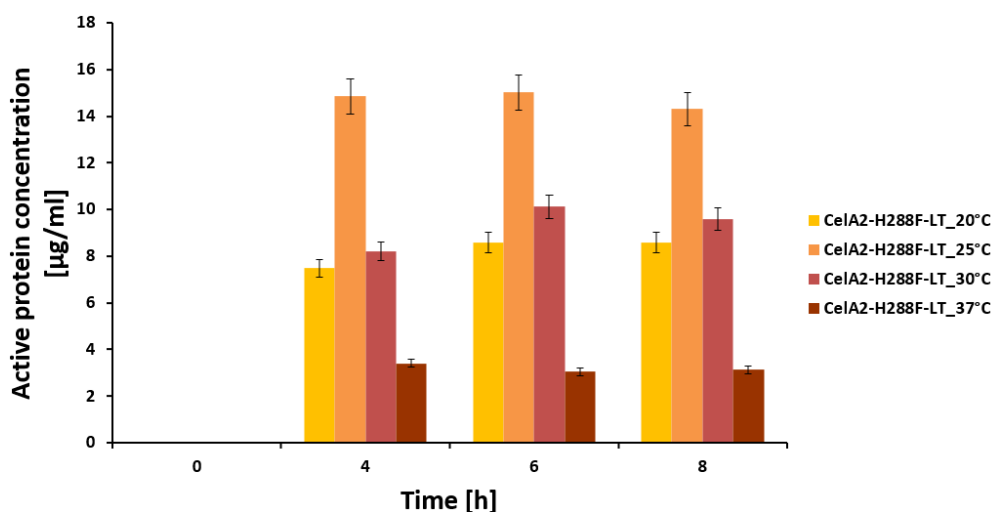
Figure 27: Optimization of incubation temperature for cell-free cellulase synthesis using CelA2-H288F LT DNA (240 ng) after incubation for 0-8 h. Samples were analyzed with 384-well-based 4-MUC activity assay after 0-8 h. Calculated active protein concentrations [µg/ml] are shown on the y-axis and the x-axis represents the time in [h]. As experiments were performed only once a standard deviation of 5 % was added.

3.4.3. *Optimization of cell-free phytase and cellulase expression by removal/addition of N-terminal His-Tag*

For further optimization of *in vitro* expression levels in order to achieve higher protein yields using the cell-free protein production system the influence of presence and absence of an N-terminal Histidin-tag (His-tag) was investigated. The two genes of interest *cellulase* and *phytase* were subjects to the addition or removal of an N-terminal His-Tag, and the obtained active protein concentrations were compared by the use of SDS-PAGE and fluorescence-based activity assays (4-MUC for cellulase and 4-MUP activity assay for phytase).

In case of cellulase, the His-tag was removed from the original construct (pIX3.0RMT7-CelA2-H288F-no His) and protein amounts obtained in cell-free reaction were compared to the original construct harboring N-terminal His-tag (pIX3.0RMT7-CelA2-H288F). The original phytase construct (pIX3.0RMT7-YmPhWT) did not contain a His-tag, so in this case the His-tag and an additional ATG was added in front of the N-terminus of the phytase gene. Cell-free phytase production using the original construct without His-tag was then compared to active phytase yield achieved with the newly generated construct (pIX3.0RMT7-YmPhWT-HIS-ATG) harboring the His-tag in front of the phytase gene sequence.

3.4.3.1. *Cell-free expression of YmPhWT construct with His-tag*

The newly generated construct harboring N-terminal His-tag was transformed into *E.coli* BL21 Gold(DE3)lacI^{Q1} cells and expressed *in vivo* to investigate the possible influence of His-tag on phytase activity. Additionally, the protein expression levels were compared between the original construct (pIX3.0RMT7-YmPhWT) and the newly generated construct (pIX3.0RMT7-YmPhWT-HIS-ATG) using SDS-PAGE and 4-MUP activity assay. Successful *in vivo* phytase production was confirmed by SDS-PAGE analysis, showing the phytase protein band at 47 kDa in the gel (data not shown), as well as with 4-MUP activity assay, confirming the functionality and catalytic activity of phytase in both constructs (Figure 28). Figure 28 shows higher amount of the active protein in pIX3.0RMT7-YmPhWT-HIS-ATG (~1.6 µg/ml) compared to protein amount in the original construct (~1 µg/ml, pIX3.0RMT7-YmPhWT).
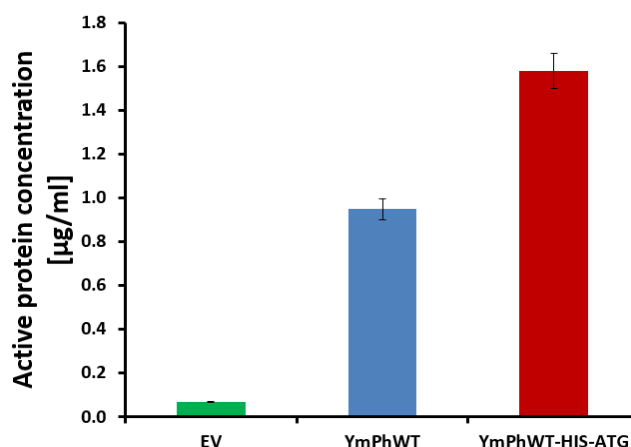
Figure 28: Activity profiles of *in vivo* expressed phytase constructs (pIX3.0RMT7-YmPhWT and pIX3.0RMT7-YmPhWT-HIS-ATG) and the empty vector (EV). All samples were diluted 1:50 with assay buffer prior the measurement with 4-MUP activity assay. Calculated active protein concentrations [µg/ml] are shown on the y-axis and the x-axis represents the different pIX3.0RMT7-constructs used for enzyme expression. As experiments were performed only once a standard deviation of 5 % was added.

Calculated protein concentrations from fluorescence measurements using 4-MUP activity assay with *in vitro* expressed pIX3.0RMT7-YmPhWT constructs after 2-24 h are shown in Figure 29. Both constructs without and with the modification of an additional His-tag show a similar increase in active protein concentration after 24 h (Figure 29). Therefore, it can be concluded that the insertion of an N-terminal His-tag into the original phytase construct is not beneficial for increasing the protein yield upon cell-free phytase expression, as *in vitro* phytase expression level is not improved by adding the His-tag (Figure 29). The significant activity in the negative control probably results from internal *E. coli* phosphatases present in the cell-free *E. coli* extract that are able to convert 4-MUP substrate.
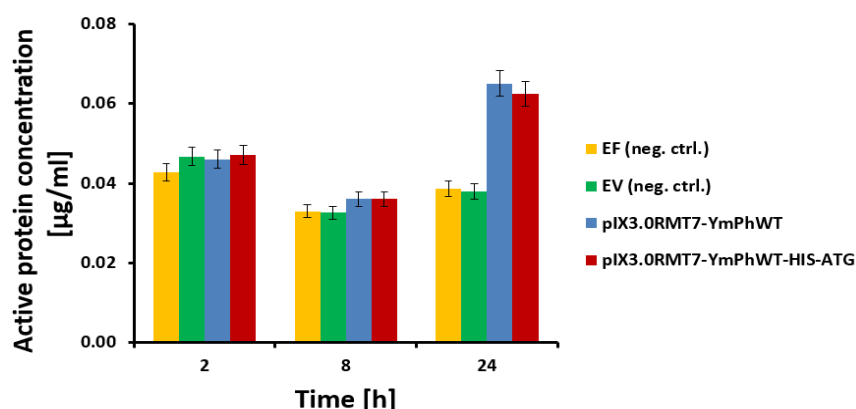
Figure 29: Analysis of cell-free expressed pIX3.0RMT7-YmPhWT constructs without and with His-tag (24 h, 20°C, 240 ng template DNA) with calculated active protein concentrations from 4-MUP activity assay measurements. The x-axis represents the time in [h] and the y-axis shows the active protein concentration in [µg/ml]. As experiments were performed only once a standard deviation of 5 % was added. EF is the cell-free expressed elongation factor, and EV is the empty vector construct without YmPhWT gene, both serving as negative controls for *in vitro* phytase production.

Abstract

Abstract

Directed evolution is a powerful algorithm for tailoring industrially important biocatalysts. Methodological and conceptual advancements in directed evolution in the past decades enabled engineering of enzymes towards unnatural substrates with catalytic efficiencies beyond natural enzymes. Despite all progress in the field of directed evolution the exploration of generated sequence space and acceleration of iterative rounds of mutagenesis and screening is limited by throughput of commonly employed screening technologies. Development of ultrahigh throughput screening (uHTS) represents the main challenge in today's state-of-the-art directed evolution and ideally, in contrast to medium screening formats, enables rapid analysis of up to 10^7 events per hour using flow cytometry. The possibility to screen millions of variants within one day using uHTS offers the opportunity to generate libraries of high mutational load to successfully identify and investigate structure-function relationships or combinatorial effects within directed evolution campaigns. The latter is up to now unobtainable employing standard screening formats due to cost and time limitations of directed evolution campaigns. Most of reported uHTS screening systems employ *in vivo* screening where the bacterial host is used as carrier of enzyme variants. The main limitation of the cell-based systems is the low transformation efficiency resulting in reduced gene diversity. In this study we developed an uHTS platform in which cells were replaced by cell-free transcription-translation machinery encapsulated within artificial cell-like compartments (water-in-oil-in-water (w/o/w) emulsions). Each emulsion compartment encapsulates a single gene copy of a highly diverse gene library (10^8). Upon cell-free transcription and translation of genes, encoding active enzyme variants, compartments are labelled by conversion of fluorogenic substrate into fluorescent product and are analyzed by flow cytometry. *In vitro* compartmentalization (IVC) enables genotype-phenotype linkage and isolation of genes encoding for active enzyme variants from an excess of genes encoding inactive enzymes. The uHTS screening platform was developed for cellulase (CelA2) using *in vitro* production of CelA2 enzyme within (w/o/w) emulsions and fluorescein-di-β-D-cellobioside (FDC)-fluorescein substrate-product pair (λ_{ex} 494 nm/λ_{em} 516 nm). The developed uHTS cell-free compartmentalization platform was validated by screening of a random epPCR cellulase mutagenesis library and yielded an improved cellulase variant within only two weeks. The identified variant CelA2-H288F-M1 (N273D/N468S) showed 5-fold reduced K_M value and 13.3-fold increased specific activity (220.60 U/mg) compared to cellulase wild type (16.57 U/mg). The uHTS *in vitro* based platform was employed to study differences between three focused mutagenesis approaches (iterative site-saturation mutagenesis, multiple site-saturation mutagenesis, and simultaneous saturation mutagenesis of four amino acid positions *i.e.* OmniChange) in order to explore combinatorial effects. As an outcome the best performing variant contained only two amino acid substitutions (F288 and Q524) and showed 8-fold improvement in specific activity compared to parent.

In order to expand the uHTS platform on other enzymes, in this study *in vitro* based screening assay for *Yersinia mollaretii* phytase wildtype (YmPhWT) was optimized. Due to discrepancies between pH optimum for cell-free expression machinery (pH 7) and acidic pH optimum of YmPhWT, optimization of YmPhWT through a directed evolution campaign was performed. The goal of the campaign was to broaden the activity of the phytase towards neutral pH. Phytase libraries were generated *via* epPCR and subsequent simultaneous site-saturation mutagenesis based on identified beneficial positions in the epPCR library screening. The variants YmPh-M10 and -M16, showing 7-fold improved specific activity at pH 6.6 towards the model substrate 4-methylumbelliferylphosphate (4-MUP) compared to YmPhWT, were isolated.

The flow cytometer-based uHTS-IVC technology platform drastically reduces time requirements for one round of directed evolution from several months to only one day and enables an efficient screening of mutant libraries (10^{10}) with high diversity covering a significant portion of the generated sequence space in a time efficient manner. Consequently, the flow cytometer-based uHTS-IVC technology platform offers an efficient solution to tackle present challenges of directed evolution, like the performance of multiple, iterative rounds of diversity generation and screening of mutant libraries with unlimited diversity, in order to successfully identify beneficial variants with altered selectivity, specificity or activity. In addition, the flow cytometer-based uHTS-IVC technology platform has an impressive potential to study demanding/challenging scientific questions *e.g.* the exploration of combinatorial effects and answering fundamental questions on structure-function relationships within biocatalysts. Additionally, the uHTS technology platform can - from our point of view - be adapted to other enzyme classes since it uses a widely applicable fluorescence sorting of cell-free expressed active enzyme variants within (w/o/w) emulsions compartments.

phytase) which is more suitable to achieve the goal to perform one directed evolution experiment within one scientific day. Therefore, the selected model enzyme for the further development of the in uHTS-IVC platform will be cellulase (CelA2-H288F).

3.4.5. Optimization of (w/o/w) emulsion generation

Essential characteristics of *in vitro* protein production in (w/o/w) emulsions used for directed evolution experiments are (1) compatibility of cell-free expression mixture and substrate/product with emulsion building components (*i.e.* oil, surfactants), (2) stability throughout the selection procedure (including sorting), (3) homogeneity in size and shape, and (4) monodispersity (not more than one inner aqueous phase per (w/o/w) emulsion droplet). For optimization of (w/o/w) emulsion generation three different techniques (stirring, homogenizing, and membrane extrusion) and three currently used emulsification systems (ABIL EM 90/CMC, decane/Tween, Span/Tween) were compared and further optimized in regard to the four above mentioned criteria (Table 2). Generated (w/o) and (w/o/w) emulsions were analyzed by fluorescence microscopy and flow cytometer.

For generation of stable (w/o/w) emulsions with a diameter of ~10 μm different emulsification speeds were tested for stirring and homogenization method. Additionally various surfactants (Tween 20, Tween 80, Span 60, Span 80, ABIL EM 90) were tested in different combinations and concentrations in combination with stirring, homogenizing and extrusion method.

Table 2: Different investigated emulsification methods for (w/o/w) emulsion generation.

Emulsion components			Emulsification technique	Application	References
Internal aqueous phase	Oil phase	External aqueous phase			
Cyt P450 BM3 cells in activity buffer (100 mM Tris-HCl pH 7.5) using 7-benzoxy-3-carboxy coumarin (BCC Acid) as a fluorogenic substrate for detection of Cyt P450 BM3 activity	2.9 % (w/w) ABIL EM90 in light mineral oil	1.5 % (w/v) carboxymethyl-cellulose and 2 % (w/v) Tween 20 in PBS	Homogenizer	Whole-cell system	Blanusa, PhD Thesis, 2009 (Blanusa, 2009)
Beta-galactosidase library in *in vitro* translation reaction mix	1 % (w/v) Span 60 and 1 % (w/v) cholesterol in decane	0.5 % (w/v) Tween 80 in PBS buffer	Extrusion	Cell-free system	Mastrobattista et al., 2005 (Mastrobattista *et al.*, 2005)
Cellulase library in cell-free expression mixture	**2 % (w/w) Span 80 and 0.4 % (w/w) Tween 80 in mineral oil**	**0.5 % (w/v) Tween 80 in PBS buffer**	**Extrusion**	**Cell-free system**	**This work**

The (w/o/w) emulsion generation by **homogenization method** using Miccra D-1 homogenizer (ART Prozess- & Labortechnik GmbH & Co. KG, Müllheim, Germany) and the protocol reported by Blanusa (Blanusa, 2009) applying ABIL EM 90/CMC resulted in polydisperse (w/o/w) emulsions of various sizes (1-20 µm) harboring more than one inner aqueous compartment (Figure 35C-D). As shown in Table 3, an increase in the emulsification speed (1,300-18,000 rpm) and an increase in emulsification time (1-5 min) resulted in a significant decrease in size of the generated (w/o) emulsions. In summary, all tested different emulsification speeds and times for generation of (w/o) and (w/o/w) emulsions using Miccra D-1 homogenizer (ART Prozess- & Labortechnik GmbH & Co. KG) did not result in monodisperse (w/o/w) emulsions (Table 3, Figure 35C-D).

Table 3: Correlation of droplet size with emulsification time and speed during (w/o) emulsion preparation by stirring and homogenizing technique. Values represent mean values of 50 measurements.

Speed [rpm]	Time [min]	Method	Mean (w/o) emulsion diameter [µm]
1300	**15**	**stirring**	**7.50**
5000	1	homogenizing	3.36
5000	5	homogenizing	2.64
7500	1	homogenizing	2.25
7500	5	homogenizing	1.38
10000	5	homogenizing	1.27
14500	5	homogenizing	1.03
18000	5	homogenizing	0.95

Therefore, **stirring method** as well as various **combinations of stirring and homogenization method** were tested using different emulsification speeds for generation of (w/o) and (w/o/w) emulsions (Table 4). The (w/o/w) emulsions obtained solely by stirring method applying ABIL EM 90/CMC emulsification system were not soluble in PBS buffer, as mainly (w/o) emulsions and no or only few (w/o/w) emulsions were generated (Table 4). Solubility of (w/o/w) emulsions in PBS buffer is a key requirement for flow cytometry analysis.

As shown in Table 4, suitable (w/o/w) emulsions with an average size of 5.3 μm were generated using a combination of stirring (1,300 rpm, 1st emulsification) and homogenizing (7,500 rpm, 2nd emulsification), which were soluble in PBS buffer and contained only one inner aqueous compartment as shown in Figure 35A-B. An increase of homogenization speed during the 2nd emulsification step resulted in a significantly lower fraction of (w/o/w) emulsions with a reduced average diameter (<4 μm) and a high fraction of (w/o) emulsions. These results indicate that speeds >7,500 rmp for the 2nd emulsification step are not suitable for generation of a high fraction of (w/o/w) emulsions (Table 4).

Table 4: Correlation of droplet size with emulsification time and speed during (w/o/w) emulsion preparation using stirring and homogenizing technique. Values represent mean values of at least 15 measurements.

Speed [rpm] 1st/2nd emulsification	Time [min] 1st/2nd emulsification	Method 1st/2nd emulsification	Mean (w/o/w) emulsion diameter [μm]
800/800	15/5	Stirring/stirring	-
1,300/1,300	15/5	Stirring/stirring	9.00
1,300/7,500	**15/3**	**Stirring/homogenizing**	**5.31**
1,300/10,000	15/3	Stirring/homogenizing	3.66
1,300/14,500	15/3	Stirring/homogenizing	3.47

In order to obtain a higher fraction of (w/o/w) emulsions the ratio of oil phase (W1) to water phase (W2), which was 1:1 in previous experiments, was optimized (Table 5). The highest fraction (30 %) of (w/o/w) emulsions was obtained using an oil phase (W1) to water (W2) phase ratio of 5:1 (Table 5).

Table 5: Ratio of oil phase (W1) to water phase (W2) during (w/o/w) emulsion generation using a combination of stirring and homogenizing technique. The inner aqueous phase was kept constant at 100 µl aqueous phase (e.g. PBS buffer). Emulsification was performed at optimized conditions (1,300 rpm stirring for 1st emulsification step and homogenizing at 7,500 rpm for 3 minutes for 2nd emulsification step).

Volume oil phase W1 [µl]	Volume aqueous phase W2 [µl]	Ratio W1:W2	Method 1st/2nd emulsification	Fraction of (w/o/w) emulsion [%]
1,000	1,000*	1:10	Stirring 15 min /homogenizing	2
200	**1,000**	**1:5**	**Stirring 5 min /homogenizing**	**30**
200	1,000	1:5	Stirring 15 min /homogenizing	10
400	1,500	1:3.75	Stirring 5 min /homogenizing	5
1,000	500**	1:1	Stirring 5 min /homogenizing	4
1,000	200	5:1	Stirring 5 min /homogenizing	0

*Take 100 µl of stirred W1 mixture and add 1000 µl W2 solution
**Take 500 µl of stirred W1 mixture and add W2 solution

Furthermore, the retention of fluorescent dyes (4-MU, fluorescein) within the (w/o/w) emulsion were tested using optimized conditions for emulsion generation by combination of stirring and homogenizing (100 µl inner aqueous phase, 200 µl W1, 5 min stirring at 1,300 rpm; addition of 1,000 µl W2, 3 min homogenizing at 7500 rpm). Fluorescence microscopic analysis showed the fluorescent product 4-methylumbelliferone (1 mM) is immediately leaking out of the emulsion droplets, generating a strong blue background fluorescence (λ_{ex} 365 nm, λ_{em} 445 nm) in the external aqueous phase, but no fluorescence within the inner aqueous phase of the (w/o/w) emulsions was observed (data not shown).

For the fluorescent product fluorescein (6 µM) (w/o/w) emulsions were observed exhibiting fluorescence (λ_{ex} 492/495 nm, λ_{em} 515/525 nm) in the inner aqueous phase of the (w/o/w) emulsions. In addition, a green background fluorescence in the external aqueous phase was observed, which was quickly bleached during fluorescence microscopic observation, indicating either diffusion of the dye into the oil and external aqueous phases or a destruction of the emulsions during the homogenization step (data not shown).

In order to further investigate the background fluorescence in the surrounding buffer, diffusion tests with the highly hydrophilic fluorescent dye 8-Hydroxypyrene-1,3,6-trisulfonic acid trisodium salt (HPTS), were performed to elucidate if the homogenization step breaks the emulsions, resulting in the observed background fluorescence. Results showed (w/o/w) emulsions incubated in PBS buffer containing high concentration of HPTS (4 mM HPTS in PBS buffer) did not reveal any fluorescence in the inner aqueous phase or oil phase (λ_{ex} 454 nm, λ_{em} 520 nm) proofing HPTS does not diffuse from the outer aqueous phase into the inner aqueous phase or to the oil phase (data not shown; similar result

1. General introduction

1.1. Directed evolution and gene diversity generation

Directed evolution is a powerful tool in protein engineering to tailor enzyme properties for the requirements of industrial or therapeutic applications (*e.g.* in chemical, pharmaceutical and food industry) and to elucidate structure-function relationships of proteins (Roccatano *et al.*, 2006; Wong *et al.*, 2006). The concept of directed evolution was introduced in 1967 by Spiegelman *et al.,* performing "evolution in a test tube" (Mills *et al.*, 1967), mimicking the process of Darwinian evolution with a few differences: in contrast to the natural Darwinian evolution, directed evolution requires an outside intelligence, not blind chance, and does not create brand new species (macroevolution), but only improvements within a species (microevolution), during a process that does not take millions of years, but happens rapidly (Stemmer *et al.*, 2003). Since then, directed evolution has been successfully applied in various fields of industry ranging from biotransformation for chemical synthesis (Jaeger *et al.*, 2004), biosensors (Doi *et al.*, 1999), bioremediation (Furukawa, 2003), and protein structure function relationships (Lee *et al.*, 2000), to medical and pharmaceutical applications in form of vaccines (Locher *et al.*, 2004) or therapeutic proteins (Roccatano *et al.*, 2006). Up to date, directed evolution matured to the point where it plays a key role in industrial biocatalyst development. In order to overcome the limitations of natural enzymes, such as low catalysis of non-natural substrates, low stability (towards temperature, pH, and oxidative agents), poor activity in non-aqueous media, requirements for expensive cofactors in protein engineering various strategies have been established to tailor enzyme properties for the requirements of industrial applications (Farinas *et al.*, 2001; Wong *et al.*, 2006). Upon generation of biocatalysts with novel or improved functions and properties, directed evolution can be either used for improvement of a specific, already exhibited enzyme property (*i.e.* activity, thermal stability, enantioselectivity) or for the combination of properties that are usually not found together in natural enzymes (Matsuura *et al.*, 2006).

In contrast to rational design, directed evolution can be applied to newly discovered enzymes with unknown crystal structures or mechanistic information, *e.g.* key amino acid positions, as it does not require an understanding of structure-function relationships of the target protein for its optimization (Wong *et al.*, 2006). A typical directed evolution experiment comprises three major steps: (I) starting from diversity generation, where mutations are randomly introduced in the gene of interest, to transformation and expression

in combination with extrusion technique entrapping fluorescein sodium salt (60 µM) showed (w/o/w) emulsion harboring only one inner aqueous phase and had a size of ~5 µm (Figure 31F).

Diffusion tests with the highly hydrophilic fluorescent dye 8-Hydroxypyrene-1,3,6-trisulfonic acid trisodium salt (HPTS), were performed in order to investigate if the background fluorescence originates from broken emulsions. Results showed (w/o/w) emulsions incubated in PBS buffer containing HPTS (4 mM HPTS in PBS buffer) did not exhibit fluorescence (λ_{ex} 454 nm, λ_{em} 520 nm) proofing HPTS does not diffuse from the buffer into the inner aqueous phase or to the oil phase (Figure 31A). Entrapment of HPTS dye within the inner aqueous phase resulted in (w/o/w) emulsions showing strong fluorescence in the inner aqueous phase but no fluorescence in the external aqueous phase or the surrounding PBS buffer (Figure 31B-C). If HPTS dye was added in the W2 solution fluorescence was observed in the outer aqueous phase and in the surrounding PBS buffer, as expected, but not in the inner aqueous phase, which clearly shows the (w/o/w) emulsions are not destroyed and stay intact during the process of extrusion (Figure 31D).

In addition, (w/o/w) emulsions were prepared by extrusion method using Span/Tween surfactant combination, incubated for 2 hours in PBS buffer containing fluorescein-di-β-D-cellobioside (FDC, 30 µM, Biomol GmbH) and purified CelA2H288F to investigate whether the converted fluorescein product diffuses into the inner aqueous phase of the emulsion (Figure 31G-H). A second sample was prepared using W2 solution containing 200 µl of FDC (30 µM)/CelA2-H288F enzyme mix and incubated in PBS buffer (2 h) to see if diffusion of the converted fluorescein product into the inner aqueous phase of the emulsion occurs (Figure 31I-J). Both samples were analyzed by fluorescence microscopy and revealed no diffusion of the converted fluorescein product from the outer aqueous phases into the inner aqueous phase of the emulsion (Figure 31G-J).

So finally, **membrane extrusion method** in combination with **Span/Tween emulsification system** seems to be a **suitable method for IVC using (w/o/w) emulsions**.
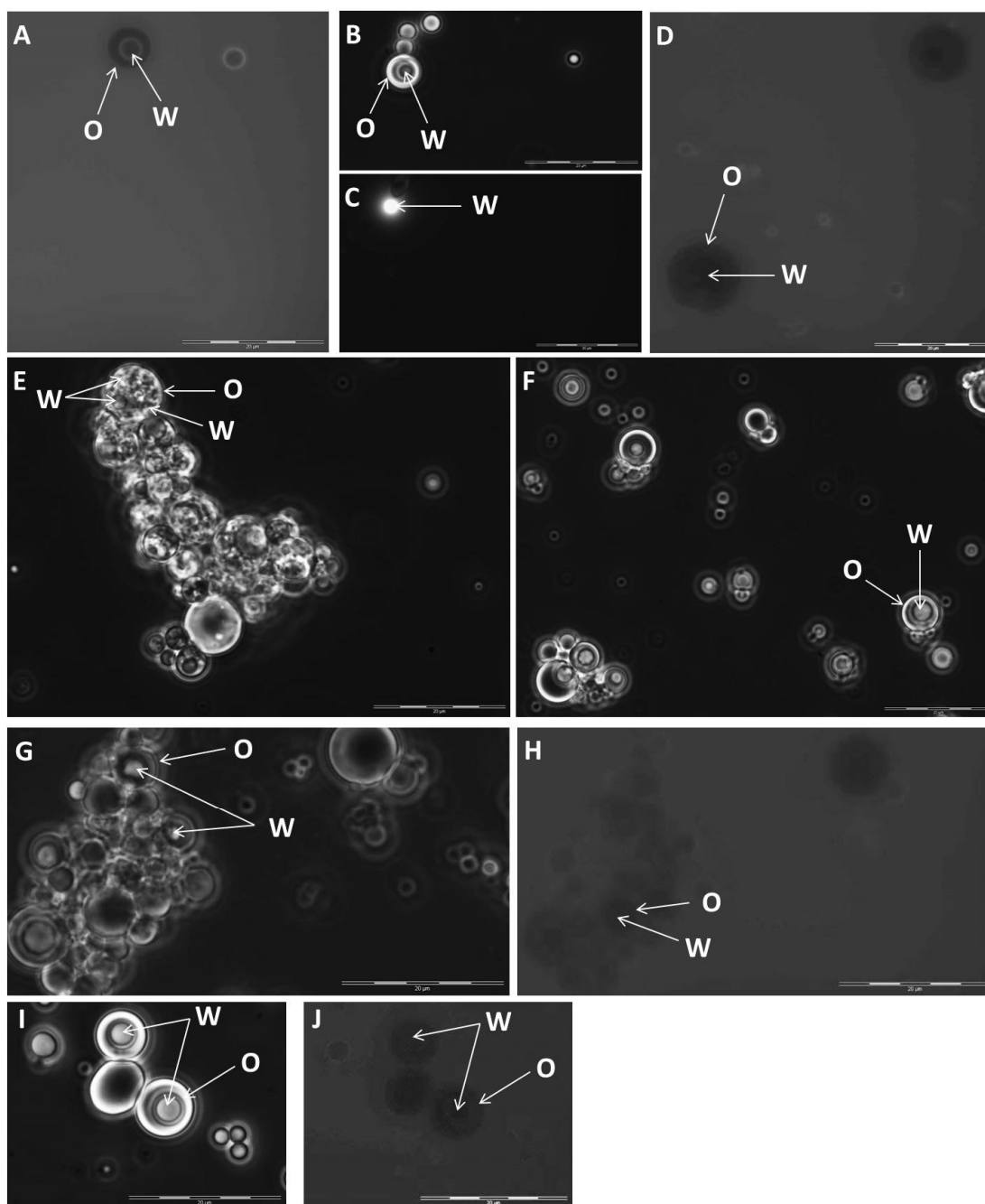
Figure 31: Diffusion tests performed with (w/o/w) emulsions generated by membrane extrusion with the Tween/Span emulsification system using different fluorescent dyes (HPTS, fluorescein) are illustrated in the microscopic pictures (100 x magnification) using light filed (**B, E, F, G, I**) and fluorescence field (**A, C, D, H, J**). (W/o/w) emulsions containing (**A**) HPTS in the surrounding PBS buffer (extrusion), (**B/C**) HPTS in the inner aqueous phase (extrusion), (**D**) HPTS in the W2 phase (extrusion), (**E**) Emulsions generated using ABIL EM 90/CMC emulsification system entrapping fluorescein sodium salt (60 µM) containing many inner aqueous phases (extrusion), (**F**) Emulsions generated using Span/Tween emulsification system in combination with extrusion technique entrapping fluorescein sodium salt (60 µM) showing (w/o/w) emulsions harboring only one inner aqueous phase with a size of ~5 µm, (G/H) (w/o/w) emulsions prepared by extrusion method in combination with Span/Tween emulsification system after incubation in PBS buffer (2 h) supplemented with FDC substrate (30 µM) and purified CelA2-H288F enzyme to see if diffusion of the converted fluorescein product takes place into the inner aqueous phase of the emulsion, (**I/J**) (w/o/w) emulsions prepared by extrusion method in combination with Span/Tween emulsification system with W2 solution containing 200 µl of FDC (30 µM) CelA2-H288F enzyme mix after incubation in PBS buffer (2 h) to see if diffusion of the converted fluorescein product takes place into the inner aqueous phase of the emulsion. **W** indicates the inner aqueous phase of the emulsion; **O** indicates the oil phase (**E**). The scale bar represents a length of 20 µm.

For further optimization of the extrusion technique using the Span/Tween emulsification system, different W1 solutions (W1.1: 2 % (w/w) Span 80, 0.4 % (w/w) Tween 80 in mineral oil, W1.2: 4.5 % (w/w) Span 80, 0.5 % (w/w) Tween 80 in mineral oil) and W2 solutions (W2.1: 2 % (w/v) Tween 20 in PBS buffer, W2.2: 0.5 % (w/v) Tween 80 in PBS buffer) containing different ratios of Span and Tween were tested. Homogeneous, monodisperse emulsions were obtained using **W1 (2 % (w/w) Span 80, 0.4 % (w/w) Tween 80 in mineral oil)** (W1.1), shown in Figure 32A-B, and **W2 (0.5 % (w/v) Tween 80 in PBS buffer)** (W2.2), as shown in Figure 32C-D (Figure 32).
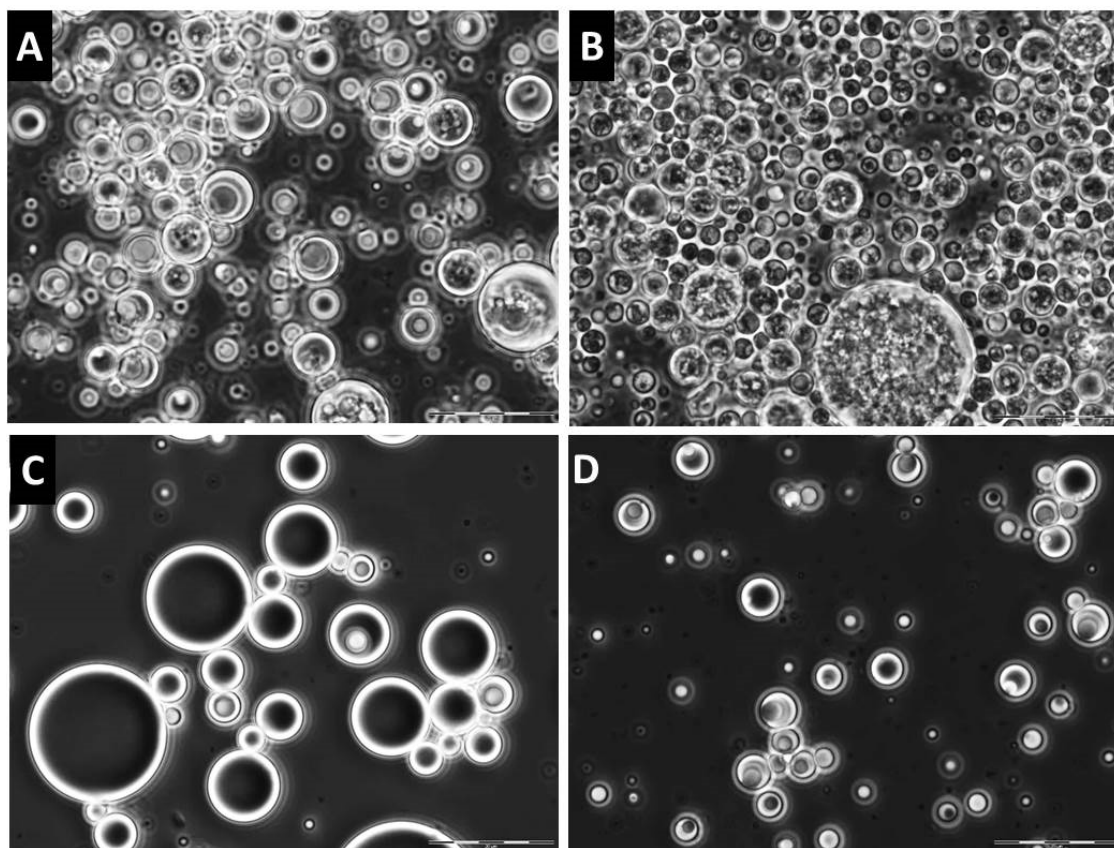
Figure 32: Tests with different W1 and W2 solutions for optimization of the extrusion technique using the span/tween emulsification system shown as transmission microscopic pictures (100 x magnification). The scale bar represents a length of 20 μm. (**A**) **W1.1 (2 % (w/w) Span 80, 0.4 % (w/w) Tween 80 in mineral oil)**, (**B**) W1.2 (4.5 % (w/w) Span 80, 0.5 % (w/w) Tween 80 in mineral oil), in combination with W2 of 0.5 % Tween 80 in PBS; W2.2 (**A-B**). (**C**) W2.1 (2 % (w/w) Tween 20 in PBS), (**D**) **W2.2 (0.5 % (w/w) Tween 80 in PBS**, in combination with W1 of 2 % (w/w) Span 80, 0.4 % (w/w) Tween 80 in mineral oil; W1.1 (**C-D**).

Furthermore optimized extrusion method was also tested using fluorinated oil (W1.3: Pico-Surf™2, 5% (w/w) in FC-40 (Dolomite Microfluidics, Charlestown, MA, USA); W2.3: 2 % (w/v) SDS in PBS buffer) with fluorescein sodium salt (60 μM) as fluorescence marker in the inner aqueous compartment. Fluorinated oil was reported to be beneficial for generation of stable emulsions and reduced leakage of aqueous phase components into the oil phase,

especially using microfluidic devices (Basova *et al.*, 2015; Mazutis *et al.*, 2013). Results showed (w/o/w) emulsions contained many small (<3 µm) inner aqueous compartments and a low fraction (~10 %) of (w/o/w) emulsions. After 24 h incubation (RT) many emulsions were fused to huge droplets (>50 µm) or were completely broken resulting in a highly fluorescent background due to released fluorescein (Figure 33A-D).

In the end, extrusion method in combination with Span/Tween emulsification system was the method of choice for IVC using (w/o/w) emulsions and applied in all later experiments with (w/o/w) emulsions.
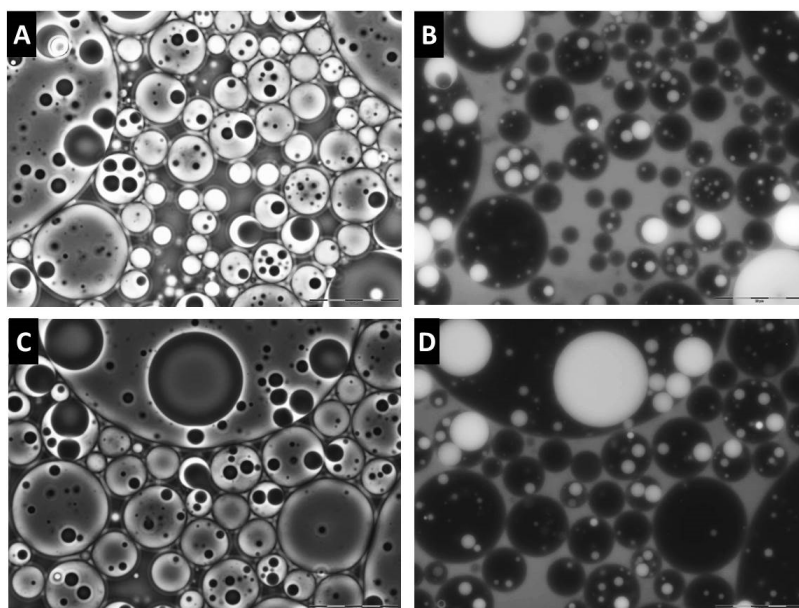
Figure 33: Emulsions extruded with fluorinated oil after 24 h incubation at RT shown as transmission microscopic pictures (**A, C**) and fluorescence pictures (**B, D**)(100 x magnification). The scale bar represents a length of 20 µm.

In order to characterize the generated (w/o/w) emulsions regarding the distribution of oil and water phases within the (w/o/w) emulsion, as well as the droplet surface, confocal microscopy was performed with (w/o) and (w/o/w) emulsion samples generated *via* optimized extrusion method. Confocal microscopy analysis using Z-scans revealed that round droplets were generated (Figure 34A, C, and D) containing an inner phase (water) which is surrounded by an second phase (oil) in case of (w/o) emulsion and for (w/o/w) emulsion the second phase was furthermore entrapped in an external aqueous phase (Figure 34B, G). Some emulsions look slightly compressed (not round) in the Z-scans due to the different reflective index of oil and water (Figure 34C, D). Fluorescence and transmission microscopic pictures confirmed the (w/o/w) emulsions of 5-10 µm in diameter (Figure 34H, I, J and K) harbor one inner aqueous phase containing fluorescein and thus exhibiting a strong green

fluorescence (Figure 34H, I, J and K). Partly destruction of the emulsions during microscopic analysis results in a green background fluorescence due to destroyed droplets releasing the fluorescein dye (Figure 34H, I, J and K).
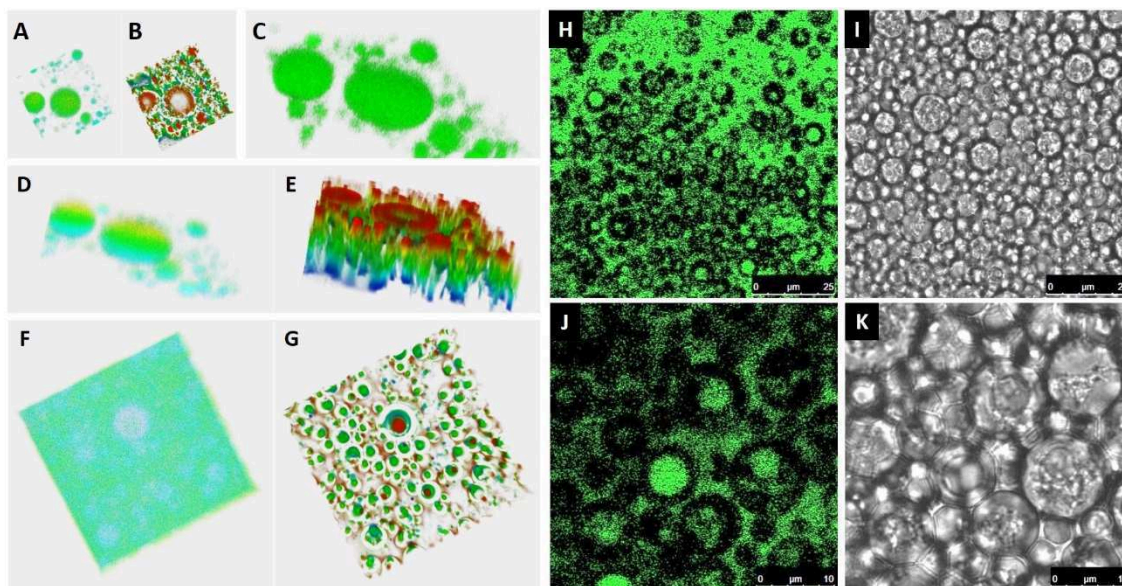
Figure 34: Confocal microscopy pictures (Leica TCS SP8, Leica Mikrosysteme Vertrieb GmbH, Wetzlar, Germany) of (w/o) and (w/o/w) emulsions containing fluorescein (6 µM fluorescein in PBS) using 63x oil immersion objective. For excitation a continuous wave laser (DPSS, 20 mW: 488 nm) was used and for emission at 510-520 nm a sensitive PMT detector was used. (**A-E**) Z-scans of (w/o) emulsion, (**F-G**) Z-scans of (w/o/w) emulsion, (**H**) and (**J**) fluorescence microscopic pictures of (w/o/w) emulsion (**I**) and (**K**) transmission microscopic pictures of (w/o/w) emulsion. The scale bar has a length of 25 µ in Figures (**H**) and (**I**) and 10 µm in Figures (**J**) and (**K**).

In the next step, cell-free cellulase production was tested in combination with the optimized protocol for (w/o/w) emulsion generation *via* extrusion (see Material and Methods, chapter 3.6.10-3.6.11) and fluorescent (w/o/w) emulsions were observed after incubation (30°C, 4 h) using fluorescence microscopy (Figure 35E-F). This result indicates the compatibility of extrusion method, using the Span/Tween emulsification system, with the cell-free expression mixture and the fluorogenic substrate (FDC). The observed fluorescence of the inner aqueous phase of the (w/o/w) emulsions confirms the successful production of active CelA2-H288F enzyme and subsequent substrate conversion within the emulsion compartments. Fluorescence microscopy of the same (w/o/w) emulsions after overnight incubation (RT) showed a similar result as illustrated in Figure 35E-F, indicating stability of the generated (w/o/w) emulsions and the entrapped fluorescent product for at least 24 hours. As fluorescence was exclusively observed in the inner aqueous phase of the (w/o/w) emulsions it can be concluded that no or only negligible leakage of the fluorescent product occurred upon incubation up to 24 hours.

of the mutated gene library in an appropriate expression host, (II) screening of the generated mutant library for improved variants using a suitable screening system mimicking conditions comparable with the final application conditions, and (III) isolation of the gene encoding for the improved protein variant, which is then usually used for iterative rounds of diversity generation and screening until a desired property or level of improvement is achieved, as shown in Figure 1 (Güven *et al.*, 2010; Wong *et al.*, 2006).

Figure 1: Principle of a directed evolution experiment. The figure is taken from Güven *et al.* (Güven *et al.*, 2010).

One of the key challenges for performance of a successful directed evolution experiment is the generation of mutant libraries of sufficient size with high diversity (Wong *et al.*, 2006). Nowadays, the large number of available chemical- or molecular-biologically-based methods for mutant library generation enables a fast and simple generation of highly diverse gene libraries (Wong *et al.*, 2006). For generation of random mutant libraries in directed evolution experiments the most commonly used method is **error prone PCR (epPCR)**. Due to incorrect incorporation of nucleotides, random mutations are introduced during PCR, by using *Taq* DNA polymerase from *Thermus aquaticus* which lacks a 3'-5' proofreading exonuclease activity and has an intrinsic high error rate of 8.0×10^{-6} mutations/bp/duplication (Cirino *et al.*, 2003). This results in around 1 substitution per 125,000 bp per PCR cycle which is insufficient for randomizing genes that seldom exceed 5,000 bp in length. In order to increase the error-rate, the fidelity of the DNA polymerase can be reduced by addition of $MnCl_2$, increasing the concentration of $MgCl_2$, supplementing unbalanced dNTP

3.4.6. Fluorimetric assay for determination of phosphatase activity

3.4.6.1. 4-MUP activity assay for determination of phosphatase activity

For phytase activity detection 4-methylumbelliferyl phosphate (4-MUP; Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany) activity assay (Shivange *et al.*, 2012) was optimized for detection of extremely low amounts of cell-free produced phytase in 384-well plate format. Optimal detection of low phytase concentrations was achieved in 384-well plate format by adding 25 µl 4-MUP (5 mM in buffer pH 5.5, 0.25 M sodium acetate, 1 mM CaCl2, 0.01 % Tween 20) to 5 µl of protein sample with a total assay volume of only 30 µl (Figure 36). Protein concentrations from 0-0.19 µg/ml were applied in optimized 4-MUP activity assay and phytase activity was determined by Tecan Infinite M1000 MTP reader (Tecan Group AG, Männedorf, Switzerland) as increase in relative fluorescence at excitation λ_{ex} 360 nm and emission λ_{em} 465 nm. The relative phytase activity was determined as 4-methylumbelliferone (4-MU) product formation by calculation of slopes from the increase in relative fluorescence in RFU (relative fluorescent units) per s (Figure 36). A linear correlation of 4-MU product formation dependent on the protein concentration is described by the equation y = 4.2659 x - 0.0157 and is shown in Figure 36. This equation (Figure 36) was used for the calculation of active phytase concentrations from measurements using optimized 4-MUP activity assay in 384-well plate.
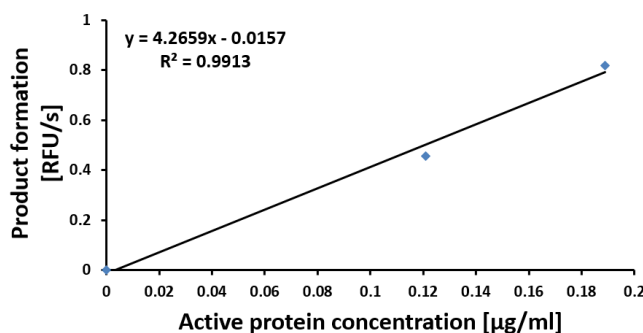
Figure 36: Linear correlation of 4-MU product formation (4-MU) in [RFU/s] dependent on the active protein concentration. The product formation, measured as slope of increase in relative fluorescence over time in [RFU/s], was calculated from relative increase in fluorescence (RFU) measured with 4-MUP activity assay over time for different concentrations of protein. The resulting equation was used for all calculations of active protein concentrations for phytase.

3.4.7. *Fluorimetric assays for determination of cellulolytic activity*

3.4.7.1. *4-MUC activity assay for determination of cellulolytic activity*

For cellulase activity detection 4-methylumbelliferyl-β-D-cellobioside (4-MUC, Sigma-Aldrich Chemie GmbH) activity assay (Lehmann *et al.*, 2012) was optimized for detection of cell-free produced cellulase in 384-well plate format. Optimal detection of low CelA2-H288F concentrations was achieved in 384-well plate format by adding 10 µl 4-MUC (0.1 mM in potassium phosphate buffer, pH 7.2, 0.2 M) to 10 µl of protein sample with a total assay volume of only 20 µl (Figure 37). Protein concentrations from 0-23 µg/ml were applied in optimized 4-MUC activity assay and product formation (4-MU) was determined by calculation of slopes [RFU/s] from the obtained fluorescence values in [RFU] over the time (Figure 37). Linear correlation of 4-MU product formation dependent on the protein concentration is described by the equation $y = 0.1101 x - 0.0048$ and is shown in Figure 37. This equation (Figure 37) was used for the calculation of active cellulase concentrations from measurements using optimized 4-MUC activity assay in 384-well plate.
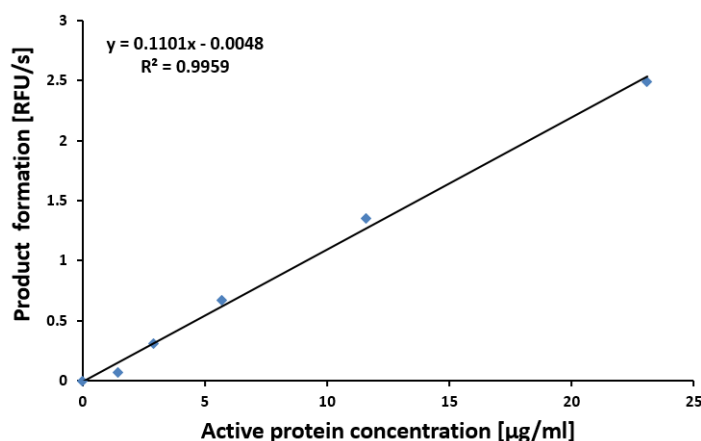
Figure 37: Linear correlation of 4-MU product formation (4-MU) in [RFU/s] dependent on the active protein concentration. Product formation measured as slope in [RFU/s] was calculated from relative increase in fluorescence (RFU) measured with 4-MUC activity assay over time for different concentrations of protein. The resulting equation was used for all calculations of active protein concentrations for cellulase.

3.4.7.2. *FDC activity assay for determination of cellulolytic activity*

Fluorescein-di-β-D-cellobioside (FDC, Biomol GmbH, Hamburg, Germany) activity assay was developed to detect cell-free produced cellulase in 384-well plate format. Optimal detection of the fluorescein product converted by active CelA2-H288F in 384-well plate format was achieved by adding 10 µl FDC (0.1 mM in KPI buffer, pH 7.2 0.2 M) to 10 µl of protein sample with a total assay volume of only 20 µl, resulting in a linear correlation of increasing fluorescence measured in relative fluorescent units (RFU) per second (Figure 38). As shown

in Figure 39 the FDC assay is compatible with the cell-free cellulase expression system and has sufficient sensitivity to detect low levels of cellulolytic activity like present in cell-free expressed samples (green curve). The cell-free cellulase production is not disturbed as FDC is not interfering with cell-free transcription and translation of cellulase which is shown by FDC addition prior 1.5 h and 4 h incubation of *in vitro* reaction mixtures, as samples exhibit fluorescence, indicating successful *in vitro* cellulase production (Figure 39, purple and red curve). The EF, serving as negative control, is not showing any conversion of FDC, indicating no conversion of the substrate by other components of the *in vitro* reaction mixture occurs (Figure 39, blue curve).
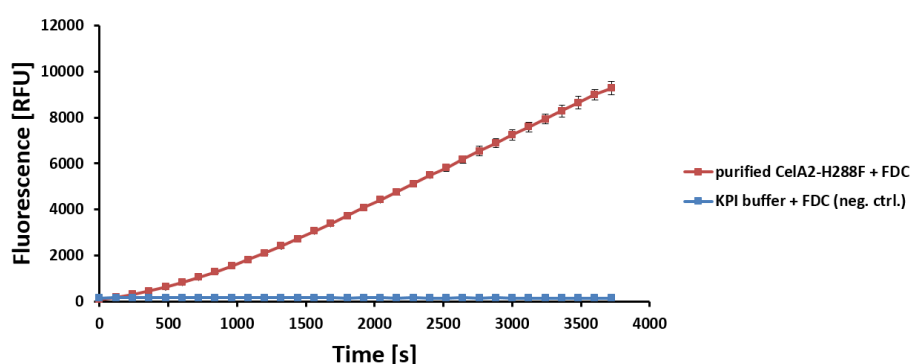
Figure 38: Linearity of fluorescein-based fluorescence over FDC conversion time are shown in relative fluorescence units (RFU). 10 µl of FDC (0.1 mM in KPI buffer, pH 7.2, 0.2 M) were added to 10 µl purified CelA2-H288F enzyme or KPI buffer (pH 7.2, 0.2 M). The reported values are the average of three measurements and deviations are calculated from the corresponding mean values.
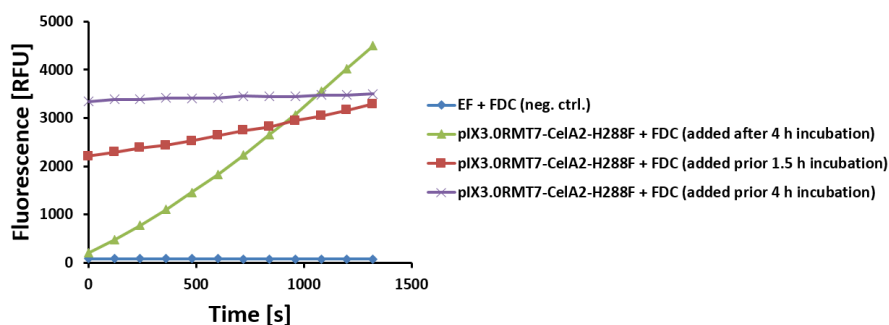
Figure 39: Conversion of FDC to fluorescein by cell-free expressed CelA2-H288F shown in RFU over the time. FDC (5 µM) was added to *in vitro* reaction samples either before incubation for 1.5 and 4 h (blue and violet curve) or afterwards prior fluorescence measurement (green curve). Cell-free expression of the EF served as negative control.

3.4.7.3. Substrate selection for flow cytometer-based in vitro compartmentalization screening platform

In order to maintain genotype-phenotype linkage between gene, the enzyme it encodes and the fluorescent product generated upon enzyme catalyzed conversion of the fluorogenic substrate within the (w/o/w) emulsion compartment a selection of most suitable substrate-product pairs was performed.

4-Methylumbelliferyl-β-D-cellobioside (4-MUC) and fluorescein-di-β-D-cellobioside (FDC) were used as substrates for cellulase and fluorescent product leakage from (w/o/w) emulsion upon substrate conversion was investigated. Fluorescence microscopic analysis of 4-methylumbelliferone, the fluorescent product generated upon conversion of 4-MUC substrate, showed that 4-MU immediately diffuses from the inner aqueous phase of the (w/o/w) emulsion into the outer aqueous phase and the PBS buffer and was quickly bleached during observation with the fluorescent microscope, so no fluorescent pictures could be taken (pictures remained all black, as during photographing the sample was already bleached). Fluorescein-based substrates showed to be much more sensitive than coumarin-based substrates and had lower product diffusion rates from the inner aqueous phase (Figure 35F). The presence of the charged carboxyl group in FDC probably reduces product leakage into the hydrophobic oil phase and makes FDC a suitable substrate for flow cytometry-based screening in (w/o/w) emulsions. In an additional optimization step, BSA (1-2 %) was added to the (w/o/w) emulsion to saturate the water-in-oil interface. Furthermore, BSA influence on product leakage after *in vitro* cellulase synthesis and FDC conversion was analyzed on flow cytometer. Figure 40 shows the flow cytometry density plots obtained after 4 h of *in vitro* production of cellulase in (w/o/w) emulsion containing varied BSA concentrations. Finally, (w/o/w) emulsion reaction sample containing 1 % BSA showed the highest number of fluorescent events (26.45 %, Figure 40B) indicating the least leakage of fluorescein from (w/o/w) emulsion compartments in comparison to samples without BSA (4.75 %, Figure 40A) and sample containing 2 % BSA (2.50 %, Figure 40C). These results suggest that concentrations of BSA >1 % have an inhibitory effect on cell-free cellulase production which is indicated by lower number of fluorescent events (Figure 40C).
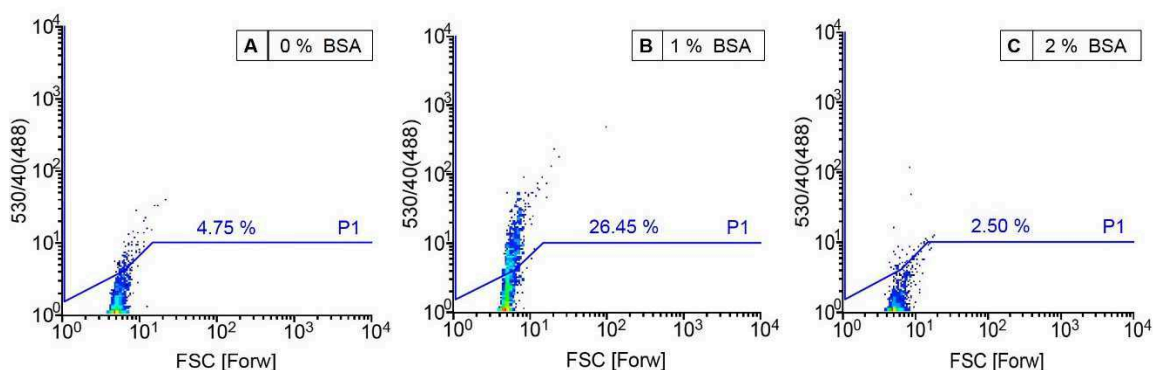
Figure 40: Flow cytometer analysis showing density plots obtained after cell-free production of cellulase from LT in (w/o/w) emulsion containing varied BSA concentrations (4 h, 25°C). (**A**) (W/o/w) emulsion reaction sample incubated without BSA, (**B**) incubated with 1 % BSA, and (**C**) incubated with 2 % BSA. The forward scatter (FSC) on the x-axis represents the size of (w/o/w) emulsions. The y-axis represents the fluorescence intensity after 488 nm excitation with a fluorescence emission detection between 530±20 nm which is annotated by 530/40(488). Lines indicate gate P1 which was set to categorize events in highly fluorescent/potentially beneficial events (appearing above) and not beneficial events. The figure is part of the published manuscript "Körfer, G., Pitzler C., Vojcic L., Martinez R., and Schwaneberg U.; *In vitro* flow cytometry-based screening platform for cellulase engineering, Scientific Reports (2016); 6: 26128" and was reprinted with permission of the Nature Publishing Group.

3.4.8. *Cell-free cellulase production within (w/o/w) emulsion, flow cytometry analysis and sorting of model libraries with defined active-to-inactive ratio using plasmid template DNA*

Cell-free cellulase production in (w/o/w) emulsion compartments generated by membrane extrusion method was performed using *in vitro* expression FastLane *E.coli* Mini Kit (RiNA GmbH). Incubation time of cell-free production mix in (w/o/w) emulsions was varied from 3 h to 24 h and after 4 h the sufficient cellulase activity for flow cytometry analysis (45.77 % active fraction) was achieved (Figure 41D). Incubation at different temperatures revealed the highest level of cellulase activity at 30°C and it was used as temperature of choice in further experiments for cell-free cellulase production from plasmid DNA (Figure 41D).
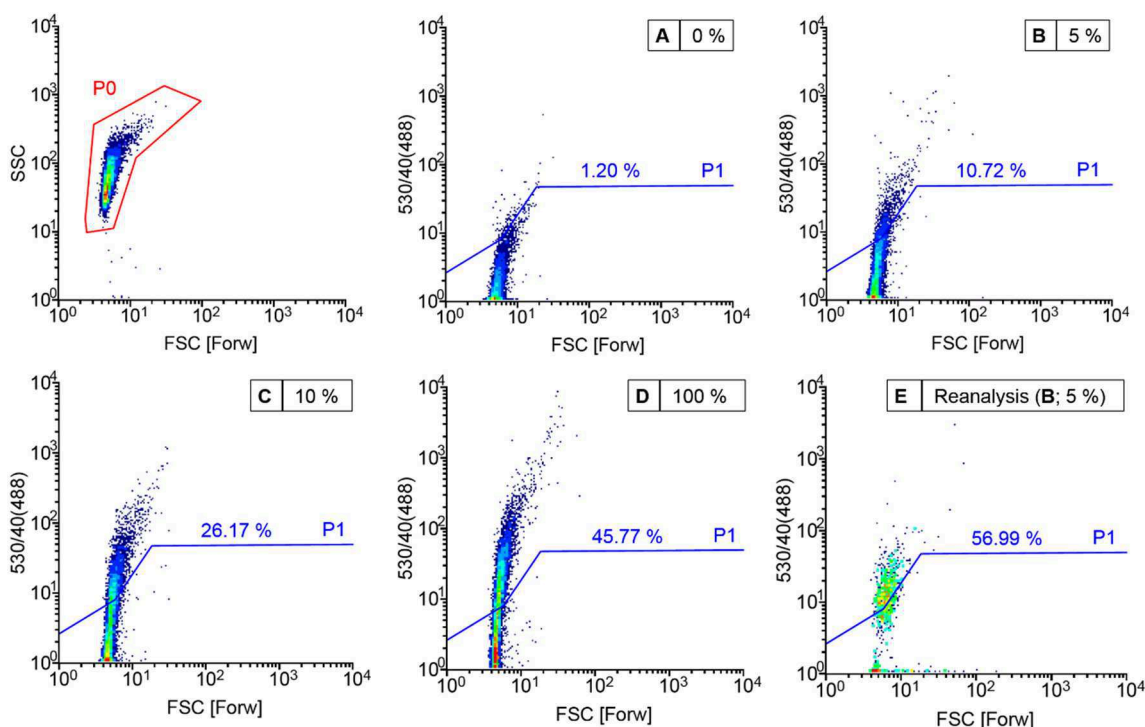
Figure 41: Flow cytometer analysis showing density plots of cellulase model libraries with defined ratios of active and inactive DNA template after *in vitro* expression in (w/o/w) emulsions. The P0 gated area represents the (w/o/w) emulsion population using the FSC representing the size of the (w/o/w) emulsion and the SSC representing the granularity of the (w/o/w) emulsion. Y-axis represents the fluorescence intensity after 488 nm excitation with a fluorescence emission detection between 530±20 nm which is annotated by 530/40(488). Lines indicate gate P1 which was set to categorize events in highly fluorescent/potentially beneficial events (appearing above) and not beneficial events. (**A**) Negative control with CelA2-H288F-H580Q$_{inactive}$ DNA: forward scatter (FSC) versus side scatter (SSC) and FSC versus fluorescence intensity (λ_{ex} 494 nm and λ_{em} 516 nm, (**B**) 5 % model library (ratio 5:95 of CelA2-H288F$_{active}$ versus CelA2-H288F-E580Q$_{inactive}$), (**C**) 10 % model library (ratio 10:90 of CelA2-H288F$_{active}$ versus CelA2-H288F-E580Q$_{inactive}$), (**D**) Positive control with CelA2-H288F$_{active}$, and (**E**) Reanalysis of the sorted 5 % model library (**B**) resulted in 5.3-fold enrichment of the active events above the line P1. The figure is part of the published manuscript "Körfer, G., Pitzler C., Vojcic L., Martinez R., and Schwaneberg U.; *In vitro* flow cytometry-based screening platform for cellulase engineering, Scientific Reports (2016); 6: 26128" and was reprinted with permission of the Nature Publishing Group.

Cell-free expressed plasmid DNA model libraries with active-to-inactive ratio 5:95 and 10:90 (CelA2-H288F$_{active}$ versus CelA2-H288F-E580Q$_{inactive}$) in (w/o/w) emulsions were analyzed for determining parameters for efficient flow cytometry sorting. Analysis and sorting was done at an event rate of 1500 events s^{-1}. An increase in the amount of CelA2-H288F$_{active}$ leads to an increase in the fluorescent signal in gate P1 (Figure 41A-D). In case of 0 % CelA2-H288F$_{active}$, 1.2 % of the fluorescent population is located inside the gate P1 that was used for sorting active variants (Figure 41A). Analysis of 5 % of a CelA2-H288F$_{active}$ populations results in approx. 11 % positive events inside the gate P1 (Figure 41B), while for 10 % CelA2-H288F$_{active}$, 26 % of the population is within the sorting gate P1 (Figure 41C) and analysis of a 100 % CelA2-H288F$_{active}$ population shows 46 % population being inside the gate P1, represented in Figure 41D. A 46 % fraction of active events for 100 % CelA2-H288F$_{active}$

suggests that not all (w/o/w) emulsions contain sufficient amount of template DNA, which is desired, as it ensures that statistically not more than one DNA molecule per droplet is entrapped (Bernath *et al.*, 2004b). Active-to-inactive model library in ratio 5:95 was sorted at an event rate of 1500 events s^{-1} with a sorting efficiency of 99.6 %. Reanalysis of the sorted fraction resulted in a 5.3-fold enrichment of the active emulsion fraction (Figure 41E). The obtained enrichment validated the flow cytometry-based *in vitro* compartmentalization cellulase screening platform.

3.4.9. *Cell-free cellulase production within (w/o/w) emulsion, flow cytometry analysis and sorting of model libraries with defined active-to-inactive ratio using linear template DNA*

Cell-free cellulase production from linear DNA template in (w/o/w) emulsion compartments generated by extrusion was performed using *in vitro* expression FastLane *E.coli* Mini Kit (RiNA GmbH). Incubation time of cell-free production mix in (w/o/w) emulsions was varied from 3 h to 24 h and after 4 h the sufficient cellulase activity for flow cytometry analysis (40.3 % active fraction) was achieved (Figure 42). Incubation at different temperatures revealed the highest level of cellulase activity at 25°C and it was used as temperature of choice in further experiments using linear DNA as template for cell-free cellulase production (Figure 42).
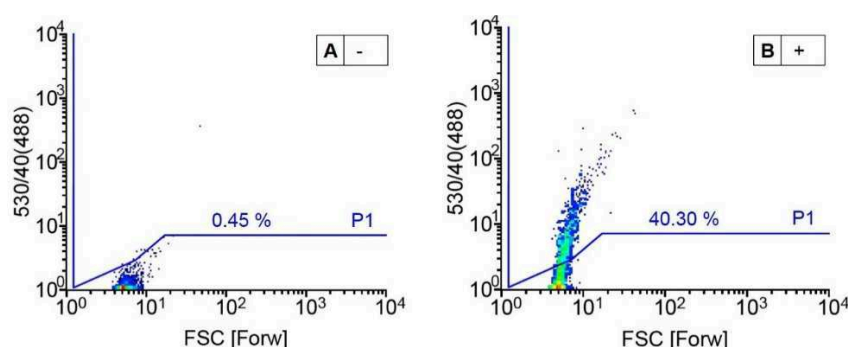
Figure 42: Flow cytometer analysis showing density plots obtained after cell-free production of cellulase in (w/o/w) emulsion after 4 h incubation at 25°C. (**A**) Negative control using CelA2-H288F-E580Q$_{inactive}$ for determination of background, (**B**) positive control using only CelA2-H288F$_{active}$. The forward scatter (FSC) on the x-axis represents the size of (w/o/w) emulsions. The y-axis represents the fluorescence intensity after 488 nm excitation with a fluorescence emission detection between 530±20 nm which is annotated by 530/40(488). Lines indicate gate P1 which was set to categorize events in highly fluorescent/potentially beneficial events (appearing above; *i.e.* 40.3 % for the positive control (**B**)) and not beneficial events. The figure is part of the published manuscript "Körfer, G., Pitzler C., Vojcic L., Martinez R., and Schwaneberg U.; *In vitro* flow cytometry-based screening platform for cellulase engineering, Scientific Reports (2016); 6: 26128" and was reprinted with permission of the Nature Publishing Group.

Cell-free expressed linear template DNA model libraries with active-to-inactive ratio 10:90, 30:70, and 50:50 (CelA2-H288F$_{active}$ versus CelA2-H288F-E580Q$_{inactive}$) in (w/o/w) emulsions were analyzed for determining parameters for efficient flow cytometry sorting. Analysis was performed at an event rate of 1500 events s^{-1}. Figure 43 shows an increase in the amount of CelA2-H288F$_{active}$ leads to an increase in the fluorescent signal in gate P1 (Figure 43). Analysis of 10 % of a CelA2-H288F$_{active}$ populations results in 1.5 % positive events inside the gate P1 (Figure 43A), whereas for 30 % CelA2-H288F$_{active}$, approx. 40 % of the population is within the sorting gate P1 (Figure 43B) and for 50 % CelA2-H288F$_{active}$, approx. 54 % of the population is within the sorting gate P1 (Figure 43C).
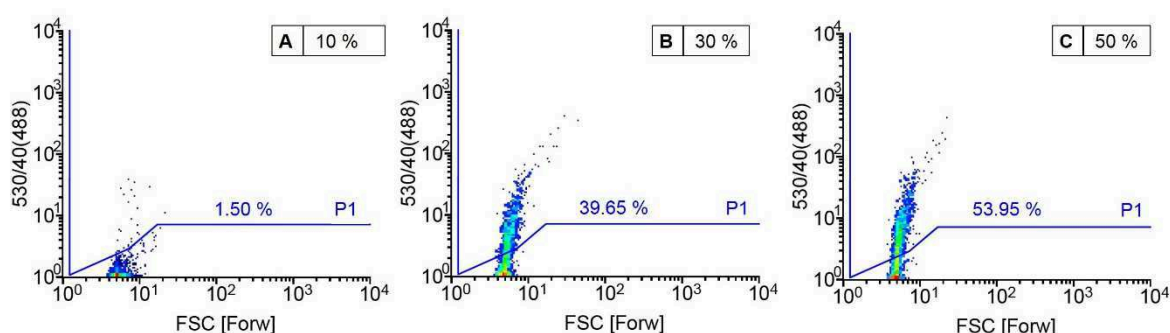
Figure 43: Flow cytometer analysis showing density plots obtained after *in vitro* expression of CelA2-H288F model libraries with defined ratios of active and inactive linear DNA template in (w/o/w) emulsions (4 h, 25°C, 0.36 mM FDC, 0.656 µM DNA). The forward scatter (FSC) on the x-axis represents the size of (w/o/w) emulsions. The y-axis represents the fluorescence intensity after 488 nm excitation with a fluorescence emission detection between 530±20 nm which is annotated by 530/40(488). Lines indicate gate P1 which was set to categorize events in highly fluorescent/potentially beneficial events (appearing above) and not beneficial events. (**A**) 10 % model library (ratio 10:90 of CelA2-H288F$_{active}$ versus CelA2-H288F-E580Q$_{inactive}$). (**B**) 30 % model library (ratio 30:70 of CelA2-H288F$_{active}$ versus CelA2-H288F-E580Q$_{inactive}$) (**C**) 50 % model library (ratio 50:50 of CelA2-H288F$_{active}$ versus CelA2-H288F-E580Q$_{inactive}$).

In order to obtain a detectable level of cellulase activity using linear DNA-based gene diversity libraries with high mutational load a model library containing active-to-inactive cellulase in the ratio 30:70 (CelA2-H288F$_{active}$ versus CelA2-H288F-E580Q$_{inactive}$) was used to mimic library screening conditions. Model library in active-to-inactive cellulase ratio 30:70 was used to optimize the amount of linear template DNA (0.164-0.656 µM) and substrate concentration (0.13-0.75 mM of FDC) for flow cytometer analysis. Samples containing <0.656 µM of template DNA during *in vitro* cellulase production in (w/o/w) emulsions showed upon flow cytometry analysis a fluorescence comparable to the negative control signal (Figure 44A). DNA template at concentration of 0.656 µM revealed 23.9 % of fluorescent events upon flow cytometry analysis which is sufficient to efficiently sort gene diversity libraries containing a fraction of active clones <30 % (Figure 44B).
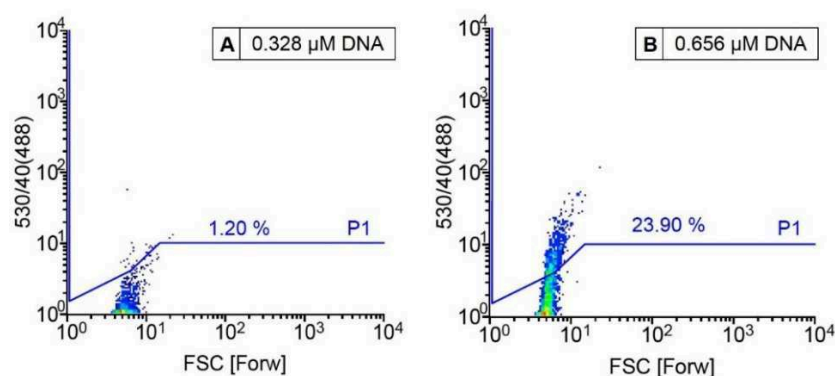
Figure 44: Flow cytometer analysis showing density plots obtained after cell-free production of cellulase in (w/o/w) emulsion with two amounts of linear DNA (**A**) 0.328 and (**B**) 0.656 µM DNA employing a model library with 30 % of active variants (ratio 3:7 of CelA2-H288F$_{active}$ versus CelA2-H288F-E580Q$_{inactive}$). The forward scatter (FSC) on the x-axis represents the size of (w/o/w) emulsions. The y-axis represents the fluorescence intensity after 488 nm excitation with a fluorescence emission detection between 530±20 nm which is annotated by 530/40(488). Lines indicate gate P1 which was set to categorize events in highly fluorescent/potentially beneficial events (appearing above) and not beneficial events. The figure is part of the published manuscript "Körfer, G., Pitzler C., Vojcic L., Martinez R., and Schwaneberg U.; *In vitro* flow cytometry-based screening platform for cellulase engineering, Scientific Reports (2016); 6: 26128" and was reprinted with permission of the Nature Publishing Group.

Using different substrate concentrations (0.13-0.75 mM FDC) revealed a highest signal-to-noise ratio (17.7-fold) at 0.46 mM FDC (Figure 45A-B). As shown in Figure 45C-D higher substrate concentrations (0.75 mM FDC) resulted in no further increase of the fluorescent signal compared to 0.46 mM FDC (Figure 45A-B), but a worse signal-to-noise ratio (9.1-fold) was obtained.

concentrations or by increasing the number of amplification cycles (Cadwell *et al.*, 1992; Cirino *et al.*, 2003). Besides the advantages like robustness and simplicity of epPCR-based methods, typical disadvantages are the limited ability to generate sufficient diversity, due to the biased mutational spectra of the DNA polymerase, the redundancy of the genetic code and the lack of subsequent mutations due to low mutation frequencies and high numbers of transitions which are leading to so called "cold or hot spots" in a protein sequence (Wong *et al.*, 2004). In the last decade, the advanced random mutagenesis method **Sequence Saturation Mutagenesis (SeSaM)** was developed, which overcomes limitations caused by the polymerase bias. SeSaM is able to saturate every single nucleotide position of the target sequence with all four standard nucleotides and introduces the subsequent nucleotide mutations, so it can be used to generate transversions with a higher probability than epPCR (Wong *et al.*, 2004). Prominent recombination-based methods in directed evolution are **DNA shuffling** (Stemmer, 1994), **Staggered Extension Process (StEP)** (Zhao *et al.*, 1998) and **Phosphorothioate-based DNA Recombination (PTRec)** (Marienhagen *et al.*, 2012), which combine and accumulate beneficial mutations from different gene variants. Prominent techniques for focused diversity generation are **Site-Directed Mutagenesis (SDM)** or **Site-Saturation Mutagenesis (SSM)**, allowing the quick introduction of point mutations, with a mutagenesis efficiency of 70-90 %, into specific DNA-sequences using a pair of complementary mutagenesis primers for the amplification of a complete plasmid in one single PCR (Wang *et al.*, 1999). SDM and SSM can be used when structure and function of the enzyme is at least partly explored and it becomes reasonable to analyze amino acid exchanges at certain positions. This method was expanded using **Iterative Saturation Mutagenesis (ISM)** which employs subsequent cycles of SSM and screening (Reetz *et al.*, 2007). A recent advancement of these focused mutagenesis methods was achieved using multiple site-saturation with **OmniChange** method which enables a simultaneous site-saturation of up to five codons, allowing engineering of "localizable" properties (e.g. activity, selectivity) and the discovery of cooperative effects between amino acid substitutions that are improbable to identify using single codon saturation methods (Dennig *et al.*, 2011). The OmniChange method enables within only four steps the generation of diverse mutant libraries in a robust, reliable and simple manner, without DNA modifying enzymes like ligases or endonucleases, using a DNA sequence independent multi DNA

mutation frequency of eight mutations per gene which was estimated by sequencing of twelve randomly selected clones. Clones were determined as "active" when the increase in fluorescence in RFU/s was at least 75 % of parent CelA2-H288F activity. In order to achieve even higher enrichment of active population a strict gating strategy was applied by sorting only the event fraction with highest fluorescence using gate P2 (8.5 %) (Figure 46A-B). Upon flow cytometry analysis the active fraction of cell-free expressed CelA2-H288F-epPCR library in (w/o/w) emulsions was determined as 17.40 % (Figure 46A), correlating with the result in MTP format (18.75 %). The fraction containing 8.50 % of the most fluorescent events was selected for sorting (Figure 46A). In only 2.25 hours in total 14,628,728 events were analyzed (2000 events s^{-1}) of which 1,026,881 were sorted with an efficiency of 93.2 %. Reanalysis of the sorted population revealed an enrichment in active fraction of 8.2-fold, determined by an increase of the population in the P2 sorting gate from 8.50 % to 69.75 % (Figure 46B).
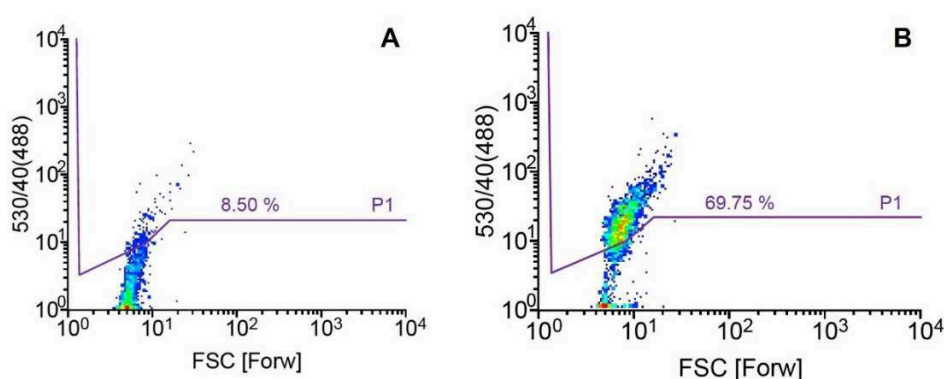
Figure 46 Flow cytometer analysis showing density plots of random *celA2-H288F* mutant library after *in vitro* expression in (w/o/w) emulsions (4 h, 25 °C). The forward scatter (FSC) on the x-axis represents the size of (w/o/w) emulsions. The y-axis represents the fluorescence intensity after 488 nm excitation with a fluorescence emission detection between 530±20 nm which is annotated by 530/40(488). Lines indicate gate P1 which was set to categorize events in highly fluorescent/potentially beneficial events (appearing above) and not beneficial events. (**A**) *celA2-H288F* random mutant library during sorting of the 8.5 % most fluorescent events. (**B**) Reanalysis of sorted library with an 8.2-fold enriched active fraction (69.75 % fluorescent events). The figure is part of the published manuscript "Körfer, G., Pitzler C., Vojcic L., Martinez R., and Schwaneberg U.; *In vitro* flow cytometry-based screening platform for cellulase engineering, Scientific Reports (2016); 6: 26128" and was reprinted with permission of the Nature Publishing Group.

3.4.11. *Recovery of sorted DNA library from (w/o/w) emulsions and transformation in expression host for MTP analysis*

Several different methods for extraction and recovery of DNA from (w/o/w) emulsions were compared regarding their efficiency and sensitivity. Reported standard methods for recovery of DNA from (w/o/w) emulsions are extraction with (a) diethyl ether and (b) hexane/mineral oil (Miller *et al.*, 2006). In this study, besides the two reported methods, purification of DNA

from (w/o/w) emulsions using NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany) was tested and compared.

For comparison of the three methods samples of sorted (w/o/w) emulsions as well as samples of different dilution of (w/o/w) emulsions ($1:10^2$-$1:10^6$) were treated with either one of the three different methods. Subsequently optimized PCR reaction with PLICing primers was performed and PCR products were analyzed on agarose gel (Figure 47). Furthermore methods (a) and (b) were optimized by introduction of a desalting step using 0.025 µm MF-Millipore^{TM} membrane filters (Merck Chemicals GmbH, Darmstadt, Germany) for microdialysis of DNA after the extraction to remove residual salts from PBS buffer interfering with the PCR reaction to improve PCR amplification efficiency.

After extraction of samples with the three different methods and performance of recovery PCRs the products were analyzed using agarose-gelelectrophoresis (Figure 47). Extraction with diethyl-ether method was not sensitive enough to isolate DNA from (w/o/w) emulsion samples, as it was not possible to amplify DNA from the extracted sample with optimized recovery PCR (data not shown). Hexane/mineral oil method showed a higher sensitivity than diethyl ether method, as it was possible to isolate and amplify DNA from sorted (w/o/w) emulsion samples in recovery PCR (Figure 47).

The most promising and fastest method with the highest sensitivity for recovery of low DNA amounts from (w/o/w) emulsions was purification of (w/o/w) emulsion samples with NucleoSpin® Gel and PCR clean up kit. CelA2-H288F DNA (1.8 kB) was successfully isolated with the kit purification method from sorted (w/o/w) emulsion samples and from diluted (w/o/w) emulsion samples (10^2 to 10^6 dilutions).
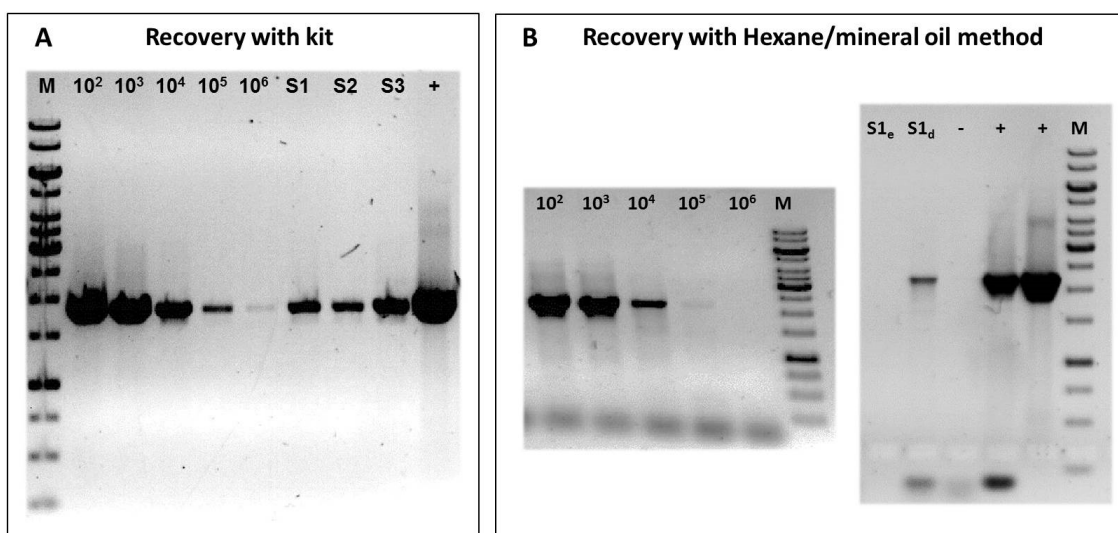
Figure 47: (**A**) Agarose-gelelectrophoresis picture of recovered PCR products after DNA isolation from (w/o/w) emulsions using NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG). **M** represents the marker (GeneRuler™ 1 kb DNA Ladder, Life Technologies GmbH, Darmstadt, Germany) used as reference; **10^2 to 10^6** are dilutions of (w/o/w) emulsions samples from $1:10^2$ up to $1:10^6$ in 500 µl PBS buffer; **S1-S3** are samples (each 500 µl) from different active emulsion fractions sorted with flow cytometer and DNA was isolated with NucleoSpin® Gel and PCR clean up kit (Machery Nagel); **+** represents the positive control containing CelA2-H288F$_{active}$ DNA as template for the PCR. (**B**) Agarose-gelelectrophoresis picture of recovery PCR products after DNA isolation from (w/o/w) emulsions using hexane/mineral oil method. **M** represents the marker (GeneRuler™ 1 kb DNA Ladder, Life Technologies GmbH, Darmstadt, Germany) used as reference; **10^2 to 10^6** are dilutions of (w/o/w) emulsions samples from $1:10^2$ up to $1:10^6$ in 500 µl PBS buffer; **S$_e$-S$_d$** are samples (each 500 µl) from different active emulsion populations sorted with flow cytometer and DNA was isolated with hexane/mineral oil method, **S$_e$** is the extracted sample without desalting step, **S$_d$** is the extracted sample with desalting step; **+** represents the positive control containing CelA2-H288F$_{active}$ DNA as template for the PCR; **-** represents the negative control containing no template DNA for the PCR.

Upon sorting, (w/o/w) emulsion the sample containing >10^6 CelA2-H288F events from gate P1 (Figure 46) was disrupted using optimized DNA recovery method based on NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG) for isolation and PCR amplification of sorted linear CelA2-H288F-epPCR DNA library. Amplified *celA2* PCR products were analyzed using agarose-gelelectrophoresis (Figure 47A; Figure A3, appendix), subsequently cloned into pET28a(+) vector backbone *via* PLICing (Blanusa *et al.*, 2010) and transformed into *E.coli* BL21 Gold (DE3) as expression host for detailed analysis in MTP.

3.4.12. *Analysis of sorted cellulase variants in MTP*

Sorted cellulase variants were analyzed for activity towards 4-MUC upon expression in 96-well MTPs (Lehmann *et al.*, 2012). MTP analysis of 528 cellulase variants revealed 33 cellulase variants showing significantly improved activity in comparison to parent CelA2-H288F (up to 7.2-fold) even though that 4-MUC and not the pair FDC/fluorescein was used in the rescreening. The latter indicates that a high quality library was isolated in which also activities for other substrates than the screening substrate can be optimized. Analysis of variant CelA2-H288F-M1 showed 17.5-fold improvement for the pair FDC/fluorescein in comparison to the parent CelA2-H288F (Figure A4, appendix). The three most promising variants were selected for rescreening and CelA2-H288F-M1 (N273D/N468S) revealed to be the best variant and was selected for purification and detailed kinetic characterization (Figure 48).
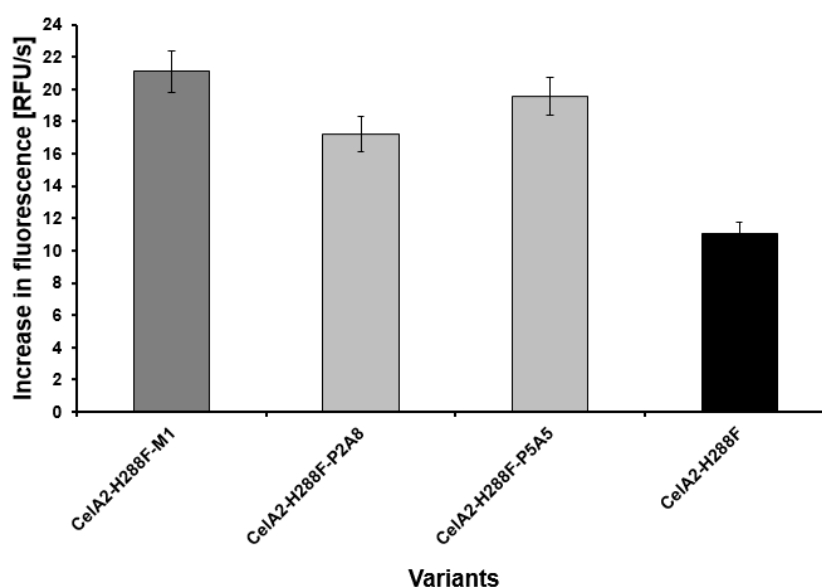
Figure 48: Results of a directed evolution campaign of CelA2-H288F using uHTS-IVC technology platform with one single round of random mutagenesis and flow cytometer-based screening. Rescreening in MTP of the three best variants in comparison to parent CelA2-H288F (black) using 4-MUC assay; the x-axis describes the different variants and the y-axis shows the increase in fluorescence in relative fluorescent units (RFU) per s. The most beneficial variant CelA2-H288F-M1 is highlighted in dark grey. The reported values are the average of ten measurements and the shown average deviations are calculated from the mean values.
The figure is part of the published manuscript "Körfer, G., Pitzler C., Vojcic L., Martinez R., and Schwaneberg U.; *In vitro* flow cytometry-based screening platform for cellulase engineering, Scientific Reports (2016); 6: 26128" and was reprinted with permission of the Nature Publishing Group.

fragment assembly, based on chemical cleavage of phosphorothioated nucleotides by ethanol-iodine (Dennig *et al.*, 2011).

The method of choice for generation of the mutant library in a directed evolution campaign strongly depends on the enzyme property which should be optimized or re-engineered as well as on the screening throughput. Error-prone based mutagenesis methods, which are leading mainly to similar amino acid substitutions, could be beneficial for tailoring of enzyme properties like enantioselectivity (Arnold, 2001; Reetz *et al.*, 1999), whereas engineering of thermostability, requiring many mutations leading to chemically different amino acid substitutions, might be achieved using SeSaM method. All in all, nowadays there are plenty diversity generation methods available (OmniChange, SeSaM, etc.) to generate huge and highly diverse mutant libraries ($\sim10^{15}$ mutants) covering and investigating the sequencing space of the gene of interest in order to tailor biocatalysts with improved and novel properties for industrial or pharmaceutical applications (Reetz *et al.*, 2008). In summary, these novel methods for diversity generation enable the generation of mutant libraries of high mutational load requiring potent screening formats to sample through the generated diversity.

1.2. Screening systems

As a huge variety of methods for the generation of highly diverse mutant libraries is already available (Ruff *et al.*, 2013), currently the biggest bottleneck of directed evolution campaigns is the ability to screen a significant portion of the generated sequence space, as for a protein of only 100 amino acids the theoretical sequence space comprises 20^{100} ($\sim1.27 \times 10^{130}$) different sequences. The first requirement for designing a screening system, is its complementarity to the final application conditions (*i.e.* substrate, temperature, environmental stress). The established screening system has to mimic the application conditions as close as possible, in order to apply a correct selection pressure (Schmidt-Dannert *et al.*, 1999). Besides identification of improved variants for the desired property according to the settled rule "you get what you screen for" (Arnold *et al.*, 1999), the developed screening system has to offer scanning through a statistically meaningful fraction of the generated sequence space in a cost and time efficient manner (Wong *et al.*, 2006). The screening system has to be individually developed for each enzyme, should be optimized for medium to high throughput screening and sensitive over the final application range of

fragment assembly, based on chemical cleavage of phosphorothioated nucleotides by ethanol-iodine (Dennig *et al.*, 2011).

The method of choice for generation of the mutant library in a directed evolution campaign strongly depends on the enzyme property which should be optimized or re-engineered as well as on the screening throughput. Error-prone based mutagenesis methods, which are leading mainly to similar amino acid substitutions, could be beneficial for tailoring of enzyme properties like enantioselectivity (Arnold, 2001; Reetz *et al.*, 1999), whereas engineering of thermostability, requiring many mutations leading to chemically different amino acid substitutions, might be achieved using SeSaM method. All in all, nowadays there are plenty diversity generation methods available (OmniChange, SeSaM, etc.) to generate huge and highly diverse mutant libraries ($\sim 10^{15}$ mutants) covering and investigating the sequencing space of the gene of interest in order to tailor biocatalysts with improved and novel properties for industrial or pharmaceutical applications (Reetz *et al.*, 2008). In summary, these novel methods for diversity generation enable the generation of mutant libraries of high mutational load requiring potent screening formats to sample through the generated diversity.

1.2. Screening systems

As a huge variety of methods for the generation of highly diverse mutant libraries is already available (Ruff *et al.*, 2013), currently the biggest bottleneck of directed evolution campaigns is the ability to screen a significant portion of the generated sequence space, as for a protein of only 100 amino acids the theoretical sequence space comprises 20^{100} ($\sim 1.27 \times 10^{130}$) different sequences. The first requirement for designing a screening system, is its complementarity to the final application conditions (*i.e.* substrate, temperature, environmental stress). The established screening system has to mimic the application conditions as close as possible, in order to apply a correct selection pressure (Schmidt-Dannert *et al.*, 1999). Besides identification of improved variants for the desired property according to the settled rule "you get what you screen for" (Arnold *et al.*, 1999), the developed screening system has to offer scanning through a statistically meaningful fraction of the generated sequence space in a cost and time efficient manner (Wong *et al.*, 2006). The screening system has to be individually developed for each enzyme, should be optimized for medium to high throughput screening and sensitive over the final application range of

Figure 51: Michaelis-Menten plots of CelA2-WT, CelA2-H288F, and variant CelA2-H288F-M3. Enzyme kinetics were obtained with 4-MUC activity assay at pH 7.2 in KPI-buffer. 4-MUC substrate concentrations between 2.5 µM and 200 µM were used for determination of kinetic parameters with 4-MUC activity assay. Initial activities were plotted against different substrate concentrations and fitted with Michaelis-Menten equation using GraphPad Prism 6 software for analysis and determination of kinetic parameters. The reported values are the average of three measurements and deviations are calculated from the corresponding mean values. The x-axis shows the 4-MUC substrate concentration in [µM] and the y-axis shows the 4-MU product formation in [µM*s^{-1}].

3.5. *Discussion*

The developed uHTS platform combining cell-free compartmentalization with flow cytometer-based analysis and sorting accomplishes a throughput of 6.5 x 10^6 events per hour, yielding over 1.4 x 10^7 screened events in one single experiment. The platform relies on compartmentalization of a highly diverse cellulase gene library (>10^8) within (w/o/w) emulsion compartments containing fluorogenic substrate and cell-free production mix. Cell-free transcribed and translated active cellulase variants catalyze the conversion of the fluorogenic substrate into the fluorescent product which is used as fluorescent marker for labelling of (w/o/w) compartments. Subsequently, fluorescent emulsion compartments can be discriminated from non-fluorescent ones by flow cytometry, offering isolation and enrichment of genes encoding for active cellulase variants by many fold, and therefore significantly minimize screening efforts in MTP formats.

Different parameters, like incubation time, addition of BSA and chaperones, substrate concentration, DNA amount, and incubation temperature were optimized in order to achieve eminent cell-free cellulase expression levels within (w/o/w) emulsions within minimal time. For the second model enzyme phytase, cell-free enzyme production was optimized in reaction tubes, but unfortunately the obtained cell-free produced enzyme concentration was too low to be applied within (w/o/w) emulsion. A possible reason for the low protein yield could be the neutral pH of the cell-free enzyme expression mixture (~pH 7.0) which is not matching to the acidic pH optimum of active phytase (Shivange *et al.*, 2012). Further reasons for the low activity level upon cell-free phytase production might be the incorrect folding of the tetrameric phytase enzyme leading to a high production of non-functional phytase or the poor signal-to-noise ratio due to the high background activity generated by *E.coli* internal phosphatases which are able to cleave the 4-MUP substrate (Figure 17). For cellulase, the obtained elevated enzyme concentration within femtoliter volume scale of (w/o/w) emulsions enables analysis and sorting by flow cytometer due to the excellent signal-to-noise ratio between active and inactive variants.

The developed IVC-flow cytometry-based screening platform was validated in a directed evolution experiment by screening of an epPCR-CelA2-H288F linear template library (0.05 mM MnCl$_2$) using cellulase as a model enzyme of industrially important biocatalysts (biofuel production; 82.7 billion USD world-market in 2011, Pike research; 95.2 USD in 2012, Clean Edge, Inc., 2013). For the 0.05 mM MnCl$_2$ CelA2-epPCR library more than 14 million

events were screened and over one million events were sorted in a single round of directed cellulase evolution. The active fraction of the library was enriched from 8.5 % to 69 % highly active variants after only one round of sorting (Figure 46). In the MTP-based rescreening, the variant M1 was finally identified and exhibits with a specific activity of 220.6 U mg^{-1} a 13.3-fold increase compared to CelA2-WT (Table 7). Our result reported here indicates that (w/o/w) emulsions are stable enough to withstand the pressure and shear forces of flow cytometer, and that no cross talk between (w/o/w) emulsions is observed. The presented technology constitutes the success story using compartmentalization and cell-free expression in (w/o/w) emulsion for direct sorting by fluorescent signal. *In vitro* enzyme expression from a gene library using cell-free extract encapsulated in aqueous droplets of single (w/o) emulsions involves no cloning or transformation and overcomes the challenge of diversity loss during the transformation into *in vivo* expression hosts, and facilitates the possibility to produce membrane bound or toxic proteins in a cell-free environment.

The frequently reported leakage of fluorescent products from the inner aqueous phase of the emulsion into the oil phase and external aqueous phase was successfully minimized by addition of BSA (1 %) to the *in vitro* reaction mixture (Figure 40), (Courtois *et al.*, 2009). In this case, BSA acts as a sacrificial substrate that saturates the water-oil interface of the (w/o/w) emulsion. This supports the concentration of components of the cell-free expression mixture, substrate and fluorescent product within the emulsion compartment and prevents their leakage from the inner aqueous phase into the oil phase and external aqueous phase of the emulsion (Courtois *et al.*, 2009). Additionally, the fluorescein-based substrate FDC was applied, which was reported to show negligible product leakage from the emulsion in comparison to coumarin- or rhodamine-based fluorescent products (Mazutis *et al.*, 2013; Skhiri *et al.*, 2012) while showing excellent sensitivity (Figure 35F). DNA was recovered with a rapid and efficient protocol from sorted (w/o/w) emulsions using DNA purification kit (Macherey-Nagel GmbH & Co. KG) which showed higher sensitivity towards low amounts of template DNA (<5 ng) in comparison to the other tested methods (hexane/oil and decane method) as shown in Figure 47. Further optimization of cell-free cellulase production was achieved by modifications on the DNA level, as removal of the N-terminal His-tag resulted in 20 % increase in active protein concentration (Figure 30), indicating there might be room for further improvements regarding addition of different tags (*e.g.* Streptavidin tag), changes in the distance and nucleotide sequence between

ribosome binding site and start codon or between promotor and start codon in order to investigate changes on the cell-free enzyme expression level. Additionally, the application potential of cell-free enzyme evolution is an emerging industrial market, especially for enzymes from human or animal origin beyond standard bacterial and yeast proteins, or toxic and membrane proteins for which *in vivo* expression remains challenging. Furthermore, most commonly used production strains in industry, *i.e. Aspergillus niger*, *Bacillus* and *Trichoderma*, are often limited in transformation efficiencies and directed evolution in *E. coli* or yeast strains can yield different results due to differences in glycosylation or result in further codon usage optimization for high level expression. A solution to overcome these limitations could be the usage of novel eukaryotic and vesicle containing cell-extracts, *e.g.* insect extracts and human-based cell-free extracts, which enable the N-linked glycosylation of the target protein and the embedding of membrane proteins into microsomal membranes.

Current state of the art on directed evolution based on *in vitro* compartmentalization technology comprise only relatively few reports in the past 10 years (Griffiths *et al.*, 2003; Mastrobattista *et al.*, 2005; Stapleton *et al.*, 2010). Reports comprise either screening system validation on model enzyme library, *i.e.* [FeFe] hydrogenase, (Stapleton *et al.*, 2010) or usage of microbead antibody-antigen based display for maintaining genotype-phenotype linkage (Griffiths *et al.*, 2003). Additionally, the described screening systems require synthesis of a specific substrate for each enzyme. Soluble commercially available substrate fluorescein-di-β-D-galactopyranoside was used for directed *in vitro* β-galactosidase evolution where the starting enzyme had negligible galactosidase activity and resulted in significantly improved enzyme variant after two rounds of sorting (Mastrobattista *et al.*, 2005).

Here we report for the first time cell-free cellulase directed evolution using (w/o/w) compartmentalization technology where fluorogenic substrate is embedded in the emulsion droplet with no need of physical linkage. The presented technology opens potential application for tailoring different enzyme classes with suitable commercial available fluorogenic substrates.

In summary, the presented technology proposes a rapid, cost efficient and nonlaborious alternative for labelling (w/o/w) emulsions and performing screening of highly diverse mutant libraries using flow cytometry. The uHTS-IVC platform can – from our point of view – be adapted to other enzyme classes since it uses a widely applicable fluorescence sorting of

the substrate concentration, enabling the identification of desired hits in the generated mutant library (Aharoni *et al.*, 2005b).

The most common screening systems are solid phase, such as agar plates (Table 1), filter papers or membranes, offering high throughput of 10^4-10^6 clones/week, but having less detail level, as they only allow differentiation between active and inactive variants and are therefore widely used as prescreening methods (Lin *et al.*, 2002). Microtiter plate (MTP) based screening systems (Table 1) offer medium throughput (10^3-10^5 clones/week) and are mainly based on colorimetric and fluorimetric detection systems (Cohen *et al.*, 2001). MTP-based systems are in contrast to solid phase based screening methods expensive and labor intensive, but more accurate than solid phase screening methods (Olsen *et al.*, 2000). Flow cytometer-based screening technologies and microfluidic devices, offering **ultrahigh throughput screening (uHTS)** capabilities (10^7 events/hour), are becoming the most promising screening format in directed evolution experiments as they are able to screen a significant fraction of the generated sequence space in a cost-efficient and non-laborious manner (Aharoni *et al.*, 2005b; Farinas, 2006).

Table 1: Comparison of screening and selection technologies (Leemhuis *et al.*, 2009).

Strategy	Library size	Advantage	Disadvantage
Selection	~10^9	Yields desirable variants only	Only possible if activity gives growth advantage
Agar plate screen	~10^5	Simple to operate	Limited dynamic range
Microtiter plate screen	~10^4	All analytical methods possible, excellent dynamic range	Relatively low screening capacity
Cell-in-droplet screen	~10^9	Large libraries	Fluorescence detection and DNA modifying enzymes only
Cell as microreactor	~10^9	Large libraries	Fluorescence detection only
Cell surface display	~10^9	Large libraries	Fluorescence detection only
In vitro compartmentalization	~10^{10}	No cloning steps, large libraries	Fluorescence detection and DNA modifying enzymes only

1.3. Ultrahigh throughput screening (uHTS)

In order to identify improved enzyme variants with the desired properties (increased pH-stability, activity or thermostability) for industrial applications it is necessary to develop efficient screening systems offering high or ultrahigh throughput for significant coverage of

3.6. Material and Methods

All chemicals used were of analytical reagent grade or higher quality and purchased from Sigma-Aldrich Chemie GmbH (Taufkirchen, Germany), Applichem (Darmstadt, Germany), or Carl Roth (Karlsruhe, Germany). All enzymes were purchased from New England Biolabs (Frankfurt, Germany). Polymerases (PfuS, Taq) were prepared in Schwaneberg group. Fluorogenic substrate fluorescein-di-β-D-cellobioside (FDC) was synthesized by AAT Bioquest, Inc. (Sunnyvale, CA, USA) and purchased from BIOMOL GmbH (Hamburg, Germany). Fluorogenic substrate 4-methylumbelliferyl-β-D-cellobioside (4-MUC) was purchased from Sigma–Aldrich Chemie GmbH (Taufkirchen, Germany). *In vitro* transcription-translation kit was purchased from RiNA GmbH (Berlin, Germany). Plasmid isolation, PCR purification and His-tag purification kits were purchased from Macherey-Nagel GmbH & Co. KG (Düren, Germany). Microtiter plates (96-wells) for cultivation (Greiner Bio-one GmbH, Frickenhausen, Germany) and *in vivo* protein expression (Corning Incorporated, New York, USA) were incubated in a Multitron II Infors shaker (Infors AG, Bottmingen, Switzerland).

In all PCRs a thermal cycler (Mastercycler gradient; Eppendorf, Hamburg, Germany) and thin-wall PCR tubes (Multi ultra-tubes; 0.2 ml; Carl Roth GmbH, Karlsruhe, Germany) were used. DNA concentrations in all experiments were quantified using a NanoDrop photometer (ND-1000, NanoDrop Technologies, Wilmington, DE, USA). For detection of fluorescence Tecan Infinite®M1000 (Tecan Group Ltd., Männedorf, Switzerland) plate reader (λ_{ex}: 488 nm, λ_{em}: 530 nm for FCB and $\lambda_{ex.}$: 330 nm, λ_{em}: 450 nm for 4-MUC) or the flow cytometer (BD Influx Cell Sorter, Becton Dickinson Biosciences, Erembodegem, Belgium) were used. The Experion system from Bio-Rad Laboratories GmbH (München, Germany) and Pierce® BCA Protein Assay Kit (No. 23225, Life Technologies GmbH, Darmstadt, Germany) were used for estimation of protein concentrations and protein purity analysis according to the manual's instructions.

3.6.1. Strains, plasmids, and target gene

E.coli DH5α (Stratagene, La Jolla, CA, USA) and *E.coli* XL10 Gold (Stratagene, La Jolla, CA, USA) were used as cloning hosts. *E.coli* BL21 Gold (DE3) (Aglient Technologies, Santa Clara, CA, USA) and *E.coli* BL21 Gold(DE3)lacI[Q1] (Blanusa *et al.*, 2010) were used for the expression and generation of epPCR cellulase libraries. Plasmids pET-28a(+) from Novagen (Darmstadt,

Germany) and pIX3.0-RMT7 modified in house using Qiagen vector pIX3.0 (Hilden, Germany), were used as *in vivo* and *in vitro* expression vectors.

The gene celA2 (GenBank: JF826524.1) with 41 % similarity to a Glycosyl Hydrolase Family 9 (GH9) cellobiosidase from *Clostridium cellulovorans* codes for a cellulase protein with a molecular weight of ~69 kDa, consisting of 604 amino acids (Ilmberger *et al.*, 2012). Positions 288 and 580 are reported to be key residues for activity, so in this study variant CelA2-H288F, showing 8.7-fold higher specific activity than CelA2-WT, was used for development of the uHTS-IVC platform (Lehmann *et al.*, 2012). The amino acid substitution from glutamic acid to glutamine at the catalytic residue 580 is reported to lead to a complete loss of activity, therefore variant CelA2-H288F-E580Q, showing a behavior like the empty vector pET28a(+), was used as negative control in this study (Lehmann, 2013).

3.6.2. *Gene cloning into expression vectors and sequencing*

The construct pET28a(+)-CelA2-H288F was modified by introducing an additional restriction *Nde*I site in front of His-tag sequence by PCR. Subsequently CelA2-H288F gene was cut out using double restriction (*Nde*I and *Xho*I) and subcloned into pIX3.0RMT7 vector backbone. The *Nde*I-CelA2-H288F insert was generated by PCR (98°C for 30 s, 1 cycle; 98°C for 15 s / 63°C for 15 s / 72°C for 90 s, 25 cycles; 72°C for 5 min, 1 cycle) using dNTP mix (0.2 mM), primers (NdeI_CelA2_FW: 5' – ata taa cat ATG GGT AGC AGC CAT CAC - 3'; T7_pET28_REV: 5' – GTT ATT GCT CAG CGG TGG CAG CAG C - 3'; 0.4 μM each), plasmid DNA template (CelA2-H288F, 20 ng), NEB Phusion HF polymerase (2.5 U) in a final volume of 50 μl. The amplified PCR product was digested (1 h, 37°C) by *Dpn*I (20 U), purified using the NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany) and eluted in 20 μl ddH$_2$O. The recombinant plasmid was named pIX3.0RMT7-CelA2-H288F.

The inactive variant CelA2-H288F-E580Q was generated by Site-Directed Mutagenesis (SDM) according to the published method (Wang *et al.*, 2002) using pIX3.0RMT7-CelA2-H288F as template DNA and specific SDM primers (SDM_E580Q_for: 5' - GCT ATG CCA CCA AT**C AG**A TTT GCA TTT ATT GGA ATA GTC CG - 3', SDM_E580Q_rev: 5' - CGG ACT ATT CCA ATA AAT GCA AAT **CTG** ATT GGT GGC ATA GC - 3'). The recombinant plasmid was named pIX3.0RMT7-CelA2-H288F-E580Q. Constructs were digested (1 h, 37°C) by *Dpn*I (20 U), purified using NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG) and transformed into chemically competent *E.coli* BL21 Gold(DE3)lacI^{Q1} cells. DNA sequencing of

the inserted gene was performed by Eurofins Genomics (Ebersberg, Germany) and Clone Manager 9 Professional Edition (Sci-Ed software, Cary, USA) was used for sequence analysis.

3.6.3. Removal and addition of His-Tag in phytase and cellulase constructs

3.6.3.1. His-tag insertion in the phytase construct

For generation of a phytase construct containing the HIS-tag the *Nde*I site was changed in the original pIX3.0RMT7-YmPhWT construct by SDM to a *Nhe*I-ATG site to obtain the construct pIX3.0RMT7-*Nhe*I-ATG-YmPhWT and to maintain the original start codon for phytase transcription and translation. The constructs were generated by two step PCR. In step one, two extension reactions were performed in separate tubes; one containing the forward primer and the other containing the reverse primer. Therefore, the master mix was split in 2 x 24.5 µl and the primers were added up to a final volume of 25 µl. In the second step, the two reactions were mixed and the PCR program was continued (Step I: 98°C for 30 s, 1 cycle; 98°C for 10 s / 55°C for 30 s / 72°C for 4 min, 3 cycles; 72°C for 3 min, 1 cycle and Step II: 98°C for 30 s, 1 cycle; 98°C for 10 s / 55°C for 1 min / 72°C for 4 min, 14 cycles; 72°C for 5 min, 1 cycle) using dNTP mix (0.2 mM), primers (YmPh-*Nhe*I-ATG-FW and -RV; 20 µM each; Table 8), plasmid DNA template (pIX3.0RMT7-YmPhWT, 50 ng), *PfuS* polymerase (2.5 U), dNTPs (10 mM) in a final volume of 50 µl. The amplified PCR product was digested (1 h, 37°C) by *Dpn*I (20 U), purified using the NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG) and eluted in 20 µl ddH$_2$O. Restriction digest was performed with the construct pIX3.0RMT7-CelA2-H288F, containing the vector backbone with His-tag, and in parallel with the PCR product pIX3.0RMT7-*Nhe*I-ATG-YmPhWT using the enzymes *Nhe*I and *Sap*I. Restriction digests were analyzed on agarose gel (0.8 %), DNA was extracted and purified from gel with NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG) and eluted in 20 µl ddH$_2$O. The vector backbone with His-tag (pIX3.0RMT7) and the insert (*Nhe*I-ATG-YmPhWT) were ligated to obtain the final construct pIX3.0RMT7-YmPhWT-HIS-ATG (Figure 52A). Ligation and restriction were performed according to the manufacturer's recommendations (New England Biolabs, Frankfurt a. Main, Germany).

3.6.3.2. His-tag removal from cellulase construct

For generation of pIX3.0RMT7-CelA2-H288F-no HIS the *Nde*I site was changed in the pIX3.0RMT7 (empty vector) construct by SDM (see 3.6.3.1, pIX3.0RMT7_GW-NheI-FW and -RV, each 20 µM, Table 8) to a *Nhe*I site to obtain the construct pIX3.0RMT7_GW-*Nhe*I without His-tag (Figure 52C). The amplified PCR product was digested (1 h, 37°C) by *Dpn*I (20 U), purified using the NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG) and eluted in 20 µl ddH₂O. Restriction digest of the PCR product pIX3.0RMT7_GW-*Nhe*I, containing the vector backbone without His-tag, and the construct pIX3.0RMT7-CelA2-H288F, containing the insert, with the enzymes *Nhe*I and *Sap*I were performed in parallel, followed by DNA extraction and purification from the agarose gel using the NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG). Subsequently, the CelA2-H288F insert was ligated with the pIX3.0RMT7_GW-*Nhe*I vector backbone to obtain the final construct pIX3.0RMT7-CelA2-H288F-no HIS without His-tag (Figure 52B). Ligation and restriction were performed according to the manufacturer's recommendations (New England Biolabs, Frankfurt a. Main, Germany).

After ligation both constructs (pIX3.0RMT7-CelA2-H288F-no HIS and pIX3.0RMT7-YmPhWT-HIS-ATG) were transformed into chemically competent *E.coli* BL21 Gold(DE3)lacI[Q1]. DNA was isolated (NucleoSpin® Plasmid kit, Macherey-Nagel GmbH & Co. KG), from transformed clones and correct sequence of the constructs was confirmed by sequence analysis.

Table 8: Primer for removal and insertion of His-tag in cellulase and phytase constructs.

Primer	Sequence 5' - 3'
YmPh-*Nhe*I-ATG-FW	CTTTAAGAAGGAGATAA**GCTAGCATG**CGATTAACTGCACTGG
YmPh-*Nhe*I-ATG-REV	CCAGTGCAGTTAATCGCATGCTAGCTTATCTCCTTCTTAAAG
pIX3.0RMT7_GW-NheI-FW	CTTTAAGAAGGAGATAA**GCTAGC**TCTGAGCGCCGCTGCG
pIX3.0RMT7_GW-NheI-REV	CGCAGCGGCGCTCAGAGCTAGCTTATCTCCTTCTTAAAG
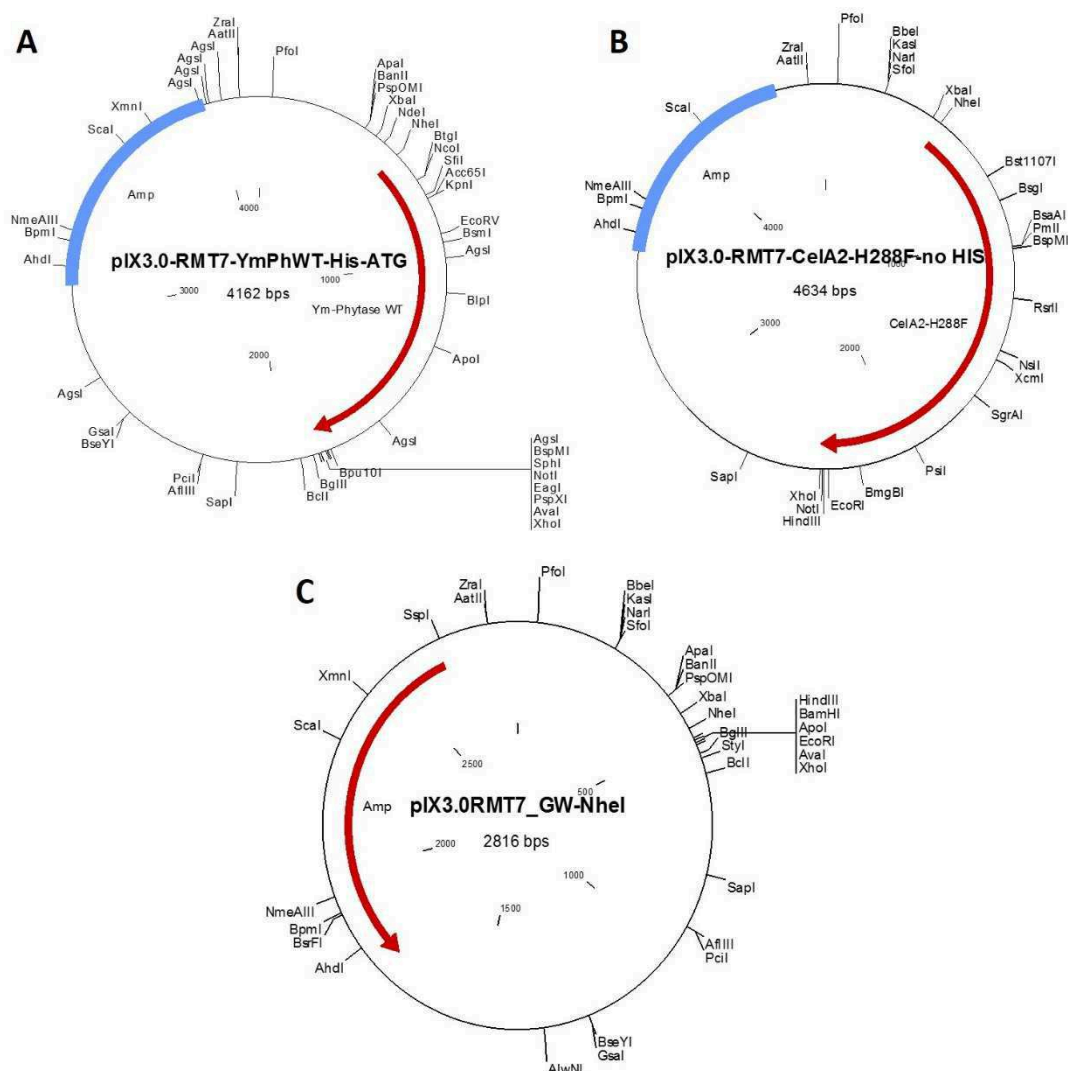
Figure 52: Vector maps of constructs (**A**) pIX3.0RMT7-YmPhWT-HIS-ATG, (**B**) pIX3.0RMT7-CelA2-H288F-no HIS and (**C**) pIX3.0RMT7_GW-NheI.

3.6.4. Fluorimetric MTP-based assay for determination of phosphatase activity

3.6.4.1. 4-MUP activity assay for 384-well plate

For continuous phytase activity detection the highly sensitive fluorescence-based 4-MUP activity assay (Shivange *et al.*, 2012) for quantification of phosphorolytic activity was adapted to 384-well MTPs (polystyrene, black, flat bottom; Greiner Bio–one GmbH, Frickenhausen, Germany). Reactions had a final volume of 30 µl. For the assay 25 µl fluorogenic substrate 4-methylumbelliferyl-phosphate (4-MUP; Sigma–Aldrich Chemie GmbH; 5 mM dissolved in buffer, 250 mM sodium acetate, 1 mM $CaCl_2$, 0.01 % Tween-20, pH 5.5) were added to 5 µl of cell supernatant, purified enzyme or *in vitro* expressed sample. Enzyme activity was measured at 37°C by monitoring the increase in relative fluorescence (λ_{ex} 360 nm, λ_{em} 465 nm), due to the hydrolysis of 4-MUP to the fluorescent product

4-methylumbelliferone (4-MU), over time using Tecan Infinite®M1000 plate reader (Tecan Group Ltd., Männedorf, Switzerland).

3.6.5. Fluorimetric MTP-based assays for determination of cellulolytic activity

3.6.5.1. 4-MUC activity assay for 384-well plate

For continuous cellulase activity detection the highly sensitive, fluorescence-based 4-MUC activity assay (Lehmann *et al.*, 2012) for quantification of cellulolytic activity was adapted to 384-well MTPs (polystyrene, black, flat bottom; Greiner Bio–one GmbH, Frickenhausen, Germany). Reactions had a final volume of 20 µl. For the assay 10 µl fluorogenic substrate 4-methylumbelliferyl-β-D-cellobioside (4-MUC; Sigma–Aldrich Chemie GmbH; 0.1 mM dissolved in potassium phosphate buffer, 0.2 M, pH 7.2) were added to 10 µl of cell supernatant, purified enzyme or *in vitro* expressed sample. Enzyme activity was measured at 30°C by monitoring the increase in relative fluorescence (λ_{ex} 330 nm, λ_{em} 450 nm), due to the hydrolysis of 4-MUC to the fluorescent product 4-methylumbelliferone (4-MU), over time using Tecan Infinite®M1000 plate reader (Tecan Group Ltd.).

3.6.5.2. FDC activity assay

For continuous cellulase activity detection a highly sensitive, fluorescence-based fluorescein-di-β-D-cellobioside (FDC) activity assay for quantification of cellulolytic activity was developed employing 384-well MTPs (polystyrene, black, flat bottom; Greiner Bio–one GmbH, Frickenhausen, Germany). Reactions had a final volume of 20 µl. For the assay 10 µl fluorogenic substrate fluorescein-di-β-D-cellobioside (BIOMOL GmbH; 0.1 mM dissolved in potassium phosphate buffer, 0.2 M, pH 7.2) were added to 10 µl of cell supernatant, purified enzyme or *in vitro* expressed sample. Enzyme activity was measured at 30°C by monitoring the increase in relative fluorescence (λ_{ex} 494 nm, λ_{em} 516 nm), due to the hydrolysis of FDC to the fluorescent product fluorescein (Figure 53), over time using Tecan Infinite®M1000 plate reader (Tecan Group Ltd.).
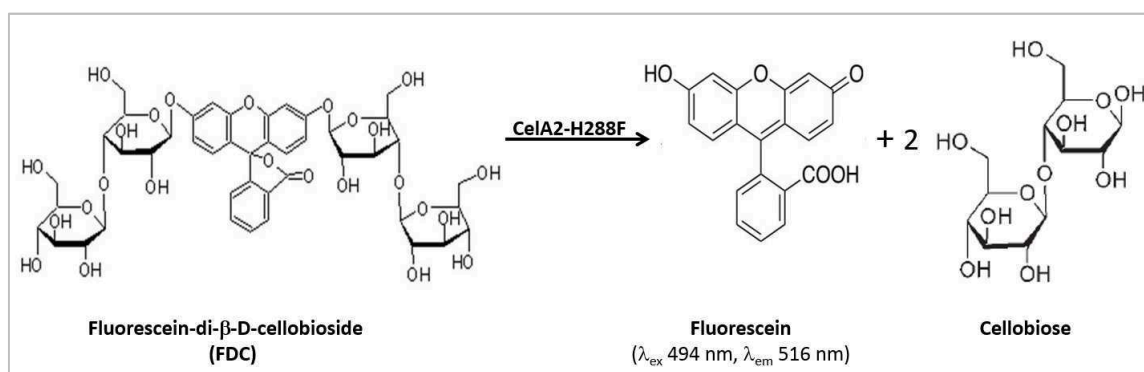
Figure 53: Principle of the fluorescein-di-β-D-cellobioside (FDC) detection system.

3.6.6. Generation of epPCR mutagenesis library of CelA2

The random mutagenesis library was constructed by standard epPCR method (Cadwell *et al.*, 1994). The epPCR library was generated by PCR (94°C for 1 min, 1 cycle; 94°C for 30 s / 51°C for 30 s / 72°C for 90 s, 25 cycles; 72°C for 5 min, 1 cycle) using dNTP mix (0.2 mM), primers (pIX3.0RMT7_FW: 5' - TGC AAG GCG ATT AAG TTG GG - 3'; pIX3.0RMT7_REV: 5' - ACC CCA GGC TTT ACA CTT TAT G - 3'; 0.4 µM each), plasmid DNA template (CelA2-H288F, 20 ng), *Taq* polymerase (2.5 U), and $MnCl_2$ (0.05 mM) in a final volume of 50 µl. The amplified epPCR product was digested (1 h, 37°C) by *Dpn*I (20 U), purified using the NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG), eluted in 15 µl ddH_2O and used for *in vitro* expression of the linear template library.

3.6.7. Detection of cellulase activity in emulsions

Upon *in vitro* transcription and translation of the cellulase DNA within the "*in vitro* expression mixture" active cellulases are able to convert the fluorogenic substrate fluorescein-di-β-D-cellobioside (FDC) into the fluorescent product fluorescein and two glucose units (comprising cellobioside) by cleavage of the β-1,4-D-glycosodic bonds. The product fluorescein can be detected within the (w/o/w) emulsions by fluorescence measurements (λ_{ex} 494 nm, λ_{em} 516 nm) in MTP- or flow cytometry-based assay formats. Due to the presence of charged COO^- group, the fluorescein showed low difussion into the oil phase of the emulsion droplet. Emulsion droplets harbouring active cellulase variants show a green fluorescence due to substrate conversion, making FDC applicable for flow cytomtery screening.

the generated sequence space (~10^9 different DNA molecules using PCR-based random mutagenesis methods for diversity generation) (Agresti *et al.*, 2010; Wong *et al.*, 2006). Therefore, ultrahigh throughput screening (uHTS) employing flow cytometer-based methods, which were developed in the 1970`s to analyze mammalian cells (Givan, 2001), are becoming more and more attractive for protein engineering and bacterial genetics (Daugherty *et al.*, 2000). The employment of uHTS in directed evolution campaigns is of significant scientific importance, as it increases the probability to efficiently identify beneficial variants from huge mutant libraries of high mutational loads and enables the identification and exploration of novel structure-function relationships and combinatorial effects in proteins. The uHTS technology offers qualitative differentiation between positive and negative activity enabling enrichments of active variants. It is a powerful analytical tool, especially in the fields of medicine, biology and biotechnology, as with this method up to 100,000 events per second can be analyzed for up to 13 different parameters per cell (Givan, 2001). With flow cytometry *e.g.* large libraries of proteins displayed on bacterial, yeast and mammalian cells can be screened quickly for binding interactions or enzymatic activity based on fluorescence signal, as one of the most sensitive technologies for the detection of activity (Bernath *et al.*, 2004a). For screening employing flow cytometry it is essential to maintain the linkage between genotype and phenotype and to enable isolation of improved variants by differentiation between positive and negative variants (Mastrobattista *et al.*, 2005). Thereby its applications are mainly based on **whole cell systems** (Table 1), in which the fluorescent product of the protein variant is either entrapped on the cell surface (Olsen *et al.*, 2000) or in the cell (Kawarasaki *et al.*, 2003). In the last decade, novel **compartmentalization-based techniques** (Table 1) are used to entrap the cell or gene together with the fluorogenic substrate and the fluorescent product (converted by active protein variants), in aqueous microdroplets or connected to microbeads residing inside the microdroplets. Flow cytometer-based screening systems enable in combination with compartmentalization techniques the rapid analysis of cell-free generated enzyme libraries, which significantly accelerates the screening efforts during directed evolution campaigns (Mastrobattista *et al.*, 2005). Employment of cell-free library generation, independent from any kind of host cells (*e.g.* bacterial or yeast), accelerates the screening efforts for directed evolution campaigns, offering qualitative differentiation between active and inactive enzyme

PBS buffer) and fluorescein (100 µl of 6 µM fluorescein sodium salt in Tris/HCl buffer pH 8.0, 0.1 M) were used.

For optimization of emulsification using stirring and homogenizing method different emulsification speeds, emulsification times and ratios of W1 to W2 solution were tested.

3.6.10. Compartmentalization in emulsions by extrusion

For optimization of emulsification using extrusion method different W1 to W2 solution were tested in various combinations (Table 9). The best results were obtained by generation of the (w/o) emulsions using extrusion method with an Avanti Mini-Extruder (Avanti, Polar Lipids, Inc., Alabaster, USA) by adding 200 µl of W1 solution (1.25 g Span 80, 0.25 g Tween 80 dissolved in 60 ml light mineral oil) to 100 µl PBS buffer (pH 7.4, 1.06 mM KH_2PO_4, 2.97 mM Na_2HPO_4, and 155.17 mM NaCl) containing fluorescein (60 µM) and pushing the solution tree times though 5 µm Whatman membrane (Whatman Nuclepore Track-Etched Membranes, Sigma-Aldrich Biochemie GmbH, Hamburg, Germany) with the extruder device. Subsequently the second emulsification was performed by addition of 700 µl of W2 solution (0.5 % Tween 80 in PBS buffer (1.06 mM KH_2PO_4, 2.97 mM Na_2HPO_4, and 155.17 mM NaCl, pH 7.4)) to the (w/o) emulsion and three times extrusion though 12 µm Whatman membrane (Sigma-Aldrich Biochemie GmbH) to obtain (w/o/w) emulsions. The optimized protocol was applied for all further experiments using emulsion generation by extrusion for cell-free protein synthesis and subsequent flow cytometry analysis and sorting.

Table 9: For optimization of extrusion method different W1 and W2 solutions were tested.

W1 solution	W2 solution	Method applied	Reference
2.9 % (w/w) ABIL EM90 in light mineral oil	1.5 % (w/v) carboxymethyl-cellulose and 2 % (w/v) Tween 20 in PBS	Homogenizing/ Stirring/ Extrusion	Blanusa, 2009, (Blanusa, 2009)
1 % (w/v) Span 60 and 1 % (w/v) cholesterol in decane	0.5 % (w/v) Tween 80 in PBS buffer	Extrusion	Mastrobattista, 2005, (Mastrobattista *et al.*, 2005)
2 % (w/w) Span 80 and 0.4 % (w/w) Tween 80 in mineral oil (W1.1)	**0.5 % (w/v) Tween 80 in PBS buffer (W2.2)**	**Extrusion**	**This work**
4.5 % Span 80, 0.5 Tween 80 in mineral oil (W1.2)	2 % (w/v) Tween 20 in PBS buffer (W2.1)	Extrusion	This work
Pico-Surf™2, 5% (w/w) in FC-40 (Dolomite Microfluidics, Charlestown, MA, USA) (W1.3)	2 % (w/v) SDS in PBS buffer) (W2.3)	Extrusion	This work

3.6.11. Compartmentalization by extrusion and cell-free production of CelA2 from linear and plasmid DNA template in emulsions

Cell-free cellulase production was performed in emulsion compartments using the FastLane *E.coli* Mini Kit (RiNA GmbH, Berlin, Germany) in combination with extrusion method. Optimal cell-free cellulase production in (w/o/w) emulsions was achieved by optimizing parameters such as incubation temperature, incubation time, amount of DNA template, substrate concentration and BSA concentration. For successful *in vitro* transcription and translation, linear CelA2-H288F DNA (0.656 µM linear PCR product) or a CelA2 linear template library (0.656 µM linear epPCR product generated using 0.05 mM $MnCl_2$) was mixed with the *in vitro* expression mixture (35 µl *E.coli* extract, 40 µl Easy Xpress reaction buffer, 0.8 µl Chaperone-Mix (DnaK, DnaJ, GroE), 1.15 µl of BSA (100 mg/ml), 27 µl substrate (2 mM; final conc. 0.46 mM FDC), RNase free water added to 115 µl and incubated within (w/o) emulsion compartments (4 h, 25°C, in water bath). If plasmid DNA or plasmid CelA2-H288F DNA libraries (0.656 µM plasmid DNA) were used instead of a CelA2 linear template DNA the incubation was performed at 30°C (instead of 25°C).

The (w/o) emulsions were generated by extrusion method with an Avanti Mini-Extruder (Avanti, Polar Lipids, Inc., Alabaster, USA) by adding 200 µl of W1 solution (1.25 g Span 80, 0.25 g Tween 80 dissolved in 60 ml light mineral oil) to the 115 µl *in vitro* reaction mixture and pushing the solution three times though 5 µm Whatman membrane (Whatman Nuclepore Track-Etched Membranes, Sigma-Aldrich Biochemie GmbH, Hamburg, Germany) with the extruder device. The (w/o) emulsion sample containing the *in vitro* reaction mixture was collected in a 0.5 ml Eppendorf tube for incubation (4 h, 25°C). Subsequently the second emulsification was performed by addition of 700 µl of W2 solution (0.5 % Tween 80 in PBS buffer (1.06 mM KH_2PO_4, 2.97 mM Na_2HPO_4, and 155.17 mM NaCl, pH 7.4)) to the (w/o) emulsion and three times extrusion though 12 µm Whatman membrane (Sigma-Aldrich Biochemie GmbH) to obtain (w/o/w) emulsions for analysis on flow cytometer.

3.6.12. Confocal microscopy

Confocal microscopy (Leica TCS SP8, Leica Microsysteme Vertrieb GmbH, Wetzlar, Germany) was performed using 63x oil immersion objective. For excitation a continuous wave laser (DPSS, 20 mW: 488 nm) was used and for emission at 510-520 nm a highly sensitive

PMT detector was used. Z-scans, transmission and fluorescence microscopic pictures were taken from (w/o) and (w/o/w) emulsions generated by extrusion according to the optimized protocol containing fluorescein (6 µM) in PBS buffer (pH 7.4, 1.06 mM KH_2PO_4, 2.97 mM Na_2HPO_4, and 155.17 mM NaCl) in the inner aqueous phase of the (w/o/w) emulsion.

3.6.13. Flow cytometry-based screening and sorting

The (w/o/w) emulsions harboring the cell-free expressed CelA2-linear template epPCR library were diluted 1000-fold in PBS buffer (pH 7.4, 1.06 mM KH_2PO_4, 2.97 mM Na_2HPO_4, and 155.17 mM NaCl), filtered through 50 µm cell trics (Sysmex Partec GmbH, Görlitz, Germany) and analyzed with BD Influx cell sorter (Becton Dickinson Biosciences, Erembodegem, Belgium) according to forward and sideward scatter as well as to its fluorescence intensity (λ_{ex} 494 nm, λ_{em} 516 nm). The BD Influx flow cytometer was operated with a 100 µm nozzle and PBS (pH 7.4, 1.06 mM KH_2PO_4, 2.97 mM Na_2HPO_4, and 155.17 mM NaCl) was used as sheath fluid. Fluorescence intensities of (w/o/w) emulsions labelled with the converted product fluorescein were recorded, and sort gates were set to collect (w/o/w) emulsions with high green fluorescence intensity. Out of 14,628,729 analyzed events 1,026,881 were sorted (Sort mode: 1.0 Drop Pure) using 1,800 events s^{-1} analysis speed allowing a throughput of 6.48 x 10^6 per hour. Reanalysis of the sorted populations with FACS showed 8-fold enrichment of the active (w/o/w) emulsion population. Sorted (w/o/w) emulsions were collected in PBS and genes were isolated by breakage of the (w/o/w) emulsions and purification of the recovered DNA.

3.6.14. Recovery of DNA from (w/o/w) emulsions

Samples after sorting were heated up (5 min, 70°C) to break the emulsions. DNA was recovered from (w/o/w) emulsion sample by purification of sorted (w/o/w) emulsions in PBS (pH 7.4, 1.06 mM KH_2PO_4, 2.97 mM Na_2HPO_4, and 155.17 mM NaCl) with the NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG) according to the user's manual. Samples were eluted from the purification column with 15 µl ddH$_2$O. DNA amount was measured by Nano Drop 2000 UV-Vis Spectrophotometer (Thermo Scientific, Wilmington, USA). The obtained DNA amount was used as a template in a subsequent recovery PCR (94°C for 1 min, 1 cycle; 94°C for 30 s / 60°C for 30 s / 72°C for 2 min, 35 cycles; 72°C for 5 min, 1 cycle) was performed, containing *Taq* DNA polymerase (2.5 U), dNTP mix (0.3 mM),

forward PLICing primer (0.4 μM) (1_FW_Cel_PLIC: 5' - AGC AGC GGT GAA AAT CTG TAT TTT CAG G - 3'), reverse PLICing primer (0.4 μM) (1_RV_Cel_PLIC: 5' - TTC GGC CAC GGT ATT CAG GCT ATA AAC AT - 3), recovered DNA from (w/o/w) emulsions (20 ng) in a final volume of 25 μl.

3.6.15. Membrane-based desalting of recovered DNA from (w/o/w) emulsions

Extracted emulsion samples were evaporated for concentration of DNA and dry DNA pellets were dissolved in ddH$_2$0 (15 μl). For micro-dialysis of DNA after the extraction to remove residual salts MF-Millipore^{TM} membrane filters of 0.025 μm (Merck Chemicals GmbH, Darmstadt, Germany) were carefully applied on ddH$_2$O surface in a petri dish and wetted in a way that the upper side of the membrane stayed dry. The extracted, evaporated emulsion sample (6 μl) was pipetted on top of the filter membrane and desalted *via* diffusion for 30 min. Subsequently samples were removed from the filter membrane, pipetted into a fresh PCR cup and stored at -20°C. For recovery PCR of desalted sample (4 μl of 1-30 ng/μl) was applied in the PCR reaction. PCR products (5 μl PCR product mixed with 1μl loading dye (6x)) were analyzed by agarose-gelelectrophoresis on 0.8 % agarose gel.

3.6.16. Cloning of enriched, active CelA2 variants into pET28a expression vector

Amplified epPCR libraries (before and after) sorting was cloned into pET28a(+) vector backbone by PLICing for expression and activity analysis in MTP (Blanusa *et al.*, 2010). For vector backbone amplification, the pET28a(+)-CelA2 DNA template was amplified using standard PCR (98°C for 30 s, 1 cycle; 98°C for 30 s / 57°C for 30 s / 72°C for 4 min 30 s, 25 cycles; 72°C for 5 min, 1 cycle), containing *PfuS* DNA polymerase (2.5 U), dNTP mix (0.2 mM), forward PLICing primer (0.4 μM) (2_FW_Cel_PLIC: 5' - ACC GTG GCC GAA TAA TAA GAA TTC GAG C - 3'), reverse PLICing primer (0.4 μM) (2_RV_Cel_PLIC: 5' – TCA CCG CTG CTA TGA TGA TGG TGG - 3), pET28a(+)-CelA2 DNA (20 ng) in a final volume of 25 μl.

The amplified PCR products were digested (1 h, 37°C) by *Dpn*I (20 U), analyzed for correct size by agarose-TAE gel electrophoresis according to the standard protocol (Sambrook *et al.*, 2001), purified from agarose gel using the NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG) and eluted in 20 μl ddH$_2$O. Purified PCR products were used for PLICing and resulting hybridization products were transformed into 100 μl chemically competent *E. coli* BL21-Gold (DE3) cells following the standard protocol (Hanahan

variants in highly diverse gene libraries in form of a pre-screening and enrichment of active variants from a vast pool of inactive variants for subsequent detailed analysis.

1.4. *Compartmentalization technologies*

The microdroplets or compartments can be generated either by dispersion in water-in-oil (w/o) emulsions (Tawfik *et al.*, 1998) or in **water-in-oil-in-water (w/o/w) emulsions** that can be sorted by flow cytometer (Bernath *et al.*, 2004a). In addition, stable micro-compartments can be generated employing block copolymers, based on polyethylene glycol and biodegradable polyesters or polycarbonates, forming hollow vesicles, *e.g.* polymersomes, and entrapping the gene or cell of interest together with fluorogenic substrate within the polymer vesicle (Discher *et al.*, 2002; Meng *et al.*, 2005). Another method for generation of micro-compartments is the preparation of liposomes using freeze-dried empty liposomes for entrapment of genes or cells (Sunami *et al.*, 2006). Hydrogels, an alternative method applicable for many hydrolases, generate a fluorescent shell around cells producing active protein variants (Pitzler *et al.*, 2014). Additionally, micro-compartments can be generated, analyzed and sorted on-chip using microfluidic devices (Brouzes *et al.*, 2009; Kintses *et al.*, 2010).

In general, as compartmentalization can be used to entrap whole cells or cell-free mutant libraries, two types of compartmentalization can be distinguished: *in vivo* and *in vitro* compartmentalization. In the case of ***in vivo* compartmentalization** a gene library is cloned into expression host cells. The single cells are entrapped, *e.g.* in (w/o/w) emulsions, and sorted by flow cytometer (Aharoni *et al.*, 2005a). As shown in Figure 2, active cells, expressing the desired protein convert the substrate inside of the emulsion droplet into the fluorescent product. The fluorescent microdroplet with the active cell is detected and sorted directly into microtiter plates or collected and plated on agar plates for further analysis (Aharoni *et al.*, 2005a).

determined using Experion^{TM} Automated Electrophoresis System (BIO-RAD Laboratories GmbH, München, Germany) and Pierce^{TM} BCA Protein Assay Kit (Life Technologies GmbH, Darmstadt, Germany).

3.6.19. Determination of kinetic parameters for CelA2 activity with 4-MUC activity assay

For determination of kinetic parameters of purified CelA2-H288F and its variants 4-MUC activity assay was performed according to the published protocol (Lehmann *et al.*, 2012).

3.6.20. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

For SDS-PAGE of CelA2-H288F and its variants stacking gel (5 % (w/v) acrylamide) and separating gel (12 % (w/v) acrylamide) were used (Sambrook *et al.*, 2001) and stained with Coomassie Brilliant Blue (R-250).

4. Comparison of Iterative Site-Saturation Mutagenesis (ISM), multiple Site-Saturation Mutagenesis (SSM) and OmniChange approach using flow cytometer-based IVC screening platform for cellulase engineering

4.1. Project objective

The goal of this work is to investigate and compare the three different focused mutagenesis approaches: **Iterative Saturation Mutagenesis (ISM)**, **multiple Site-Saturation Mutagenesis (multiple SSM)** and **OmniChange**. In order to enable efficient sampling through the generated diversity two different screening systems will be used. Depending on the diversity generated by the respective mutagenesis method either **MTP-based screening formats** or the novel **flow cytometer-based ultrahigh throughput screening platform using *in vitro* compartmentalization technology (uHTS-IVC) in (w/o/w) emulsions** will be employed. In order to ensure the coverage of a meaningful fraction of the generated diversity upon usage of powerful mutagenesis methods like OmniChange (4 simultaneously saturated positions, 1.04^6 different gene variants) the employment of the uHTS-IVC screening platform is crucial. For all mutagenesis approaches four positions in close proximity to the substrate binding pocket (H288, Y368, H524 and Y576) in the CelA2-H288F cellulase will be saturated. Due to their location, particularly upon employment of simultaneous Site-Saturation Mutagenesis, it is possible to identify combinatorial effects among the selected positions. In the end, the beneficial amino acid substitutions and combinatorial effects identified in the three different approaches will be compared to evaluate the power of the respective mutagenesis method in combination with the employed screening technology.

The first approach standard ISM method (Reetz *et al.*, 2007) for mutagenesis of four positions in close proximity to the substrate binding pocket (288, 368, 524 and 576) in the CelA2-H288F cellulase will be used. ISM comprises single site-saturation library screening and identification of the most improved variant as template for the subsequent round of site-saturation mutagenesis. The generated site-saturation libraries will be screened employing standard MTP-based screening systems (4-MUC-activity assay and FDC activity assay). In contrast to the ISM approach, in the second approach (multiple SSM) the entire isolated active fraction will serve as a template for the next round of SSM and not only the best identified variant. Multiple SSM libraries will be screened with the uHTS-IVC platform in subsequent rounds of screening and enrichment of active cellulase variants for further

analysis in MTP. The third approach will use simultaneous SSM at the targeted positions employing the OmniChange technique (Dennig *et al.*, 2011) in combination with the uHTS-IVC platform for screening of the generated library and enrichment of active cellulase variants for further analysis in MTP. The three different highly diverse libraries will be screened for improved CelA2-H288F cellulase variants to explore possible combinatorial effects.

4.2. Introduction

4.2.1. Diversity generation using focused mutagenesis and ultrahigh throughput screening

In contrast to random mutagenesis methods for directed enzyme evolution, rational design and semi-rational protein engineering, using focused mutagenesis approaches are recommended for tailoring of biocatalysts for industrial applications, when already a certain knowledge about the enzyme was generated or reported, *e.g.* in form of crystal structures, reliable structural models or knowledge and understanding of beneficial mutations improving a specific enzyme property (Shivange *et al.*, 2009; Wong *et al.*, 2007). Several success stories were reported, which show that the enzyme properties activity and selectivity can be successfully improved using focused mutagenesis methods (Reetz *et al.*, 2010; Reetz *et al.*, 2006). The most frequently used method for single Site-Directed and Site-Saturation Mutagenesis is the principle of mutagenic plasmid amplification which was commercialized in form of a kit by Stratagene (QuickChange Site-Directed Mutagenesis Kit, La Jolla, CA, USA). This method was expanded using iterative saturation mutagenesis (ISM) which employs subsequent cycles of SSM and screening (Reetz *et al.*, 2007). A further development of iterative saturation mutagenesis is the CASTing method, a combinatorial active site-saturation test (CAST), which uses the 3D structure of the enzyme to identify amino acid positions with side chains in close proximity to the active site (Reetz *et al.*, 2006). Small mutant libraries which are generated by randomization of the respective identified positions allow the focused screening around the entire substrate binding pocket (Reetz *et al.*, 2006). The simultaneous saturation of up to three positions, in iterative rounds of CASTing at several randomization sites around the substrate binding pocket, significantly increases the possibility to identify combinatorial effects (Reetz *et al.*, 2006). During iterative rounds of Site-Saturation Mutagenesis always the best identified variant of the previous

round serves as template for the next mutagenesis cycle, like illustrated in the example scheme in Figure 54 (Reetz *et al.*, 2007).
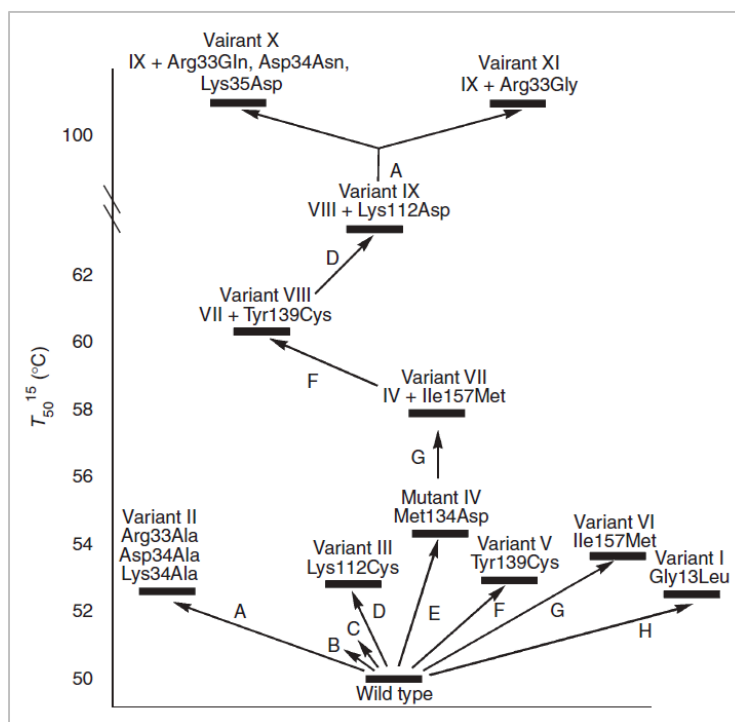
Figure 54: Scheme for iterative rounds of saturation mutagenesis at different randomization sites (A-H) on the example of the evolution of a lipase from *Bacillus subtilis* for improved thermostability. The figure is taken from Reetz *et al.* (Reetz *et al.*, 2007).

The main limitation of many multi-codon mutagenesis methods are the restricted number of codons that can be saturated simultaneously, e.g. by CASTing only two or three amino acids can be randomized by SSM simultaneously (Reetz *et al.*, 2006). In addition, there are often restrictions by the gene sequence or the requirement of subsequent PCR steps or additional enzymatic modifications (Seyfang *et al.*, 2004; Shen *et al.*, 2010) . In 2011, a rapid and sequence independent method for simultaneous site-saturation of up to five codons, which do not require a minimal distance between each other, was reported by Dennig *et al.* (Dennig *et al.*, 2011). As proof of principle, in a one-pot-reaction five chemically cleaved DNA fragments of the phytase gene from *Yersinia mollaretii*, each containing one degenerated NNK-codon, were generated and assembled without additional PCR amplification steps or use of additional enzymes (Dennig *et al.*, 2011). For these reasons simultaneous performance of multiple SSMs employing OmniChange method is a promising strategy to enhance the chances to identify combinatorial effects for successful tailoring of industrial important biocatalysis.

Figure 2: Compartmentalization of single cells in (w/o/w) emulsions and sorting with flow cytometer. (1.) A gene library is cloned and transformed into host cells resulting in a cell mutant library. (2.) Cells are expressed for production of the enzyme of interest. (3.) The mutant library is encapsulated in (w/o) emulsions and subsequently (4.) in (w/o/w) emulsions. (5.) Active cells, expressing and secreting the desired enzyme, convert the substrate inside of the emulsion droplet into the fluorescent product. (6.) The fluorescent microdroplet with the active cell is detected and sorted directly into microtiter plates or collected and plated on agar plates for further analysis.

In case of **_in vitro_ compartmentalization (IVC)** single genes are cell-free transcribed and translated within a micro-compartment, _e.g._ a (w/o/w) emulsion (Figure 3). Especially for cell-free protein production, most commonly water-in-oil-in-water (w/o/w) emulsions, shortly called double emulsions, are employed which consists of an inner aqueous phase, surrounded by an oil phase (primary (w/o) emulsion) and an additional water phase. The gene encoding the enzyme of interest is entrapped together with the substrate inside the (w/o/w) emulsion compartment. Upon the protein expression, the substrate is converted by the _in vitro_ produced enzyme into the fluorescent product and the fluorescent microdroplet can be detected by the optical and electronics system of the flow cytometer for identification of active enzymes for further enrichment (Figure 3), (Griffiths _et al._, 2003; Mastrobattista _et al._, 2005). Conversion of the primary emulsion into (w/o/w) emulsion, mimicking the lipid bi-layer of the cell-membrane, makes it possible to sort the droplets by flow cytometry (Bernath _et al._, 2004a; Bernath _et al._, 2004b). Using the IVC-based approach, the limitations of cell-based systems, _e.g._ low transformation efficiency of the expression hosts or low protein yields upon _in vivo_ expression of membrane or toxic proteins, are addressed since the genetic material is directly put into the compartment, making this technique advantageous and interesting for research and industry (Blanusa, 2009).

Figure 55: (**A**) Hybrid homology model of CelA2, which was generated with YASARA 16.3.5 software package. The four targeted positions H288, Y368, H524 and Y576, close to the active site of CelA2-WT, for focused site-saturation mutagenesis are highlighted in magenta. Amino acid E580 is a catalytically important residue (highlighted in blue), as amino acid substitution E580Q leads to complete inactivation of the enzyme. Position 288 is located at the entrance of the active site and is like the other three positions 368, 524 and 576 in close proximity to the substrate isofagomine (ball-stick representation) at the active site of the enzyme. (**B**) The surface representation of the CelA2-WT homology model shows the substrate isofagomine (PDB ID: 3RX7) in the substrate binding pocket at the active site of the enzyme, which is in close proximity to the four selected residues, highlighted in magenta.

4.3. Results

The result section is divided into three parts in order to compare the different approaches for diversity generation and screening employing the novel uHTS-IVC-based flow cytometry technology platform (Chapter I, 3.3.1) besides standard MTP-based screening formats for analysis of the ISM and OmniChange libraries. In the first section, the ISM approach is presented, in which four single Site-Saturation Mutagenesis CelA2-ISM libraries are screened and analyzed using conventional MTP-based activity assays (4.4). In the second section, the analysis of a CelA2-multiple SSM library, generated by performance of subsequent single saturations of four positions, using the uHTS-IVC-based flow cytometry technology platform is shown. The resulting libraries are analyzed using MTP-based activity assays before and after each round of SSM, analysis and sorting with the flow cytometer. In contrast to the ISM approach, here the entire sorted active fraction serves as a template for the next round SSM (4.5) and not only the best identified variant. The third section presents the screening and analysis of a CelA2-OmniChange library, generated by performance of simultaneous saturation of four positions, using the uHTS-IVC-based flow cytometry technology platform. The CelA2-OmniChange library is analyzed using MTP-based activity assays before and after sorting with the flow cytometer (4.6). The best identified variants of each approach of

diversity generation and screening are analyzed by sequencing and resulting beneficial amino acid substitutions are compared regarding their activity towards the substrate 4-MUC. Finally, the best cellulase variant is purified and characterized in MTP format, and its kinetic parameters were reported (4.6.6-4.6.7).

4.4. CelA2-ISM library

4.4.1. CelA2-ISM library generation, screening and analysis with 4-MUC and FDC activity assay in MTP

The standard ISM strategy was applied in this focused mutagenesis approach for generation of the CelA2-ISM libraries in order to reveal the structure-function relationship between these residues and the 4-MUC or FDC substrate and to identify possible combinatorial effects. Therefore, the SSM libraries were generated subsequently by targeting the amino acids F288, Y368, H524, and Y576 in close proximity to the active center of CelA2-H288F which were selected by previous computational analysis (see chapter 9.2.2). The selected positions were saturated subsequently using "Quick-Change PCR" with degenerated NNK-primers for saturation of each position with all 20 amino acids (see Material and Methods 4.8.3, Table 20) as illustrated in Figure 56. In the ISM approach, the best identified variant from the first/previous round of screening serves as template for the next round of SSM (Figure 56).
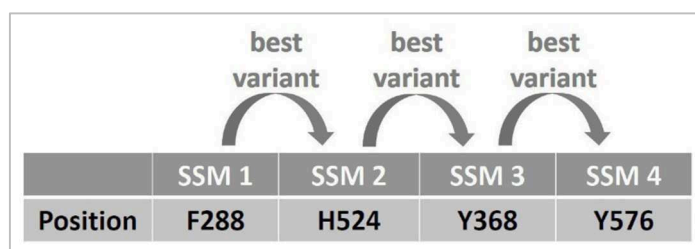
Figure 56: Focused mutagenesis strategy for saturation of four positions (F288, H524, Y368 and Y576) in the *celA2* gene using subsequent SSM in the respective order for generation of ISM libraries.

The resulting PCR products were analyzed by agarose-gelelectrophoresis, purified with NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG), *Dpn*I digested and transformed into chemically competent *E.coli* BL21 Gold(DE3)lacI[Q1] cells for expression and activity tests. Activity determination of 176 randomly picked clones was performed using 4-MUC (Lehmann *et al.*, 2012) and FDC activity assay (see Material and Methods 4.8.15) in parallel for screening of the obtained libraries. As only a diversity of 96 different cellulase

variants can be generated by single SSM and degenerated NNK codons (Tab. 10), analysis of 176 variants (2 MTPs) ensures screening of all possible generated amino acids by achieving full coverage of the generated diversity (with a >95 % confidence) due to oversampling (Martinez *et al.*, 2013; Nov, 2012; Reetz *et al.*, 2008). The best identified variant from the previous round of saturation mutagenesis was always used as template for the next round of SSM as depicted in the scheme (Figure 56, 57). At the end, four rounds of subsequent SSM enable the investigation of beneficial amino acid substitutions at the targeted positions and reveal possible combinatorial effects (Figure 57).
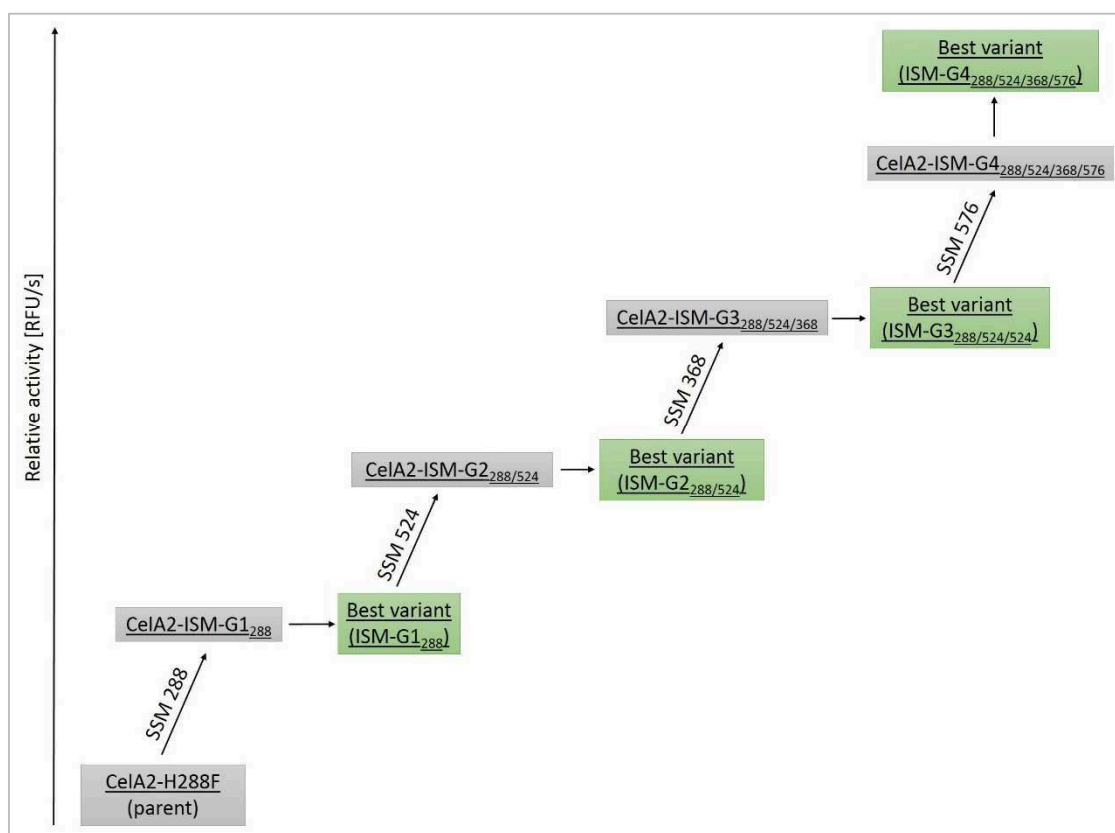
Figure 57: Scheme for ISM at four positions (F288, H524, Y368 and Y576) starting with the parent pIX3.0RMT7-CelA2-H288F using subsequent SSM for generation of CelA2-ISM libraries in iterative rounds of SSM and screening for cellulase variants showing improved activity compared to the parent.

4.4.1.1. CelA2-ISM-G1$_{288}$ library

The CelA2-ISM library was generated using pIX3.0RMT7-CelA2-H288F as template in combination with specific NNK-primers (see Material and Methods 4.8.3, Table 20) for saturation of position F288 in the first step. Activity determination of 176 randomly picked clones revealed 43.75 % active clones for 4-MUC and 37.5 % active clones for FDC activity assay (Figures 58-61). Clones were determined as "active" when the increase in fluorescence was at least 0.2 RFU/s for 4-MUC which is 10-fold the average activity of the negative control

(average activity CelA2-H288F-E580Q for 4-MUC: 0.02 RFU/s) and at least 1.1 RFU/s for FDC which is 10-fold the average activity of the negative control (average activity CelA2-H288F-E580Q for FDC: 0.11 RFU/s). The standard deviation obtained for 4-MUC activity assay was <17 % and for FDC activity assay was ~20 %. Screening with both assays resulted in general in the same activity pattern for CelA2-ISM-G1$_{288}$ library, *i.e.* variants exhibiting activity towards 4-MUC also show activity towards FDC substrate (Figures 58, 60). Whereas the relative activity (slope in RFU/s) of the active variants towards FDC is slightly higher than towards 4-MUC substrate (Figures 58, 60).

For 24 randomly selected clones sequence analysis was performed for evaluation of the quality of the CelA2-ISM-G1$_{288}$ library (Table 11). The sequencing analysis of only 24 clones revealed an exchange of F288 to T, R, S, D, V, I, F, M, N, H, W, L, and K, and already 65 % coverage of all possible amino acid exchanges (Table 11). Variant F1 (MTP No. 2), showing the highest activity towards the substrates 4-MUC (12.68 RFU/s) and FDC (19.36 RFU/s), was sequence analyzed and taken as template for the second round of SSM. Sequencing reveled that variant F1 (Figures 59, 61) contains the amino acid phenylalanine at position 288, which is in consistence with the previously reported data on the beneficial substitution H288F which increased the activity of CelA2 variant towards 4-MUC (Lehmann *et al.*, 2012). Additionally the sequences of variants A4, D12, F10, H10 (MTP No. 1) and D11, E4, E5, E11, H6, H8 (MTP No. 2) were analyzed and reveled all the amino acid phenylalanine at position 288.
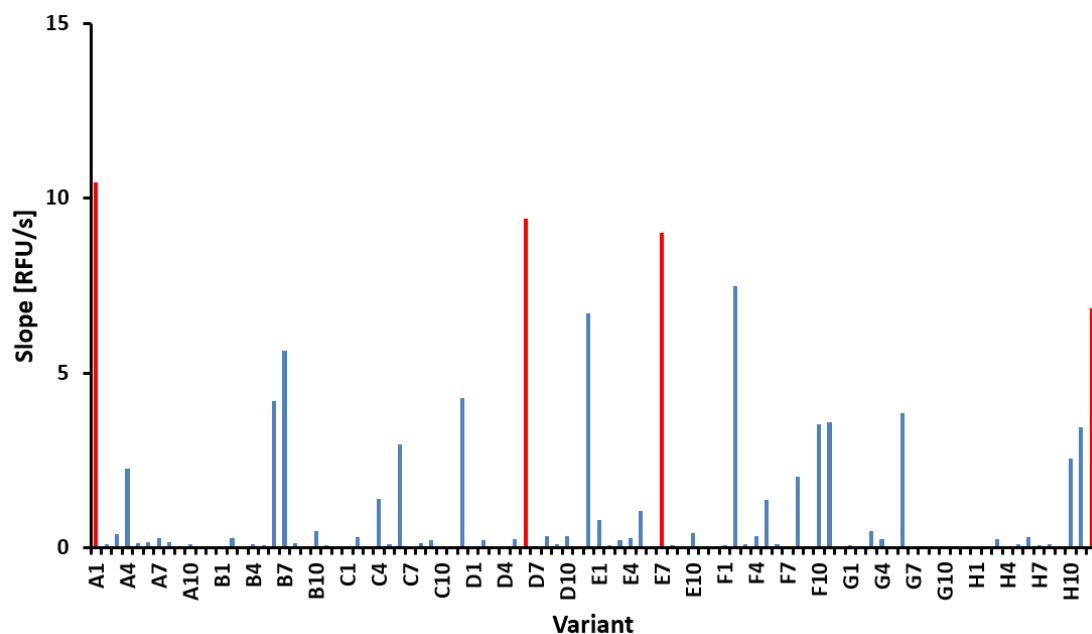
Figure 58: Analysis of CelA2-ISM-G1$_{288}$ library (MTP No. 1) with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s).
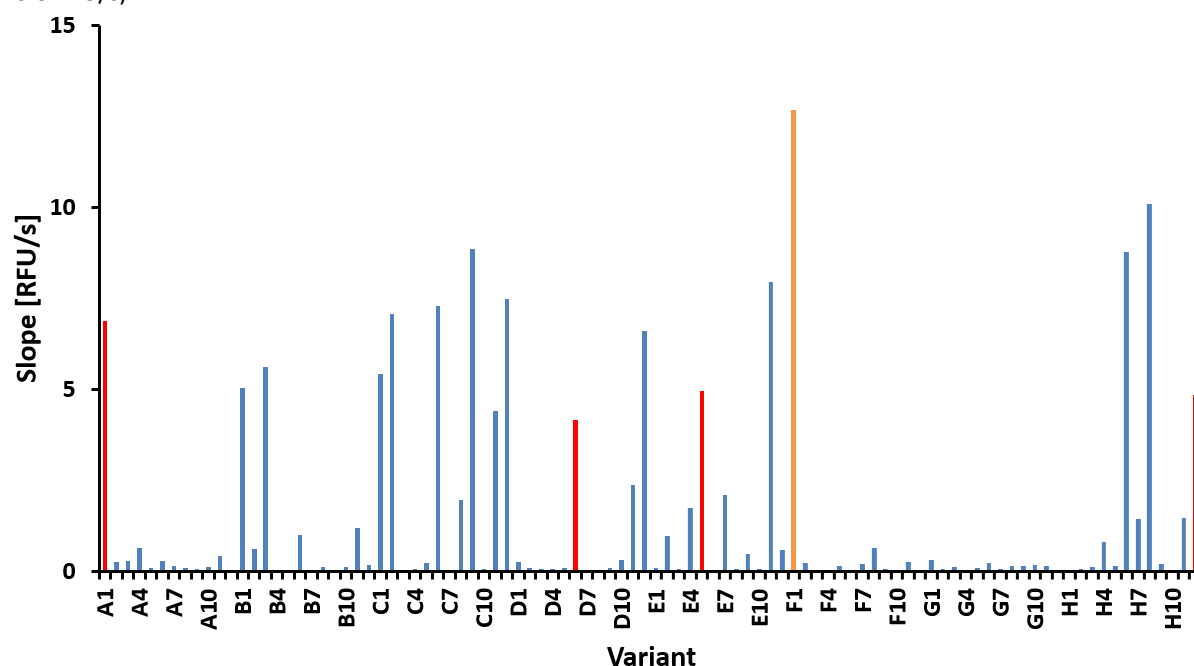
Figure 59: Analysis of CelA2-ISM-G1$_{288}$ library (MTP No. 2) with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s). Variant F1, showing the highest activity, is highlighted in orange.
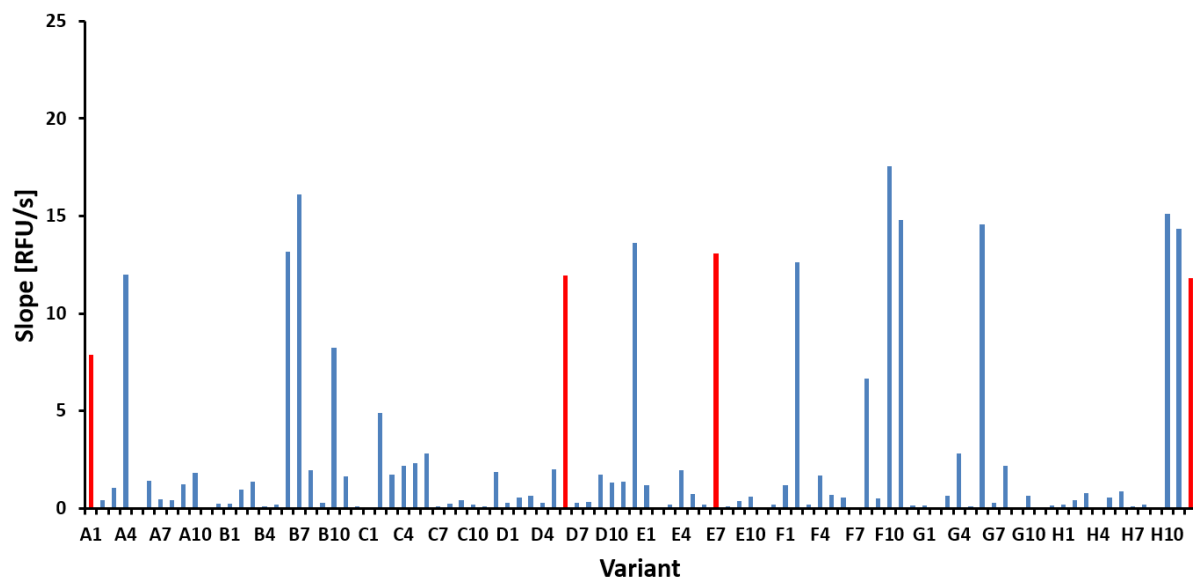
Figure 60: Analysis of CelA2-ISM-G1$_{288}$ library (MTP No. 1) with FDC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate FDC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s).
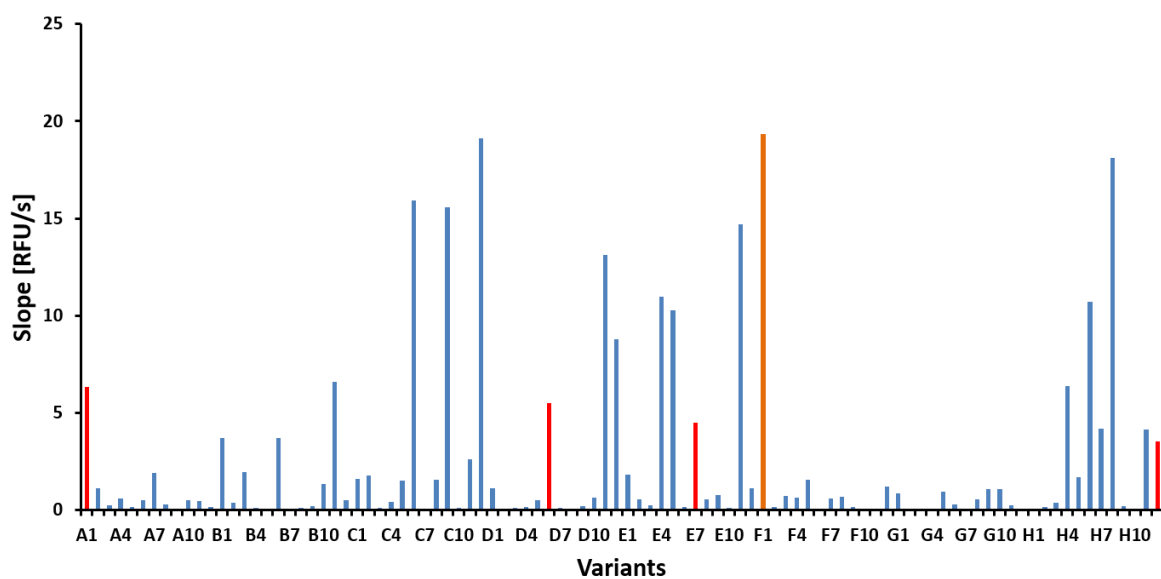
Figure 61: Analysis of CelA2-ISM-G1$_{288}$ library (MTP No. 2) with FDC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate FDC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s). Variant F1, showing the highest activity, is highlighted in orange.

Table 11: Sequence analysis of 24 clones of CelA2-ISM-G1$_{288}$ library.

Clone	Amino acid exchanges	Activity for 4-MUC in slope [RFU/s]	Activity for FDC in slope [RFU/s]
ISM 1.1 P1B1	F288T	0.01	0.27
ISM 1.1 P1B2	F288R	0.26	0.99
ISM 1.1 P1B3	F288S	-0.07	1.40
	N579D		
ISM 1.1 P1B4	F288V	0.08	0.10
ISM 1.1 P1B5	F288I	0.08	0.20
ISM 1.1 P1B6	F288F	4.20	13.18
ISM 1.1 P1B7	F288F	5.63	16.12
ISM 1.1 P1B8	F288S	0.13	1.98
ISM 1.1 P1B9	F288V	0.02	0.30
ISM 1.1 P1B10	F288M	0.46	8.23
ISM 1.1 P1B11	F288N	0.07	1.66
ISM 1.1 P1B12	F288T	-0.04	0.09
	T284I		
ISM 1.1 P1C1	F288Stop	0.04	0.00
ISM 1.1 P1C2	F288H	0.29	4.91
	A284S		
ISM 1.1 P1C3	F288S	0.02	1.74
ISM 1.1 P1C4	F288W	1.41	2.21
ISM 1.1 P1C5	F288N	0.10	2.32
ISM 1.1 P1C6	F288W	2.94	2.83
ISM 1.1 P1C7	F288V	-0.05	0.11
ISM 1.1 P1C8	F288I	0.12	0.25
ISM 1.1 P1C9	F288L	0.21	0.45
ISM 1.1 P1C10	F288I	0.00	0.19
ISM 1.1 P1C11	F288K	-0.05	0.13
ISM 1.1 P1C12	F288W	4.29	1.88
CelA2-H288F	-	**8.94**	**11.19**
CelA2-H288F-E580Q	-	**0.01**	**0.24**

4.4.1.2. CelA2-ISM-G2$_{288/524}$ library

The CelA2-ISM library was generated using the most active variant from CelA2-ISM-G1$_{288}$ (pIX3.0RMT7-CelA2-H288F) as template and specific NNK-primers (see Material and Methods 4.8.3, Table 20) for saturation of position 524 in the second round of SSM. Activity determination of 176 randomly picked clones revealed 69.32 % active clones for 4-MUC, illustrated as an example in Figure 62, and 43.18 % active clones for FDC activity assay. Clones were determined as "active" when the increase in fluorescence was at least 0.2 RFU/s for 4-MUC which is 10-fold the average activity of the negative control (average activity CelA2-H288F-E580Q for 4-MUC: 0.02 RFU/s) and at least 1.1 RFU/s for FDC which is 10-fold the average activity of the negative control (average activity CelA2-H288F-E580Q for FDC: 0.11 RFU/s). The standard deviations obtained for 4-MUC activity assay are <10 % and

Figure 3: *In vitro* compartmentalization (IVC) of a gene library in (w/o/w) emulsions and sorting with flow cytometer. (1.) A gene library is encapsulated in (w/o) emulsions and subsequently (2.) in (w/o/w) emulsions. Genes encoding for active enzyme variants are cell-free transcribed and translated within the *in vitro* expression mixture and produce the desired enzyme. Active enzyme variants convert the substrate inside of the emulsion droplet into the fluorescent product. (3.) The fluorescent microdroplet with the gene encoding for the active enzyme variant is detected and sorted with the flow cytometer. (4.) The sorted DNA encoding for active enzyme variants are recovered from the emulsions for the next round of diversity generation and sorting.

The generation of (w/o/w) emulsions can be performed by different techniques consisting of two emulsification steps. An elaborate and complex technique is the generation of (w/o/w) emulsions using microfluidic devices (Okushima *et al.*, 2004; Shah *et al.*, 2008), (Figure 4). More simple and rapid methods for generation of (w/o/w) emulsions are emulsification by ultra turrax (Hai *et al.*, 2004), by stirring, or by membrane extruder (Miller *et al.*, 2006), (Figure 4).

Different detergent and oil systems have been tested while the most promising ones for compartmentalization in (w/o/w) emulsions seem to be: Tween 80, Span 80 in light mineral oil (LMO)/Tween 20 in PBS (Hai *et al.*, 2004); ABIL EM 90 in LMO/CMC, triton-X102 in PBS (Aharoni *et al.*, 2005a) and decane, cholesterol and Span 60/Tween 20 in PBS (Miller *et al.*, 2006).

Figure 62: Analysis of CelA2-ISM-G2$_{288/524}$ library (MTP No. 1) with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s). Variant C12, showing the highest activity, is highlighted in orange.

4.4.1.3. CelA2-ISM-G3$_{288/524/368}$ library

The CelA2-ISM library was generated using pIX3.0RMT7-CelA2-H288F-H524Q as template, showing the highest activity towards both substrates after the second round of mutagenesis CelA2-ISM-G2$_{288/524,}$ in combination with degenerated NNK-primers (see Material and Methods 4.8.3, Table 20) for saturation of position 368 in the third round of SSM. Activity determination of 176 randomly picked clones revealed 30.11 % active clones for 4-MUC (Figure 63) and 25.57 % active clones for FDC activity assay. Clones were determined as "active" when the increase in fluorescence was at least 0.2 RFU/s for 4-MUC which is 10-fold the average activity of the negative control (average activity CelA2-H288F-E580Q for 4-MUC: 0.02 RFU/s) and at least 1.1 RFU/s for FDC which is 10-fold the average activity of the negative control (average activity CelA2-H288F-E580Q for FDC: 0.11 RFU/s). The standard deviations obtained for 4-MUC activity assay and for FDC activity assay are <15 %.

For 24 randomly selected clones sequence analysis was performed to evaluate the quality of the CelA2-ISM-G3$_{288/524/368}$ library (Table 13). The sequencing analysis of 24 clones revealed a 45 % coverage of all possible amino acid exchanges, showing an exchange of Y368 to L, N, F, W, A, S, T, V and K (Table 13). Among the analyzed clones no variant showing higher activity than CelA2-H288F/H524Q towards 4-MUC or FDC substrate could be identified (Figure 63).

All analyzed amino acid substitutions different from tyrosine (Y) at position 368 lead to inactive or low active variants, *e.g.* Y368N/L/F/W/S (Table 13), indicating position Y368 plays a key role for maintaining CelA2 activity.

Table 13: Sequence analysis of 24 clones of CelA2-ISM-G3$_{288/524/368}$ library.

| Clone	Amino acid exchanges	Activity for 4-MUC in slope [RFU/s]	Activity for FDC in slope [RFU/s]
ISM 1.3 P1A2	Y368Y	19.77	4.37
ISM 1.3 P1A3	Y368L	0.08	-0.01
ISM 1.3 P1A4	Y368N	0.04	-0.01
ISM 1.3 P1A5	Y368F	0.07	-0.02
ISM 1.3 P1A6	Y368L	0.08	-0.01
ISM 1.3 P1A7	Y368W	0.07	-0.02
ISM 1.3 P1A8	Y368A	0.07	-0.02
ISM 1.3 P1A9	Y368S	0.06	-0.02
ISM 1.3 P1A10	Y368F	0.05	-0.02
ISM 1.3 P1A11	Y368T	0.05	-0.01
ISM 1.3 P1B1	Y368L	0.05	0.00
ISM 1.3 P1B2	Y368Y	21.50	5.04
ISM 1.3 P1B3	Y368L	0.00	0.00
ISM 1.3 P1B4	Y368S	0.02	-0.02
ISM 1.3 P1B5	Y368V	0.05	0.00
ISM 1.3 P1B6	Y368Y	0.05	-0.01
ISM 1.3 P1G4	Y368T	0.00	0.02
ISM 1.3 P1G5	Y368S	0.02	0.00
ISM 1.3 P1G6	Y368S	0.01	-0.01
ISM 1.3 P1G10	Y368K	-0.04	-0.02
ISM 1.3 P1G11	Y368L	-0.01	0.00
ISM 1.3 P2C2	Y368L	0.03	-0.01
ISM 1.3 P2C8	Y368Y	7.79	1.16
ISM 1.3 P2C10	Y368N	0.03	-0.01
CelA2-H288F/H524Q	**-**	**16.76**	**2.21**
CelA2-H288F-E580Q	**-**	**0.00**	**0.05**
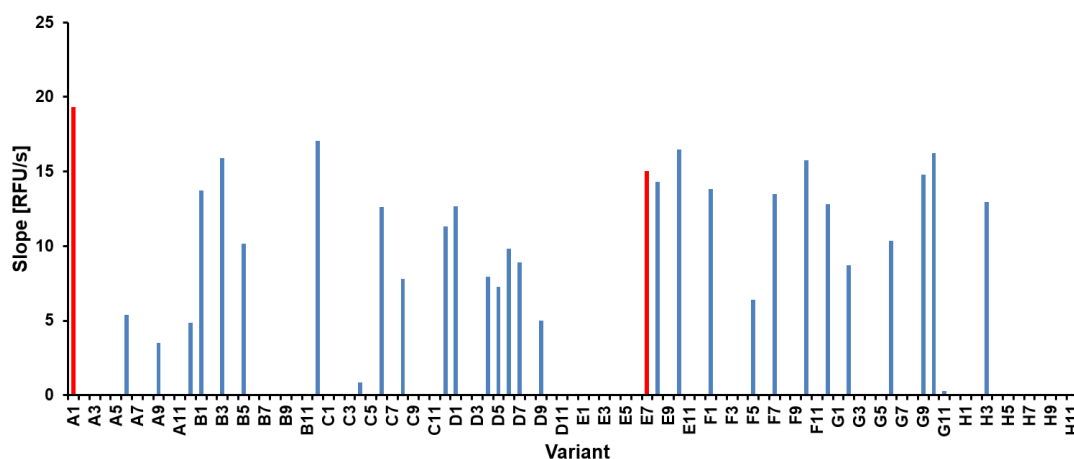
Figure 63: Analysis of CelA2-ISM-G3$_{288/524/368}$ library (MTP No. 2) with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F/H524Q, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s).

4.4.1.4. *CelA2-ISM-G4$_{288/524/368/576}$ library*

The CelA2-ISM library was generated using pIX3.0RMT7-CelA2-H288F-H524Q as template, showing the highest activity towards both substrates after the second and third round of mutagenesis CelA2-ISM-G3$_{288/524/368}$, in combination with degenerated NNK-primers (see Material and Methods 4.8.3, Table 20) for saturation of position 576 in the fourth round of SSM. Activity determination of 176 randomly picked clones revealed 30.68 % active clones for 4-MUC (Figure 64) and 21.59 % active clones for FDC activity assay. Clones were determined as "active" when the increase in fluorescence was at least 0.2 RFU/s for 4-MUC which is 10-fold the average activity of the negative control (average activity CelA2-H288F-E580Q for 4-MUC: 0.02 RFU/s) and at least 1.1 RFU/s for FDC which is 10-fold the average activity of the negative control (average activity CelA2-H288F-E580Q for FDC: 0.11 RFU/s). The standard deviations obtained for 4-MUC activity assay and for FDC activity assay are <20 %.

For 24 randomly selected clones sequence analysis was performed to evaluate the quality of the CelA2-ISM-G4$_{288/524/368/576}$ library (Table 14). The sequencing analysis of 24 clones revealed 50 % coverage of all possible amino acid exchanges, showing an exchange of Y576 to S, E, L, Q, T, D, C, M, I and F (Table 14). Among the analyzed clones no variant with higher activity towards MUC or FDC could be identified from SSM library Y576. Still, the most active variant was CelA2-H288F/H524Q from the CelA2-ISM-G2$_{288/524}$ library. As indicated on the example in Figure 64, the expression of the parent, used as reference in the MTPs, was in

general very low (4.78 RFU/s for 4-MUC; 0.67 RFU/s for FDC). Some variants which were showing higher activity than the parent CelA2-H288F/H524Q/Y368Y/Y576Y had either additional mutations, *e.g.* variant P2C8 (F288/Y368/L495M/Q524/Y576), showing increased activity towards 4-MUC (22.07 RFU/s) and FDC (4.48 RFU/s), containing the additional mutation L495M (Figure 64, Table 14), or they seem to be expression mutants, *e.g.* variant P1A5 (Table 14) or P2G10 (Figure 64, Table 14), as they contain the same amino acid sequence as the parent. Many amino acid substitutions at position 576 lead to inactive or low active variants, *e.g.* Y576S/L/Q/T/D (Table 14), indicating a key role of position 576 for cellulolytic activity.

Table 14: Sequence analysis of 24 clones of CelA2-ISM-G4$_{288/524/368/576}$ library.

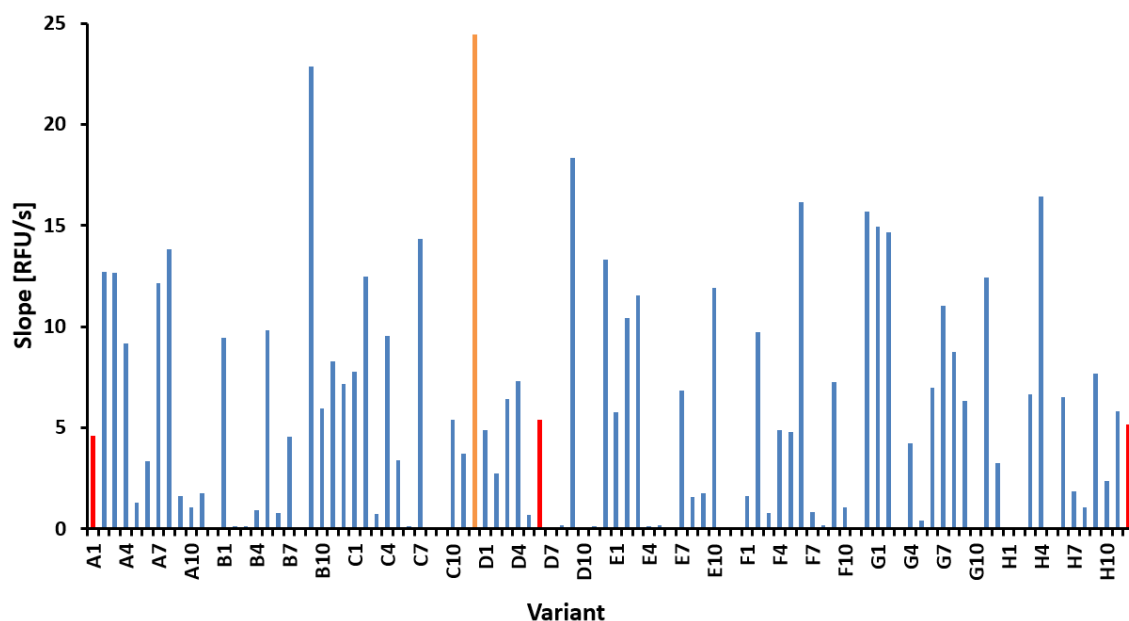| Clone	Amino acid exchanges	Activity for 4-MUC in slope [RFU/s]	Activity for FDC in slope [RFU/s]
ISM 1.4 P1A5	Y576Y	30.60	3.68
ISM 1.4 P1A7	Y576Y	26.32	3.12
ISM 1.4 P1B4	Y576S	0.26	-0.01
ISM 1.4 P1B6	Y576E	0.08	0.01
ISM 1.4 P1B7	Y576L	0.12	0.01
ISM 1.4 P1B9	Y576Y	25.06	3.06
ISM 1.4 P1B10	Y576Q	0.21	-0.03
ISM 1.4 P1B11	Y576Y	20.05	2.08
ISM 1.4 P1B12	Y576S	0.10	-0.02
ISM 1.4 P1C3	Y576Y	29.42	3.33
ISM 1.4 P1D8	Y576Y	21.63	2.45
ISM 1.4 P1D11	Y576Y	24.55	2.75
ISM 1.4 P1D12	Y576Y	22.44	3.04
ISM 1.4 P1F3	Y576S	0.00	-0.03
ISM 1.4 P1F11	Y576T	0.06	-0.02
ISM 1.4 P2A6	Y576Y	9.20	1.41
ISM 1.4 P2B3	Y576D	0.18	-0.01
ISM 1.4 P2B9	Y576C	0.09	0.01
ISM 1.4 P2C5	Y576L	10.18	2.11
	I538M		
ISM 1.4 P2C8	Y576Y	22.07	4.48
	L495M		
ISM 1.4 P2F7	Y576I	0.04	-0.03
ISM 1.4 P2F11	Y576F	5.28	0.59
ISM 1.4 P2G10	Y576Y	23.76	4.37
CelA2-H288F/H524Q	-	**4.78**	**0.67**
CelA2-H288F-E580Q	-	**-0.04**	**0.02**

Figure 64: Analysis of CelA2-ISM-G4$_{288/524/368/576}$ library (MTP No. 2) with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F/H524Q, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s).

4.5. CelA2-multiple SSM library

4.5.1. CelA2-multiple SSM library generation and screening strategy

In addition to ISM as focused mutagenesis strategy, multiple Site-Saturation Mutagenesis (multiple SSM) libraries can be analyzed through employment of flow cytometer-based ultrahigh throughput screening systems offering a throughput of up to 10^7 events per hour. In contrast to the ISM approach, in the CelA2-multiple SSM approach the library was analyzed by flow cytometer, active variants were enriched by sorting, and the entire sorted, active fraction served as template for the subsequent round of SSM and not only the best identified variant. In this approach, highly diverse CelA2-multiple SSM libraries were generated by targeting the amino acids F288, Y368, H524, and Y576 in close proximity to the active center of CelA2-H288F, which were selected by previous computational analysis (see 4.2.2). The selected positions were saturated using "Quick-Change PCR" with degenerated NNK-primers for saturation of each position with all 20 amino acids (see Material and Methods 4.8.3, Table 20). The resulting PCR products were analyzed by agarose-gelelectrophoresis, purified with NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG), *Dpn*I digested and transformed into chemically competent *E.coli* BL21 Gold(DE3)lacI^{Q1} cells for expression and activity tests in MTP-based format. For analysis with the uHTS-IVC platform, transformation was performed in

commercial, ultracompetent *E.coli* XL10-Gold cells (Aglient technologies, Santa Clara, CA, USA) with a transformation efficiency of $\geq$5 x 10^9 transformants/µg pUC18 control plasmid DNA. Activity determination of 352 randomly picked clones using 4-MUC activity assay (Lehmann *et al.*, 2012) was performed to determine the percentages of active variants before and after sorting. The generated CelA2-multiple SSM plasmid libraries were subsequently analyzed with FDC substrate using the developed uHTS-IVC platform in iterative cycles for enrichment of beneficial cellulase variants. The beneficial variants from the first round of saturation mutagenesis were enriched, by sorting of active variants, pooled and used as templates for the next round of SSM as depicted in the scheme (Figure 65). The advantage of the flow cytometer-based screening of multiple SSM libraries is the possibility to use many different beneficial enzyme variants/mutants as templates for the next round of mutagenesis and not only the best identified one, like in standard ISM approaches. For this reason the multiple SSM method is a promising strategy to enhance the chances to identify combinatorial effects.
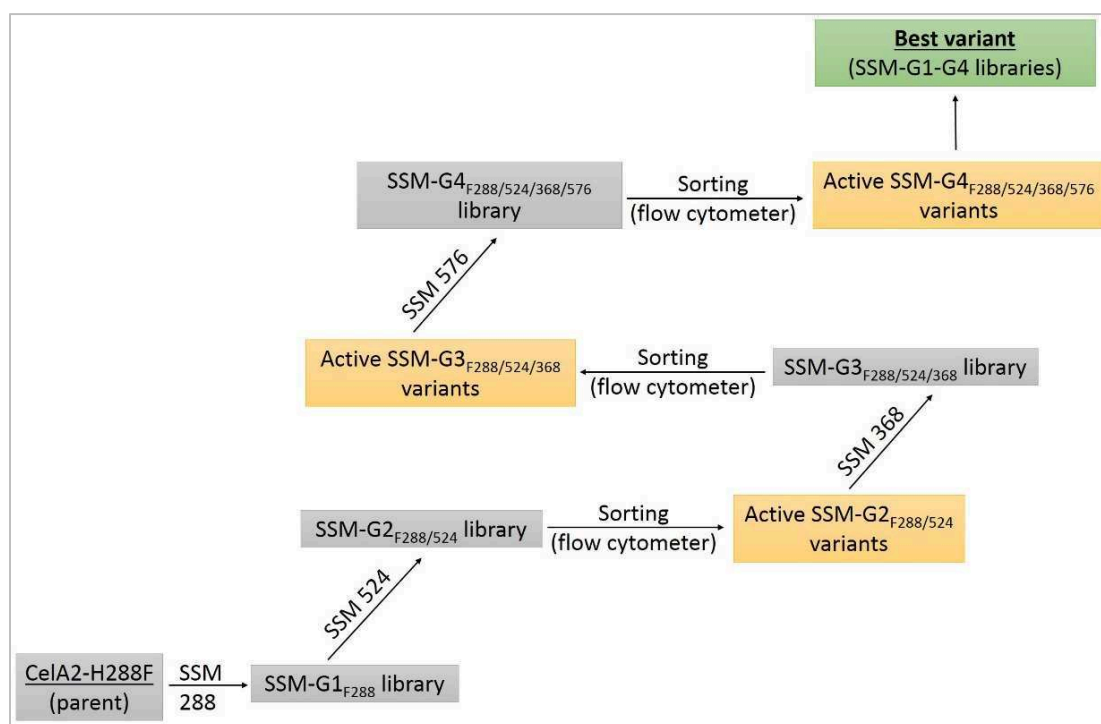
Figure 65: Scheme illustrating the saturation of four positions (F288, H524, Y368 and Y576), starting with the parent pIX3.0RMT7-CelA2-H288F, using subsequent SSM for generation of CelA2-multiple SSM libraries in iterative rounds of SSM and screening for cellulase variants which show improved activity compared to the parent, *e.g.* variant M3 (CelA2-H288F/H524Q).

Figure 4: Preparation of (w/o) and (w/o/w) emulsion compartments, entrapping gene or cell mutants, by different techniques (*i.e.* stirring, extrusion, homogenizing or microfluidics). The (w/o/w) emulsions are prepared in two steps, first a high-shear emulsification step with lipophilic surfactants for generation of the (w/o) emulsion, and second a low shear emulsification step with hydrophilic surfactants for generation of the (w/o/w) emulsion. The (w/o) and (w/o/w) emulsion pictures were taken by transmission microcopy and the scale bars represent a length of 20 µm.

1.5. Cell-free protein expression

For the *in vitro* expression of proteins it is necessary to perform transcription and translation of a DNA molecule into mRNA molecules and the subsequent translation of the mRNA molecules into polypeptides in a cell-free environment, e.g. within a laboratory mixture containing ribosomes, polymerases, chaperones etc. (Figure 5). The first successful attempts to translate an exogenous mRNA molecule in a cell-free environment were performed by Nirenberg and Matthaei in 1961 (Nirenberg *et al.*, 1961). Six years later, Zubay and his team managed to couple transcription and translation and developed in this way an *in vitro* system for translation of DNA molecules (DeVries *et al.*, 1967). Nowadays *in vitro* translation is an important technique to analyze the process of mRNA translation into functional proteins, especially in the field of functional and structural proteomics (Kigawa *et al.*, 1999).

4.5.3. CelA2-multiple SSM library generation, flow cytometer-based analysis and sorting of CelA2-multiple SSM library libraries in emulsions and analysis of libraries with 4-MUC activity assay in MTP before and after sorting

4.5.3.1. CelA2-multiple SSM-G1$_{288}$ library

The CelA2-multiple SSM-G1$_{288}$ library was generated using pIX3.0RMT7-CelA2-H288F as template in combination with degenerated NNK-primers (see Material and Methods 4.8.3, Table 20) for saturation of position 288 in the first step. The resulting CelA2-multiple SSM-G1$_{288}$ library was transformed in competent *E.coli* BL21 Gold(DE3)lacI^{Q1} for expression and activity tests. Activity determination of 176 randomly picked clones using 4-MUC activity assay (Lehmann *et al.*, 2012) revealed 34 % active clones (Figure 67; Figure A5, appendix). Clones were determined as "active" when the increase in fluorescence was at least 0.2 RFU/s which is 13-fold the average activity of the negative control (average activity CelA2-H288F-E580Q: 0.0077 RFU/s). Colonies were pooled from the CelA2-multiple SSM-G1$_{288}$ agar plates and plasmids were isolated with Nucleo Spin® Plasmid Kit (Macherey-Nagel GmbH & Co. KG) and served as template for the second round of SSM at position 524.

Figure 67: Analysis of CelA2-multiple SSM-G1$_{288}$ library (MTP No. 1) with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s).

4.5.3.2. CelA2-multiple SSM-G2$_{288/524}$ library

The CelA2-multiple SSM-G2$_{288/524}$ library was generated using CelA2-multiple SSM-G1$_{288}$ library as template in combination with degenerated NNK-primers (see Material and Methods 4.8.3, Table 20) for saturation of position H524. The resulting CelA2-multiple SSM-G2$_{288/524}$ library was transformed in competent *E.coli* BL21 Gold(DE3)lacI^{Q1} for expression and activity tests. Activity determination of 352 randomly picked clones using 4-MUC activity assay (Lehmann *et al.*, 2012) revealed 21.87 % active clones (Figure 68; Figure A6, appendix). Clones were determined as "active" when the increase in fluorescence was at least 0.2 RFU/s which is 13-fold the average activity of the negative control (average activity CelA2-H288F-E580Q: 0.0077 RFU/s).
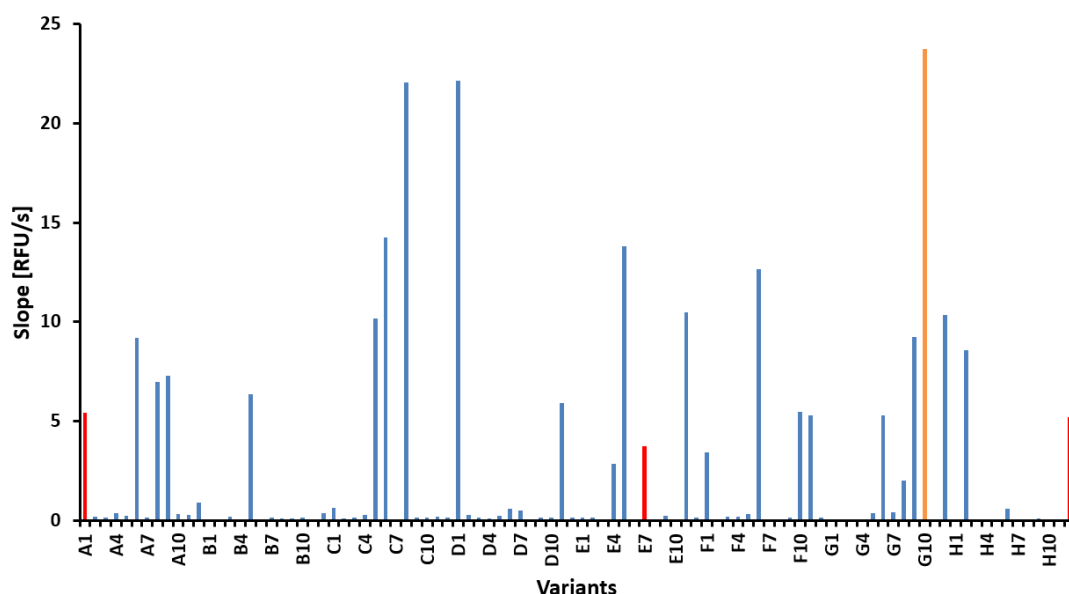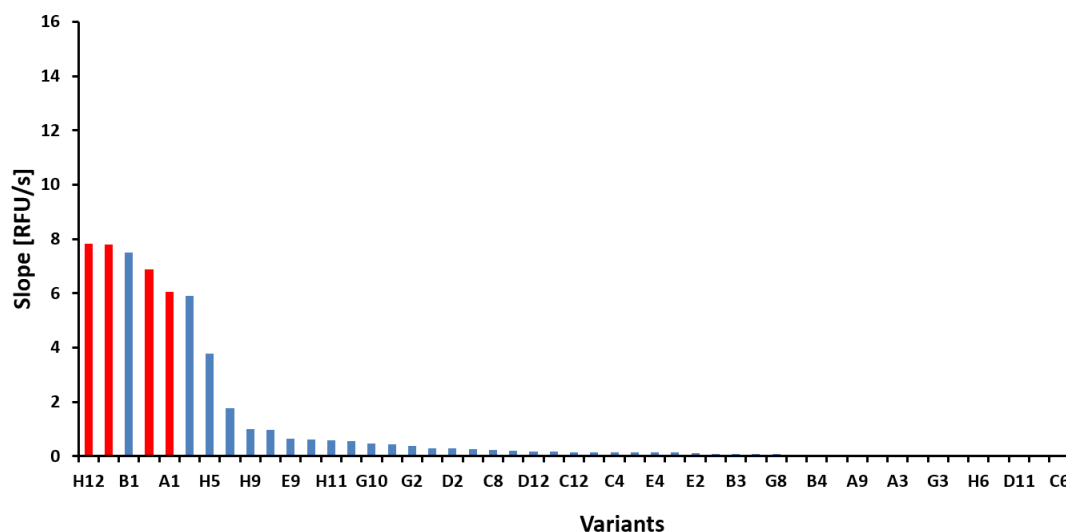

Figure 68: Analysis of CelA2-multiple SSM-G2$_{288/524}$ library (MTP No. 4) with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s).

The CelA2-multiple SSM-G2$_{288/524}$ library was transformed in competent *E.coli* XL10 Gold cells (2.63 x 10^5 CFU), plasmids were isolated with Nucleo Spin® Plasmid Kit (Macherey-Nagel GmbH & Co. KG) and the resulting plasmid library was cell-free expressed within (w/o/w) emulsions employing optimized conditions (0.656 µM DNA, 0.46 mM FDC, 1 % BSA, 30°C, 4 h) and subsequently analyzed on flow cytometer (Figure 69). The in chapter 9.5.2 optimized amount of template DNA of 0.164 µM, was not applicable to be used for screening of the generated SSM-G2$_{288/524}$ library, as upon flow cytometer analysis of the CelA2-multiple SSM-G2$_{288/524}$ library the fluorescent signal was comparable to the negative control, due to the low percentage of active variants in this library (Figure 68). Therefore, the

amount of employed template DNA had to be increased to 0.656 µM again in order to obtain a fluorescent signal upon flow cytometry analysis which is sufficient to efficiently sort CelA2-multiple SSM libraries with less than 20 % active variants, showing low activity levels in comparison to the parent pIX3.0RMT7-CelA2-H288F (Figures 68, 72, 73). Upon flow cytometry analysis the active fraction of cell-free expressed CelA2-multiple SSM-G2$_{288/524}$ library in (w/o/w) emulsions was determined as 21.4 % (Figure 69C), correlating with the result in MTP format (21.87 %). The fraction containing 4.8 % of the most fluorescent events was selected for sorting (Figure 70). In only 2.28 hours in total 8,399,643 events were analyzed (1,000 events s^{-1}) of which 739,814 events were sorted with an efficiency of 96 %. Reanalysis of the sorted population revealed an enrichment in active fraction of 9-fold, determined by an increase of the population in the P2 sorting gate from 4.8 % to 45.6 % (Figure 70).
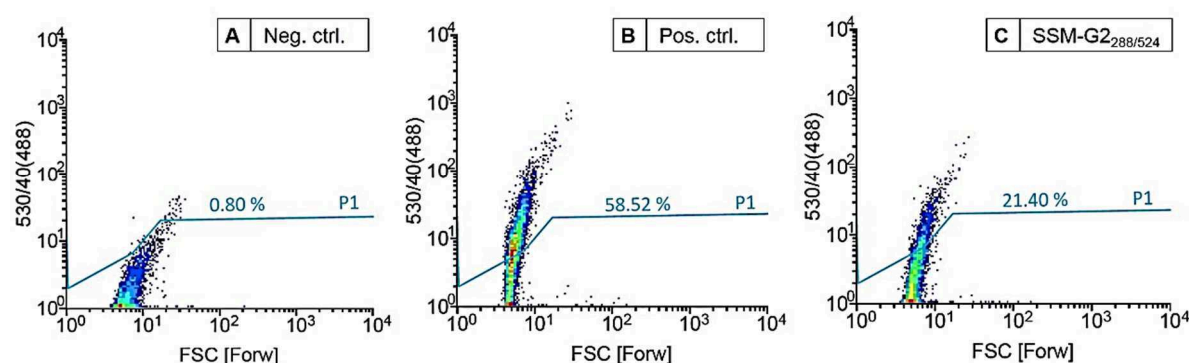
Figure 69: Flow cytometer analysis showing density plots of CelA2-multiple SSM-G2$_{288/524}$ library after *in vitro* expression in (w/o/w) emulsions (4 h, 30 °C). The FSC represents the size of the emulsion on the x-axis and 530/40 (488) represents the fluorescence intensity (λ_{ex} 494 nm and λ_{em} 516 nm) of the fluorescein as reaction product on the y-axis. Lines indicate gate P1 separating fluorescent events (appearing above) from non-fluorescent events. (**A**) Negative control (pIX3.0RMT7-CelA2-H288F-E580Q$_{inactive}$), (**B**) Positive control (pIX3.0RMT7-CelA2-H288F$_{active}$), and (**C**) CelA2-multiple SSM-G2$_{288/524}$ plasmid DNA library showing 21.40 % fluorescent events.
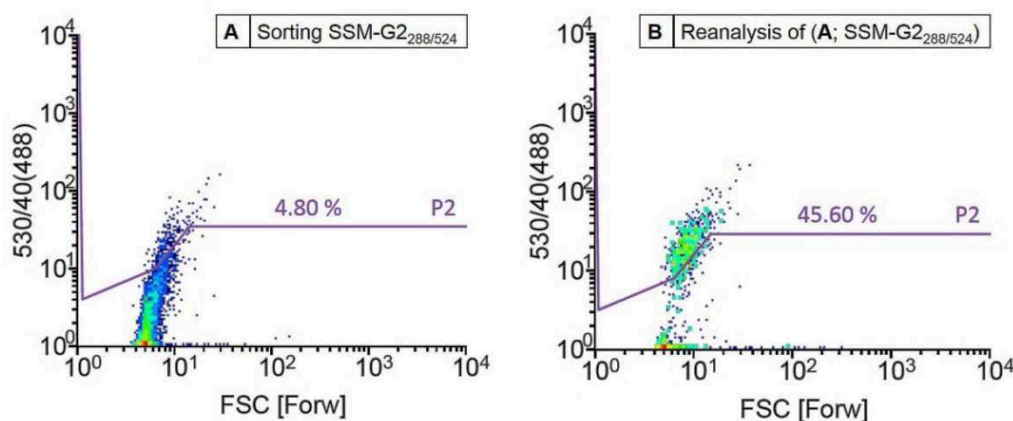
Figure 70: Flow cytometer analysis showing density plots of CelA2-multiple SSM-G2$_{288/524}$ plasmid DNA library after *in vitro* expression in (w/o/w) emulsions (4 h, 30 °C). The FSC represents the size of the emulsion on the x-axis and 530/40 (488) represents the fluorescence intensity (λ_{ex} 494 nm and λ_{em} 516 nm) of the fluorescein as reaction product on the y-axis. Lines indicate sorting gate P2 separating the most fluorescent events (appearing above) from non-/weak-fluorescent events. (**A**) CelA2-multiple SSM-G2$_{288/524}$ plasmid DNA library during sorting of the 4.8 % most fluorescent events. (**B**) Reanalysis of sorted library with a 9-fold enriched active fraction (45.60 % fluorescent events).

The plasmid DNA of the active variants of the CelA2-multiple SSM-G2$_{288/524}$ library was recovered from the sorted (w/o/w) emulsion fraction (see Figure A7, appendix) using NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG) in combination with subsequent PCR amplification according to the optimized protocol (see Material and Methods 4.8.11). The recovered plasmid DNA from the sorted active fraction of SSM-G2$_{288/524}$ library served as template for the third round of SSM at position 368. In addition, recovered plasmid DNA from the sorted active fraction of SSM-G2$_{288/524}$ library were transformed in *E.coli* BL21 Gold(DE3)lacI^{Q1} for determination of enrichment ratios for the active fraction in MTP (Figure 71; Figure A8, appendix). Activity determination of 352 randomly picked sorted clones was performed using 4-MUC activity assay (Lehmann *et al.*, 2012) and revealed 13.64 % active clones (Figure 71; Figure A8, appendix). Clones were determined as "active" when the increase in fluorescence was at least 0.2 RFU/s which is 13-fold the average activity of the negative control (average activity CelA2-H288F-E580Q: 0.0077 RFU/s).

Figure 71: Analysis of the sorted CelA2-multiple SSM-G2$_{288/524}$ library (MTP No. 4) with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s).

4.5.3.3. CelA2-multiple SSM-G3$_{288/524/368}$ library

The CelA2-multiple SSM-G3$_{288/524/368}$ library was generated using variants of the sorted, active fraction of CelA2-multiple SSM-G2$_{288/524}$ library as template in combination with degenerated NNK-primers (see Material and Methods 4.8.3, Table 20) for saturation of position 368. The resulting CelA2-multiple SSM-G3$_{288/524/368}$ library was transformed in competent *E.coli* BL21 Gold(DE3)lacI^{Q1} for expression and activity tests. Activity determination of 352 randomly picked clones using 4-MUC activity assay (Lehmann *et al.*, 2012) revealed 14.20 % active clones (Figure 72; Figure A9, appendix). Clones were determined as "active" when the increase in fluorescence was at least 0.2 RFU/s which is 13-fold the average activity of the negative control (average activity CelA2-H288F-E580Q: 0.0077 RFU/s).
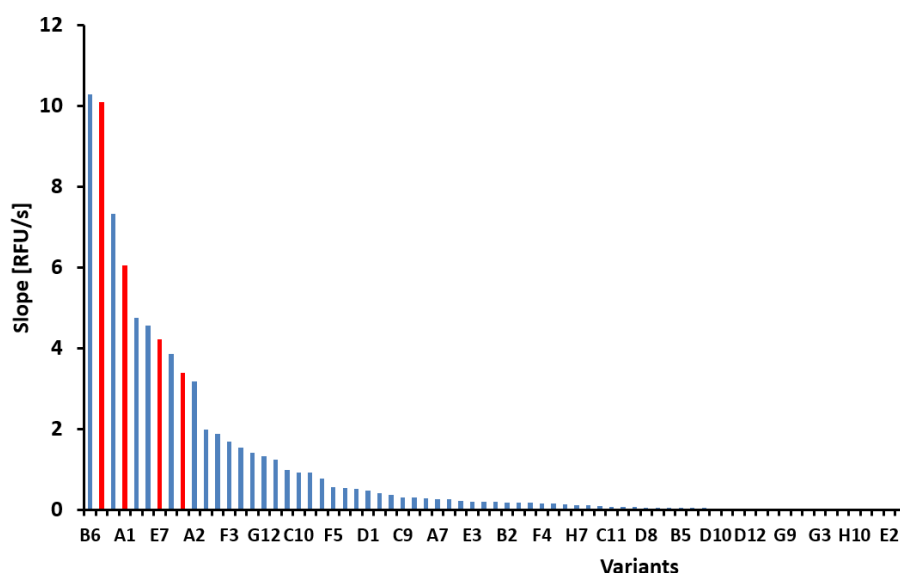
Figure 72: Analysis of CelA2-multiple SSM-G3$_{288/524/368}$ library (MTP No. 3) with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s).

The CelA2-multiple SSM-G3$_{288/524/368}$ library was transformed in competent *E.coli* XL10 Gold cells (4.51 x 10^5 CFU), plasmids were isolated with Nucleo Spin® Plasmid Kit (Macherey-Nagel GmbH & Co. KG) and the resulting plasmid library was cell-free expressed within (w/o/w) emulsions employing optimized conditions (0.656 µM DNA, 0.46 mM FDC, 1 % BSA, 30°C, 4 h) and subsequently analyzed on flow cytometer (Figure 73). Upon flow cytometry analysis the active fraction of cell-free expressed CelA2-multiple SSM-G3$_{288/524/368}$ library in (w/o/w) emulsions was determined as 8.80 % (Figure A10, appendix), correlating with the result in MTP format (14.20 %). The fraction containing 3.4 % of the most fluorescent events was selected for sorting (Figure 73). In around two hours in total 17,787,080 events were analyzed (2,360 events s^{-1}) of which 336,971 events were sorted with an efficiency of 95 %. Reanalysis of the sorted population revealed an enrichment in active fraction of 8.5-fold, determined by an increase of the population in the P2 sorting gate from 3.4 % to 29.7 % (Figure 73).
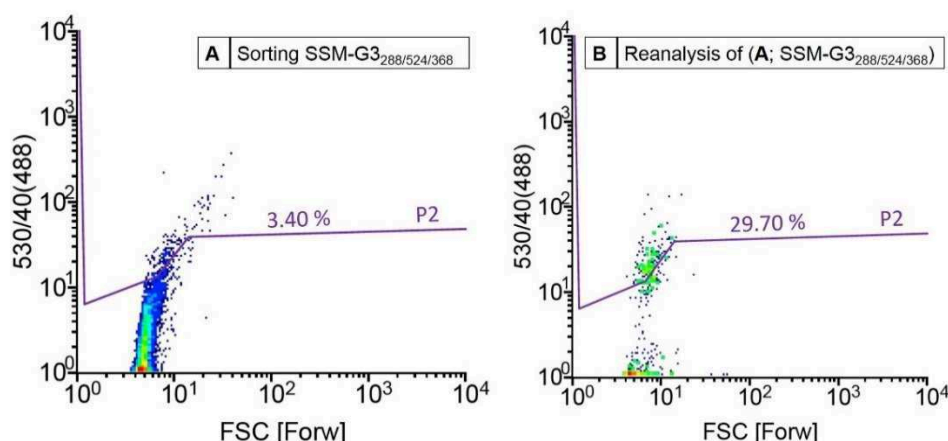
Figure 73: Flow cytometer analysis showing density plots of CelA2-multiple SSM-G3$_{288/524/368}$ plasmid DNA library after *in vitro* expression in (w/o/w) emulsions (4 h, 30 °C). The FSC represents the size of the emulsion on the x-axis and 530/40 (488) represents the fluorescence intensity (λ_{ex} 494 nm and λ_{em} 516 nm) of the fluorescein as reaction product on the y-axis. Lines indicate sorting gate P2 separating the most fluorescent events (appearing above) from non-/weak-fluorescent events. (**A**) CelA2-multiple SSM-G3$_{288/524/368}$ plasmid DNA library during sorting of the 3.40 % most fluorescent events. (**B**) Reanalysis of sorted library with an 8.5-fold enriched active fraction (29.70 % fluorescent events).

The plasmid DNA of the active variants from the CelA2-multiple SSM-G3$_{288/524/368}$ library was recovered from the sorted (w/o/w) emulsion fraction using NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG) in combination with subsequent PCR amplification according to the optimized protocol (see Material and Methods 4.8.11). The recovered plasmid DNA from the sorted active fraction of SSM-G3$_{288/524/368}$ library served as template for the fourth round of SSM at position 576. In addition, recovered plasmid DNA from the sorted active fraction of SSM-G3$_{288/524/368}$ library was transformed in *E.coli* BL21 Gold(DE3)lacI^{Q1} for determination of enrichment ratios for the active fraction in MTP. Activity determination of 352 randomly picked sorted clones was performed using 4-MUC activity assay (Lehmann *et al.*, 2012) and revealed with 34.66 % active clones a 2.4-fold enrichment of the active fraction compared to the unsorted library containing 14.20 % active fraction (Figure 72, 74; Figure A11, appendix). Clones were determined as "active" when the increase in fluorescence was at least 0.2 RFU/s which is 13-fold the average activity of the negative control (average activity CelA2-H288F-E580Q: 0.0077 RFU/s).
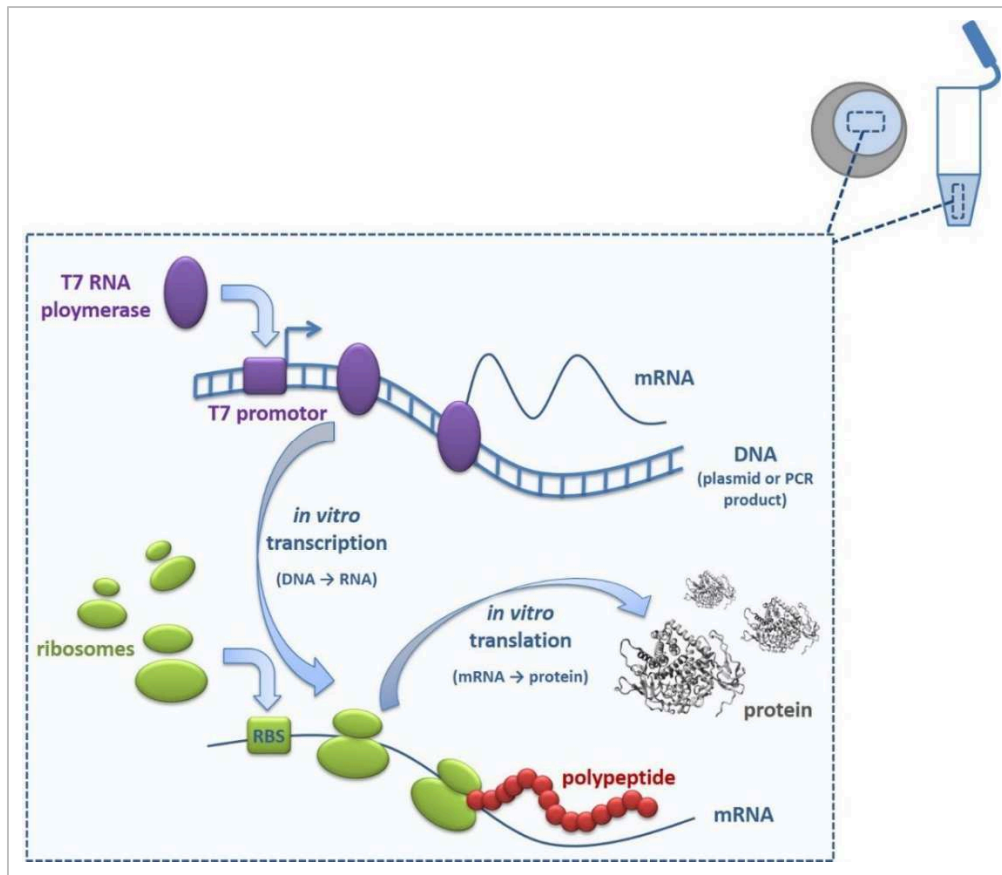
Figure 5: Principle of *in vitro* expression of proteins in a PCR-tube or within emulsion compartments. For the *in vitro* expression of proteins it is necessary to perform transcription and translation of a DNA molecule into mRNA molecules and the subsequent translation of the mRNA molecules into polypeptides in a cell-free environment, *e.g.* within a laboratory *in vitro* expression mixture containing components like ribosomes, polymerases, and chaperones.

Cell-free protein expression has several advantages, as it is possible to express and characterize toxic proteins, membrane proteins or insoluble proteins that are difficult to express *in vivo*. *In vitro* expression technology offers applications in the pharmaceutical field, *e.g.* in the synthesis of toxic compounds or in antibiotic drug discovery (Kigawa *et al.*, 1999). Another advantage of *in vitro* translation is the opportunity to change reaction components and conditions easily making it interesting for protein arrays or *in vitro* evolution experiments (Jackson *et al.*, 2004; Nakano *et al.*, 2004; Spirin, 2004). By addition of helper molecules, e.g. chaperones, or chemicals, the environment of the reaction can be controlled and modified, to enable proper protein folding (Kang *et al.*, 2005). The expressed protein can be modified during translation by including modified tRNAs to incorporate non-natural or chemically modified amino acids into the protein expressed and in this way generate novel molecules, as well as labelling of proteins with fluorescent or biotinylated amino acids (Gerrits M., 2000-2013; Quast *et al.*, 2015b; Singh-Blom *et al.*, 2014). *In vitro* protein
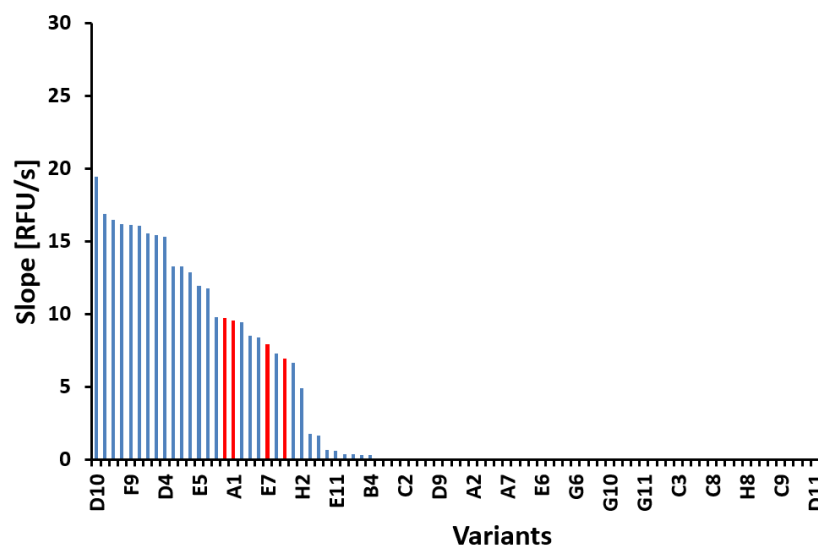
Figure 75: Analysis of CelA2-multiple SSM-G4$_{288/524/368/576}$ library (MTP No. 3) with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s).

The CelA2-multiple SSM-G4$_{288/524/368/576}$ library was transformed in competent *E.coli* XL10 Gold cells (5.25 x 10^5 CFU), plasmids were isolated with Nucleo Spin® Plasmid Kit (Macherey-Nagel GmbH & Co. KG) and the resulting plasmid library was cell-free expressed within (w/o/w) emulsions employing optimized conditions (0.656 µM DNA, 0.46 mM FDC, 1 % BSA, 30°C, 4 h) and subsequently analyzed on flow cytometer (Figure 76). Upon flow cytometry analysis the active fraction of cell-free expressed CelA2-multiple SSM-G4$_{288/524/368/576}$ library in (w/o/w) emulsions was determined as 5.6 % (Figure A13, appendix). The fraction containing 4.4 % of the most fluorescent events was selected for sorting (Figure 76). In only 3.2 hours in total 25,067,704 events were analyzed (2,200 events s^{-1}) of which 1,027,518 events were sorted with an efficiency of 95 %. Reanalysis of the sorted population revealed an enrichment in active fraction of 12-fold, determined by an increase of the population in the P2 sorting gate from 4.4 % to 53.4 % (Figure 76).
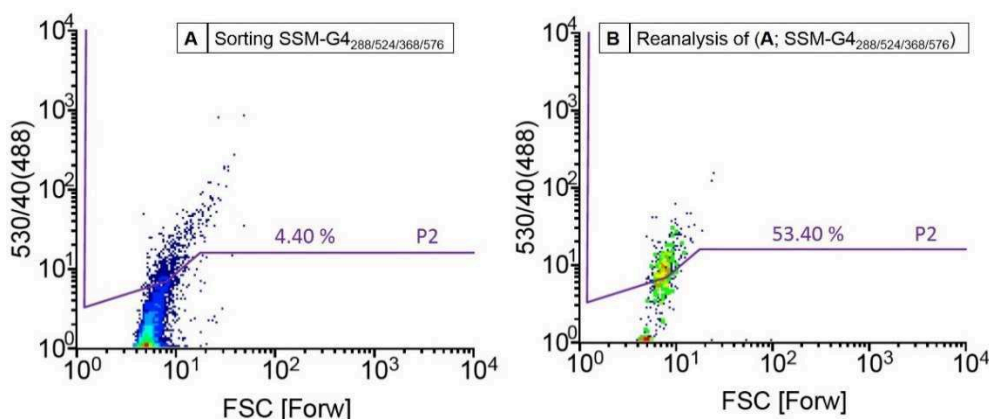
Figure 76: Flow cytometer analysis showing density plots of CelA2-multiple SSM-G4$_{288/524/368/576}$ plasmid DNA library after *in vitro* expression in (w/o/w) emulsions (4 h, 30 °C). The FSC represents the size of the emulsion on the x-axis and 530/40 (488) represents the fluorescence intensity (λ_{ex} 494 nm and λ_{em} 516 nm) of the fluorescein as reaction product on the y-axis. Lines indicate sorting gate P2 separating most fluorescent events (appearing above) from non-/weak-fluorescent events. (**A**) CelA2-multiple SSM-G4$_{288/524/368/576}$ plasmid DNA library during sorting of the 4.40 % most fluorescent events. (**B**) Reanalysis of sorted library with a 12-fold enriched active fraction (53.40 % fluorescent events).

The plasmid DNA of the active variants of the CelA2-multiple SSM-G4$_{288/524/368/576}$ library was recovered from the sorted (w/o/w) emulsion fraction using NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG) in combination with subsequent PCR amplification according to the optimized protocol (see Material and Methods 4.8.11). The recovered plasmid DNA of variants from the sorted active fraction of SSM-G4$_{288/524/368/576}$ library was transformed in *E.coli* BL21 Gold(DE3)lacI^{Q1} for determination of enrichment ratios for the active fraction in MTP. Activity determination of 352 randomly picked sorted clones was performed using 4-MUC activity assay (Lehmann *et al.*, 2012) and revealed with 53.12 % active clones a 1.3-fold enrichment of the active fraction compared to the unsorted library containing 39.49 % active fraction (Figures 75, 77; Figure A14, appendix). Clones were determined as "active" when the increase in fluorescence was at least 0.2 RFU/s which is 13-fold the average activity of the negative control (average activity CelA2-H288F-E580Q: 0.0077 RFU/s).
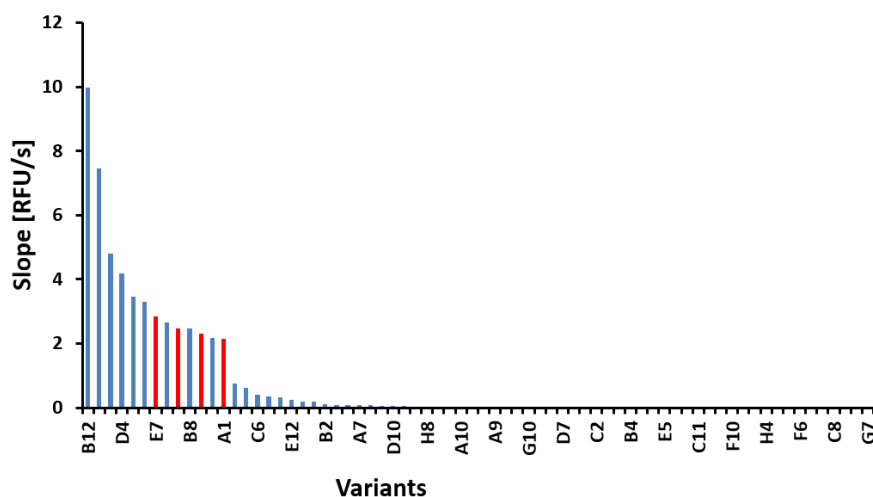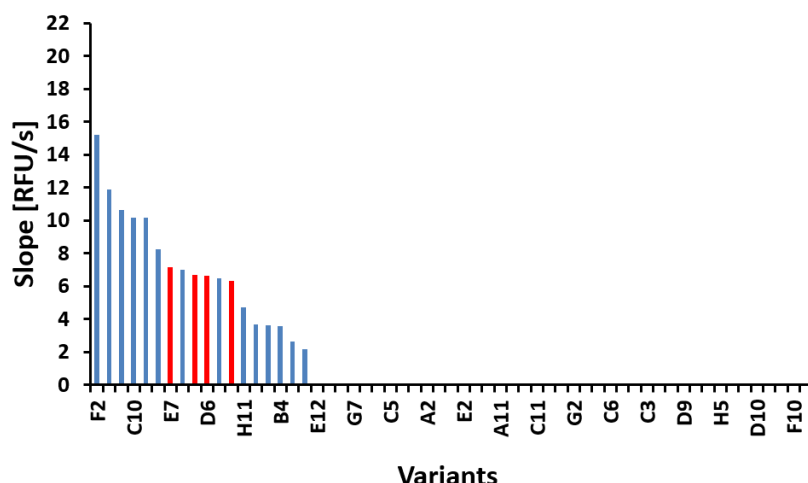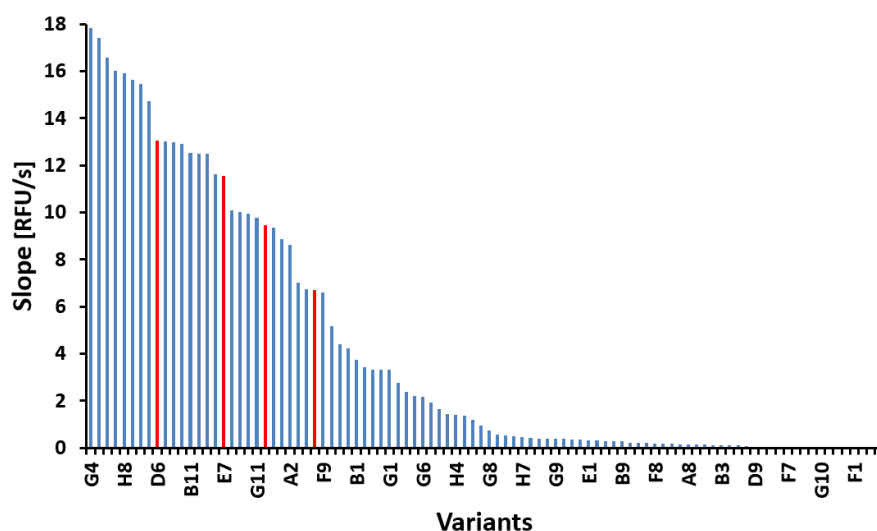
Figure 77: Analysis of the sorted CelA2-multiple SSM-G4$_{288/524/368/576}$ library (MTP No. 1) with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s).

4.5.4. Sequence analysis of beneficial variants of CelA2-multiple SSM libraries

4.5.4.1. CelA2-multiple SSM-G2$_{288/524}$ library

MTP analysis of CelA2-multiple SSM-G2$_{288/524}$ library revealed 24 cellulase variants showing significantly improved activity in comparison to parent CelA2-H288F (up to 4.8-fold) and the 13 best variants were selected for rescreening (Figure 78).
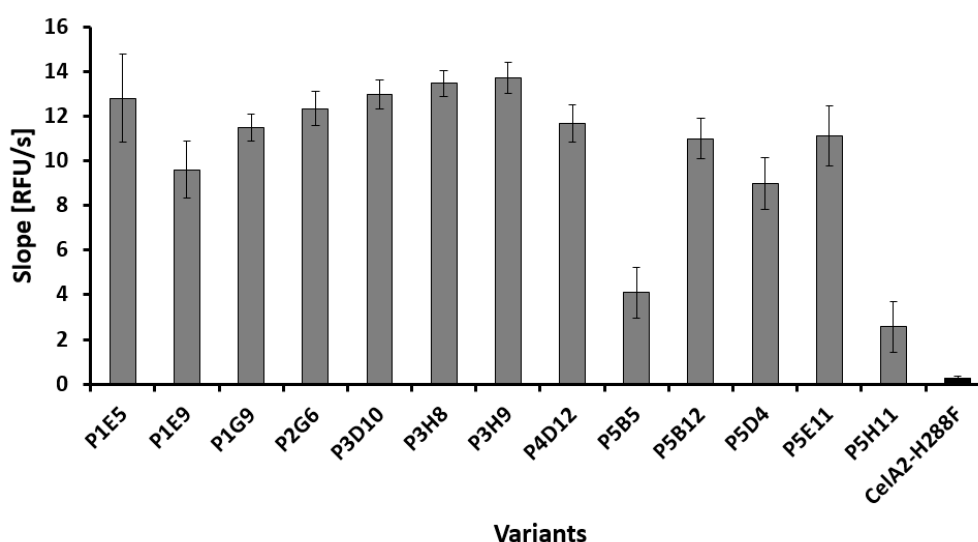
Figure 78: Analysis of the rescreening of the 13 best variants from the CelA2-multiple SSM-G2$_{288/524}$library with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as reference is highlighted in black.

Sequencing analysis of the improved variants P3D10 (H288F/M483T/H524M) and P3H9 (H288F/C337R/H524H) and of three randomly picked clones from the CelA2-multiple SSM-G2$_{288/524}$ library revealed substitutions F288G, F288L and H524S. Variants showed additional amino acid substitutions besides the targeted positions, *e.g.* M483T and C337R.

4.5.4.2. CelA2-multiple SSM-G3$_{288/524/368}$ library

MTP analysis of CelA2-multiple SSM-G3$_{288/524/368}$ library revealed 69 cellulase variants showing significantly improved activity in comparison to parent CelA2-H288F (up to 4.2-fold). The 30 most beneficial variants were selected for rescreening (Figures 79, 80).
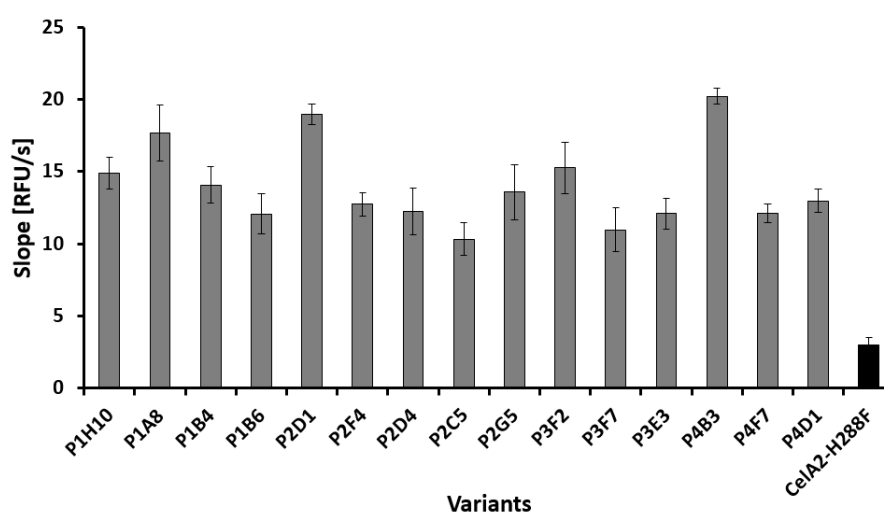
Figure 79: Analysis of the rescreening of the 15 best variants from the CelA2-multiple SSM-G3$_{288/524/368}$ library before sorting with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as reference is highlighted in black.
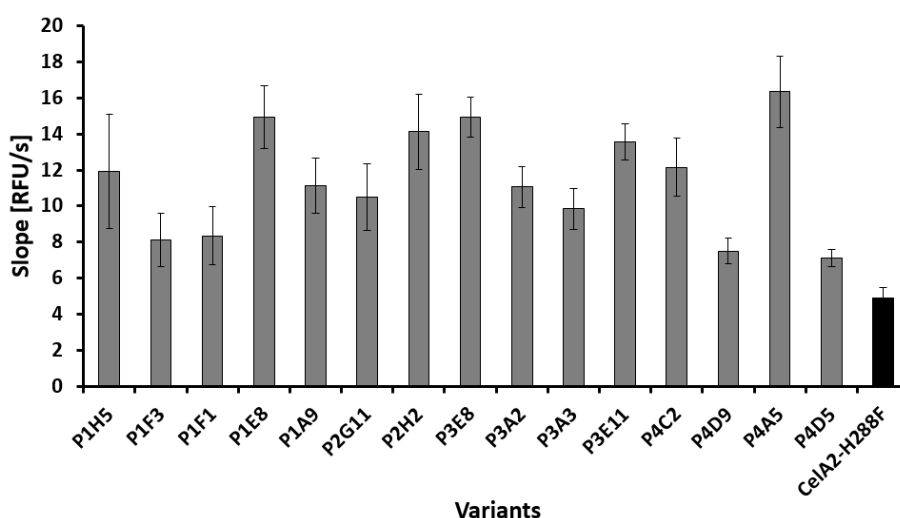
Figure 80: Analysis of the rescreening of the 15 best variants from the sorted CelA2-multiple SSM-G3$_{288/524/368}$ library with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as reference is highlighted in black.

Sequencing analysis of the beneficial variants of the rescreening from the CelA2-multiple SSM-G3$_{288/524/368}$ library revealed substitutions H524Q, H524M, H524N in combination with H288F and additional amino acid substitutions besides the targeted positions, *e.g.* C337R, M483T and V540I (Table 15). Position F288 and Y368 was kept like in the parent CelA2-H288F and was not substituted in any variant from the SSM-G3$_{288/524/368}$ library rescreening, which showed improved activity towards 4-MUC substrate compared to the parent CelA2-H288F (Table 15). Therefore, it can be concluded that phenylalanine and tyrosine at positions F288 and Y368 are the most beneficial for cellulolytic activity.

Table 15: Sequence analysis of 19 clones from the rescreening plates of CelA2-multiple SSM-G3$_{288/524/368}$ library. The variants showing the highest improvement for 4-MUC substrate in comparison to CelA2-H288F are bolded.

Clone	Amino acid exchanges
CelA2-multiple SSM-G3$_{288/524/368}$ library before sorting	
P1A8	**H288F/Y368Y/H524Q**
P1B4	H288F/C337R/Y368Y
P1B6	H288F/R324R/Y368Y/M483T/H524M
P2D1	**H288F/Y368Y/H524N**
P2F4	H288F/R324R/Y368Y/M483T/H524M
P2C5	H288F/Y368Y/H524M/L302P
P3E3	H288F/Y368Y/H524M
P4B3	**H288F/Y368Y/M483T/H524M/I558M**
P4F7	H288F/Y357H/Y368Y/H524N
CelA2-multiple SSM-G3$_{288/524/368}$ library after sorting	
P1H5	H288F/Y368Y/P424P/E485G/A545A
P1F1	M50V/H288F/Y368Y
P1E8	**V37A/V69V/K141K/H288F**
P1A9	H288F/S300G/Y368Y/K449K
P3E8	**H288F/Y368Y/H524N/V540I**
P3A2	H288F/Y368Y/K404K/K453K/E485G/T533A
P4C2	E85E/R187R/H288F/Y368Y
P4D9	F185S/H288F/Y368Y/H525H
P4A5	**M14V/T192A/H288F/Y368Y/E427G/H524N**
P4D5	T292A/H288F/Y368Y/E485G/I551I

4.5.4.3. CelA2-multiple SSM-G4$_{288/524/368/576}$ library

MTP analysis of CelA2-multiple SSM-G4$_{288/524/368/576}$ library revealed 82 cellulase variants showing significantly improved activity in comparison to parent CelA2-H288F (up to 3.9-fold) and the 14 best variants were selected for rescreening (Figures 81, 82).
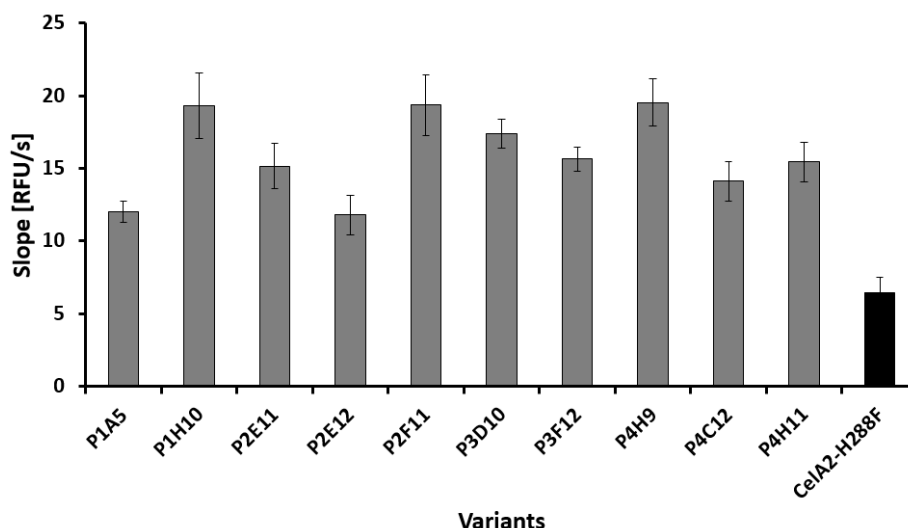
Figure 81: Analysis of the rescreening of the ten best variants from the CelA2-multiple SSM-G4$_{288/524/368/576}$ library before sorting with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as reference is highlighted in black.
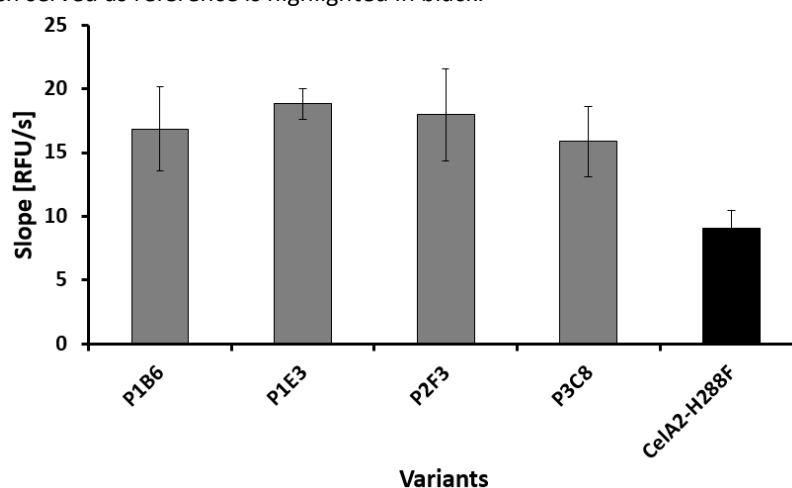
Figure 82: Analysis of the rescreening of the four best variants from the sorted CelA2-multiple SSM-G4$_{288/524/368/576}$ library with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as reference is highlighted in black.

Sequencing analysis of the beneficial variants of the rescreening from the CelA2-multiple SSM-G4$_{288/524/368/576}$ library revealed substitution H524N in combination with F288, Y368, Y576 and additional amino acid substitutions besides the targeted positions, *e.g.* C337R, E485G and V540I (Table 16). Positions H288F Y368, and Y576 were mostly kept like in the parent CelA2-H288F and were not substituted in any variant from the SSM-G3$_{288/524/368/576}$ library rescreening, which showed improved activity towards 4-MUC substrate compared to the parent CelA2-H288F (Table 16). Additionally, randomly selected inactive or low active variants (<CelA2-H288F activity) showed amino acid substitutions at position Y576 to the

amino acids Q, L, D and K (Table 16). Therefore, it can be concluded replacement of phenylalanine and tyrosine by any other amino acid at positions F288 and Y368 is not beneficial in order to achieve improved cellulolytic activity.

Table 16: Sequence analysis of 14 clones from the rescreening plates of CelA2-multiple SSM-G4$_{288/524/368/576}$ library and four randomly chosen clones. The variants showing the highest improvement for 4-MUC substrate in comparison to CelA2-H288F are bolded.

Clone	Amino acid exchanges
CelA2-multiple SSM-G4$_{288/524/368/576}$ library before sorting	
P1A6	I226V/H288F/Y368Y/H524H/Y576Y
P1H10	**H288F/Y368Y/H524N/V540I/Y576Y**
P1E1	H288F/Y263H/Y368Y/H524N/V540I/Y576Y
P2F11	**H288F/Y368Y/P424P/E485G/H524N/V540I/Y576Y**
P2C6	V37A/I69I/K141K/H288F/D339D/Y368Y/R377H/H524H/Y576Y
P3D10	F251L/H288F/Y368Y/H524N/Y576Y
P3F12	T276A/H288F/Y368Y/H524H/Y576Y
P3A8	L97L/H288F/C337R/Y368Y/H524H/Y576Y
P4H9	**H288F/C337R/Y368Y/H524N/V540I/Y576Y**
P4D2	F185S/H288F/Y368Y/H524H/Y576Y
CelA2-multiple SSM-G4$_{288/524/368/576}$ library after sorting	
P1G4	L97L/H288F/C337R/Y368Y/H524H/Y576Y
P1G3	H288F/Y368Y/H524H/E532K/Y576Y
P1B6	**L97L/H288F/C337R/Y368Y/H524H/Y576Y**
P1E3	**F185S/H288F/Y368Y/E485G/H524H/A545AY576Y**
P2F3	**H288F/C337R/Y368Y/H524H/Y576Y**
P3C8	L97L/H288F/G190R/C337R/Y368Y/H524H/Y576Y
P4B9	T276A/H288F/Y368Y/H524H/V540G/Y576Y
P4F2	H288F/Y368Y/H524H/E532K/Y576Y
Additional randomly chosen variants	
P3C1	H288F/Y357H/Y368Y/H524H/Y576Q
P3C2	H288F/T292A/Y368Y/E485G/H524H/Y576L
P3C4	H288F/Y368Y/P424P/E485G/H524H/A545A/Y576D
P3C5	I226V/H288F/T292A/Y368Y/E485G/H524H/Y576K

4.6. CelA2-OmniChange library

4.6.1. CelA2-OmniChange library generation and screening strategy

The third approach uses simultaneous SSM at the targeted positions (F288, Y368, H524, and Y576) as focused mutagenesis strategy, employing the OmniChange (Dennig *et al.*, 2011) technique in combination with the uHTS-IVC platform for screening of the generated library and enrichment of active cellulase variants for further analysis in MTP. In this approach, highly diverse CelA2-OmniChange library was generated using pIX3.0RMT7-CelA2-H288F as template and simultaneous saturation of amino acids F288, Y368, H524, and Y576 in close

proximity to the active center of CelA2-H288F, which were selected by previous computational analysis (see 4.2.2). The selected positions were saturated according to the OmniChange protocol (Dennig *et al.*, 2011) with specific degenerated, phosphorothioated NNK-primers for saturation of each position with all 20 amino acids and subsequent ligase free cloning *via* iodine cleavage and hybridization (see Material and Methods 4.8.5, Table 22). The resulting PCR products were analyzed by agarose-gelelectrophoresis, purified with NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG), *Dpn*I digested and transformed into chemically competent *E.coli* BL21 Gold(DE3)lacI^{Q1} cells for expression and activity tests in MTP-based format. For analysis with the uHTS-IVC platform, transformation was performed in commercial, ultracompetent *E.coli* XL10-Gold cells (Aglient technologies) with a transformation efficiency of ≥5 x 10^9 transformants/µg pUC18 control plasmid DNA. Activity determination of 352 randomly picked clones using MTP-based 4-MUC activity assay (Lehmann *et al.*, 2012) was performed to determine the percentages of active variants before and after sorting. The generated CelA2-OmniChange-G4$_{288/524/368/576}$ plasmid library was subsequently analyzed with FDC substrate using the developed uHTS-IVC platform for enrichment of beneficial cellulase variants. The beneficial variants of the simultaneous CelA2-OmniChange-G4$_{288/524/368/576}$ saturation mutagenesis were enriched, by sorting of active variants, pooled and recovered from the emulsions for transformation and analysis with 4-MUC activity assay in MTP. The screening strategy is depicted in the scheme (Figure 83). In this approach, more than 1.04 x 10^6 different variants were generated (Table 10) using simultaneous SSM *via* OmniChange and flow cytometer-based screening with the developed uHTS-IVC platform, which facilitates to screen the full generated diversity (1.04 x 10^6 different variants). For the generated OmniChange library with four simultaneously saturated positions an 3-times oversampling is mandatory to screen the whole diversity of the library with a ~95 % confidence (Reetz *et al.*, 2008). Therefore, at least 3.15 x 10^6 variants have to be screened (Table 10), which is only possible using rapid and efficient screening technologies like flow cytometry (Martinez *et al.*, 2013; Reetz *et al.*, 2008). For this reason the simultaneous SSM method OmniChange in combination with the throughput achieved with the uHTS-IVC technology platform is a powerful tool to obtain full coverage upon screening of the generated sequence space, which enhances the chances to identify combinatorial effects and to elucidate structure-function relationships for successful protein engineering.

synthesis is completely independent of host cells and therefore significantly faster than protein production in living cells, so synthesis and assays can be carried out within a few hours (Sawasaki *et al.*, 2002). As no living cells are needed, these experiments can be carried out in laboratories without S1 restrictions, making the work much easier and cheaper. There are several eukaryotic and prokaryotic systems available for *in vitro* expression of proteins: *E.coli-*, wheat germ- and rabbit reticulocytes-system (Nakano *et al.*, 1998). The prokaryotic *E.coli* systems are mainly used for structural studies of proteins, as they have higher yields and homogeneous samples while the eukaryotic ones usually have lower yields and are used mainly for functional studies on co- or post-translationally modified proteins (Jackson *et al.*, 2004). The general procedure for cell-free protein expression is the incubation of the template DNA (linear or circular DNA) together with the *in vitro* expression mixture, (containing the *E.coli* extract, including ribosome, polymerases, etc., buffer, chaperones and nuclease free water), for one hour or longer, yielding several micrograms of the protein. The successful cell-free production of active protein can be analyzed subsequently on SDS-PAGE and/or with colorimetric or fluorimetric enzyme-based activity assays.

A system for *in vitro* expression of proteins in emulsion compartments was first developed by Tawfik and Griffiths in 1998, using DNA-binding enzymes as proof-of-principle (Tawfik *et al.*, 1998). Further developments of this system were realized by Leemhuis and his colleagues, using emulsion technology for the evolution of catalytic properties of enzymes (Leemhuis *et al.*, 2005). This technique can be used for HTS, as 10^{10} individual compartments, all entrapping different genes, can be generated (Bernath *et al.*, 2004a). In this context the method of emulsion generation and the used chemicals play an important role, as some homogenization methods cause a loss of activity of the *in vitro* expressed proteins and oxidize the translational machinery resulting in a shutdown of protein expression (Miller *et al.*, 2006). A solution for this problem was found by using silicone surfactants, like ABIL EM 90, as oxidation of the translation mix components was prevented and the protein expression was even increased with 40 % in comparison to non-emulsified samples (Miller *et al.*, 2006). Until now only few proteins, like β-galactosidase (Mastrobattista *et al.*, 2005), could be successfully expressed inside emulsion compartments in directed evolution campaigns, so the development of a screening platform suitable for different classes of enzymes is necessary, for engineering enzymes with novel properties and catalytic characteristics comparable to those of natural enzymes.

36,757,972 events were analyzed (4,275 events s^{-1}) of which 395,229 events were sorted with an efficiency of 90 %. Reanalysis of the sorted population revealed a 30-fold enrichment in active fraction, determined by an increase of the population in the P2 sorting gate from 1.8 % to 54.0 % (Figure 84B).
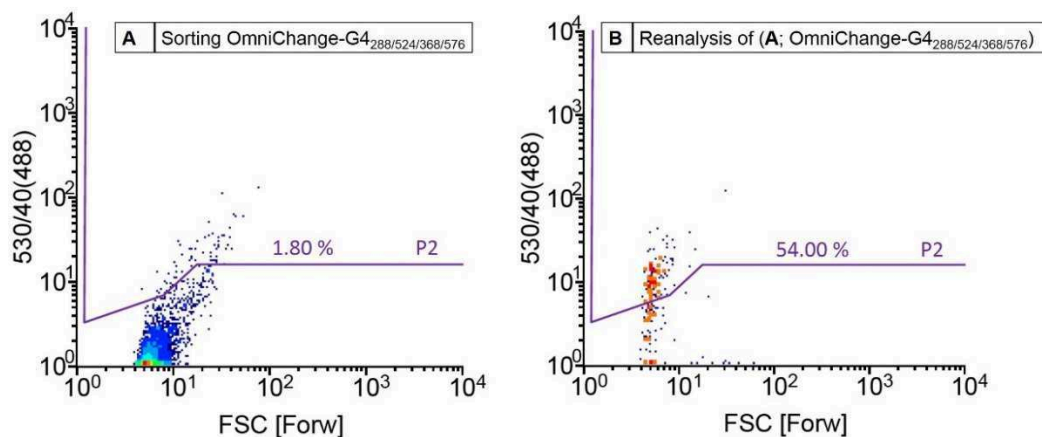
Figure 84: Flow cytometer analysis showing density plots of CelA2-OmniChange-G4$_{288/524/368/576}$ plasmid DNA library after *in vitro* expression in (w/o/w) emulsions (4 h, 30 °C). The FSC represents the size of the emulsion on the x-axis and 530/40 (488) represents the fluorescence intensity (λ_{ex} 494 nm and λ_{em} 516 nm) of the fluorescein as reaction product on the y-axis. Lines indicate sorting gate P2 separating the most fluorescent events (appearing above) from non-/weak-fluorescent events. (**A**) CelA2-OmniChange-G4$_{288/524/368/576}$ plasmid DNA library during sorting of the 1.8 % most fluorescent events. (**B**) Reanalysis of sorted library with a 30-fold enriched active fraction (54.00 % fluorescent events).

4.6.4. Analysis of direct CelA2-OmniChange library with 4-MUC and FDC activity assay in MTP format before and after sorting

The plasmid DNA of the active variants of the CelA2-OmniChange-G4$_{288/524/368/576}$ library was recovered from the sorted (w/o/w) emulsion fraction (Figure A17, appendix) using NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG) in combination with subsequent PCR amplification according to the optimized protocol (see Material and Methods 4.8.11). The recovered plasmid DNA of variants from the sorted active fraction of CelA2-OmniChange-G4$_{288/524/368/576}$ library was transformed in *E.coli* BL21 Gold(DE3)lacI^{Q1} for determination of enrichment ratios for the active fraction in MTP. Activity determination of 352 randomly picked sorted clones was performed using 4-MUC activity assay (Lehmann *et al.*, 2012) and revealed with 78.13 % active clones a 30-fold enrichment of the active fraction compared to the unsorted library containing 2.56 % active fraction (Figure 85B; Figure A18, appendix). Clones were determined as "active" when the increase in fluorescence was at least 0.2 RFU/s which is 10-fold the average activity of the negative control (average activity CelA2-H288F-E580Q: 0.019 RFU/s).
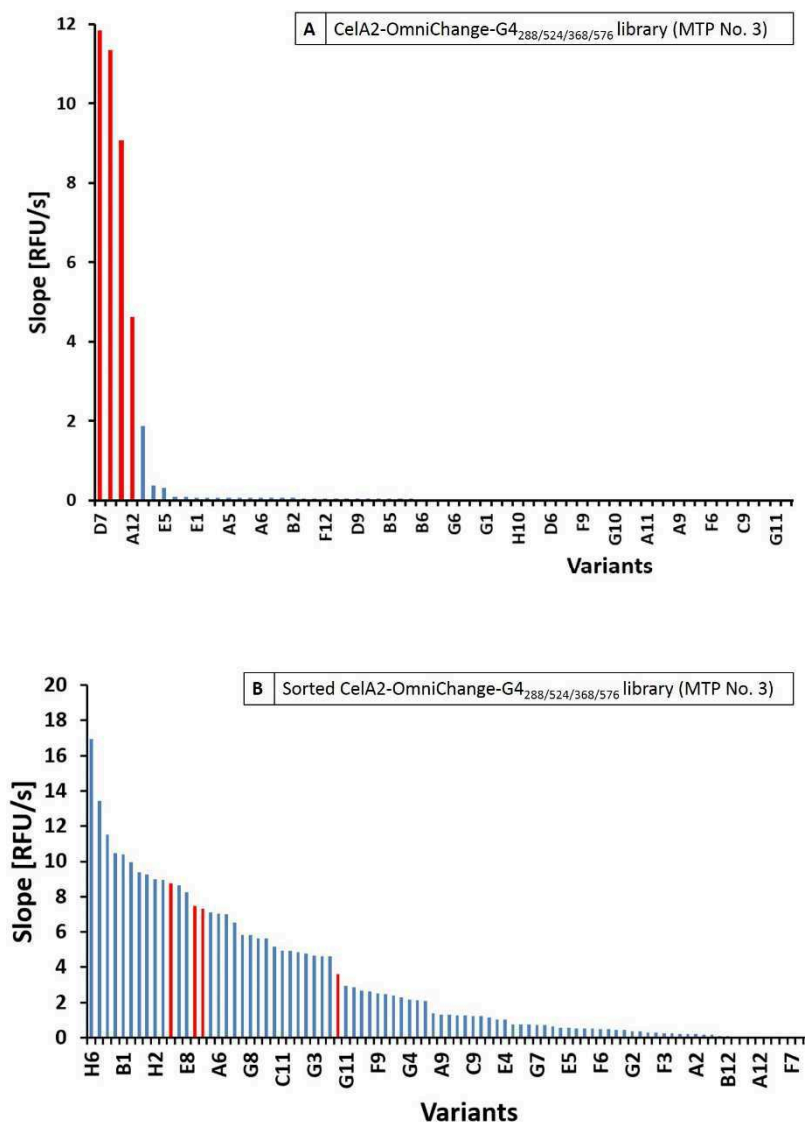
Figure 85: Analysis of the CelA2-OmniChange-G4$_{288/524/368/576}$ library before and after sorting, (**A**) CelA2-OmniChange-G4$_{288/524/368/576}$ library (MTP No. 3), (**B**) sorted CelA2-OmniChange-G4$_{288/524/368/576}$ library (MTP No. 3), with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s).

4.6.5. Sequence analysis of CelA2-OmniChange library

MTP analysis of CelA2-OmniChange-G4$_{288/524/368/576}$ library revealed 25 cellulase variants showing significantly improved activity in comparison to parent CelA2-H288F (up to 2.5-fold) and the four best variants were selected for rescreening (Figure 86).
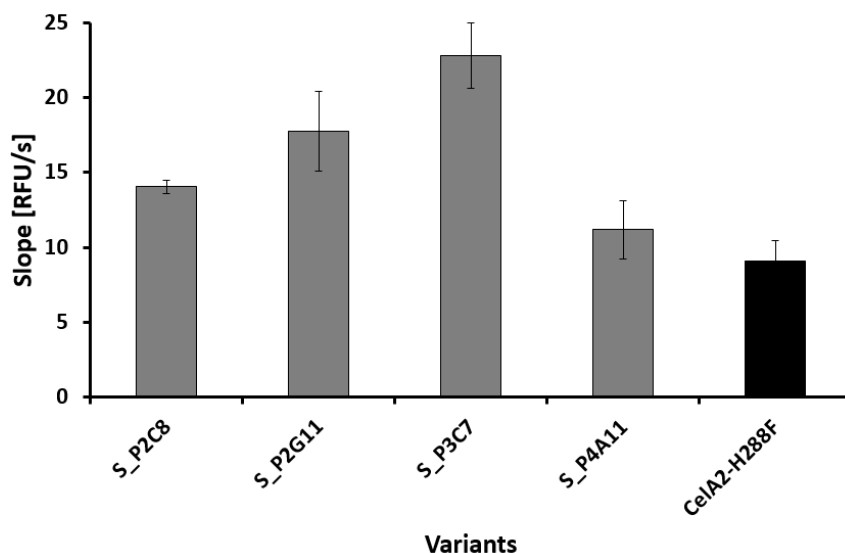
Figure 86: Analysis of the rescreening of the four best variants from the sorted CelA2-OmniChange-G4$_{288/524/368/576}$ library with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as reference is highlighted in black.

Sequencing analysis of the rescreened beneficial variants, sorted from CelA2-multiple SSM-G4$_{288/524/368/576}$ library, was performed revealing mainly additional amino acid substitutions, *e.g.* G151S, T276A, Y357H, and L21P besides the targeted positions F288, Y368, H524 and Y576 (Table 17). In general all four positions were harboring amino acid substitutions indicating a good library quality. For the rescreened variants showing an increased activity towards 4-MUC compared to parent CelA2-H288F, no substitutions at targeted positions were observed different from substitutions in parent, indicating beneficial effects of additional amino acid substitutions for cellulolytic activity towards MUC (Table 17).

Table 17: Sequence analysis of 4 best clones from the rescreening plate of sorted CelA2-OmniChange-G4$_{288/524/368/576}$ library and additional inactive and low active clones from the library before sorting. The variant showing the highest improvement for 4-MUC substrate in comparison to CelA2-H288F is bolded.

Clone	Amino acid exchanges
CelA2-OmniChange-G4$_{288/524/368/576}$ library before sorting	
P4G1	F288F/Y368Y/H524 Stop/Y576S
P4G2	F288Stop/Y368N/H524K/Y576H
P4G3	F288W/Y368 Stop/H524Y/Y576Y
P4G4	F288F/Y368L/H524V/Y576I
P4G5	F288A/Y368N/H524L/Y576L
P4G6	D221V/F222I/F288V/Y368G/H524P/Y576P
P4G7	F288 Stop/Y368A/H524 Stop/Y576 Stop
P4G8	Q6 Stop/F288F/Y368Y/H524A/Y576T
P4G9	L97L/F288F/C337R/Y368Y/H524H/Y576G
P4G10	M50V/ F288F/G313G/Y368Y/H524H/Y576S
P1A2	F288V/Y368V/H524K/Y576H
P1A3	F288V/Y368V/H524H/Y576Y
P1A5	F288Stop/Y368D/H524Y/Y576I
P1A6	F288N/Y368K/H524T/Y576Q
P1A7	F288R/Y368S/H524T/Y576S
P1A8	F288K/Y368Y/H524E/Y576R
P1A9	D287N/F288K/Y368F/H524F/Y576R
P1A11	F288S/Y368K/H524T/Y576H
CelA2-OmniChange-G4$_{288/524/368/576}$ library after sorting	
S_P2C8	F288F/Y357H/Y368Y/H524H/Y576Y
S_P3H6	L21P/T276A/F288F/Y368Y/H524H/Y576Y
S_P3C7	**G151S/F288F/Y368Y/H524H/Y576Y**
S_P4A11	F288F/Y368Y/H524H/Y576Y

4.6.6. Purification and kinetic characterization of CelA2 and its variants

As an outcome of the screening of three different types of gene diversity libraries a cellulase variant CelA2-H288F-M3 (H524Q) with 5-fold increase in activity towards 4-MUC compared to the parent was identified (Table 12). CelA2-H288F-M3 (H524Q), revealed to be the best variant, which contained no additional amino acid substitutions besides the targeted ones and was selected for purification and detailed kinetic characterization. CelA2 wild type, CelA2-H288F parent, and variant CelA2-H288F-M3 (H524Q) were purified using the Protino® Ni-TED 2000 packed columns (Macherey-Nagel GmbH & Co. KG), the collected, purified fractions were analyzed with 4-MUC activity assay (Lehmann *et al.*, 2012), to confirm the activity of the cellulase in the purified samples, and by SDS-PAGE (Figure 87). Sample preparation and purification procedure is described in detail in Material and Methods, section 4.8.14.
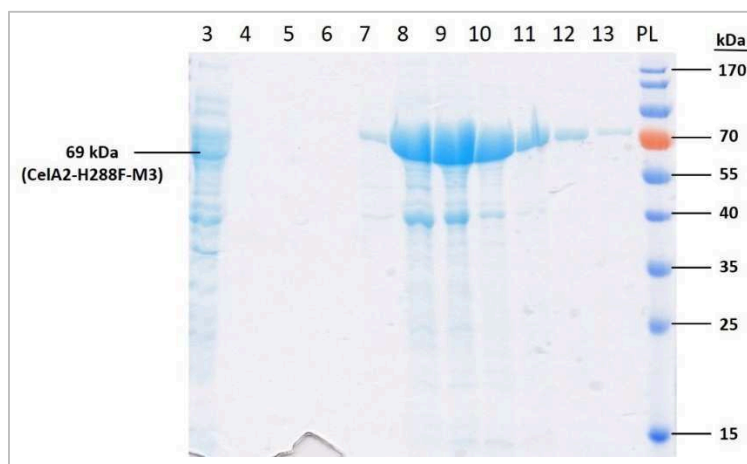
Figure 87: Result of the SDS-PAGE, showing the bands of the collected, purified fractions (10-13) from CelA2-H288F-M3 at the expected size of 69 kDa. PL indicates the protein ladder. Fractions (10-13) of highest purity were pooled together.

The purest fractions were pooled together (Figure 87) and the purity of the variants was determined with the Experion™ Automated Electrophoresis System from Bio-Rad Laboratories GmbH (Figure 88).
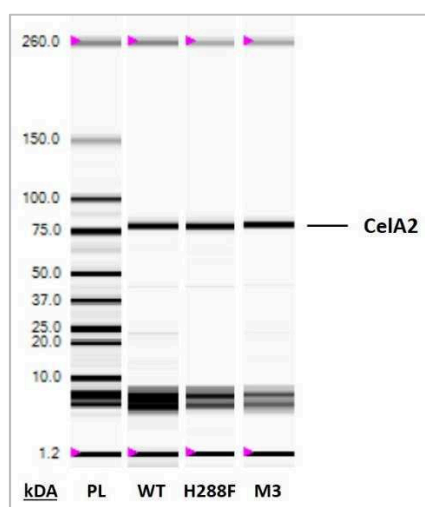
Figure 88: Result of the Experion™ analysis, showing the bands for purified CelA2-WT (WT), CelA2-H288F (H288F) parent, and variant CelA2-H288F-M3 (M3). PL indicates the protein ladder showing the respective sizes of the bands in kDa. The samples were prepared according to the protocol of the manufacturer and analyzed in duplicate.

For all variants >89 % purity was obtained and the total protein concentration of the purified samples was determined using the Pierce™ BCA Protein Assay Kit (Life Technologies GmbH). Specific protein concentrations were calculated for each sample as listed in Table 18. For determination of kinetic parameters CelA2-WT, CelA2-H288F and variant CelA2-H288F-M3 were diluted to an equal protein concentration of 10 µg/ml.

Table 18: Calculated specific enzyme concentrations and obtained purity for purified CelA2-WT, CelA2-H288F (parent) and variant CelA2-H288F-M3. The specific enzyme concentrations were calculated according to the total protein concentration obtained with the Pierce^{TM} BCA Protein Assay Kit (Life Technologies GmbH) and the purity determined by analysis with Experion^{TM} Automated Electrophoresis System (Bio-Rad Laboratories GmbH).

Sample	Specific protein concentration [µg/ml]	Purity [%]
CelA2-WT	332.04	89.03
CelA2-H288F	423.86	95.69
CelA2-H288F-M3	481.40	94.95

The kinetic parameters (k_{cat}, K_M) were calculated and compared and the results are summarized in Table 19 and Figure 89. The best identified variant, CelA2-H288F-M3 (H524Q), showed 8-fold increased K_{cat} value compared to CelA2-H288F parent (Table 19). The variant CelA2-H288F-M3 (H524Q) showed drastically increased K_M (6.7-fold) in comparison to parent (Table 19). The determined specific activity value of CelA2-H288F-M3 (H524Q) is 41-fold higher compared to CelA2 wild type and 8-fold higher compared to its parent CelA2-H288F (Table 19). The calculated specific activity of CelA2-WT is in agreement with the published data from Lehmann *et al.* (15.1 U mg^{-1}).

Table 19: Kinetic characterization of improved variant CelA2-H288F-M3 with 4-MUC activity-based assay. Cellulase kinetics were determined in 0.2 M KPI-buffer containing different concentrations of 4-MUC (2.5-200 µM) at 30°C and pH 7.2 for CelA2-WT, CelA2-H288F, and –M3. One Unit (U) of enzyme was defined as the amount of cellulase that catalyzes the conversion of 1 µmol of 4-MUC per minute. All values reported are the average of three measurements. Standard errors determined in triplicate measurements are represented in the brackets. Values are normalized to protein content based on analysis with Experion^{TM} Automated Electrophoresis System (Bio-Rad Laboratories GmbH).

	k_{cat} [min^{-1}]	K_M [µM]	k_{cat}/K_M [min^{-1} µM^{-1}]	Specific activity [U mg^{-1}]	Amino acid exchanges
CelA2-WT	0.11 (± 0.007)	22.53 (± 4.78)	0.005	16.27 (± 1.04)	-
CelA2-H288F	0.58 (± 0.020)	7.66 (± 1.12)	0.075	83.87 (± 2.89)	H288F
CelA2-H288F-M3	4.65 (± 0.263)	51.19 (± 7.55)	0.091	674.50 (± 38.17)	H288F/H524Q
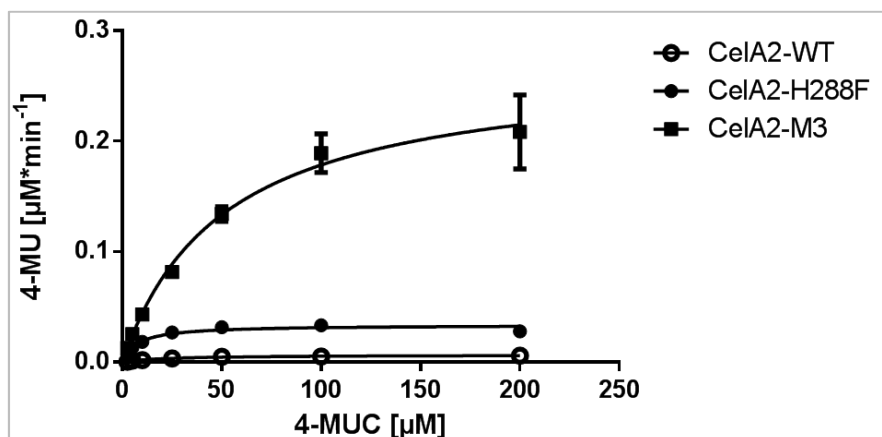
Figure 89: Michaelis-Menten plots of CelA2-WT, CelA2-H288F, and variant CelA2-H288F-M3. Enzyme kinetics were obtained with 4-MUC activity assay at pH 7.2 in KPI- buffer. 4-MUC substrate concentrations between 2.5 and 200 µM were used for determination of kinetic parameters with 4-MUC activity assay. Initial activities were plotted against different substrate concentrations and fitted with Michaelis-Menten equation using GraphPad Prism 6 software for analysis and determination of kinetic parameters. The reported values are the average of three measurements and deviations are calculated from the corresponding mean values. The x-axis shows the 4-MUC substrate concentration in [µM] and the y-axis shows the 4-MU product formation in [µM*s^{-1}].

4.6.7. Analysis of beneficial amino acid exchanges using the homology model of CelA2-WT

The analysis of the beneficial additional amino acid substitutions H524N, H524M, H524Q and V540I, illustrated in the hybrid homology model for CelA2 (Figure 90; see Material and Methods, 4.8.17), revealed that identified amino acid substitutions at the saturated position 524 show improved cellulolytic activity. A hypothesis for this phenomenon could be that substitution of histidine by asparagine or glutamine at position 524 (H524N or H524Q) might be beneficial, as they are polar and have uncharged side chains, in contrast to the original amino acid histidine, which contains an aromatic ring. In addition, it could be speculated that amino acid substitution V540I, which is also located in close proximity to the active site of the enzyme, might be beneficial because isoleucine and valine are both amino acids containing hydrophobic side chains, but isoleucine is longer than valine which could lead to improved hydrophobic interaction with the substrate upon substrate cleavage reaction.

1.6. *Enzymes*

1.6.1. *Cellulases*

Cellulases (EC-number: 3.2.1.4) catalyze the hydrolysis of β-1,4-glycosidic bonds in cellulose and cellulose compounds (polymers build-up of anhydro-β-1,4-glucose units connected by β-1,4-D-glycosidic bonds). There are three different types of cellulases (Percival Zhang *et al.*, 2006): (1) endoglucanases (EC-number: 3.2.1.4), hydrolyzing internal β-1,4-glycosidic bonds, (2) exoglucanases (EC-number: 3.2.1.91), hydrolyzing β-1,4-glycosidic bonds at the non-reducing end and (3) β-glucosidases (EC-number: 3.2.1.21), cleaving cellobiose and oligosaccharides to glucose. Figure 6 shows the hydrolysis of cellulose by the cellulolytic system of *Trichoderma reesei* in form of a simplified scheme. At internal amorphous sites in the cellulose polysaccharide chain endoglucanases cleave the β-1,4-D-glycosidic bonds to oligosaccharides of different lengths and chain ends (reducing or non-reducing ends) (Figure 6). In contrast to endoglucanases, exoglucanases are able to cleave microcrystalline cellulose, releasing cellobiose as reaction product, acting either from the reducing end (CBHI) or from the non-reducing end (CBHII) (Teeri, 1997). The soluble cellobiose can be converted subsequently to glucose by β-glucosidase. As for complete cellulose conversion the accessibility of the cellulase to its substrate is main limitation, depending on physical properties like polymerization, crystallinity and accessibility, determined by size and porosity of the cellulose, most cellulases have different domain structures, *e.g.* the cellulose binding domain (Chandra *et al.*, 2007). The cellulose binding domain (CBD) is connected by a flexible linker to the catalytic domain and enables the cellulase to recognize the solid surface of the cellulose and facilitates the movement across a single chain of decrystallized cellulose fiber, which is immediately moving into the active site tunnel of the catalytic domain. It was reported that the binding efficiency of the cellulase on the cellulose surface directly correlates with the efficiency of crystalline cellulose degradation (Esteghlalian *et al.*, 2001; Linder *et al.*, 1995) and a removal of the CBD results in a drastic decrease (80 %) in cellulase activity on solid substrates (Gilkes *et al.*, 1988).

4.7. Discussion

In this chapter the developed flow cytometer-based uHTS-IVC platform was validated by screening of two different powerful focused mutagenesis libraries (CelA2-multiple SSM and OmniChange) using cellulase as model enzyme. The obtained improved enzyme variants were compared to the cellulase variant identified upon screening of standard Iterative Saturation Mutagenesis (ISM) library. The selected positions for SSM (F288, Y368, H524, and Y576) are located in close proximity to the active site, which increases the probability of discovering combinatorial effects among the targeted amino acid positions. The targeted positions were chosen based on computational analysis using a homology model of CelA2-WT which was generated with YASARA software.

Determination of kinetic parameters for the best identified variant from all three approaches, CelA2-H288F-M3 (F288/Q524), showed 8-fold increased K_{cat} value compared to CelA2-H288F parent (Table 19), 41-fold higher specific activity compared to CelA2-WT and 8-fold higher specific activity compared to its parent CelA2-H288F (Table 19). These results show clearly that a combinatorial effect between F288 and Q524 could be successfully identified.

Analysis of the CelA2-ISM libraries showed that positions F288 and Y368 and Y576 (Tables 11, 13, 14) are crucial for keeping the activity at the level of the highly active parent (CelA2-H288F), as the majority of amino acid exchanges at these positions result in a drastic loss of activity towards 4-MUC (<1 RFU/s) and FDC substrate (<2 RFU/s), except for substitutions F288W (4.29 RFU/s) and Y576F (5.28 RFU/s). Simultaneous saturation of positions H524 and F288 yielded variants with improved activity compared to the parent, *e.g.* F288/H524M (1.9-fold), H524C (2.8-fold) and F288/H524Q (4.8-fold). Due to the fact, that substitutions at positions F288, Y368 and Y576 result exclusively in variants exhibiting low activity levels, the best identified variant of the screened CelA2-ISM libraries is CelA2-H288F-M3 (F288/Q524). The quality of CelA2-ISM libraries was proofed by sequencing analysis of 24 clones/library (Tables 11-14), showing an amino acid substitution pattern at all four targeted positions (F288, Y368, H524, and Y576).

For validation of the flow cytometer-based uHTS-IVC technology platform, a CelA2-OmniChange library and several subsequent CelA2-multiple SSM libraries were prepared. For generation of each CelA2-multiple SSM library the most active variants from the previous round of mutagenesis and sorting were pooled as a template for the next round

of SSM. This resulted in higher initial percentages of active variants (14.20-39.49 %, Figures 67, 68, 72, 75) compared to generated CelA2-OmniChange library (2.56 %, Figure 85A), due to the lower mutational load, resulting from the subsequent, iterative rounds of single Site-Saturation Mutagenesis and enrichment of active variants. For the CelA2-multiple SSM libraries up to 12-fold enrichment of active variants was achieved using flow cytometer-based analysis (Figure 76). Analysis of the sorted fraction using 4-MUC activity assay in MTP showed only a slight enrichment of active variants (up to 2.4-fold, Figures 72, 74) indicating the entrapment of more than one gene mutant per compartment. Additionally, a stricter sorting strategy (*i.e.* sorting of the 1 % most fluorescent fraction instead of sorting of the 4-5 % most fluorescent fraction) was expected to improve the enrichment in MTP and subsequently proven to be successful upon sorting the 1.8 % most fluorescent fraction of the CelA2-OmniChange-G4$_{288/524/368/576}$ library (Figure 84).

The generated CelA2-OmniChange library, with four simultaneously saturated positions (F288, Y368, H524, and Y576), attaining over 1×10^6 different mutants, was analyzed by flow cytometer-based uHTS-IVC platform. The highly diverse mutant library displaying a low percentage of active variants (2.56 %, Figure 85A) was successfully enriched by impressive 30-fold using flow cytometer-based analysis of over 36×10^6 events and sorting of 3.9×10^5 active events, ensuring full coverage of the generated diversity due to oversampling (Figure 84). Subsequent analysis of the sorted fraction with 4-MUC activity assay in MTP confirmed the enrichment obtained upon reanalysis of active fraction on flow cytometer (Figure 85B).

Sequence analysis of cellulase variants with improved performance compared to the highly active parent CelA2-H288F, obtained by screening with the flow cytometer-based uHTS-IVC platform, revealed amino acid substitutions H524M, H524N and H524Q in combination with phenylalanine at position 288 (H288F, parent), indicating these positions are essential for substrate conversion and cellulase activity. Variant CelA2-H288F/H524M (P3E3) showed 4-fold increase in activity and CelA2-H288F/H524Q (P1A8) and CelA2-H288F/H524N (P2D1) showed 6-fold increase in activity compared to parent using 4-MUC activity assay (Figure 79). These results proof that a combinatorial effect between F288 and Q524, M524 and N524 could be successfully identified, indicating the importance of the generated diversity by the respective mutagenesis method, since some variants, *e.g.* N524 were only identified using multiple SSM library screening employing flow cytometer-based screening.

In all three approaches, the substitution of positions F288, Y368, and Y576 with other amino acids resulted in a loss of activity, indicating the data obtained from CelA2-ISM libraries are in agreement with the data obtained in CelA2-multiple-SSM libraries and in CelA2-OmniChange library, fostering the hypothesis that F288, Y368 and Y576 are essential amino acids for CelA2 activity, which play a key role during substrate binding and catalysis. Due to the loss of activity upon substitution of amino acids at positions F288, Y368 and Y576 no combinatorial effects were observed among all four targeted positions, but only between positions F288 and H524. The quality of CelA2-OmniChange library was proofed by sequencing analysis of 22 clones showing an amino acid substitution pattern at all four targeted positions (F288, Y368, H524, and Y576) and also additional substitutions were identified (Table 17). The same phenomenon was observed during iterative rounds of SSM and screening with uHTS-IVC platform, which showed besides the targeted substitutions (F288, H524, Y368, and Y576) additional mutations, *i.e.* V540I, C337R, or I558M (Table 15). The origin of these additional mutations might be due to the usage of *Taq*-polymerase in the step of DNA recovery from sorted (w/o/w) emulsions. In this way, new positions *e.g.* H288F/C337R, besides the targeted ones were identified, showing an increase in activity of up to 4.7-fold (P1B4, Figure 79; P2F3, Figure 82).

The results derived from the three focused mutagenesis approaches show the developed flow cytometer-based uHTS-IVC technology platform is a powerful pre-screening system, especially for the analysis of highly diverse mutant libraries ($>10^5$ variants) that are not possible to screen with common medium to high throughput screening formats. The uHTS-IVC technology platform enables the simple and rapid enrichment of active variants from a huge fraction ($>10^7$) of mainly inactive variants for further studies in a time and cost efficient manner. Detailed analysis of the sorted active fraction revealed beneficial amino acid substitutions H524N and H524Q, showing high activity towards 4-MUC substrate (6-fold improved compared to CelA2-H288F), in contrast to standard ISM using only screening in MTP, where only one beneficial substitution was discovered (H524Q), leading to 6-fold increased activity towards 4-MUC substrate. Other amino acid substitutions identified during ISM-library screening showed only slightly improved activity (up to 2.8-fold) in comparison to CelA2-H288F. This demonstrates the reliability and robustness of the flow cytometer-based uHTS-IVC technology platform to identify and enrich different highly active enzyme variants from a vast pool of inactive variants, which are not easily obtained using standard

MTP-based screening systems. In addition to mutant libraries using linear DNA templates (see Chapter I), also mutant libraries using plasmid DNA as template for cell-free expression within (w/o/w) emulsion compartments can be successfully pre-screened by flow cytometer. In summary, the uHTS-IVC technology platform is a profitable pre-screening technology for the tailoring and engineering of biocatalysts for industry, as it can be applied to different classes of enzymes, including toxic or membrane bound proteins that are difficult to express *in vivo*. Especially for screening of highly diverse mutant libraries (>10^5 mutants), generated with novel mutagenesis methods like OmniChange, the benefits employing uHTS-IVC technology as pre-screening system are outstanding in regard to throughput, as well as to time and cost efficiency of the screening system.

Figure 6: Schematic representation of the hydrolysis of amorphous and microcrystalline cellulose with the cellulolytic system of *T. reesei*. The solid circles represent non-reducing ends and the open circles represent reducing ends. The figure is modified from Lynd *et al.* (Lynd *et al.*, 2002).

Cellulases can be isolated from fungi, bacteria and protozoa and play an important role in coffee production, biofuel production, laundry, textile, paper and detergent industry. In order to meet the increasing global energy demands it is necessary to find alternatives to fossil fuels, like the production of tailor made biofuels. For the production of liquid biofuels an efficient conversion of lignocellulosic material into ethanol is necessary which can be performed by cellulases, which are hydrolyzing cellulose polymers in biomass into monomeric sugars that can be easily fermented to ethanol (Kumar *et al.*, 2009). The cellulase used for this project is a **CelA2 cellulase** (celA2, GenBank: JF826524.1), consisting of 604 amino acids and having a molecular weight of around 69 kDa. CelA2 was isolated from a metagenomic library from a biogas plant sample (Ilmberger *et al.*, 2012) and was cloned into *E. coli* BL21-Gold(DE3) strain. The variant **CelA2-H288F** revealed an outstanding performance in ionic liquids (IL) and sodium chloride and is therefore an excellent candidate to be applied in the cellulose hydrolysis process in the generation of biofuels (Lehmann *et al.*, 2012).

Table 20: Degenerated primers for generation of Site-Saturation Mutagenesis libraries (CelA2-ISM and CelA2-multiple SSM).

Primer	Sequence (5'→ 3')	Size [nt]
SSM_288_fwd	CATAAAGTTACCGCAAAAGAT**NNK**GCACCGATGACCATTCTG	42
SSM_288_rev	CAGAATGGTCATCGGTGC**MNN**ATCTTTTGCGGTAACTTTATG	42
SSM_368_fwd	TACCACCGGCACC**NNK**AATGATGTTAGCG	29
SSM_368_rev	CGCTAACATCATT**MNN**GGTGCCGGTGGTA	29
SSM_524_fwd	GCACAGAATCCG**NNK**CATCGTCCGAG	26
SSM_524_rev	CTCGGACGATG**MNN**CGGATTCTGTGC	26
SSM_576_fwd	GATGTTACCGGTAGC**NNK**GCCACCAATG	28
SSM_576_rev	CATTGGTGGC**MNN**GCTACCGGTAACATC	28

Table 21: Reaction mixture for performance of the 2-step PCR.

Components (Master mix)	Volume in [µl]
Template (50 ng/µl)	1.0
dNTP-mix	1.0
PfuS polymerase 2 U/µl	0.5
10 X *PfuS* polymerase buffer	5.0
H_2O	41.5
Total	**49.0**
FW primer*	0.5
RV primer*	0.5

*The master mix was split in 2 x 24.5 µl and the primers were added up to final volume of 25 µl.

4.8.4. Generation of CelA2-multiple SSM mutagenesis libraries

See Material and Methods of chapter 4.8.3.

In the first round (CelA2-multiple SSM-G1$_{288}$) pIX3.0RMT7-CelA2-H288F was used as plasmid template DNA. For the second round of mutagenesis the entire isolated CelA2-multiple SSM G1$_{288}$ library was used as plasmid template DNA. For the subsequent rounds of mutagenesis always the plasmid DNA isolated from active cellulase variants from the previous round of flow cytometer-based screening was pooled and used as plasmid template DNA for the next round of Site-Saturation Mutagenesis.

4.8.5. Generation of CelA2-OmniChange mutagenesis libraries

The CelA2-OmniChange mutagenesis library was generated according to the published protocol (Dennig *et al.*, 2011). OmniChange fragments 1-3 were amplified by PCR 1 (98°C for 30 s, 1 cycle; 98°C for 15 s / 60°C for 15 s / 72°C for 1 min, 25 cycles; 72°C for 2 min, 1 cycle) and fragment 4 was amplified by PCR 2 (98°C for 30 s, 1 cycle; 98°C for 30 s / 60°C for 30 s /

72°C for 4 min, 25 cycles; 72°C for 5 min, 1 cycle) using dNTP mix (0.2 mM), primers (Table 22; 0.4 µM), plasmid DNA template (20 ng), *PfuS* polymerase (2 U) in a final volume of 50 µl. Fragment sizes of the four amplified fragments (Figure 19, appendix) are: 258 bp (fragment 1), 474 bp (fragment 2), 173 bp (fragment 3) and 3840 bp (fragment 4). Table 22 shows the different degenerated, phosphorothioated primers used for the amplification and cloning (*via* PLICing) of the different fragments 1-4 and pIX3.0RMT7-CelA2-H288F was used as plasmid template DNA. The amplified PCR products (fragments) were digested (1 h, 37°C) by *Dpn*I (20 U), purified using the NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG) and eluted in 15 µl ddH₂O. PLICing reaction was performed for cloning of the CelA2-OmniChange mutagenesis library (Dennig *et al.*, 2011) and generated constructs were transformed into competent *E.coli* XL10 Gold (Stratagene, La Jolla, CA, USA). Plasmids were isolated with Nucleo Spin® Plasmid Kit (Macherey-Nagel GmbH & Co. KG) and used for *in vitro* expression of the plasmid CelA2-OmniChange library.

Table 22: Degenerated primers for generation of the OmniChange library.

Primer	Sequence (5' → 3')	Size [nt]	Fragment Amplification
288_FW_MNN_Omni_PTO	gcaccgatgaccATTCTGCCGC	22	1
288_RV_MNN_Omni_PTO	ggtcatcggtgcMNNATCTTTTGCG	25	4
368_FW_MNN_Omni_PTO	gttagcgcaaccGATGAACGTTATTGG	27	2
368_RV_MNN_Omni_PTO	ggttgcgctaacATCATTMNNGGTGCCGG	29	1
524_FW_MNN_Omni_PTO	catcgtccgagcGCAGTTCTG	21	3
524_RV_MNN_Omni_PTO	gctcggacgatgMNNCGGATTCTG	24	2
576_FW_MNN_Omni_PTO	caatgaaatttgCATTTATTGGAATAGTCCGCTG	34	4
576_RV_MNN_Omni_PTO	caaatttcattgGTGGCMNNGCTACCGGTAAC	32	3

4.8.6. *Detection of cellulase activity in emulsions*

See Material and Methods of chapter 3.6.7.

4.8.7. *Compartmentalization in emulsions by extrusion*

See Material and Methods of chapter 3.6.10.

4.8.8. *Compartmentalization by extrusion and cell-free production of CelA2 from plasmid DNA template in emulsions*

See chapter 3.6.11.

4.8.9. Flow cytometry-based screening and sorting of CelA2-multiple SSM libraries

The (w/o/w) emulsions harboring the cell-free expressed CelA2-multiple SSM plasmid DNA library as template were diluted 1000-fold in PBS buffer (pH 7.4, 1.06 mM KH_2PO_4, 2.97 mM Na_2HPO_4, and 155.17 mM NaCl), filtered through 50 µm cell trics (Sysmex Partec GmbH, Görlitz, Germany) and analyzed with BD Influx cell sorter (Becton Dickinson Biosciences, Erembodegem, Belgium) according to forward and sideward scatter as well as to its fluorescence intensity (λ_{ex} 494 nm, λ_{em} 516 nm). The BD Influx flow cytometer was operated with a 100 µm nozzle and PBS (pH 7.4, 1.06 mM KH_2PO_4, 2.97 mM Na_2HPO_4, and 155.17 mM NaCl) was used as sheath fluid. Fluorescence intensities of (w/o/w) emulsions labelled with the converted product fluorescein were recorded, and sort gates were set to collect (w/o/w) emulsions with high green fluorescence intensity. Samples were analyzed with the flow cytometer and the active fraction was sorted (Sort mode: 1.0 Drop Pure). Reanalysis of the sorted populations with the flow cytometer showed up to 12-fold enrichment of the active (w/o/w) emulsion population. Sorted (w/o/w) emulsions were collected in PBS and plasmids were isolated by breakage of the (w/o/w) emulsions and purification of the recovered plasmid DNA.

4.8.10. Flow cytometry-based screening and sorting of CelA2-OmniChange library

The (w/o/w) emulsions harboring the cell-free expressed CelA2-OmniChange plasmid template library were diluted 1000-fold in PBS buffer (pH 7.4, 1.06 mM KH_2PO_4, 2.97 mM Na_2HPO_4, and 155.17 mM NaCl), filtered through 50 µm cell trics (Sysmex Partec GmbH, Görlitz, Germany) and analyzed with BD Influx cell sorter (Becton Dickinson Biosciences, Erembodegem, Belgium) according to forward and sideward scatter as well as to its fluorescence intensity (λ_{ex} 494 nm, λ_{em} 516 nm). The BD Influx flow cytometer was operated with a 100 µm nozzle and PBS (pH 7.4, 1.06 mM KH_2PO_4, 2.97 mM Na_2HPO_4, and 155.17 mM NaCl) was used as sheath fluid. Fluorescence intensities of (w/o/w) emulsions labelled with the converted product fluorescein were recorded, and sort gates were set to collect (w/o/w) emulsions with high green fluorescence intensity. Out of 36,757,972 analyzed events 395,229 were sorted (Sort mode: 1.0 Drop Pure) using 4,275 events s^{-1} analysis speed. Reanalysis of the sorted populations with FACS showed 30-fold enrichment of the active (w/o/w) emulsion population. Sorted (w/o/w) emulsions were collected in PBS and plasmids

were isolated by breakage of the (w/o/w) emulsions and purification of the recovered plasmid DNA.

4.8.11. *Recovery of DNA from (w/o/w) emulsions*

See Material and Methods of chapter 3.6.14.

4.8.12. *Cloning of enriched, active CelA2 variants into pIX3.0RMT7 in vitro expression vector*

Amplified, recovered CelA2-multiple SSM or OmniChange libraries (before and after) sorting (see Material and Methods of chapter 3.6.14) were cloned into pIX3.0RMT7 vector backbone by PLICing for expression and activity analysis in MTP (Blanusa et al 2010). For vector backbone amplification, the pIX3.0RMT7-CelA2 DNA template was amplified using standard PCR (98°C for 30 s, 1 cycle; 98°C for 30 s / 55°C for 30 s / 72°C for 2.5 min, 25 cycles; 72°C for 5 min, 1 cycle), containing *PfuS* DNA polymerase (2.5 U), dNTP mix (0.2 mM), forward PLICing primer (0.4 µM) (2_FW_Cel_PLIC: 5'-ACCGTGGCCGAATAATAAGAATTCGAGC -3'), reverse PLICing primer (0.4 µM) (2_RV_Cel_PLIC: 5'-TTCACCGCTGCTATGATGATGGTGG-3), pIX3.0RMT7-CelA2 DNA (20 ng) in a final volume of 25 µl.

The amplified PCR products were digested (1 h, 37°C) by *Dpn*I (20 U), analyzed for correct size by agarose-gel electrophoresis according to the standard protocol (Sambrook *et al.*, 2001), purified from agarose-gel using the NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG) and eluted in 20 µl ddH$_2$O. Purified PCR products were used for PLICing and resulting hybridization products were transformed into 100 µl chemically competent *E.coli* BL21 Gold(DE3)lacI^{Q1}-cells following the standard protocol (Hanahan et al 1991). Transformants were grown on LB$_{Amp}$ agar plates for further analysis in 96-well plate.

4.8.13. *Cultivation and expression of CelA2 and its variants in 96-well MTP*

Transformants were transferred *via* toothpick into 96-well flat bottom MTPs (Greiner Bio-one GmbH, Frickenhausen, Germany) containing 150 µL Luria Broth medium supplemented with ampicillin (50 µg/ml, LB$_{Amp}$) and cultivated (16 h, 37°C, 900 rpm, 70 % humidity). Plates were stored at -80°C after addition of 100 µl glycerol (50 % w/v) as master plates.

For preculture preparation a 96-well flat bottom MTP (Greiner Bio-one GmbH, Frickenhausen, Germany) containing 150 µL LB$_{Amp}$ was inoculated with 96-well replicator from master plate and incubated (16 h, 37°C, 900 rpm, 70 % humidity).

For main culture preparation a 96-well V-bottom expression MTP (Greiner Bio-one GmbH, Frickenhausen, Germany) containing 150 µL LB$_{Amp}$ medium was inoculated with 20 µl of the preculture. After cultivation (2 h, 37°C, 900 rpm, 70 % humidity) 15 µl IPTG (0.01 mM) per well was added for induction of protein expression, plates were incubated (4 h, 30°C, 900 rpm, 70 % humidity), centrifuged (4000 g, 10 min, 4°C) and pellets were stored at -20°C.

4.8.14. Cultivation, expression in flask and purification of CelA2 and its variants

For cultivation Luria Broth (LB) supplemented with ampicillin (100 µg/ml) and a shaking incubator (Multitron II; Infors GmbH, Einsbach, Germany) was used for incubation (20 h, 37°C, 900 rpm).

For *in vivo* cellulase expression the main expression culture was inoculated with 1 % of the preculture and incubated until OD$_{600}$ of 0.6 was reached (37°C, 250 rpm) in 100 ml Luria Broth (LB) supplemented with appropriate antibiotics. Cellulase expression was induced by addition of 100 µl IPTG (0.1 M). Cells were harvested after incubation (4 h, 30°C, 250 rpm) by centrifugation (4000 g, 10 min, 4°C) and pellets were stored at -20°C.

For cell disruption pellets were resuspended in Tris-HCl buffer (pH 7.4, 50 mM), sonicated 3 x 30 s and lysates were centrifuged (13,000 g, 15 min, 4°C). Supernatants were filtered 0.45 µm before His-tag purification with Protino® Ni-IDA 2000 Packed Columns (Macherey-Nagel GmbH & Co. KG) as described in the manufacturer's protocol. Purity was determined using Experion™ Automated Electrophoresis System (Bio-Rad Laboratories GmbH) and Pierce™ BCA Protein Assay Kit (Life Technologies GmbH).

4.8.15. Determination of kinetic parameters for CelA2 activity with 4-MUC and FDC activity assay

For determination of kinetic parameters of CelA2-H288F and its variants 4-MUC activity assay was performed according to the published protocol (Lehmann *et al.*, 2012).

For determination of kinetic parameters of CelA2-H288F and its variants FDC activity assay was performed in 96-well plate by mixing 40 µl KPI buffer, 40 µl cell lysate and 20 µl 0.05 mM FDC (Biomol GmbH) dissolved in KPI buffer (pH 7.2, 0.2 M). Fluorescence was

measured (λ_{ex} 494 nm/λ_{em} 516 nm; Z-position: 19500) in time intervals of 30 seconds over a total time of 15 minutes.

4.8.16. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

See Material and Methods of chapter 3.6.20.

4.8.17. Homology modeling of CelA2 cellulase

Since a crystal structure of CelA2 is currently not available, in order to gain first insights on the molecular level, homology modelling of the structure of CelA2 was performed by using YASARA Structure version 16.3.8 (Krieger *et al.*, 2002) with the default settings (PSI-BLAST iterations: 6 (Altschul *et al.*, 1997), E value cutoff: 0.5, templates: 5, OligoState: 4). A position-specific scoring matrix (PSSM) was used to score the obtained template structures (Jones, 1999; Qiu *et al.*, 2006). Four X-ray templates were selected for modelling the CelA2 structure (550 aa): a Cel9A in complex with cellotetraose-like isofagomine (530 residues with quality score 1.000, PDB ID: 3RX7), Endoglucanase D (541 residues with quality score 0.586, PDB ID: 1CLC), The cellobiohydrolase Cbha from *C. Thermocellum* (605 residues with quality score 0.551, PDB ID: 1UT9), and the cellobiohydrolase Cbha from *C. Thermocellum* (602 residues with quality score 0.413, PDB ID: 1RQ5). The best-scoring model (Z-score= -1.370) was a hybrid model build on X-ray structure 3RX7 and consisted of a monomer with bound cellotetraose-like isofagomine.

1.6.2. *Phytases*

Phytases, (EC-number: 3.1.3.8), also called *myo*-inositol hexakis phosphohydrolases, are mainly found in plants, fungi, microorganisms, and animals and belong to the enzyme group of phosphomonoester hydrolases, which catalyze the hydrolytic cleavage of phytates (*myo*-inositol-1,2,3,4,5,6 hexakis dihydrogen phosphate) into inositol, phosphates, divalent cations and proteins (Yao *et al.*, 2012), as shown in Figure 7.

Figure 7: The hydrolysis of phytate by phytase into inositol, phosphate and other divalent elements. The cleavage of phosphate groups by phytase results in the release of metals, metal-binding enzymes and proteins. Many phytases however cleave only five phosphates. The figure is taken from Yao *et al.* (Yao *et al.*, 2012).

Phosphorous and inositol are stored in plants or seeds (*i.e.* cereals, oil seeds or legumes) in form of phytate which cannot be hydrolyzed in the gut of monogastric animals, like dogs, birds and humans due to low level of phytase activity (Bitar *et al.*, 1972). This leads to phosphorous deficiency which requires supplementation of phosphorous in form of expensive feed additives (Konietzny, 2004). Additionally, phytates are strong chelating agents as they have the ability to fix divalent cations (*e.g.* Ca^{2+}, Zn^{2+}, Fe^{2+}) and proteins in stable complexes making them unavailable for the organism (Rao *et al.*, 2009). Phytases catalyze the step-wise hydrolysis of phosphomonoester bonds by release of inorganic phosphate, enabling monogastric animals to assimilate phosphorus. Therefore, phytases are attractive food supplements for the treatment and prevention of iron and mineral deficiency caused by phytate in humans and animals as they significantly improve the food and feed nutritional quality (Sandberg *et al.*, 1996). Beneficial uses of phytase comprise phosphate-mobilizing feed additives (*i.e.* swine, poultry or fish), development of transgenic organisms, and soil amendment for intensive crop production (Shivange *et al.*, 2012; Yao *et al.*, 2012). Furthermore, unabsorbed phosphates in form of phytate are usually excreted and released to the soil. The phytate is subsequently converted by soil-microorganisms which leads to phosphate run-off and environmental pollution (Konietzny, 2004). In this case,

5.2. Introduction

5.2.1. Phytases

Phytases (*myo*-inositol hexakis phosphohydrolase) are mainly found in plants, fungi, microorganisms, and animals and belong to the enzyme group of phosphomonoester hydrolases, which catalyze the hydrolytic cleavage of phytates (*myo*-inositol-1,2,3,4,5,6 hexakis dihydrogen phosphate) into inositol, phosphates, divalent cations and proteins (Yao *et al.*, 2012), as shown in Figure 91.
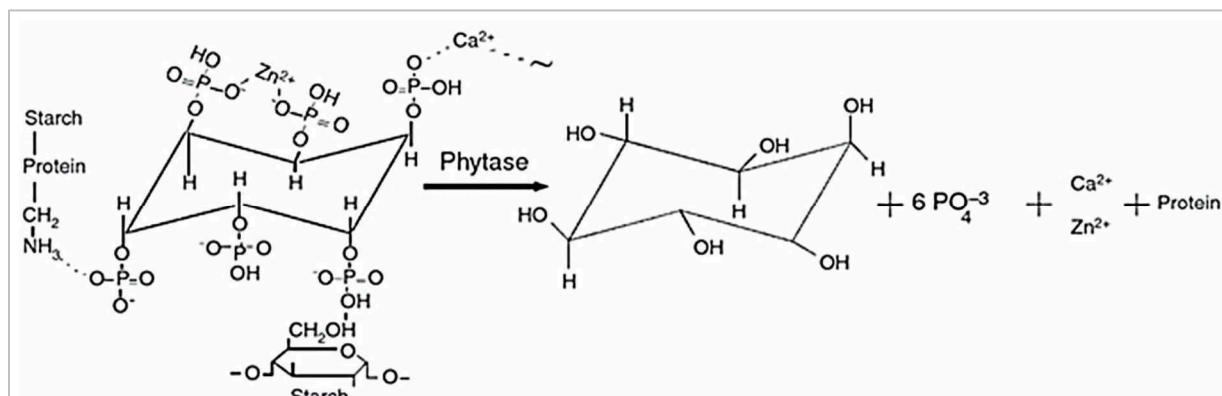
Figure 91: The hydrolysis of phytate by phytase into inositol, phosphate and other divalent elements. The cleavage of phosphate groups by phytase results in the release of metals, metal-binding enzymes and proteins. Some phytases however cleave only five phosphates. The figure is taken from Yao *et al.* (Yao *et al.*, 2012).

Phosphorous and inositol are stored in plants or seeds (*i.e.* cereals, oil seeds or legumes) in form of phytate which cannot be hydrolyzed in the gut of monogastric animals. This leads to phosphorous deficiency making supplementation of phosphorous in form of expensive feed additives necessary (Konietzny, 2004). Phytases catalyze the step-wise hydrolysis of phosphomonoester bonds by release of inorganic phosphate, enabling monogastric animals to assimilate phosphorus. Additionally, phytates are strong chelating agents as they have the ability to fix divalent cations (*e.g.* Ca^{2+}, Zn^{2+}, Fe^{2+}) and proteins in stable complexes making them unavailable for the organism (Rao *et al.*, 2009). Therefore, phytases are attractive food supplements for the treatment and prevention of iron and mineral deficiency caused by phytate in humans and animals as they significantly improve the food and feed nutritional quality (Sandberg *et al.*, 1996). Beneficial uses of phytase comprise phosphate-mobilizing feed additives (*i.e.* swine, poultry, fish), development of transgenic organisms, and soil amendment for intensive crop production (Yao *et al.*, 2012). Additionally, unabsorbed phosphates in form of phytate are usually excreted and released to the soil. The phytate is subsequently converted by soil-microorganisms which leads to phosphate run-off and environmental pollution (Konietzny, 2004). Phytases significantly decrease the

environmental pollution by 30-50 % caused by excretion of phosphorous in form of phytate by monogastric animals (Konietzny, 2004; Shivange *et al.*, 2012; Yao *et al.*, 2012). The usage of cost-intensive phosphorous fertilizers in agriculture can be decreased upon usage of phytases in transgenic plants or for soil amendment on the fields to increase the crop production by improved phosphate assimilation from phytate in plants and animals (Hong *et al.*, 2008). Furthermore, phytases can potentially produce versatile lower *myo*-inositol phosphates of pharmaceutical importance (Rao *et al.*, 2009).

Phytases have been grouped into three main classes (3-phytases; 5-phytases and 6-phytases) dependent on the carbon atom in the *myo*-inositol ring where the dephosphorization is initiated. Additionally, phytases can be divided into acid and alkaline phytases dependent on their pH optima. Furthermore, phytases can be grouped according to their catalytic mechanism into three classes (1) histidine acid phosphatase (HAPs), (2) ß-propeller phytase, and (3) cysteine phosphatase or purple acid phosphatase (Rao *et al.*, 2009; Tye *et al.*, 2002; Yao *et al.*, 2012). Many phytases isolated from fungi, bacteria and plants belong to the HAPs class of enzymes (Wyss *et al.*, 1999) and share the same conserved active site amino acid motif (RHGXRXP) and the catalytic dipeptide (HD) (Shivange *et al.*, 2012; Van Etten *et al.*, 1991). The HAPs class of enzymes catalyzes the phytate hydrolysis in a two-step process: a nucleophilic attack on the phosphorous atom by the histidine in the active site, followed by hydrolysis of the resulting phospho-histidine intermediate (Vincent *et al.*, 1992).

5.2.1.1. *Yersinia mollaretii* phytase

The *Yersinia mollaretii* phytase (YmPh) investigated in this work belongs to the HAP group of enzymes with an optimum acidic pH. YmPh expressed in *E. coli* has a tetrameric structure (tetramer, ~188 kDa, monomer ~47 kDa) and shows sigmoidal kinetics with positive cooperativity (Hill coefficient 2.3). The specific activity of YmPh towards the natural substrate phytate is 1,073 U/mg (determined in 0.25 sodium acetate, pH 4.5, 37 °C) which is around ten times higher than commonly used fungal phytases (Shivange *et al.*, 2012). The YmPh possesses a conserved α/β-domain and a variable α-domain, represented in Figure 92. The α/β-domain underneath has a central β-propeller with two α-Helixes on each side (Shivange *et al.*, 2012) and resembles the structure of many well-known phytases (Lim *et al.*, 2000). The α-domain of the YmPhytase is composed of two central α-Helixes and considered as part of the substrate binding region, which is surrounded by a series of helixes (Shivange

et al., 2012). The active site is located in the interface between both areas of the α-Helixes and is composed of four residues (R37, R41, E241, D327; Figure 92), which are involved in the substrate binding (Shivange *et al.*, 2010; Shivange *et al.*, 2012). The structure of the phytase has a substantial influence on its biophysical properties, such as the substrate binding (Rao *et al.*, 2009). Since the main industrial application of phytases is as feed additives, the elucidation of the behavior of the enzyme in contact to the gastric acids is of great interest. Therefore, high pH-stability under acidic conditions and a high catalytic efficiency are of significant relevance and make *Yersinia mollaretii* phytase an attractive target for investigation (Konietzny, 2004). In general, the scientific research mainly concentrates on the further optimization of the thermostability of phytases for the pelleting process for application in the food industry (Shivange *et al.*, 2012). A relevant property of this enzyme is that it exhibits low activity in neutral or alkaline environment (Lee *et al.*, 2007; Mullaney *et al.*, 2010). Therefore an adaption of phytases for specific uses, *i.e.* as food additive for fishes in aquacultures, towards improved activity around neutral pH is required (Cao *et al.*, 2007; Fu *et al.*, 2008; Wang *et al.*, 2011).
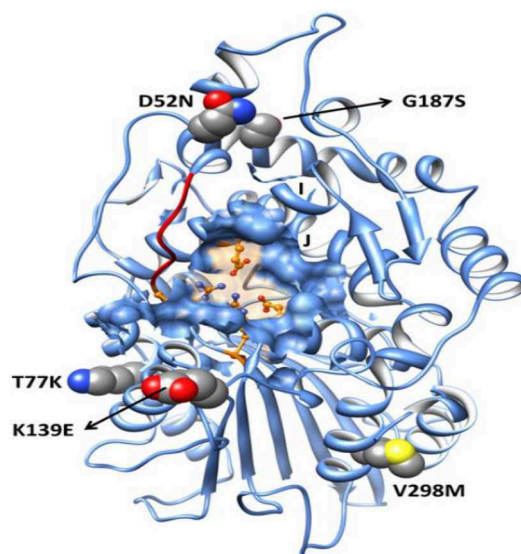
Figure 92: Homology model of the YmPh: The active site loop is highlighted in red. Important catalytic amino acids (R37, R41, E241, D327) involved in the binding of the substrate are highlighted in orange. The figure is taken from Shivange *et al.* (Shivange *et al.*, 2012).

5.2.2. Screening systems

The success of a directed enzyme evolution campaign is to a great extend dependent on selection of a suitable screening system based on monitoring enzyme activity at conditions as close as possible to the final application. Enzyme assays mediate the measurement of enzymatic activities (Arnold *et al.*, 1997; Reymond *et al.*, 2008), by either measuring the

consumption of the substrate or the release of the product in function of time (Goddard *et al.*, 2004). The assay selection is depending on the properties of the enzyme that will be tested (Reymond *et al.*, 2008). When the enzymatic reaction is accompanied by a change in the color, colorimetric assays are appropriate, while spectrophotometric assays detect changes in the light absorption (Reymond *et al.*, 2008; Sapan *et al.*, 1999). A further method to detect enzyme activity is based on monitoring increase of fluorescence due to conversion of a fluorogenic substrate by the enzyme of interest in a time dependent manner (Reymond *et al.*, 2008; Shivange *et al.*, 2012). The key advantage of the fluorescence-based activity assay is its significantly higher sensitivity compared to spectrophotometric assays (Passonneau *et al.*, 1993). On the other hand, the main drawback of the fluorescence-based assays is often high interference by impurities on the fluorescence obtained by target enzyme conversion and fluorogenic substrate instability (Shivange *et al.*, 2012).

5.3. Results

The result section is divided into three parts in order to perform successful directed phytase evolution for improved activity at neutral pH. In the first section optimization of 4-MUP activity assay for screening at neutral pH is performed (5.3.1). In detail, optimization steps include: a) selection of a suitable assay buffer (5.3.1.1), b) optimization of substrate and enzyme concentration (5.3.1.2) and c) selection of a suitable pH for screening of improved YmPh variants (5.3.1.3). In the second result section directed phytase evolution is performed using YmPh-epPCR library generation and screening (5.3.2-5.3.4), followed by YmPh-SSM library generation and screening, based on computational analysis of improved phytase variants identified in the YmPh-epPCR library (5.3.5-5.3.7). Finally, in the third result section pH profiles of YmPh variants with increased activity at pH 6.6 are determined (5.3.8) and the two best improved YmPh variants were purified and kinetically characterized (5.3.9-5.3.10). The identified beneficial positions of the amino acid exchanges in the improved variants are analyzed by homology model generation using YASARA software (5.3.11).

Parts of this work were previously published in the master thesis of Janina Kern *"Entwicklung einer Yersinia mollaretii Phytase mit verbesserter Aktivität bei höherem pH-Wert"* (2013).

5.3.1. Optimization of 4-MUP activity assay for screening at neutral pH

5.3.1.1. Comparison of phytase activity in acetate and Tris maleate buffer

The fluorescence-based 4-MUP activity assay was recently reported by Shivange *et al.* (Shivange *et al.*, 2012) for screening of YmPh variants with improved activity at pH 5.5. The assay is based on the conversion of the fluorogenic substrate 4-methylumbelliferylphosphat (4-MUP) into the fluorescent 4-methylumbelliferone (4-MU) product and phosphate (Figure 93) in acetate buffer (pH 5.5). The measurements were performed in 384- and 96-well MTP-based format at λ_{ex} 360 nm and λ_{em} 465 nm.

In this work, the established 4-MUP activity assay was optimized for screening of an YmPh mutant library towards improved activity at neutral pH (~7.0). Therefore, the acetate buffer having a pH range from 3.6 to 5.6 was replaced by Tris maleate buffer with a pH range from 5.2 to 8.6. Optimization of the activity assay comprised buffer selection, substrate and enzyme concentration and selection of the pH for screening of improved phytase variants with activity at neutral pH.
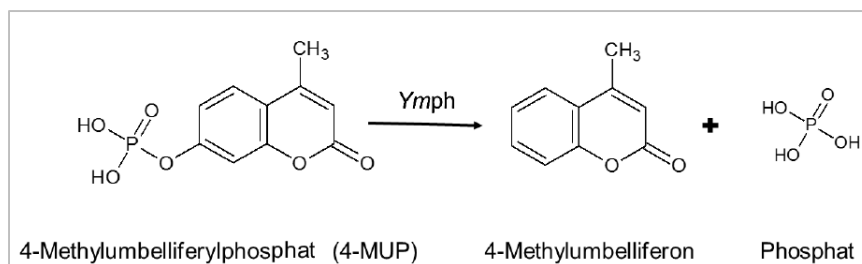
Figure 93: Conversion of 4-methylumferylphosphat into 4-methylumbelliferone and the release of phosphate.

Therefore, the reported 4-MUP activity assay was performed in the standard acetate buffer and in Tris maleate buffer at pH 5.5 with cell lysates from *E.coli* BL21 Gold(DE3)lacI^{Q1}-pALXtreme-5b-YmPhWT and -empty vector (EV) after expression in 96-well MTP to compare the phytase performance in the different buffers. The activities were determined as slope in RFU/s over time. Mean values and standard deviations were calculated from twelve measurements. As illustrated in Figure 94, the activity of YmPhWT in Tris maleate buffer was 3-fold lower than in the standard acetate buffer at pH 5.5.
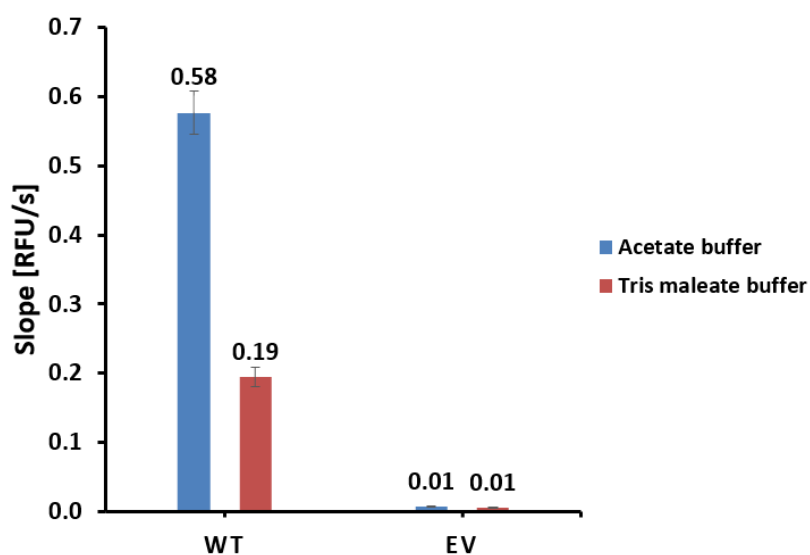
Figure 94: Activity of *E.coli* BL21 Gold(DE3)lacI^{Q1}-pALXtreme-5b-YmPhWT and -EV cell lysates in Tris maleate and acetate buffer (pH 5.5; 1 mM 4-MUP) after expression in 96-well MTP. The activities were determined as slope in RFU/s over time. Mean values and standard deviations were calculated from 12 measurements.

5.3.1.2. Optimization of substrate and enzyme concentration

For further optimization of the 4-MUP activity assay in Tris maleate buffer the amount of cell lysate (10-20 µl) and the substrate concentration (1-5 mM 4-MUP) were varied to increase the fluorescence signal of the YmPhWT cell lysate in the assay. An increase in activity of 55 % was observed for 20 µl of cell lysate applied in the 4-MUP activity assay in comparison to 10 µl of cell lysate. The activities were determined as slope in RFU/s over time. Mean values and standard deviations were calculated from ten measurements (Figure 95).
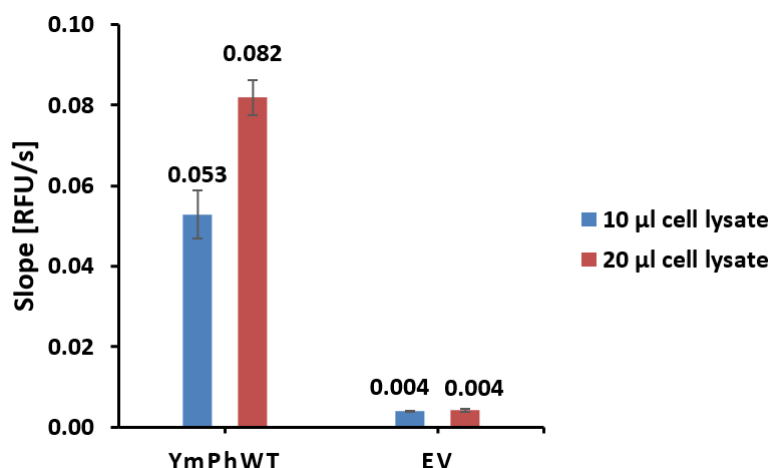
Figure 95: Activity of *E.coli* BL21 Gold(DE3)lacI^{Q1}-pALXtreme-5b-YmPhWT and -EV in Tris maleate buffer at pH 5.5 after expression in 96-well MTP using different amounts of cell lysate (10-20 µl, 1 mM 4-MUP). The activities were determined as slope in RFU/s over time. Mean values and standard deviations were calculated from ten measurements.

Additionally, the substrate concentration was varied from 1 mM to 5 mM 4-MUP (in Tris maleate buffer, pH 5.5; 20 µl cell lysate) to determine the optimal substrate concentration for detection of phytase activity. The activities were determined as slope in RFU/s over time. Mean values and standard deviations were calculated from six measurements. The highest fluorescence signal was observed for 5 mM 4-MUP (Figure 96) which was chosen as optimal substrate concentration.
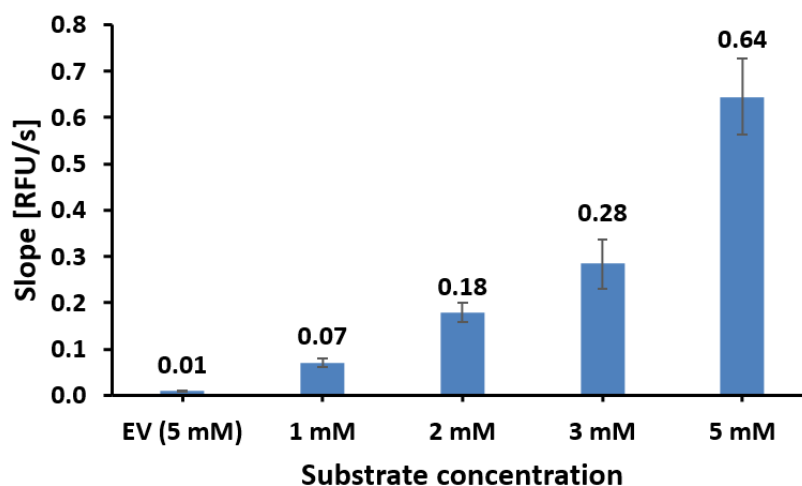
Figure 96: Activity of *E.coli* BL21 Gold(DE3)lacI^{Q1}-pALXtreme-5b-YmPhWT and -EV after expression in 96-well MTP in Tris maleate buffer (pH 5.5, 20 µl cell lysate) using different substrate concentrations (1-5 mM 4-MUP). The activities were determined as slope in RFU/s over time. Mean values and standard deviations were calculated from six measurements.

In summary, optimal assay conditions for monitoring YmPh activity using Tris maleate buffer at pH 5.5 were 20 µl *E.coli* cell lysate containing YmPh and 5 mM MUP substrate.

treatment of agricultural soils with phytases can significantly decrease the environmental pollution by 30-50 % caused by excretion of phosphorous in form of phytate by monogastric animals (Konietzny, 2004; Shivange *et al.*, 2012; Yao *et al.*, 2012). In addition, the usage of cost-intensive phosphorous fertilizers in agriculture can be decreased upon usage of phytases in transgenic plants or for soil amendment on the fields to increase the crop production by improved phosphate assimilation from phytate in plants and animals (Hong *et al.*, 2008). Phytases can either be divided into acid and alkaline phytases dependent on their pH optima or they can be grouped according to their catalytic mechanism into three classes (1) histidine acid phosphatase (HAPs), (2) ß-propeller phytase, and (3) cysteine phytase or purple acid phosphatase (Rao *et al.*, 2009; Tye *et al.*, 2002; Yao *et al.*, 2012). Many phytases isolated from fungi, bacteria and plants belong to the HAPs class of enzymes (Wyss *et al.*, 1999) and share the same conserved active site amino acid motif (RHGXRXP) and the catalytic dipeptide (HD) (Shivange *et al.*, 2012; Van Etten *et al.*, 1991). The enzyme of choice for this project is a highly active tetrameric phytase, which belongs to the group of acidic phosphatases, with a size of 43 kDa (188 kDa in native page electrophoresis) and a specific activity of 1073 U/mg which was isolated from *Yersinia mollaretii* (Ym-phytase) and cloned into *E.coli* BL21-Gold(DE3)lacI^{Q1} strain (Shivange *et al.*, 2012).

1.6.3. *Fluorescence-based assays for detection of phytase and cellulase activity*

As the proposed ultrahigh throughput screening (uHTS) system is based on flow cytometry, for development of this screening platform it is necessary to find suitable fluorescent assays to detect phytase and cellulase activity with the available BD Influx flow-cytometer-based sorting device in our laboratory. This flow cytometer, in our case, includes three excitation lasers, a UV laser for excitation at 355 nm, a blue laser for excitation at 488 nm and a yellow-green laser for excitation at 651 nm, as well as emission detection filters for 450/50 nm, 530/40 nm, 585/29 nm, 612/20 nm, 670/30 nm, 673/30 nm, 710/50 nm, 750/LP.

Several available fluorescent assays, *e.g.* 4-MUC activity (Lehmann *et al.*, 2012) assay or 4-MUP activity assay (Shivange *et al.*, 2012), adapted for MTP screening have to be tested and optimized for *in vitro* compartmentalization systems. The main requirements for fluorescent assays employed in a flow cytometer screening platform are: (1) dye diffusion (*e.g.* fluorescent substrate/product of the enzymatic reaction should stay inside the inner water phase of the emulsion droplet), which can be prevented by charged moieties into

Figure 98: Activity of *E.coli* BL21 Gold(DE3)lacl[Q1]-pALXtreme-5b-YmPhWT after expression in 96-well MTP measured in 384-well MTP with optimized conditions (5 µl Tris maleate buffer pH 6.6, 20 µl cell lysate, 25 µl 5 mM 4-MUP in Tris maleate buffer pH 6.6). The activities were determined as slope in RFU/s over time. Mean values and standard deviations were calculated from 160 measurements.

5.3.2. Generation of the epPCR library

The generation of a random YmPh mutant library was performed by epPCR using different concentrations of $MnCl_2$ (0-0.4 mM). The generated PCR products (expected size of 1.3 kb for the phytase gene) were analyzed by agarose-gelelectrophoresis and are illustrated in Figure 99. The band intensity of the PCR products decreases with increasing $MnCl_2$ concentration during PCR amplification (Figure 99). The negative control (*Dpn*I religation control) of the PCR amplification of the *ymph* insert, as well as the negative control for the vector amplification (NV) without primers shows no band. For the amplification of the pALXtreme-5b vector backbone (V1, V2) bands can be observed at the expected size of 2 kb (Figure 99).
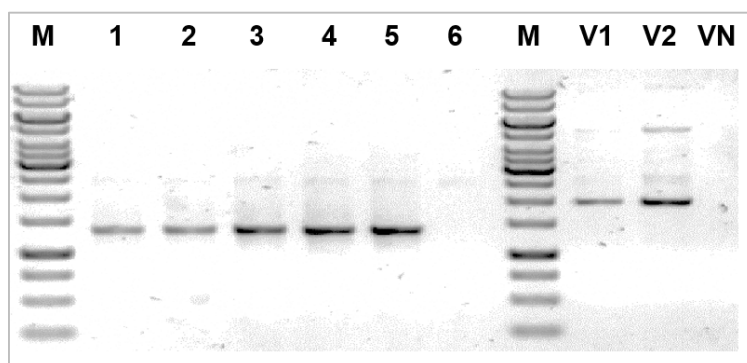
Figure 99: PCR products of insert (*ymph*) amplification using 0-0.4 mM $MnCl_2$ and vector (pALXtreme-5b) amplification. M: Marker, 1-6: epPCR with *ymph*, 1: 0.4 mM $MnCl_2$, 2: 0.2 mM $MnCl_2$, 3: 0.1 mM $MnCl_2$, 4: 0.05 mM $MnCl_2$, 5: 0.0 mM $MnCl_2$ (positive control), 6: 0.2 mM $MnCl_2$ without primers (negative control), V1/V2: Vector, VN: Vector without primer (negative control).

For the screening of mutant libraries the optimal balance between the number of active clones and the mutational load (*i.e.* MnCl$_2$ concentration) has to be chosen. In order to achieve this goal, a library should contain ~50 % of active clones to minimize screening efforts in MTP and to maximize, at the same time, the possibility to find improved variants which is highly dependent on the mutational load applied. Therefore, epPCR of the *ymph* gene was performed using different concentrations of MnCl$_2$ (0-0.4 mM), the phytase gene libraries were inserted into the vector backbone (pALXtreme-5b) *via* pLICing and subsequently transformed into chemically competent *E.coli* BL21 Gold(DE3)lacI^{Q1} cells (see Material and Methods, chapter 5.5.2.1). For the screening the epPCR-YmPh library generated with 0.1 mM MnCl$_2$ was selected, as it showed the best ratio of active (60 %) to inactive (40 %) phytase clones and with 2216 colonies growing on the agar plates a suitable number of transformants for further analysis (Table 23).

Table 23: Total number of colonies counted on the agar plates and percentage of active of pALXtreme-5b-YmPh clones generated using epPCR for *ymph* amplification with different amounts of MnCl$_2$, after transformation in chemically competent *E.coli* BL21 Gold(DE3)lacI^{Q1}.

MnCl$_2$ concentration in [mM]	Number of colonies	Number of active clones in [%]
Religation (control)	25	0
0.4	110	45
0.2	270	35
0.1	**2216**	**60**
0.05	3572	70
0	1232	76

Table 23 shows the total number of transformants and the percentage of active clones in dependence on the MnCl$_2$ concentration used during epPCR-YmPh library generation. The number of colonies as well as the percentage of active YmPh clones decreases with increasing MnCl$_2$ concentration. For determination of the percentage of active clones for each MnCl$_2$ concentration 90 clones were expressed in 96-well MTP and analyzed with optimized 4-MUP activity assay using Tris maleate buffer pH 5.5 and compared to the YmPhWT. Variants were regarded as active if they showed at least 20 % of the YmPhWT activity (YmPhWT activity was set to 100 %). The YmPh-epPCR library generated with 0.1 mM MnCl$_2$ was selected for further screening as 60 % of its clones are active (Table 23).

5.3.3. Screening of YmPh-epPCR library

Screening of 3600 clones of the YmPh-epPCR library (0.1 mM MnCl$_2$) was performed with optimized 4-MUP activity assay at pH 5.6 and 6.6 in 384-well plates (Figure 100) after expression in 96-well plate. *E. coli* BL21 Gold(DE3)lacI^{Q1}-pALXtreme-5b-YmPhWT (WT) served as positive control and *E. coli* BL21 Gold(DE3)lacI^{Q1}-pALXtreme-5b (EV) as negative control in each MTP (in triplicate). The activity was determined by measuring the increase in fluorescence in time (RFU/s). Variants showing significantly higher slopes (*i.e.* activity) in comparison to the YmPhWT at pH 6.6, were selected for rescreening. Therefore, activity ratios were determined by computing the ratio of the slope of the variant and the slope of YmPhWT (Figure 100).
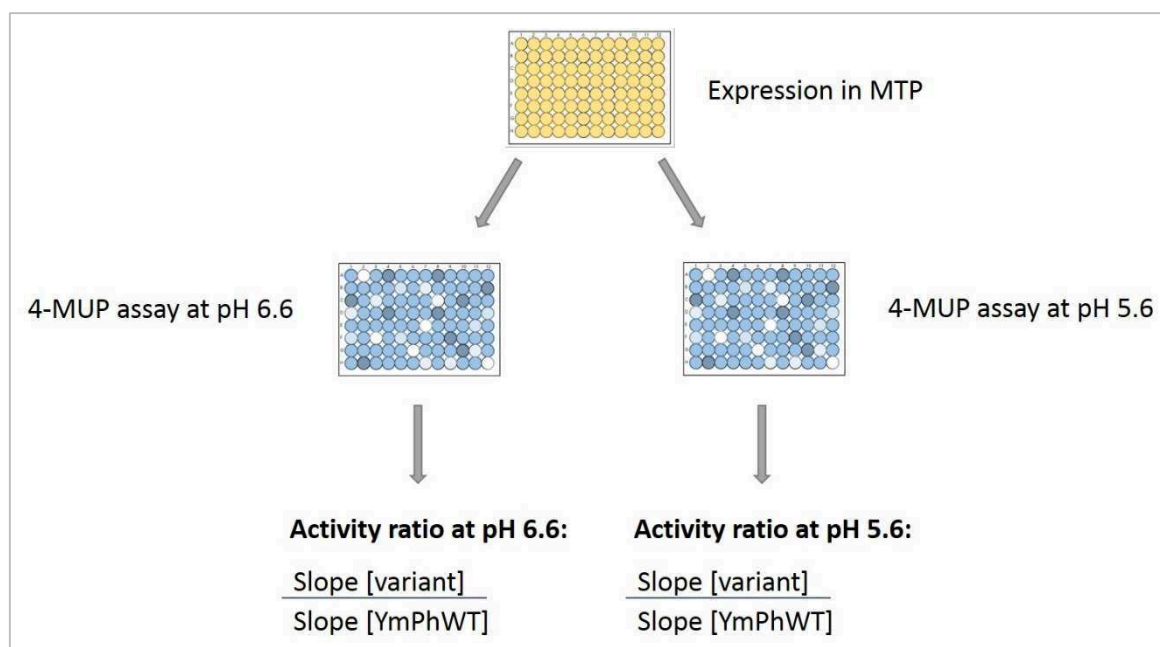
Figure 100: Screening strategy for selection of improved variants for rescreening from the YmPh-epPCR library.

Figure 101 shows exemplary the result of the 4-MUP activity test of plate 23 of the YmPh-epPCR library analyzed at pH 6.6, presenting on the y-axis of the graph the obtained activity ratios in [%] for each variant, YmPhWT and EV. Variant YmPh-M2, highlighted in orange, has a 2.6-fold increased activity ratio at pH 6.6 compared to YmPhWT (268 %; Figure 101), and a 1.3-fold increased activity ratio at pH 5.6 (126 %) compared to YmPhWT (Table 24), so this variant was selected for rescreening. As the activity profile towards pH 6.6 is broadened (and not completely shifted from pH 5.6 to 6.6), while keeping a high activity at pH 5.6, the activity ratio of pH (6.6/5.6) is only slightly increased for variant (2-fold) compared to YmPhWT (Table24).
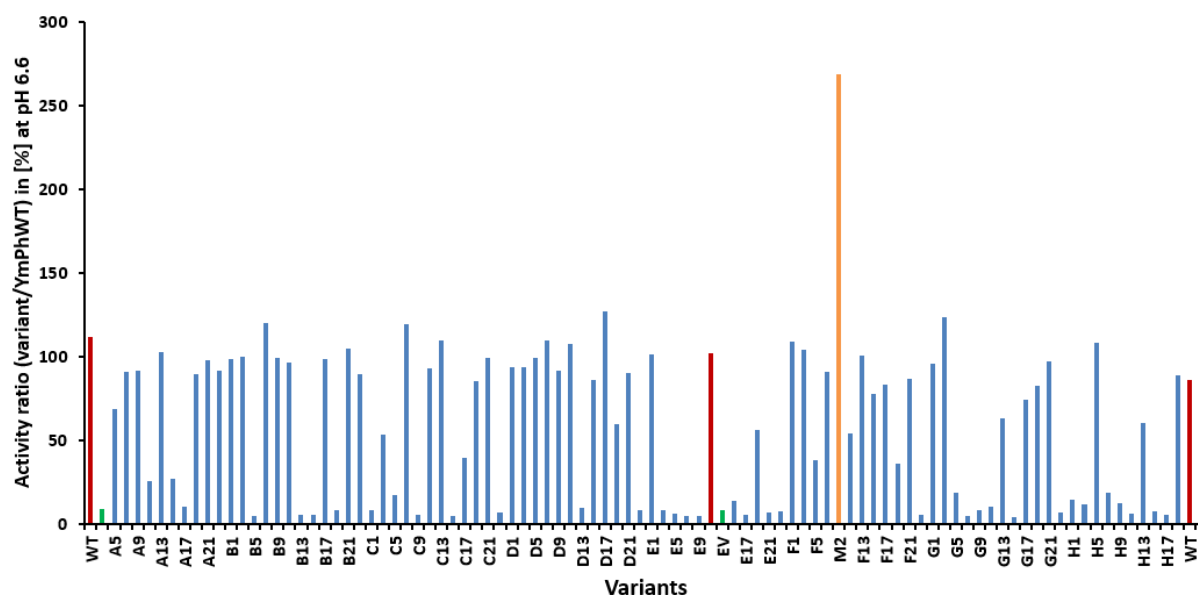
Figure 101: Screening result of YmPh-epPCR library plate 23 analyzed with 4-MUP activity assay at pH 6.6. The YmPhWT is highlighted in red, the EV in green and the variant YmPh-M2, showing improved activity ratio in comparison to the YmPhWT, is highlighted in orange. The y-axis shows the activity ratio of the slope [variant/YmPhWT] at pH 6.6. Slopes were measured as fluorescence over time in RFU/s. The x-axis illustrates the screened variants.

Table 24: Result of the analysis of EV, YmPhWT and variant YmPh-M2 from MTP 23 of the YmPh-epPCR library with 4-MUP activity assay using Tris maleate buffer at pH 5.6 and 6.6. For EV and WT the mean values for activity (slope) were calculated from three measurements.

epPCR-library (MTP 23)	pH 5.6		pH 6.6		Ratio slope (pH 6.6/pH 5.6)
	Slope in [RFU/s]	Activity ratio (variant/YmPhWT) in [%]	Slope in [RFU/s]	Activity ratio (variant/YmPhWT) in [%]	
EV (mean)	0.02	2.45	0.02	8.45	0.92
YmPhWT (mean)	0.88	100.00	0.24	100.00	0.27
YmPh-M2	1.12	126.32	0.63	268.31	0.57

5.3.4. *Rescreening of best variants of YmPh-epPCR library and sequence analysis*

In total 24 variants from the YmPh-epPCR library showed significantly higher activity than YmPhWT at pH 6.6 and the eight most beneficial variants were selected for rescreening. Table 25 and Figures 102-103 summarize the result of the analysis with 4-MUP activity assay at pH 5.6 and 6.6 of the eight best variants (M2-M9) of the YmPh-epPCR rescreening. Mean values were calculated from five measurements (EV and YmPhWT four measurements). The rescreening results confirm that all nine variants show improved activity at pH 6.6 in comparison to the YmPhWT (Table 25, Figures 102, 103). At pH 5.6 YmPh-M1 shows lower activity than the YmPhWT and all other variants YmPh-M2-M9 show elevated activity levels, indicating the broadening of the pH spectrum to more neutral pH while keeping a good performance at acidic pH. Variants M2-M9 show up to 3-fold increased activity in

substrate/product molecule. Keeping the fluorescent product in the droplet keeps the connection between gene and product, (2) substrate interference with the *in vitro* expression mixture must be minimal to avoid inhibition of *in vitro* expression by the substrate or product, (3) sensitivity of the assay (detectable concentration should be very low, as only small amounts of proteins will be produced within the emulsions) and (4) wide applicability (principle of the system should be applied for screening different classes of enzymes). Two most common fluorescent substrates employed for screening in the directed evolution experiment are based on coumarin and fluorescein derivatives offering a wide substrate spectrum for different classes of enzymes. For the detection of phytase activity the 4-methylumbelliferyl phosphate (Shivange *et al.*, 2012) and for cellulase the 4-methylumbelliferyl-β-D-cellobioside (Lehmann *et al.*, 2012) are currently employed in microtiter plate screening formats.

1.6.3.1. Fluorescence-based assays for detection of phytase activity

For fluorescence-based microtiterplate and flow cytometric assays, the 4-MUP activity assay, established by Shivange *et al.* will be used for monitoring Ym-phytase activity (Shivange *et al.*, 2012). The non-fluorescent substrate 4-methylumbelliferyl phosphate (4-MUP) is converted by the enzyme to the blue fluorescent product 4-methylumbelliferone. Formation of the product can be detected by fluorescence measurement at an excitation of 360 nm and an emission of 465 nm, after cleavage of the phosphate group from the substrate, as shown in Figure 8.
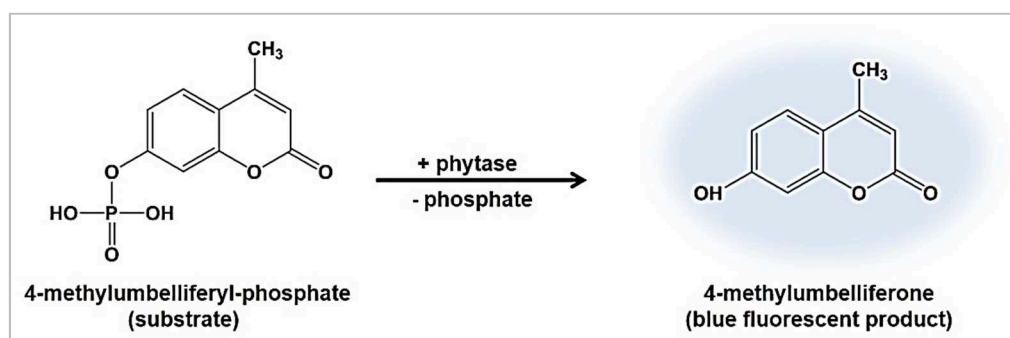
Figure 8: Conversion of 4-methylumbelliferyl-phosphate (4-MUP) substrate into the fluorescent product 4-methylumbelliferone (4-MU) by phytase.

Table 25: Result of the sequencing analysis of the eight YmPh-epPCR variants (M2-M9) from the rescreening.

Variant	Amino acid substitutions
WT	-
M2	Y27C/K45E/K306E
M3	L389S/C432G
M4	I15T/R162G
M5	I15V/P62S
M6	M1T/H146R
M7	T44M
M8	I15T
M9	I15Y/P94S

5.3.5. *YmPh-SSM library generation based on computational analysis*

A phytase homology model was generated using YASARA software (version 16.3.5, www.yasara.org) and obtained amino acid substitutions for activity improved YmPh variants at pH 6.6 were shown (Figure 104). The positions T44 and K45 having amino acid exchanges in improved YmPh variants were selected for simultaneous Site-Saturation Mutagenesis (SSM), as they are in close proximity to the active center of the enzyme and to the active site loop.

The other amino acid exchanges are not investigated in this study, as they are located far from the active site of the enzyme which makes an interaction with the substrate unlikely (Figure 104).
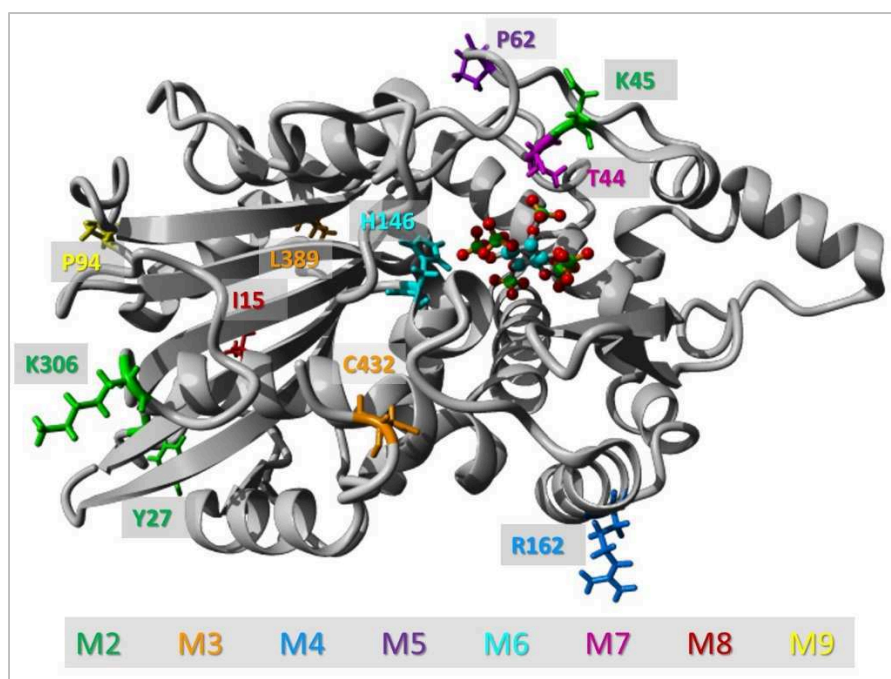
Figure 104: Hybrid homology model of YmPhWT generated with YASARA (Version 16.3.5, www.yasara.org), illustrating beneficial positions which were identified in epPCR-YmPh variants M2-M9. The substituted positions of variant M2 are highlighted in green, of variant M3 in orange, of variant M4 in blue, of variant M5 in purple, of variant M6 in cyan, of variant M7 in magenta, of variant M8 in red and of variant M9 in yellow. The substrate phytate is shown in ball-stick representation.

5.3.6. Screening of YmPh-SSM library

An YmPh-SSM library was generated by simultaneous SSM at positions T44 and K45. The library was analyzed with optimized 4-MUP activity assay at pH 5.6 and 6.6 in Tris maleate buffer in 384-well plates, to identify variants that show higher activity than YmPhWT at pH 6.6. *E. coli* BL21 Gold(DE3)lacI^{Q1}-pALXtreme-5b-YmPhWT (WT) served as positive control and *E. coli* BL21 Gold(DE3)lacI^{Q1}-pALXtreme-5b (EV) as negative control in each MTP (in triplicate). The activity was determined by monitoring a fluorescence increase over time (slope [RFU/s]). Variants showing significantly increased slopes in comparison to the YmPhWT at pH 6.6, were selected for rescreening.

For the YmPh-SSM library at simultaneously saturated positions 44/45 in total 900 clones were screened and 56 variants showed significantly improved activity ratios at pH 6.6 in comparison to YmPhWT. Figures 105 and 106 show exemplary the results of the 4-MUP activity test of plate 5 of the YmPh-SSM-44/45 library, analyzed at pH 5.6 and 6.6. Variants showing the highest activity ratios at pH 6.6 compared to YmPhWT are highlighted in orange and were selected for rescreening (Figure 106). Figure 105 shows, the activity ratios of the different variants for pH 5.6, *e.g.* variant E1 with an activity ratio of 94 %, is comparable to YmPhWT (100 %), highlighted in red. For pH 6.6 these variants show up to 5.8-fold improved

activity ratios, *e.g.* variant E1 with an activity ratio of 582 %, compared to YmPhWT (Figure 106). Despite that, some variants (*e.g.* C23) show significantly lower activity ratios at pH 5.6 (61.88 %), but higher activity ratios at pH 6.6 (411 %) compared to YmPhWT (Figures 105, 106).
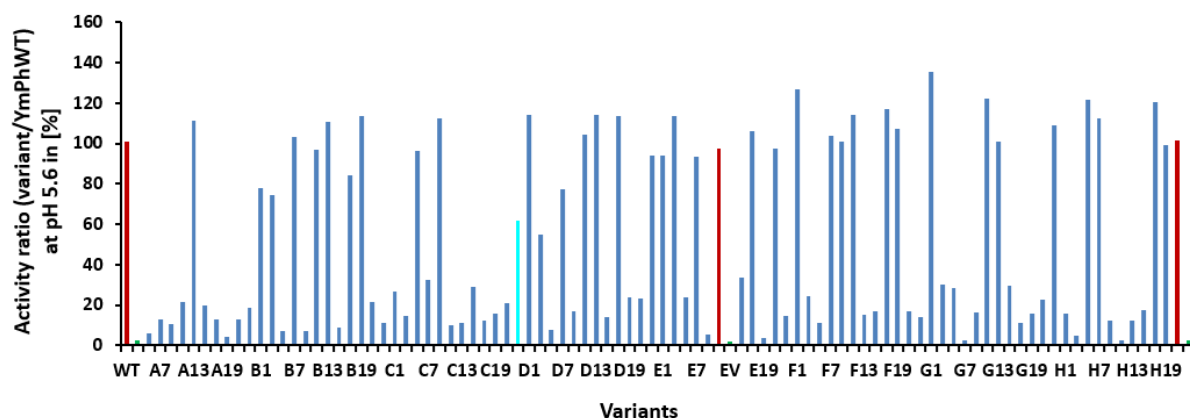
Figure 105: Screening result of YmPh-SSM-44/45 library plate 5 analyzed with 4-MUP activity assay at pH 5.6. The YmPhWT is highlighted in red and the EV in green. Variant C23 is highlighted in cyan. The y-axis shows the activity ratio in [%] and the x-axis illustrates the screened variants. The standard deviation of the 4-MUP activity measurement at pH 5.6 was 2.19 %.
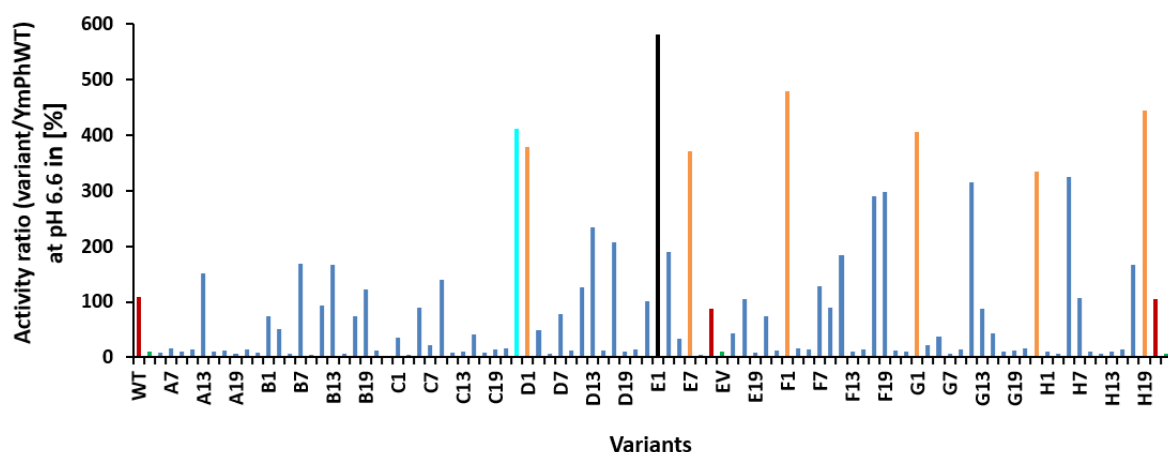
Figure 106: Screening result of YmPh-SSM-44/45 library plate 5 analyzed with 4-MUP activity assay at pH 6.6. The YmPhWT is highlighted in red, the EV in green and the variants selected for the rescreening, showing improved activity in comparison to the YmPhWT, are highlighted in orange. Variant C23 is highlighted in cyan and variant E1 is highlighted in black. The y-axis shows the activity ratio in [%] and the x-axis illustrates the screened variants. The standard deviation of the 4-MUP activity measurement at pH 6.6 was 11.17 %.

5.3.7. Rescreening of the best variants of YmPh-SSM library and sequence analysis

MTP analysis revealed 56 variants showing significantly increased activity at pH 6.6 in comparison to YmPhWT, while keeping the activity at pH 5.6 comparable to YmPhWT. The eight most beneficial variants (M10-M17) were selected for rescreening and were analyzed with 4-MUP activity assay at pH 5.6 and 6.6. Table 26 and Figure 107 summarize the result of the eight best variants (M10-M17) of the YmPh-SSM library rescreening.
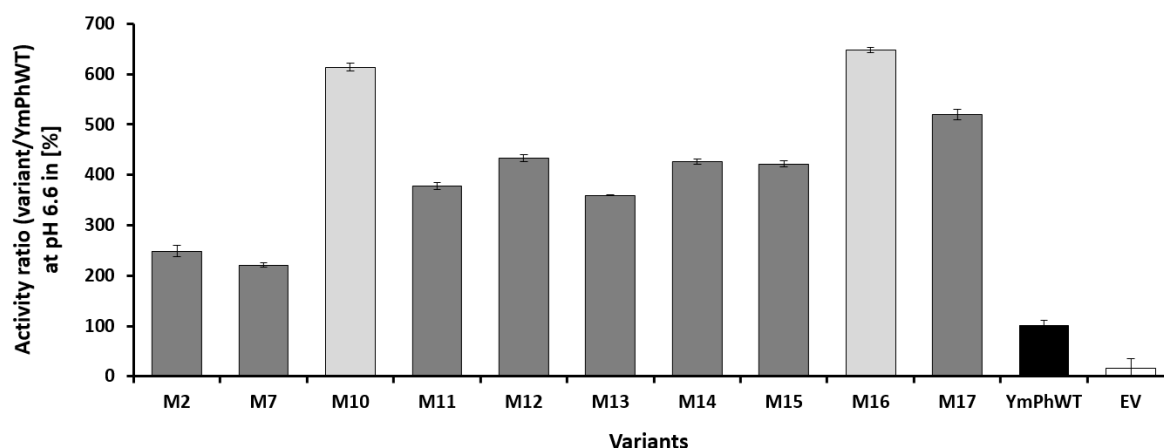
Figure 107: Results of the YmPh-SSM library rescreening performed with selected variants YmPh-M2, -M7, -M10-M17 after analysis with 4-MUP activity assay at pH 6.6. The y-axis shows the mean values of the obtained activity ratios at pH 6.6 in [%] which were calculated from five measurements (EV and YmPhWT four measurements). The x-axis shows the different variants in comparison to the YmPhWT and EV.

The variants were analyzed with optimized 4-MUP activity assay at pH 5.6 and 6.6 in Tris maleate buffer after expression in 96-well plates. *E. coli* BL21 Gold(DE3)lacI^{Q1}-pALXtreme-5b-YmPhWT, as well as -YmPh-M2, -M7 and -M9 (identified in the YmPh-epPCR library) served as positive controls and *E. coli* BL21 Gold(DE3)lacI^{Q1}-pALXtreme-5b (EV) as negative control in each MTP. Variants YmPh-M2 and –M7 were chosen for comparison as they are parents for position 44/45 in the YmPh-SSM library. The activity of the variants was determined by fluorescence measurements over time in RFU/s resulting in the slope. Mean values were calculated from five measurements (EV and YmPhWT four measurements). Activity ratios were determined which are illustrated in Figure 107 to compare the activity of the rescreened variants to YmPhWT.

The rescreening results confirm that all of the selected YmPh-SSM-44/45 variants M10-M17 show improved activity at pH 6.6 in comparison to YmPhWT and compared to the YmPh-epPCR variants YmPh-M2 and -M7 (Figure 107). Variants YmPh-M10 and -M16 show 6-fold improved activity in comparison to YmPhWT at pH 6.6 (Figure 107) and more than 5-fold increased activity in comparison to YmPhWT at pH 5.6 (Figure A20, appendix). The eight best YmPh-SSM-44/45 variants (M10-M17) compared to YmPhWT were subsequently sent for sequence analysis (Table 26).

Table 26: Result of the sequence analysis and the rescreening of YmPh-SSM-44/45 variants using 4-MUP activity assay using Tris maleate buffer at pH 5.6 and 6.6. For EV and WT the mean values were calculated from four measurements. For all other variants mean values were calculated from five measurements. The variants showing the highest improvement for 4-MUP substrate in comparison to WT are bolded. Variants M2 and M7 from the YmPh-epPCR library serve as additional reference.

Variant	Amino acid substitutions
WT	-
M2	Y27C/K45E/K306E
M7	T44M
M10	**T44I/K45E**
M11	T44V/K45W
M12	T44R/K45M
M13	T44V/K45S
M14	T44V/K45A
M15	T44V/K45D
M16	**T44V/K45E**
M17	T44V/K45Y

The sequencing results of the nine investigated variants are summarized in Table 26. All variants show point mutations at the targeted positions. The best variants of the YmPh-SSM rescreening were YmPh-M10 (T44I/K45E) and M16 (T44V/K45E), showing 6-fold improved activity at pH 6.6 compared to YmPhWT and having increased activity at pH 5.6 (Figure A20, appendix).

Additionally, SDS-PAGE was performed with the variants YmPh-M10, -M12, -M14 to -M17, which showed the highest activity at pH 6.6 in comparison to YmPhWT, in order to exclude expression mutants (Figure 108). As shown in Figure 108, the band for the phytase around 47 kDa has a comparable intensity for all variants and the YmPhWT, indicating similar protein expression levels, so expression mutants can be excluded.
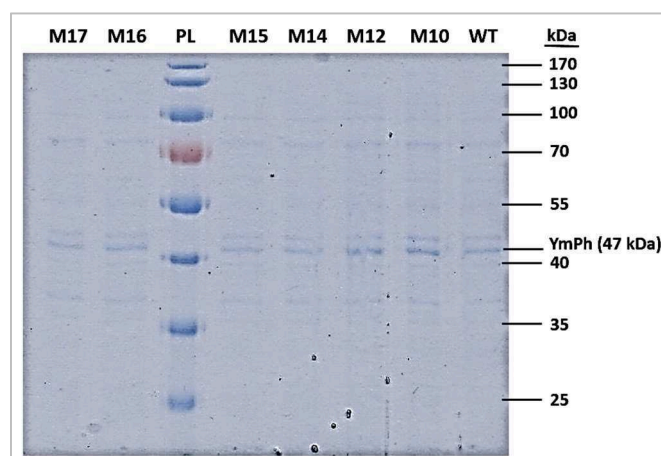
Figure 108: SDS-PAGE of YmPhWT-M10, -M12, -M14-M17 cell lysates. PL indicates the protein ladder and M10 to M17 describe the different YmPh variants. For the YmPhWT and the variants the phytase band is visible at the expected size of 47 kDa.

5.3.8. Determination of pH-profiles of YmPh variants with increased activity at pH 6.6

For characterization of the different YmPh variants pH profiles were performed with the best variants from the directed evolution campaign (YmPh-M10, -M12, -M14, -M15, -M16, -M17). Therefore, optimized 4-MUP activity assay was performed at various pHs (pH 3.2-8.6) according to the described protocol (see 5.5.2.6), using three different buffers (glycine buffer, sodium acetate buffer and Tris maleate buffer). The average slopes in RFU/s were calculated from the mean values of three measurements (Figure 109).
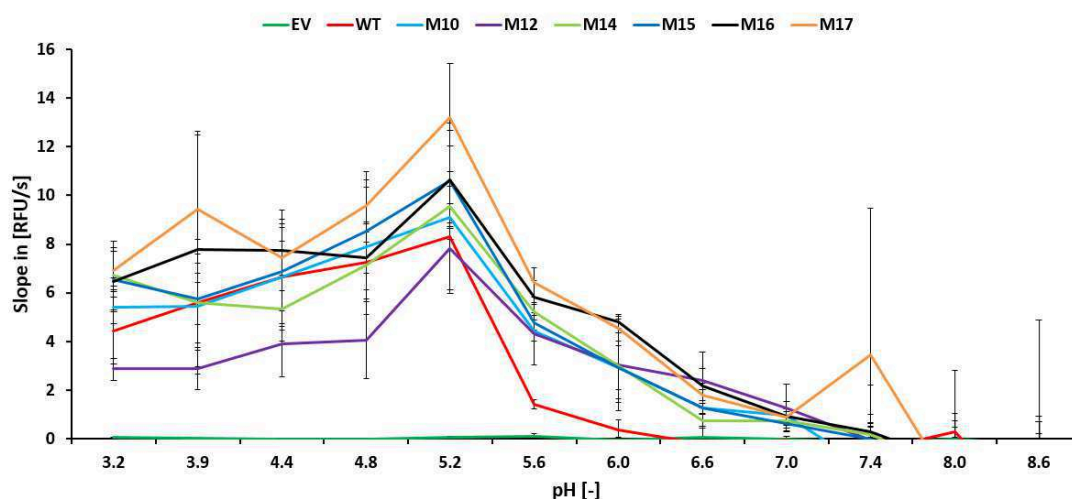
Figure 109: Result of the 4-MUP activity assay at different pHs (3.2 - 8.6) for determination of pH-profiles for the six best variants from the directed evolution campaign. The pH-profiles of EV (dark green) and YmPhWT (red) serve as comparison. The y-axis shows the slopes calculated from fluorescence measurements over time in [RFU/s] and the x-axis shows the different pHs. The slopes and standard deviations are calculated from mean values of three measurements.

The activity profile of all variants and the YmPhWT is characterized by a drastic loss in activity at pHs >5.2 (Figure 109). All six variants, especially variant YmPh-M16 shows a broadened pH profile towards neutral pH and a significantly improved activity at pH 6.6 to 7.0 in comparison to YmPhWT which is illustrated in Figure 109. At pH 6.6 and 7.0 the YmPhWT is completely inactive and YmPh-M16 still shows a slope of 2.15 RFU/s (for pH 6.6) and 0.91 RFU/s (for pH 7.0) respectively (Figure 109). Due to high activity at neutral pH in comparison to YmPhWT and low standard deviations, variants YmPh-M10 and -M16 were selected for purification and detailed kinetic characterization in terms of 4-MUP conversion.

5.3.9. Purification of YmPhWT and its variants

The YmPhWT and the best performing variants, YmPh-M10 and YmPh-M16, in terms of increased activity at pH 6.6, were purified using cation-exchange chromatography (ÄKTA prime plus, GE Healthcare Europe GmbH, Munich, Germany) according to the published protocol (Shivange *et al.*, 2012). Fractions from the phytase peak (Figure A21, appendix) were collected and analyzed using 4-MUP activity assay (see 5.5.2.7; Figure A22, appendix). SDS PAGE (Figure A23, appendix) was performed with the active fractions and the purest active samples were pooled together, desalted and concentrated with 10 kDA Amicon. The Experion system (Bio-Rad Laboratories GmbH) and Pierce® BCA Protein Assay Kit (Life Technologies GmbH) were used for estimation of the final protein concentrations and protein purity analysis of the purified phytase samples according to the manual's instructions. Figure 110 shows the analysis of the purified phytase samples (YmPhWT, -M10, -M16) undiluted and 1:3 diluted using Experion analysis. The phytase band can be observed at ~47 kDa in all analyzed samples. Finally, analysis of samples with Pierce[TM] BCA Protein Assay Kit (Life Technologies GmbH; Figure A24, appendix) and Experion[TM] Automated Electrophoresis System (Bio-Rad Laboratories GmbH; Figure 110), resulted in protein yields of 50.0 µg/ml for YmPhWT, 65.4 µg/ml for YmPh-M10 and 64.3 µg/ml for YmPh-M16 with an average purity of >80 %.
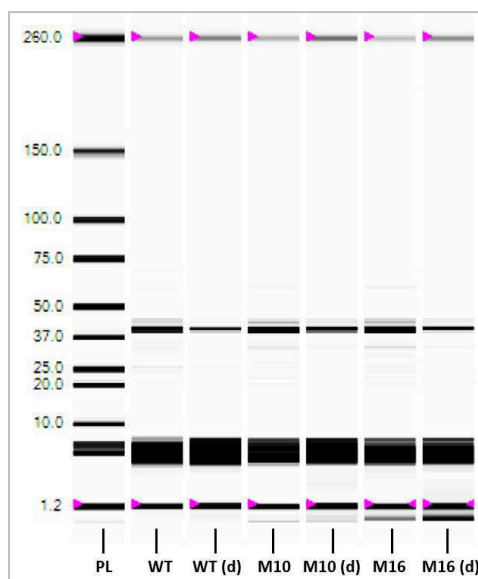
Figure 110: Experion analysis of purified YmPhWT, -M10 and M16 protein samples. **PL:** Protein ladder (Standard); **WT:** purified YmPhWT; **WT (d):** purified YmPhWT (1:3 diluted); **M10:** purified YmPh-M10; **M16:** purified YmPh-M16; **M16 (d):** purified YmPh-M16 (1:3 diluted).

Since 4-MU is not charged, it is possible that it might diffuse into the oil phase. In this case several alternative substrates that cannot diffuse out of the emulsion can be used: 1.) 4-methylumbelliferyl phosphate-6-carboxylic acid (4-MUPCA). In this reaction the phytase converts the substrate to the blue fluorescent product 4-methylumbelliferone-6-carboxylic acid (4-MUCA) (λ_{ex} 360 nm; λ_{em} 465 nm), by cleavage of the phosphate group (Figure 9).

Figure 9: Conversion of 4-methylumbelliferyl-phosphate-6-carboxylic-acid (4-MUPCA) substrate into the fluorescent product 4-methylumbelliferone-6-carboxylic-acid (4-MUCA) by phytase.

2.) Fluorescein di-phosphate triammonium salt (FDP) is another alternative substrate, which should not likely diffuse into the oil phase. In this reaction the phytase converts the FDP to the green fluorescent fluorescein (λ_{ex} 490nm; λ_{em} 514nm), by cleavage of the two phosphate groups (Figure 10).

Figure 10: Conversion of fluorescein di-phosphate triammonium salt (FDP) substrate into the fluorescent product fluorescein by phytase.

1.6.3.2. Fluorescence-based assays for detection of cellulase activity

For the cellulase, 4-methylumbelliferyl-β-D-cellobioside (4-MUC) is used as fluorogenic substrate for fluorescence based microtiterplate and flow cytometric assays (Lehmann *et al.*, 2012). The conversion of the blue fluorescent product 4-methylumbelliferone is detected by increase of fluorescence (ex. 360 nm, em. 465 nm), see Figure 11.

Figure 111: Enzyme kinetics of YmPhWT, -M10 and -M16 with 4-MUP substrate in Tris maleate buffer at pH 5.6. Initial activities plotted against different substrate concentrations and fitted with Hill equation using GraphPad Prism 6 software. The reported values are the average of three measurements and deviations are calculated from corresponding mean values. The x-axis shows the 4-MUP substrate concentration in [μM] and the y-axis shows the 4-MU product formation in [μM*s^{-1}].
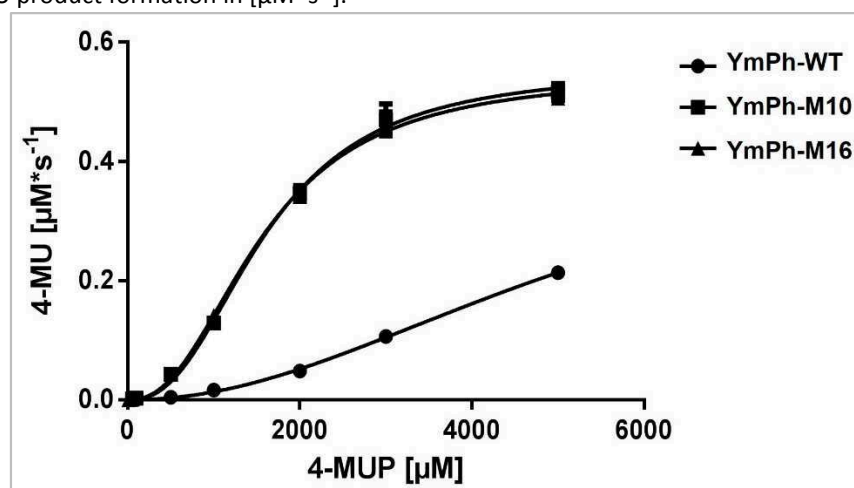
Figure 112: Enzyme kinetics of YmPhWT, -M10 and -M16 with 4-MUP substrate in Tris maleate buffer at pH 6.6. Initial activities plotted against different substrate concentrations and fitted with Hill equation using GraphPad Prism 6 software. The reported values are the average of three measurements and deviations are calculated from corresponding mean values. The x-axis shows the 4-MUP substrate concentration in [μM] and the y-axis shows the 4-MU product formation in [μM*s^{-1}].

5.3.11. Homology model generation using YASARA software

As shown in Figure 113, a structural model of YmPhWT was generated (see Material and Methods, chapter 5.5.2.6) and the identified beneficial positions T44 and K45 of YmPh-M10 and -M16 are illustrated. The phytase adopts a closed active site loop upon substrate binding, highlighted in cyan, and an open active site loop when no substrate is present at the active site (Lim *et al.*, 2000), highlighted in magenta (Figure 113). The identified substitutions in variant YmPh-M10 (T44I/K45E) and -M16 (T44V/K45E) are located at the active site loop (Figure 113A) and are expected to interact with the substrate upon binding. Figure 113B shows a schematic representation of the 4-MUP substrate (ball-stick representation) placed

manually at the active site of the phytase, which illustrates the proposed hypothesis of interaction between substrate and amino acid residues upon substrate conversion. The polar amino acid threonine, with uncharged side chain, is substituted by the amino acids valine and isoleucine, having hydrophobic side chains, which could probably lead to a better interaction with the -CH$_3$ group at the aromatic ring of the 4-MUP substrate (Figure 113B). In addition, the amino acid lysine, with a positively charged side chain, was replaced by glutamic acid in YmPh-M10 and -M16, harboring a negatively charged side chain, which might play a role upon hydrogen bond formation with the oxygen at the aromatic ring of the 4-MUP substrate, during conversion of the substrate by the putative catalytic residues (H38, D235, T326), highlighted in dark blue (Figure 113B).
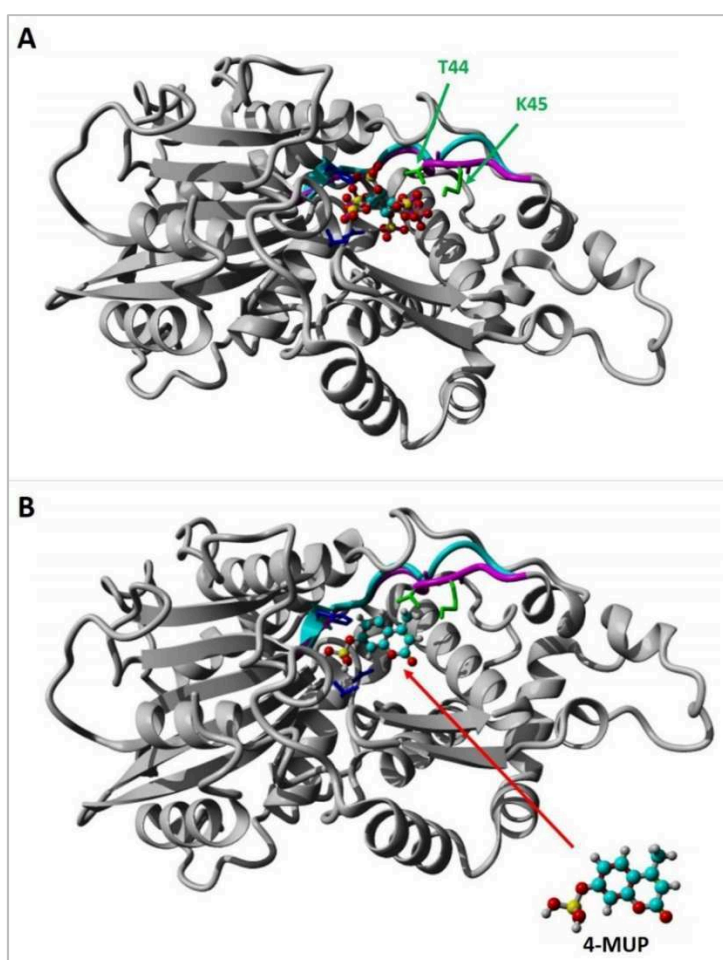
Figure 113: Hybrid homology model of YmPhWT generated with YASARA (Version 16.3.5, www.yasara.org), illustrating positions T44 and K45 (highlighted in green) which were identified as targets for beneficial amino acid substitutions in YmPh-M10 and -M16. Essential catalytic residues (H38, D325, T326) in the center of the enzyme are highlighted in dark blue. (**A**) Positions T44 and K45 are located in the active site loop of the phytase and are expected to interact with the substrate upon binding. Phytase adopts a closed (highlighted in cyan) and open active site loop (highlighted in magenta). The phytase substrate (phytate) is illustrated in ball-stick representation. (**B**) Schematic representation of the 4-MUP substrate (ball-stick representation) placed manually at the active site of the phytase, which illustrates the proposed hypothesis of substrate/amino acid interaction upon substrate conversion.

5.4. Discussion

In this study several *Yersinia mollaretii* phytase variants with broadened pH profile and improved activity towards 4-MUP substrate were identified, showing up to 7-fold increased specific activity at pH 6.6 in Tris maleate buffer (Table 27), significantly reduced K_{half} (up to 2.2-fold) and increased k_{cat} (up to 2-fold) values. Starting point for the directed evolution campaign was the highly active YmPhWT with a specific activity of 1,073 U/mg which is around 10-fold higher than widely used fungal phytases (Shivange *et al.*, 2012). Phytases, mainly of fungal origin, exhibiting specific activities of ≥100 U/mg, *i.e. Aspergillus niger* phytase, and are widely used as a feed supplement ($\sim$70 % of global monogastric feed products contain phytases) with a market value of around US $350 million per year and an impressive growth rate of 10 % per year (A. Cowieson, 2010; Shivange *et al.*, 2012). The identified bacterial phytase variants represent advantageous alternatives to fungal phytases for application in agriculture and monogastric feed products (*e.g.* fish industry) for which high phosphatase activity at a broad pH range is desired (Sandberg, 2002; Zhang *et al.*, 2011).

For successful identification of improved phytase variants at neutral pH, the reported 4-MUP screening system, which allows a continuous and sensitive monitoring of phytase activity (Shivange *et al.*, 2012), was optimized regarding pH, buffer, substrate and cell lysate concentration and validated for simultaneous screening at pH 5.6 and 6.6 in MTP-based format. Instead of the sodium acetate buffer (buffering capacity pH 3.7-5.6) Tris maleate buffer (buffering capacity pH 5.2-8.6) was chosen for the experiments enabling activity measurements around neutral pH values (pH 5.6 and 6.6).

The directed evolution campaign, comprising generation and screening of an YmPh-epPCR library in the first round and YmPh-SSM libraries in the second round, resulted in significantly improved phytase variants (M2-M17, Tables 25-26). After homology modelling and structural analysis of the identified, beneficial amino acid substitutions in the variants M2 and M7, exhibiting improved phytase activity at pH 6.6, the positions T44 and K45 located in close proximity to the active site of the enzyme (Figure 104) were selected for Site-Saturation Mutagenesis. The rescreening yielded five phytase variants (YmPh-M12, -M14, -M15, -M16 and -M17) with improved activity at pH 6.6 and overall broadened pH profile compared to YmPhWT (Figure 109). The obtained improvements are most probably due to the location of the identified amino acid substitutions in the active site

loop of the phytase. This might probably have an influence on substrate binding, *i.e.* stronger binding of 4-MUP substrate, which could enhance the substrate conversion at pH 5.6 and 6.6 (Figure 113). Especially, amino acid substitution K45E seems to be essential for both improved variants YmPh-M10 and -M16 from the YmPh-SSM library, showing significantly higher activity compared to the rest of the identified variants from the YmPh-epPCR library (Table 25). The activity tests at pH 5.5 using sodium acetate buffer resulted in comparable activities of the improved YmPh variants to the YmPhWT. Therefore, it can be concluded that positions T44 and K45 are the key positions for increasing phytase activity at acidic and neutral pH. Additionally, these results suggest that the active site loop might play a key role in modulating catalytic activity for substrate conversion. A highly flexible and conserved active site loop is proposed to be a prerequisite for very high catalytic efficiencies of *Enterobacteriaceae* phytases, *e.g.* YmPhWT, (Bohm *et al.*, 2010; Shivange *et al.*, 2012). The drastic decrease in activity (up to 4-fold) at pH >5.2 (Figure 109), which is characteristic for all obtained pH profiles ranging from pH 3.2 to 8.6 (for variants M10-M17) might be due to the pH optimum of these acid phosphatases (pH 5; Figure 109), therefore the loss of activity at pHs >5.2 is expected. Additionally the buffer is switched from sodium acetate buffer (for measurements at pH 3.9-5.2) to Tris maleate buffer (pH 5.6-8.6) using optimized protocol (see 5.5.2.6) which could be an additional reason for the drop in activity (Figure 109). On the other hand, variants YmPh-M10 and -M16 showed high activity levels in the acid pH range (pH 3.9-5.2) with simultaneously increased activity at neutral pH (Figure 109).

As members of the histidine acid phosphatase (HAPs) family have attracted increased interest from the animal feed industry because of their high specific activity and their ability to efficiently hydrolyze phytate (Shivange *et al.*, 2012), YmPh-M10 and -M16 are promising candidates for various industrial application (*e.g.* in aquaculture or agriculture). Due to their broad pH range and high specific activity, they are appropriate for a wide range of applications in the food industry, such as general agricultural additives for soil amendments (for soils of various pHs) or as feed additive in poultry and particularly for fishes due to their satisfactory activity in neutral pH range.

Additionally, variants YmPh-M10 and -M16 can be further evolved and tailored for industrially desired properties (*i.e.* thermostability, specific activity) by employing them in novel ultrahigh throughput screening platforms using *in vitro* compartmentalization techniques. For directed phytase evolution using cell-free enzyme expression systems,

phytase activity at neutral pH range is essential to obtain sufficient protein yields, as cell-free enzyme production systems are efficient at neutral pH range. Flow cytometer-based ultrahigh throughput screening platforms using *in vitro* compartmentalization techniques offer screening of up to 10^7 enzyme (phytase) variants per hour which provides an efficient coverage of the generated protein sequence space and facilitates a great potential to evolve new tailor made biocatalysts for industrial applications in a cost and time efficient manner.

5.5. Material and Methods

See Material and Methods chapter 3.6. Fluorogenic substrate 4-methylumbelliferylphosphate (4-MUP) and product 4-methylumbelliferone (4-MU) was purchased from Sigma–Aldrich Chemie GmbH (Taufkirchen, Germany).

5.5.1. Material

5.5.1.1. Strains, plasmids, and target gene

The *Yersinia mollaretii* ATCC 43969 phytase (*appA*) gene (GenBank: JF911533.1), codes for a phytase protein (*Yersinia mollaretii* phytase wild-type, YmPhWT) with a molecular weight of 47 kDa (monomer), consisting of 434 amino acids. YmPhWT expressed in *E. coli* has a tetrameric structure (~188 kDa) with a specific activity of 1,073 U/mg towards the natural substrate phytate (determined in 0.25 sodium acetate, pH 4.5, 37°C) which is around 10-fold higher than commonly used fungal phytases (Shivange *et al.*, 2012). The epPCR and SSM phytase libraries were cloned into expression vector pALXtreme-5b (Dennig *et al.*, 2011) and expression was performed using *E. coli* BL21 Gold(DE3)lacI^{Q1} cells (Blanusa *et al.*, 2010).

5.5.1.2. Media, solutions and buffers

LB-medium		TY-medium	
Trypton	10 g/l	Trypton	10 g/l
Yeast extract	5 g/l	Yeast extract	5 g/l
Sodium cloride	10 g/l	ddH$_2$O	add to 1 l
(Agar-Agar	10 g/l)		
ddH$_2$O	add to 1 l		

Autoclave (121°C, 1.4 bar, 20 min).

TYM-5052 medium (100 ml)		TYM-505 medium (100 ml)	
TY-medium	95.78 ml	TY-medium	95.78 ml
50x M	2 ml	50x M	2 ml
50x 5052	2 ml	50x 505	1 ml
1 M MgSO$_4$	0.2 ml	1 M MgSO$_4$	0.2 ml
Ampicillin sodium salt	0.1 ml	Ampicillin sodium salt	0.1 ml
Trace element solution	0.02 ml	Trace element solution	0.02 ml

Combine all solutions under sterile conditions in an appropriate vessel.

50x M stock solution (1x M solution: 50 mM PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4)

Na_2HPO_4	177.452 g/l
KH_2PO_4	170.112 g/l
NH_4Cl	133.724 g/l
Na_2SO_4	35.512 g/l
ddH_2O	add to 1 l

Autoclave (121°C, 1.4 bar, 20 min).

50x 5052 stock solution (1x 5052 solution: 0.5 % glycerol, 0.05 % glucose, 0.2 % α-lactose)

Glycerol	250 g/l
Glucose	27.5 g/l
α-Lactose	105.26 g/l
ddH_2O	add to 1 l

Autoclave (121°C, 1.4 bar, 20 min).

$MgSO_4$ (500x stock solution)

Dissolve 61.62 g of $MgSO_4$ x 7 H_2O in 210.5 ml ddH_2O and sterilize by autoclaving (121°C, 1.4 bar, 20 min).

100x Trace element solution

$CaCl_2$ x 2 H_2O	0.5 g/l
$ZnSO_4$ x 7 H_2O	0.18 g/l
$MnSO_4$ x H_2O	0.1 g/l
Na_2-EDTA 20	10 g/l
$FeCl_3$ x 6 H_2O	16.7 g/l
$CuSO_4$ x 5 H_2O	0.16 g/l
$CoCl_2$ x 6 H_2O	0.18 g/l

Filter sterile (0.2 μm) and store at 4°C.

Ampicillin stock solution (1000x)

Dissolve 100 mg/ml of ampicillin sodium salt in ddH_2O, filter sterile and store at -20°C.

4-Methylumbelliferylphosphate solution (10 mM)

Dissolve 2.5615 mg/ml of 4-methylumbelliferylphosphat (4-MUP) in acetate buffer (pH 5.5) or tris maleate buffer (pH 5.6 or 6.6) and store at -20°C.

4-Methylumbelliferone solution (100 mM)

Dissolve 17.62 mg/ml of 4-methylumbelliferone (4-MU) in 1 ml methanol and store at RT.

Lysozyme solution

Dissolve 2 mg/ml of lysozyme in Tris/HCl buffer (pH 7.4).

TFB1-solution (100 ml)		TFB2- solution (100 ml)	
Potassium acetat	2.394 g/l	MOPS	2.93 g/l
$MnCl_2$	9.89 g/l	$CaCl_2$	11.026 g/l
$RbCl_2$	12.09 g/l	$RbCl_2$	1.209 g/l
$CaCl_2$	1.11 g/l	Glycerol (w/v)	15 %

Dissolve in Tris/HCl buffer (pH 7.4), filter sterile (0.2 µm) and store at 4°C.

50 % (w/v) Glycerol

Dissolve 125 g of glycerol in 150 ml in ddH_2O and autoclave (121°C, 1.4 bar, 20 min).

50 mM Tris-HCl-Puffer pH 7,4

Dissolve 6.058 g/l Tris-ultrapur in ddH_2O and adjust the pH to 7.4 with 2 M HCl.

Glycine buffer (0.5 M)

Dissolve 37.54 g/l glycine in ddH_2O and adjust the pH to 3.2 with HCl.

Sodium acetate buffer (0.5 M)

Acetic acid (100 %)	1.76 g/l
Sodium acetate	18 g/l
$CaCl_2$ x 2 H_2O	0.147 g/l
Tween 20	10 %

Dissolve in ddH_2O and adjust the pH to 3.9-5.5 with acetic acid/NaOH.

Figure 11: Conversion of 4-methylumbelliferyl-β-D-cellobioside (4-MUC) substrate into the fluorescent product 4-methylumbelliferone (4-MU) by cellulase.

Since 4-MU is not charged, it is possible that it might diffuse into the oil phase. In this case fluorogenic substrate fluorescein-di-β-D-cellobioside (FDC) can be used as alternative substrate for flow cytometer-based screening within (w/o/w) emulsion compartments, which should not likely diffuse into the oil phase. In this reaction, cellulase converts the FDC substrate to the green fluorescent product fluorescein (λ_{ex} 494 nm/λ_{em} 516 nm), by cleavage of the β-1,4-D-glycosodic bonds (Figure 12).

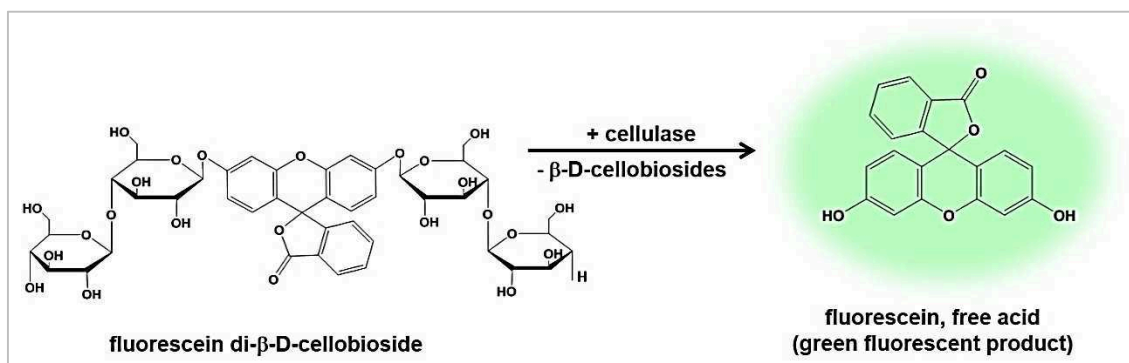**Figure 12:** Conversion of fluorescein di-β-D-cellobioside (FDC) substrate into the fluorescent product fluorescein by cellulase.

5.5.2. Methods

5.5.2.1. Error-prone PCR (epPCR)

The random mutagenesis library was constructed by standard epPCR method (Cadwell *et al.*, 1994). To obtain linearized template DNA for the PCR reactions, the DNA for insert amplification with epPCR was double digested with restriction enzymes *Pst*I and *Bse*YI (10 U) using pALXtreme-5b-YmPhWT as template according to the protocol of the manufacturer and purified from agarose-gel with Nucleo Spin® Gel and PCR Clean-up kit (Macherey-Nagel GmbH & Co. KG). The vector DNA was linearized respectively, with restriction enzyme *Kpn*I/*Nar*I (10 U) using pALXtreme-5b-YmPhWT as template.

The epPCR-YmPh library was generated by PCR (94°C for 1 min, 1 cycle; 94°C for 30 s / 57°C for 30 s / 72°C for 1.45 min, 25 cycles; 72°C for 5 min, 1 cycle) using dNTP mix (0.2 mM), primers (Plic_*Ym*ph-WT_FP: 5' – ctt taa gaa gga GAT ATA CAT ATG CGA TTA ACT G - 3'; Plic_*Ym*ph-WT_RP: 5' – ggc ttt gtt agc AGC CGG ATC- 3'; 0.4 µM each; Table 29, Material and Methods chapter 5.5.1.4), plasmid DNA template (linearized pALXtreme-5b-YmPhWT, 20 ng), *Taq* polymerase (2.5 U), and MnCl$_2$ (0.05 mM - 0.4 mM) in a final volume of 50 µl.

The vector backbone (pAlXtreme-5b) for the PLICing reaction (Blanusa *et al.*, 2010) was generated by PCR (98°C for 2 min, 1 cycle; 98°C for 30 s / 57°C for 1.30 min / 72°C for 5 min, 25 cycles; 72°C for 5 min, 1 cycle) using dNTP mix (0.2 mM), primers (Plic_pALX-5b_FP: 5' - gct aac aaa gcc CGA AAG GAA GCT - 3'; Plic_pALX-5b_RP: 5' - tcc ttc tta aag TTA AAC AAA ATT ATT TCT AGA GGG G - 3'; 0.4 µM each; Table 29, Material and Methods chapter 5.5.1.4), plasmid DNA template (linearized pALXtreme-5b-YmPhWT, 20 ng), and *PfuS* polymerase (2.5 U) in a final volume of 50 µl.

The amplified (ep)PCR products were digested (1 h, 37°C) by *Dpn*I (20 U) according to the protocol of the manufacturer. *Dpn*I was inactivated (20 min, 80°C) and PCR products were purified using the NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG), eluted in 20 µl ddH$_2$O and used for subcloning of the insert DNA (YmPh) into the vector (pALXtreme-5b) *via* PLICing reaction, performed according to the published protocol (Blanusa *et al.*, 2010). Subsequently, pALXtreme-5b-YmPh constructs were transformed into chemically competent *E. coli* BL21 Gold(DE3)lacI^{Q1} cells. DNA sequencing of the inserted gene was performed by Eurofins Genomics (Ebersberg, Germany) and Clone Manager 9 Professional Edition (Sci-Ed software, Cary, USA) was used for sequence analysis.

5.5.2.2. Colony-PCR

Successful cloning of YmPh gene into pAlXtreme-5b vector backbone was checked after transformation by colony PCR (94°C for 1 min, 1 cycle; 94°C for 30 s / 57°C for 30 s / 72°C for 1.45 min, 25 cycles; 72°C for 5 min, 1 cycle) using dNTP mix (0.05 mM), primers (Plic_*Ym*ph-WT_FP: 5' – ctt taa gaa gga GAT ATA CAT ATG CGA TTA ACT G - 3'; Plic_*Ym*ph-WT_RP: 5' - ggc ttt gtt agc AGC CGG ATC - 3'; 0.1 µM each; Table 29, Material and Methods chapter 5.5.1.4), *Taq* polymerase (0.625 U) and ½ colony from agar plate was transferred into the PCR reaction mixture with a final volume of 12.5 µl. PCR products were analyzed by agarose-gelelectrophoresis.

5.5.2.3. Expression of YmPh in MTP

Clones were randomly picked into 96-well flat bottom MTPs (Greiner Bio-one GmbH) containing 150 µL Luria broth medium supplemented with ampicillin (100 µg/ml, LB$_{Amp}$) and cultivated (overnight, 37°C, 900 rpm, 70 % humidity). Plates were stored at -80°C after addition of 100 µl glycerol (50 % w/v) as master plates.

For preculture preparation a 96-well flat bottom MTP (Greiner Bio-one GmbH) containing 150 µL LB$_{Amp}$ was inoculated with 96-well replicator from master plate and incubated (overnight, 37°C, 900 rpm, 70 % humidity).

For main culture preparation a 96-well V-bottom expression MTP (Greiner Bio-one GmbH) containing 150 µL TYM-5052 auto-induction medium supplemented with ampicillin (100 µg/ml, TYM-5052$_{Amp}$) was inoculated with the preculture using a 96-well replicator (mechanical shop of RWTH Aachen, Aachen, Germany). After cultivation (8 h, 37°C, 900 rpm, 70 % humidity) plates were centrifuged (4000 *g*, 20 min, 4°C) and pellets were stored at -20°C.

5.5.2.4. Expression of YmPh in Erlenmeyer-flasks

For preculture preparation TYM-505$_{Amp}$ (10 ml in 50 ml Erlenmeyer-flask) was inoculated with 4 µl of a respective cryoculture and incubated (8 h, 37°C, 250 rmp). For phytase expression, the main culture containing 200 ml TYM-5052$_{Amp}$ was inoculated with 1 % of the preculture and incubated (8 h, 37°C, 250 rpm). Cells were harvested by centrifugation (4000 *g*, 20 min, 4°C) and pellets were stored at -20°C.

5.5.2.5. Screening and rescreening with 4-MUP activity assay

For continuous phytase activity detection the highly sensitive fluorescence-based 4-MUP activity assay (Shivange *et al.*, 2012) for quantification of phosphorolytic activity was optimized for measurements at pH >5.6. Cell pellets were resuspended in 100 µl lysozyme solution (1 mg/ml Tris/HCl buffer, 50 mM, pH 7.4) and incubated (1 h, 37°C, 900 rmp, 70 % humidity). The plates were centrifuged (4000 *g*, 20 min, 4°C), 10 µl supernatant was added to 40 µl Tris maleate buffer (0.2 M; pH 5.6/6.6) and reactions were started by addition of 50 µl fluorogenic substrate 4-MUP (Sigma–Aldrich Chemie GmbH; 2 mM dissolved in Tris maleate buffer, pH 5.6/6.6). Reactions had a final volume of 100 µl.

Phytase activity was determined using a Tecan Infinite®M1000 plate reader (Tecan Group Ltd.; gain: 140; z-position: 19500; interval: 30 s, time: 20 min, temperature: 37 °C) as increase in the relative fluorescence at (λ_{ex} 360 nm, λ_{em} 465 nm) over time in RFU/s, due to the hydrolysis of 4-MUP to the fluorescent product 4-methylumbelliferone (4-MU). Three YmPhWT and empty vector controls were included in each MTP. Relative activity for each clone was defined by calculating the increase of fluorescence over time (slope in RFU/s) of each variant compared to the average slope of YmPhWT (determined from measurements in triplicates).

5.5.2.6. Determination of pH-profiles using 4-MUP activity assay

The 4-MUP activity assay (see 5.5.2.5) was further optimized regarding gain (140, 100), 4-MUP concentrations (25 mM; 50 mM; 100 mM) and cell lysate dilutions (1:2; 1:3; 1:6), for determination of pH-profiles of the different variants using three different buffers (glycine buffer, pH 3.2; sodium acetate buffer, pH 3.9-5.2; Tris maleate buffer pH 5.6-8.6). For measurement of pH profiles, the above mentioned assay conditions (see 5.5.2.5) were optimized as following. To 88 µl of buffer (glycine buffer/sodium acetate buffer/Tris maleate buffer) at the respective pH, 10 µl of cell lysate (1: 6 diluted) were added and reactions were started by addition of 2 µl 4-MUP (Sigma–Aldrich Chemie GmbH; 100 mM; in respective buffer and pH). Reactions had a final volume of 100 µl. Phytase activity was measured by monitoring the increase in RFU (λ_{ex} 360 nm, λ_{em} 465 nm), due to the hydrolysis of 4-MUP to the fluorescent product 4-methylumbelliferone (4-MU), over time (s) using a Tecan Infinite®M1000 plate reader (Tecan Group Ltd.; gain: 100; z-position: 19500; interval: 20 s, time: 20 min, temperature: 37 °C).

5.5.2.7. Activity measurement of purified YmPh fractions using 4-MUP activity assay

To 50 µl of purified phytase 50 µl 4-MUP (Sigma–Aldrich Chemie GmbH; 1 mM; in Tris maleate buffer, 0.2 M; pH 5.6/6.6) in were added prior measurement (final reaction volume 100 µl). Phytase activity was measured by monitoring the increase in RFU (λ_{ex} 360 nm, λ_{em} 465 nm), due to the hydrolysis of 4-MUP to the fluorescent product 4-methylumbelliferone (4-MU), over time (s) using a Tecan Infinite®M1000 plate reader (Tecan Group Ltd.; gain: 100; z-position: 19500; interval: 30 s, time: 20 min, temperature: 37°C).

5.5.2.8. Kinetic characterization of YmPh variants at pH 5.6 and 6.6 using 4-MUP activity assay

To 20 µl of purified phytase enzyme (YmPhWT, YmPh-M10 and YmPh-M16, see 5.3.9), 30 µl of respective buffer (sodium acetate buffer, pH 5.5; Tris maleate buffer, pH 5.6/6.6) and 50 µl 4-MUP (Sigma–Aldrich Chemie GmbH; 0.01-5.0 mM; in respective buffer, pH 5.6/6.6) were added prior measurement. Reactions had a final volume of 100 µl. Phytase activity was measured by monitoring the increase in RFU (λ_{ex} 360 nm, λ_{em} 465 nm), due to the hydrolysis of 4-MUP to the fluorescent product 4-methylumbelliferone (4-MU), over time (s) using a Tecan Infinite®M1000 plate reader (Tecan Group Ltd.; gain: 100; z-position: 19500; interval: 30 s, time: 20 min, temperature: 37 °C). The obtained data were analyzed using GraphPad Prism 6 software (GraphPad software, San Diego, CA, USA) according to sigmoidal kinetics (Hill coefficient >1) in order to obtain kinetic data (K_{half}, k_{cat}) for each protein variant, as it was reported previously that YmPhWT is a tetrameric enzyme showing allosteric interactions (Shivange *et al.*, 2012). Initial velocity data were obtained and fitted with the following equation: $Y = V_{max} * X^h/(K_{half}^h + X^h)$, where Y is the initial activity, V_{max} is the maximum velocity, h is the Hill coefficient, X is the substrate concentration, and K_{half} is the concentration of substrate that will afford half maximal velocity using GraphPad Prism software (GraphPad software). The k_{cat} was calculated from the ratio of V_{max} and the phytase concentration. The catalytic activity (k_{cat}) was calculated by considering four active sites in one tetrameric molecule of phytase. One unit (U) of enzyme was defined as the amount of phytase that catalyzes the conversion of 1 µmol of 4-MUP per minute.

Calibration curves were prepared for different concentrations (0.01-1.8 mM; 50 µl) of the fluorescent product 4-methylumbeliferone [4-MU] in dependency to the relative fluorescence in [RFU/s] in three different buffers (50 µl; sodium acetate buffer pH 5.5;

Tris maleate buffer pH 5.6; Tris maleate buffer pH 6.6). Measurements were performed with Tecan Infinite®M1000 plate reader (Tecan Group Ltd.; gain: 100; z-position: 19500; interval: 30 s, time: 20 min, temperature: 37°C) resulting in linear correlation of relative fluorescence versus product formation (Figure A27-A29, appendix).

5.5.2.9. Purification, desalting and concentration of phytase

Purification of YmPhWT, YmPh-M10 and YmPh-M16 was performed using cation-exchange chromatography (Äkta prime, Toyopearl SP650C column) according to the published protocol (Shivange *et al.*, 2012). Active, pure protein fractions were pooled together, desalted and concentrated in 10 kDa Amicons and stored at 4°C. Protein purity was determined using Experion™ Automated Electrophoresis System (Bio-Rad Laboratories GmbH, München, Germany) and Pierce™ BCA Protein Assay Kit (Life Technologies GmbH, Darmstadt, Germany). The purified YmPhWT protein and the variants were used for detailed kinetic characterization (see 5.3.10 and 5.5.2.8).

5.5.2.10. Homology modeling of Yersinia mollaretii phytase

Homology modelling of the structure of *Yersinia mollaretii* phytase was performed by using YASARA Structure version 16.3.8 (Krieger et al., 2002) with the default settings (PSI-BLAST iterations: 6 (Altschul et al., 1997), E value cutoff: 0.5, templates: 5, OligoState: 4). A position-specific scoring matrix (PSSM) was used to score the obtained template structures (Jones, 1999; Qiu et al., 2006). Five X-ray templates were selected for modelling the YmPhWT structure (417 aa): *Yersinia Kristensenii* phytase apo-form (409 residues with quality score 0.562, PDB ID: 4ARV), *Hafnia Alvei* phytase in complex with tartrate (405 residues with quality score 0.547, PDB ID 4ARU), *Hafnia Alvei* phytase in complex with myo-inositol hexakis sulphate (401 residues with quality score 0.530, PDB ID: 4ARO), *Hafnia Alvei* phytase apo-form (407 residues with quality score 0.487, PDB ID: 4ARS), phytate complex *Escherichia coli* phytase with bound phytate in the active site (410 residues with quality score 0.605, PDB ID: 1DKQ). The best-scoring model (Z-score= -0.196) was a hybrid model build on X-ray structure 4ARV, 4ARU, and 4ARO and consisted of a monomer with a bound tartrate. For location of active site loop the structural alignment of homology model with X-ray structures of open and closed conformation of *Escherichia coli* phytase (PDB ID: 1DKP and 1DKL) was performed (Lim *et al.,* 2000).

6. Summary and conclusions

The main limitation of protein engineering by directed evolution is the development of a suitable screening system in order to achieve sufficient coverage of the generated protein sequence space. A peptide consisting of only five amino acids can already yield 3.2 million different protein variants. In conventional directed evolution campaigns rather small libraries with only a few thousand clones and low mutational loads are screened using MTP or agar-based screening formats to achieve modest improvements by laborious, iterative rounds of mutagenesis and screening (Ruff *et al.*, 2013). Detailed studies showed that >50 % of beneficial amino acid positions influence the enzyme property while only a small portion of beneficial positions was identified in traditional rounds of epPCR using medium screening formats (Frauenkron-Machedjou *et al.*, 2015; Zhao *et al.*, 2014) indicating the demand for an ultrahigh throughput screening system. This emphasizes the importance to develop widely applicable routine ready ultrahigh throughput screening (uHTS) systems, offering a throughput of >10^7 events per hour and the rapid identification of beneficial variants to bridge the limited screening capacity to the generated protein sequence space. Consequently, uHTS-IVC technology offers the opportunity to employ novel mutagenesis strategies, which enable generation of mutants with up to five multiple simultaneous site-saturations (*e.g.* OmniChange) in directed evolution campaigns, to efficiently capitalize on nature's diversity. The development of routine ready uHTS-IVC technologies combined with novel mutagenesis methods for generation of highly diverse mutant libraries provide a powerful tool to efficiently tailor enzymes for specific properties, *e.g.* activity, selectivity, pH stability or substrate specificity. Beyond that, flow cytometer-based IVC technology enables to overcome the limitations of cell-based systems, *e.g.* gene diversity loss due to low transformation efficiency of the expression host, or low protein yields upon *in vivo* expression of membrane or toxic proteins, since the genetic material is directly put into the artificial compartment, making this technique advantageous and interesting for research and industry (Blanusa, 2009). For these reasons, the employment of uHTS-IVC technology upon screening of highly diverse enzyme libraries plays a key role to facilitate the probability to identify beneficial variants, as well as to investigate combinatorial effects between selected amino acid positions, or to explore complex structure-function relationships in enzymes to generate a deeper understanding of functional genomics which in the end is essential for an

efficient tailoring of enzyme properties and to adapt the obtained knowledge to other enzymes or enzyme classes.

In this work, for the first time a flow cytometer-based uHTS-IVC screening platform for *in vitro* directed cellulase (CelA2-H288F) evolution was developed and validated, employing fluorescein-di-β-D-cellobioside (FDC) as substrate within (w/o/w) emulsion compartments (Chapter I). The development of the flow cytometer-based uHTS-IVC screening platform comprised the optimization of (w/o/w) emulsion generation, cell-free expression of cellulases within emulsion compartments, and establishment a flow cytometer-based sorting strategy using libraries with defined active-to-inactive ratios. The combination of cell-free compartmentalization with flow cytometer-based analysis achieves a throughput of 6.5×10^6 events per hour. In one single round of directed cellulase evolution more than 1.4×10^7 events were analyzed in only one day. The validation of the developed flow cytometer-based uHTS-IVC technology platform was performed by identification of an improved cellulase variant upon flow cytometry screening of a random epPCR-CelA2 mutant library. The rescreening in MTP format revealed the highly improved cellulase variant M1 (N273D/H288F/N468S) with a specific activity of 220.6 U mg^{-1} compared to CelA2 wild type with 16.57 U mg^{-1}, and the parent CelA2-H288F with 72.62 U mg^{-1} despite that a coumarin derivative (4-MUC) and not the flow cytometer prescreening substrate (pair FDC/fluorescein) was used for the MTP-based rescreening. The establishment of the flow cytometer-based uHTS-IVC screening platform for *in vitro* directed cellulase evolution offers novel routes to in depth study structure-function relationships and elucidate cooperative effects in cellulase which is of great scientific impact for the scientific community.

In a second approach (Chapter II), the uHTS-IVC technology platform was successfully employed and validated for screening of different Site-Saturation Mutagenesis libraries (*e.g.* CelA2-OmniChange library) with four simultaneously saturated positions, yielding over one million different mutants which are inaccessible to screen using standard MTP-based screening formats. The targeted positions, located in close proximity to the active site, were selected using a homology model of CelA2-WT which was generated with YASARA software. The highly diverse mutant library displaying a low percentage of active variants (2.56 %) was successfully enriched by 30-fold using flow cytometry-based analysis of over 36×10^6 events and subsequent sorting of 3.9×10^5 active events, ensuring full coverage of the generated diversity due to oversampling. Subsequent analysis of the sorted fraction with 4-MUC

2. Goal of this work

The main goal of this project was the development of a novel and routine-ready flow cytometer-based ultrahigh throughput *in vitro* compartmentalization (uHTS-IVC) screening platform for directed protein evolution (Figure 13). The project was financed by the German Federal Ministry of Education and Research (BMBF) within the framework of *"Basistechnologien für die nächste Generation biotechnologischer Verfahren"* (FKZ 031A165), which is supporting the development of novel, robust technology platforms for industrial applications in the field of biotechnology.

Screening throughput is one of the key criteria for successful directed enzyme evolution and discovery of novel enzymes. Medium to high throughput screening systems based on microtiter plates (MTP) or agar plates usually offer analysis of 10^4-10^5 variants, which is still significantly lower compared to the size (10^8-10^9) and potential complexity of random mutagenesis libraries (Agresti *et al.*, 2010; Wong *et al.*, 2006). As nowadays the development of rapid screening systems for efficient coverage of the generated sequencing space is the main limitation of directed evolution campaigns, ultrahigh throughput enzyme screening (uHTS) technologies offering a throughput of up to 10^7 events per hour are a key tool for today's state-of-the-art directed evolution experiments, as they enable rapid analysis of millions of variants (Griffiths *et al.*, 2006; Mayr *et al.*, 2009; Miller *et al.*, 2006). Ultrahigh throughput of 10^7 events per hour enables efficient coverage of the generated sequence space in highly diverse mutant libraries (Dennig *et al.*, 2011; Reetz *et al.*, 2010; Shivange *et al.*, 2009) and allows successful re-engineering of biocatalysts with high industrial application potential (*e.g.* for chemical, pharmaceutical, food industry) through iterative rounds of directed evolution in a time and cost efficient manner (Prodanovic *et al.*, 2011).

A recent highlight in directed enzyme evolution was the development of an alternative ultrahigh throughput screening platform employing *in vitro* enzyme library production within water-in-oil-in-water (w/o/w) emulsion compartments (*in vitro* compartmentalization) and subsequent analysis by flow cytometry (Bernath *et al.*, 2005). *In vitro* compartmentalization (IVC) technology enables the miniaturization of reaction volumes to microliter scale, by production of 10^{10} reaction compartments per milliliter of reaction with diameters ranging from 0.5-10 µm, resulting in a dramatic reduction of consumable costs, workload and assay time, compared to standard medium to high throughput screening systems (Agresti *et al.*, 2010; Bernath *et al.*, 2005; Mastrobattista *et al.*, 2005; Taly *et al.*, 2007). Upon *in vitro*

numbers that are order of magnitude higher than standard whole cell screening systems (>10^3-fold) (Leemhuis *et al.*, 2009; Packer *et al.*, 2015). Despite that uHTS-IVC technology requires oversampling, as only every 100th emulsion droplet contains a gene template to ensure that sorted emulsions contain genes, which encode in the majority beneficial variants in order to minimize diversity loss in subsequent PCR amplification. In addition, due to the miniaturization in (w/o/w) emulsions, the volumes of reaction components are drastically minimized to only a few microliters of volume (115 µl) in comparison to MTP-based screening volumes, which leads to a drastic reduction in assay costs.

In summary, the developed cellulase uHTS-IVC technology platform is a rapid, cost-efficient and non-laborious screening platform for directed cellulase evolution which offers to cover a significant fraction of the generated sequence space and to identify beneficial positions beyond the possibilities of traditional screening formats. Especially, the broad applicability of the uHTS-IVC technology is of great scientific value, as it offers the great advantage to employ cell-free expression systems based on various cell extracts, *e.g.* wheat germ and other eukaryotic cell extracts, which opens many exciting possibilities to evolve besides cell toxic proteins (Worst *et al.*, 2015) and membrane proteins (Quast *et al.*, 2015a; Sachse *et al.*, 2013) also proteins from human or animal origin for which *in vivo* expression remains challenging, with great application potential for many different enzyme classes and various industrial opportunities (Stech *et al.*, 2015; Zemella *et al.*, 2015). We see the main application of the uHTS-IVC technology as pre-screening system enabling researchers to isolate the most active variants from a vast pool of variants and to explore novel directed evolution strategies with high mutational loads in which only a small fraction of a populations is active. The minimized time effort, enabling the performance of one round of directed evolution in less than one week, is a prerequisite for the implementation of the routine-ready uHTS-IVC technology platform in directed evolution campaigns for industrial or pharmaceutical purposes.

7. References

A. Cowieson, R. C. (2010). *Introduction to the event and overview of the phytase market.* International Phytase Summit, Washington, DC.

Agresti, J. J., Antipov, E., Abate, A. R., Ahn, K., Rowat, A. C., Baret, J. C., Marquez, M., Klibanov, A. M., Griffiths, A. D., & Weitz, D. A. (2010). Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc Natl Acad Sci U S A*, 107(9): 4004-4009.

Aharoni, A., Amitai, G., Bernath, K., Magdassi, S., & Tawfik, D. S. (2005a). High-throughput screening of enzyme libraries: thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments. *Chem Biol*, 12(12): 1281-1289.

Aharoni, A., Griffiths, A. D., & Tawfik, D. S. (2005b). High-throughput screens and selections of enzyme-encoding genes. *Curr Opin Chem Biol*, 9(2): 210-216.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25(17): 3389-3402.

Arnold, F. H. (2001). Combinatorial and computational challenges for biocatalyst design. *Nature*, 409(6817): 253-257.

Arnold, F. H., & Moore, J. C. (1997). Optimizing industrial enzymes by directed evolution. *Adv Biochem Eng Biotechnol, 58:* 1-14.

Arnold, F. H., & Volkov, A. A. (1999). Directed evolution of biocatalysts. *Curr Opin Chem Biol*, 3(1): 54-59.

Basova, E. Y., & Foret, F. (2015). Droplet microfluidics in (bio)chemical analysis. *Analyst*, 140(1): 22-38.

Bernath, K., Hai, M., Mastrobattista, E., Griffiths, A. D., Magdassi, S., & Tawfik, D. S. (2004a). In vitro compartmentalization by double emulsions: sorting and gene enrichment by fluorescence activated cell sorting. *Anal Biochem*, 325(1): 151-157.

Bernath, K., Magdassi, S., & Tawfik, D. S. (2004b). In vitro compartmentalization (IVC): A high-throughput screening technology using emulsions and FACS. *Discov Med*, 4(20): 49-53.

Bernath, K., Magdassi, S., & Tawfik, D. S. (2005). Directed evolution of protein inhibitors of DNA-nucleases by in vitro compartmentalization (IVC) and nano-droplet delivery. *J Mol Biol*, 345(5): 1015-1026.

Bitar, K., & Reinhold, J. G. (1972). Phytase and alkaline phosphatase activities in intestinal mucosae of rat, chicken, calf, and man. *Biochim Biophys Acta*, 268(2): 442-452.

Blanusa, M. (2009). *Flow cytometry based screening system for finding and improving bioindustrial important biocatalysts - Cytochrome P450 BM3.* Unpublished Dissertation, Jacobs University, Bremen, Germany, School of Engineering and Science, Jacobs University, Bremen, Germany.

Blanusa, M., Schenk, A., Sadeghi, H., Marienhagen, J., & Schwaneberg, U. (2010). Phosphorothioate-based ligase-independent gene cloning (PLICing): An enzyme-free and sequence-independent cloning method. *Anal Biochem*, 406(2): 141-146.

Bohm, K., Herter, T., Muller, J. J., Borriss, R., & Heinemann, U. (2010). Crystal structure of Klebsiella sp. ASR1 phytase suggests substrate binding to a preformed active site that meets the requirements of a plant rhizosphere enzyme. *FEBS J*, 277(5): 1284-1296.

Brouzes, E., Medkova, M., Savenelli, N., Marran, D., Twardowski, M., Hutchison, J. B., Rothberg, J. M., Link, D. R., Perrimon, N., & Samuels, M. L. (2009). Droplet microfluidic technology for single-cell high-throughput screening. *Proc Natl Acad Sci U S A*, 106(34): 14195-14200.

References

Cadwell, R. C., & Joyce, G. F. (1992). Randomization of genes by PCR mutagenesis. *PCR Methods Appl*, 2(1): 28-33.

Cadwell, R. C., & Joyce, G. F. (1994). Mutagenic PCR. *PCR Methods Appl*, 3(6): S136-140.

Cao, L., Wang, W., Yang, C., Yang, Y., Diana, J., Yakupitiyage, A., Luo, Z., & Li, D. (2007). Application of microbial phytase in fish feed. *Enzyme Microb Technol*, 40(4): 497-507.

Chandra, R. P., Bura, R., Mabee, W. E., Berlin, A., Pan, X., & Saddler, J. N. (2007). Substrate pretreatment: the key to effective enzymatic hydrolysis of lignocellulosics? *Adv Biochem Eng Biotechnol*, 108: 67-93.

Cirino, P. C., Mayer, K. M., & Umeno, D. (2003). Generating mutant libraries using error-prone PCR. *Methods Mol Biol*, 231: 3-9.

Cohen, N., Abramov, S., Dror, Y., & Freeman, A. (2001). In vitro enzyme evolution: the screening challenge of isolating the one in a million. *Trends Biotechnol*, 19(12): 507-510.

Courtois, F., Olguin, L. F., Whyte, G., Theberge, A. B., Huck, W. T. S., Hollfelder, F., & Abell, C. (2009). Controlling the Retention of Small Molecules in Emulsion Microdroplets for Use in Cell-Based Assays. *Anal Chem*, 81(8): 3008-3016.

Daugherty, P. S., Iverson, B. L., & Georgiou, G. (2000). Flow cytometric screening of cell-based libraries. *J Immunol Methods*, 243(1-2): 211-227.

Dennig, A., Shivange, A. V., Marienhagen, J., & Schwaneberg, U. (2011). OmniChange: the sequence independent method for simultaneous site-saturation of five codons. *PLoS One*, 6(10): e26222.

DeVries, J. K., & Zubay, G. (1967). DNA-directed peptide synthesis. II. The synthesis of the alpha-fragment of the enzyme beta-galactosidase. *Proc Natl Acad Sci U S A*, 57(4): 1010-1012.

Discher, D. E., & Eisenberg, A. (2002). Polymer Vesicles. *Science*, 297(5583): 967-973.

Doi, N., & Yanagawa, H. (1999). Design of generic biosensors based on green fluorescent proteins with allosteric sites by directed evolution. *FEBS Lett*, 453(3): 305-307.

Esteghlalian, A. R., Srivastava, V., Gilkes, N. R., Kilburn, D. G., Warren, R. A., & Saddle, J. N. (2001). Do cellulose binding domains increase substrate accessibility? *Appl Biochem Biotechnol*, 91-93: 575-592.

Farinas, E. T. (2006). Fluorescence activated cell sorting for enzymatic activity. *Comb Chem High Throughput Screen*, 9(4): 321-328.

Farinas, E. T., Bulter, T., & Arnold, F. H. (2001). Directed enzyme evolution. *Curr Opin Biotechnol*, 12(6): 545-551.

Frauenkron-Machedjou, V. J., Fulton, A., Zhu, L., Anker, C., Bocola, M., Jaeger, K. E., & Schwaneberg, U. (2015). Towards understanding directed evolution: more than half of all amino acid positions contribute to ionic liquid resistance of Bacillus subtilis lipase A. *Chembiochem*, 16(6): 937-945.

Fu, S., Sun, J., Qian, L., & Li, Z. (2008). Bacillus phytases: present scenario and future perspectives. *Appl Biochem Biotechnol*, 151(1): 1-8.

Furukawa, K. (2003). 'Super bugs' for bioremediation. *Trends Biotechnol*, 21(5): 187-190.

Gerrits M., S. J., Claußnitzer I., von Goll U., Schäfer F., Rimmele M. and Stiege W. (2000-2013). *Cell-Free Synthesis of Defined Protein Conjugates by Site-directed Cotranslational Labeling.* Austin (TX): Landes Bioscience: Madame Curie Bioscience Database [Internet].

Gilkes, N. R., Warren, R. A., Miller, R. C., Jr., & Kilburn, D. G. (1988). Precise excision of the cellulose binding domains from two Cellulomonas fimi cellulases by a homologous protease and the effect on catalysis. *J Biol Chem*, 263(21): 10401-10407.

References

Givan, A. L. (2001). Principles of flow cytometry: An overview. ***Method Cell Biol, Vol 63***, 63: 19-50.

Goddard, J.-P., & Reymond, J.-L. (2004). Enzyme assays for high-throughput screening. ***Curr Opin Biotechnol***, 15(4): 314-322.

Griffiths, A. D., & Tawfik, D. S. (2003). Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. ***EMBO J***, 22(1): 24-35.

Griffiths, A. D., & Tawfik, D. S. (2006). Miniaturising the laboratory in emulsion droplets. ***Trends Biotechnol***, 24(9): 395-402.

Güven, G., Prodanovic, R., & Schwaneberg, U. (2010). Protein Engineering - An Option for Enzymatic Biofuel Cell Design. ***Electroanal***, 22(7-8): 765-+.

Hai, M., Bernath, K., Tawfik, D., & Magdassi, S. (2004). Flow cytometry: a new method to investigate the properties of water-in-oil-in-water emulsions. ***Langmuir***, 20(6): 2081-2085.

Hanahan, D., Jessee, J., & Bloom, F. R. (1991). Plasmid transformation of Escherichia coli and other bacteria. ***Methods Enzymol***, 204: 63-113.

Hong, Y. F., Liu, C. Y., Cheng, K. J., Hour, A. L., Chan, M. T., Tseng, T. H., Chen, K. Y., Shaw, J. F., & Yu, S. M. (2008). The sweet potato sporamin promoter confers high-level phytase expression and improves organic phosphorus acquisition and tuber yield of transgenic potato. ***Plant Mol Biol***, 67(4): 347-361.

Ilmberger, N., Meske, D., Juergensen, J., Schulte, M., Barthen, P., Rabausch, U., Angelov, A., Mientus, M., Liebl, W., Schmitz, R. A., & Streit, W. R. (2012). Metagenomic cellulases highly tolerant towards the presence of ionic liquids--linking thermostability and halotolerance. ***Appl Microbiol Biotechnol.***, 95(1): 135-146.

Jackson, A. M., Boutell, J., Cooley, N., & He, M. (2004). Cell-free protein synthesis for proteomics. ***Brief Funct Genomic Proteomic***, 2(4): 308-319.

Jaeger, K. E., & Eggert, T. (2004). Enantioselective biocatalysis optimized by directed evolution. ***Curr Opin Biotechnol***, 15(4): 305-313.

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. ***J Mol Biol***, 292(2): 195-202.

Kang, S. H., Kim, D. M., Kim, H. J., Jun, S. Y., Lee, K. Y., & Kim, H. J. (2005). Cell-free production of aggregation-prone proteins in soluble and active forms. ***Biotechnol Prog***, 21(5): 1412-1419.

Kawarasaki, Y., Griswold, K. E., Stevenson, J. D., Selzer, T., Benkovic, S. J., Iverson, B. L., & Georgiou, G. (2003). Enhanced crossover SCRATCHY: construction and high-throughput screening of a combinatorial library containing multiple non-homologous crossovers. ***Nucleic Acids Res***, 31(21): e126.

Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T., & Yokoyama, S. (1999). Cell-free production and stable-isotope labeling of milligram quantities of proteins. ***FEBS Lett***, 442(1): 15-19.

Kintses, B., van Vliet, L. D., Devenish, S. R., & Hollfelder, F. (2010). Microfluidic droplets: new integrated workflows for biological experiments. ***Curr Opin Chem Biol***, 14(5): 548-555.

Konietzny, U. G., R. (2004). Bacterial phytase: potential application, in vivo function and regulation of its synthesis. ***Braz. J Microbiol***, 35: 11-18.

Krieger, E., Koraimann, G., & Vriend, G. (2002). Increasing the precision of comparative models with YASARA NOVA--a self-parameterizing force field. ***Proteins***, 47(3): 393-402.

transcription-translation of the gene library into enzyme variants, active variants convert fluorogenic substrate into fluorescent product thus labelling water-in-oil (w/o) single emulsions. In order to enable subsequent ultrahigh throughput analysis and sorting by flow cytometer in an aqueous environment, (w/o) single emulsions are dispersed in an external water phase resulting in water-in-oil-in-water (w/o/w) emulsions (Bernath *et al.*, 2005; Ruff *et al.*, 2012; Taly *et al.*, 2007). IVC technology is completely independent of host cells, enabling the production of toxic and membrane bound proteins (Merk *et al.*, 2015; Worst *et al.*, 2015). Additionally, cell-free technology overcomes the challenge of diversity loss in directed evolution experiments due to low transformation efficiency of expression hosts (*i.e. E.coli* or *Bacillus*) since the DNA material is directly added in the inner aqueous phase of the compartment (Tu *et al.*, 2011). Several criteria must be fulfilled in order to achieve successful qualitative discrimination between active and inactive enzyme variants, by using *in vitro* enzyme compartmentalization-based flow cytometry screening platform: (1) reduced diffusion rates of fluorogenic substrate and/or fluorescent-based product from (w/o/w) emulsion compartments, (2) reduced diffusion rates of cell-free reaction components from (w/o/w) emulsion compartments, (3) monodispersity of (w/o/w) emulsion, and (4) stability of (w/o/w) emulsion compartments.

In order to establish the proposed flow cytometry-based uHTS-IVC screening platform and to make it applicable in directed evolution campaigns it is necessary, in the first step, to develop a flow cytometer-based ultrahigh throughput screening combined with *in vitro* compartmentalization (IVC) technology. In the second step, one or two industrial important enzymes (*i.e.* phytase and/or cellulase) have to be optimized for desired properties, *e.g.* activity, by application of the established screening platform as a proof of concept. The strategy to develop and validate the flow cytometer-based uHTS-IVC technology platform comprises the following steps: (1) development of an efficient protocol for emulsion generation, (2) selection of the fluorescent assays suitable for flow cytometric application in emulsions, (3) optimization of *in vitro* expression for enzymes (phytase and/or cellulase) in emulsion droplets, (4) sorting of populations with high mutational load and low numbers of active variants and (5) performing the directed evolution experiment for phytase and/or cellulase as a proof of principle for validation the novel developed technology platform. In total up to now, only three success stories for directed evolution of β-galactosidase (Mastrobattista *et al.*, 2005), [FeFe] Hydrogenase (Stapleton *et al.*, 2010) and

References

Meng, F., Engbers, G. H., & Feijen, J. (2005). Biodegradable polymersomes as a basis for artificial cells: encapsulation, release and targeting. *J Control Release*, 101(1-3): 187-198.

Merk, H., Rues, R. B., Gless, C., Beyer, K., Dong, F., Dotsch, V., Gerrits, M., & Bernhard, F. (2015). Biosynthesis of membrane dependent proteins in insect cell lysates: identification of limiting parameters for folding and processing. *J Biol Chem*, 396(9-10): 1097-1107.

Miller, O. J., Bernath, K., Agresti, J. J., Amitai, G., Kelly, B. T., Mastrobattista, E., Taly, V., Magdassi, S., Tawfik, D. S., & Griffiths, A. D. (2006). Directed evolution by in vitro compartmentalization. *Nat Methods*, 3(7): 561-570.

Mills, D. R., Peterson, R. L., & Spiegelm.S. (1967). An Extracellular Darwinian Experiment with a Self-Duplicating Nucleic Acid Molecule. *Proc Natl Acad Sci U S A*, 58(1): 217-&.

Mullaney, E. J., Locovare, H., Sethumadhavan, K., Boone, S., Lei, X. G., & Ullah, A. H. (2010). Site-directed mutagenesis of disulfide bridges in Aspergillus niger NRRL 3135 phytase (PhyA), their expression in Pichia pastoris and catalytic characterization. *Appl Microbiol Biotechnol*, 87(4): 1367-1372.

Nakano, H., Kawarasaki, Y., & Yamane, T. (2004). Cell-free protein synthesis systems: increasing their performance and applications. *Adv Biochem Eng Biotechnol*, 90: 135-149.

Nakano, H., & Yamane, T. (1998). Cell-free protein synthesis systems. *Biotechnol Adv*, 16(2): 367-384.

Nirenberg, M. W., & Matthaei, J. H. (1961). The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proc Natl Acad Sci U S A*, 47: 1588-1602.

Nov, Y. (2012). When second best is good enough: another probabilistic look at saturation mutagenesis. *Appl Environ Microbiol*, 78(1): 258-262.

Okushima, S., Nisisako, T., Torii, T., & Higuchi, T. (2004). Controlled production of monodisperse double emulsions by two-step droplet breakup in microfluidic devices. *Langmuir*, 20(23): 9905-9908.

Olsen, M., Iverson, B., & Georgiou, G. (2000). High-throughput screening of enzyme libraries. *Curr Opin Biotechnol*, 11(4): 331-337.

Packer, M. S., & Liu, D. R. (2015). Methods for the directed evolution of proteins. *Nat Rev Genet*, 16(7): 379-394.

Passonneau, J. V., & Lowry, O. H. (1993). *Enzymatic analysis: a practical guide*: Springer Science & Business Media.

Percival Zhang, Y. H., Himmel, M. E., & Mielenz, J. R. (2006). Outlook for cellulase improvement: screening and selection strategies. *Biotechnol Adv*, 24(5): 452-481.

Pitzler, C., Wirtz, G., Vojcic, L., Hiltl, S., Boker, A., Martinez, R., & Schwaneberg, U. (2014). A fluorescent hydrogel-based flow cytometry high-throughput screening platform for hydrolytic enzymes. *Chem Biol*, 21(12): 1733-1742.

Prodanovic, R., Ostafe, R., Scacioc, A., & Schwaneberg, U. (2011). Ultrahigh-throughput screening system for directed glucose oxidase evolution in yeast cells. *Comb Chem High Throughput Screen*, 14(1): 55-60.

Qiu, J., & Elber, R. (2006). SSALN: An alignment algorithm using structure-dependent substitution matrices and gap penalties learned from structurally aligned protein pairs. *Proteins: Struct Funct Bioinf*, 62(4): 881-891.

Quast, R. B., Kortt, O., Henkel, J., Dondapati, S. K., Wustenhagen, D. A., Stech, M., & Kubick, S. (2015a). Automated production of functional membrane proteins using eukaryotic cell-free translation systems. *J Biotechnol*, 203: 45-53.

Quast, R. B., Mrusek, D., Hoffmeister, C., Sonnabend, A., & Kubick, S. (2015b). Cotranslational incorporation of non-standard amino acids using cell-free protein synthesis. *FEBS Lett*, 589(15): 1703-1712.

Rao, D. E., Rao, K. V., Reddy, T. P., & Reddy, V. D. (2009). Molecular characterization, physicochemical properties, known and potential applications of phytases: An overview. *Crit Rev Biotechnol*, 29(2): 182-198.

Reetz, M. T., & Carballeira, J. D. (2007). Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat Protoc*, 2(4): 891-903.

Reetz, M. T., & Jaeger, K. E. (1999). Superior biocatalysts by directed evolution. *Biocatalysis - from Discovery to Application*, 200: 31-57.

Reetz, M. T., Kahakeaw, D., & Lohmer, R. (2008). Addressing the numbers problem in directed evolution. *Chembiochem*, 9(11): 1797-1804.

Reetz, M. T., Prasad, S., Carballeira, J. D., Gumulya, Y., & Bocola, M. (2010). Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: rigorous comparison with traditional methods. *J Am Chem Soc*, 132(26): 9144-9152.

Reetz, M. T., Wang, L. W., & Bocola, M. (2006). Directed evolution of enantioselective enzymes: iterative cycles of CASTing for probing protein-sequence space. *Angew Chem Int Ed Engl*, 45(8): 1236-1241.

Reymond, J.-L., Fluxa, V. S., & Maillard, N. (2008). Enzyme assays. *Chem Commun* (1): 46.

Roccatano, D., Wong, T. S., Schwaneberg, U., & Zacharias, M. (2006). Toward understanding the inactivation mechanism of monooxygenase P450 BM-3 by organic cosolvents: a molecular dynamics simulation study. *Biopolymers*, 83(5): 467-476.

Ruff, A. J., Dennig, A., & Schwaneberg, U. (2013). To get what we aim for-progress in diversity generation methods. *Febs J*, 280(13): 2961-2978.

Ruff, A. J., Dennig, A., Wirtz, G., Blanusa, M., & Schwaneberg, U. (2012). Flow Cytometer-Based High-Throughput Screening System for Accelerated Directed Evolution of P450 Monooxygenases. *ACS Catal*, 2(12): 2724-2728.

Sachse, R., Wustenhagen, D., Samalikova, M., Gerrits, M., Bier, F. F., & Kubick, S. (2013). Synthesis of membrane proteins in eukaryotic cell-free systems. *Engineering in Life Sciences*, 13(1): 39-48.

Sambrook, J., & Russell, D. W. (2001). *Molecular cloning: a laboratory manual* (3rd ed.). Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Sandberg, A.-S. (2002). Bioavailability of minerals in legumes. *Br J Nutr*, 88: S281-S285.

Sandberg, A. S., Hulthen, L. R., & Turk, M. (1996). Dietary Aspergillus niger phytase increases iron absorption in humans. *J Nutr*, 126(2): 476-480.

Sapan, C. V., Lundblad, R. L., & Price, N. C. (1999). Colorimetric protein assay techniques. *Biotechnol Appl Biochem*, 29(2): 99-108.

Sawasaki, T., Ogasawara, T., Morishita, R., & Endo, Y. (2002). A cell-free protein synthesis system for high-throughput proteomics. *Proc Natl Acad Sci U S A*, 99(23): 14652-14657.

Schmidt-Dannert, C., & Arnold, F. H. (1999). Directed evolution of industrial enzymes. *Trends Biotechnol*, 17(4): 135-136.

Seyfang, A., & Jin, J. H. Q. (2004). Multiple site-directed mutagenesis of more than 10 sites simultaneously and in a single round. *Anal Biochem*, 324(2): 285-291.

Shah, R. K., Shum, H. C., Rowat, A. C., Lee, D., Agresti, J. J., Utada, A. S., Chu, L.-Y., Kim, J.-W., Fernandez-Nieves, A., Martinez, C. J., & Weitz, D. A. (2008). Designer emulsions using microfluidics. *Mater Today*, 11(4): 18-27.

Shen, Z., Qu, W., Wang, W., Lu, Y., Wu, Y., Li, Z., Hang, X., Wang, X., Zhao, D., & Zhang, C. (2010). MPprimer: a program for reliable multiplex PCR primer design. *BMC Bioinform*, 11: 143.

Shivange, A. V., Marienhagen, J., Mundhada, H., Schenk, A., & Schwaneberg, U. (2009). Advances in generating functional diversity for directed protein evolution. *Curr Opin Chem Biol*, 13(1): 19-25.

Shivange, A. V., Schwaneberg, U., & Roccatano, D. (2010). Conformational dynamics of active site loop in Escherichia coli phytase. *Biopolymers*, 93(11): 994-1002.

Shivange, A. V., Serwe, A., Dennig, A., Roccatano, D., Haefner, S., & Schwaneberg, U. (2012). Directed evolution of a highly active Yersinia mollaretii phytase. *Appl Microbiol Biotechnol*, 95(2): 405-418.

Singh-Blom, A., Hughes, R. A., & Ellington, A. D. (2014). An amino acid depleted cell-free protein synthesis system for the incorporation of non-canonical amino acid analogs into proteins. *J Biotechnol*, 178: 12-22.

Skhiri, Y., Gruner, P., Semin, B., Brosseau, Q., Pekin, D., Mazutis, L., Goust, V., Kleinschmidt, F., El Harrak, A., Hutchison, J. B., Mayot, E., Bartolo, J. F., Griffiths, A. D., Taly, V., & Baret, J. C. (2012). Dynamics of molecular transport by surfactants in emulsions. *Soft Matter*, 8(41): 10618-10627.

Spirin, A. S. (2004). High-throughput cell-free systems for synthesis of functionally active proteins. *Trends Biotechnol*, 22(10): 538-545.

Stapleton, J. A., & Swartz, J. R. (2010). Development of an in vitro compartmentalization screen for high-throughput directed evolution of [FeFe] hydrogenases. *PLoS One*, 5(12): e15275.

Stech, M., & Kubick, S. (2015). Cell-Free Synthesis Meets Antibody Production: A Review. *Antibodies*, 4(1): 12-33.

Stemmer, W., & Holland, B. (2003). Survival of the fittest molecule. *Am Sci*, 91(6): 526-533.

Stemmer, W. P. (1994). Rapid evolution of a protein in vitro by DNA shuffling. *Nature*, 370(6488): 389-391.

Sunami, T., Sato, K., Matsuura, T., Tsukada, K., Urabe, I., & Yomo, T. (2006). Femtoliter compartment in liposomes for in vitro selection of proteins. *Anal Biochem*, 357(1): 128-136.

Taly, V., Kelly, B. T., & Griffiths, A. D. (2007). Droplets as microreactors for high-throughput biology. *Chembiochem*, 8(3): 263-272.

Tawfik, D. S., & Griffiths, A. D. (1998). Man-made cell-like compartments for molecular evolution. *Nat Biotechnol*, 16(7): 652-656.

Tee, K. L., & Wong, T. S. (2013). Polishing the craft of genetic diversity creation in directed evolution. *Biotechnol Adv*, 31(8): 1707-1721.

Teeri, T. T. (1997). Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. *Trends Biotechnol*, 15(5): 160-167.

Tu, R. (2010). *Flow cytometry based screening systems for directed evolution of proteases*. Unpublished Dissertation, Jacobs University, Bremen, Germany, School of Engineering and Science, Jacobs University, Bremen, Germany.

Tu, R., Martinez, R., Prodanovic, R., Klein, M., & Schwaneberg, U. (2011). A flow cytometry-based screening system for directed evolution of proteases. *J Biomol Screen*, 16(3): 285-294.

References

Tye, A. J., Siu, F. K., Leung, T. Y., & Lim, B. L. (2002). Molecular cloning and the biochemical characterization of two novel phytases from B. subtilis 168 and B. licheniformis. *Appl Microbiol Biotechnol*, 59(2-3): 190-197.

Van Etten, R. L., Davidson, R., Stevis, P. E., MacArthur, H., & Moore, D. L. (1991). Covalent structure, disulfide bonding, and identification of reactive surface and active site residues of human prostatic acid phosphatase. *J Biol Chem*, 266(4): 2313-2319.

Vincent, J. B., Crowder, M. W., & Averill, B. A. (1992). Hydrolysis of phosphate monoesters: a biological problem with multiple chemical solutions. *Trends Biochem Sci*, 17(3): 105-110.

Wang, Q., Fu, S.-J., Sun, J.-Y., & Weng, X.-Y. (2011). Characterization of a thermostable alkaline phytase from Bacillus licheniformis ZJ-6 in Pichia pastoris. *World J Microbiol Biotechnol*, 27(5): 1247-1253.

Wang, W., & Malcolm, B. A. (1999). Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange Site-Directed Mutagenesis. *Biotechniques*, 26(4): 680-682.

Wang, W., & Malcolm, B. A. (2002). Two-stage polymerase chain reaction protocol allowing introduction of multiple mutations, deletions, and insertions, using QuikChange site-directed mutagenesis. *Methods Mol Biol*, 182: 37-43.

Wong, T. S., Roccatano, D., & Schwaneberg, U. (2007). Steering directed protein evolution: strategies to manage combinatorial complexity of mutant libraries. *Environ Microbiol*, 9(11): 2645-2659.

Wong, T. S., Tee, K. L., Hauer, B., & Schwaneberg, U. (2004). Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. *Nucleic Acids Res*, 32(3): e26.

Wong, T. S., Zhurina, D., & Schwaneberg, U. (2006). The diversity challenge in directed protein evolution. *Comb Chem High Throughput Screen*, 9(4): 271-288.

Worst, E. G., Exner, M. P., De Simone, A., Schenkelberger, M., Noireaux, V., Budisa, N., & Ott, A. (2015). Cell-free expression with the toxic amino acid canavanine. *Bioorg Med Chem Lett*, 25(17): 3658-3660.

Wyss, M., Pasamontes, L., Friedlein, A., Remy, R., Tessier, M., Kronenberger, A., Middendorf, A., Lehmann, M., Schnoebelen, L., Rothlisberger, U., Kusznir, E., Wahl, G., Muller, F., Lahm, H. W., Vogel, K., & van Loon, A. P. (1999). Biophysical characterization of fungal phytases (myo-inositol hexakisphosphate phosphohydrolases): molecular size, glycosylation pattern, and engineering of proteolytic resistance. *Appl Environ Microbiol*, 65(2): 359-366.

Yao, M. Z., Zhang, Y. H., Lu, W. L., Hu, M. Q., Wang, W., & Liang, A. H. (2012). Phytases: crystal structures, protein engineering and potential biotechnological applications. *J Appl Microbiol*, 112(1): 1-14.

Zemella, A., Thoring, L., Hoffmeister, C., & Kubick, S. (2015). Cell-Free Protein Synthesis: Pros and Cons of Prokaryotic and Eukaryotic Systems. *Chembiochem*, 16(17): 2420-2431.

Zhang, R., Yang, P., Huang, H., Yuan, T., Shi, P., Meng, K., & Yao, B. (2011). Molecular and biochemical characterization of a new alkaline beta-propeller phytase from the insect symbiotic bacterium Janthinobacterium sp. TN115. *Appl Microbiol Biotechnol*, 92(2): 317-325.

Zhao, H., Giver, L., Shao, Z., Affholter, J. A., & Arnold, F. H. (1998). Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat Biotechnol*, 16(3): 258-261.

phosphotriesterase (Griffiths *et al.*, 2003), employing IVC-based flow cytometry in (w/o/w) emulsions have been reported.

In this thesis, for the first time the development of an uHTS technology platform-based on IVC (uHTS-IVC) in (w/o/w) emulsion for directed cellulase (CelA2-H288F) evolution will be presented. The cell-free transcribed and translated cellulase mutant libraries will be encapsulated together with fluorescein-di-β-D-cellobioside (FDC) as substrate within (w/o/w) emulsion compartments. Translated active cellulase variants catalyze the cleavage of the β-1,4-D-glycosodic bonds and the resulting fluorescein will be detected by fluorescence measurements (λ_{ex} 494 nm/λ_{em} 516 nm) using flow cytometry to discriminate between active and inactive cellulase variants. The validation of the developed flow cytometry-based uHTS-IVC platform will be performed in form of a CelA2-directed evolution campaign by screening an error-prone PCR (epPCR) cellulase mutant library using linear DNA template. In addition, three different types of focused mutagenesis (*e.g.* OmniChange) on positions in close proximity to the active site of cellulase will be screened using uHTS-IVC platform, by using *in vitro* transcription-translation of plasmid DNA, and combinatorial effects will be investigated.

Additionally, the highly active *Yersinia mollaretii* phytase will be evolved for higher activity at neutral pH (*i.e.* pH 6.6), in order to broaden the application potential of this acid phosphatase in various industries (*e.g.* feed industry, agriculture) and to make it applicable in cell-free expression systems employed in the flow cytometry-based uHTS-IVC screening platform. Therefore, the pH profile of the phytase has to be expanded to neutral pH which will be realized through a directed evolution campaign using epPCR as random mutagenesis method for diversity generation, followed by focused mutagenesis using SSM of the identified beneficial positions from the random approach.

Compared to existing *in vivo* and *in vitro* flow cytometer-based screening systems, the novel flow cytometry-based uHTS-IVC screening platform inaugurates applications for various enzymes, especially hydrolases, overcoming the limitations of previous flow cytometer-based screening systems, *i.e.* leakage of substrate/product due to cell membrane permeability, it uses completely cell-free expression systems instead of frequently used lysed cells in the compartments, no additional microbeads entrapped together with the protein in the compartment, no complicated cell surface display technology, and it is completely independent of host cells fostering the production of toxic or membrane bound

Appendix
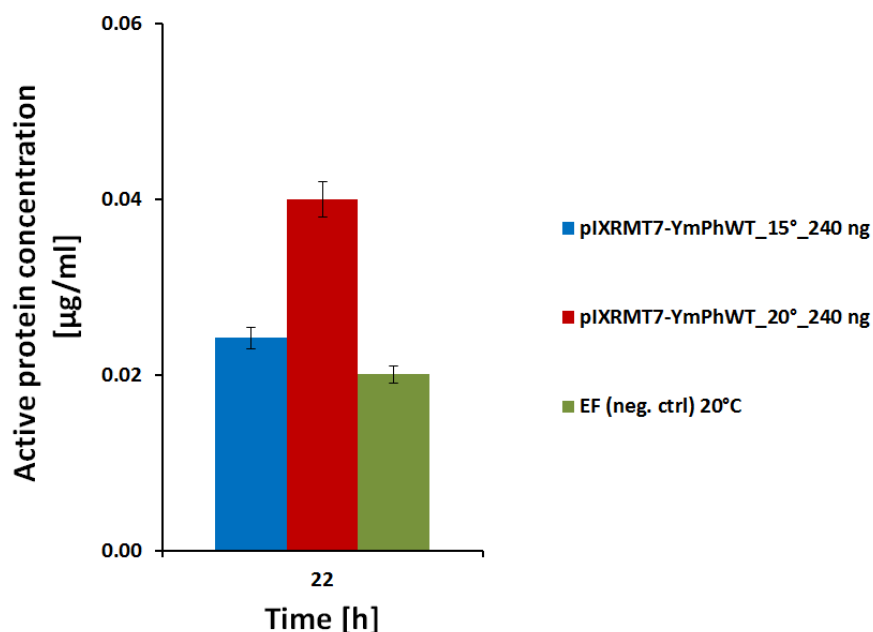
Figure A1: *In vitro* expression of pIX3.0RMT7-YmPhWT after incubation at 15-20 °C (240 ng template DNA pIX3.0RMT7-YmPhWT). Samples were analyzed with 384-well MTP-based 4-MUP activity assay after 22 h and obtained active protein concentrations were calculated using calibration curve of YmPhWT activity compared to protein concentration.
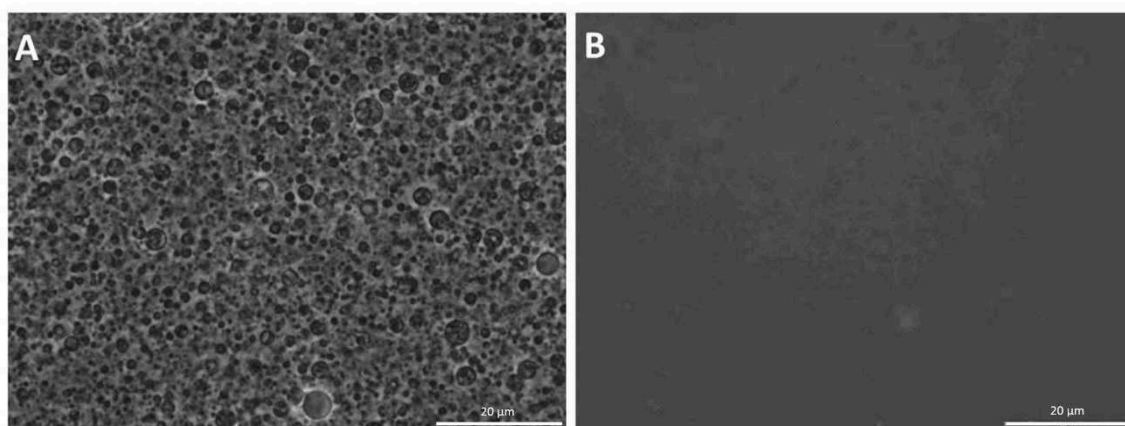

Figure A2: (W/o/w) emulsion generation using membrane extrusion method employing the decane/Tween emulsification system. Samples were extruded 14 times. (**A**) Transmission microscopy picture of the generated (w/o/w) emulsions (**B**) fluorescence microscopy picture of the generated (w/o/w) emulsions. The generated emulsions were small (<5 µm), not stable and precipitation was observed at temperatures lower than 37°C resulting in a release of the entrapped fluorescein dye causing a high background fluorescence. The scale bar represents a length of 20 µm.
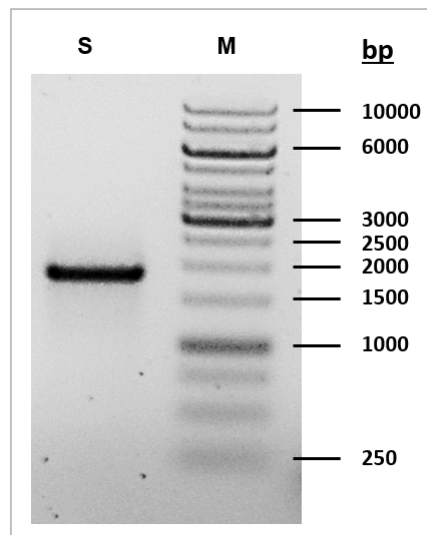
Figure A3: Results of sorted, recovered DNA of a directed evolution campaign of cellulase using uHTS-IVC technology platform**.** Agarose-gelelectrophoresis picture of recovered and amplified PCR products after DNA isolation from sorted (w/o/w) emulsions using NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG). **M** represents the marker (GeneRuler[TM] 1 kb DNA Ladder, Life Technologies GmbH, Darmstadt, Germany) used as reference. **S** represents the recovered amplified DNA (~1.8 kb) from the active emulsion fractions sorted with flow cytometer from an epPCR library (0.05 mM MnCl$_2$). The DNA was isolated with NucleoSpin® Gel and PCR clean up kit (Machery-Nagel).

The figure is part of the published manuscript "Körfer, G., Pitzler C., Vojcic L., Martinez R., and Schwaneberg U.; *In vitro* flow cytometry-based screening platform for cellulase engineering, Scientific Reports (2016); 6: 26128" and was reprinted with permission of the Nature Publishing Group.
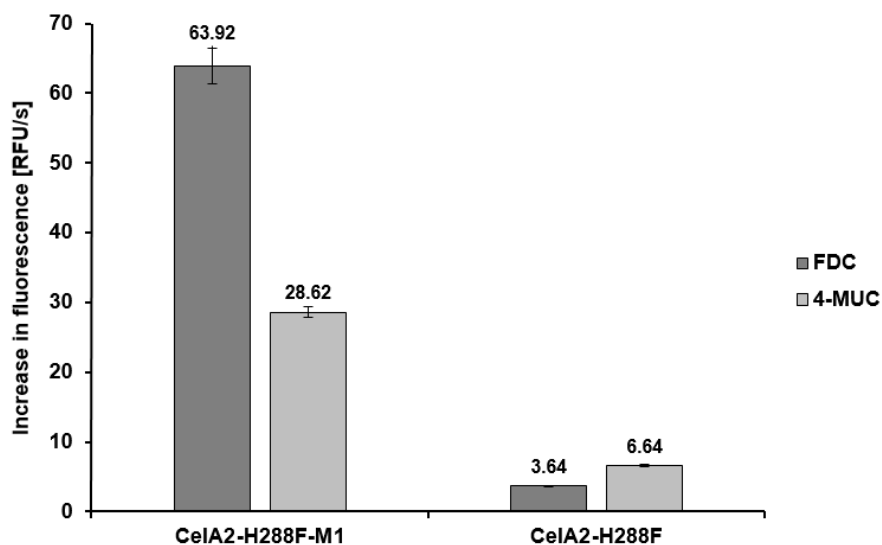
Figure A4: Results of CelA2-H288F-M1 and parent CelA2-H288F analysis with FDC and 4-MUC activity assay in MTP format using crude cell extracts. The x-axis describes the two different variants and the y-axis shows the increase in fluorescence in relative fluorescent units (RFU) per s. The reported values are the average of two measurements and the shown average deviations are calculated from the mean values. For the activity assay crude cell extract (40 µl) was mixed with FDC or 4-MUC substrate (0.05 mM; 20 µl) and potassium phosphate buffer (pH 7.2, 0.2 M; 40 µl).

The figure is part of the published manuscript "Körfer, G., Pitzler C., Vojcic L., Martinez R., and Schwaneberg U.; *In vitro* flow cytometry-based screening platform for cellulase engineering, Scientific Reports (2016); 6: 26128" and was reprinted with permission of the Nature Publishing Group.
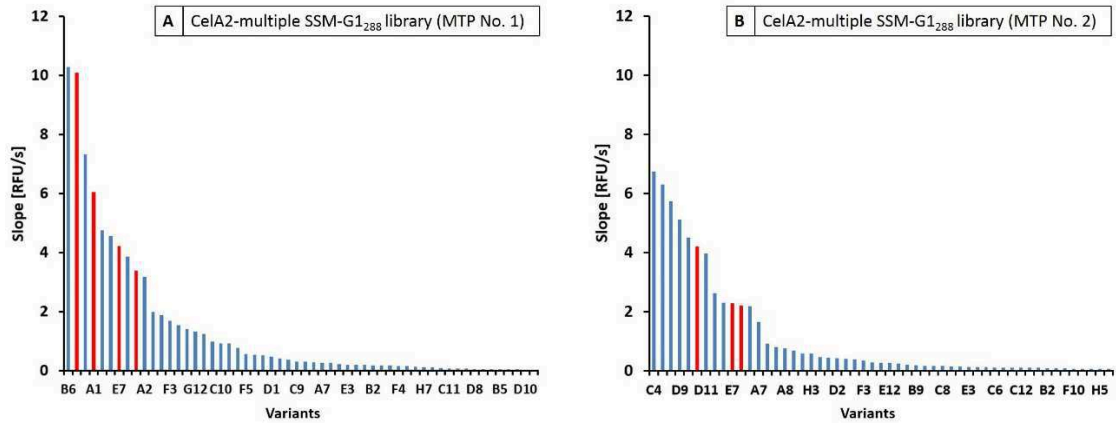
Figure A5: Analysis of CelA2-multiple SSM-G1$_{288}$ library (**A**) MTP No. 1, (**B**) MTP No. 2 with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s).
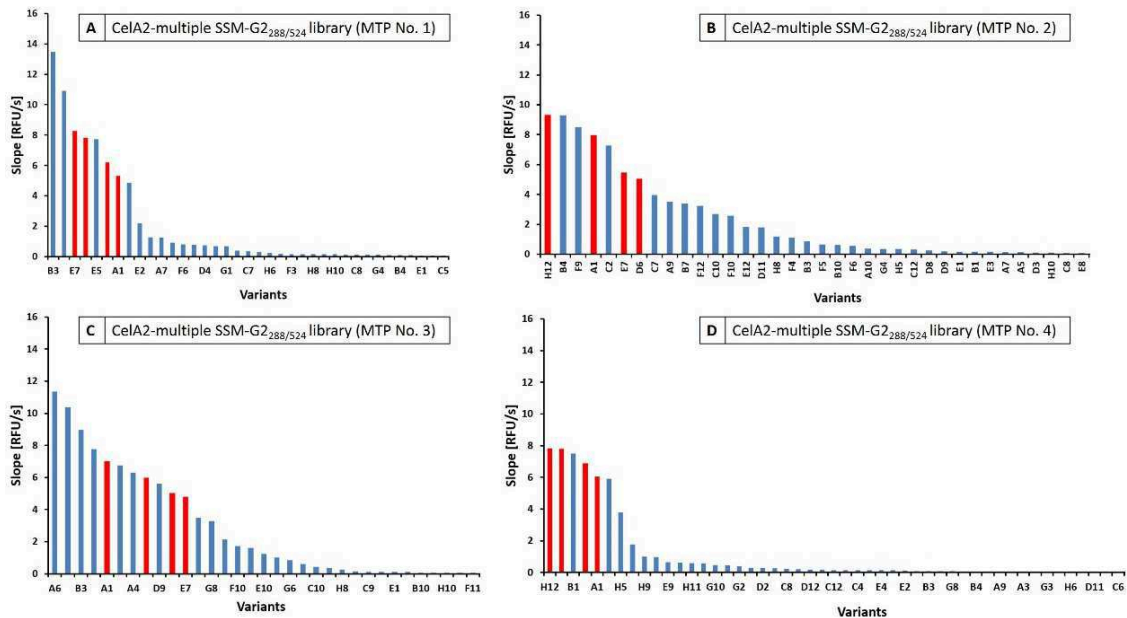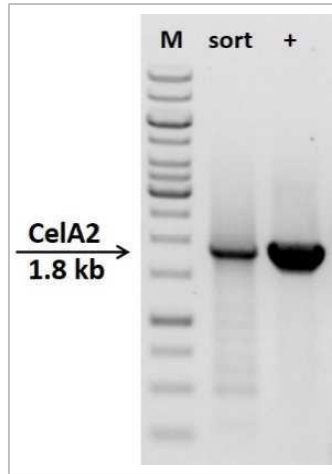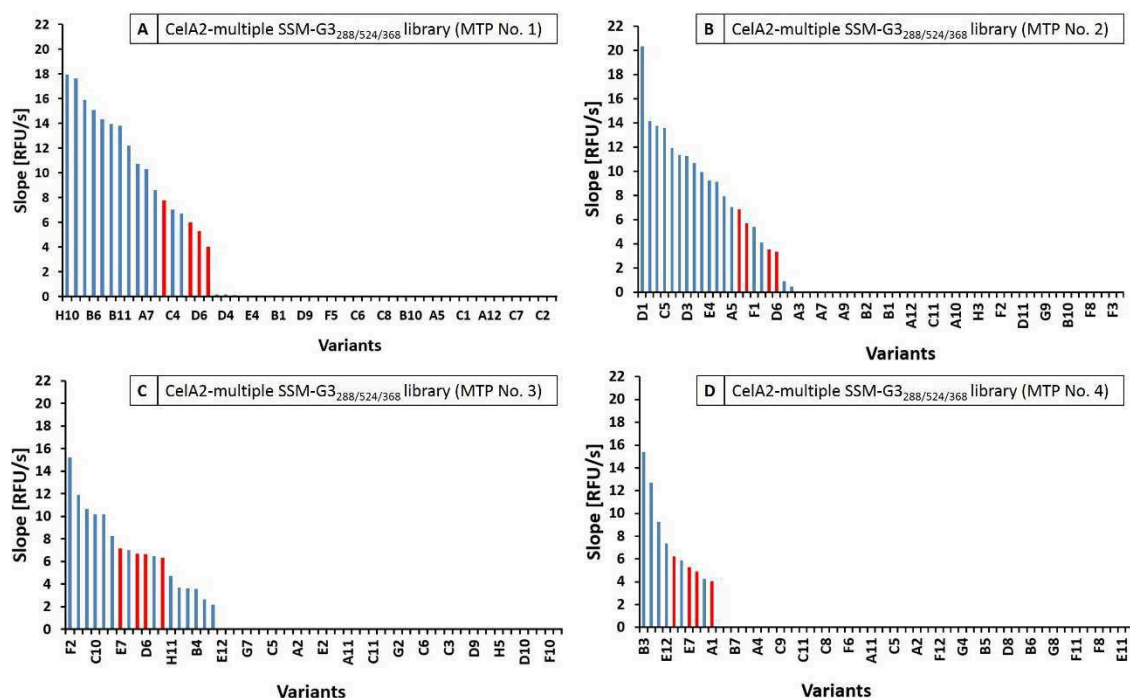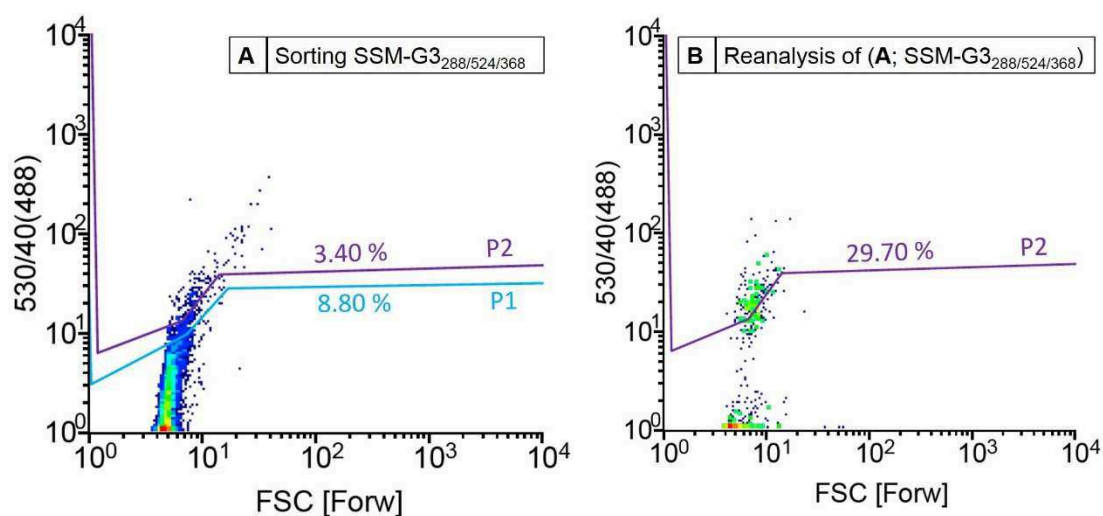

Figure A6: Analysis of CelA2-multiple SSM-G2$_{288/524}$ library (**A**) MTP No. 1, (**B**) MTP No. 2, (**C**) MTP No. 3, and (**D**) MTP No. 4 with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s).

Figure A7: Agarose-gelelectrophoresis picture of recovered sorted CelA2-multiple SSM-G2$_{288/524}$ library PCR products at ~1.8 kb after DNA isolation from (w/o/w) emulsions using NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG). **M** represents the marker (GeneRuler^{TM} 1 kb DNA Ladder, Life Technologies GmbH, Darmstadt, Germany) used as reference; **sort** represents the sorted, recovered fraction of the CelA2-multiple SSM-G2$_{288/524}$ library; **+** represents the positive control using 10 ng of pIX3.0RMT7-CelA2-H288F plasmid DNA as template for the PCR reaction.
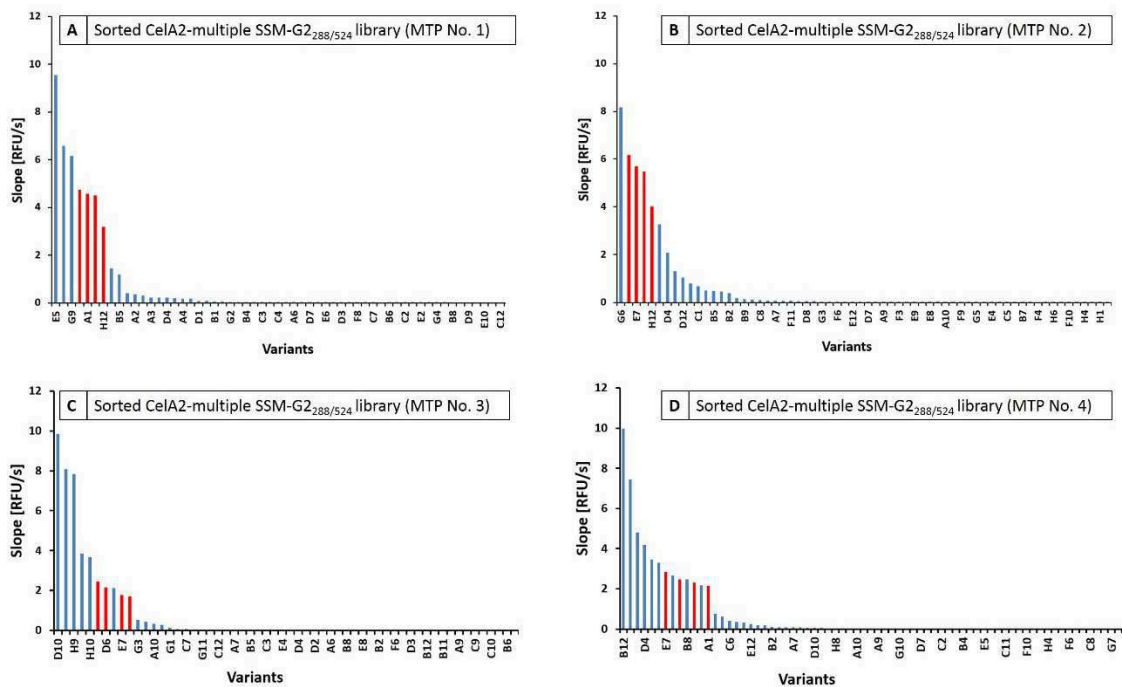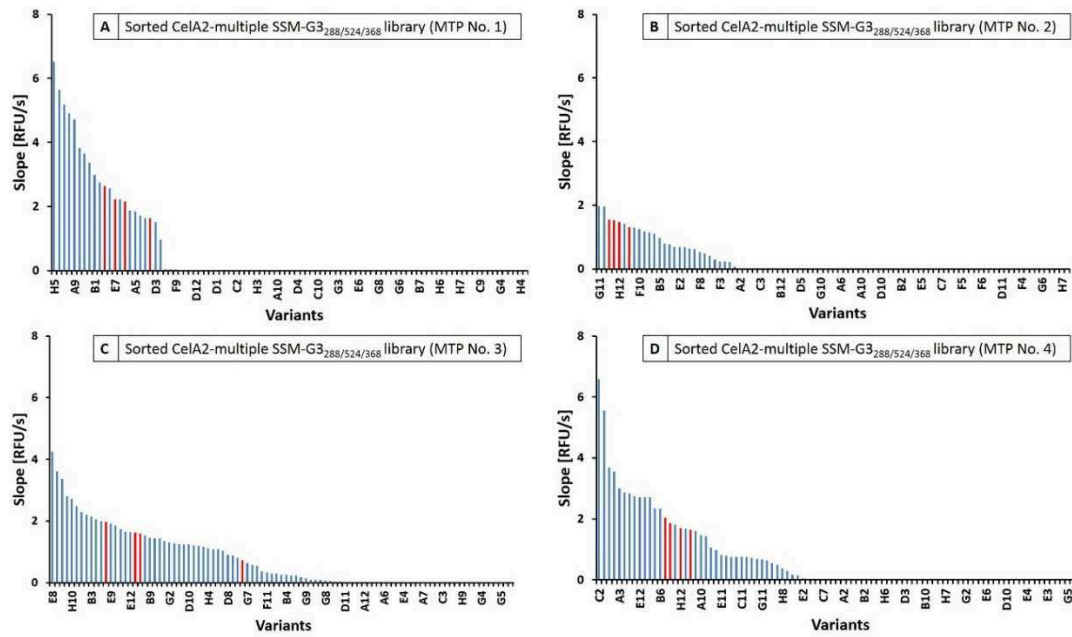

Figure A8: Analysis of the sorted CelA2-multiple SSM-G2$_{288/524}$ library (**A**) MTP No. 1, (**B**) MTP No. 2, (**C**) MTP No. 3, and (**D**) MTP No. 4 with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2 H288F, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s).
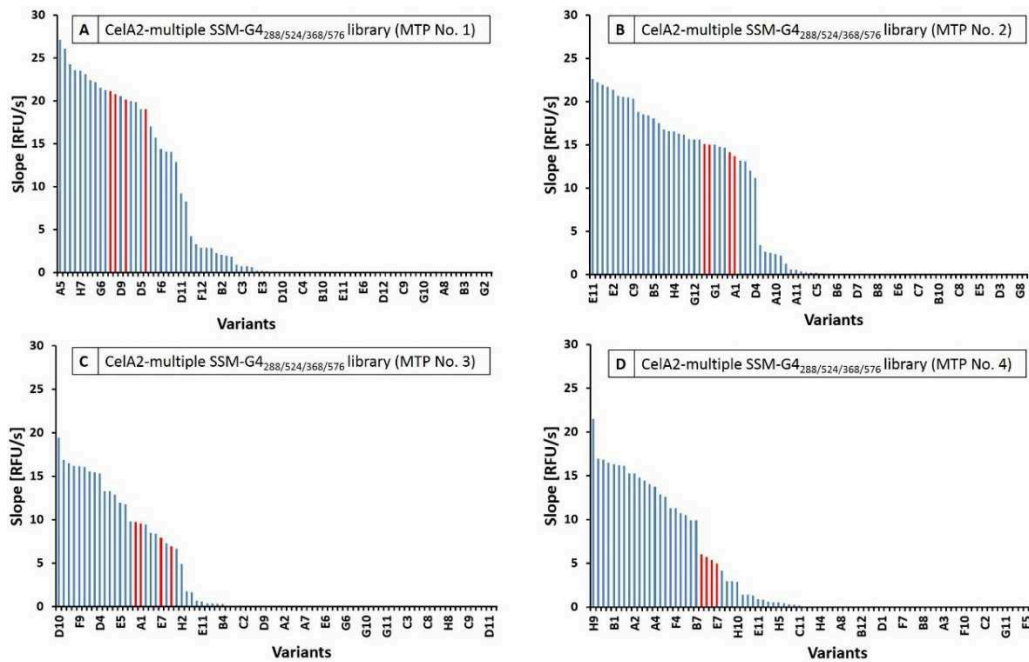
Figure A9: Analysis of CelA2-multiple SSM-G3$_{288/524/368}$ library (**A**) MTP No. 1, (**B**) MTP No. 2, (**C**) MTP No. 3, and (**D**) MTP No. 4 with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s).
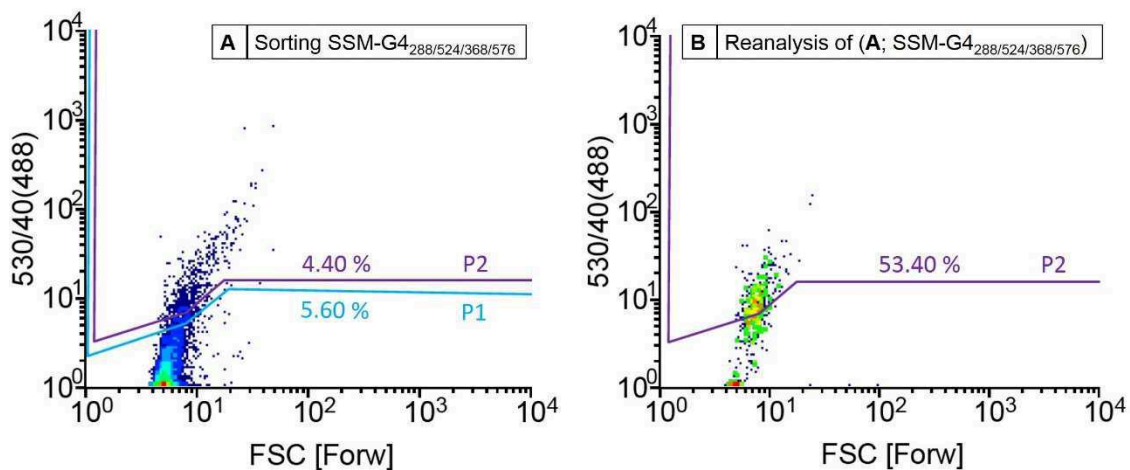

Figure A10: Flow cytometer analysis showing density plots of CelA2-multiple SSM-G3$_{288/524/368}$ library after *in vitro* expression in (w/o/w) emulsions (4 h, 30°C). The FSC represents the size of the emulsion on the x-axis and 530/40 (488) represents the fluorescence intensity (λ_{ex} 494 nm and λ_{em} 516 nm) of the fluorescein as reaction product on the y-axis. Lines indicate gate P1 separating fluorescent events (appearing above) from non-fluorescent events and sorting gate P2 separating the most fluorescent events (appearing above) from non-/weak-fluorescent events. (**A**) CelA2-multiple SSM-G3$_{288/524/368}$ library (8.80 % fluorescent events) during sorting of the 3.4 % most fluorescent events. (**B**) Reanalysis of sorted library with an 8.5-fold enriched active fraction (29.70 % fluorescent events).
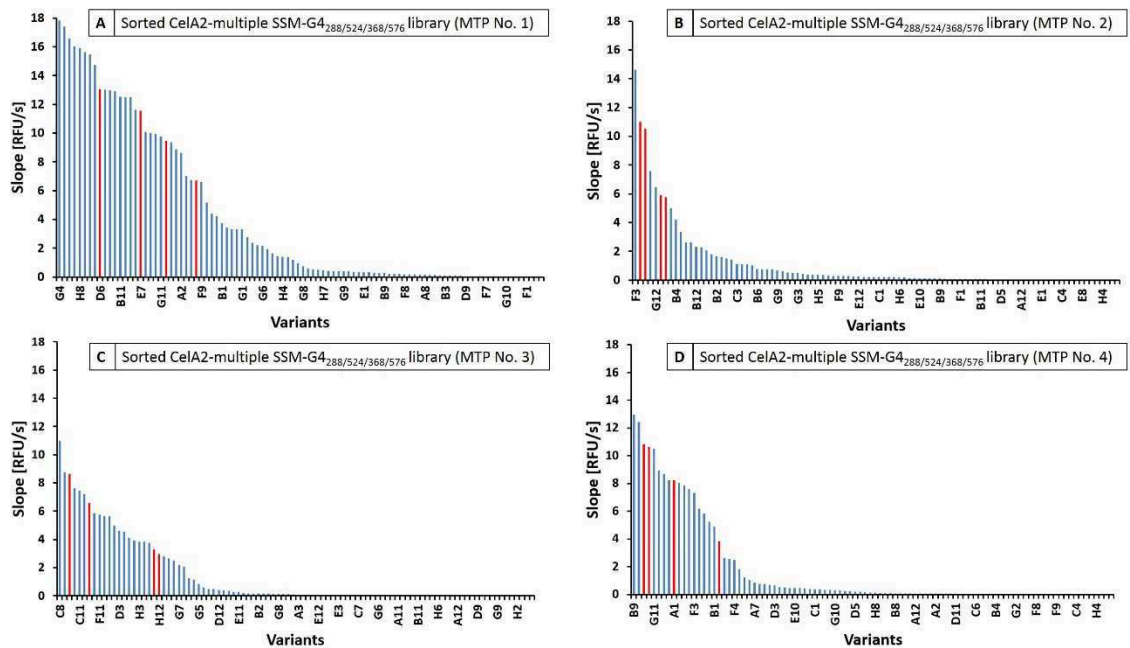
Figure A11: Analysis of the sorted CelA2-multiple SSM-G3$_{288/524/368}$ library (**A**) MTP No. 1, (**B**) MTP No. 2, (**C**) MTP No. 3, and (**D**) MTP No. 4 with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s).
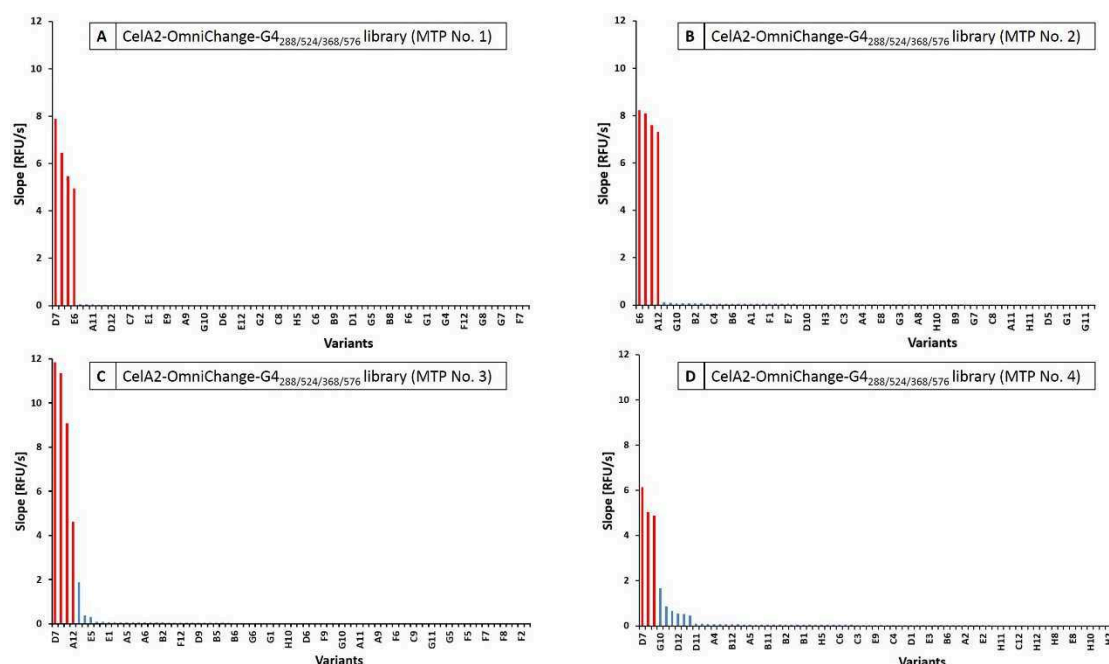

Figure A12: Analysis of CelA2-multiple SSM-G4$_{288/524/368/576}$ library (**A**) MTP No. 1, (**B**) MTP No. 2, (**C**) MTP No. 3, and (**D**) MTP No. 4 with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s).
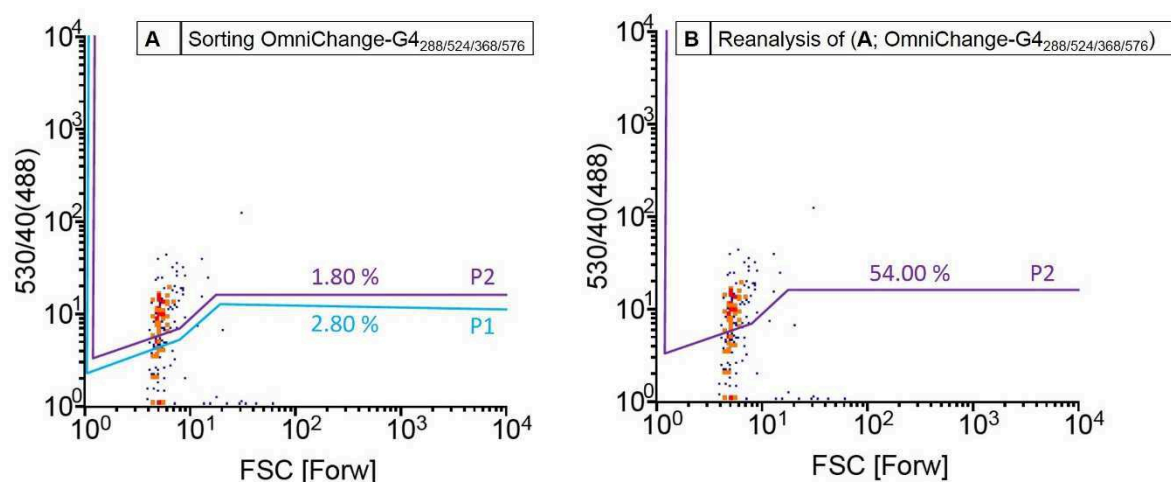
Figure A13: Flow cytometer analysis showing density plots of CelA2-multiple SSM-G4$_{288/524/368/576}$ library after *in vitro* expression in (w/o/w) emulsions (4 h, 30 °C). The FSC represents the size of the emulsion on the x-axis and 530/40 (488) represents the fluorescence intensity (λ_{ex} 494 nm and λ_{em} 516 nm) of the fluorescein as reaction product on the y-axis. Lines indicate gate P1 separating fluorescent events (appearing above) from non-fluorescent events and sorting gate P2 separating the most fluorescent events (appearing above) from non-/weak-fluorescent events. (**A**) CelA2-multiple SSM-G4$_{288/524/368/576}$ library (5.60 % fluorescent events) during sorting of the 4.40 % most fluorescent events. (**B**) Reanalysis of sorted library with a 12-fold enriched active fraction (53.40 % fluorescent events).
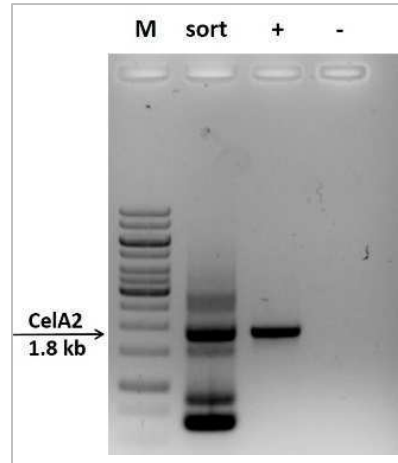

Figure A14: Analysis of the sorted CelA2-multiple SSM-G4$_{288/524/368/576}$ library (**A**) MTP No. 1, (**B**) MTP No. 2, (**C**) MTP No. 3, and (**D**) MTP No. 4 with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s).
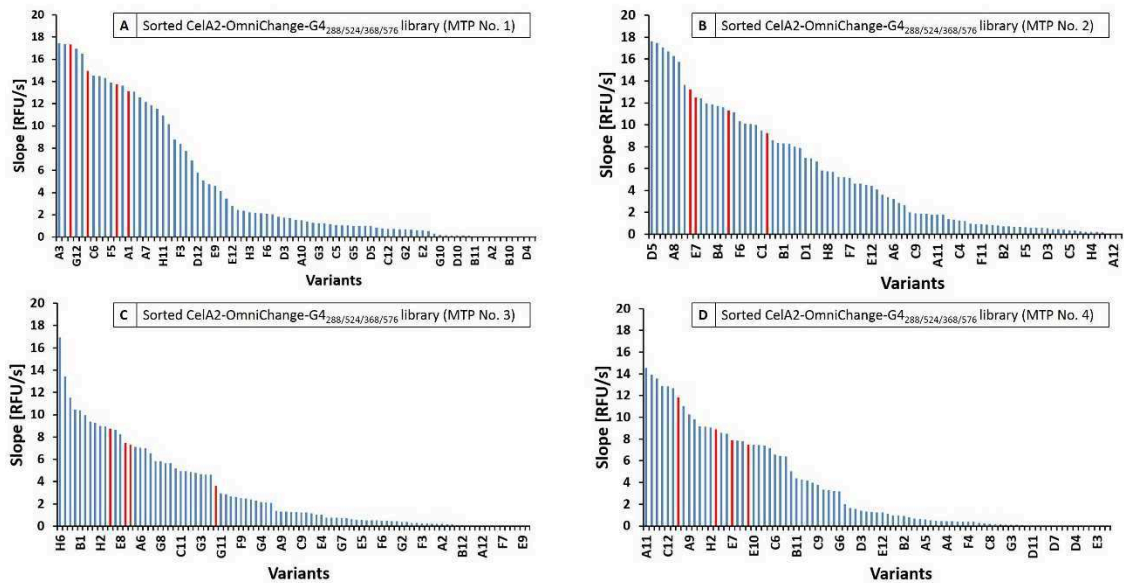
Figure A15: Analysis of CelA2-OmniChange-G4$_{288/524/368/576}$ library (**A**) MTP No. 1, (**B**) MTP No. 2, (**C**) MTP No. 3, and (**D**) MTP No. 4 with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s).


Figure A16: Flow cytometer analysis showing density plots of CelA2-OmniChange-G4$_{288/524/368/576}$ plasmid DNA library after *in vitro* expression in (w/o/w) emulsions (4 h, 30 °C). The FSC represents the size of the emulsion on the x-axis and 530/40 (488) represents the fluorescence intensity (λ_{ex} 494 nm and λ_{em} 516 nm) of the fluorescein as reaction product on the y-axis. Lines indicate gate P1 separating fluorescent events (appearing above) from non-fluorescent events and sorting gate P2 separating the most fluorescent events (appearing above) from non-/weak-fluorescent events. (**A**) CelA2-OmniChange-G4$_{288/524/368/576}$ plasmid DNA library during sorting of the 1.8 % most fluorescent events. (**B**) Reanalysis of sorted library with a 30-fold enriched active fraction (54.00 % fluorescent events).

Figure A17: Agarose-gelelectrophoresis picture of recovered sorted CelA2-OmniChange-G4$_{288/524/368/576}$ library PCR products at ~1.8 kb after DNA isolation from (w/o/w) emulsions using NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel GmbH & Co. KG). **M** represents the marker (GeneRuler^{TM} 1 kb DNA Ladder, Life Technologies GmbH, Darmstadt, Germany) used as reference; **sort** represents the sorted, recovered fraction of the CelA2-OmniChange-G4$_{288/524/368/576}$ library; **+** represents the positive control using 10 ng of pIX3.0RMT7-CelA2-H288F plasmid DNA as template for the PCR reaction; **-** represents the negative control using 10 ng of pIX3.0RMT7-CelA2-H288F plasmid DNA as template for the PCR reaction and no primers.

Figure A18: Analysis of the sorted CelA2-OmniChange-G4$_{288/524/368/576}$ library (**A**) MTP No. 1, (**B**) MTP No. 2, (**C**) MTP No. 3, and (**D**) MTP No. 4 with 4-MUC activity assay in 96-well MTP. The y-axis describes the activity towards the substrate 4-MUC as slope in RFU/s. The x-axis shows the different variants in the MTP. The variant CelA2-H288F, which served as template for the SSM is highlighted in red. The inactive variant CelA2-H288F-E580Q served as negative control and is highlighted in green (not visible as slopes are ~0.0 RFU/s).
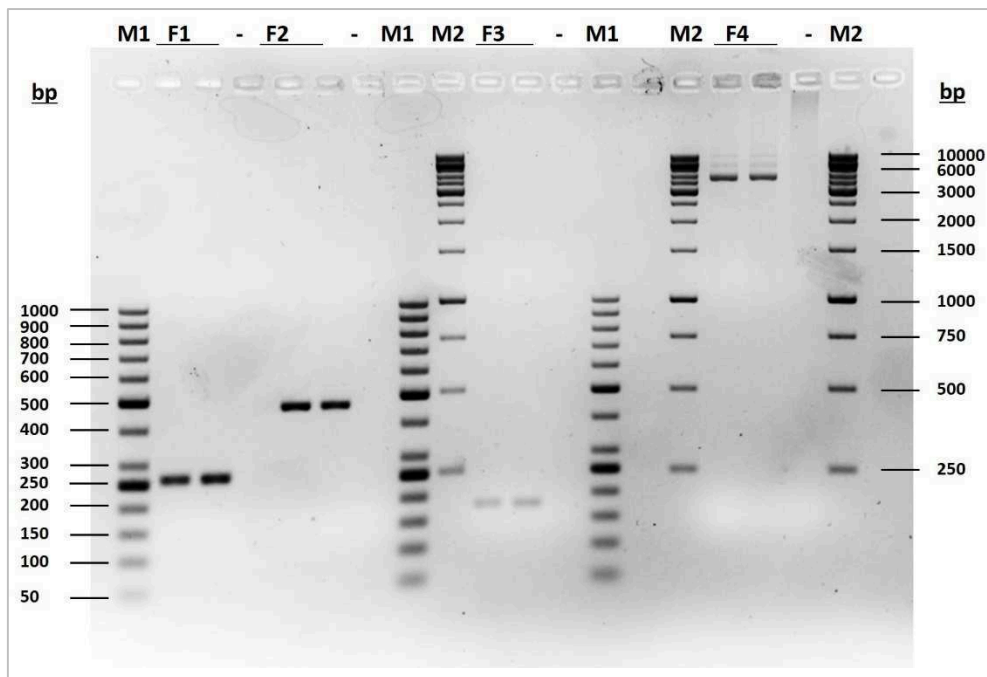
Figure A19: Agarose-gelelectrophoresis picture of the different fragments for generation of the CelA2 OmniChange-G4$_{288/524/368/576}$ library. Samples were analyzed on 2 % agarose-gel. **M1** represents the 50 bp marker (GeneRuler^{TM} 50 bp DNA Ladder, Life Technologies GmbH, Darmstadt, Germany) and **M2** represents the 1 kb marker (GeneRuler^{TM} 1 kb DNA Ladder, Life Technologies GmbH, Darmstadt, Germany) used as reference; **F1-F4** indicate the four different fragments for generation of the CelA2-OmniChange-G4$_{288/524/368/576}$ library (F1: 258 bp, F2: 474 bp, F3: 173 bp, F4: 3840 bp); **-** represents the negative control using 10 ng of the respective sample as template for the PCR reaction but no primers.
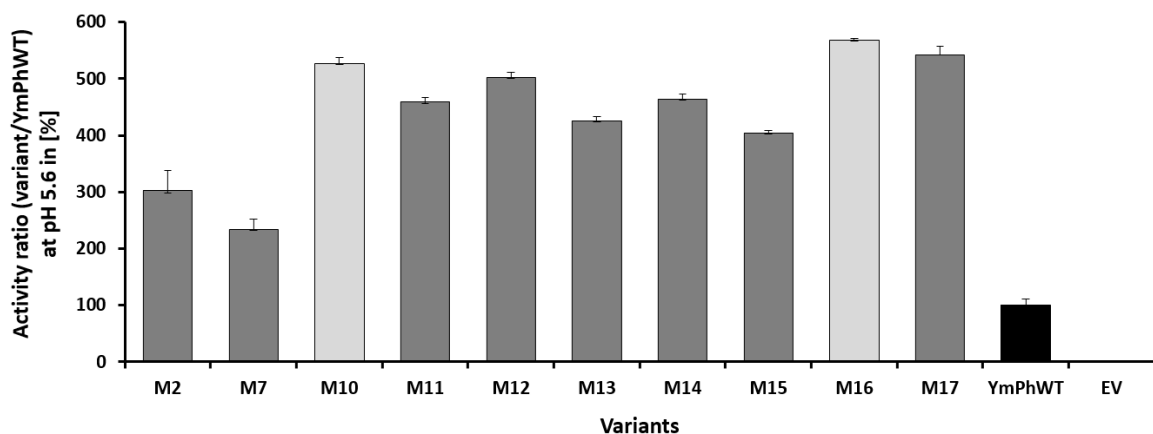
Figure A20: Results of the YmPh-SSM-44/45 library rescreening performed with selected variants YmPh-M10-M17, after analysis with 4-MUP activity assay at pH 5.6. The x-axis shows the different variants in comparison to YmPhWT, -M2, -M7 and empty vector (EV). The y-axis shows the mean values of the obtained activity ratios at pH 5.6 in [%] which were calculated from five measurements (EV and YmPhWT four measurements).
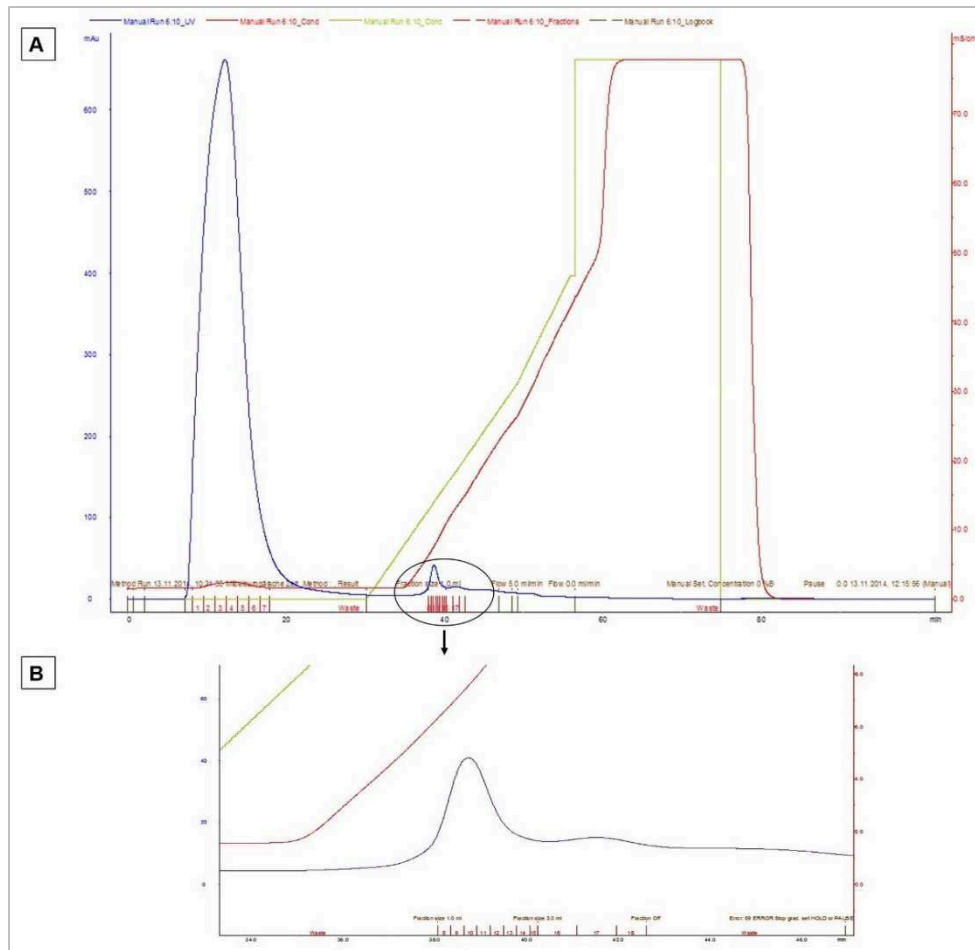
Figure A21: Chromatogram of the cation-exchange chromatography with YmPh-M10 as example. (**A**) Chromatogram of YmPh-M10 purification. The phytase peak is encircled. The y-axis shows the protein detection by UV in [mAu] on the left and the conductivity in [mS/cm] on the right hand side. The x-axis indicates the time in [min] and the collected fractions. The green curve indicates the concentration of the elution buffer containing 1 M sodium chloride. (**B**) Enlarged detail of the phytase peak from which the purified phytase fractions were collected (indicated by the numbers on the x-axis).
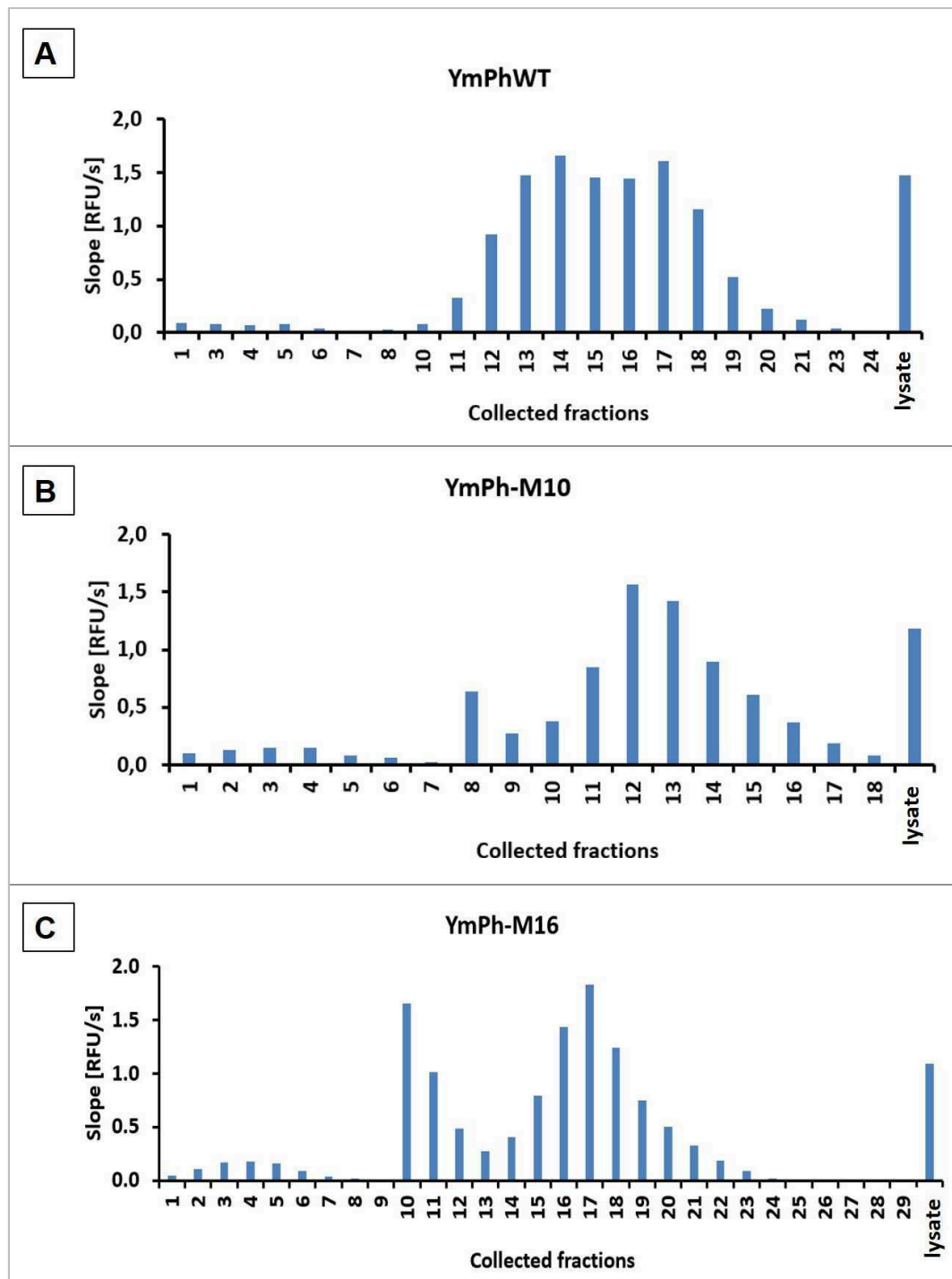
Figure A22: Result of the 4-MUP activity assay performed with the collected purified fractions from the phytase peak of (**A**) YmPhWT, (**B**) YmPh-M10 and (**C**) YmPh-M16. Lysate indicates the cell lysate before purification. The y-axis shows the slopes calculated from fluorescence measurements over time in [RFU/s] and the x-axis shows the different collected purified fractions.
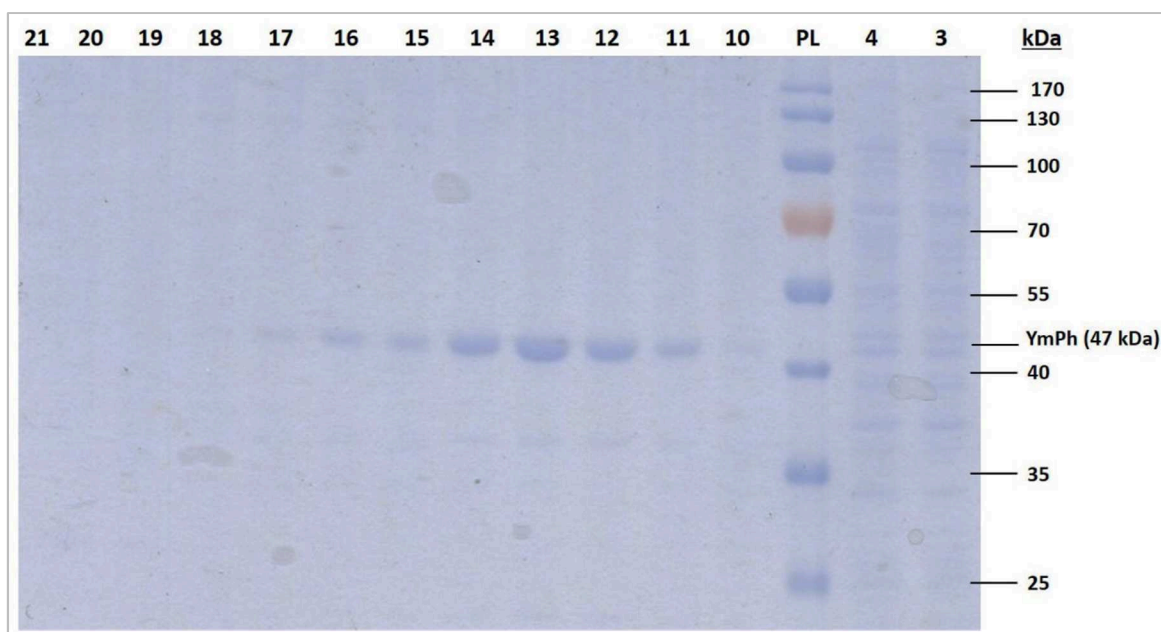
Figure A23: Result of the SDS-PAGE, showing the bands of the collected, purified fractions (10-21) from YmPh-M16 (exemplary) at the expected size of 47 kDa. Fractions 3-4 show the bands for the YmPh-M16 cell lysate before purification. PL indicates the protein ladder. Fractions (10-18) of highest purity were pooled together for concentration and desalting.
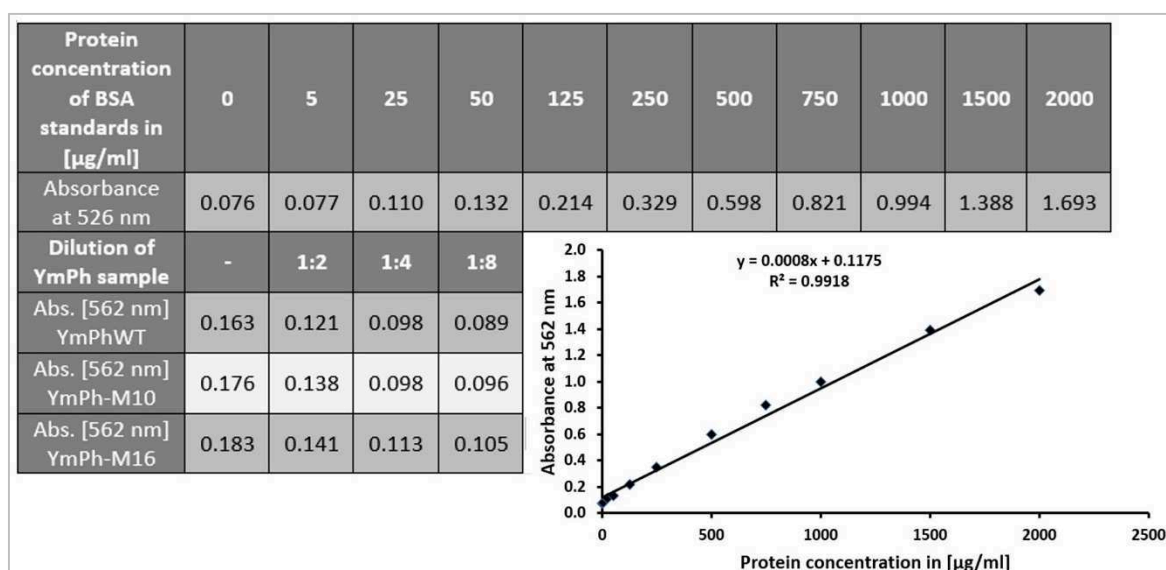
Figure A24: Result of the estimated protein concentrations for YmPhWT, -M10 and -M16 using Pierce™ BCA Protein Assay Kit (Life Technologies GmbH). BSA standards and purified proteins from YmPhWT, -M10 and -M16 (undiluted and dilutions 1:2; 1:4; 1:8) were analyzed with the kit and absorbance was measured at 562 nm. The calibration curve showing the protein concentration (x-axis) in correlation to the absorbance at 562 nm (y-axis) resulted in the linear function as illustrated.

Table A1: Kinetic characterization of improved variants with 4-MUP activity assay in sodium acetate buffer at pH 5.5. One unit (U) of enzyme was defined as the amount of phytase that catalyzes the conversion of 1 µmol of 4-MUP per minute. All values reported are the average of three measurements. Standard errors determined in triplicate measurements are represented in the brackets. Values are normalized to protein content based on Bio-Rad chip analysis. Phytase kinetics were determined in sodium acetate buffer containing different concentrations of 4-MUP (10-5000 µM) at 37°C and pH 5.6 and 6.6 for both phytase variants and the WT.

Sodium acetate buffer (pH 5.5)	*k_{cat} [s^{-1}]	K_{half} [µM]	Specific activity [U mg^{-1}]	Hill coefficient	Amino acid exchanges
YmPhWT	91.99 (± 0.19)	1044 (± 67.3)	189.02 (± 3.10)	1.68 (± 0.14)	-
YmPh-M10	64.87 (± 0.27)	929 (± 178.1)	123.78 (± 19.87)	0.92 (± 0.09)	T44I/K45E
YmPh-M16	66.24 (± 0.21)	893 (± 125.1)	115.31 (± 6.42)	0.96 (± 0.07)	T44V/K45E

*Catalytic center activity: enzyme concentration (moles) multiplied by the number of active centers (four).
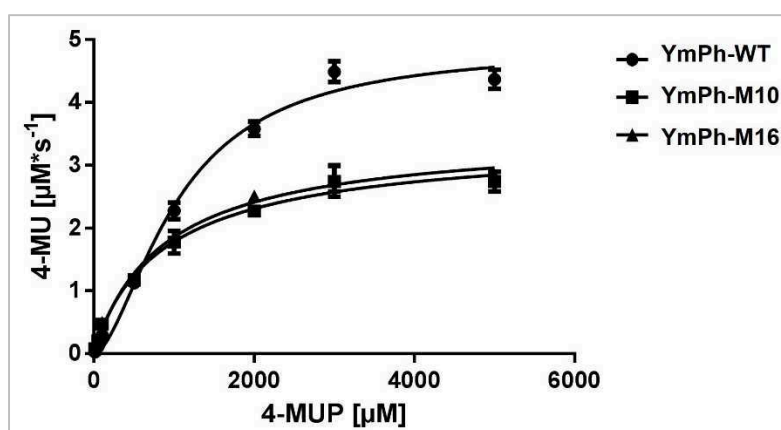
Figure A25: Enzyme kinetics of YmPhWT, -M10 and -M16 with 4-MUP substrate in sodium acetate buffer at pH 5.5. Initial activities were plotted against different substrate concentrations and fitted with Hill equation using GraphPad Prism 6 software. The reported values are the average of three measurements and deviations are calculated from corresponding mean values. The x-axis shows the 4-MUP substrate concentration in [µM] and the y-axis shows the 4-MU product formation in [µM*s^{-1}].
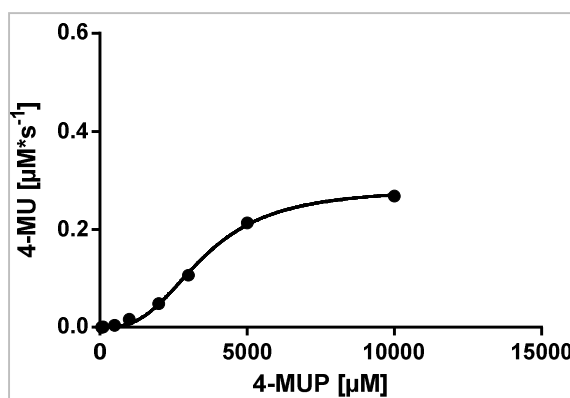

Figure A26: Enzyme kinetic of YmPhWT with 4-MUP substrate in Tris maleate buffer at pH 6.6. Initial activities were plotted against different substrate concentrations and fitted with Hill equation using GraphPad Prism 6 software. The reported values are the average of three measurements and deviations are calculated from corresponding mean values. The x-axis shows the 4-MUP substrate concentration in [µM] and the y-axis shows the 4-MU product formation in [µM*s^{-1}].

proteins, which are difficult to express *in vivo*. The application of the simple and rapid flow cytometry-based uHTS-IVC screening platform for directed cellulase evolution reveals revolutionary advances in regard to screening capabilities matching to requirements of novel methods for diversity generation in today's directed evolution campaigns. The quick and simple encapsulation of DNA using optimized extrusion technique, as well as DNA recovery after sorting, and the good compatibility between cell-free extracts and emulsion components (*i.e.* surfactants, oils) for generation of robust and stable emulsion compartments, that are able to withstand the shear forces upon flow cytometer-based analysis and sorting, demonstrate the broad applicability of the routine-ready flow cytometry-based uHTS-IVC technology platform.

Consequently, the development of the novel flow cytometry-based uHTS-IVC technology platform will be a major breakthrough for the directed evolution of proteins, as it overcomes the current limitations of today's state-of-the-art directed evolution such as insufficient throughput capacities and diversity losses due to limited transformation efficiencies of the host cells. The uHTS-IVC technology will offer significant advancements regarding reduction of time, screening efforts and expenses facilitating the performance of iterative rounds of mutagenesis and screening. The powerful uHTS-IVC technology platform will enable studying cooperative effects of amino acid side chains by screening advanced mutant libraries of high mutational loads with up to five simultaneous site-saturations (*i.e.* OmniChange). The latter will increase the possibility to identify beneficial variants showing altered activity, specificity or selectivity. The development of a robust uHTS-IVC technology screening platform will increase success probabilities to identify beneficial variants from diverse mutant libraries containing high numbers of inactive variants, which is unfeasible using standard screening systems due to time and cost limitation of directed evolution campaigns. Additionally, the use of IVC technology will facilitate the performance of iterative rounds of mutagenesis and screening and enable a rapid engineering of proteins exhibiting improved properties in a cost and time efficient manner. In addition, the employment of the flow cytometry-based uHTS-IVC technology platform will open up new exciting opportunities to study challenging scientific questions like the exploration of non-understood enzyme properties (*e.g.* organic solvent, ionic liquid resistance, pH stability).

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.

Aachen, 3rd of May 2016

_____

Georgette Körfer

Publications

_____ Journals

04/2016   **Körfer G.**, Pitzler C, Vojcic L., Martinez R., and Schwaneberg U.; *In vitro* flow cytometry-based screening platform for cellulase engineering; *Scientific Reports* (2016); 6: 26128

01/2015   Vojcic L.[§], Pitzler C.[§], **Körfer G.**[§], Jakob F. [§], Martinez R., Maurer K.-H., and Schwaneberg U.; Advances in protease engineering for laundry detergents; *New Biotechnology* (2015); 32(6) 629-634

12/2014   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

11/2012   Ruff A. J.[§], Dennig A.[§], **Wirtz G.**, Blanusa M., and Schwaneberg U.; Flow Cytometer-Based High-Throughput Screening System for Accelerated Directed Evolution of P450 Monooxygenases; *ACS Catalysis*, (2012); 2(12), 2724-2728

[§] shared first authorship

_____ Patents

Application patent   „Kompartiment-basierte, zellfreie Ultra-Hochdurchsatz-Durchmusterungssysteme für das Protein Engineering" (RA 1697-01EP GDO/thh; patent filing in process)

Application patent   „Lipasen für den Einsatz in Wasch- und Reinigungsmitteln" (PT033756; patent filed)

_____ Conferences

07/2014   16th European Congress on Biotechnology (Edinburgh Scotland)
Poster presentation: "Development of a Flow Cytometer-based *In Vitro* Compartmentalization Screening Platform for Directed Protein Evolution"

11/2012   Belgrade Food International Conference (Belgrade, Serbia)
Talk: "Development of a Flow Cytometer-based *In Vitro* Compartmentalization Screening Platform for Directed Protein Evolution"

Curriculum Vitae

_____ ## Georgette Körfer (geb. Wirtz)

Grenzweg 46a
52224 Stolberg
+49 2409-7017420
geoccino@web.de

Date of birth   20th March 1984
Place of birth   Eschweiler, Germany
Nationality   German

_____ ## Academic education

Since 11/2011   PhD studies of biotechnology, RWTH Aachen University
                     PhD Thesis entitled *"Development of a flow cytometer-based in vitro compartmentalization screening platform for directed protein evolution"* (Institute of Biotechnology, Prof. Dr. Ulrich Schwaneberg, RWTH Aachen University)

10/2009 - 09/2011   Master studies of biology, RWTH Aachen University
                     (Degree: Master of Science)
                     Master Thesis entitled *"Directed evolution of cytochrome P450 BM3 for mediated electron transfer"* (Institute of Biotechnology, Prof. Dr. Ulrich Schwaneberg, RWTH Aachen University)

10/2006 - 09/2009   Bachelor studies of biology, RWTH Aachen University
                     (Degree: Bachelor of Science)
                     Bachelor Thesis entitled *"Auswirkungen verschiedener Stressfaktoren auf das Redoxpotential von Glutathion bei Schizosaccharomyces pombe"* (Institute of Applied Microbiology, Prof. Dr. Klaus Wolf, RWTH Aachen University)

_____ ## Education

06/2006 - 09/2006   Technical draftswoman (Bombardier Transportation Germany GmbH & Co. KG, Aachen)
09/2003 - 06/2006   Apprenticeship as technical draftswoman (Bombardier Transportation Germany GmbH & Co. KG, Aachen)
1994 - 2003   Goethe-Gymnasium Stolberg (bilingual German-English classes)
                     Degree: Allgemeine Hochschulreife (with bilingual certificate)
1990 - 1994   Grundschule Gressenich

3. A flow cytometer-based in vitro compartmentalization screening platform for directed protein evolution

3.1. Declaration

Parts of this chapter have been published in the journal *Scientific Reports* and are adapted to this thesis with permission of the Nature Publishing Group: "Körfer, G., Pitzler C., Vojcic L., Martinez R., and Schwaneberg U.; *In vitro* **flow cytometry-based screening platform for cellulase engineering**, Scientific Reports (2016); **6**: 26128."

3.2. Project objective

As the challenge of diversity generation for directed evolution campaigns was already tackled by methods like OmniChange and SeSaM, the main bottleneck nowadays is the development of suitable screening systems offering ultrahigh throughput to sample through a statistically meaningful fraction of the generated sequence space. An excellent option to achieve this goal is to eliminate the transformation step into the expression host, which is the main limitation upon screening of highly diverse mutant libraries, as the diversity is drastically reduced during cloning and transformation. Besides this aspect, cell-free protein expression enables to broaden the screening to toxic or membrane proteins, which are difficult to express *in vivo*. In addition, the employment of flow cytometer-based screening formats, in combination with *in vitro* compartmentalization (IVC) technology, enables a drastic reduction of time requirements for one round of directed evolution from several weeks to only few days.

In order to overcome the limitations of todays directed evolution campaigns, the goal of this work is to develop a widely applicable, robust flow cytometer-based ultrahigh throughput screening (uHTS) platform for directed enzyme evolution using *in vitro* compartmentalization (IVC) technology in water-in-oil-in-water (w/o/w) emulsions. For development of the uHTS-IVC platform the two model enzymes phytase (YmPhWT) and cellulase (CelA2-H288F) will be used. The main goal is to employ the developed technology platform using compartmentalization of cell-free expressed enzymes in (w/o/w) emulsions as pre-screening for analysis of huge gene libraries (10^8), facilitating a rapid discrimination between active and inactive enzyme variants. In order to validate the developed uHTS-IVC platform it is necessary to perform at least one round of directed evolution for one model enzyme by screening of highly diverse mutant libraries and isolation of improved variants.