

DISS. ETH NO. 19176

**NOVEL HIGH THROUGHPUT TECHNOLOGIES FOR BACTERIAL  
LIBRARY SCREENING IN LIFESCIENCE**

A dissertation submitted to

**ETH ZURICH**

For the degree of

**Dr. Sc. ETH Zurich**

Presented by

**MARCEL WALSER**

Dipl. Natw. ETH

Born November 14<sup>th</sup>, 1976

Citizen of Rehetobel (AR), Switzerland

Accepted on the recommendation of

Prof. Dr. Sven Panke, examiner

Prof. Dr. Willhelm Grissem

Dr. Richard Reinhardt

2010

## **Acknowledgements**

We are indebted to the R'equipe program of the Swiss National Science Foundation for generous support in acquiring the COPAS Plus Biosorter, and to the Swiss Commission for Technology and Innovation for grants to this thesis.



Eidgenössische Technische Hochschule Zürich  
Swiss Federal Institute of Technology Zurich

## **Abstract**

Availability of strains and enzymes with a certain catalytic property is a key requirement for the development of biotechnological manufacturing processes. Appropriate candidates can either be selected or screened out of a collection of natural isolates or recombinant clones. However, identification of a desired activity is challenging because, due to enormous variety of proteins and organisms, millions of samples may have to be analyzed in order to finally identify the desired property. A multitude of technologies has been developed in order to negotiate processing and analysis of the large numbers of required samples within considerable time and reasonable costs.

The aim of this thesis was to develop novel high throughput screening technologies for microbial libraries with a focus on sample processing in microcapsules of a volume of not more than a few nanoliters. Hydrogel-microcapsules were adapted to the purpose of compartmentalizing single cells followed by their proliferation to microcolonies. Protocols for the analysis of larger populations of colonized microcapsules by flow cytometry were successfully developed and applied for the enrichment of microcapsules colonized by a single microcolony (“monoclonals”).

Afterwards, protocols for colony-based polymerase chain reaction (PCR) were adapted to the task of performing PCR within the microcapsules. The protocol was used for PCR-screening of microsatellites in 20'000 *E. coli* clones containing a genomic library of cassava. PCR-positive microcapsules were identified and isolated by flow cytometry and the extracted plasmids were sequenced. As a result, 11 novel microsatellite markers were obtained in less than a week with an optimized protocol.

In order to further explore the limits of the microcapsule technology for screening, a second selection and screening protocol for microbial colonies

grown within microcapsules was successfully developed. The underlying principle used for both, analysis and separation of microcapsules with positive clones from those with negative clones or empty microcapsules was density difference of the respective microcapsules. The key element of the protocol is the use of catalase as a reporter protein. Presence of the reporter allows the formation of an oxygen bubble inside microcapsules that induces the density difference and makes microcapsules with positive colonies move upwards to the surface of a suspension, thereby allowing facile separation. This novel reporter-system is widely scalable and allows both, analysis and separation in a single, highly parallelized step.

A second aspect of microcapsule technology is that the semi-permeable nature of the capsule allows rapid exchange of solutes between the exterior and the interior and thus to use the microcapsules as microreactors in which many processing steps can be carried out successively. We investigated whether this property could be used to miniaturize an exceptionally laborious protocol, the protocol for the generation of Sanger-sequencing fragments. Even though all single steps of the protocol could be shown to work in the microcapsules, the developed protocol delivered only low signal intensities in subsequent capillary electrophoresis analyses, pointing towards a currently poorly understood limitation of the microcapsule technology.

In summary, the work of this thesis established the processing and analytical fundamentals of a highly efficient, hydrogel microcapsule-based screening technology for biotechnological applications.

## Zusammenfassung

Die Verfügbarkeit von Stämmen und Enzymen mit bestimmten katalytischen Eigenschaften ist eine der Hauptanforderungen für die Entwicklung von biotechnologischen Fertigungsprozessen. Geeignete Kandidaten werden dabei aus einer Sammlung von natürlichen Isolaten oder rekombinanten Klonen entweder selektiert oder ausgesondert. Indes ist die Identifizierung einer gewünschten Aktivität eine Herausforderung, weil aufgrund der enormen Vielfalt an Proteinen und Organismen Millionen von Proben analysiert werden müssten, um letztendlich die gewünschte Eigenschaft zu ermitteln. Viele Technologien wurden entwickelt, um das Prozessieren und Analysieren dieser grossen Zahl an gewünschten Proben innerhalb nützlicher Frist und zu angemessenen Kosten zu bewältigen.

Es war das Ziel dieser Arbeit neue Hochdurchsatztechnologien für mikrobielle Bibliotheken, mit Fokus auf die Probenprozessierung in Mikrokapseln mit einem Volumen von nicht mehr als ein paar Nanolitern, zu entwickeln. Zur Kompartimentalisierung von einzelnen Zellen, welche anschliessend zu Mikrokolonien proliferierten, wurden Hydrogelmikrokapseln angewendet. Zur Analyse von grösseren Populationen kolonisierter Mikrokapseln mittels Durchflusszytometer wurden erfolgreich Protokolle entwickelt und zur Anreicherung von Mikrokapseln, welche von einer einzigen Mikrokolonie besetzt waren, angewendet.

Danach wurden Protokolle für Kolonie-basierende Polymerase Kettenreaktionen (PCR) angepasst, um PCR in Mikrokapseln durchzuführen. Das PCR-basierende Protokoll wurde für die Suche nach Mikrosatelliten in einer genomischen Maniok Bibliothek in 20'000 *E. coli* Klonen verwendet. PCR-positive Mikrokapseln wurden identifiziert, mittels Durchflusszytometer isoliert, die Plasmide extrahiert und anschliessen sequenziert. Als Resultat

wurden mit diesem optimierten Protokoll elf neue Mikrosatellitenmarker in weniger als einer Woche entwickelt.

Um weiter Grenzen der Mikrokapselftechnologie für Screenings zu erkunden, wurde erfolgreich ein zweites Selektions- und Screening Protokoll für mikrobielle Kolonien in Mikrokapselfn entwickelt. Das Grundprinzip für beides ist die Analyse und die Separation von Mikrokapselfn mit positiven Klonen von denen mit negativen Klonen oder leeren Mikrokapselfn. Dabei ist das Schlüsselement des Protokolls die Katalase als Reporter Protein. Die Anwesenheit des Reporters erlaubt die Bildung von Sauerstoffblasen im Innern der Mikrokapselfn, was zu einem Dichteunterschied führt und die positiven Mikrokapselfn dazu veranlasst zur Suspensionsoberfläche auszusteigen, was eine vereinfachte Separierung erlaubt. Dieses neuartige Reporter-System ist weiträumig skalierbar und erlaubt sowohl die Analyse wie auch die Separierung in einem einzigen, hochparallelen Schritt. Ein zweiter Aspekt der Technologie mit Mikrokapselfn kommt mittels der halbdurchlässigen Natur der Kapselfn zum Vorschein, was einen schnellen Austausch von gelösten Substanzen zwischen dem Äusseren und dem Inneren der Kapselfn und somit den Gebrauch von Mikrokapselfn als Mikroreaktoren, in denen viele aufeinanderfolgende Prozessschritte ausgeführt werden können, erlaubt. Wir prüften, ob diese Eigenschaft gebraucht werden kann, um ein ausserordentlich aufwändiges Protokoll zu etablieren, das Protokoll zur Synthese von Sanger Sequenzierungsfragmenten. Wenn auch gezeigt wurde, dass die einzelnen Schritte der Protokolle funktionieren könnten, lieferte doch das entwickelte Protokoll lediglich schwache Signalintensitäten bei anschliessenden Kapillarelektrophoreseanalysen, was auf eine bis jetzt eher spärlich verstandene Limitierung der Mikrokapselfntechologie hinweist.

Zusammengefasst wurden während der Arbeit dieser These Prozess- wie auch Analysegrundsätze für hoch effiziente, auf Hydrogelmikrokapseln basierende, Screening-Technologien für biotechnologische Anwendungen, ermittelt.

## Table of Contents

I. General introduction .....	- 8 -
Introduction .....	- 9 -
Selection Strategies .....	- 12 -
Cellular library screening .....	- 14 -
Clone processing technologies .....	- 19 -
Scope of this thesis .....	- 24 -
II. Isolation of monoclonal microcarriers colonized by fluorescent <i>E. coli</i> ....	- 28 -
Abstract.....	- 29 -
Introduction .....	- 31 -
Materials and Methods .....	- 34 -
Results and Discussion.....	- 43 -
Summary & Conclusions .....	- 55 -
III. Novel method for high throughput colony PCR screening in nanoliter-reactors .....	- 59 -
Abstract.....	- 60 -
Introduction .....	- 61 -
Materials and Methods .....	- 66 -
Results.....	- 72 -
Discussion and Conclusions .....	- 81 -
IV. Novel screening platform employing catalase-induced density separation of microcapsules.....	- 88 -
Introduction .....	- 89 -
Materials and methods.....	- 94 -
Results.....	- 98 -
Discussion and Conclusion.....	- 108 -
V. Dye-end-labeled DNA synthesis in microcapsules .....	- 111 -
1. Introduction .....	- 112 -
2. Materials and Methods .....	- 114 -
3. Results.....	- 121 -
4. Discussion and outlook .....	- 127 -
VI. Concluding remarks and outlook .....	- 133 -
References.....	- 138 -



## **I. General introduction**

## Introduction

In the course of evolution, biological systems have acquired an extreme variety of molecular interactions and macromolecules. Irrespective of the specific field (such as metagenomics, systems biology, ecology, or protein engineering), we are only at the beginning of understanding this complexity of interactions and of exploring the molecular variety. Just to quote one example, even in a microbiological habitat such as the bovine rumen, whose general functional principles are known already for a long time, only the minority of harbored microorganism species has been described and the number of totally inhabiting microorganisms can only be estimated very roughly by classification in general.<sup>1-3</sup>

However, such complex habitats can harbor cells with very interesting properties, such as cellulose hydrolysis. However, large efforts are required to identify and classify unknown and potentially interesting organisms as these organisms might either be not culturable or, in terms of cell numbers, make up only a very small fraction.<sup>3-7</sup>

Rather than isolating novel activities from nature, novel enzyme functionalities can be obtained by directed evolution of a protein with known function. However, the corresponding effort can be huge: for a middle-sized protein of, say, 300 amino acids, only producing all variants in which 1% of the amino acids (3 amino acids) were altered, relative to the parent molecule, results already in more than a billion different combinations ( $L = (X^N K!)/N!(K-N)!$ ; where L is the total number of combinations, N the number of varied amino acids (e. g. 3), K the sequence length (e. g. 300) and X the number of amino acids available (i. e. 20)).<sup>8-10</sup> Even if in recent times the use of focused libraries has led to the use of substantially smaller libraries, there are many cases in which the 3D-structure

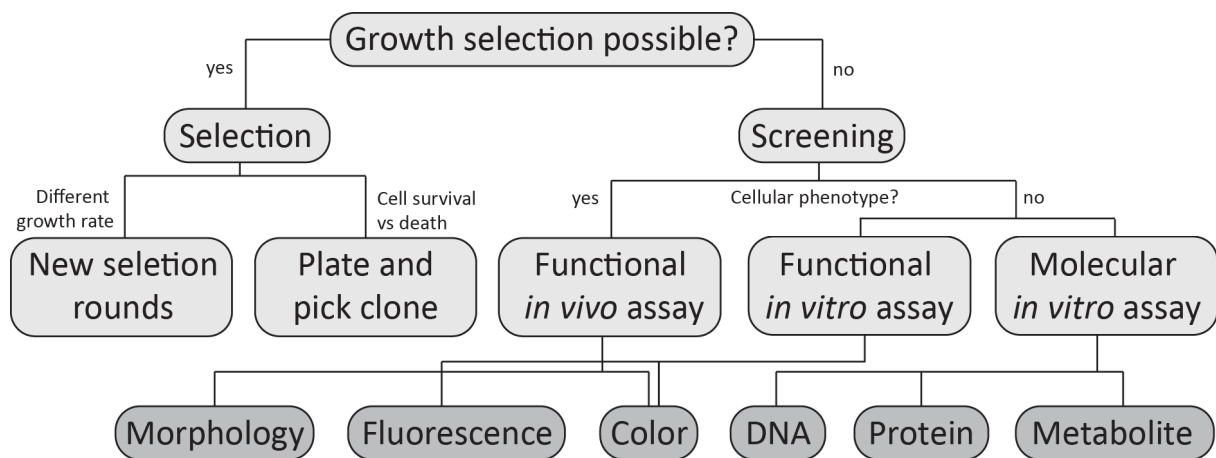
of the protein is not available and at least at the beginning, large populations of molecules need to be screened.

Out of this large number of events, certainly not all enzymes or protein interactions, whether explored or engineered, match the criteria that were put forward. In fact, only a tiny minority of possibilities can be expected to have the required characteristics. For that purpose, high throughputs need to be achieved when searching for such few desired variants in a pool of undesired variants. In general two main strategies are available. [i] Selection for the cell whose phenotype matches best a specific requirement, in which the desired phenotype gives a cell the possibility to survive, while cells carrying undesired variants do not, or [ii] the screening for novel candidates by analyzing large numbers of experimental events and isolating those matching a given criterion (see figure 1).<sup>11-14</sup> Here, selection is defined as the process by which certain heritable traits make it more likely for an organism to survive and successfully reproduce to become more abundant in a population over successive generations. A screening is understood as the investigation to isolate target clones, harboring particular features, from a vast community of clones.

Within the last two decades, many screening and selection technologies have been developed.<sup>13, 15-20</sup> Even though these strategies differ in detail, they rely either on selection or screening of single clones based on similar analyzable properties (e. g. fluorescence, color, structure).<sup>21</sup> In our context, a clone is defined as a group of identical cells or molecules that are in close proximity – for example a group of bacteria in a colony that go back to one progenitor cell or a group of DNA molecules that were produced by polymerase chain reaction (PCR) from one DNA molecule and are kept together, such as in an alginate microcapsule. Due to the inherently large numbers of clones that need to be handled in screening, clone handling technologies (e. g. fluorescence assisted

cell sorting (FACS), complex object parametric analysis and sorting (COPAS), mass spectrometry (MS), plate-handler, colony pickers) are an integral part of the screening process.<sup>22</sup> In order to identify the desired “needle in the haystack”, the experimenter has thus a large variety of techniques available (see figure 1).

Due to the number of the various available techniques, the scope of this introduction is limited to the basic principles of selection and screening technologies, subdivided into groups according to various criteria such as in-vivo or in-vitro processing of clones in respect to analyzable properties<sup>21-30</sup> (see figure 1). A more detailed introduction will then be given in the field of microdroplet processes, the core technology used to separate and process single clones in this thesis.



**Figure 1:** Search strategies for the discovery of novel microbial species or enzymes interesting for the development of biotechnological manufacturing processes (light grey boxes). According to the chosen strategy different paths for analysis may be employed (dark grey boxes – DNA, protein and metabolite stand for analyses that are typical for these molecule types, such as PCR for DNA).

## **Selection Strategies**

In general, cellular selection strategies are favored due to their simple experimental setup when compared with highly sophisticated screening strategies. As an example, Entcheva et al. 2001<sup>31</sup> selected clones containing biotin operons from different soil libraries. The cosmid libraries contained up to 35'000 clones and suitable cosmids were selected by complementation of a biotin auxotrophy in an *Escherichia coli* strain. This resulted in the isolation of seven cosmids which contained operons for biotin synthesis. Here, the evaluation of 35'000 different cosmids was efficiently reduced to the evaluation of only 7. Such selection experiments can be performed by batch-cultures and/or -two-dimensional (2D) array plate assays (such as Petri dishes). In selection processes, usually the positive clones will be enriched, ideally exclusively, and the unwanted background of cells becomes diminishingly small which shortens the time to find positive clones.<sup>8</sup> The selective advantage of positive clones in a screening can be made very stringent, for example by selecting for antibiotic resistance so that all cells that remain below a certain level of resistance die and are hence removed from the remaining pool. Alternatively, selection can be coupled as well to utilization of a selective carbon source. Kachroo et al. 2007<sup>32</sup> selected and characterized 42 *E. coli* mutants with the ability to grow on mineral medium plates containing cellobiose as a sole carbon source, which is generally not used by *E. coli*. *E. coli* possesses five genes in one operon for cellobiose conversion into glucose and glucose phosphate: three for membrane transport (chbA, chbB and chbC), one for hydrolysis (chbF) and one for operon regulation (chbR). Since the operon is usually silenced and not inducible by cellobiose, mutations found were mainly in the operator region of the operon or in the regulator gene responsible for upregulation of the operon.

However, selection conditions can also be less stringent, for example when cells compete on specific growth rate in a liquid culture. Depending on the growth period, the culture might still contain substantial number of alternative clones.

In selection processes, a growth benefit is commonly described by the selection ratio ( $R_s$ ).  $R_s$  describes the ratio of the growth rate of one strain compared to another one ( $R_s = \mu_1 / \mu_2$ ). As a rule of thumb, a value  $R_s > 1.2$  is enough to distinguish between two clones in cultivations over 30-40 generations.<sup>8</sup> Below that, continuous cultures are required to enable enough generations for the detection of a difference in cell numbers for the best performing strains.

In order to find more than just the best growing strain, novel array-based platforms for the compartmentalization of a larger number of single clones (e. g.  $10^7 - 10^9$  cells) were developed.<sup>26, 33-35</sup> This compartmentalization in combination with high-throughput imaging technologies enables precise screening of minor growth rate differences as well as the handling of cellular entities in parallel.<sup>16, 20, 28, 36</sup> However, many of these novel miniaturizing technologies potentially comprise the ability to screen as well for alternative cellular properties than growth rate (e. g. color, fluorescence) as discussed in the next sections (see figure 1).

## Cellular library screening

In case no selection mechanism can be applied in order to identify the required clones, highly sophisticated library screening strategies are available which were developed within the last two decades and which resulted in increasing sample throughput and information content. In general, two main strategies are pursued for the screening of cellular libraries: firstly, functional *in vivo* screenings and secondly, whenever no cellular phenotype can be exploited, molecular *in vitro* screenings (see figure 1).<sup>37</sup>

In function-based cellular screenings, desired candidates are mostly discovered by in-vivo assays. A cell displays a certain phenotype (e.g. fluorescence or color signal or simply a characteristic morphology) that can be screened directly. In case no cellular approach is possible (e. g. when looking for enzymes which are toxic for the screening host),<sup>38</sup> or when a molecular approach is more practical or faster,<sup>39</sup> the entire library needs to be analyzed by in-vitro processes suitable for the detection of specific molecules, such as a specific DNA sequence, a specific protein, or a specific metabolite or chemical product) (see figure 1).<sup>10, 36</sup> On DNA level, for example, such strategies are usually performed by recently developed ultra-high-throughput technologies such as next generation DNA-sequencing or micro-arrays for DNA-hybridization. However, in order to obtain the actual clone after the screening process, one needs to go back to the original library or even synthesize it from scratch. Hence, conventional methods in which the connection between a living clone and the screening signal can be maintained remain standard procedures for some applications (e.g. membrane blotting, PCR based technologies and MT plates handling platforms).

## Analyzable properties in function-based in-vivo screenings

Fluorescence-based visualization of cellular properties has become of central importance since the discovery of fluorescent proteins (FP) in the luminous hydromedusa *Aequorea* in 1962 by Shimomura.<sup>40</sup> Fluorescent proteins display a number of desirable properties. Most importantly, they emit light after optical excitation without any additional requirement. This is in contrast to other light-related assays, such as for example luminescence assays, where the substrate for the light reaction has to be added or even synthesized by the cell first.<sup>41</sup> Moreover, fluorescent proteins are relatively small and can be engineered such that they act as monomers, which facilitates the formation of functional fusion proteins.<sup>42</sup> Finally, the light emission properties makes the detection of FPs rather sensitive.<sup>43</sup> These advantages have triggered the use of FPs in many molecular engineering projects. Engineered fusion proteins cannot only be used to localize a certain protein within an organism or a cell<sup>41, 44</sup> but also for a connection to a certain protein function<sup>15</sup> or simply to stain a specific cellular subpopulation.<sup>45</sup> As an example, Miesenböck et al. 1998<sup>46</sup> developed in such way a labeling system to visualize secretion and synaptic transmission with pH-sensitive green fluorescent proteins linked to a vesicle membrane protein. Florescent proteins are not the only way for fluorescent labeling of cells.<sup>47</sup> Fluorescent dyes can be attached to affinity tags which specifically bind to certain structures. One example here is the screening of inhibitor peptides as potential new drug candidates after peptide-display on the surface of *E. coli* in 2005 by Jose et al.<sup>48</sup> These peptides were screened for immobilizing biotinylated human cathepsin G, which in turn was detected by a streptavidin-fluorescein and sorted by FACS. However, such assays are often difficult to be performed intracellularly, as fluorescent dyes may not penetrate the cell-membranes respectively walls.



A frequently used color assay in lifesciences is the blue/white screening where bromo-chloro-indolyl-galactopyranoside (X-Gal) is cleaved by a functional beta-galactosidase (a homo tetrameric protein) yielding galactose and 5-bromo-4-chloro-3-hydroxyindole (a blue insoluble stain). This reaction is employed in screenings when combined with *E. coli* harboring a mutation in the 5'-region of *lacZ* which prevents subunit association and down-regulates enzyme activity. Subunit assembly (and enzyme activity) is restored by the presence of a small (26 amino acids) amino-terminal fragment of the *lacZ* product (the so-called alphapolypeptide), which is small enough to be encoded on cloning vehicles to signal inactivation by insertion of a recombinant DNA-sequence. This way, insertion can be easily screened for the absence of blue color. Likewise, sugar-hydrolysing enzymes can be detected by hydrolysis of a chromogenic sugar-based substrate,<sup>13</sup> This concept can be easily extended to other groups of molecules.<sup>21, 49</sup> Finally, many cell staining dyes that were discovered in cell biology's early days are still in use to visualize cellular differences after for example mutagenesis.<sup>50</sup>

The detection of morphological differences requires the integration of a microscopic analysis into the screening process. This extends the scope of "analyzable" properties far beyond what can be detected through dyes. Here, speed is often limited due to manual handling of samples or by the robot-device which picks the positive clone after scanning each individual compartment.<sup>50</sup>

### **Molecular analyzes for in-vitro screening platforms**

Not all functional assays can be performed in-vivo and some enzyme classes, DNA structures or cellular metabolites are screened in-vitro. For molecular screenings, typically standard methods and commercially available devices may

be implemented, which helps to set up a screening experiment within a relatively short time. For this reason, even though throughputs might be lower than in a function-based screenings, the time required until identification of a positive clone might effectively be shorter (see figure 1).

Within the last five years novel sequencing technologies (“next-generation sequencing technologies”) increased sequencing throughputs up to several giga-basepairs per run.<sup>36, 51-54</sup> However, the length of the sequenced part of any specific DNA molecule has actually decreased (relative to Sanger-type sequencing) and, due to the preparation of the samples for sequencing, no direct link between the determined sequence and the original clone remains. Due to rather short read-length between 20-250 bp (depending on used technology) up to 30-fold sequence coverage is required for proper alignment of single reads *in silico* and in some cases it is impossible to close sequence gaps with such small reads.<sup>55</sup> Nevertheless, costs per sequenced megabase drastically decreased to below 1 USD and next-generation sequencing has the potential to substantially accelerate research in lifescience.<sup>54</sup> However, it should be mentioned that fast screening strategies based on membrane blotting or PCR still fill certain niches when it comes to screening for rare events in a large sequence space,<sup>56, 57</sup> such as screening for gene families or microsatellites. Thus, Kim et al. 2003<sup>58</sup> screened more than 250’000 individual clones of a gorilla fosmid library with an average insert size of 40’000 bp to find 9 clones that cover the entire HOX A gene cluster. This corresponds to a screened sequencing space of more than 10 Gbp which is infeasible even with next-generation sequencing technologies.

Many of the above mentioned cell labeling techniques for functional screens can as well be used in in-vitro screens. So, Lee et al. 2002 performed for example in vitro transcription and translation reactions in micro emulsion

droplets to screen for activities of methyltransferases encoded in an error prone PCR library. In case an active methyltransferase mutant was translated, a restriction site of its own gene construct was methylated. The emulsion was afterwards broken and the gene constructs were treated with a methylation sensitive restriction endonuclease. Digested vectors could not enter into a new round of selection resulting in an enrichment of active methyltransferases. In addition, other analytic technologies can be integrated into the screening process, such as MS, which becomes important when the enzyme activity cannot be coupled to a fluorogenic or chromogenic signal output. In such case the enzyme is analyzed solely upon its amino acid sequence and generally protein analysis platforms sample less than one event per second, which reduces throughput of the experiments when compared to solely function-base experiments.

Cellular metabolite screening is of major importance in the pharmaceutical industry.<sup>13, 59</sup> Metabolites may be screened by secondary enzymatic reactions, high pressure liquid chromatography (HPLC), gas chromatography (GC), MS, nucleospin magnet-resonance spectroscopy (NMR), or infra-red spectroscopy (IR).<sup>10</sup> The corresponding instruments can be operated in routine fashion and combined with robotic stations available on the market. However, these instruments have relatively low throughputs and are generally standardized to 2D-plate formats operating with rather high volumes per sample.<sup>18</sup>

## Clone processing technologies

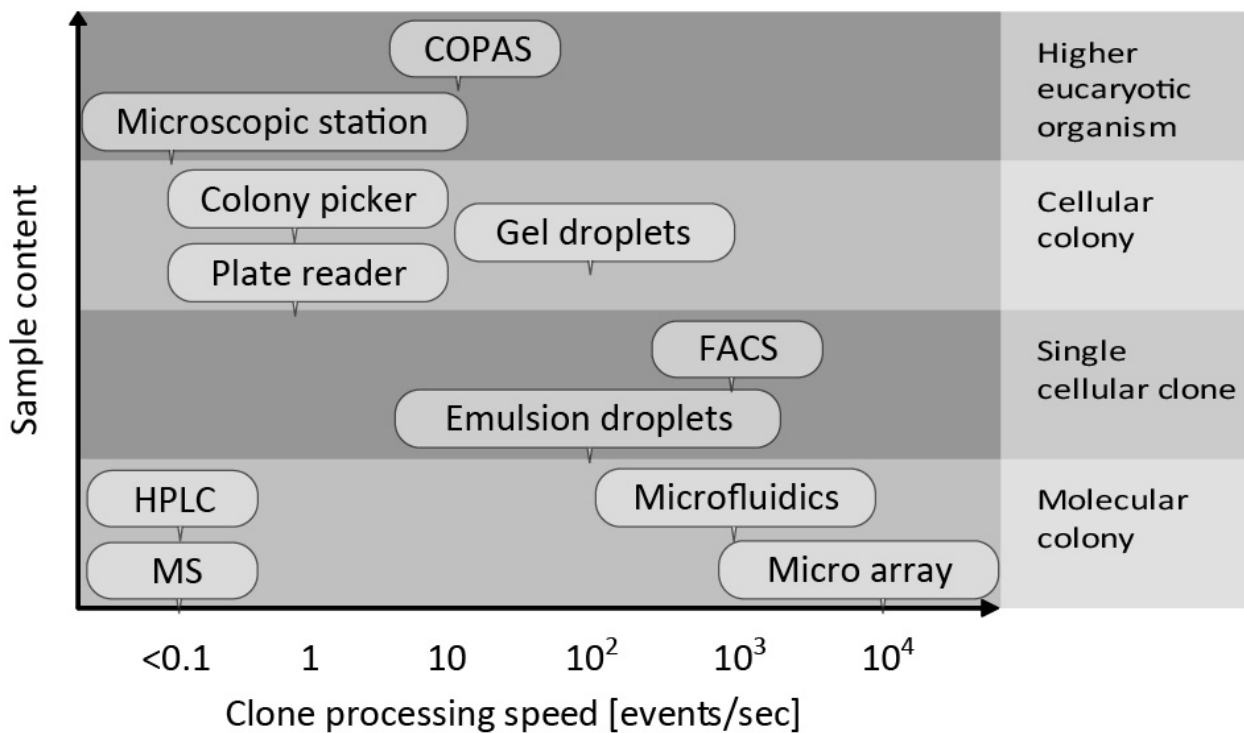
Influenced by the decision what selection or screening strategy is chosen, there is a multitude of technologies for processing of the large number of single clones. Often these technologies are based on random arraying of individual clones of a library and rely on Poisson distribution of samples within the array.<sup>51, 60, 61</sup> The available technologies vary drastically by throughput and information content and might be separated in two categories according to their geometrical properties: two dimensional arrays, where the samples are attached to a surface and three dimensional suspension arrays, where the clones are confined to a compartment but the compartment is freely suspended in an emulsion or a suspension.<sup>22, 47</sup> A brief listing of the most common clone processing technologies is listed in figure 2, with a rough comparison of throughput and information content per analyzed clone. In terms of processed information content, technologies differ drastically from each other. Thus, there is for example more information to gain from the analysis of an entire organism than from a molecular clone. Therefore, several subdivisions for the information content of a single clone need to be taken into account when comparing throughputs of different technologies. These subdivisions of presented technologies range from molecular colonies to single cells, cellular colonies and multicellular organisms (see figure 2). Some of these classifications are somewhat arbitrary. For example, flowcytometers focussing on single cell sorting (here referred to as fluorescence activated cell sorters; FACS)<sup>22</sup> have rather high throughputs up to  $10^4$  clones per second. Other flowcytometers with larger flow capillary diameters (COPAS – complex object parametric analysis and sorting) can analyze objects in the range of hundreds of microns (e. g. worms, insect eggs)<sup>62</sup> but sort with smaller frequencies up to 10 Hz. While this is still rather high with respect to sorted information content

of an entire eucaryotic organism,<sup>44</sup> this number is orders of magnitude smaller than with conventional FACS machines.

Whole organisms can also be analyzed with microscopes coupled to imaging software and robotic stations that process the positive hits according to the application. Cellular colonies may be analyzed in plate readers and processed with microtiter plates or on agar plates in combination with a colony picking robot<sup>63</sup>.

If there is no possibility to analyze the samples functionally, like for non-cellular molecular clones (e. g. DNA, proteins, or metabolites),<sup>27</sup> the technologies of choice are next- generation sequencers, MS or HPLC. Even though MS and HPLC devices are highly standardized, their throughput is generally below one event per second and hence not applicable for ultra high-throughput screenings.<sup>10</sup>

As an alternative to plate formats (microtiter plates or agar plates), the encapsulation of single cells in microdroplet compartments and their use in 3D suspension arrays gains increasing importance in high throughput screenings as throughputs of cell or even colony analysis can be increased due to combination possibilities with other clone processing devices (e. g. flowcytometers, microscopic stations) in standardized fashion.<sup>21, 64</sup> Since the focus of this thesis is based on such microdroplet processing applications we will discuss some of the possibilities in the next sections.<sup>65</sup>



**Figure 2 :** Classification of clone processing technologies according to throughput and sample information content. Numbers for throughput are taken from the literature which is referenced in the text.

### Microdroplet based technologies

A microdroplet is defined as any droplet generated with a diameter in the micron or sub-micron range, without regard for the droplet production method used.

Microdroplets can compartmentalize cellular and molecular clones so that various reactions can occur and be physically linked to the originating clone. Main advantages lie in the drastic reaction volume reduction when compared with other reaction compartmentalization technologies (e. g. micro-titer plates).<sup>30</sup> These properties allowed the development of a number of applications, such as therapeutic delivery, biomedical imaging, drug discovery, biomolecule synthesis, and screenings. Different microdroplet materials and production procedures are briefly presented in the next sections.

The experiments of this thesis are based mainly on technologies for encapsulation of microbial cells in hydrogel microdroplets. Hydrogels are linked polymer meshes in aqueous solutions which may be complexed and gellified. For example, alginate is a sugar polymer containing mannuronic- and guluronic acid units, whose negatively charged carboxyl groups can be complexed with various bivalent cations, such as calcium, strontium or barium. These technologies have attracted high interest for use in cellular screenings<sup>65</sup>. Hydrogels such as alginate and agarose are the most commonly used polymers to produce hydrogel microdroplets because of their relatively simple preparation, easy handling and biocompatibility.<sup>66</sup> Cell encapsulators employing laminar-jet-break-up<sup>67</sup> or flow focusing technologies<sup>68</sup> can create thousands of uniform liquid microdroplets per second which, upon contact with polycations or cooling, immediately gellify into semi-solid gel capsules (see movie 1 and caption at the end of this chapter).<sup>66</sup> Cells remain in principle viable during this process. Their survival is attributed to the fact that the droplet shell can isolate the cells from harsh environment (e. g. xenotransplants),<sup>69</sup> and needs to be ensured by incubating the droplets in culture medium. The “pseudo-one-phase” character and the aqueous environment enables not only washing and transfer into new media by batch-wise sieving but also penetration of low molecular weight molecules<sup>45</sup> (details see figure 3).

Other strategies to process single cells or molecular samples in microdroplets are water-in-oil emulsions which have been utilized in a variety of novel biological applications.<sup>70-74</sup> So called *in vitro* compartmentalization (IVC) has been demonstrated to be an effective tool for increased throughput in molecular evolution experiments. As an example for IVC, Nakano et al. demonstrated in 2005<sup>75</sup> single molecule reverse transcription polymerase chain

reaction (RT-PCR) in emulsion droplets and thus confirmed sufficient sensitivity to detect single RNA molecules. The main difference between emulsion and hydrogel microdroplets is the multiphase character of emulsions which prevents any transfer of hydrophilic agents. Still, single droplets in emulsions can be fused, e.g. as described by Tewhey et al. 2009 performing PCR in microdroplets.<sup>76</sup>

Emulsion microdroplets may be produced by stirring or shearing forces or by microfluidic devices.<sup>34</sup> Microfluidic devices have been gaining increasing importance for droplet-applications over the past decade, since tailored control of fluids (i. e. fission, fusion, sorting, and mixing of the droplet) adds potential of this technology for numerous biochemical applications.<sup>74, 77</sup> In addition to these manipulations, aqueous droplets may as well be solidified (e. g. by addition of alginate to the aqueous phase which may be complexed at any time during the emulsion process by bivalent cations) enabling encapsulation of cells, proteins, and DNA as well as the synthesis of micro/nano particles.



## **Scope of this thesis**

Sample processing by the aid of cell encapsulation in hydrogel microcapsules is a widely used measure due to relatively high production speeds and the aqueous “pseudo-one-phase” character. Latter enables penetration of low molecular weight molecules throughout the microcapsules and retention of high molecular weight molecules inside the capsules. Production rates of several thousand Hertz demand high throughput microcapsule processing techniques when screening for microbial libraries based on protein-function or DNA-sequence information. These challenges were addressed in the following chapters.

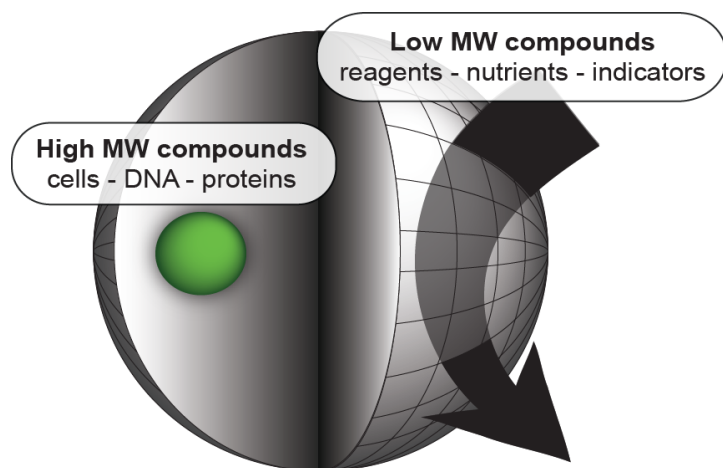
When it comes to cell encapsulation and processing of formed capsules, cells generally distribute randomly over the entire 3D suspension array. This distribution of individual samples according to Poisson is a well-known challenge in many processes that rely on random arraying.<sup>78</sup> In chapter II of this thesis, individual bacterial clones were embedded in alginate microcapsules and proliferated to colonies (see Figure 3). Processing of these colonies with a COPAS Plus Biosorter was achieved upon fluorescent signals generated by either fluorescent proteins expressed by the cells them self or by fluorescent dyes commonly employed in staining processes. This sorting strategy contains additionally a quality control protocol enabling the determination of the clonality state of individual nL-reactors and sorting of monoclonal microcapsules.

Even though novel high throughput sequencing technologies evolved within the past 5 years (see above), the demand for sequence based screenings remains, especially when large libraries with analyzable sequence space of more than 500 bp per clone are screened which are above the current limits of next-generation sequencing platforms (up to 250 bp). In chapter III, 20'000

individual colonies of a genomic cassava library were monoclonally enriched and screened for conserved DNA motifs (e.g. microsatellites) with a newly developed PCR-based screening technology.

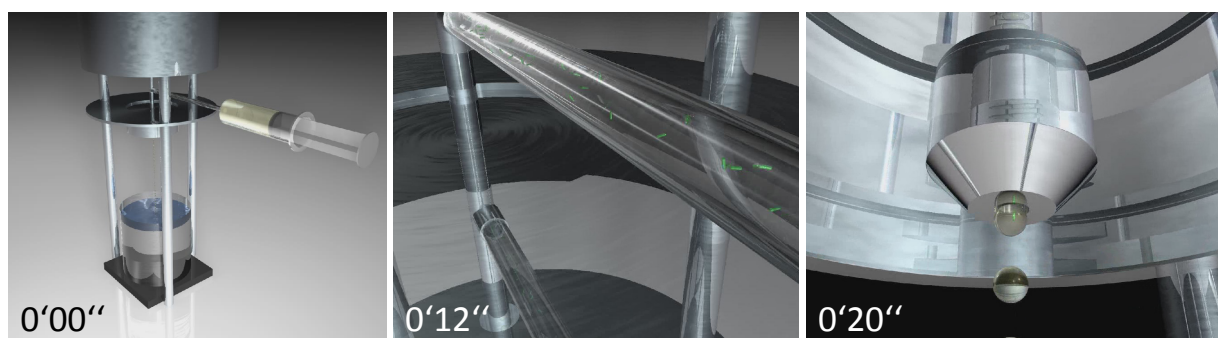
High throughput screening platforms are generally expensive and separation of positive clones from a background of unwanted clones is limited especially when sorting entire colonies. In this chapter, a novel separation technology for encapsulated colonies based on the production of a catalase enzyme as a reporter protein releasing gaseous oxygen upon induction with hydrogen peroxide was developed. The gas bubble formation within the microcapsules, suspended in an aqueous solution, causes a density reduction of the carriers leading to floatation of the carriers containing the highest levels of catalase enzymes inside (see movie 2 & 3 and the caption at the end of this chapter) A proof of concept for separation of positive candidates from 500'000 negative carriers within a time-span of a few seconds will be demonstrated with regard to survival rate and sorting efficiencies.

Finally, a fifth chapter documents efforts towards the application of alginate-based micropodroplets for a Sanger sequencing platform. Since quality as well as reproducibility was not competitive enough to cope with next-generation sequencing platforms, efforts were abandoned. However, some results and protocols shall not be deprived and are presented in this chapter.

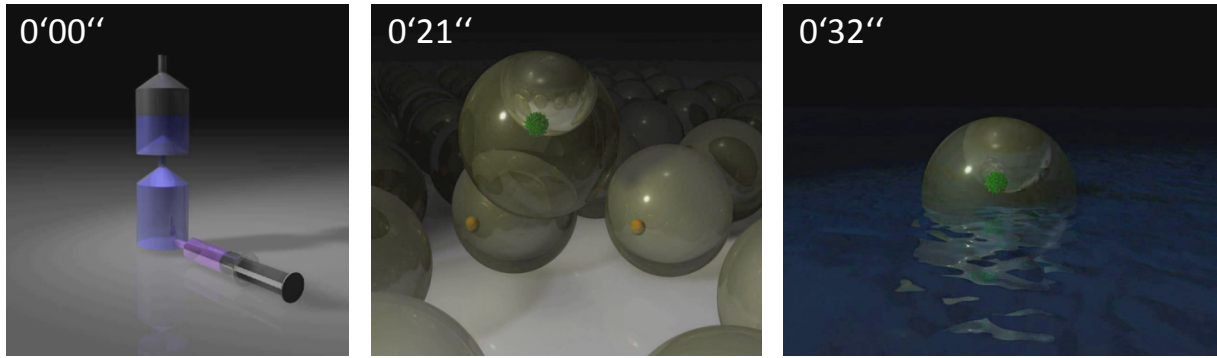


**Figure 3:** Hydrogel microcapsule produced with frequencies of >1kHz with a volume of a few nanoliters. The porosity of the capsule allows free diffusion of low molecular weight compounds whereas high molecular weight structures cannot penetrate nor escape. In such a manner nutrients

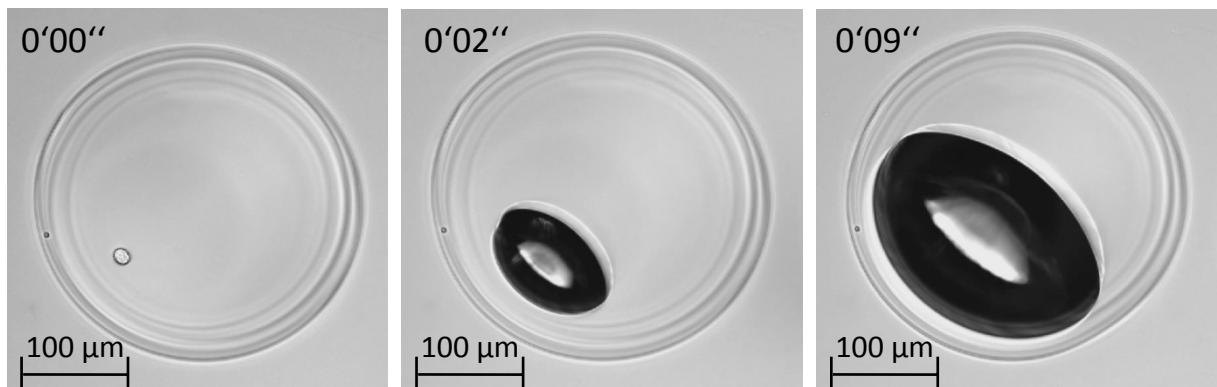
can diffuse through the hydrogel microcarrier and embedded cells of a library can therefore readily expand to colonies which remain immobilized within the microcarrier. Compartmentalized colonies are subsequently batch-wise washed and incubated according to the screening's requirements.



**Movie 1** (the movies are on the supplementary CD as well as online on: <http://www.bsse.ethz.ch/bpl/media/mwalser>): Animation movie illustrating the cell encapsulation process by laminar-jet break-up technology. Cells (green rods) mixed with alginate are continuously injected into an encapsulator by syringe handled with a syringe pump (not shown). Generated laminar-jet of the alginate/cell suspension is disrupted when passing a vibrating nozzle. This process generates highly monodisperse droplets with randomly distributed cells inside. The droplets fall down into a hardening solution (usually one of the bi-valent cations, calcium, strontium, or barium) and gellify instantaneously upon contact with the solution's surface. Suspended microcapsules may afterwards be washed in a sieve and used for appropriate applications (not shown).



**Movie 2** (the movies are on the supplementary CD as well as online on: <http://www.bsse.ethz.ch/bpl/media/mwalser>): Animation movie of an example for catalase separation of positive (green colony) and negative (brown) colonies in hydrogel microcapsules in a separation chamber. Upon hydrogen peroxide injection with a syringe into the lower chamber of the separation device, the positive colonies, synthesizing substantially higher amounts of catalase than the negative colonies, begin to form an oxygen gas bubble and will ascend into the upper chamber (middle panel). Since the hydrogen peroxide is injected into the lower chamber, the concentration of hydrogen peroxide (an antibiotic agent) is kept at a minimum in the upper chamber. As soon as all positive microcapsules have reached the upper chamber, one may close the interconnection between the two chambers and harvest the positive fraction of the upper chamber by sieving and washing of the microcapsules.



**Movie 3** (the movies are on the supplementary CD as well as online on: <http://www.bsse.ethz.ch/bpl/media/mwalser>): Microscopic image of an agarose-embedded alginate microcapsule containing a colony with cells over-producing catalase. As soon as hydrogen peroxide is added to the sample, an oxygen bubble is formed within the microcapsule immediately. The movie is kept in real time illustrating that bubble formation takes only about a few seconds.

## **II. Isolation of monoclonal microcarriers colonized by fluorescent *E. coli***

## Abstract

Microencapsulation gains increasing importance for processing of bacterial libraries and especially in high throughput (HT) environments where  $>10^6$  samples per day are studied. As a rule, a one-to-one relationship between an individual cell and analytical results is of key importance. Ideally, each microcarrier would therefore contain exactly one cell or colony. However, synthesis of larger numbers of capsules containing exactly one cell is not feasible as during carrier-production cells are randomly distributed. The dilemma is that high dilution conditions will yield a satisfactory degree of monoclonality, but also a very large fraction of empty compartments, while distribution under low dilution generates unacceptable numbers of polyclonal compartments for whose removal no satisfactory technologies exist. Hydrogel carriers with a volume of 35nL were used as growth compartments for individual microbial colonies. *E. coli* cells expressing green fluorescent protein (GFP) were encapsulated at low dilution thereby intentionally producing a considerable amount of polyclonal microcarriers. Empty and polyclonal microcarriers were then removed from the desired monoclonal fraction by a COPAS Plus particle analyzer. The results were compared to model predictions in order to investigate possible limitations in the analysis & sorting of monoclonal microcarriers by COPAS.

Fluorescent *E. coli* cells (GFP) distributed randomly throughout the microcarrier population. Cells were successfully propagated to colonies in the microcarriers and enriched to 95% monoclonality by a COPAS sorter. Enrichment-efficiency was found to mainly depend on the colony diameter. With increasing colony size two contrary effects were observed: First, improved sorting efficiency due to increased fluorescence intensity and therefore higher detection efficiency,

and second, deterioration of sorting efficiency due to occlusion occurring in polyclonal carriers.

The combination of microencapsulation under low dilution conditions followed by HT sorting procedures is an efficient way for isolating larger amount of monoclonal carriers from bacterial libraries while concomitantly keeping the amount of empty carriers at a moderate level.

## Introduction

The enormous complexity of living systems makes HT-experimenting and analyses an indispensable tool for research and development in the life- and biosciences. Sample numbers exceeding  $10^6$  per day<sup>79-82</sup> require new technologies for sample miniaturization and highly parallelized processing.<sup>83</sup> Naturally, most experimental strategies rely on formation of spatially separated monoclonal cell assemblies at the beginning. If microbial species are studied, an approach that is frequently employed for generating larger numbers of monoseptic compartments is to first cultivate on agar plates and then to either experiment directly on the plate or to pick and incubate the colonies in a secondary array (e.g. a microtiter plate), that is then used in subsequent steps.<sup>80, 84</sup> A more space-saving method for cell separation and cultivation is microencapsulation in which single cells or colonies are trapped in miniature gel-microcarriers.<sup>85-89</sup> These microcarriers readily allow exchange of nutrients or reactants through the gel-structure but are well suited to retain cells. Microcarriers thus comprise the key properties of all cultivation vessels typically employed in biosciences, specifically allow biomass to grow under monoseptic conditions.

Preparation of microcapsules involves the mixing of a cell-suspension and a gel-precursor solution, typically an alginate or an agarose solution.<sup>90</sup> This suspension is then compartmentalized into single small droplets by one of several available techniques.<sup>91-93</sup> Next, the droplets are reinforced by gelation which in the case of agaroses is induced by cooling<sup>94</sup> and in the case of alginates by chemically induced cross linking.<sup>95, 96</sup> As a result, cells become immobilized within the microcarrier.

A direct consequence of this procedure is that cells are distributed at random (Poisson distribution) throughout the microcarrier population.<sup>97, 98</sup> An intrinsic



disadvantage is therefore that the initial number of cells per capsule can only be controlled by means of statistics. However, as microcarriers containing more than one monoclonal cell are prone to lead to ambiguous results during analysis, for HT experimenting usually compartments containing exactly one cell are desired. Especially for approaches where positive cells or clones are later subjected to rather laborious secondary screening or analysis methods, (e.g. biopanning, protein/DNA extraction or sequencing)<sup>39, 80, 84, 99-103</sup> larger number of false positives will result in costly and non-productive secondary analyses which in severe cases may off-set the advantages of microcarrier-based approaches.

A simple way to increase the fraction of monoclonal microcarriers is keeping the average degree of occupation low, i.e. to highly dilute the cells in the gel-precursor solution prior to microcarrier formation. For example, at an initial average cell concentration of 0.01 cells per microcarrier and a random distribution of cells over all microcarriers, almost 99.4% of all colonized capsules will harbor exactly one cell. On the other hand, this high degree of dilution leads to a high number of empty microcarriers and consequently to undesired large production volumes and long sorting and analysis times. In the example mentioned above, for instance, approximately 99.8% of all synthesized capsules will remain empty and  $5 \times 10^7$  microcarriers will have to be screened in order to analyze  $10^5$  monoclonal microcarriers. Thus, high dilution is only an option if very fast analysis methods are applicable.

A powerful method for rapid analysis is fluorescent-activated cell sorting (FACS), allowing more than  $10^5$  analytic events per second.<sup>104, 105</sup> To this end, FACS is a frequently employed technology for e.g. sorting of polymer particles and mammalian cells.<sup>104, 106</sup> However, if it comes to hydro-gel carrier sorting, FACS is rarely an option as only very small microcarriers (typically  $< 50 \mu\text{m}$ ) can

be processed, which – for this purpose – usually have to be prepared from a limited number of materials such as agaroses.<sup>89, 97, 107</sup>

Furthermore, FACS has been developed for cell- rather than for capsule sorting and is therefore susceptible to clogs if stiff matter such as hydro-gel carriers is processed. Continuous operation without occasional nozzle-clogging by hydro-gel carriers can therefore hardly be avoided, and consequently the very high effective rates may not materialize. Finally, the typical detection systems of FACS machines would not allow differentiating between monoclonal and polyclonal capsules, reinforcing the requirement to go to high dilutions in order to obtain suitable homogeneous monoclonal microcarrier populations.

In this paper, we describe a technology for isolation of monoclonal microcarriers colonized by green fluorescent *E. coli*. The microcarriers comprise highly monodisperse alginate hydrogel capsules of a volume of 35 nL and an average diameter of 400µm. The large dimensions of the capsules have the advantage that upon incubation in appropriate media, relatively large numbers of cells can be accumulated within the capsule's interior while keeping the capsule intact. However, production of larger numbers of monoclonal capsules by means of high initial dilution is prohibitive as production volumes and processing times would rapidly become too high for large-scale laboratory applications. Therefore, we embedded cells under conditions of low dilution adapting COPAS-technology (complex object parametric analyzer and sorter)<sup>108-110</sup> featuring analyses rates of 20-30 compartments per second as a tool for the differentiation between monoclonal microcarriers and empty or polyclonal ones after cell cultivation.

## Materials and Methods

### Chemicals

Bacto Yeast Extract (BYE), Bacto Tryptone (BT) and Bacto Agar used for media production were supplied by Becton Dickinson (Le Pont de Claix, France). Chloramphenicol (Cm) was obtained from GERBU (Gaiberg, Germany), Pronova ultra pure LVG alginate solution from Novamatrix (Oslo, Norway). All other chemicals, media, and media components were obtained from Fluka (Buchs, Switzerland).

### Preparation of cell suspension in alginate

Cultures of *E. coli* Top10 (F- *mcrA*  $\Delta$ (*mrr-hsdRMS-mcrBC*)  $\phi$ 80/*lacZ* $\Delta$ M15  $\Delta$ *lacX74* *recA1* *araD139*  $\Delta$ (*ara-leu*)7697 *galU* *galk* *rpsL* *endA1* *nupG*, Invitrogen, Carlsbad, CA) harboring plasmid pMMB207-Km14-GFPc with a *gfp-mut2* gene<sup>111</sup> constitutively expressed from an unregulated  $P_{tac}$  promoter were grown in LB medium<sup>112</sup> supplemented with 30 mg/L of Cm at 30°C in a test-tube to an OD<sub>600</sub> of 0.5. Next, the cells were 5-fold diluted with the same medium. An aliquot of 20  $\mu$ L of this diluted cell suspension was immediately diluted a second time to the required concentration into 31.5 ml salt-solution (9 g/L NaCl, 1 g/L KCl) and mixed with a graphite suspension. The addition of graphite became necessary because previous results had indicated that without it, microcarriers were too transparent for reliable subsequent COPAS analysis. To prepare the graphite suspension, an aliquot of 3.5 mL of a crude graphite suspension consisting of 10 g/L graphite powder (<1  $\mu$ m particle size), 20 mL/L ethanol, and 1g/L Triton X-100, was centrifuged, the supernatant was removed, the graphite-containing pellet was resuspended in 3.5 mL salt-solution (see above), and the resulting final suspension was then added to the cell suspension. After gentle inversion of the cell-graphite suspension for 2 min, it

was added to 35 mL of a sterile-filtered 3% alginate solution. The liquids were mixed by gentle inversion for 2 min and the resulting 1.5% alginate solution containing cells and graphite was rapidly transferred to 10 mL syringes for encapsulation.

### **Production of microcarriers by encapsulation**

For droplet formation by laminar jet break-up, 10 mL aliquots of cells suspended in the gel precursor solution were passed through an encapsulator (Nisco, Zurich, Switzerland) operated with a nozzle diameter of 150  $\mu\text{m}$  at 1050 Hz. The feed rate into the encapsulator was maintained at 3.8 mL/min by a syringe pump (TSE Systems, Bad Homburg, Germany). The resulting droplets of approximately 500  $\mu\text{m}$  diameter and 60 nL volume were collected in a continuously stirred beaker (120 mL maximal filling volume; magnetic stirring bar, 200 rpm) filled with 100 mL of a hardening solution (100mM  $\text{CaCl}_2$ ). The beads were allowed to mature in the hardening solution for 40 min which caused the droplets to shrink to diameters of approx. 400  $\mu\text{m}$  and a volume of approx. 35 nL. All steps described here were carried out at room temperature.

### **Proliferation of *E. coli* in microcarriers and determination of colony diameter**

**( $\phi_{\text{colony}}$ )**

The microcarriers were recovered from the hardening solution by sieving (100  $\mu\text{m}$  Falcon sieve; BD, Franklin Lakes, NJ) and washed three times in 150 mL of growth-medium (4 g/L BYE; 1 g/L BT; 1 g/L glycerol, 4 mM  $\text{CaCl}_2$ ; 10 mM Tris-HCl pH of 7.0; 30 mg/L Cm). Aliquots of approximately 2.5 g of wet microcarriers were added into Petri dishes containing 25 mL of growth-medium. The plates were covered with their lids and incubated at 30°C in an incubator.

For sampling, generally 50'000 microcarriers (3 mL) were withdrawn from the liquor, sieved, and washed three times with 50 mL washing solution (10 mM Tris pH 8.0, 10 mM CaCl<sub>2</sub>). The various fractions of washed microcarriers were then stored at 4°C, and analyzed together on the following day.

The average  $\phi_{\text{colony}}$  was generally determined by analyzing at least 30 colonies that have been grown within the microcarriers by fluorescence microscopy with a Zeiss Axiostar Plus microscope and the AxioVision software (50-fold magnification; excitation filter at 450/50 nm in combination with non filtered white light for illumination of the non fluorescent microcarrier structure; longpass emission filter at 520 nm ).

Cell survival rate upon encapsulation and cell expansion were determined by plating of an aliquot of the alginate-cell suspension on LB agar plates (30 mg/L of Cm; incubation over night at 30°C). Living cell counts were then compared with the number of *E. coli* colonies detected in the microcarriers by microscopy after 15 h of growth.

### **COPAS analysis of colonized microcarriers**

Microcarriers were spectrophotometrically imaged with a COPAS Plus particle analyzer and sorter (1000  $\mu\text{m}$  flow-cell) equipped with the "Profiler" software upgrade for pulse shape analysis (Union Biometrica, Holliston, MA). Pulse shape diagram-recording was triggered by the opacity signal (threshold > 25 AU; signal gain factor 1.5; measuring range 0 to 65'000 AU). Microcarrier's size is expressed as time-of-flight ([ToF]; gated range 400 to 750 ToF). Fluorescence signals were generally recorded at 510 nm, appropriate for the emission maximum of GFP at 510 nm (ex 488 nm, photon multiplier settings [PMT] 800 V; gain factor 1.0; measuring range 0 to 65'500 AU; peak-profiling was initiated

at > 20'000 AU). In case of a signal overlap, the relative minimum had to be more than 10'000 AU lower than the next relative or absolute maximum in order to still assign two peaks instead of one. The COPAS-device was operated at frequencies between 30 and 40 Hz.

### **Stringent sorting conditions**

In order to improve sorting quality, the COPAS particle analyzer can be operated in modes ensuring that only one microcarrier is in the sorting chamber at a given time, which is synonymous to saying that there should not be more than one microcarrier in the volume segment that is diverted into the collecting vessels after the analysis of one microcarrier. As access to the sorting chamber is not specifically regulated, there are a number of events where more than one microcarrier is in the sorting chamber. COPAS eliminates this source of error by transferring these events to the waste (coincidence settings on "pure"). This leads to a fraction of monoclonal microcarriers of approximately 20% that is lost during each sorting step when sorted at frequencies of 30-40 Hz.

### **Experimental determination of the colony detection efficiency (CDE)**

In COPAS experiments for CDE determination, typically 20'000 microcarriers were analyzed. As a first step, the average degree of occupation  $DO_{det}$  as measured by COPAS was determined by relating the number of COPAS-determined occupied microcarriers ( $n^{k>0}$ ) to the total number of measured microcarriers ( $n_{total}$ ).

$$DO_{det} = \frac{n^{k>0}}{n_{total}} \text{ (eq. 1)}$$

Here,  $k$  indicates the clonality class of a specific microcarrier and the notation as superscript indicates measurement by COPAS, i.e. superscripts of  $k=1$  and  $k=2$  indicate the number of monoclonal and biconal microcarriers as counted by the particle analyzer, respectively, and  $k>0$  includes all microcarriers for which at least one fluorescent colony peak was detected. As  $n^{k>0}$  increases over time (with increasing colony diameter, see below),  $DO_{det}$  is a function of time.

In order to define the absolute degree of occupation  $DO_{abs}$ , the analyses were performed on samples of microcarriers isolated after the maximally applied incubation time of 15 h. When occupied, these carriers contained large colonies of approximately 30  $\mu\text{m}$  in diameter that could be detected with a high degree of reliability. From this data set, 600 randomly chosen COPAS-recorded fluorescence pulse shapes were manually inspected to quantify microcarriers that had been erroneously counted as “unoccupied”. On the basis of these values, the colony detection efficiency CDE was then defined as:

$$CDE = \frac{DO_{det}}{DO_{abs}} \text{ (eq. 2)}$$

### **Quantification of false monoclonal microcarriers due to detector-specific limitations and colony occlusion**

In order to estimate the number of microcarriers of a certain clonality class, we assumed a distribution of colonies over the microcarriers according to Poisson:

$$f(\lambda; k) = \frac{e^{-\lambda} \lambda^k}{k!} \text{ (eq. 3)}$$

where  $\lambda$  indicates the mean number of colonies per microcarrier. This mean number was calculated either based on data obtained directly from the particle analyzer as

$$\lambda_{det} = \ln \left( \frac{n_{total}}{n^{k=0}} \right) \text{ (eq. 4)}$$

or as  $\lambda_{abs}$  with the corrected absolute number of empty microcarriers due to manual pulse shape inspection as described above.

The detector specific limitation

$$\alpha_{det} = \frac{n_2^{1,det}}{n^1} \text{ (eq. 5)}$$

is defined as the ratio of biclonal microcarriers (subscript k=2) that were mistakenly assigned to the monoclonal fraction (superscript k=1), e.g. due to the fact that only one of the two colonies was detected, to the total number of microcarriers sorted into the monoclonal fraction. However,  $n_2^1$ , the number of biclonal microcarriers erroneously sorted into the monoclonal fraction, is not available and needs to be estimated. Therefore, we can calculate the probability to detect one of the two colonies in a biclonal microcarrier as

$$P_{det1/2} = 2(CDE(1-CDE)) \text{ (eq. 6)}$$

If  $P_{det1/2}$  is multiplied by the absolute number of biclonal microcarriers ( $n_2$ , estimated by multiplying the fraction of biclonal microcarriers according to eq. 3 with the overall number of microcarriers in a sample) we obtain an estimate for the number of erroneously assigned microcarriers  $n_2^1$  due to low CDE. This number delivers an estimate for the detector-specific limitation  $\alpha_{det}$ , to the total number of monoclonally detected microcarriers  $n^1$ :

$$\alpha_{det} = \frac{P_{det1/2} f(\lambda; 2) n_{total}}{n^1} = \frac{2(CDE(1-CDE)) e^{-\lambda_{abs}} \lambda_{abs}^2 n_{total}}{2! n^1} \text{ (eq. 7)}$$

In order to account for the situation in which one colony was occluding a second when viewed along the axis of analysis, the probability  $\alpha_{occ}$  with which a microcarrier was assigned to the monoclonal fraction although it was biclonal was defined as

$$\alpha_{occ} = \frac{n_2^{1,occ}}{n^1} \text{ (eq. 8)}$$



This value could be accessed from experimental data (see Results section). To compare this experimentally accessible value with our expectations, we needed to estimate  $n_2^{1,occ}$ . Therefore, we made the following assumptions: The shape of the two colonies  $S^1$  and  $S^2$  is spherical, the two colonies are located inside the enveloping, spherical microcarrier with radius  $r > 0$ , they have the same diameter  $d$  with  $d > 0$  and  $d \leq r$ , and the centers of  $S^1$  and  $S^2$  are independently distributed inside the enveloping sphere. The probability for an overlapping projection of  $S^1$  and  $S^2$  onto a plane parallel to the flight axis and perpendicular to the axis of analysis, can be estimated as follows (figure 1A): Let

$$X^i = \begin{pmatrix} X_1^i \\ X_2^i \\ X_3^i \end{pmatrix} \text{ (eq. 9)}$$

be the random coordinates of the center of  $S^i$  with the suffices 1 to 3 as the three Cartesian axes. Without loss of generality it can be assumed that the given flight axis is the first Cartesian axis and the problem of having two colonies within the same scanning segment with thickness  $d$  reduces to determine

$$P[|X_1^1 - X_1^2| < d] \text{ (eq. 10)}$$

Note that in order to have an enveloping sphere without outgrowth of the colonies, we have (figure 1B)

$$\begin{aligned} X_1^i &\in [-r', r'] \\ r' &= r - \frac{1}{2}d \end{aligned} \text{ (eq. 11)}$$

The distribution function of  $X_1^1$  and  $X_1^2$  is given by:

$$F_{X_1^i}(x) = \frac{V(x, r')}{V(r', r')}$$

$$V(x, r') = \begin{cases} 0 & \text{if } -r' > x \\ \frac{\pi}{3} (3r'^2 x + 2r'^3 - x^3) & \text{if } -r' \leq x \leq r' \text{ (eq. 12)} \\ \frac{4}{3} \pi r'^3 & \text{if } x > r' \end{cases}$$

where  $V(x, r')$  is the volume of a sphere-segment with  $x$  as the axial position of the colony. The corresponding density function is:

$$f_{X_1^1}(x) = \frac{3}{4r'} - \frac{3}{4r'^3} x^2 \text{ (eq. 13)}$$

$-r' \leq x \leq r'$ , and zero otherwise

Using the assumption of stochastic independence between  $X_1^1$  and  $X_1^2$ , the bivariate density function is

$$f_{X_1^1, X_1^2}(x, y) = \frac{9}{16r'^6} (r'^2 - x^2)(r'^2 - y^2) \text{ (eq. 14)}$$

$[-r', r']^2$ , zero otherwise

Finding the probability of the event  $\{|X_1^1 - X_1^2| < d\}$  reduces then to:

$$P_{occ^*} = P[|X_1^1 - X_1^2| < d] = \int_{-r}^{r} \int_{-r-x-d}^{r-x-d} \frac{9}{16r'^6} (r'^2 - x^2)(r'^2 - y^2) dy dx \text{ (eq. 15)}$$

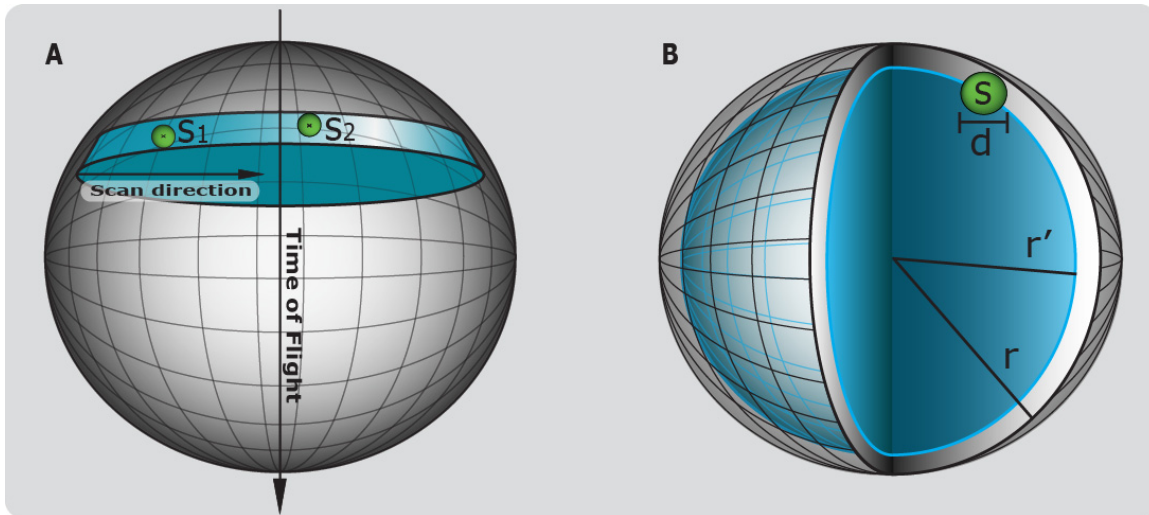
It can be shown that this simplifies to:

$$P_{occ^*} = P[|X_1^1 - X_1^2| < d] = \frac{6d}{5r'} - \frac{d^3}{2r'^3} + \frac{3d^4}{16r'^4} - \frac{d^6}{160r'^6} \text{ (eq. 16)}$$

This probability as a function of the colony diameter  $d$  and the microcarrier radius  $r$ , is multiplied with the probability to detect two colonies ( $P_{det\ 2/2} = CDE^2$ ) and the ratio of the absolute number of biclonal microcarriers  $n_2$  to the total number of detected microcarriers in the monoclonal fraction  $n^1$ , in order to obtain the estimated occlusion error  $\alpha_{occ^*}$ :

$$\alpha_{occ^*} = \frac{P_{occ^*} P_{det\ 2/2} f(\lambda; 2) n_{total}}{n^1} = \frac{P_{occ^*} CDE^2 e^{-\lambda_{abs}} \lambda_{abs}^2 n_{total}}{2! n^1} \text{ (eq. 17)}$$

Due to the small portion of microcarriers with more than two colonies, generally no efforts were undertaken to account for falsely assigned monoclonal microcarriers with  $k > 2$ .



**Figure 1: Colony occlusion in one-dimensionally scanned microcarriers.**

A) The two spheres  $S_1$  and  $S_2$  representative for two bacterial colonies are in the same scanning segment and can thus not be resolved in a COPAS profile.

B) Calculation of the probability that a microcarrier with two overlapping colonies is mistakenly classified as a monoclonal microcarrier: The enveloping sphere with radius  $r$  corresponds to a microcarrier. Based on experimental observations, it was assumed that the colonies (small spheres;  $s$ ) do not grow out of the enveloping sphere's periphery. Hence, the inner sphere with radius  $r'$  equaling  $r - \frac{1}{2}d$  was defined as the volume segment that can be occupied by colonies with diameter  $d$  (for details, see Materials and Methods).

## Results and Discussion

The use for cell libraries encapsulated into microcarriers is a powerful alternative to two dimensional screening arrays, e.g. microtiter plates. However, a requirement for application of encapsulated cell libraries is the reduction of experimental errors, i.e. wrong negative or positive results, inevitably occurring if larger numbers of macrocarriers contain more than one cell or colony are employed.

Thus, elimination of polyclonal and empty carriers prior to a screening is a key success factor. We therefore developed a sorting technology, based on a COPAS device, for the isolation of monoclonal microcarriers against a large background of polyclonal and non-colonized ones. Physical as well as device-specific parameters were determined and optimal operation conditions were defined.

### Growth kinetics of fluorescent *E. coli* in microcarriers

As a first step, growth of fluorescently labeled *E. coli* colonies in microcarriers was qualitatively and quantitatively described. Indicated by light microscopy studies, all colonies were generally globularly shaped which suggests that cells expanded equally in all three dimensions (see figure 2A). Survival rates measured before encapsulation and after cell-growth within the microcarriers indicated that generally about 95 % of the encapsulated cells survived the encapsulation & growth procedure.

Colony diameters ( $\varnothing_{\text{colony}}$ ) were measured only after 10 h of incubation as at incubation times of less than 10 h, an increasing number of encapsulated colonies (> 50%) cannot be detected by COPAS due to a too low fluorescence intensity (< 20'000 AU). However, the results of hourly measurements between

10 and 15 h of incubation (table 1) and indicated a steady increased of the biomass from  $\varnothing_{\text{colony}} 12.7 \mu\text{m} \pm 2.9 \mu\text{m}$  (10 h) to  $30.1 \mu\text{m} \pm 6.2 \mu\text{m}$  (15 h).

Assuming that colonies are generally shaped as either a sphere or an ellipsoid, and based on an average cell volume for single *E. coli* cells of approx. 2 fL,<sup>113</sup> the number of cells per colony could be estimated to be 250-1000 cells for  $\varnothing_{\text{colony}} = 12.7 \mu\text{m} \pm 2.9 \mu\text{m}$  and 3600-12500 cells for  $\varnothing_{\text{colony}} = 30.1 \mu\text{m} \pm 6.2 \mu\text{m}$ . These data suggest a doubling time between 1 and 2 h, which is in the range of the doubling time of 80 min measured ex-bead in a shaking flask with the same medium and growth temperature.

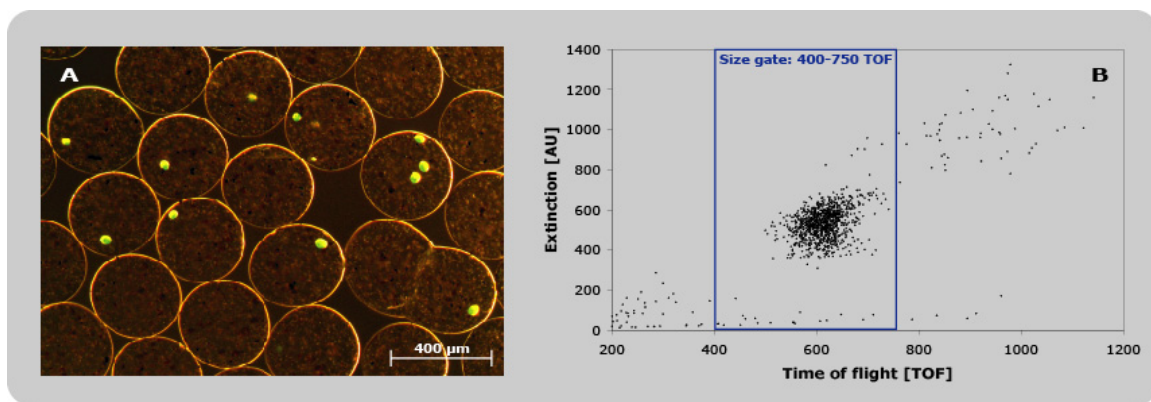
### **Colony-detectability is a function of the growth time**

As longer incubation times result in more biomass and the expression from the *gfp-mut2* gene is supposed to be constitutive, an increased intensity of fluorescence over time and therefore an improved detectability was expected. The results indicated further that the fraction of COPAS-detected occupied microcarriers ( $\text{DO}_{\text{det}}$ ) in a population of ~20'000 microcarriers indeed increased, from 36.7% after 10 h of growth to 54.7% after 15 h for a mean occupation of 0.8 colonies per microcarriers. For interpretation of all COPAS-results generally only single microcarriers (ToF 400 – 750) but not the multiplets (approx. 5%) were analyzed. A representative histogram is shown in figure 2B.

For determination of the COPAS detection efficiency, COPAS results were reevaluated by visual inspection of 600 consecutively recorded pulse-shape diagrams of microcarriers (15 h incubation time). The inspection clearly indicated that 14 colonized microcarriers had remained undiscovered by the analyzer as the fluorescence intensity did not reach the threshold criteria of 20'000 AU. The absolute degree of occupation  $\text{DO}_{\text{abs}}$  was therefore 57.0%. The detection efficiency (CDE; expressed as the ratio of COPAS assigned and

manually inspected DOs) was calculated for each growth time. As expected, the results (table 1a) indicated that the CDE increased with increasing growth time namely from 64.3% after 10 h ( $\varnothing_{\text{colony}}=12.7 \mu\text{m} \pm 2.9 \mu\text{m}$ ) to 95.9% after 15 h ( $\varnothing_{\text{colony}}=30.1 \mu\text{m} \pm 6.2 \mu\text{m}$ ).

As the precise correlation between the growth time,  $\varnothing_{\text{colony}}$  and fluorescence signal intensity are of no importance for this work, no efforts were undertaken to further quantify this effect.



**Figure 2: Microscope and COPAS-analyses of microcarriers harboring green fluorescent *E. coli* colonies**

*A: Hydrogel microcarriers harboring GFP-labeled *E. coli* colonies after 15 h of growth. As indicated, the colonies are generally of globularly shaped and rather similar in size. Microcarriers were produced by laminar jet break-up technology which results in bead-structures of very high monodispersity. However, statistically a minor fraction of multiplets is formed. A duplet is seen in the lower right corner.*

*B: Time of flight vs. extinction dot-plot of COPAS-analyzed microcarriers. Microcarrier-singlets are characterized by a TOF of 500-700. Above these values, microcarrier-multiplets are found. Signals below indicate very small particles such as “satellite droplets” formed during microcarrier-synthesis and other, not further characterized impurities.*

Growth time [h] /	10	11	12	13	14	15
$\varnothing_{\text{colony}} [\mu\text{m}]$	12.7±2.9	14.3±4.1	17.8±4.4	20.2±3.9	27.6±4.0	30.1±6.2
<b>a) COPAS data</b>						
$n^{k=0} [\%]^1$	63.3	57.8	53.6	50.8	45.5	45.3
$\text{DO}_{\text{det}} [\%]^1$	36.7	42.2	46.4	49.2	54.5	54.7
$\text{DO}_{\text{abs}} [\%]^1$	-	-	-	-	-	57.0
$\text{CDE} [\%]^2$	64.3	74.0	81.3	86.2	95.5	95.9
$\text{Lambda} [-]^3$	0.46	0.55	0.62	0.68	0.79	0.79
<b>b) Technical limitations</b>						
$\alpha_{\text{det}} [\%]^4$	24.1	18.3	13.6	10.1	3.4	3.1
$\alpha_{\text{occ}} [\%]^4$	0.3	1.1	1.9	3.9	6.6	7.5
$\alpha_{\text{occ}^*} [\%]^4$	1.7	2.3	3.3	4.0	6.4	7.0
$\alpha_{\text{occ}^*} / \alpha_{\text{occ}} [-]^5$	5.7	2.1	1.7	1.0	1.0	0.9

<sup>1</sup> Values expressed as the percentage of  $n_{\text{total}}$ .

<sup>2</sup> Expressed as the percentage of  $\text{DO}_{\text{abs}}$ .

<sup>3</sup> Lambda was calculated from the fraction of empty detected microcarriers for the determination of the colony distribution within the microcarriers according to Poisson.

<sup>4</sup> Values expressed as the percentage of  $n^{k=1}$ .

<sup>5</sup> Factor by which modeled  $\alpha_{\text{occ}^*}$  differs from the empirical data  $\alpha_{\text{occ}}$

**Table 1:** COPAS analyses of Poisson-distributed colonies of fluorescently labeled *E. coli* Top 10 [pMMB207-Km14-GFPc] in microcarriers.

A) COPAS analyses of fluorescent *E. coli* grown in microcarriers for 10 to 15 h, respectively.

B) Underlying effects influencing COPAS-performance during enrichment of monoclonal microcarriers.

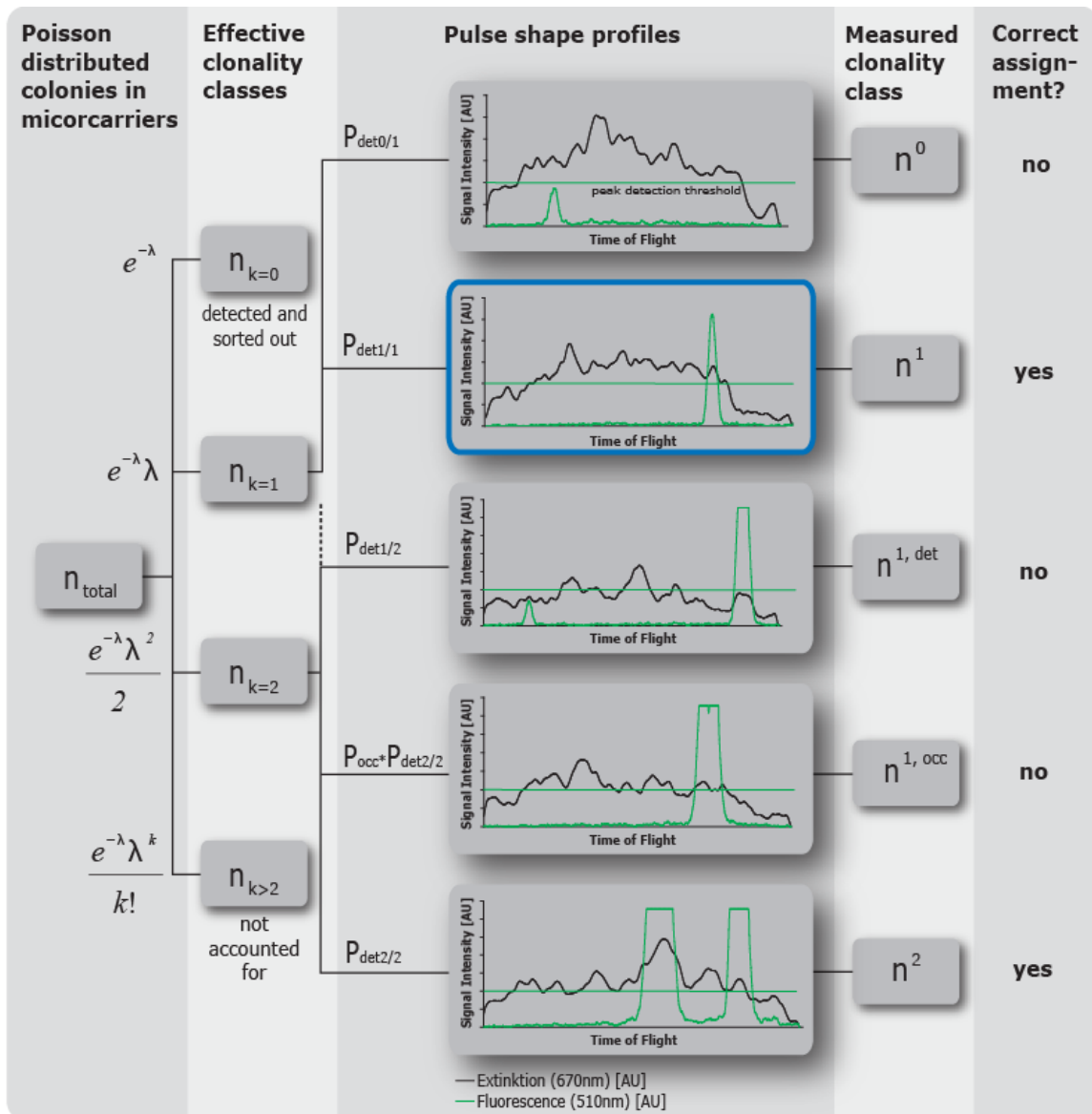
## Determination of clonality class

Several problems can be anticipated that might prevent the correct assignment of a clonality class by COPAS to a microcarrier (figure 3). First, it might be expected that colony size in polyclonal microcarriers is in general smaller than in monoclonal microcarriers, e.g. due to competition for nutrients, which might

limit colony detectability. Second, the observed distribution of colony diameters might lead to situations where a large colony might be paired with a small colony in one microcarrier and assigned to the monoclonal class, effectively making this a problem of CDE. Alternatively, two colonies in one microcarrier increase the chance that one of them is large enough to be detected thereby effectively increasing the chance that polyclonal microcarriers are detected at all. And third, the physical setting of the optical analysis along only one axis might lead to occlusion phenomena, where actually two colonies are in one microcarrier but lie – in the axis of analysis – behind each other and therefore to wrong assignment of polyclonals to the monoclonal fraction.

We can confidently exclude the first scenario, because we could in none of the numerous microscopic observations, carried out in the course of this study, observe a correlation of average colony size and number of colonies per microcarrier (data not shown). Problems with clonality class assignment due to CDE or occlusion phenomena are discussed in the following.





**Figure 3: Microcarrier’s clonality state is not always identified correctly upon COPAS pulse shape analyzes**

The event tree illustrates the underlying effects influencing the results obtained during enrichment of monoclonally colonized microcarriers. Randomly distributed colonies are subdivided into certain clonality classes (left side). Upon COPAS analyses (pulse shapes obtained after 15 h of incubation are shown) statistically always a certain fraction of microcarriers are assigned to a wrong clonality class (right side). An exemplary pulse shape profile of a correctly assigned monoclonal microcarrier is shown in the second box from the top.

### **Errors in clonality class assignment due to CDE**

As the probability of detecting at least one colony is higher in a microcarrier containing more than one colony than in a microcarrier with exactly one colony, an increase of the CDE will lead to a reduced number of polyclonals in the fraction of monoclonal microcarriers ( $n_{k>1}^1$ ). In order to estimate the error in clonality class assignment, we used the fraction of microcarriers detected as empty for calculation of  $\lambda$  and the corresponding expected numbers of mono- and polyclonal microcarriers assuming a Poisson distribution (table 1A). The probability  $\alpha_{\text{det}}$  to mistakenly classify biclonal microcarriers as monoclonal was calculated by comparing the results of manually counted microcarriers with the COPAS detected numbers (Tab 1B). After 10 h of growth and a CDE of 63.3%,  $\alpha_{\text{det}}$  was 24.1%, i.e. approximately every fourth microcarrier is expected to be erroneously assigned to the monoclonal category. With increasing CDE's due to higher fluorescence of the embedded colonies, the accuracy of the clonality class-assignment increased and  $\alpha_{\text{det}}$  is expected to decrease to 3.1% after 15 h of incubation time and a CDE of 95.9%.

In principle,  $\alpha_{\text{det}}$  could be reduced further by an improvement of the CDE. This might be achieved by increasing the cell-intrinsic fluorescence, e.g. by using more potent expression systems or fluorescent marker proteins.

### **One-dimensional projection leads to colony occlusion phenomena**

The assumption of random distribution of cells over hydro-gel microcarriers is experimentally firmly established.<sup>97</sup> Nevertheless, we observed in our analyses consistently more monoclonal microcarriers than expected (see supplement 1), while the numbers of polyclonal microcarriers remained smaller than expected, which indicated that the remaining error had to have a different source. As the objects in the analysis chamber of the COPAS are only analyzed in one and not

three dimensions a potential error-source is that colonies occlude each other. Two or more colonies within the same volume segment can thus not be resolved by the COPAS detector (see figure 1B). Therefore, it appears likely that a certain fraction of microcarriers containing more than one colony will be generally mistakenly regarded as monoclonal. This difference between the observed and the expected fraction for monoclonal microcarriers was termed the probability for occlusion  $\alpha_{occ}$  (table 1B). It can also be calculated as a function of the colonies' diameter under the reasonable assumption that the shape of the two colonies is spherical, the two colonies are located inside the enveloping spherical microcarrier, they have the same diameter, and their centers are independently distributed inside the enveloping sphere (see Materials and Methods). This calculated probability for occlusion was termed  $\alpha_{occ^*}$ .

According to this reasoning,  $\alpha_{occ^*}$  should gradually increase with increasing  $\phi_{colony}$  from 1.7% after 10 h of growth ( $\phi_{colony}$   $12.7 \mu\text{m} \pm 2.9 \mu\text{m}$ ) to 7.0% after 15 h ( $\phi_{colony}$  of  $30.1 \mu\text{m} \pm 6.2 \mu\text{m}$ ), indicating that, as expected, increasing colony size increases the chance of occlusion. The results presented in tabel 1b indicate that the numerically modeled  $\alpha_{occ^*}$  indeed follows the trend of the experimentally determined  $\alpha_{occ}$ . For  $\phi_{colony}$  between 20.2 and 30.1  $\mu\text{m}$ , the two results hardly deviate.

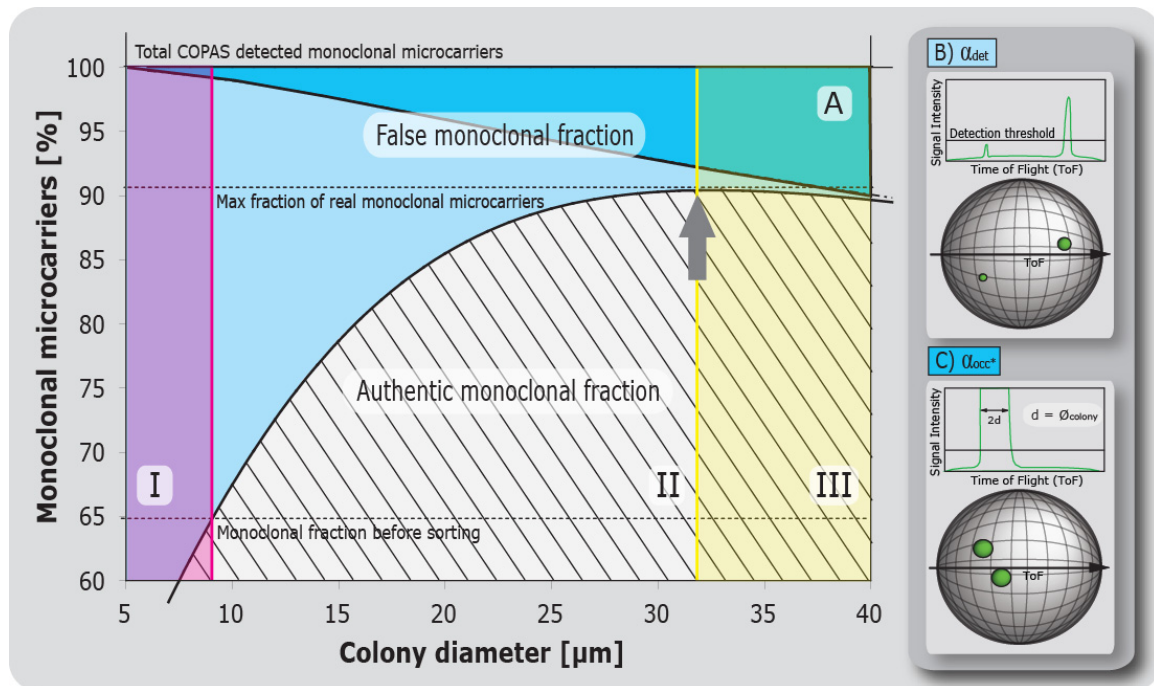
However, at  $\phi_{colony}$  smaller than 20.2  $\mu\text{m}$  modeled and experimental data deviate significantly and indicate an improved performance of the COPAS particle analyzer compared to the modeled data. The reasons for this improvement have not been firmly established, but can be rationalized as follows: At  $\phi_{colony}$  smaller than approx. 20  $\mu\text{m}$ , the peak resolution-capacity of the COPAS particle analyzer exceeds the predicted performance because small colonies (such as those with mean  $\phi_{colony}$  of 12.7  $\mu\text{m}$  and 17.8  $\mu\text{m}$ ) generate

relatively small peaks that might still be separable in spite of partial occlusion: In other words, their centers might be so far apart that the fluorescence intensities show a clear minimum between colony centers and two colonies can therefore be resolved. When the colonies become larger, the fluorescence intensity becomes larger, and even though there is again a fluorescence intensity minimum between the colony centers, the intensity in the minimum is so high that it is beyond the upper boundary of the detector. The latter situation is essentially exactly reflected in our model, where we assume that any kind of overlap, no matter how small, leads to complete occlusion. The former situation would allow a significant portion of partially occluded colonies to still be resolved, explaining the improved performance such as reflected by the experimentally determined  $\alpha_{occ}$ . The minor aberrations still observed for  $\varnothing_{colony} > 20 \mu\text{m}$  might be caused by a virtual peak-broadening due to scattered fluorescence radiated from larger colonies thereby causing an effective increase of the occluded area.

### **Operating parameters for an enrichment of monoclonal microcarriers**

With increasing colony size, two reciprocal effects influencing the accuracy of the COPAS-based particle analysis occur: First, improved CDE leading to a high  $\alpha_{det}$  and, second, colony occlusion a high  $\alpha_{occ}^*$ . Evidently the definition of an optimal protocol is therefore not trivial. For  $\varnothing_{colony} > 32 \mu\text{m}$  (sector III in figure 4), the negative effects caused by occlusion errors are expected to outweigh the benefits of increased CDE. For small  $\varnothing_{colony} (< 9 \mu\text{m})$ , the low CDE leads to very large errors due to the inability to detect all occupied microcarriers (sector I) and thus even to a depletion of monoclonal microcarriers compared to the fraction of monoclonal microcarriers before COPAS sorting (65%).

In sector II ( $9 \mu\text{m} < \phi_{\text{colony}} < 32 \mu\text{m}$ ), enrichment of monoclonal microcarriers is feasible but performance is strongly influenced by the average  $\phi_{\text{colony}}$ -values. Clearly, results improve significantly once  $\phi_{\text{colony}}$  exceeds  $25 \mu\text{m}$ . According to our estimates, optimum performance requires a  $\phi_{\text{colony}}$  of  $32 \mu\text{m}$ .



**Figure 4: Sorting efficiency of monoclonal microcarriers by a COPAS biosorter.**

A) The dark areas represent the total portion of carriers mistakenly identified as monoclonal as calculated via  $\alpha_{\text{occ}}^*$  and  $\alpha_{\text{det}}$ . The striped area indicates the authentic monoclonal carrier fraction within the totally sorted microcarriers as a function of the  $\phi_{\text{colony}}$ . The lower dashed horizontal line represents the portion of monoclonal carriers (as a percentage of  $n^{f-1}$ ) before sorting while the upper dashed line indicates the maximal portion of authentic monoclonal microcarriers that can be isolated by a single COPAS run at a given DO.

B & C) Physical limitations influencing  $\alpha_{\text{det}}$  and  $\alpha_{\text{occ}}^*$  (see results).

### Enrichment of monoclonal microcarriers by multiple sorting steps

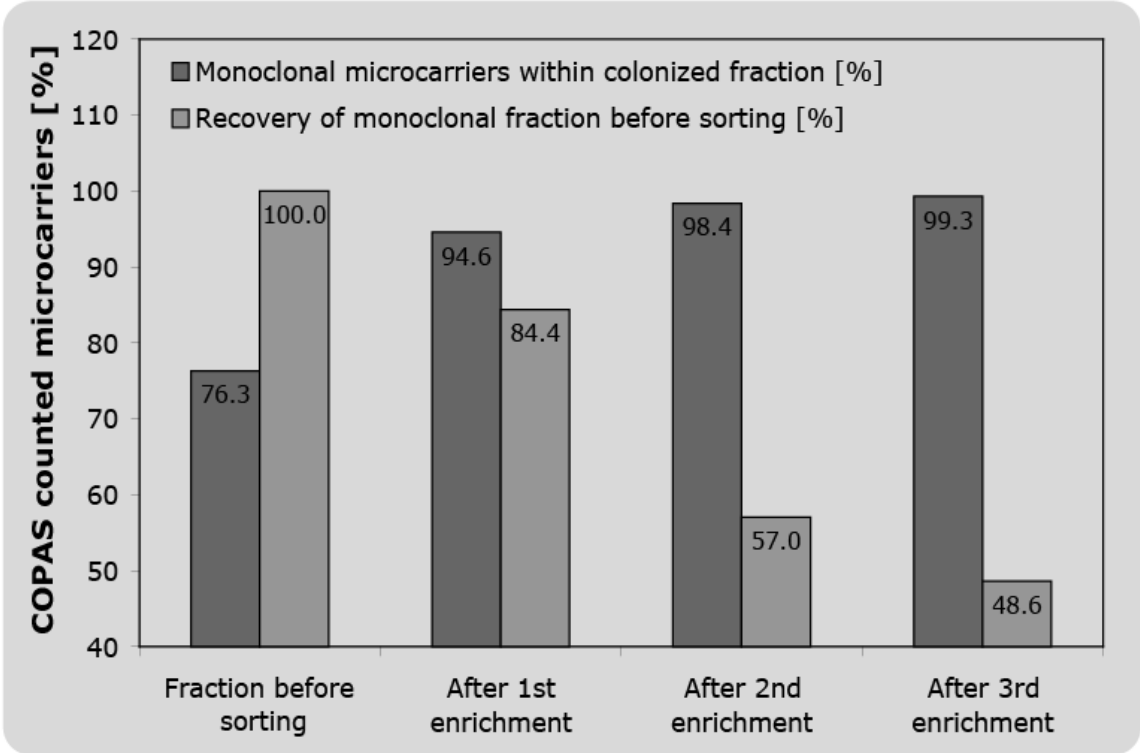
As occlusion of colonies during one-dimensional analysis is likely to become the most important source of error in analysis for large  $\phi_{\text{colony}}$ , we investigated whether repetitive sorting of the monoclonal fraction could improve the

situation. Geometrically, two colonies occluding each other in one axis cannot do so in the other two axes of a Cartesian system. As there is no preferred orientation of microcarriers in the COPAS detector, the chances that the same polyclonal microcarrier is erroneously sorted to the monoclonal fraction in a second pass through the analysis chamber is small (though of course not 0).

An experiment with three repetitive sorting steps, the results of which are shown in figure 5, was carried out. We used large  $\varnothing_{\text{colony}}$  of  $40 \mu\text{m} \pm 10.6 \mu\text{m}$  and assumed CDE to be essentially 100% ( $\alpha_{\text{det}} = 0\%$ ). During the first sorting step under stringent sorting criteria (see Materials and Methods), 76.3% of all colonized microcarriers of the initial fraction obtained after growth were classified as monoclonal ( $\mu=0.65$ ). Next, the monoclonal fraction of this first pass was re-sorted, and 94.6% of it was analyzed as monoclonal. A second re-sorting indicated that the percentage of analytically confirmed monoclonal microcarriers after the first re-sorting had increased to 98.4%, and a third re-sorting indicated the percentage in the fraction after the second re-sorting to be 99.3%, broadly confirming that re-sorting was an effective method to overcome occlusion phenomena.

However, the re-sorting steps disproportionately reduced also the total number of microcarriers in the monoclonal fraction. For example, the first sorting led to the loss of 16.6% occupied microcarriers from the monoclonal fraction obtained after growth. This problem becomes evidently accumulative with subsequent re-sorting steps: 27.4 and 51.4% of the initial monoclonal microcarrier fraction were lost during the first and second re-sorting, respectively. The most obvious explanation here is photo bleaching and stringent sorting criteria (Materials and Methods) lead to loss of actually occupied microcarriers during multiple resorting. Obviously, the exact number of resorting steps – the importance of weeding out false positives - needs to be

weighed against the danger of losing too many microcarriers, and the selection will depend on the downstream experimental requirements.



**Figure 5: Multiple re-sorting bypasses colony occlusion error**

Multiple sorting for enrichment of monoclonal microcarriers with colonies of a mean diameter of 40.0µm. Erroneously detected overlapping colonies are efficiently sorted out by repetitive COPAS runs until more than 99% of all sorted microcarriers are monoclonal (dark grey columns). Because of stringent sorting criteria approx. 50% of all monoclonal microcarriers are lost after three sorting steps (light grey columns).

## Summary & Conclusions

Microcarriers, composed of hydro-gel carriers that support microbial growth within the interior of the capsule, are a versatile tool for the highly parallelized cultivation of bacterial libraries. With an average diameter of 400  $\mu\text{m}$ , microcarriers by far exceed the size-limit of particles that can be sorted by FACS. Therefore, we used a COPAS Plus Biosorter, mostly employed for sorting of *D. melanogaster* eggs or *C. elegans* nematodes, as a tool for high-throughput analyses and imaging of microcarriers at rates of up to 50 Hz.

Here, we benchmarked and optimized the COPAS sorting technology for the isolation of large quantities of monoclonal microcarriers, i.e. microcarriers containing a single fluorescently labeled *E. coli* colony, from a background of empty and polyclonal microcarriers. The results indicate that enrichment efficiency is strongly affected by the colony detection efficiency (CDE). For *E. coli* Top-10 [pMMB207-Km14-GFPc], producing the green fluorescent protein *gfp-mut2*, efficient colony detection required an average colony diameter ( $\varnothing_{\text{colony}}$ ) of at least 30  $\mu\text{m}$ . Below this value, sorting performance started to readily deteriorate due to weak CDEs.

With increasing  $\varnothing_{\text{colony}}$ , colony occlusion becomes the main limitation for correct assessment of the clonality class, i.e. for the differentiation between monoclonal and biclonal microcarriers. On the other hand, colony detection efficiency improves with higher  $\varnothing_{\text{colony}}$  due to a stronger fluorescence signal. Therefore, the two major effects are countercurrent suggesting an optimum  $\varnothing_{\text{colony}}$  for sorting, which was identified to be between 30 to 35  $\mu\text{m}$ . Starting at a mean cell density of  $\lambda=0.85$  (~36% monoclonal microcarriers; ~19% polyclonal microcarriers), monoclonal microcarriers could be separated from empty ones in single run while keeping the background of unintentionally co-isolated polyclonal microcarriers at a minimum of ~10%. This indicates an enrichment



factor of 5.3 if the ratio between monoclonal and polyclonal microcarriers before and after sorting is considered.

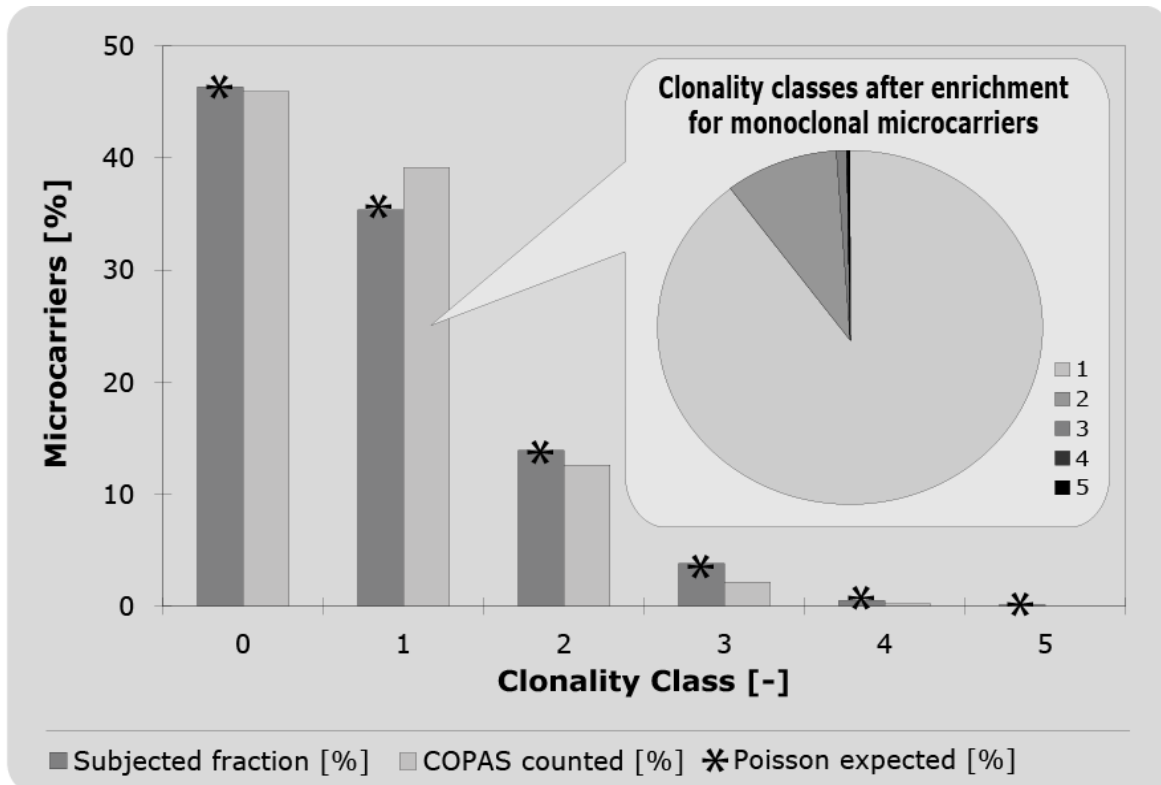
At  $\varnothing_{\text{colony}}$  of  $\sim 40 \mu\text{m}$ , where the CDE is supposed to be 100%, colony occlusion became the sole reason for mistakenly assigning polyclonal microcarriers to the monoclonal fraction. As expected, the purity of monoclonal microcarrier fractions can be improved by multiple sorting. This indicates that colony occlusion is a statistic process only occurring under conditions where microcarriers coincidentally pass the flow-cell in an orientation that prohibits resolution, i.e. an orientation in which colonies are staggered in a parallel plane relative to the sensor/laser-path. Furthermore, re-sorting lowers the isolated yield by roughly 20% per sorting cycle due to two reasons: First, GFP-bleaching due to repetitive radiation with the excitation laser, which in turn leads to lower CDE and therefore to decreasing isolated yields. Second, due to technical limitations, i.e. losses that generally occurred due to the concomitant occupation of volume segments within the flow cell by more than one microcarrier. However, if not yields but a very high fraction of isolated monoclonal microcarriers are prioritized, repetitive sorting is an interesting option as indicated by an enrichment-factor of 57 (99.3% monoclonal microcarriers in an isolated fraction) scored after three consecutive sorting steps.

All results were supported by simulations based on physical models which form the basis to readily adapt and predict COPAS performance for a variety of other sorting routines.

### **Acknowledgements**

We are indebted to the Swiss National Science Foundation's R'equipe program for generous support in acquiring the COPAS Plus Biosorter, to the Swiss

Commission for Technology and Innovation for grants to the project, and to H. Hilbi for the gift of plasmid pMMB207-Km14-GFPc.



**Supplementary material 1: Isolation of monoclonal microcarriers from a population of *E. coli*-colonized microcarriers by COPAS is biased by uncharacterized effects**

The following experiment inspired us for the paper: A batch of approx. 100'000 microcarriers harboring fluorescent *E. coli* cells at an average cell density of 0.77 was incubated for 17 h (average colony diameter of 40.2  $\mu\text{m}$ ). Of this batch, 1000 colonized carriers were isolated by COPAS (size gate: 400-750 TOF) and the clonality class of 820 randomly chosen microcarriers was determined manually under the light microscope. Our hypothesis was that *E. coli* cells should be distributed randomly (Poisson distribution). The results indicated that both set of data – those obtained by light microscopy (dark grey bars) and the values for a Poisson distribution (asterisk) – correlated very nicely thereby providing experimental evidence that cells were indeed distributed randomly.

Afterwards, another 10'000 of the microcarriers of the same batch were analyzed by COPAS (light grey bars) and the fraction detected as monoclonal was concomitantly isolated. Interestingly, about 9.6% more microcarriers ended up in the monoclonal fraction than one

would expect to find on the basis of the calculated values. To this end, COPAS performance was reevaluated using light microscopy as an orthogonal assay and the clonality state of 716 of the COPAS-sorted monoclonal microcarriers (see pie-diagram) was reassessed.

Interestingly, these results indicate that larger numbers (67 biclonal, 6 tri-clonal, and 1 penta-clonal microcarriers) of polyclonal carriers had been mistakenly attributed to the monoclonal fraction. As manual inspection under the light microscopic allows an error-free assignment of the clonality state, COPAS sorting-performance is therefore clearly influenced by certain effects that were identified and characterized further in the following paper.

### **III. Novel method for high throughput colony PCR screening in nanoliter-reactors**

## **Abstract**

We introduce a technology for the rapid identification and sequencing of conserved DNA-elements employing a novel suspension array based on nanoliter (nL)-reactors made from alginate. The reactors have a volume of 35 nL and serve as reaction compartments during monoseptic growth of microbial library clones, colony lysis, thermocycling, and screening for sequence motifs via semi-quantitative fluorescence analyses. nL-Reactors were kept in suspension during all high-throughput steps which allowed performing the protocol in a highly space-effective fashion and at negligible expenses of consumables and reagents. As a first application, 11 high quality microsatellites for polymorphism studies in cassava were isolated and sequenced out of a library of 20'000 clones in two days.

The technology is widely scalable and we envision that throughputs for nL-reactor based screenings can be increased up to 100'000 and more samples per day thereby efficiently complementing protocols based on established deep sequencing technologies.

## **Introduction**

Identification of conserved DNA-fragments in poorly characterized DNA samples (e.g. metagenomic or genomic libraries) is a challenge.<sup>114, 115</sup> We introduce a novel technology based on nL-reactors for screening of large libraries. The technology was adapted to the identification of conserved sequence motifs in the genome of cassava and we argue that any microbial library can be screened in this way. As key steps are based on well established and robust techniques (e.g. DNA amplification by microbial growth, PCR, and Sanger sequencing) the protocol can be easily extended to other applications.

### **Fishing for conserved DNA motifs**

A prerequisite for the successful sequencing of an individual DNA molecule out of a mixture of DNA-molecules is the dilution of the DNA molecules down to a level at which single molecules are obtained. Afterwards these single molecules are arrayed in a fashion guaranteeing that no cross-contamination between samples occurs. One of the most efficient methods for compartmentalization of single DNA-molecules is to clone them into replicating elements (e.g. plasmids, fosmids, or BACs)<sup>39, 116, 117</sup> and to transform microbial hosts with these. By doing so, single DNA molecules are arrayed in a very small volume which in the case of single *E. coli* cells, for instance, is in the order of a few femtoliter. The DNA is then rapidly amplified upon growth of the clones. However, in order not to mix the clones and therefore also the genetic information during growth, a secondary array for monoseptic (i.e. comprising offspring of one clone only) cell cultivation has to be employed. Such arrays are typically agar plates providing convenient means for monoseptic growth of single cells on solid support.<sup>118, 119</sup> However, processing and handling of larger quantities of plates employed for colony expansion and picking is rather laborious and requires a highly

automated environment.<sup>39</sup> Furthermore, agar plates cannot be used as reaction vessels for *in vitro* DNA manipulations (e.g. PCR or sequencing). Thus, cells have to be transferred into another array, typically a microtiter plate, thereby slowing down the process and requiring additional equipment. As a result, processing of 100'000 clones per day is probably already close to the practicable limit for most screening applications<sup>58</sup> employing microbial clones. This is one of the reasons why novel technologies<sup>114, 120</sup> from within the field of ultra-high throughput screening do not rely on the arraying of single molecules in cells anymore but employ immobilization on surfaces (e.g. microbeads<sup>121, 122</sup> or microchips<sup>123</sup>) or in microemulsions<sup>51, 124-127</sup> as main compartmentalization mechanisms.

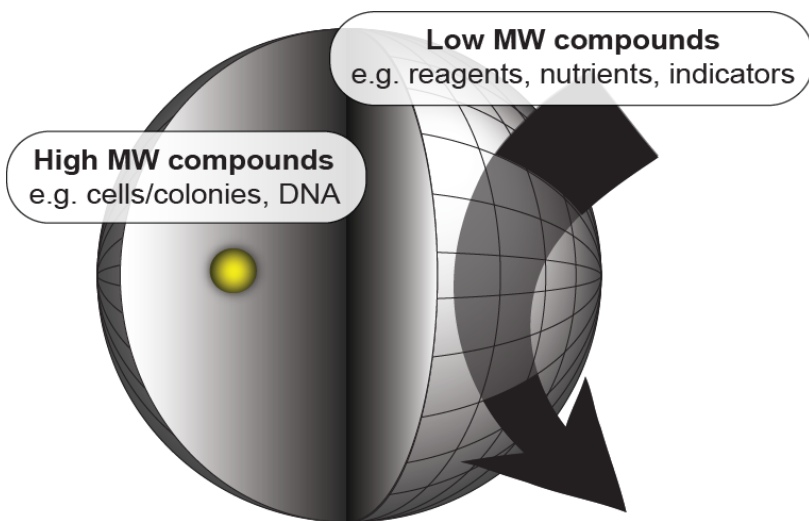
Still, single molecule arraying in cells has a number of advantages: First, error rates during amplification of the DNA by cell growth are much lower than *in vitro*.<sup>39</sup> Second, the information content per array element, i.e. the length of the screened fragment is much larger than in the competing *in vitro* systems.<sup>116</sup> Third, by the simple application of antibiotic selection pressure, array elements with contents (transformed clones) can be very efficiently separated from empty ones (i.e. clones that do not contain a DNA fragment).<sup>118</sup> Fourth, amplification of certain sequences such as repeating regions requires specific conditions *in vitro*<sup>120</sup> whereas cells generally provide an appropriate environment for polynucleotide amplification probably with the exception of a few cases where coincidentally toxic gene products are synthesized.<sup>128</sup> Fifth, cells can also be used for functional screenings, i.e. for screenings where first the phenotype is screened and only clones featuring the desired properties are sequenced.<sup>81</sup>

We demonstrate a PCR-based method for high throughput screening (HTS) of microbial DNA libraries at screening rates of at least 20'000 samples per run.

This requires the combination of conventional cloning and bacterial library construction methods with high throughput technologies for thermocycling and hit-candidate isolation. The core of the technology is the use of hydrogel microcarriers that are produced by laminar jet break-up technology and serve as growth and reaction compartments. The microcarriers have a diameter of  $\sim 400 \mu\text{m}$  and a volume of  $\sim 35 \text{ nL}$  and are subsequently referred to as nL-reactors. The key feature of nL-reactors is their capability to retain high molecular weight components such as cells, plasmid constructs or PCR products while at the same time allowing low molecular weight components (i.e. nutrients, or mono- and oligonucleotides for PCR) as well as some proteins (in this case DNA polymerase) to continuously access and leave the reactor (see figure 1). Due to these properties, PCR products are accumulated within the nL-reactors during thermocycling and positive samples can be straightforwardly isolated on the basis of the fluorescence-intensity of nL-reactors after staining of the embedded double stranded DNA.

In this article, we report a sequence-based screening of a cassava (*Manihot esculenta*) shot-gun library in *E. coli* for identification of microsatellites for gene-marker development (see figure 2). An equivalent of  $\sim 8$  million base pairs was screened within two days and it can be foreseen that the method can be readily adapted to uncover the diversity of gene sequences sharing conserved motifs, e.g. within non-sequenced organisms and environmental libraries. The process is easy to scale-up and we envisage that throughput can be easily increased further to 200 million bases or more per day.

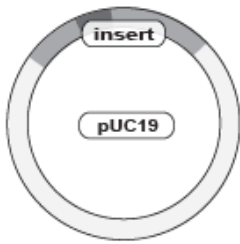




**Figure 1: nL-Reactors' characteristics.** nL-Reactors consist of a highly thermostable hydrogel and have a volume of 35 nL (diameter of 400  $\mu\text{m}$ ). The porosity of the capsule allows free diffusion of low molecular compounds whereas high

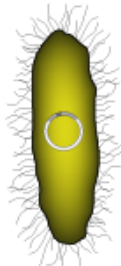
molecular weight structures cannot penetrate. Embedded *E. coli* cells can therefore feed on externally provided nutrients and readily expand to colonies that remain immobilized within the nL-reactor. Similarly, larger DNA elements such as plasmids that were liberated from the colonies upon lysis or PCR products formed by thermocycling are also retained, while PCR reagents can continuously access and leave the nL-reactor.

**FIGURE 2: FLOW CHART - HIGH THROUGHPUT COLONY PCR SCREENING IN NL-REACTORS**



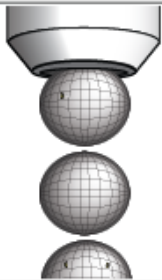
**1. Library construction**  
Digestion and ligation of target-DNA into plasmid DNA (according to current standard protocols).

14 h



**2. Transformation of fluorescent *E. coli***  
Transformation & selection in liquid culture. Transformed cells are suspended in alginate precursor solution for encapsulation.

1.5 h



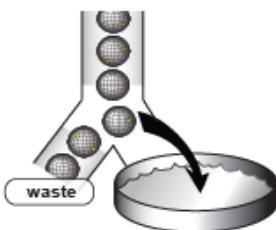
**3. Cell encapsulation in nL-reactors**  
Formation of alginate microdroplets containing Poisson-distributed cells by laminar jet break-up.

0.5 h (660'000 nL-reactors)



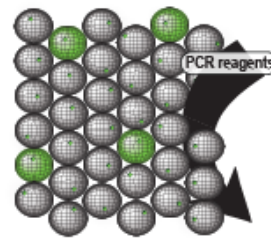
**4. Proliferation of individual clones in nL-reactors**  
Colony expansion within nL-reactors until ~50'000 cells are obtained.

15.5 h



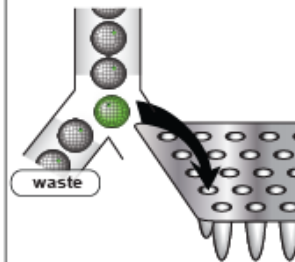
**5. Enrichment of monoclonal nL-reactors**  
Monoclonal nL-reactors are separated from empty and polyclonal reactors with a particle analyzer and sorter (COPAS).

3 h (200'000 nL-reactors)



**6. PCR screening of compartmentalized colonies**  
Monoseptic nL-reactors are washed & PCR is performed employing a customized thermo-cycler.

3 h (20'000 nL-reactors)



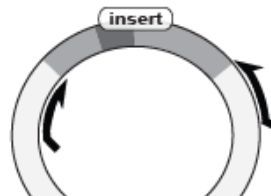
**7. Analysis & dispensing of PCR positives into 96-well plates**  
nL-Reactors are washed and stained for visualization of PCR positives. Positive nL-reactors are dispensed into 96-well plates.

0.5 h (20'000 nL-reactors)



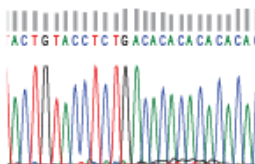
**8. Recovery of plasmid DNA**  
DNA of PCR-positive nL-reactors is extracted in 96-well plates.

3 h



**9. Plasmid-insert amplification**  
The entire inserts of the extracted plasmids are amplified by PCR.

2 h



**10. Sequence analysis with Sanger sequencing reaction**  
The PCR products are cleaned up and sequenced by conventional Sanger sequencing protocols employing capillary electrophoresis.

5 h

~48 h total screening time

## Materials and Methods

Unless mentioned otherwise, media and media components were obtained from Fluka (Buchs, Switzerland).

### Preparation of cell suspension in alginate and production of nL-reactors

Two 40  $\mu$ L aliquots of electrocompetent *E. coli* Top10 (F- *mcrA*  $\Delta$ (*mrr-hsdRMS-mcrBC*)  $\phi$ 80/*lacZ* $\Delta$ M15  $\Delta$ *lacX74* *recA1* *araD139*  $\Delta$ (*ara-leu*)7697 *galU* *galk* *rpsL* *endA1* *nupG*, Invitrogen, Carlsbad, CA) harboring plasmid pMMB207-Km14-GFPc<sup>129</sup> with a *gfp-mut2* gene constitutively expressed from an unregulated P<sub>tac</sub> promoter were transformed with 2  $\mu$ L of ligation mixture of a cassava library (gDNA digested with *Bsp143I* (Fermentas International, Burlington, Canada), size selected (350-700 bp) and ligated into a commercially *Bam*HI-digested and dephosphorylated pUC19 vector (Fermentas International, Burlington, Canada). Transformation was carried out with a Genepulser (Biorad, Hercules, CA) at 1.25 kV in a 1 mm cuvette. Cells were incubated directly after the pulse in 1 mL SOC medium (30 min; 37°C) and centrifuged (5 min; 5'000 g) in order to remove non-transformed plasmid DNA. The cell-pellet was resuspended in 1 mL SOC medium supplemented with antibiotics (30 mg/L chloramphenicol; 100mg/L ampicillin), incubated (45 min; 37°C) and centrifuged (5 min; 5'000 g). The supernatant was discarded. The cell-pellet was resuspended in 20 mL salt solution (9 g/L NaCl, 1 g/L KCl) containing 25 mg autoclaved graphite powder, added to 20 mL of a sterile-filtered 3% alginate solution in a 50 mL screw-top tube, and mixed by gently inverting the tube manually (5 min; room temperature).

The alginate-cell suspension was immediately processed with a laminar jet break-up encapsulator (Nisco Engineering AG; Zurich) with a nozzle of 150  $\mu$ m in diameter at a frequency of 1050 Hz. The resulting droplets of approximately

500  $\mu\text{m}$  diameter (65 nL) were collected in a continuously stirred beaker (120 mL maximal filling volume; 200 rpm) filled with 100 mL hardening solution (50mM  $\text{BaCl}_2$ ). The nL-reactors were allowed to mature for 30 min which caused the gel-capsules to shrink to diameters of approx. 400  $\mu\text{m}$  (35 nL).

### **Proliferation of *E. coli* in nL-reactors and determination of colony diameter**

The nL-reactors were recovered from the hardening solution by sieving (100  $\mu\text{m}$  Falcon sieve; BD, Franklin Lakes, NJ) and washed three times in 150 mL of growth-medium (4 g/L Bacto yeast extract; 1 g/L Bacto tryptone; 1 g/L glycerol, 1 mM  $\text{BaCl}_2$ ; 10 mM Tris-HCl pH 7.0; 30 mg/L chloramphenicol; 100 mg/L ampicillin). Aliquots of approximately 3 g wet nL-reactors were added to Petri dishes containing 30 mL of growth medium. The plates were covered with their lids and incubated for 14 h at 30°C. Afterwards ampicillin was added to the medium to 100 mg/L and the plates were incubated for an additional 1.5 h (37°C). nL-Reactors were recovered from the dishes, sieved, washed three times with 50 mL washing solution (10 mM Tris pH 8.0, 0.1 mM  $\text{BaCl}_2$ ; pH 8.0) and kept on ice until processing by COPAS (Complex Object Parametric Analyzer and Sorter; Union Biometrica, Holliston, MA). Afterwards the average colony diameter was determined by analyzing 30 colonies that had been grown within the nL-reactor by fluorescence microscopy under the assumption that a typical *E. coli* cell volume is 2 fL.

### **COPAS analysis for enrichment of monoclonal microcarriers**

Monoseptic nL-reactors harboring only one colony were enriched by COPAS sorting employing the “Profiler” software<sup>130</sup>. Pulse shape diagram recording was triggered by the opacity signal (threshold >25 AU (arbitrary units); signal gain factor 1.5; measuring range 0 to 65'000 AU). nL-Reactor size is expressed

as time-of-flight ([ToF]; gated range 400 to 750 ToF, arbitrary units). Fluorescence signals for colony detection were generally recorded at 510 nm, which is the emission maximum of the employed GFP (ex 488 nm, photon multiplier settings [PMT] 800 V; gain factor 1.0; measuring range 0 to 65'500 AU; peak-profiling was initiated at >20'000 AU; gated range 60'000 to 65'500). The COPAS-device was operated at an average frequency of 30 Hz and coincidence settings for dispensing into microtiter plates were adjusted to "pure"-mode which guarantees that nL-reactors are only sorted-out if no other reactor is coincidentally within the same droplet. A detailed description of the cell encapsulation, proliferation, and isolation procedure employed for enrichment of monoseptic nL-reactors by COPAS, was published elsewhere<sup>130</sup>.

### **In-bead PCR screening of embedded colonies**

A total number of 20'000 enriched monoseptical nL-reactors was suspended in a 50 mL Falcon tube in 10 mL of ddH<sub>2</sub>O. The cells were lysed by heat (10 min in a water bath at 96°C). The microcarriers were washed once in 50 mL washing solution and three times in 50 mL ddH<sub>2</sub>O and directly added to a poly(fluoro acrylate) PCR-bag (Welch Fluorocarbon, Dover, NH; 100 x 100 mm; thickness: 50 µm) containing 10 mL PCR reagent (ddH<sub>2</sub>O containing: 1 M betaine; 5 % DMSO; 1 x PCR buffer; 0.1 mM BaCl<sub>2</sub>; 0.2 mM of each dNTPs; 0.2 µM primer M13F (-21); 0.2 µM primer GA<sub>9</sub>; 700 Units *Taq* polymerase; Genscript, Piscataway, NJ). All gas bubbles were removed, the PCR-bag was air-tightly sealed, and cycled on a custom-built flat-screen thermocycler that consisted of a chamber flanked by aluminum plates which contained channels for water circulation for temperature control (32 cycles; 90 sec at 96°C; 180 sec at 55°C) exerted by two water-baths (Huber Polystat CC3; set to 96°C and 56°C, respectively). Numerically, controlled thermo-switching between the two

water-cycles was realized with four valves (m&m international, Bedford, UK) controlling the in- and outlet of the water-baths and a customized LabView (National Instruments; Austin; TX) program controlling the relay-switch station USB-Erb24 (Measurement Computing Corp., Norton, MA). After cycling, the beads were recovered from the bag by sieving, washed with 50 mL washing solution and 250 mL ddH<sub>2</sub>O and suspended in a Petri dish with 20 mL washing solution containing 1x SYBR Green I dye (Invitrogen; Carlsbad; CA) in order to stain PCR-products prior to another COPAS analysis. Microscopic pictures were taken by a Zeiss Axio Star Plus fluorescence microscope (Carl Zeiss AG; Göttingen; Germany) using an excitation filter at 488 nm and an emission longpass filter >520 nm in combination with phase contrast microscopy.

### **Screening for PCR positive nL-reactors**

All nL-reactors recovered after thermocycling were added to the COPAS sample cup and analyzed. Both, object diameter (based on extinction) as well as fluorescence intensity (510 nm) of each nL-reactor, were measured concomitantly by pulse shape analysis (COPAS profiler software). First, the system was calibrated by analyzing 300 nL-reactors in order to set an adaptive fluorescence intensity threshold just above the main, non-fluorescent, population, which was evaluated to be at 780 AU (at 510 nm). Next, gates for a ToF value of 600-1000 AU were applied for dispensing of nL-reactors into a 96-well microtiter plate. For sampling of the control group, the same ToF gates were applied and 48 putatively PCR-negative nL-reactors featuring signal intensities below the adaptive threshold (600-780 AU) were sorted out. All samples were immediately processed further in order to recover the plasmids (see next section).

### **Plasmid extraction with cetyl trimethylammonium bromide (CTAB)**

For plasmid extraction, we adapted a method developed by Allen *et al.*<sup>131</sup>. An aliquot of 90 µL of freshly prepared and pre-warmed (50°C) CTAB DNA extraction buffer (100 mM Tris-HCl (pH 8.0); 1.4 M NaCl; 2.5% w/v CTAB; 0.5% w/v N-lauryl sarcosine) was added to a well containing a single isolated nL-reactor (one reactor per MT-plate well), plates were air-tightly sealed by 8-cap strips, incubated in a PCR cycler (Mastercycler, Eppendorf, 65°C; 45 min) and every 5 min vigorously agitated by hand for 10 s in order to dissolve the nL-reactor. After centrifugation at 3'200 g for 30 s, 90 µL of chloroform/isoamylalcohol (24/1) were added and plates were agitated by hand until turbid emulsions were formed in the wells (approx. 15 sec). Plates were centrifuged (30 min at 3'200 g) and an aliquot of 33 µL was withdrawn from the aqueous phase, transferred to another microtiter plate, and DNA was precipitated by ethanol (77 µL per well). After over night incubation at -80 °C, samples were centrifuged (45 min; 3'220 g; 0°C) and supernatants were discarded prior to washing (100 µL of 70% EtOH, chilled to -20°C and centrifuged for 30 min; 3'220 g; 0°C). Once more, supernatants were discarded and the remaining liquid was allowed to evaporate at room temperature.

### **Insert amplification by PCR**

PCR reagents (25 µL comprising: 1x PCR Buffer, 200 µM of each dNTP, 200 nM of primer M13F(-43) and M13R(+86), 1.75 U *Taq* DNA polymerase) were added to the dried pellet and the plasmid inserts were amplified by PCR (denaturation at 95°C for 1 min, then 40 cycles of : 95°C for 30 s, 55°C for 30 s, 72°C for 1 min, and final elongation: 72°C for 2 min). After cycling, 5 µL of each sample were loaded on a 1.5 % agarose gel in order to determine the number of PCR-amplified fragments per sample, the insert size as well as the approximate PCR

product concentration. All samples yielding a single band of at least 100 ng of DNA on the gel (equivalent to 20 ng/μL concentration in the well) were sent out for Sanger sequencing with a M13 reverse primer (GATC Biotech, Konstanz, Germany).

### **Polymorphism test**

Clone-redundancy was tested with the DNA Star Software (DNASTAR inc., Madison, WI). Default assembly parameters were used: match size: 12 bp; minimum match percentage: 80%; minimum sequence length 100 bp; maximum added gaps per kb in contig: 70 bp; maximum added gaps per kb in sequence: 70 bp; last group considered: 2; gap penalty: 0.00; gap length penalty: 0.70.

Only sequences with a match to the screening primer of 10 bases or more were used to subsequently design microsatellite markers. Next, PCR primers flanking the CT-rich region were designed by the primer-3 software; (<http://fokker.wi.mit.edu/primer3/input.htm>). Amplicon length polymorphism of microsatellite markers was tested by PCR on 7 different cassava cultivars (91/02322; TMS60444; 95/0306; 98/0002; TAI-8; PER-183; COL-1505) using FAM-labeled primers. PCRs were performed in a final volume of 20 μL: 1x PCR Buffer, 200 μM each dNTP, 200 nM of each primer, 5 ng template; 1.75 U *Taq* DNA polymerase; cycling conditions: initial denaturation: 95°C for 1 min; 35 cycles: 95°C for 30 s, 56°C for 30 s, 72°C for 40 s; final elongation: 72°C for 2 min. Fragments were separated by capillary electrophoresis on an ABI 3700 (Applied Biosystems; Foster City; CA) and analyzed by the GeneMarker (Softgenetics; State College; PA) software. Polymorphic information content (PIC) was calculated according to Anderson *et al.*<sup>132</sup>.



## Results

The objective of this study was the detection and sequence determination of specific DNA-motifs from a genomic cassava library using an essentially freely scalable approach employing gel-like, suspended nL-reactors as reaction compartments.

### Cell encapsulation and proliferation in nL-reactors

The feasibility of the approach was investigated for a genomic cassava library (average insert length 350-700 bp) in *E. coli*. The sample-throughput was monitored over all process steps (see table 1). Freshly transformed cells synthesizing a green fluorescent protein (GFP) from *Aequorea victoria* were embedded into nL-reactors (660'000 reactors total; average degree of occupation with colonies 10.8 %) and propagated from single cell status to colonies of approx. 50'000 cells. This was confirmed manually by fluorescence microscopy. Next, 20'000 monoseptic nL-reactors were isolated by COPAS-sorting<sup>130</sup>. During this step, the number of reactors containing no or more than one colony, and reactor multiplets<sup>133</sup>, inevitably formed during nL-reactor production, was reduced to a minimum (table 1).

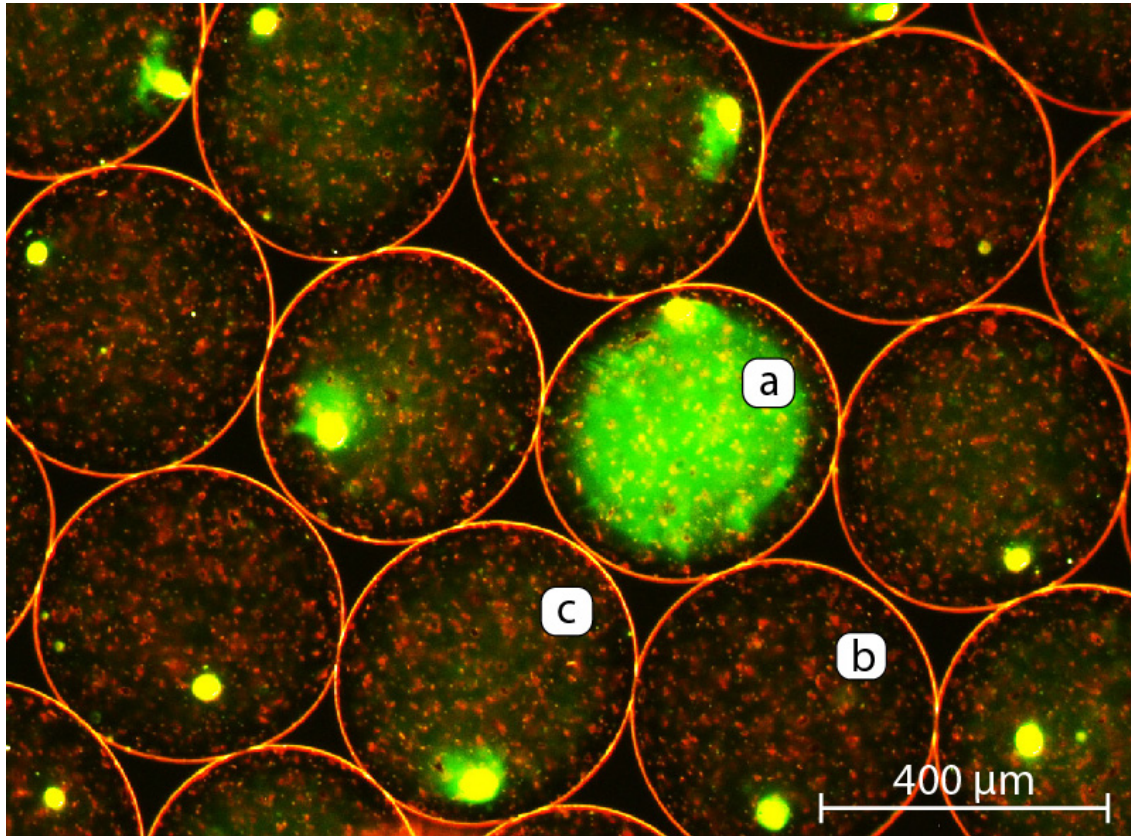
Step	Processed nL-reactors	Selected events	Array format
Enrichment	193'000 <sup>(a)</sup>	20'000 <sup>(b)</sup>	Suspension
PCR & hit selection	20'000	66 <sup>(c)</sup>	Suspension
Plasmid isolation	66	52 <sup>(d)</sup>	96-well plate
Insert sequencing	52	52 <sup>(e)</sup>	96-well plate
Polymorphic microsatellites	37 <sup>(f)</sup>	11 <sup>(g)</sup>	96-well plate

**Table 1: Sample flow over all process steps.** Overview over sample streams during HTS for novel SSRs (short sequence repeats) in a cassava genomic library.

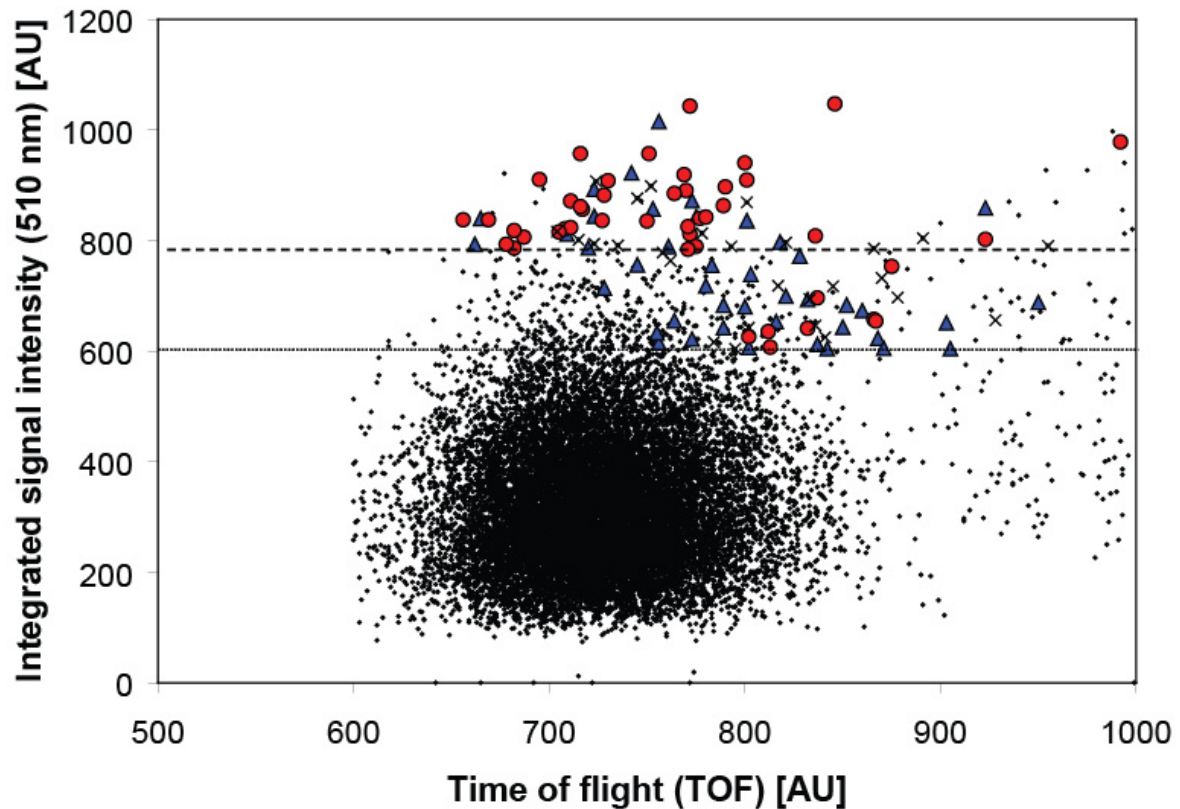
- (a) The subset of the 660'000 empty and colonized nL-reactors that was analyzed and sorted in order to isolate 20'000 monoseptic nL-reactors; colony diameter of  $32 \mu\text{m} \pm 7 \mu\text{m}$
- (b) Sorted and monoseptically enriched nL-reactors (still containing approximately 1 to 2% polyclonal reactors and roughly 2% multipliers)<sup>130</sup>
- (c) nL-reactors displaying fluorescence intensity >780 AU dispensed into microtiter plate
- (d) Plasmid extraction confirmed by PCR
- (e) Sanger sequencing
- (f) Sequences matching the screening primer by 10 or more bases  
(8 sequences were redundant and not used for polymorphism studies)
- (g) Based on polymorphism tests employing seven cassava cultivars (for details see text)

### **Highly parallelized nL-PCR and PCR-product detection**

All 20'000 monoseptic nL-reactors were pooled in one reaction compartment and subjected to 32 rounds of 2-step thermocycling with a vector-specific forward primer and a reverse primer with specificity for the desired SSR-sequences, i.e. poly(CT) repeats (see supplementary figure. 1). After thermocycling, the bead population was stained with the DNA double-strand specific dye SYBR Green I (see figure 3) and 66 fluorescent reactors were isolated in another COPAS run (figure 4).



**Figure 3: Monoseptic nL-reactors after 2-step PCR and staining of double-stranded DNA with SYBR Green I.** Double-stranded DNA entangled in nL-reactors was stained with SYBR Green I dye and fluoresces in a green-yellowish color. a) nL-reactor filled with fluorescent PCR product. b) nL-reactor without PCR product, only the SYBR Green I dye binding to the DNA of the lysed colonies fluoresces, though small coronas around the colony debris might develop (c). The small reddish spots within the reactors are graphite particles (see Materials & Methods).



**Figure 4: nL-Reactor analysis by COPAS after thermocycling.** nL-Reactor fluorescence (excitation: 488 nm; emission: 510 nm) and size (indexed by ToF) were COPAS-analyzed and 66 reactors above an adaptive fluorescence threshold of 780 AU (dotted line) were isolated. Due to apparatus specific limitations of the COPAS sorter, not all but approximately 80 % of all reactors above the dotted line could be isolated. In the region between 600 (solid line) and 780 AU (dotted line), 48 nL-reactors were randomly collected and served later as a control. The number of nL-reactors containing inserts that matched to the SSR-specific primer by at least 10 bases (red circles) amounted on 56 % above the threshold and on 17 % only in the control. Similarly, 23 % of all inserts in the putative positive fraction yielded a match of less than 10 bp (blue triangle) while in the control 56 % belonged to that class. nL-Reactors from which plasmid isolation and insert amplification by PCR did not succeed after liquefaction of the gel-carrier are labeled by an "X". Signals beyond TOF = 900 AU stem from nL-reactors with multiple volumes (about 2 % of the total) that have not been completely eliminated by COPAS during the first enrichment of monoseptic nL-reactors.

### **Identification of SSRs of at least 10 base pairs**

After isolation of all 66 putatively positive nL-reactors (fluorescent intensity >780 AU) and dispersion into a micro-titer plate, the plasmid content of the reactors was isolated and presence or absence of the plasmid, presumably harboring the SSR sequence, was verified by amplification of the insert with PCR using two vector-specific primers (see supplementary figure 1). Length and quantity of the PCR products was determined and the results indicated that 52 of the 66 reactors (79 %) carried a single PCR-product in sufficient quantity (>100 ng) to allow sequencing (supplementary figure 1). The sequencing results indicated that 37 plasmids (71%) showed a match of at least 10 bp between sequence and SSR-specific primer, which was defined as a criterion for the presence of a microsatellite. The 15 remaining nL-reactors carried an insert with less than 10 matching bases (4 matches: 6, 5 matches: 1, 7 matches: 3, 8 matches: 3, 9 matches: 2) while of the 37 sequences matching by 10 or more eight were redundant. In summary, the procedure delivered thus 29 potentially unique and useful SSR markers out of an original library of 20'000 monoseptically enriched nL-reactors. These sequences were then subjected to additional polymorphism studies in seven cassava cultivars.

Within the set of the 48 randomly collected nL-reactors with a fluorescence intensity below the adaptive threshold (i.e. collected from the interval between 600 and 780 AU, figure 4) that served as a control, plasmid isolation and insert amplification succeeded in 35 cases (73 %), and of these only eight clones (23 % of 35) had a primer match of 10 or more bases while the remaining 27 clones had a match of 9 or less. The results from this control group indicate that tailoring of the adaptive threshold is important in order to obtain a high ratio of nL-reactors featuring a one-to-one ratio between fluorescence intensity and primer match.

## **Design of microsatellite markers and polymorphism studies**

The sequence information obtained from all 29 clones (fluorescence intensity >780 AU) featuring a primer match of 10 or more was utilized for the design of primers for polymorphism studies. Generally a sufficiently large number of bases could be extracted from the data to allow the design of primers with melting temperatures of about 60°C, which were likely to target a unique address in the poorly characterized genome of cassava<sup>134</sup>.

The polymorphic information content (PIC), a value commonly used in population studies as a measure of genetic marker polymorphism, was then determined in 7 cassava cultivars (see Supplementary figure 1). In total, eleven high quality microsatellites with a PIC of more than zero had thus been isolated (table 2). Several microsatellite primer pairs (2, 11, 22, 24, and 27) were polymorphic in only two loci and therefore resulted in a PIC of 0.41. Samples 13, 16, 21, 26, and 28, led to the isolation of markers with an increasing PIC while one marker (sample 29) was found to be indeed polymorph in all cultivars and is therefore rather valuable for population studies.

Thus, 29 individual CT-containing clones had been isolated (table 2) from a sequence space of at least 7.8 Mbp (see discussion and conclusions for the rationale). This suggests an SSR frequency of one in 270 kbp. This frequency is well within the range of the CT-repeat abundances typically found in higher plants (one in 290 kbp)<sup>57</sup>. The overall efficiency of the protocol from PCR-product-sensing to polymorphism-isolation is 11 in 66 (17 %) which is comparable to process efficiency of approx. 20 % generally found in the field of microsatellite marker development.<sup>135</sup>

In the sampling window below the adaptive intensity threshold, 8 sequences still had a primer match of 10 or more. However, in only 4 cases, primer match regions were located at a sufficient distance from the multiple cloning site of

the vector to enable the straightforward design of primers for polymorphism studies in cassava. Based on the PIC-values, however, none of these microsatellites was of sufficient quality to serve as a marker for future polymorphisms studies in cassava.

Interestingly, the average of the polymorphic information content (PIC) as well as the percentage of polymorphic sequences therefore increased with an increasing number of bases in the insert matching with the screening primer.

	SSR match <sup>1</sup>	Integrated signal	Sensor fragment <sup>2</sup>	Sequence length <sup>3</sup>	PIC <sup>4</sup>	Subgroups of Cultivars <sup>5</sup>
#	[bp]	[AU]	[bp]	[bp]	[-]	[#]
1	10	811	59	245	0.00	(7)
<b>2</b>	<b>10</b>	<b>837</b>	<b>269</b>	<b>325</b>	<b>0.41</b>	<b>2; 5</b>
3	10	837	322	338	0.00	(7)
4	10	978	133	308	0.00	(7)
5	10	808 / 856*	133	174	0.00	(7)
6	11	789	396	546	0.00	(7)
7	11	806	59	418	0.00	(7)
8	11	890	171	504	0.00	(7)
9	11	940 / 840 / 786*	133	236	0.00	(7)
10	12	802	117	397	0.00	(7)
<b>11</b>	<b>12</b>	<b>818</b>	<b>200</b>	<b>222</b>	<b>0.41</b>	<b>2; 5</b>
12	12	821	149	610	0.00	(7)
<b>13</b>	<b>12</b>	<b>825</b>	<b>133</b>	<b>277</b>	<b>0.49</b>	<b>3; 4</b>
14	12	835	124	261	0.00	(7)
15	12	836	313	384	0.00	(7)
<b>16</b>	<b>12</b>	<b>863</b>	<b>168</b>	<b>384</b>	<b>0.73</b>	<b>3; 1; 1; 1; 1</b>
17	12	897	149	323	0.00	(7)
18	12	910	112	261	0.00	(7)
19	12	882 / 885 / 784 / 861*	104	332	0.00	(7)
20	13	842	118	263	0.00	(7)
<b>21</b>	<b>13</b>	<b>919</b>	<b>210</b>	<b>280</b>	<b>0.78</b>	<b>1; 1; 1; 2; 2</b>
<b>22</b>	<b>13</b>	<b>1047</b>	<b>157</b>	<b>416</b>	<b>0.41</b>	<b>2; 5</b>
23	13	909 / 654 / 793*	86	315	0.00	(7)
<b>24</b>	<b>14</b>	<b>815 / 871 / 696*</b>	<b>117</b>	<b>348</b>	<b>0.41</b>	<b>2; 5</b>
25	14	823 / 607*	117	420	0.00	(7)
<b>26</b>	<b>15</b>	<b>957</b>	<b>127</b>	<b>237</b>	<b>0.65</b>	<b>2; 2; 3</b>
<b>27</b>	<b>15</b>	<b>1043</b>	<b>117</b>	<b>462</b>	<b>0.41</b>	<b>2; 5</b>
<b>28</b>	<b>17</b>	<b>908</b>	<b>163</b>	<b>262</b>	<b>0.69</b>	<b>1; 1, 2; 3</b>
<b>29</b>	<b>17</b>	<b>957</b>	<b>157</b>	<b>283</b>	<b>0.86</b>	<b>1; 1; 1; 1; 1; 1; 1;</b>

**Table 2: Sequencing results of all clones featuring a primer match of at least 10 bases.**

*Bold font: Polymorphic microsatellite markers; primers and sequences are listed in Supplementary Material 1.*



- <sup>1</sup> *Number of base pairs matching with the sensing primer.*
- <sup>2</sup> *Length of the sensor fragment employed for identification of PCR positive clones expressed as the distance of the M13 forward primer to the location at which the screening-primer putatively bound to the insert.*
- <sup>3</sup> *The insert was sequenced employing a vector specific M13 reverse primer.*
- <sup>4</sup> *PIC = polymorphic information content; Polymorphisms were tested in seven different cassava cultivars.*
- <sup>5</sup> *Total number of cultivars carrying the same microsatellite allele.*
- \* *Redundant clones*

## Discussion and Conclusions

nL-Reactors were employed in a 3D suspension array for the compartmentalized growth of an *E. coli* library starting from single clones and directly afterwards for a PCR-based colony screening. Large numbers of both, colonized nL-reactors and post-PCR samples were analyzed and sorted by COPAS and the desired fractions of monoseptically inoculated nL-reactors and PCR-positive samples were directly isolated. Furthermore, plasmids identified on the basis of the PCR signal could be easily recovered from the nL-reactors and were subsequently subjected to further characterizations, in this case by Sanger-sequencing. The entire procedure did not require more than 48 h with an effective hands-on time of approximately 10 hours. It required only three microtiter plates, in contrast to current protocols for microsatellite identification which are substantially longer and more expensive (cells are plated on solid support followed by individual PCR screenings or colony hybridizations).

The longest distance between the vector-specific primer and an insert-sequence that was complementary to the screening primer was 391 bp (see table 2). This suggests that during PCR at least 391 bases of each of the 20'000 inserts harbored by the screened nL-reactors were scanned for regions matching the screening primer. Therefore, we reason that at least 7.8 Mbp were sampled during analysis of 20'000 colonies. However, as PCR in solution can be employed for amplification of fragments of several thousand base pairs, we believe that the length of the screened fragments can be improved further. If compared to the classical approaches with two dimensional arrays (i. e. microtiter plates or agar plates), the volumes employed here for 3D-arrays for cell expansion (600 mL) and PCR (10 mL), were rather low and could easily be scaled up by a factor of at least 10 without any substantial change of the

technology. Furthermore, all samples can be handled in one compartment during each single step in the entire screening protocol, which greatly facilitated the required manipulations such as addition of reagents or incubations. For that reason, no sophisticated robotics had to be employed.

The time-limiting step of the protocol is the standardization of the library after growth by COPAS sorting, which proceeds at a frequency of approximately 30 Hz. From our experience, COPAS devices can be operated for at least seven hours during an eight hour working shift. Still, throughput for this step will remain limited to roughly 750'000 events per shift. All other steps that are carried out in suspensions (nL-reactor production, cell growth, suspension PCR, and isolation of PCR-positives) require approximately another 2 days. Under the conservative assumption that inserts of 2 kb could be screened during PCR and assuming further that roughly 15 % of all synthesized nL-reactors contain single colonies (as it was the case in this example), the total number of screened bases could thus easily amount to approx. 200 Mbp per day.

Another obvious measure for further increasing the throughput is of course to replace COPAS-sorting by another technology such as fluorescence-assisted cell sorting (FACS). Sorting particles by FACS will however have to provide a volume in the lower picoliter (pL)-range and such particles can not be reliably produced by the laminar jet technology that was employed for capsule synthesis here. However, we are currently working on the development and standardization of the technologies required for the synthesis of hydrogel pL-reactors.

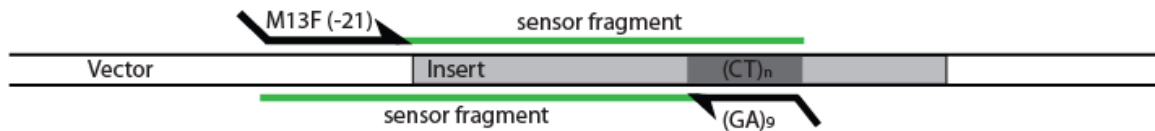
We argue that the sequence-specific amplification combined with high throughput screening makes this technology suitable for other protocols and generally for all cases where “rare” events are sought. Especially screening of gene homologues or miRNA precursors<sup>136, 137</sup> in non-sequenced species and particularly metagenomic approaches, e.g. “fishing” for new enzymes in

metagenomic libraries, or localization of insertion sequences in mutagenized genomes (e. g. transposons for random mutagenesis) will be greatly facilitated.

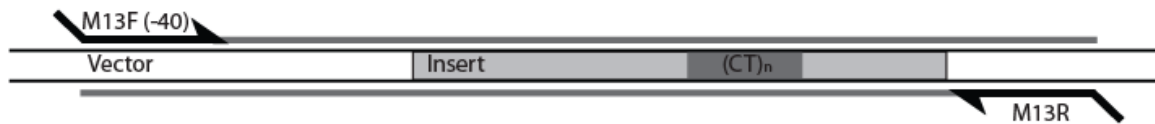
### **Acknowledgements**

We are indebted to the R'equipe program of the Swiss National Science Foundation for generous support in acquiring the COPAS Plus Biosorter, and to the Swiss Commission for Technology and Innovation for grants to the project. We also indebted to H. Hilbi for the gift of plasmid pMMB207-Km14-GFPc, to Bernhard Koller for the support with SSR libraries in preliminary tests, and to Maria Domenica Moccia for the gift of plasmid pCT16 used in optimization experiments. Additionally, we would like to thank the four anonymous reviewers for the valuable input.

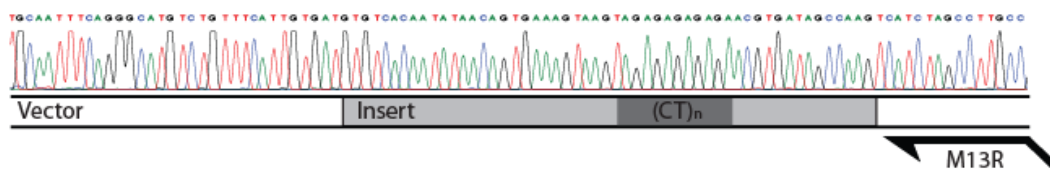
### I) Screening PCR



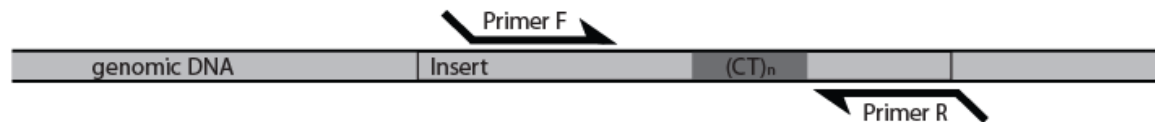
### II) Insert amplification



### III) Sequence analysis



### IV) Polymorphism test



## Supplementary Figure 1: Strategy for microsatellite identification

(I) *Identification of CT-repeats:* In order to identify inserts carrying a CT-repeat, a forward primer M13F (-21) complementary to the plasmid backbone and a screening primer (GA) complementary to the putative microsatellite sequence were used. In the presence of a CT-rich region, PCR-products (sensor fragments) accumulated in the nL-reactors. These fragments could be detected by fluorescence analysis after staining with SYBR Green I dye.

(II) *Insert amplification:* The plasmid were recovered and the entire inserts were amplified employing the two vector-specific primers M13F (-40) and M13R. In this way, flanking regions of the CT-repeat were read.

(III) *Sequencing:* The PCR product was sequenced using the M13R primer and fluorescently labeled ddNTPs.

(IV) *Polymorphism studies:* Primers were designed on the basis of the sequence information and the CT-rich region was amplified in seven different cassava cultivars in order to determine the cultivar-specific length of the repeat as a measure for degree of polymorphism.

**Supplementary Material 1: primer sequences and sequence information of polymorphic loci described in Table 2**

**Locus #2**

designed primers (5'-3'): CCAGAAGACTGTGCCACTGA / GAGGCTGCACTCCAGGTATT  
TTAGGGTTTTNCCCAGTCACGACGTTGTA AACGACGGCCAGTGAATTCGAGCTCGGTACCCGGGGA  
TCTTGTTCCAGA ACTTATTCTTTTTAGATTTTCTGTGGCCTCTGTCCCACTTCTTACTGCCTGCAACTGC  
TGCACCCAGAGAAGAAGAGTCTAACCTTTCCACCTAGGGTCTTGGAACCAGAAGACTGTGCCACT  
GACTGTTTTACCTTCCCCTCAACGATAGCACTAGCCTCCATTCTCCTAGCCATATCCACTATGGCATGA  
AAACTCTCCCTCTCTGCTGACTGTATCAAAGAGGAATACCTGGAGTGCAGCCTCATAATATATCTCCC  
TTGACTTCTTTGATCCTCTAGAGTCGACCTGCAGGCNGNAGT

**Locus #11**

designed primers (5'-3'): GGAACCATCGTTCACTGGTC / GATCAGCGAGATAGAGAGGCTTA  
TTAGGGTTTTCCCAGTCACGACGTTGTA AACGACGGCCAGTGAATTCGAGCTCGGTACCCGGGGA  
TCATATTGATGGTTCCACTGGAACCATCGTTCACTGGTCTTGCTCCCGTCCTCCGAGGTCGTTCTGTTG  
CAGGATTCTGCTGAGGCCTCTCTCCCTCCGATTTTTTACAAAGTTCCTGAGATGTCCCCTTTTTATAA  
GCCTCTCTATCTCGCTGATCCTCAGAGTCGACTGCAGGNAGAGN

**Locus #13**

designed primers (5'-3'): ACCTTCTTTGAGCCCTTGC / TAACGACGTGCTGTCTCACC  
TAGGGTTTTCCCAGTCACGACGTTGTA AACGACGGCCAGTGAATTCGAGCTCGGTACCCGGGGATC  
TCTCTGACATAGGCTGGGCCACCTTCTTTGAGCCCTTGCCCCACTTTTCCCAAAAAGTGGAGTTCCAG  
TCTGTCTCTCTCCACACCTGCAACCCTAGCCGGTCCAGTCGGTCGCGATGCTGGTCCAAGAAGTCGT  
CAAAGGCTTCTTGCGGGTGGTATGGGGCCTGGTCGGACAGGGTAGGATGGTGAGACAGCACGTCCG  
TTAAAAAGACAGAGACACTCATGTTCCGCTCTCGGATCCTCAGAGTCACTGCGGCGNAGT

**Locus #16**

designed primers (5'-3'): GGTGATGCAGGACCTCAGAT / ACAGAGGGGTACCGGTTTTTC  
NCTNCAGCCTGCNNGTCTGACTCTAGAGGATCCTCTCATTACGGTGGGAAGCCCTAACACAATAAATG  
GAGGTATTTAAATATTCCCGAGAAAAGAAGTTGGTAAGGGAAGCAGTGGTGATGCAGGACCTCAGA  
TTTGCAACATCAAGAAGACTAGGGGTAGCCATTAGAGAACAGAAGACGTACTTGAAAGAAATTTTGC  
ATAAGAAAATAGAGACAAAGT[ACAGAGAGCAAAGCGAGAAGAAGAGTCTAGAGAGAAATGGAGA  
G]TTTGAATATAAAAAGAGTAAAGGCAGAGAAAGGATGTTTTGACAGAAGTTGTAACAGCCCGAAA  
ACCGGTACCCCTCTGTAACGGCCCCAACTGCTCGGCGCTAGGATCCCCGGGTACCGAGCTCGAATT  
CACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAACCCTAN

**Locus #21**

designed primers (5'-3'): ACCTCTGGGAAGTCCTCGT / CCCATTTAACTTTTCGACATGC

NGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGAGCTCGGTACCCGGGGATC  
CCATCAGAAGTCCGAAGTTAAACGTGCTTGGGCGAGAGCAGTACTAAGATGGGTGACCTCTTGGGA  
AGTCCTCGTGTTCACCCCTTTCTTTTTCTTTCTCCGCAGCGCTCCTCACCGCCGGCGGTGCCATTT  
CGGGTTTTTATTTCTCGCTCTGTATCGCTATCGATTTAGATGCATGTGCATGTGAAAAGTTAAATGG  
GAATGAAAAAAGTAAAACGAAGATCCTCAGAGTCGACTGCAGGCGCAGTN

**Locus #22**

designed primers (5'-3'): TCTCATAGCACCCTCTCGAA / CGATAGGACCCAGACAAGTCA

TTAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGAGCTCGGTACCCGGGGA  
TCACTCTCCCTATACCTGGATAAAAAGCATCTCATAGCACCCTCTCGAATATCTTTCTGGTTCACTA  
TCTCGTATTCTCATTTGGCTTTTCTATCTCTCGCACTCTCCTACCATTCTTCGTTCTCACTCTCTTC  
GTTTAGCGGGTGGTACACCTGATGACTTGTCTGGGTCCTATCGATACCTTGTCTCTGGCTGTGGGCAC  
TACAACTGTTCCCCTTCTAATCTGACTAACTACGCTGTTCTTAACACAAGCAAGAAACCTTTGATACC  
CTCTCTAAACAGCGTAAGAGCCTCTAGGGCTGATAACAACTTCTAGGCATCTCTACTTTGTCCCCTC  
ACACATTACCTGTGATCCTCAGAGTCGACCTGCGGNTGAGN

**Locus #24**

designed primers (5'-3'): GCTGATGGACCTTACCTCT / CATTTGGCTCGAGACTGTCC

TTAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGAGCTCGGTACCCGGGGA  
TCCGTGCCGGAGCTGATGGACCTTACCTCTGTATCCTGCTGATGAAGAGGCTGACCCTCTCTCTCTG  
CCTCTGCCACTGGCCTGTGTTGTGGCTGGAGCTACTGGCTGCACCACACTGCCGGAGGTTGTCTGCT  
GAGACGGTGCTGTAAAGGCTCCTCTAGGACAGTCTCGAGCCAAATGACCCGCCTGTCCGCATCTGAA  
ACATGTATTTGTCCCTGCCAGACATACTCCCTTGTGTGGTCTCCACATCTCATACATGCTGTGACCTC  
TGCGCCAGAGCTTGAGCCACTGCCTAAACCCAGNCCTGACTTGATCCTCAGAGTCACTGCAGGNGA  
GT

**Locus #26**

designed primers (5'-3'): AAGGAAGCGGCCATGGTT / GAGCACAGCGGTAATAATCG

TTAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGAGCTCGGTACCCGGGGAT  
CGACAAGGAAGCGGCCATGGTTTCCCAGACCATCCGCAACATGCTCACCTCTCCAGG[TCCCTCTCTG  
TCTCTCTC]AATTTTGTGAGGCCTGCTTTGATTGGTAGTATTATTTGACTCGTTCTATCGGGGTTCGA  
TTAGTTACCGCTGTGCTCTACTCAGTTGAGGATGATCCTCAGAGTCACTGCGGCGAGN

**Locus #27**

designed primers (5'-3'): TTGGGAACCTGCTGAAAAAG / GGTAAGCCGCACAAAGAGTT  
AGGGTTTTTTNCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGAGCTCGGTACCCGGGGAT  
CCGTGCTGAGGCTGAAGGACCTTCCCCTTGGGAACCTGCTGAAAAAGAGGTTATCCCTCTCTCTAT  
CTCGCCCTCTACCCTGTGATGGCTAAAGCTGCTGGCTGAGCCACGCTGCTAGAAGCTGTTTATTAGGA  
CTGTGCCATCCTAGCTGCATTGGGACACTCATGTGCCATGTATCCCTCTTATCCACATTTGTAACAAGC  
TGTAGTCCCAAAGTACAAAACCTTTTGTGCGGCTTACCACACCTTATACTTCTAGAGCTATCCGAACC  
AGAGCTTGAAACACTACCCATTCCTAGACCAAACCTTTAATATGTTCCAGAAGCTGTTCTTCTTAGTCTT  
CCTAGTGGATTTGTTCTACTTTTTTAACCTTCTGTGGCAGCTGCACTCAGAGAAGAGAGATCCTCTAG  
AGTCGACCNGCGGCNGNAGN

**Locus #28**

designed primers (5'-3'): TCTCTCCGCTGTCCTTCTTC / AAAAGTCGCCATTTTTGCAT  
TTTAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGAGCTCGGTACCCGGGG  
ACCAATGCAAGAGCTATAGTATTGTCGTCATTCTCTAGGCTCTGTGTTTCTATAAAACAAAATCTCTCC  
GCTGTCCTTCTTCTTTACGGCCACCTCCGGTGTTCTCTCTCTCTGTCTCCGCCTCTTTGTAGCCA  
TGAGGCGAATCAAGCAAGCTTTTTGCGCCCTAAAAGAGCACAGCTCTGTAAGCTATGCAAAAATGGC  
GACTTTTGGATGGCTTTTGCATCCTCTAGAGTCGACCTGCAGGCNGNAGN

**Locus #29**

designed primers (5'-3'): CTTCGTTTCCCACCACAACT / GTCTGGACGCGTAACCATTT  
TTTANGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGAGCTCGGTACCCGGGCA  
TCCTTCGTTTCCCACCACAACTCTGGAACATTCCAGGGGGGTGAGGTAAGTGAATAAAAACCT  
AAGCTACCAAGCCTCTGTACTACCCTTACTCTCTCTCTCTCTCTTTCTCTCTGCAGGTGCCATTCTG  
TAGGGAAAGGAAAGAAATGGTTACGCGTCCAGACACCACTCCTATTCCATACTTTGTGTAACAGCCC  
GCTTACCGGACCATCACNGGCACTAGGATCCTCTAGAGTCGACTGCAGCNGCAGT



#### **IV. Novel screening platform employing catalase-induced density separation of microcapsules**

## Introduction

The availability of protocols for rapid and cost effective discovery<sup>12</sup> or optimization<sup>10</sup> of enzymes, pathways or whole cell biocatalysts is key for development of competitive synthesis routes in industrial biotechnology.<sup>13, 138</sup>

From a technical point of view, two principal discovery strategies can be pursued: screening and selection.

Screening<sup>139</sup> refers to a process where each candidate cell, clone, or colony is generally analyzed one after the other by specific assays for the desired property.<sup>140</sup> In most cases these assays are colorimetric<sup>49</sup> or fluorometric.<sup>41</sup> The analyzable substance (i.e. the fluorescent or colored substance) is usually a protein, an antibody, or a small molecule indicative of the compound of interest. Numerous technologies are available in order to either quantitatively or qualitatively measure interesting phenotypes such as cell specific product formation rates (e.g. plate readers, flow cytometers, imagers, and colony pickers).<sup>18, 141</sup> However, all currently available technologies share the same general limitation of screening in that each sample is analyzed and isolated in a one-by-one fashion.<sup>11</sup> Depending on the applied readout and isolation technology, throughput and assay-content can thus become low while cost and logistic can become very large. On the other hand, assays based on flow cytometry are rather fast.<sup>22</sup> However, throughput, though already large, still remains limited to some 100 millions of analyzed or approx. 250'000 of sorted single cells per day and requires the equivalent of at least one operation day of the cytometer and one full time employee.<sup>60, 78</sup>

The second principal assay type is based on selection.<sup>139</sup> Selection refers to a much simpler process:<sup>8</sup> growth rate and the expression of the phenotype that is desired (typically the capability to form a specific compound at an elevated rate) are positively coupled to each other.<sup>12</sup> "Positive" cells thus grow faster

than “negative” cells and therefore readily overgrow the negative population. After or during the selection, the positive fraction can be recovered by a simple step such as plating on a solid support such as an agar plate. Selection is a basic technique in molecular biological protocols where selection-markers such as antibiotic-resistance genes are used to select transformed cells. A very large number of phenotypes can be straightforwardly selected for.<sup>13</sup> However, in selection-based approaches usually only the fastest growing fraction of the population is obtained, even though this population is not necessarily the desired fraction and promising activities might be overlooked due to e.g. suboptimal cultivation conditions or poor expression.

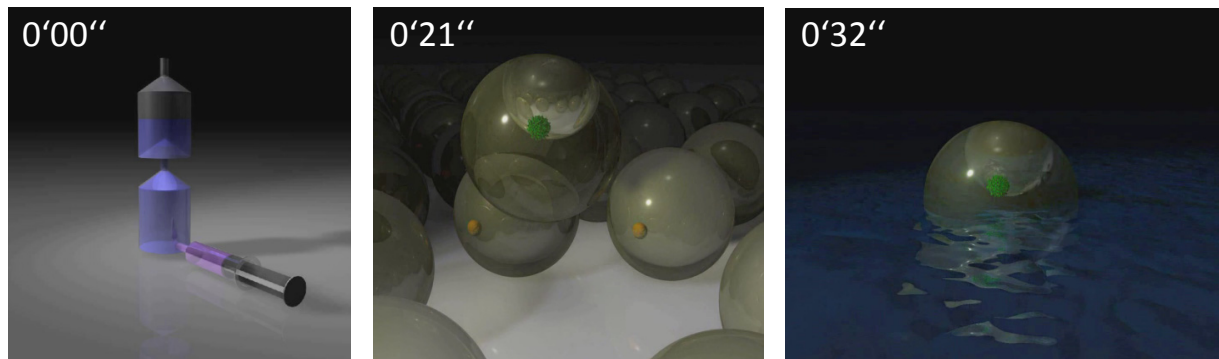
Selection-based assays can however also be performed with compartmentalized single clones or cells. Due to compartmentalization, rapidly growing clones can no longer overgrow clones that grow more slowly. A routinely applied compartmentalization technique is plating all of the cells on a solid support (e.g. a large number of agar-plates) that contains the medium used for selection.<sup>10</sup> Here, both, fast growing and slow growing cells can be detected, classified (e.g. on the basis of the approximate colony diameter), and re-isolated by automated solutions such as picking robots.<sup>63</sup> Another approach has been described by Zengler et al. 2002,<sup>142</sup> where single cells were embedded into microcapsules which were then immobilized in a bioreactor with the required medium components.<sup>26, 30</sup> With this design, both fast and slow growing fractions were obtained and separated from each other by regularly sampling and quantification of the biomass present within the microcapsules by a flow cytometer.<sup>143</sup> The approach was successfully used for isolation of organisms that would have been overlooked by classical enrichment protocols,<sup>38</sup> i.e. by plating, since large growth rate differences lead to overgrowing colonies at some time, or a combination of selection in liquid followed

by plating. However, a limitation of the system is that even though cultivation is done in a highly parallelized way, analysis and isolation is still done one after the other<sup>65, 107, 144</sup> and in case very fast analysis and sorting rates of 10 Hz are applied, the processing of 1 Mio. clones still amounts to 28 h.

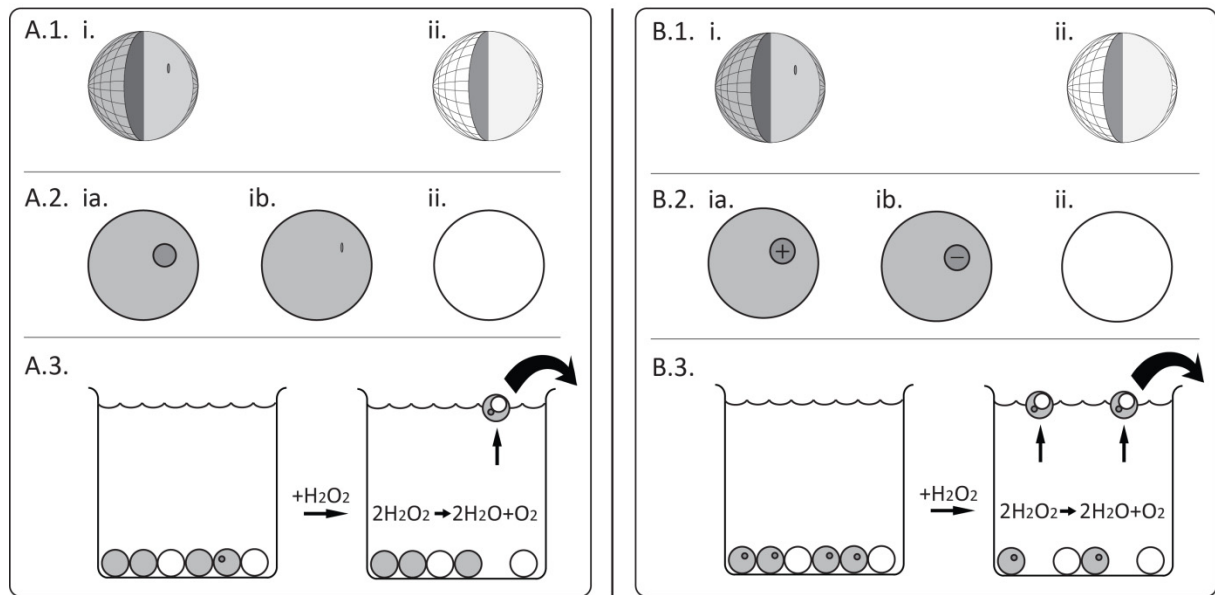
We describe here a novel technology for analyzing and isolation of recombinant library clones in a single, highly parallelized step using one and the same mechanism for both steps. Positive and negative clones are separated from each other by the aid of the specific density of their compartment in a suspension which can be influenced as a function of the desired phenotype. As a reporter protein, the catalase of *L. seeligeri*<sup>145</sup> was applied, which converts hydrogen peroxide to water and molecular oxygen.<sup>146</sup> Catalases are among the most active enzymes known today with turnover numbers of up to 1 million events per second.<sup>146</sup> If the catalase gene is expressed, massive amounts of oxygen can be released this way in microcapsules while only small amounts are liberated in capsules containing a low amount of catalase. If the system is appropriately parameterized, an oxygen-bubble is thus selectively formed within those capsules that contain a positive clone. The positive population can then be easily recovered from the suspension by flotation (see movies 2 for an illustration of the principle).

In principle, both selection and screening-based protocols could be developed with this technology. For screening, catalase induction would have to be coupled to a regulatory element<sup>147</sup> (e.g. a regulatory protein) which senses the product of interest and couples it to the production of catalase. For selection, for example in a selective medium, catalase production would simply be coupled to biomass formation. We envision a number of scenarios the novel technology could be applied for (see figure 1).

In this chapter, we describe a proof-of principle for such a technology that demonstrates that catalase-induced, density-dependent separation is feasible, it can be used to enrich positive clones, and the desired cells can be recovered despite the potential toxicity of the oxygen source, hydrogen peroxide.



*Movie 2 (the movies are on the supplementary CD as well as online on: <http://www.bsse.ethz.ch/bpl/media/mwalser>): Animation movie of an example for catalase separation of positive (green colony) and negative (brown) colonies in hydrogel microcapsules in a separation chamber. Upon hydrogen peroxide injection with a syringe into the lower chamber of the separation device, the positive colonies, synthesizing substantially higher amounts of catalase than the negative colonies, begin to form an oxygen gas bubble and will ascend into the upper chamber (middle panel). Since the hydrogen peroxide is injected into the lower chamber, the concentration of hydrogen peroxide (an antibiotic agent) is kept at a minimum in the upper chamber. As soon as all positive microcapsules have reached the upper chamber, one may close the interconnection between the two chambers and harvest the positive fraction of the upper chamber by sieving and washing of the microcapsules.*



**Figure 1: Isolation of clones via catalase reaction-induced density differences**

A) Selection: Recombinant cells expressing catalase are grown under selection pressure within microcapsules. A.1. After encapsulation some carriers contain a cell (i) while others are empty (ii). A.2. After growth under selective pressure only positive cells form a colony (ia) while negative ones do not (ib). A.3. Microcapsules containing a positive clone (colony) are separated from ib and ii by catalase reaction-induced buoyancy differences in a hydrogen peroxide suspension. B) Screening: Recombinant strains expressing a catalase gene under the control of a regulatory protein responding to the product of interest are embedded into microcapsules. B.1. Microcapsules contain cells (i) or remain empty (ii). B.2. Depending on the concentration of the product of interest within the micro-colony formed after growth, three fractions are obtained: Colonies with high amounts of catalase (ia), colonies with low amounts of catalase (ib) and empty carriers (ii). B.3. Upon incubation in a hydrogen peroxide solution only carriers containing a sufficiently high amount of catalase (i.e. the positive strains) start to float and can be readily recovered from the top.

## Materials and methods

### Chemicals

All chemicals were purchased from Fluka (Buchs, Switzerland). Molecular biology enzymes and electrocompetent cells were obtained from New England Biolabs (Ipswich, MA).

### Construction of plasmids and strains

In order to over-produce the catalase of *L. seeligeri*, the *Ptet* promoter and its cognate *tetR* repressor gene were excised from plasmid pKTS<sup>147</sup> by digestion with restriction endonucleases *Bam*HI and *Nde*I and the catalase gene was excised from vector pAHA1<sup>145</sup> by an *Nde*I and *Nsi*I digestion. The fragments were ligated simultaneously into the *Bam*HI and *Nsi*I sites of the low copy vector pAct3<sup>148</sup> (p15A). The resulting vector pAct3-Cat-TetR was used to transform *E. coli* Top-10 cells (F<sup>-</sup> *mcrA*  $\Delta$ (*mrr-hsdRMS-mcrBC*)  $\Phi$ 80*lacZ* $\Delta$ M15  $\Delta$ *lacX74* *recA1* *araD139*  $\Delta$ (*ara leu*) 7697 *galU* *galk* *rpsL* (StrR) *endA1* *nupG*; Invitrogen, Carlsbad, USA). To obtain a simple marker to differentiate between different *E. coli* strains, three pUC18 plasmids with different insert sizes in their multiple cloning site were constructed. For this, *Bam*HI-digested pUC18 was ligated to *Sau*III A1-digested DNA of phage lambda (New England Biolabs, Ipswich, MA) and the resulting mix was used to transform *E. coli* Top-10 cells and different insert lengths were determined by polymerase chain reaction (PCR). Three constructs were selected which could be distinguished easily by PCR resulting in amplicon lengths of 250 bp (pUC18-(S)), 550 bp (pUC18-(M)) and 750 bp (pUC18-(L)). The two smaller constructs were then used to transform the catalase over-expressing strain Top-10[pAct3-Cat-TetR], while plasmid pUC18-(L) was used to transform Top-10[pAct3] which does not carry a recombinant catalase gene. The resulting strains were labelled S-cat<sup>+</sup>, M-cat<sup>+</sup>,

and L-cat<sup>-</sup>. Here, it is to mention that the two housekeeping catalases of *E. coli* were not knocked-out since it was shown in previous experiments that expression of the two housekeeping catalase genes of *E. coli* (when grown in microcapsules) is neglectable.

### **Encapsulation procedure, cell proliferation, and induction**

Prior to encapsulation, cells were grown in 5 mL LB medium (1 % tryptone, 0.5 % yeast extract, 0.5 % NaCl) in 15 mL centrifugation tubes shaken at 200 rpm to an OD<sub>600</sub> of around 1. Then 3 µL of a culture were gently mixed with 3 mL of an alginate solution (2% Pronova ultra pure LVG alginate, Novamatrix, Drammen, Norway) for encapsulation with a flow focussing<sup>68</sup> Nisco encapsulator (Nisco Engineering, Zürich, Switzerland). The following conditions were applied for encapsulation: alginate flow: 0.4 mL/min, controlled by a syringe pump (TSE Systems; Bad Homburg; Germany); 350 µm nozzle diameter; air pressure: 70 mbar. When strains containing a recombinant catalase gene were encapsulated, fluorescein-labeled alginate<sup>149</sup> was used. When strains without recombinant catalase gene were encapsulated, non-labeled alginate was used. Produced microcapsules were stirred for 20 min at room temperature in 100 mL of hardening solution (50 mM BaCl<sub>2</sub> in ddH<sub>2</sub>O). Afterwards, microcapsules were recovered by sieving (40µm Falcon cell sieve; BD Biosciences; San Jose; USA), washed three times in 100 mL growth medium (0.4 % yeast extract; 0.1 % tryptone; 0.1 % glycerol, 0.1 mM BaCl<sub>2</sub>; 10 mM Tris-HCl pH 7.0; 100 mg/L ampicillin; 30 mg/L chloramphenicol) and incubated in two Petri dishes containing 25 mL of growth medium. The Petri dishes were incubated for 15 h at 30°C. Where required, anhydrotetracyclin (aTc) was added to the Petri dish to a final concentration of 50 ng/mL to induce catalase production.



### **Microcapsule characterization**

Microcapsules were characterized for microcapsule diameter, colony diameter, the degree of occupation, and fluorescence using a microscope equipped with the appropriate software (Axiovision software with AxioStar microscope; Zeiss; Göttingen; Germany).

### **Microcapsule dispensing**

Microcapsules were dispensed by a Complex Object Parametric Analyzer and Sorter (COPAS, Union Biometrica; Holliston; USA) with the following settings: Trigger channel at 545 nm; trigger channel threshold at 50 AU; PMT at 545 nm of 500 V; coincidence settings button on “pure”; sorting speed of approx. 1 Hz. The microcapsules were dispensed into 96-deep-well plates (1.2 mL square well plate, Thermo Scientific - ABgene, Epsom, UK). The wells of the plate contained 200  $\mu$ L of LB-medium supplemented with phosphate (0.58 g/L  $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$ , 1.54 g/L  $\text{Na}_2\text{HPO}_4 \cdot 7 \text{H}_2\text{O}$ , pH 7.0) in order to dissolve the microcapsule and release the cells.

### **Growth of individual clones in deep-well plates**

The deep-well plates were closed by a gas permeable adhesive seal (Thermo Scientific - ABgene, Epsom, UK) and incubated at 25°C for 3 h at 200 rpm in order to dissolve the microcapsules and release the cells. Afterwards the plate was incubated on a shaker for another 12 h at 37°C at 200 rpm. As a control, the two catalase-positive strains were inoculated (1/50 from a liquid culture) and grown to equal optical densities in two wells of the plate. After incubation, cultures were scored as “proliferated” or “non-proliferated” depending on whether their  $\text{OD}_{600}$  had advanced beyond a value of 0.5.

### **Analysis of cross-contamination by PCR**

To test cross-contaminations, culture-PCR was applied using the following conditions in multi-well plates: 1x NEB-PCR-Buffer, 200 $\mu$ M dNTPs, 100nM forward primer M13F(-111)(5'-CGGGCCTCTTCGCTATT-3'), 100nM reverse primer M13R(+86)(5'-CGGCTCGTATGTTGTGTG-3') (Microsynth; Balgach; Switzerland); 1.25U Taq polymerase; 1  $\mu$ L cell suspension; 25  $\mu$ L reaction volume. The plates were closed with 8 strip-caps and thermocycling was performed (Eppendorf MasterCycler, Hamburg; Germany) under the following conditions: Initial denaturation of 2 min at 95°C followed by 30 cycles of: denaturation at 95°C for 30 s, primer annealing at 55°C for 30 s, elongation at 72°C for 1 min; final elongation at 72°C for 1 min. The fragments were analyzed on a 2% agarose gel.

## Results

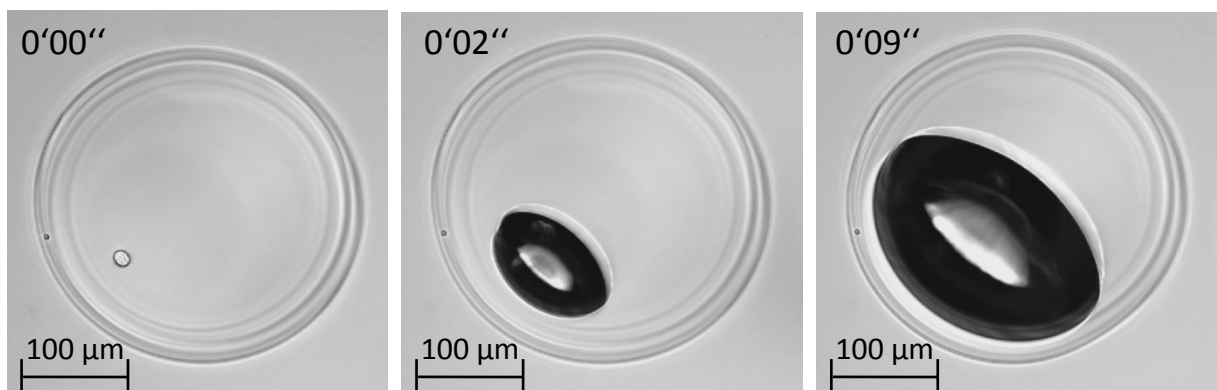
### **Catalase-induced bubble formation is sufficient to separate colonized microdroplets by density difference.**

In order to test whether catalase production-induced density differences could serve as a discriminating property in microcapsules and could additionally be regulated with aTc, we cloned the catalase gene under the control of the *tet* regulatory system on a low-copy plasmid vector into a suitable *E. coli* strain. The resulting cells were encapsulated into alginate microcapsules and allowed to expand into colonies in Petri dishes in rich medium with and without added inducer. Since microcapsules were occupied by relatively low cell numbers, the microcapsules contained in most cases a single colony. When hydrogen peroxide was added up to a final concentration of 3 % to Petri dishes containing microcapsules with uninduced colonies, no bubble formation was observed. However, microcapsules with colonies that had been induced with aTc readily displayed a population of floating microdroplets when hydrogen peroxide was provided to a final concentration of 1 %. Microscopic analysis could easily show that the behaviour correlated with the formation of a bubble inside the colonized microcapsules (see movie 3). This suggested that it was indeed possible to use density difference in microdroplets as a potentially very efficient tool to separate microdroplets according to specific properties of the colonies contained inside them.

### **Analysis of the suitability of density difference-induced separation for high-throughput screening**

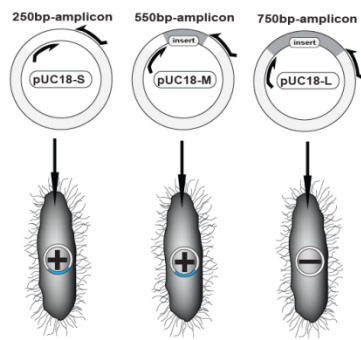
In order to convert this observation into a more general tool, one main point to clarify is whether it is possible to isolate specific populated microcapsules by density difference and retrieve living cells from them (after the hydrogen

peroxide treatment), while ensuring conditions throughout the entire process to allow that the cells that are finally obtained stem indeed from the colony that led to the density difference in the first place. In other words, there should be a one-to-one relationship between the cells in the colony that display the desired phenotype and the cells that are finally retrieved from the microcapsules after separation. This led us to carry out a proof-of-principle experiment, for which an overview is given in figure 2.



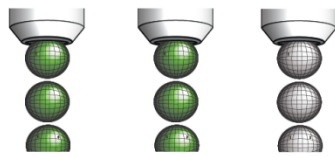
**Movie 3** (the movies are on the supplementary CD as well as online on: <http://www.bsse.ethz.ch/bpl/media/mwalser>): Microscopic image of an agarose-embedded alginate microcapsule containing a colony with cells over-producing catalase. As soon as hydrogen peroxide is added to the sample, an oxygen bubble is formed within the microcapsule immediately. The movie is kept in real time illustrating that bubble formation takes only about a few seconds.

**Figure 2: Flow chart for microcapsule selection by catalase**



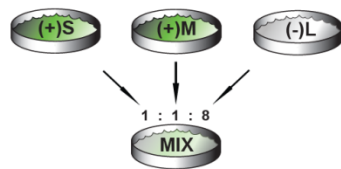
**1. Transformation of test strains**

Recombinant *E. coli* carrying pUC18 plasmids with differently sized inserts. Strains S-Cat<sup>+</sup> and M-Cat<sup>+</sup> also expressed a catalase gene while strain L-cat<sup>-</sup> did not.



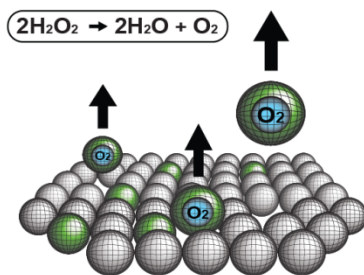
**2. Cell encapsulation**

Two catalase positive strains (S-cat<sup>+</sup> and M-cat<sup>+</sup>) were encapsulated into fluorescein-labeled alginate microcapsules and the negative strain (L-cat<sup>-</sup>) into non-labeled microcapsules.



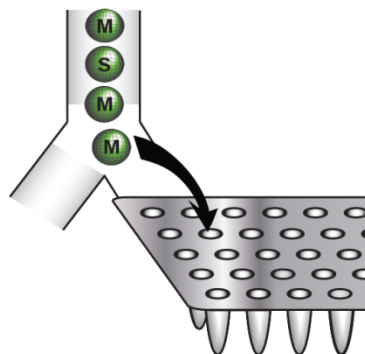
**3. Batch-wise cell proliferation**

All three microcapsule batches were incubated separately (15h incubation, 30°C) and afterwards mixed at a ratio of 1:1:8 (S-Cat<sup>+</sup> : M-Cat<sup>+</sup> : L-Cat<sup>-</sup>)



**4. Separation of catalase positive strains**

Catalase positive strains (S-cat<sup>+</sup> and M-cat<sup>+</sup>) were separated from catalase negative strains (L-cat<sup>-</sup>) and the empty microcapsules by density difference after the addition of hydrogen peroxide. The entire upper fraction was harvested by pipette and washed immediately in medium.



**5. Dispensing the catalase-positive fraction by COPAS**

After brief washing, 80 microcapsules of the floating fraction were dispensed one-by-one into separate wells of a 96-deep-well-plate. The microcapsules were disintegrated and the released cells were proliferated.



**6. Test for cross-contamination by PCR**

The genotype of the proliferated clones was determined by PCR using the size of the insert as a marker.

### **Construction of a suitable control system**

First of all, three different *E. coli* strains were constructed that could be easily differentiated. The first two contained the already mentioned low-copy vector with the catalase gene from *L. seeligeri*, while the third strain received the empty low-copy vector. Next, three fragments with different sizes from phage lambda-DNA were cloned into the high copy number plasmid pUC18. These three different plasmids were used to construct the three *E. coli* strains S-cat<sup>+</sup>, M-cat<sup>+</sup>, and L-cat<sup>-</sup>, so that each of the three strains could be clearly identified by the length of a PCR fragment obtained from the present plasmid.

The strains with the recombinant catalase gene were then encapsulated into fluorescently labelled alginate and the strain with the empty low-copy vector was encapsulated into non-labelled alginate microcapsules. The three different microcapsule populations were incubated separately (to control that colonies of all batches grow to equal sizes) in Petri dishes with growth medium and inducer. At the end of the incubation phase, microcapsules with the three different strains obtained from the three different Petri dishes were microscopically analyzed. The results are summarized in table 1 and indicate an average microcapsule volume of 930 pL and an average colony content of approx. 1'000 cells (assuming that an *E. coli* cell occupies a volume of 2 fL).

### **Catalase-induced separation by density difference**

Next, the different fractions were united by adding two times 100 µL of microcapsules derived from the strains with recombinant catalase gene and 800 µL of microcapsules derived from the strain without recombinant catalase. Assuming that the microcapsules were closely packed (spheres occupying 74% of an available volume), the microcapsule suspension obtained after mixing the

3 aliquots contained approximately  $7.5 \times 10^5$  microcapsules suspended in growth medium. From these,  $5 \times 10^5$  (675  $\mu\text{L}$ ) microcapsules were transferred to a small Petri dish (40 mm in diameter) filled with 6 mL of growth medium. Microscopic pictures were taken and the ratio of fluorescein-labeled and non-labeled microcapsules as well as the percentage of colonized microcarriers was determined once more. These results are summarized in the top half of table 2. These data were in good agreement with the values expected according to the data of the source populations (table 1).

Subsequently, oxygen-bubble formation was initiated by addition of a 30% solution of hydrogen peroxide to a final concentration of 1%. Within 20 s, a substantial fraction of fluorescent microcapsules floated on top of the suspension. The floating fraction was sucked off with a pipette, transferred to a sieve, and washed with growth medium to remove remaining hydrogen peroxide and cells that may have escaped from the microcapsule. The entire separation took about 30 s. Finally, representative samples were taken from the microcapsules in the top fraction and the bottom phase and microscopically analyzed. These results are summarized in the bottom half of table 2. They suggest that before the separation, 9.2 % of the fluorescent microcapsules and 21.5 % of the non-labelled microcapsules were occupied, amounting to a total of roughly 9'000 microcapsules with a colony over-expressing the catalase of *L. seeligeri* and approx. 86'000 with a colony that did not. The remaining fraction, slightly more than 400'000 microcapsules, were unoccupied (see figure 2A). After the separation, in an aliquot of 97 capsules derived from the upper fraction, all microcapsules were fluorescent (figure 2C and 2D). In an aliquot of 677 microcapsules from the lower fraction, 16.8 % were fluorescent and 83.2 % were non-labeled. Of the fluorescent microcapsules, only 3 were colonized, which amounts to a percentage of 2.6 %,

a factor of 3.5 lower than the 9.2 % prior to separation. On the other hand, the percentage of colonized microcapsules in the non-labeled microdroplet fraction remained rather constant (19.4 % after and 21.5 % prior to the sorting step, figure 2B).

### **Cell proliferation after the separation step**

After separation of the upper and the lower fraction by catalase, 80 microcapsules of the upper fraction were dispensed into a 96-deep-well plate containing a rich growth medium and additional phosphate to dissolve the microcapsule and release the cells. After incubation over night, 59 (74%) of the wells showed proliferation.

### **Effect of the procedure on culture purity**

In order to determine the state of purity of the cultures we had obtained after overnight incubation, the cultures were subjected to PCR. As each of the three strains had a differently sized insert in the recombinant pUC vector, which could be amplified with the same primer pairs, the three strains could be easily told apart and the cultures could be analyzed by performing one PCR on each culture and analyzing the obtained pattern of amplicons. For this, we had to first determine the sensitivity of the PCR assay: Both catalase positive strains (containing plasmids with two different amplicons, 250 bp and 550 bp) were first grown separately and then two dilution series (1:1; 1:2; 1:4; 1:8; 1:16; 1:32; 1:64, 1:128, starting from cultures at the same OD) of one strain in the other and *vice versa* were made. As shown in figure 3, proof of the presence of one strain in a culture of the other was reliably possible above a ratio of 1:32 of the two strains. Therefore, the assay allowed reliable detection of a contamination if at least 3% of the cells came from a different strain.



When PCR was performed with the supernatants from all 80 occupied wells, no PCR product could be obtained from those samples where the strains had failed to proliferate. From all 59 samples that had proliferated after overnight incubation, a PCR product was obtained (see figure 3). In 47 cases a 550 bp amplicon (indicative for strain M-cat<sup>+</sup>) and in 12 cases 250 bp amplicon (indicative for strain S-cat<sup>+</sup>) was obtained. In no case could we obtain the two PCR product of S-cat<sup>+</sup> and M-cat<sup>+</sup> in one reaction. As the longest amplicon (indicative for the catalase negative strain L-cat<sup>-</sup>) with a size of 750 bp is at the same size as the PCR by-product of M-cat<sup>+</sup> (see faint band in positive control of dilution series in figure 4), we can not entirely exclude it's generation in reactions where we detected the 550 bp fragment. But since the 750 bp fragment was not detected in the reactions with the 250 bp fragment, we have strong evidence that contaminations of the catalase negative strain is not a major issue. Within the limits of the PCR-based assay, all re-cultured strains can thus be considered as pure cultures.

**Table 1: Characterization of individual batches prior to catalase-induced separation**

	L-cat <sup>-</sup>	M-cat <sup>+</sup>	S-cat <sup>+</sup>
Amplicon size [bp]	750	550	250
Overexpression of heterologous catalase	No	Yes	Yes
Produced microcapsules [-]	3'000'000	3'000'000	3'000'000
Alginate labeling	none	fluorescein	fluorescein
Proliferation time [h]	15h	15h	15h
Average microcapsule diameter [ $\mu\text{m}$ ] <sup>1</sup>	121±23	121±23	121±23
Average colony diameter [ $\mu\text{m}$ ] <sup>1</sup>	14±3	14±3	14±3
Average cell number per colony [-]	1'000	1'000	1'000
Total counted microcapsules for [-] <sup>1</sup>	671	578	582
Colonized microcapsules [-] <sup>1</sup>	145	79	29
Colonized microcapsules [% of fraction] <sup>2</sup>	21.6	13.7	5.0
Microcapsules mixed for catalase separation process [-]	400'000	50'000	50'000

<sup>1</sup>Determined by microscopy (see material and methods for details)

<sup>2</sup>Percentage of colonized microcapsules in each fraction.

**Table 2: Characterization of mixed populations before and after density-induced fluorescein-labeled non-labeled**

	fluorescein-labeled	non-labeled
<b>Before catalase separation Stepp</b>		
Total microcapsules [-] <sup>1</sup>	100'000	400'000
Total counted microcapsules per fraction[-]	119	594
Colonized microcapsules [-] <sup>2</sup>	11	128
Percentage of colonized microcapsules per fraction [%]	9.2	21.5
Total colonized microcapsules [-] <sup>4</sup>	9'200	86'000
Total empty carriers [-] <sup>4</sup>	90'800	314'000
Colonized microcapsules [% of all microcapsules] <sup>5</sup>	1.8	17.2
<b>After catalase separation step - bottom fraction</b>		
Total counted microcapsules per fraction[-]	114	563
Colonized microcapsules [-] <sup>2</sup>	3	109
Percentage of colonized microcapsules per fraction [%] <sup>3</sup>	2.6	19.4
Removed colonized carriers [%] <sup>6</sup>	71.5	10.1
<b>After catalase separation step - top fraction</b>		
Total counted microcapsules per fraction[-]	97	0
Colonized microcapsule [-] <sup>2</sup>	97	N/A
Percentage of colonized microcapsules [%]	100 %	N/A

<sup>1</sup>Number of mixed microcapsules as estimated based on data from table 1

<sup>2</sup>Determined by microscopy (see material and methods).

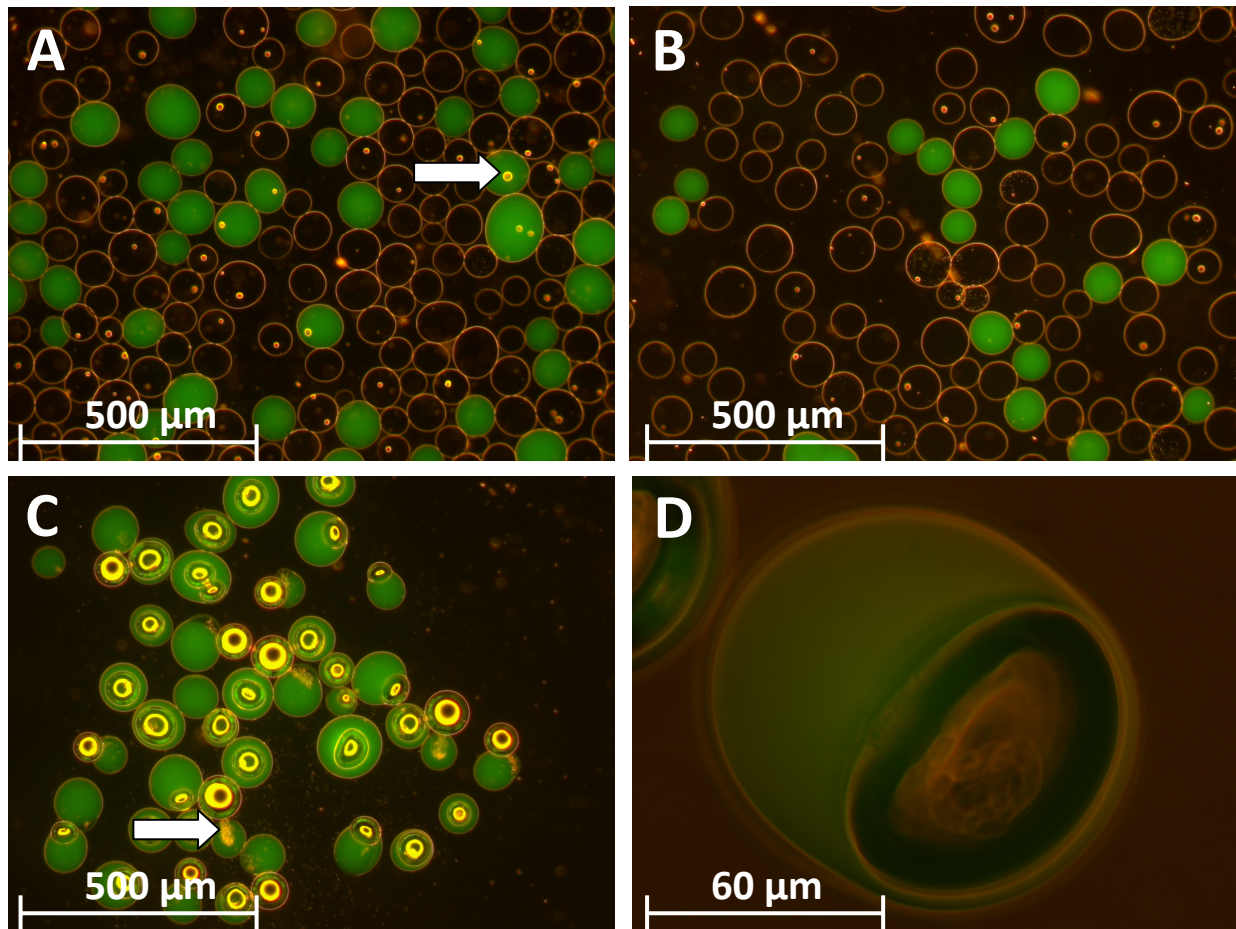
<sup>3</sup>Calculated as the percentage of colonized microcapsules in the total amount of microcapsules in each fraction.

<sup>4</sup>Calculated on the basis of the ratio of colonized to empty microcapsules and the number of total microcapsules.

<sup>5</sup>Calculated on the bases of the amount colonized microcapsules in each fraction divided by all microcapsules.

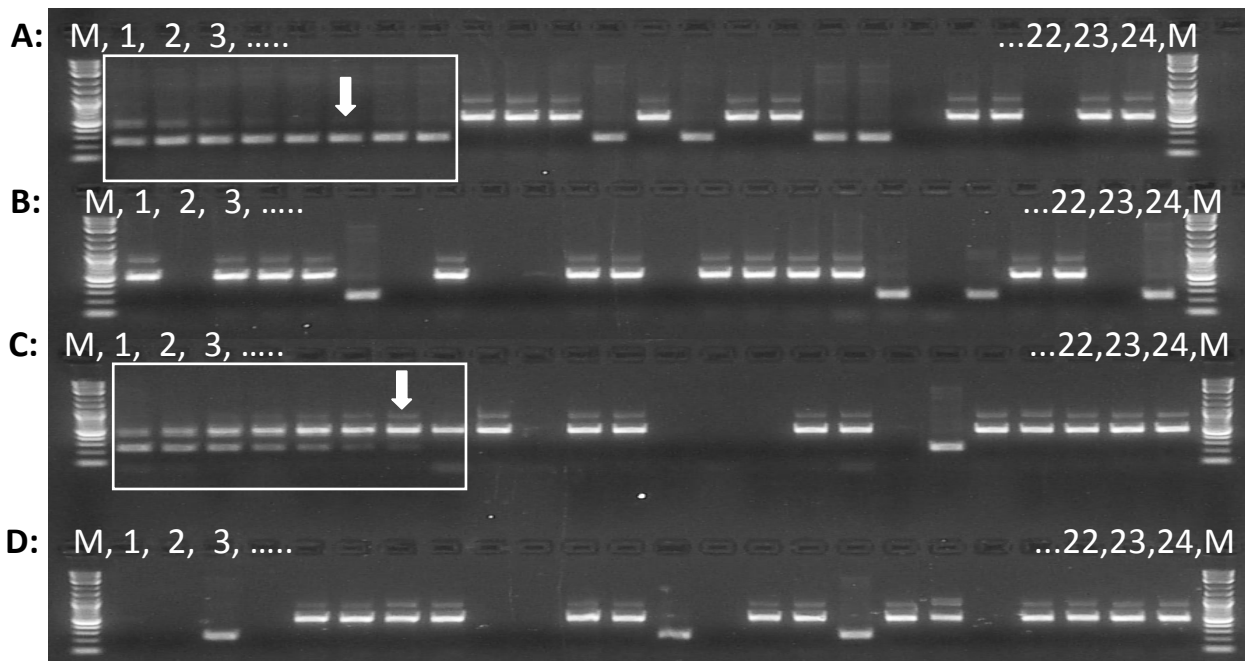
<sup>6</sup>Calculated via the different numbers of colonized microcapsules before and after sorting.

N/A: not applicable



**Figure 3: Separation of microcapsules by catalase-reaction induced density differences**

A) Mixed population of fluorescently-labelled microcapsules containing statistically distributed catalase-positive strains  $S\text{-cat}^+$  and  $M\text{-cat}^+$  (green circles) and transparent microcapsules containing the statistically distributed catalase-negative strain  $L\text{-cat}^-$  (orange circles with black interior). One colony inside a microdroplet is marked by an arrow. B) Bottom fraction after separation. C) Top fraction after separation. Some of the carriers still carry an oxygen bubble inside. Arrow: disrupted microcapsule. D) Magnified single microcapsule after the catalase separation step carrying an oxygen bubble and a colony, visible (off-focus) behind the oxygen bubble.



M: DNA-marker: fragment lengths [bp] from bottom to top: 100, 200, 300, 400, **500**, 600, 700, 800, 900, **1'000**, 1'200, 1'500, 2'000, 2'500, **3'000**, 4'000, 5'000, 6'000, 8'000, 10'000.

**Figure 4: PCR analysis of culture supernatants**

Panel A, lane 1-8: Dilution of stain M-cat<sup>+</sup> in strain S-cat<sup>+</sup> (lane 1: 1:1, lane 8: 1:128). Panel C, lane 1-8: Dilution of stain S-cat<sup>+</sup> in strain M-cat<sup>+</sup> (lane 1: 1:1, lane 8: 1:128). The arrows mark the limit of detection for the diluted strain. Remaining lanes: PCR results from culture supernatants after microcapsule dissolution and incubation of released cells. Three types of results can be observed: [i] no PCR product (21 cases, coinciding exactly with the cases of non-proliferation); [ii] detection of a small PCR fragment of 250 bp, no indication of a 550 bp fragment (12 cases); [iii] detection of a medium PCR fragment of 550 bp, no indication of a 250 bp fragment (47 cases).

## Discussion and Conclusion

Catalase reaction-induced density differences were used – to the best knowledge of the author – for the first time for colony-sorting and analysis of recombinant clones in a single step. The results presented here demonstrate successful enrichment of microcapsules harbouring catalase-producing strains against a background of roughly 500'000 microcapsules that were empty or contained colonies without recombinant catalase over-production. The fraction of catalase over-producing colonies in the bottom phase was efficiently reduced. The top phase consisted, within the scope of the microscopic analysis, only of such microcapsules which had contained catalase over-producing colonies, at the time point of separation. This also indicated that without catalase over-production, the oxygen-liberation capacity of *E. coli* is too low to interfere with the enrichment scheme. This hypothesis is also supported by a large number of microscopic observations made by us in the course of the protocol development. In no case, bubble formation could be observed in microcapsules that harboured an *E. coli* isolate synthesizing only the house-keeping catalase.

The cells in the colonies in the microcapsules isolated from the top-fraction could be recovered alive in most cases. Given the fact that hydrogen peroxide is a widely used antimicrobial agent, this is somewhat surprising. Previously we had observed that the survival rate of *E. coli* (without recombinant catalase production) incubated for 10 min in LB medium in the presence of 1 % hydrogen peroxide is approx. 10 % (data not shown). On the other hand, the cells in the microcapsules are organized as a colony (figure 2D). This might help to insulate the inner cell layers of the colony against the effects of the hydrogen peroxide. The high ratio of proliferating against non-proliferating colonies after microcapsule dissolution suggests that this protection is rather

efficient, but not perfect. However, we suspect that the ratio is actually higher: Clearly, some of the microcapsules recovered from the top-phase were disrupted in the course of the rapid formation of the oxygen bubble (figure 2C). We further observed that in the course of the disruption, larger numbers of cells escaped from the respective microcapsules (figure 2C). This phenomenon may have contributed to the failure to have proliferation in 100 % of the cultures started from top-fraction microcapsules.

Clearly, escape of cells from microcapsules is a potential source of contamination of the cultures recovered after microcapsule dissolution. However, based on the results of the PCR analysis all recovered cultures were pure at least down to the sensitivity of the PCR assay. A substantial contamination should have led at least in some of the wells to a mixed PCR-signal – in particular given the fact that of the approximately 1000 cells in the colony before separation an (unknown) percentage would have died, so that the initial ratio of contaminating cells to cells recovered from the microcapsule could be expected to be in the area of the sensitivity of the assay.

We suggest that the approach described in this chapter is generally a potent alternative to the currently applied selection and / or screening protocols (e. g. fluorescent proteins, blue-white screenings, *lux*-genes, antibiotic resistance genes, auxotrophy complementation selections, etc.). However, two features of the system seem to be unique when compared to other selection markers or reporter proteins: First, the indicative signal allowing the identification of the positive fraction is not only an indicator but also the underlying physical principle used for separation of a positive colony without the aid of any supplementary device (e. g. FACS). Second, colonies are separated in a single, highly parallelized step.

Due to these two features of the system, there is no need for identification and isolation to take place one after the other. The throughput of the system is thus limited by size of the batch only but not by the time frame between analysis and sorting. In this experiment microcapsules of 1 nL were used. Based on the results reported in here, we submit that upon further scaling to a total volume of 1 L sorting analysis of up to 1 billion samples should be feasible in a single run of a few minutes.

### **Acknowledgements**

We are grateful for the support with electro-competent *E. coli* NEB-10-beta cells by Thomas Causey (New England Biolabs). Plasmids pAHA1 and pAct3 were generously provided by Albert Haas and Tom Linn. We are indebted to Sonja Billerbeck for her valuable suggestions. This work was supported by grants from the CTI and the Swiss National Fund (SNF).

## **V. Dye-end-labeled DNA synthesis in microcapsules**



## 1. Introduction

One of the most sample-intensive applications in lifesciences is shotgun-Sanger-sequencing<sup>150</sup> employing capillary electrophoresis sequencers (CE) as a readout.<sup>151, 152</sup> On the one hand, the sequencing of an entire genome simply requires a large number of samples to produce a complete set of data. On the other hand, preparation of the Sanger sequencing fragments is a multistep procedure.<sup>39</sup> An abbreviated version of an exemplary protocol looks as follows: 1) Size-standardized DNA is ligated into vectors. 2) Vectors are used to transform competent cells. 3) The library is plated and non-transformed cells are eliminated by antibiotic selection. 4) Cells are picked (while discriminating cells containing a self-ligated vector for example in blue-white selections) and grown in liquid media. 5) Cells are lysed and cell proteins and RNAs are removed in a sequence of precipitations and resuspensions. 6) The DNA is obtained in purified form from an ethanol-precipitation. 7) Sanger sequencing fragments are obtained by thermal cycling with heat-stable DNA polymerase and one sequencing primer. From the fourth step on, all manipulations are done in microtiter (MT)-plates as spatial arrays.<sup>153</sup>

Such multistep protocols are rather laborious and costly. Substantial amounts of consumables (tips, MT-plates, etc) are needed and a considerable degree of automation is required in order to repetitively perform all these steps on a larger scale. In the Joint Genome Institute for instance, more than 1.2 Billion bases, synthesised by Sanger sequencing, were called solely in the first quarter of 2010.<sup>154</sup> This enormous number of analyzed bases corresponds to a total number of at least 2 million analyzed samples per month. translating into at least 14 million single manipulations per month in order to supply the samples.<sup>155</sup>

We developed a protocol for rapid PCR-screening of bacterial colonies in nL-sized barium-alginate microcapsules (see chapter III). The protocol has been specifically tailored to the identification of “rare events” using conserved primers in *E. coli* libraries of up to 20'000 cells and has the potential for further scaling. All compartmentalized steps typically required for colony screenings, i.e. growth of single clones followed by PCR and analysis, could be performed while always using one and the same microcarrier for one colony. Due to their very small size and the possibility to treat all microcarries in one batch at the same time in the course of colony-growth and subsequent manipulations (such as PCR), rather fast and inexpensive protocols can be developed. Another advantage of the microcapsule approach stems from the fact that virtually no sophisticated secondary reaction arrays (e.g. microtiter plates) are required but all manipulations are done in rather simple reaction containers (e.g. beakers of plastic-bags). We therefore explored the development of a protocol for preparation of Sanger fragments based on microcapsules. A first effort towards the application of microcapsules as reaction compartments for growth of *E. coli*, amplification of heterologous plasmid DNA, followed by synthesis of dye-terminated Sanger fragments is reported in here.

## **2. Materials and Methods**

### **Chemicals**

All chemicals, media, and media components were obtained from Fluka (Buchs, Switzerland) unless stated differently.

### **Determination of the detection limit and influence of barium chloride in an ABI3100 sequencer**

The detection limit, in terms of minimal analyzable amount of Sanger fragments injected into an ABI3100 sequencer (Applied Biosystems; Foster City; USA; injection conditions: run temperature of 60°C, run voltage of 15 kV; injection voltage 1 kV; injection time of 15 sec; 50 cm array length) was experimentally determined by analyzing a dilution-series (concentrations were always reduced by half down to 1:128) of Sanger fragments obtained by a standard reaction performed according to the BigDye V3.1 manual (Applied Biosystems; Foster City; USA). The concentration of the undiluted dye terminated fragments was determined by a spectrophotometer at 260 nm (Perkin Elmer Lambda 25; Waltham, MA). The CE-detection limit was then determined as the concentration at which no base could be called at a  $QV > 20$  by the SeqAnalysis software (KB basecaller algorithm; Applied Biosystems; Foster City; USA), where  $QV > 20$  is a value for the number of bases called correctly with a probability of  $> 99.0\%$ . The arbitrary units (AU) at this concentration were taken from the CE raw data and tentatively defined as the lower limit of signal intensities that could be still deconvoluted by the software. The sensitivity of the CE measurements to barium chloride was determined with the same undiluted positive sample but at increasing amounts of barium chloride (0 mM, 0.1, 0.2, 0.4, and 0.8 mM) to samples prior to injection to CE analysis.

### **Pre-culturing of microorganism used for encapsulation**

All *E. coli* isolates used in here were derivatives of Top10 (F<sup>-</sup> *mcrA*  $\Delta$ (*mrr-hsdRMS-mcrBC*)  $\Phi$ 80*lacZ* $\Delta$ *M15*  $\Delta$ *lacX74* *recA1* *araD139*  $\Delta$ (*ara leu*) 7697 *galU galK rpsL* (Str<sup>R</sup>) *endA1 nupG*; Invitrogen, Carlsbad, USA). The strain harbored a pACYC184 plasmid containing a gene for GFP-mut2 for cell-labeling (see also chapter II) and a pUC18 derivative containing inserts as templates for Sanger sequencing reactions. The insert DNA was obtained from phage lambda digested by *Sau*III A1 and inserted into the multiple cloning site (*Bam*HI) of the pUC18 vector. Prior to encapsulation, cells were generally grown in 15 mL centrifugation tubes shaken at 200 rpm to an OD<sub>600</sub> of ~1. Cells were then diluted in LB to an OD<sub>600</sub> of 0.1 and 40  $\mu$ l were suspended in 8.960 mL salt solution (0.9 % w/v NaCl; 0.1 % w/v KCl; 15 mL Falcon tube), supplemented by 1 mL graphite solution (1 % w/v graphite, 2 % v/v ethanol, 0.1 % v/v Triton X-100) in order to increase microcapsule opaqueness, and 10 mL of a 3 % alginate solution (Pronova ultra pure LVG; Novamatrix; Drammen; Norway) were added. The suspension was homogenized by gentle reversion of the tube for a few times and then immediately transferred into two 10 mL syringes (Terumo Corporation; Tokio; Japan). Microcapsules were then formed by laminar jet break-up technology<sup>67</sup> (Nisco Encapsulator, Nisco Engineering AG; Zürich; Switzerland; nozzle diameter 150 $\mu$ m) at 1050Hz and a laminar alginate solution flow of 3.8 mL / min (syringe pump from TSE Systems; Bad Homburg; Germany). The resulting droplets of 60 nL were cross-linked at room temperature for 30 min in 100 mL hardening solution (50 mM BaCl<sub>2</sub>), during which the beads shrank to a volume of approx 35 nL (diameter of 400  $\mu$ m). Afterwards, microcapsules were recovered by decanting (100  $\mu$ m Falcon cell sieve; BD Biosciences; San Jose; USA), briefly washed (100 mL washing solution;

0.1 mM BaCl<sub>2</sub>, 10 mM Tris-HCl pH 7.0; Fluka Chemie; Buchs; Switzerland) and recovered in a sieve again. The microcapsules in the sieve were then distributed over the required number of Petri dishes with a spatula (2.5 g of wet microcapsules per dish). The dishes were filled with 25 mL of growth medium (0.4 % w/v yeast extract; 0.1 % w/v tryptone; 0.1 % v/v glycerol, 0.1 mM BaCl<sub>2</sub>; 10 mM Tris-HCl pH 7.0, all in water) and incubated at 30°C. The incubation time was chosen according to the experimental boundary conditions and grown batches were stored on 4°C for a maximum of 24 h until further use. The microcapsules were then recovered separately from each of the dishes by sieving, washed in 50 mL washing solution and analyzed by COPAS. Colony sizes in one population of microcapsules were determined by fluorescence microscopy (Axiovision; Zeiss; Göttingen; Germany) by analyzing 10 occupied colonies.

### **PCR in microcapsules**

Generally, 20 microcapsules were dispensed into one MT-plate well by a COPAS flow cytometer (Union Biometrica; Holliston; USA). The COPAS profiler software was used to characterize the embedded colonies in terms of size, optical density, and fluorescent signal emitted from them (for details see chapter II) according to the experimental requirements (see below).

After removal of the excess sorter liquid (ddH<sub>2</sub>O, remaining with the microcapsules from the sorting and dispensing process) by pipetting, 24 µL PCR reaction mix was added to each of the wells. In short: 1x PCR buffer (Fermentas, Burlington; Canada), 1.5mM MgCl<sub>2</sub> (Fermentas; Burlington; Canada), 100µM BaCl<sub>2</sub>, 200µM dNTPs (Fermentas; Burlington; Canada), 100nM fluorescently labeled forward primer M13F(-40)-FAM (5'-GTTTCCAGTCACGAC-3') and 100nM reverse primer M13R (5'-

CAGGAAACAGCTATGAC-3') (Microsynth; Balgach; Switzerland), and 1.6U *Taq* polymerase (GenScript; Piscataway; USA). The plates were closed with 8 strip-caps and thermocycling was started (Eppendorf MasterCycler, Hamburg; Germany) under the following reactions conditions: Initial denaturation 2 min at 95°C followed by 31 cycles of: denaturation 95°C for 30 sec, primer annealing at 55°C for 30 sec, elongation at 72°C for 1 min; final elongation at 72°C for 1min. The remaining PCR-mix was removed by pipetting, the microcapsules were resuspended in ddH<sub>2</sub>O and recovered by sieving (100 µm Falcon cell sieve; BD Biosciences; San Jose; USA). The microcapsules were briefly rinsed with 50 mL of ddH<sub>2</sub>O, transferred into the COPAS sample-cup and dispensed in a 10 µL water droplet individually into MT-plate wells as described in chapter II. Numbers of sorted microcapsules per well are described in the individual experiment sections below.

### **Determination of the influence of the colony size on PCR efficiency**

The influence of the colony-size on PCR efficiency was investigated at a fragment size of 440 bp. Cell cultivation and encapsulation conditions were the same as described above. A total number of 100'000 microcapsules inoculated by *E. coli* (degree of occupation of 75 %) were produced, split into 10 Petri dishes (10'000 microcapsules each) and incubated at 30°C for up to 24 h in order to accumulate an increasing amount of biomass within the microcapsules. PCR conditions were analogous to those described above with the following changes: Since the colonies with cells grown for less than 10 h were hardly detectable by COPAS (see chapter II), no enrichment for monoclonal beads was performed and generally 20 microcapsules were dispensed into one MT-plate well independently of whether or not they harbored a colony. For each growth time, 16 wells were filled resulting in a

total of 320 PCR-cycled microcapsules per growth time. After PCR microcapsules were washed as described above and 40 individual microcapsules per growth time point were added individually to single wells of an MT-plate. The PCR-product was recovered by diffusion and analyzed by an ABI 3100 sequencer (Applied Biosystems; Foster City; USA) directly from the same well. PCR fragments were then quantified using peak height (PH) as a measurant (SeqAnalysis software; Applied Biosystems; Foster City; USA). Afterwards, the percentage of PCR-positive probes was determined using the experimentally specified degree of occupation (75%, determined after growth for 24 h; see chapter II) as a reference.

This protocol was slightly modified when the influence of colony growth on the production of Sanger sequencing fragments was investigated: The investigated *E. coli* strain carried only one plasmid, a 390 bp amplicon obtained from a *SauIII*A1 digest of genomic DNA of *Streptococcus epidermis* randomly ligated into pGEM7z(+) (Invitrogen, Carlsbad, USA). Next, monoclonal microcapsules were sorted out by COPAS prior to dispensation into MT-plates. In this way, 8 wells per time point were filled with 20 microcapsules each resulting in a grand total of 160 occupied microcarriers per time point (see figure 1). The degree of occupation was determined by COPAS as described before (chapter II) after 22 h of growth and was 78%.

### **Production of Sanger sequencing-fragments in microcapsules**

The influence of the colony growth time (and initial colony size) on the success of producing sequencing fragments after the PCR step was also investigated (for an experimental workflow see also figure 1).

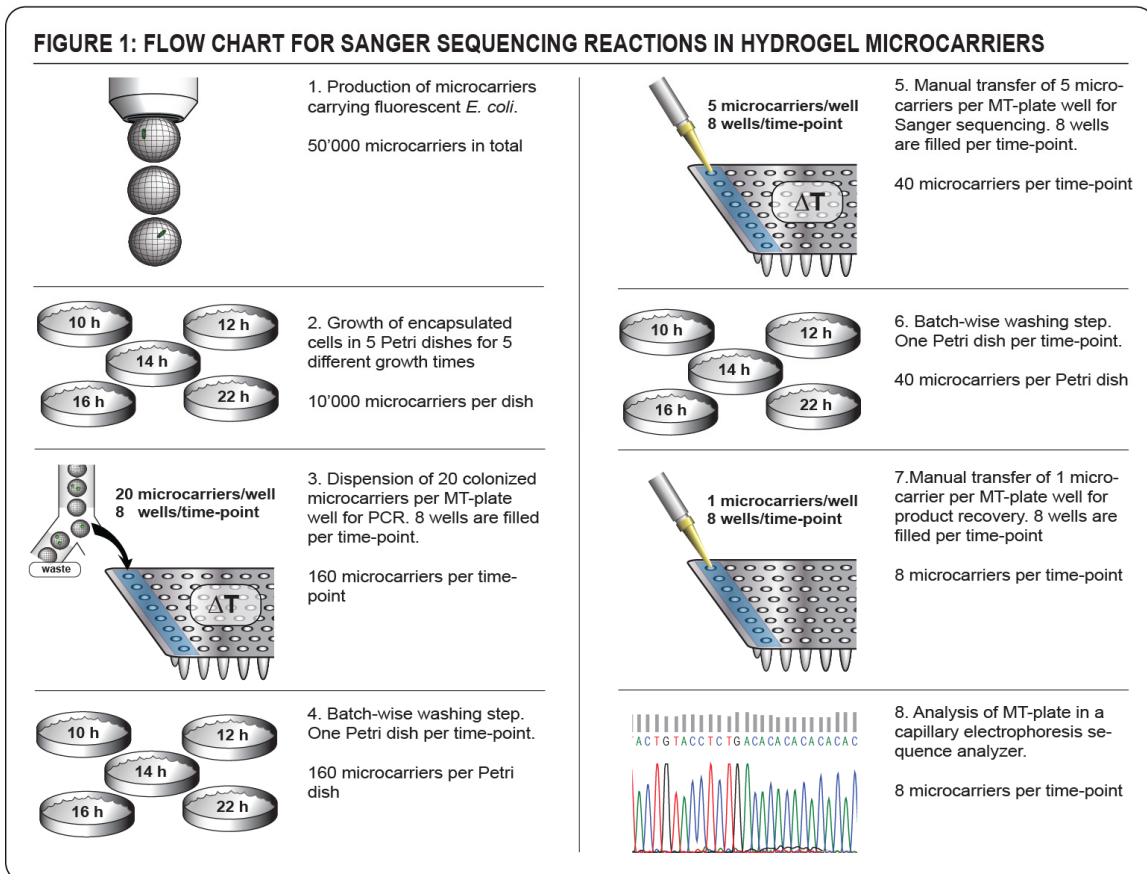
After the PCR-step (see above), 160 microcapsules per time point were transferred to a sieve (100µm Falcon cell sieve; BD Biosciences; San Jose; USA) by pipetting (using a decapitated pipette tip) and washed with 5 mL washing solution (0.1 mM BaCl<sub>2</sub> and 10 mM Tris-HCl pH 7.0). Afterwards, microcapsules were recovered in a Petri dish and re-incubated for 10 min at 25°C in 50 mL washing solution.

Five randomly selected microcapsules were then transferred to one 200 µL well in an MT-plate. Co-transferred superfluous washing solution was immediately afterwards removed by pipetting. In this manner, eight wells containing 5 microcapsules each were prepared per time point.

The wells were then filled with 10 µL sequencing mastermix containing 2 µL buffer, 2 µL BigDye v3.1 sequencing reagents (both from Applied Biosystems, Foster City, CA) and 1 M betaine (from a 2M stock solution), and 160 nM of primer M13F(-40) (5'-GTTTTCCCAGTCACGAC-3') (Microsynth; Balgach; Switzerland). The mixture was heat cycled (Eppendorf Mastercycler, Hamburg, Germany) with the following protocol: initial denaturation at 95°C for 2 min; 26 cycles: 95°C for 10 s; 50°C for 5 s; 60°C for 2 min; cooling down to 4°C. The fragments were rendered fluorescent by chain termination with a fluorescently labeled terminator nucleotide. After the thermal cycling, the superfluous liquid was removed from the microcapsules by pipetting. All microcapsules from the same time-point were pooled in a single vessel in 200 µL H<sub>2</sub>O, washed three times with 1 mL of H<sub>2</sub>O, and transferred to a Petri dish, from where eight randomly selected microcapsules could be recovered after microscopic inspection and transferred individually into a well of an MT-plate (one microcarrier per well), taking with them 10 µL of water needed for recovery of the fragments by diffusion. The plate was then incubated for 20 h at 4 °C and the supernatant was analyzed with an ABI3100 capillary electrophoresis



sequencer (Applied Biosystems, Foster City, CA; same parameters as indicated above), which allowed to determine the peak height of the FAM-labeled PCR products as well as the quality of the sequencing read (QV's > 20; corresponding to the number of bases correctly assigned by a probability of more than 99%) and the relative signal intensities.



### 3. Results

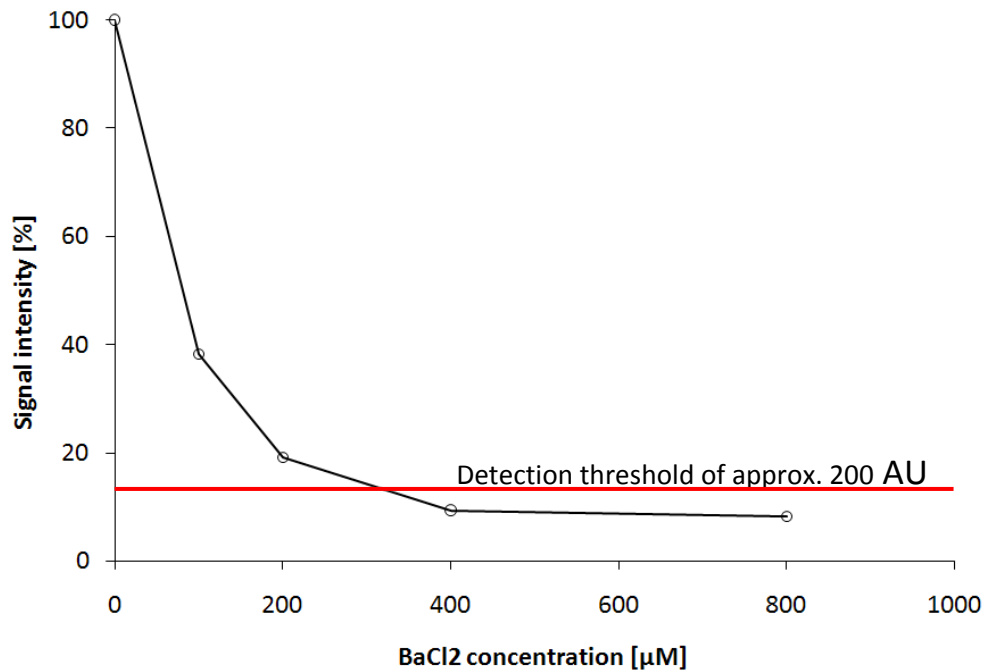
#### **Performance of the ABI3100 capillary electrophoresis sequencer (CE) in a context of bead-derived Sanger-fragments**

In order to define the lowest requirement for the synthesis of Sanger-fragments in microcarriers, we wanted to determine the detection limits of the ABI3100 sequencer. In addition, it was important to understand the influence of barium chloride on the analysis, as barium ions could be expected to leach from the microcapsules during the final diffusion step of Sanger-fragments out of the capsules.

While using a mixture of Sanger sequencing fragments obtained from a plasmid with an 800 bp insert as a template, a detection limit of 2.5 ng containing all 800 fragments in different length from the entire Sanger sequencing reaction dissolved in 10  $\mu$ L of demineralized water was estimated. At this concentration, signal intensities of the base-peaks in the raw-data were in the order of 100-200 AU and deconvolution of the rather complex information obtained during analysis by CE was no longer possible.

The influence of barium chloride on CE-performance was determined by the artificial addition of barium chloride salts to an undiluted sample synthesized according to the BigDyeV3.1 standard protocol (40 ng/ $\mu$ L of purified Sanger sequencing product dissolved in demineralized water; 10  $\mu$ L reaction volume) prior to CE analysis (figure 2). Even at the lowest tested concentration of 100  $\mu$ M barium chloride a considerable loss of signal intensity (38 %; from 2000 AU down to 700 AU) was observed if compared to the results obtained with a sample dissolved in demineralized water only (see figure 2). With increasing concentrations, signal intensity readily deteriorated and approached 9 % and 8 % at the highest tested barium chloride concentrations of 400  $\mu$ M and 800  $\mu$ M. At both concentrations, the signal intensity dropped below 200 AU and the

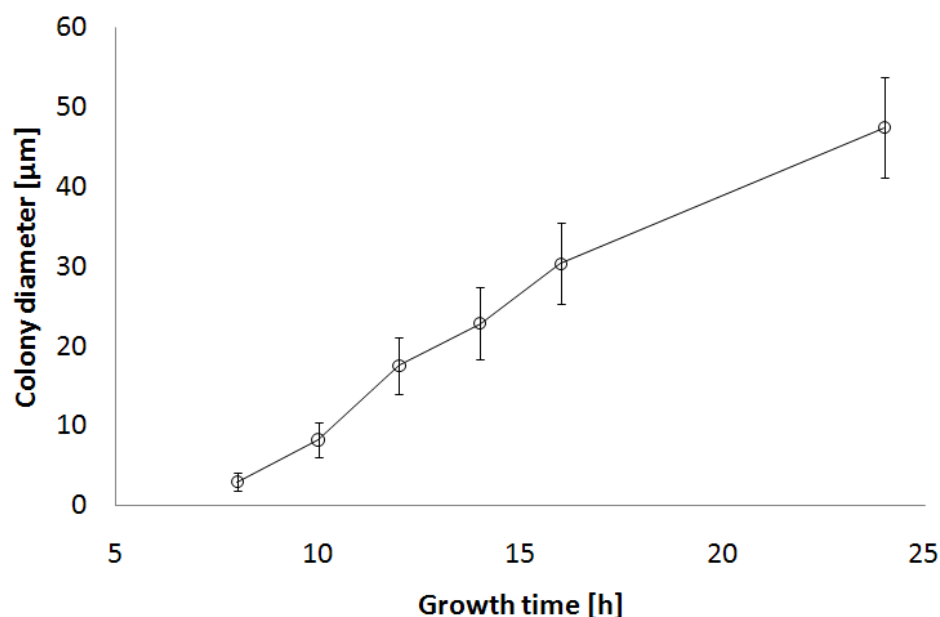
software algorithm was not able to assign a sequence anymore. Hence, we conclude that Sanger fragment synthesis in microcapsules had to deliver at least 2.5 ng of Sanger fragments for a successful analysis by CE, if no barium is present.



**Figure 2: Influence of barium chloride on CE performance.** Barium chloride was added to an undiluted Sanger sequencing sample (40 ng of purified Sanger sequencing product dissolved in demineralized water / µL; 10 µL reaction volume) prior to CE analysis.

### **Synthesis of dye-end labeled PCR-fragments in microcapsules**

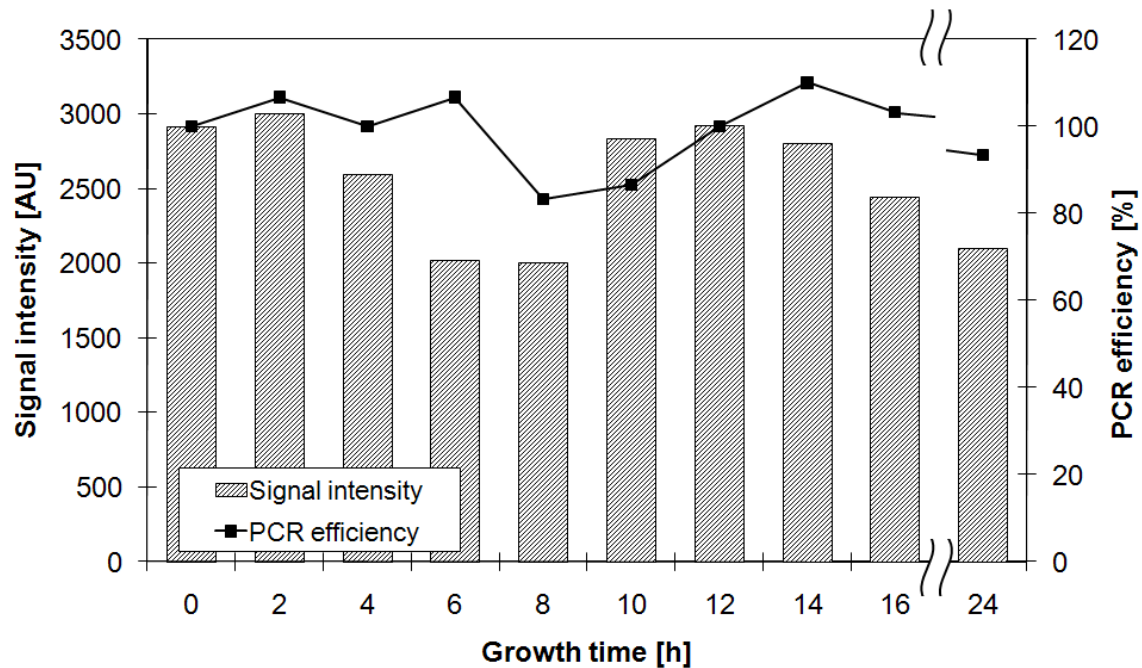
Our working hypothesis was that PCR-performance is positively affected by a high concentration of plasmid per microcapsule and negatively by a high concentration of other cells components inevitably formed during lysis of the microcolony embedded in the microcapsule. Therefore, we investigated the influence of colony growth time on the efficiency of PCR-based processes in the beads. For this, colony diameters for growth in microcapsules were microscopically determined as a function of time (see figure 3).



**Figure 3: Growth of the cells in microcapsules.**

Colony diameters were measured by microscopy at each growth time. Since the colonies at growth times below 8 h were microscopically not detectable, these colony diameters were not measured. However, at 8 h, the colonies had an average diameter of  $2.9 \pm 1.1 \mu\text{m}$ , at 10 h the diameter was  $8.2 \pm 2.2 \mu\text{m}$ , at 12 h  $17.5 \pm 3.6 \mu\text{m}$ , at 14 h  $22.8 \pm 4.5 \mu\text{m}$ , at 16 h  $30.3 \pm 5.1 \mu\text{m}$  and after 24 h of incubation the average diameter was determined to be  $47.4 \pm 6.3 \mu\text{m}$ .

As colonies, proliferated for less than 10 h, cannot be detected by COPAS (see chapter II), the microcarriers had not been curated for monoclonality prior to distribution and PCR, but the degree of occupation was at the end of the growth phase (24 h) determined as 75%. This led us to expect on average 30 successful PCR-reactions from a set of 40 beads. In fact, as shown in figure 4, PCR fragments could be produced from microcapsules with a high reliability of close to 100 % even if only a single cell was present at the reaction start. Neither the number of PCR-positive reactions nor the amount of detected PCR product was significantly influenced by the amount of cells present at the reaction start. The results of this experiment thus indicate that within the investigated range, PCR performance is not affected by the colony size.



**Figure 4: Relationship between colony size and growth time and PCR efficiency.** Microcapsules harvested after different growth times were distributed in MT-plate wells and subjected to PCR. An efficiency of 100% was defined when for one a population of 40 beads 30 positive PCR reactions were obtained (occupation ratio of 75%). As the microcapsules from the separate time points had not been curated for monoclonality, the numbers of occupied beads fluctuate around the average value of 30, giving rise to the values higher and lower than 100%.

### Investigation of growth time on sequencing product quality

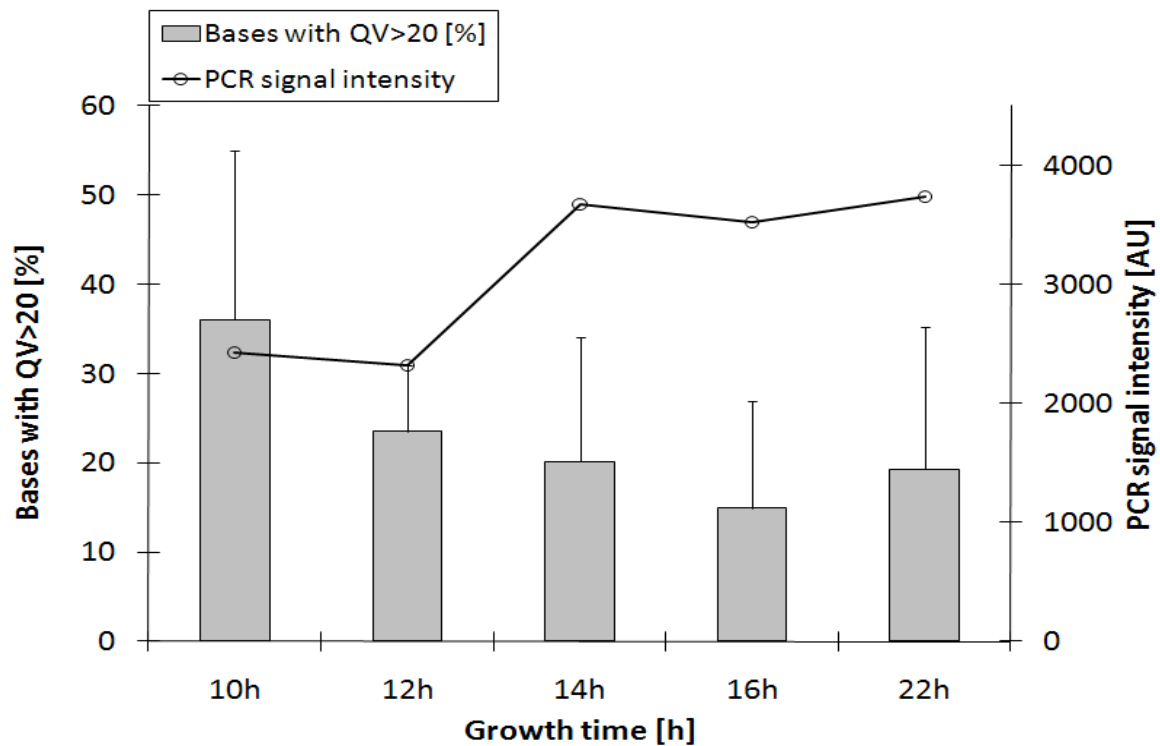
After clarifying the influence of cell growth on the efficiency of the PCR, we carried out a similar investigation for this influence on the production of Sanger sequencing fragments using a PCR-amplified fragment as a template. For this, recombinant cells carrying an amplicon of 390 bp were first grown in beads for 10 to 22 h subjected to PCR and then subjected to thermal cycling for Sanger-fragment production as described in the materials & methods section (see figure 1 for an overview of the experimental workflow).

As both, the PCR products formed during the first thermal cycling and the Sanger-sequencing fragments formed during the second thermal cycling, were

labeled with a fluorescent dye, they could both be quantified by CE from the supernatant in the well after the bead had been incubated over night in water and the fragments had time to diffuse out of the bead. As shown in figure 5, signal intensities obtained from the recovered PCR product increased from 2'400 AU for cells grown for 10 h (colony diameter of 8  $\mu\text{m}$ ) to 3'700 AU when cell were proliferated for 22 h (colony diameter of 45  $\mu\text{m}$ ). These values were broadly within the range of the signals obtained for the PCR products recovered after PCR in the previous section (2000 – 3000 AU; see figure 4), indicating that the PCR-product that served as the template remained immobilized during sequencing. Furthermore, a trend towards an increasing amount of PCR product with an increasing size of the colony subjected to PCR was observed.

Furthermore, sequencing products were detected in all samples (figure 5). The highest percentage of putatively correctly called bases ( $QV>20$ ) was 36 % (139 bases of the 390 bp fragment) and was found after 10 h of growth (colony diameter of 8  $\mu\text{m}$ ). The values decreased with increasing growth time to 15 % (58 bases; 16 h of growth; colony diameter of 30  $\mu\text{m}$ ). Interestingly, results improved for longer growth times again to 19 % correctly called bases (75 bases , 22 h of growth, 45  $\mu\text{m}$  colony diameter). The average read-length followed a similar trend, from 179 bp for 10 h of growth to 103 bp after 16 h of growth, and 124 bp after 22 h.

Both, the low number of  $QV>20$  bases as well as the short read-length, most likely resulted from the low quantities of Sanger fragments injected into the capillary electrophoresis unit. For example, we observed a raw data signal for Sanger-sequencing fragments after 10h of growth (198 bases  $>QV20$ ) not higher than 200 AU which is close to the previously defined threshold value at which signal deconvolution was still possible.



**Figure 5: Obtained bases with QV>20 and PCR signal of recovered products from microcarries.**

Left axis/grey bars: Ratio of bases with a quality value (QV) > 20 from Sanger-fragment producing reactions in microcapsules after specific colony growth times. Right axis: Total amounts of labeled PCR-products that had diffused out of the microcapsule after PCR and Sanger-fragment reaction (line).

#### **4. Discussion and outlook**

In order to discuss the results of this chapter, we would like to briefly point out some findings of the results of the previous section and of chapter II and III: [i] no or a negligible influence of the colony size on PCR efficiency; [ii] a rapid deterioration of COPAS sorter performance for enrichment of monoclonal microcarriers at less than 10 h of cultivation time corresponding to a colony size of  $< 8 \mu\text{m}$  (see chapter II); [iii] a superior PCR-performance for amplification of 440 bp double stranded PCR fragments within the microcapsules; [iv] a sufficient degree of immobilization of 440 bp double stranded fragments in order to allow dispersion of microbeads after PCR by COPAS; [v] we can reliably recover PCR-product from a microcapsule by diffusion into water over night.

Referring to these findings, we would like to discuss the influence of the average size of microcolonies grown within microcapsules on colony-PCR efficiency performed directly within the same capsules. Based on the results, an experiment aiming at first performing colony-PCR and second a sequencing-reaction of the PCR-product still immobilized within microcapsules was parameterized and carried out. All relevant steps, i.e. inoculation, growth, PCR, and (in the second experiment) also sequencing were performed using one and the same set of microcapsules during all process steps, thus effectively implementing the application of microcapsule technology to real-world problems.

Remarkably, PCR-efficiency was not influenced by the colony size. According to the results obtained by CE-analyses similar quantities of PCR product were detected independently on whether PCR-reactions started from a single cell (approx. 200 plasmid-copies) or from microcolonies comprised of



approximately 50'000 cells. This rather surprising result could be due to anyone of the following reasons: First, the final amount of PCR-product might not be limited by the amount of template (plasmid-DNA) present at the reaction start but by other reaction components, such as nucleotides, PCR primers, or polymerase. Second, the capsule geometry might limit the yield, so that the PCR process might proceed only until PCR-inhibitory concentrations of polynucleotides are accumulated within the capsule interior. Third, the protocol used for extraction of the PCR-products might fail to recover all PCR-product accumulated within the microcapsules but instead could deliver a constant amount of PCR-product as the driving concentration gradient leads in the different cases to similar steady-state diffusion for the various beads. At this point, we cannot differentiate between these possible explanations.

Also Sanger-fragment generation performed on the PCR product generated previously could be demonstrated for colonies obtained in the investigated operation window of colony sizes of 8  $\mu\text{m}$  to 45  $\mu\text{m}$  (growth time of 10 h to 22 h). However, the signals obtained during the CE-analysis of the recovered Sanger-fragments were generally rather low. Signal intensities remained close to the previously defined lower limit of quantification, i.e. to the minimal signal intensities required by the sequencer software in order to correctly deconvolute the raw data. Only a limited number of bases could be called reliably with a  $QV > 20$ . However, similar amounts of PCR-products were recovered from the microcarriers that were thermocycled twice (PCR followed by Sanger fragment generation) as from those cycled only once (PCR only). These results suggest that the PCR product remained immobilized in the course of the Sanger sequencing reaction and should thus have been available as a template for the Sanger sequencing reaction within the microcapsule.

A first potential explanation for the rather low quantities of isolated Sanger fragments is thus that the Sanger reaction proceeded at rather low efficiency within the microcapsules. Even though PCR worked efficiently within the microcapsules, it cannot be excluded that Sanger fragmentation generation does not. One reason for the deteriorated performance could be the rather high amounts of PCR products and / or the accumulated Sanger sequencing products present in the microcapsules compete with the sequencing-primer for binding and therefore inhibit the generation of Sanger fragments. Other possibilities are inhibition of the polymerase used for sequencing (AmpliTaq) by other substances present in the microcapsules (e.g. alginate, barium, or still immobilized cell debris), for which there are indications in the literature.<sup>156</sup>

A second potential reason for the low amounts of isolated dye-terminated Sanger fragments could be that the fragments were not retained by the microcapsule but lost during synthesis or in the course of the subsequent washing steps. We indeed found some evidence that loss by diffusion took place (data not shown). Interestingly, no bases could be called close to the region directly after the primer up to a fragment size of 140 bp in any of the sequencing runs. We therefore tentatively suggest that these fragments were indeed lost at some stage of the protocol. As the diffusion losses are most likely not occurring in an all-or-nothing reaction but gradually, loss of considerable amounts of Sanger fragments exceeding the size of those that could not be detected at all is therefore rather likely. This could explain the generally low signal intensities of the sequencing products. The putatively poor retention of smaller fragments is of minor importance for PCR-screening applications as here usually large fragments are synthesized and statistical loss of smaller fragments is of minor importance. However, in case of sequencing applications every called base counts. Fragment loss is thus a severe limitation.<sup>157</sup>

A third explanation for the poor amount of detected Sanger sequencing fragments is that fragments were not recovered from the microcapsules by the method applied for recovery of double stranded PCR products (recovery by diffusion into pure water over night). However, the fact that no smaller fragments were detected after CE analysis is an indirect proof for the potency of the recovery method and we consider this possibility as rather unlikely.

We tentatively suggest that poor retention of the sequencing products is likely to be the main problem. Potential strategies for avoiding the loss by diffusion are the use of a gel with a lower mesh width or the suppression of diffusion losses by performing the reaction in a suspension-emulsion in which the microcapsules are immersed in oil. However, each of these potential solutions requires considerable development work or is made difficult by other constraints. The mesh width of alginate, for instance, can be influenced by the concentration and the type of metal ions.<sup>66</sup> However, the barium-gel used here has already been optimized for maximal stability (results not shown) which is synonymous for the smallest possible mesh-width.<sup>66</sup> Alternatively, alginates of higher guluronic acid content could be used but these materials are not easily available and would have to be synthesized, which in turn is difficult as alginates are essentially insoluble in any solvent except water and in addition are prone to hydrolysis at high temperature and low or high pH. For these reasons, chemical modification of alginates is rather difficult. Alternative other gels could be used but from our experience with the development of PCR protocols in alginates, such a material change brings with it an enormous amount of development work.<sup>158, 159</sup>

For the sequencing application, additional development effort would also have to be directed towards optimization of the PCR conditions. As shown above, PCR also starts from a single cell. A positive aspect is that as growth prior to

PCR is not required rather time-effective PCR-screening protocols employing just two key-steps namely encapsulation of transformants followed directly by PCR could be developed.

However, for sequencing applications, it is important to ensure that not more than one cell or one clone is available in one bead. Efficient separation of monoclonal microcarriers of the size used here from empty or polyclonal reactors is only feasible by COPAS technology and only in a certain operation window (see chapter II). However, sensitivity of the COPAS is not sufficient in order to detect a single cell. Furthermore, approximately 5 to 10 % of the encapsulated cells either do not survive the encapsulation procedure or fail to grow afterwards even under fully optimized conditions (results not shown). Therefore, we cannot exclude that in some scenarios a non-proliferating cell might be present next to a detectable colony, which might present a problem for some applications. The fact that the PCR protocol described in here also starts from single cells can be a limitation if it comes to the preparation of samples for Sanger fragment generation.

However, in addition to the development challenge discussed above, microcapsule-supported protocols for synthesis of Sanger sequencing fragments suffer from some intrinsic limitations. The most severe one is most likely the impossibility of producing a backup copy of the clones prior to encapsulation. In the course of the processing steps, in particular during thermocycling all viable cells are killed. Therefore, the only way of again getting hold of the genetic information that led to the analytic result is recovery of the genetic information, i.e. the plasmid-DNA (see chapter III) or the PCR fragments from the well after CE-analysis. The respective procedures are however rather laborious and applications relying on the availability of a copy of a given clone

(e.g. paired-end sequencing or re-sequencing) are therefore rather unlikely to ever work at acceptable costs.

Taken together, development of microcapsule-based protocols for synthesis of Sanger fragments translates into a considerable development challenge that was not pursued further in the course of this thesis. We see a number of indications that the Sanger-fragment generation protocol can be performed within microcapsules. Still, the potential impact on the field of shotgun sequencing remains questionable due to intrinsic limitation of microcapsule based approaches as well as due to the general progress made in the field of automated sequencing (e.g. pyrosequencing),<sup>36, 51, 52, 55</sup> which will presumably outcompete Sanger sequencing for large scale genomic applications in the near future.

## **VI. Concluding remarks and outlook**

In screening and selection processes applied in the course of biotechnological research and development, a general trend towards increased throughput and decreased sample volumes can be observed. Amongst one of the fastest analysis and sorting technologies is fluorescence activated cell sorting (FACS), which can analyze single cells for fluorescent signals. But since sample processing volumes of a FACS are generally too low for processing of entire colonies, two-dimensional array formats, such as agar plates or microtiter-plates, are employed. For handling of individual cultures, a 1536-well plate is one of the formats with the highest colony densities still enabling the processing of individual clones by robot-stations while for lower volumes, random sample distribution into individual compartments is applied (e. g. Gigamatrix-plates). Hence, high throughput sample processing technologies for the analysis of individual colonies in low volume-compartments still is a key demand of today's lifescience. A well established technique for cultivation of cells in formats other than in two-dimensional arrays is cell encapsulation and proliferation in hydrogel microcapsules. Due to the nature of the capsule, small molecules can penetrate through the gel and thus supply nutrients and / or reagents to the embedded organism. However, in the field of microencapsulation no high throughput method for controlled inoculation of microcapsules is known. Instead, cells are generally randomly distributed within the microcapsules. Production of a large fraction of microcapsules inoculated by exactly one clone requires therefore either high dilution of the cells prior to encapsulation - thereby leaving the bulk of the microcapsules empty - or procedures for the specific isolation of the monoseptically colonized fraction after the embedment. On the other hand, availability of "pure" colonies, i.e. colonies originating from a single cell and not from a mixture of cells, is a key requirement for a large number of applications.

We addressed the challenge of isolating larger amounts of monoseptically colonized microcapsules by eliminating the unwanted background of empty and overpopulated samples in the course of this thesis in chapter II, where we exploited a particle sorter (COPAS) for enriching monoclonal microcapsules. The device allows recording of one-dimensional microcapsule profiles, which can be used to detect monoclonal microcapsules, but entails the potential for occlusion of colonies by each other on a statistical basis. The underlying phenomena were modeled and verified experimentally. Two main factors were identified as crucial: First, the efficiency at which a single colony can be detected and, second, the probability that two colonies cannot be resolved in a one-dimensional scan. Both phenomena have an opposing effect: While the efficiency of detection increases with an increasing colony-size, the increase in colony size decreases the possibility that two colonies can still be discriminated, leading to an optimal “operation point” of the system.

In chapter III, we developed a first application for the hydrogel-microcapsule technology. It was used to develop a colony-PCR screening technology for identification of conserved sequence elements of the cassava genome. Large fractions of monoclonal microcapsules were produced and then subjected to colony-PCR directly in the microcapsules. PCR-positive capsules could be identified by staining and COPAS-based flow cytometry. Positive microcapsules could thus be isolated and the underlying plasmid could be recovered and sequenced. While screening a total sequence space of 8 Mbp, 11 high-quality microsatellites for polymorphism studies in cassava were isolated from a library of 20'000 clones in only 2 days.

However, we reason that even higher throughputs should be feasible. The volumes of the microcapsules (35 nL), for instance, could still be reduced



further and we demonstrated that colony-PCR also works in capsules of a volume of approx. 1 nL (results not shown). Further possible improvements could be achieved by the application of fluorescently labeled deoxyuridine triphosphate (dUTPs) during PCR product synthesis. Since SYBR Green stains the entire double stranded DNA present in the microcapsule (including genomic and plasmid DNA of embedded colonies), we believe that application of labeled dUTPs may allow improving the signal to noise ratio to the extent required in order to reliably detect PCR products even in very small microcapsules. Furthermore, the screened sequence space per microcapsule could most likely be increased further by optimization of the PCR protocols and we argue that protocols for PCR screening of up to 4 Gbp could be developed without a principal change of the technology.

In chapter IV, we present a protocol for highly parallelized, one-step analysis and isolation of *E. coli* clones in microcapsules based on catalase as a reporter protein. Catalase converts hydrogen peroxide to water and gaseous oxygen at very high rates. By coupling of the desired phenotype to the amount of catalase present within a microcapsule, gas-bubble formation-induced density differences could be exploited for separation of microcapsules. The catalase-based analysis and isolation protocol was demonstrated for the separation of encapsulated catalase-positive strains from a background of catalase negative strains. Based on the results, the catalase positive fraction was isolated at a purity of 100 % and a sorting efficiency of approx 70 %. The protocol has substantial benefits when compared to existing systems as the indicative event (i.e. the analysis) and the isolation happen at the same time which allows analysis and isolation of a very large numbers of positive clones in a few minutes. The microcapsules used had a volume of 1 nL which suggests that

sample numbers of up to 1 billion could be straightforwardly analyzed and isolated in a single run while still keeping the total volume close to 1 L.

In chapter V, attempts towards synthesis of dye-terminated DNA-fragments for Sanger sequencing are summarized. However, the protocol was not optimized to the extent required for its application as the obtained product yields were too low in order to reliably detect them by capillary electrophoresis analyzer favored as a suitable readout. Fragment-loss during synthesis of the dye terminated fragments occurred to be the most likely cause for failure of the protocol but a final proof has not been made. This result illustrates that the transfer even of well established laboratory protocols into a miniaturized and highly parallelized state is not trivial, but rather a complex procedure. Still, the obtained results and insights will be important in the further developments of high-throughput screening protocols.

In summary, this thesis presents the processing of hydrogel microcapsules with a commercially available flow cytometer as well as a novel ultra high throughput screening technique relying on catalase-induced density differences. The protocols build a new base for implementation of high throughput screening platforms in lifescience such as demonstrated with a newly developed colony-PCR screening method.

## References

1. Tajima, K., Aminov, R.I., Nagamine, T., Ogata, K., Nakamura, M., Matsui, H. & Benno, Y. Rumen bacterial diversity as determined by sequence analysis of 16S rDNA libraries. *FEMS Microbiol. Ecol.* 29, 159-169 (1999).
2. Krause, D.O., Denman, S.E., Mackie, R.I., Morrison, M., Rae, A.L., Attwood, G.T. & McSweeney, C.S. Opportunities to improve fiber degradation in the rumen: microbiology, ecology, and genomics. *FEMS Microbiol. Rev.* 27, 663-693 (2003).
3. Stevenson, D.M. & Weimer, P.J. Dominance of *Prevotella* and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. *Appl. Microbiol. Biotechnol.* 75, 165-174 (2007).
4. Hall, B.G. & Faunce III, W. Functional genes for cellobiose utilization in natural isolates of *Escherichia coli*. *J. Bacteriol.* 169, 2713-2717 (1987).
5. Galbraith, E.A., Antonopoulos, D.A. & White, B.A. Suppressive subtractive hybridization as a tool for identifying genetic diversity in an environmental metagenome: the rumen as a model. *Env. Microbiol.* 6, 928-937 (2004).
6. Hall, B.G., Pikis, A. & Thompson, J. Evolution and biochemistry of family 4 glycosidases: implications for assigning enzyme function in sequence annotations. *Mol. Biol. Evol.* 26, 2487-2497 (2009).
7. Blow, N. Exploring unseen communities. *Nature* 453, 687-690 (2008).
8. Zhang, P.Y.H., Himmel, M.E. & Mielenz, J.R. Outlook for cellulase improvement: screening and selection strategies. *Biotechnol. Adv.* 24, 452-481 (2006).
9. Wang, F., Li, F., Chen, G. & Liu, W. Isolation and characterization of novel cellulase genes from uncultured microorganisms in different environmental niches. *Microbiol. Res.* 164, 650-657 (2009).
10. Dietrich, J.A., McKee, A.E. & Keasling, J.D. High throughput metabolic engineering: advances in small molecule screening and selection. *Annu. Rev. Biochem.* 79, 21.21-21.28 (2010).
11. Dove, A. Screening for content - the evolution of high throughput. *Nat. Biotechnol.* 21, 859-864 (2003).

12. Lorenz, P. & Eck, J. Metagenomics and industrial applications. *Nature Reviews* 3, 510-516 (2005).
13. Reymond, J.C. & Babiak, P. Screening systems. *Adv. Biochem. Eng. Biotechnol.* 105, 31-58 (2007).
14. Schmeisser, C., Steele, H. & Streit, W. Metagenomics, biotechnology with non-culturable microbes. *Appl. Microbiol. Biotechnol.* 75, 955-962 (2007).
15. Aharoni, A., Amitai, G., Bernath, K., Magdassi, S. & Tawfik, D.S. High-throughput screening of enzyme libraries: Thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments. *Chembiol* 12, 1281-1289 (2005).
16. Sauer, S., Lange, B.M.H., Gobom, J., Nyarsik, L., Seitz, H. & Lehrach, H. Miniaturization in functional genomics and proteomics. *Nat. Rev. Genet.* 6, 465-476 (2005).
17. Fan, J.B. & Chee, M.S. Highly parallel genomic assays. *Nat. Rev. Genet.* 7, 632-644 (2006).
18. Inglese, J., Johnson, R.L., Simeonov, A., Xia, M., Zheng, W., Austin, C.P. & Auld, D.S. High-throughput screening assays for the identification of chemical probes. *Nat. Chem. Biol.* 3, 466-479 (2007).
19. Kim, B.S., Kim, S.Y., Park, J., Park, W., Hwang, K.Y., Yoon, Y.J., Oh, W.K., Kim, B.Y. & Ahn, J.S. Sequence-based screening for self-sufficient P450 monooxygenase from a metagenomic library. *J. Appl. Microbiol.* 102, 1392-1400 (2007).
20. Uchiyama, T. & Watanabe, K. Substrate-induced gene expression (SIGEX) screening of metagenome libraries. *Nature Protocols* 3, 1202-1212 (2008).
21. Khattak, S.F., Spataro, M., Roberts, L. & Roberts, S.C. Application of colorimetric assays to assess viability, growth and metabolism of hydrogel-encapsulated cells. *Biotechnol. Lett.* 28, 1361-1370 (2006).
22. Ibrahim, S.F. & van den Engh, G. High-speed cell sorting: fundamentals and recent advances. *Curr. Opin. Biotechnol.* 14, 5-12 (2003).
23. Ryan, C., Nguyen, B.T. & Sullivan, S.J. Rapid assay for mycobacterial growth and antibiotic susceptibility using gel microdrop encapsulation. *J. Clin. Microbiol.* 33, 1720-1726 (1995).
24. Becker, S., Schmoldt, H.U., Adams, T.M., Wilhelm, S. & Kolmar, H. Ultra-high-throughput screening based on cell-surface display and fluorescence-activated cell

- sorting for the identification of novel biocatalysts. *Curr. Opin. Biotechnol.* 15, 323-329 (2004).
25. Eisenstein, M. Divide and conquer. *Nature* 441, 1179-1185 (2006).
  26. Leamon, J.H., Link, D.R., Egholm, M. & Rothberg, J.M. Overview: methods and applications for droplet compartmentalization of biology. *Nat. Meth.* 3, 541-543 (2006).
  27. Link, A.J., K.J., J. & Georiou, G. Beyond toothpicks: new methods for isolating mutant bacteria. *Nature* 5, 680-688 (2007).
  28. Zengler, K., Walcher, M., Clark, G., Haller, I., Toledo, G., Holland, T., Mathur, E.J., Woodnutt, G., Short, J.M. & Keller, M. in *Methods in Enzymology*, Edn. Volume 397. (ed. J.R. Leadbetter) 124-130 (Academic Press, 2005).
  29. Taly, V., Kelly, B.T. & Griffiths, A.D. Droplets as microreactors for high-throughput biology. *Chembiochem* 8, 263-272 (2007).
  30. Kelly, B.T., Baret, J.C., Taly, V. & Griffiths, A.D. Miniaturizing chemistry and biology in microdroplets. *Chem. Commun.* 18, 1773-1788 (2007).
  31. Entcheva, P., Liebl, W., Johann, A., Hartsch, T. & Streit, W.R. Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *App. Env. Microbiol.* 67, 89-99 (2001).
  32. Kachroo, A.H., Kancharla, A.K., Singh, N.S., Varshney, U. & Mahadevan, S. Mutations that alter the regulation of the *chb* operon of *Escherichia coli* allow utilization of cellobiose. *Mol. Microbiol.* 66, 1382-1395 (2007).
  33. Sahar, E., Nir, R. & Lamed, R. Flow cytometric analysis of entire microbial colonies. *Cytometry* 15, 213-221 (1994).
  34. Teh, S.Y., Lin, R., Hung, L.H. & Lee, A.P. Droplet microfluidics. *Lab Chip* 8, 198-220 (2008).
  35. Walsler, M., Leibundgut, R., Pellaux, R., Panke, S. & Held, M. Isolation of monoclonal microcarriers colonized by fluorescent *E. coli*. *Cytometry Part A* 73A, 788-798 (2008).
  36. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* 26, 1135-1145 (2008).
  37. Cowan, D., Meyer, Q., Stafford, W., Muyanga, S., Cameron, R. & Wittwer, P. Metagenomic gene discovery: past, present, future. *Trends Biotechnol.* 23, 321-329 (2005).

38. Green, B.D. & Keller, M. Capturing the uncultivated majority. *Curr. Opin. Biotechnol.* 17, 236-240 (2006).
39. Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W. et al. Environmental genome shotgun sequencing of the sargasso sea. *Science* 304, 66-74 (2004).
40. Shimomura, O., Johnson, F.H. & Saiga, Y. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*. *J. Cell. Comp. Physiol.* 59, 223-239 (1962).
41. Stewart, C.N. Go with the glow: fluorescent proteins to light transgenic organisms. *Trends Biotechnol.* 24, 155-162 (2006).
42. Arum, K.H.S., Kaul, C.L. & Ramarao, P. Green fluorescent proteins in receptor research: An emerging tool for drug discovery. *J. Pharmacol. Toxicol. Methods* 51, 1-23 (2005).
43. Shaner, N.C., Steinbach, P.A. & Tsien, R.Y. A guide to choosing fluorescent proteins. *Nat. Meth.* 2, 905-909 (2005).
44. Hellrung, D.J., Rossi, G. & Link, C.J. High-Throughput Fluorescent Screening of Transgenic animals: Phenotyping and Haplotyping. *Cytometry Part A* 69A, 1092-1095 (2006).
45. Williams, G.B., Weaver, J.C. & Demain, A.L. Rapid microbial detection and enumeration using gel microdroplets and colorimetric or fluorescence indicator systems. *J. Clin. Microbiol.* 28, 1002-1008 (1990).
46. Miesenböck, G., De Angelis, D.A. & Rothman, J.E. Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins. *Nature* 394, 192-195 (1998).
47. Dressman, D., Yan, H., Traverso, G., Kinzler, K.W. & Vogelstein, B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc. Natl. Acad. Sci. U. S. A.* 100, 8817-8822 (2003).
48. Jose, J., Betscheider, D. & Zangen, D. Bacterial surface display screening by target enzyme labeling: Identification of new human cathepsin G inhibitors. *Anal. Biochem.* 346, 258-267 (2005).

49. Perry, J.D., Morris, K.A., James, A.L., Oliver, M. & Gould, F.K. Evaluation of novel chromogenic substrates for the detection of bacterial beta-glucosidase. *J. Appl. Microbiol.* 102, 410-415 (2007).
50. Carpenter, A.E. Image-based chemical screening. *Nat. Chem. Biol.* 3, 461-465 (2007).
51. Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Z., Dewell, S.B. et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376-380 (2005).
52. Hutchison, C.A., III DNA sequencing: bench to bedside and beyond. *Nucleic Acids Res.* 35, 6227-6237 (2007).
53. Hodges, E., Xuan, Z., Balija, V., Kramer, M., Molla, M.N., Smith, S.W., Middle, C.M., Rodesch, M.J., Albert, T.J., Hannon, G.J. et al. Genome-wide in situ exon capture for selective resequencing. *Nat. Genet.* 39, 1522-1527 (2008).
54. Shendure, J. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 309, 1728-1732 (2005).
55. Holt, K.E., Parkhill, J., Mazzoni, C.J., Roumagnac, P., Weill, F.-X., Goodhead, I., Rance, R., Baker, S., Maskell, D.J., Wain, J. et al. High-throughput sequencing provides insights into genome variation and evolution in *Salmonella Typhi*. *Nat. Genet.* 40, 987-993 (2008).
56. Farrar, K. & Donnison, I. Construction and screening of BAC libraries made from *brachypodium* genomic DNA. *nature Protocols* 2, 1661-1674 (2007).
57. Gupta, P.K. & Varshney, R.K. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* 113, 163-185 (2000).
58. Kim, C.G., Fujiyama, A. & Saitou, N. Construction of a gorilla fosmid library and its PCR screening system. *Genomics* 82, 571-574 (2003).
59. Burns, A.R., Kwok, T.C.Y., Howard, A., Houston, E., Johanson, K., Chan, A., Cutler, S.R., McCourt, P. & Roy, P.J. High-throughput screening of small molecules for bioactivity and target identification in *Caenorhabditis elegans*. *Nature Protocols* 1, 1906-1914 (2006).
60. Hahne, F., Arlt, D., Sauermann, M., Majety, M., Poustka, A., Wiemann, S. & Huber, W. Statistical methods and software for the analysis of highthroughput reverse genetic assays using flow cytometry readouts. *Genome Biol.* 7, R77.71-R77.12 (2006).

61. Uchiyama, T., Takashi, A., Toshimichi, I. & Watanabe, K. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nat. Biotechnol.* 23, 88-93 (2005).
62. Dupuy, D., Bertin, N., Hidalgo, C.A., Venkatesan, K., Tu, D., Lee, D., Rosenberg, J., Svrzikapa, N., Blanc, A., Carnec, A. et al. Genome-scale analysis of in vivo spatiotemporal promoter activity in *Caenorhabditis elegans*. *Nat. Biotechnol.* 25, 663-668 (2007).
63. Jones, P., Watson, A., Davies, M. & Stubbings, S. Integration of image analysis and robotics into a fully automated colony picking and plate handling system. *Nucleic Acids Res.* 20, 4599-4606 (1992).
64. Gift, E.A., Park, H.J., Paradies, G.A., Demain, A.L. & Weaver, J.C. FACS-based isolation of slowly growing cells: Double encapsulation of yeast in gel microdrops. *Nat. Biotechnol.* 14, 884-887 (1996).
65. Weaver, J.C., McGrath, P. & Adams, S. Gel microdrop technology for rapid isolation of rare and high producer cells. *Nat. Meth.* 3, 583-585 (1997).
66. Bienaimé, C., Barbotin, J.N. & Nava-Saucedo, J.E. How to build an adapted and bioactive cell microenvironment? A chemical interaction study of the structure of Calcium alginate matrices and their repercussion on confined cells. *J. Biomed. Mater. Res. Part A* 67A, 376-388 (2003).
67. Brandenberger, H. & Widmer, F. Immobilization of highly concentrated cell suspensions using the laminar jet breakup technique *Biotechnol. Prog.* 15, 366-372 (1999).
68. Martin-Banderas, L., Flores-Mosquera, M., Riesco-Chueca, P., Rodriguez-Gil, A., Cebolla, A., Chavez, S. & Ganán-Calvo, A.M. Flow focusing: a versatile technology to produce size-controlled and specific-morphology microparticles. *Small* 1, 688-692 (2005).
69. Beck, J., Angus, R., Madsen, B., Britt, D., Vernon, B. & Nguyen, K.T. Islet encapsulation: Strategies to enhance islet cell functions. *Tissue Eng.* 13, 589-599 (2007).
70. Diehl, F., Li, M., He, Y., Kinzler, K.W., Vogelstein, B. & Dressman, D. BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions. *Nat. Meth.* 3, 551-559 (2006).



71. Li, M., Diehl, F., Dressman, D., Vogelstein, B. & Kinzler, K. BEAMing up for detection and quantification of rare sequence variants. *Nat. Meth.* 3, 95-97 (2006).
72. Kiss, M.M., Ortoleva-Donnelly, L., Beer, N.R., Warner, J., Bailey, C.G., Colston, B.W., Rothberg, J.M., Link, D.R. & Leamon, J.H. High-throughput quantitative polymerase chain reaction in picoliter droplets. *Anal. Chem.* 80 (2008).
73. Williams, G.J. & Thorson, J.S. A high-throughput fluorescence-based glycosyltransferase screen and its application in directed evolution. *Nature Protocols* 3, 357-362 (2008).
74. Brouzes, E., Medkova, M., Savenelli, N., Marran, D., Twardowski, M., Hutchison, J.B., Rothberg, J.M., Link, D.R., Perrimon, N. & Samuels, M.L. Droplet microfluidics technology for single-cell high-throughput screening. *Proc. Natl. Acad. Sci. U. S. A.* (2009).
75. Nakano, M., Nakai, N., Kurita, H., Komatsu, J., Takashima, K., Katsura, S. & Mizuno, A. Single-molecule reverse transcription polymerase chain reaction using water-in-oil emulsion. *J. Biosci. Bioeng.* 99, 293-295 (2005).
76. Tewhey, R., Warner, J., Nakano, M., Libby, B., Medkova, M., David, P.H., Kotsopoulos, S.K., Samuels, M.L., Hutchison, J.B., Larson, J.W. et al. Microdroplet-based PCR enrichment for large-scale target sequencing. *Nat. Biotechnol.* 27, 1025-1031 (2009).
77. Mazutis, L., Araghi, A.F., Miller, O.J., Baret, J.C., Frenz, L., Janoshazi, A., Taly, V., Miller, B.J., Hutchison, J.B., Link, D.R. et al. Droplet-based microfluidics systems for high-throughput single DNA molecule isothermal amplification and analysis. *Anal. Chem.* 81, 4813-4821 (2009).
78. Keij, J.F., van Rotterdam, A., Groenewegen, A.C., Stokdijk, W. & Visser, J.W.M. Coincidence in high-speed flow cytometry: models and measurements. *Cytometry* 12, 398-404 (1991).
79. Pereira, D.A. & Williams, J.A. Origin and evolution of high throughput screening. *Br J Pharmacol* 152, 53-61 (2007).
80. Reymond, J.C. & Babiak, P. Screening systems. *Adv Biochem Engin/Biotechnol* 105, 31-58 (2007).
81. Inglese, J., Johnson, R.L., Simeonov, A., Xia, M., Zheng, W., Austin, C.P. & Auld, D.S. High-throughput screening assays for the identification of chemical probes. *Nat Chem Biol* 3 (2007).

82. Carpenter, A.E. Image-based chemical screening. *Nat Chem Biol* 3, 461-465 (2007).
83. Sauer, S., Lange, B.M.H., Gobom, J., Nyarsik, L., Seitz, H. & Lehrach, H. Miniaturization in functional genomics and proteomics. *Nat Rev Genet* 6, 465-476 (2005).
84. Wan, K.H., Yu, C., George, R.A., Carlson, J.W., Hoskins, R.A., Svirskas, R., Stapleton, M. & Celniker, S.E. High-throughput plasmid cDNA library screening. *Nat Protocols* 1, 624-632 (2006).
85. Zimmermann, H., Zimmermann, D., Reuss, R., Feilen, P.J., Manz, B., Katsen, A., Weber, M., Ihmig, F.R., Ehrhart, F., Gessner, P. et al. Towards a medically approved technology for alginate-based microcapsules allowing long-term immunoisolated transplantation. *J Mater Sci - Mater Med* 16, 491-501 (2005).
86. Khattak, S.F., Spatara, M., Roberts, L. & Roberts, S.C. Application of colorimetric assays to assess viability, growth and metabolism of hydrogel-encapsulated cells. *Biotechnol Lett* 28, 1361-1370 (2006).
87. Champagne, C.P. & Fustier, P. Microencapsulation for the improved delivery of bioactive compounds into foods. *Curr Opin Biotechnol - Plant Biotechnol - Food Biotechnol* 18, 184-190 (2007).
88. Weaver, J.C., McGrath, P. & Adams, S. Gel microdrop technology for rapid isolation of rare and high producer cells. *Nat Med* 3, 583-585 (1997).
89. Zengler, K., Walcher, M., Clark, G., Haller, I., Toledo, G., Holland, T., Mathur, E.J., Woodnutt, G., Short, J.M. & Keller, M. in, Edn. Volume 397 124-130 (Academic Press, 2005).
90. Rabanel, J.M., Bertrand, N., Sant, S., Louati, S. & Hildgren, P. Polysaccharide hydrogels for the preparation of immunoisolated cell delivery systems. *ACS Symposium Series* 934, 305-339 (2006).
91. Ryan, C., Nguyen, B.T. & Sullivan, S.J. Rapid assay for mycobacterial growth and antibiotic susceptibility using gel microdrop encapsulation. *J Clin Microbiol* 33, 1720-1726 (1995).
92. Serp, D., Cantana, E., Heinzen, C., von Stockar, U. & Marison, I.W. Characterization of an encapsulation device for the production of monodisperse alginate beads for cell immobilization. *Biotechnol Bioeng* 70, 41-53 (2000).
93. Jayasinghe, S.N. & Suter, N. Aerodynamically assisted jetting: a pressure driven approach for processing nanomaterials. *Mic Nano Lett* 1, 35-38 (2006).

94. Paepe, I.D., Declercq, H., Cornelissen, M. & Schacht, E. Novel hydrogels based on methacrylate-modified agarose. *Polymer Int* 51, 867-870 (2002).
95. Bienaimé, C., Barbotin, J.N. & Nava-Saucedo, J.E. How to build an adapted and bioactive cell microenvironment? A chemical interaction study of the structure of Ca-alginate matrices and their repercussion on confined cells. *J Biomed Mater Res Part A* 67A, 376-388 (2003).
96. Heinemann, M., Meinberg, H., Büchs, J., Koss, H.J. & Ansorge-Schumacher, M.B. Method for quantitative determination of spatial polymer distribution in alginate beads using raman spectroscopy. *Appl Spect* 59, 280-285 (2005).
97. Nir, R., Lamed, R., Gueta, L. & Sahar, E. Single-cell entrapment and microcolony development within uniform microspheres amenable to flow cytometry. *Appl Environ Microbiol* 56, 2870-2875 (1990).
98. Rosenblatt, J.I., Hokanson, J.A., McLaughlin, S.R. & Leary, J.F. Theoretical basis for sampling statistics useful for detecting and isolating rare cells using flow cytometry and cell sorting. *Cytometry* 27, 233-238 (1997).
99. Izumikawa, M., Murata, M., Tachibana, K., Ebizuka, Y. & Fujii, I. Cloning of modular type I polyketide synthase genes from salinomycin producing strain of *Streptomyces albus*. *Bioorgan Med Chem* 11, 3401-3405 (2003).
100. Sergeeva, A., Kolonin, M.G., Mollrem, J.J., Pasqualini, R. & Arap, W. Display technologies: Application for the discovery of drug and gene delivery agents. *Adv Drug Deliv Rev* 58, 1622-1654 (2006).
101. Green, B.D. & Keller, M. Capturing the uncultivated majority. *Curr Opin in Biotechnol/Env biotechnol/En biotechnol* 17, 236-240 (2006).
102. Huang, J.K., Cui, Y., Chen, C.H., Clampitt, D., Lin, C.T. & Wen, L. Molecular cloning and functional expression of bovine deoxyhypusine hydroxylase cDNA and homologs. *Protein Express Purif* 54, 126-133 (2007).
103. Farrar, K. & Donnison, I.S. Construction and screening of BAC libraries made from *Brachypodium* genomic DNA. *Nat Protocols* 2, 1661-1674 (2007).
104. Ibrahim, S.F. & van den Engh, G. High-speed cell sorting: fundamentals and recent advances. *Curr Opin Biotechnol* 14, 5-12 (2003).

105. Sklar, L.A., Carter, M.B. & Edwards, B.S. Flow cytometry for drug discovery, receptor pharmacology and high-throughput screening. *Curr Opin Pharmacol* 7, 527-534 (2007).
106. Nolan, J.P. & Mandy, F.F. Multiplexed and microparticle-based analyses: Quantitative tools for the large-scale analysis of biological systems. *Cytometry Part A* 69A, 318-325 (2006).
107. Weaver, J.C., Bliss, J.G., Powell, K.T., Harrison, G.I. & Williams, G.B. Rapid clonal growth measurements at the single-cell level: Gel micordroplets and flow cytometry. *Biotechnology* 9, 873-877 (1991).
108. Dupuy, D., Bertin, N., Hidalgo, C.A., Venkatesan, K., Tu, D., Lee, D., Rosenberg, J., Svrzikapa, N., Blanc, A., Carnec, A. et al. Genome-scale analysis of in vivo spatiotemporal promoter activity in *Caenorhabditis elegans*. *Nat Biotechnol* 25, 663-668 (2007).
109. Quinones-Coello, A.T., Petrella, L.N., Ayers, K., Melillo, A., Mazzalupo, S., Hudson, A.M., Wang, S., Castiblanco, C., Buszczak, M., Hoskins, R.A. et al. Exploring strategies for protein trapping in *Drosophila*. *Genetics* 175, 1089-1104 (2007).
110. Miller, R.L., Zhang, P., Chen, T., Rohrwasser, A. & Nelson, R.D. Automated method for the isolation of collecting ducts. *Am J Physiol Renal Physiol* 291, F236-245 (2006).
111. Mampel, J., Spirig, T., Weber, S.S., Haagensen, J.A.J., Molin, S. & Hilbi, H. Planktonic replication is essential for biofilm formation by *Legionella pneumophila* in a complex medium under static and dynamic flow conditions. *Appl Environ Microbiol* 72, 2885-2895 (2006).
112. Bertani, G. Studies on lysogenesis I: The mode of phage liberation by lysogenic *Escherichia coli*. *J Bacteriol* 62, 293-300 (1951).
113. Neidhardt, F.C. *Escherichia coli* and *Salmonella typhimurium*. Cellular and molecular biology, Edn. 1. (American Society for Microbiology, Washington DC; 1987).
114. Hutchison, C.A., III DNA sequencing: bench to bedside and beyond. *Nucl. Acids Res.* 35, 6227-6237 (2007).
115. Lorenz, P. & Eck, J. Metagenomics and industrial applications. *Nat. Rev.* 3, 510-516 (2005).
116. Farrar, K. & Donnison, I. Construction and screening of BAC libraries made from *brachypodium* genomic DNA. *Nat. Protoc.* 2, 1661-1674 (2007).

117. Beja, O. To BAC or not to BAC: marine ecogenomics. *Curr. Opin. Biotechnol.* 15, 187-190 (2004).
118. Dahlroth, S.L., Nordlund, P. & Cornvik, T. Colony filtration blotting for screening soluble expression in *Escherichia coli*. *Nat. Protoc.* 1, 253-258 (2006).
119. Jones, P., Watson, A., Davies, M. & Stubbings, S. Integration of image analysis and robotics into a fully automated colony picking and plate handling system. *Nucleic Acids Res.* 20, 4599-4606 (1992).
120. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* 26, 1135-1145 (2008).
121. Freeman, A., Cohen-Hadar, N., Abramow, S., Modai-Hod, R., Dror, Y. & Georgiou, G. Screening of large protein libraries by the "cell immobilized on adsorbed bead" approach. *Biotechnol. Bioeng.* 86 (2004).
122. Dressman, D., Yan, H., Traverso, G., Kinzler, K.W. & Vogelstein, B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc. Natl. Acad. Sci.* 100, 8817-8822 (2003).
123. Fan, J.B. & Chee, M.S. Highly parallel genomic assays. *Nature Rev.* 7, 632-644 (2006).
124. Lee, Y.-F., Tawfik, D.S. & Griffiths, A.D. Investigating the target recognition of DNA cytosine-5 methyltransferase HhaI by library selection using in vitro compartmentalisation. *Nucl. Acids Res.* 30, 4937-4944 (2002).
125. Miller, O.J., Bernath, K., Agresti, J., Amitai, G., Kelly, B.T., Mastrobattista, E., Taly, V., Magdassi, S., Tawfik, D.S. & Griffiths, A.D. Directed evolution by in vitro compartmentalization. *Nat. Methods* 3, 561-570 (2006).
126. Leamon, J.H., Link, D.R., Egholm, M. & Rothberg, J.M. Overview: methods and applications for droplet compartmentalization of biology. *Nat. Methods* 3, 541-543 (2006).
127. Diehl, F., Li, M., He, Y., Kinzler, K.W., Vogelstein, B. & Dressman, D. BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions. *Nat. Methods* 3, 551-559 (2006).
128. Cowan, D., Meyer, Q., Stafford, W., Muyanga, S., Cameron, R. & Wittwer, P. Metagenomic gene discovery: past, present, future. *Trends Biotechnol.* 23, 321-329 (2005).

129. Mampel, J., Spirig, T., Weber, S.S., Haagensen, J.A.J., Molin, S. & Hilbi, H. Planktonic replication is essential for biofilm formation by legionella pneumophila in a complex medium under static and dynamic flow conditions. *Appl. Environ. Microbiol.* 72, 2885-2895 (2006).
130. Walser, M., Leibundgut, R., Pellaux, R., Panke, S. & Held, M. Isolation of monoclonal microcarriers colonized by fluorescent *E. coli*. *Cytometry Part A* 73A, 788-798 (2008).
131. Allen, G.C., Flores-Vergara, M.A., Krasnyanski, S., Kumar, S. & Thompson, W.F. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* 1, 2320-2325 (2006).
132. Anderson, J.A., Churchill, G.A., Autrique, J.E., Tanksley, S.D. & Sorrells, M.E. Optimizing parental selection for genetic linkage maps. *Genome* 36, 181-186 (1993).
133. Brandenberger, H. & Widmer, F. Immobilization of highly concentrated cell suspensions using the laminar jet breakup technique *Biotechnol Prog.* 15, 366-372 (1999).
134. Okogbenin, E., Marin, J. & Fregene, M. An SSR-based molecular genetic map of cassava. *Euphytica* 147, 433-440 (2006).
135. Squirrell, J., Hollingsworth, P.M., Woodhead, M., Russell, J., Lowe, A.J., Gibby, M. & Powell, W. How much effort is required to isolate nuclear microsatellites from plants? *Mol. Ecol.* 12, 1339-1348 (2003).
136. Terai, G., Komori, T., Asai, K. & Kin, T. miRRim: A novel system to find conserved miRNAs with high sensitivity and specificity. *RNA* 13, 2081-2090 (2007).
137. Ohler, U.W.E., Yekta, S., Lim, L.P., Bartel, D.P. & Burge, C.B. Patterns of flanking sequence conservation and a characteristic upstream motif for microRNA gene identification. *RNA* 10, 1309-1322 (2004).
138. Eijsink, V.G.H., Vaaje-Kolstad, G., Varum, K.M. & Horn, S.J. Towards new enzymes for biofuels: lessons from chitinase research. *Trends Biotechnol.* 26, 228-235 (2008).
139. Taylor, S.V., Kast, P. & Hilvert, D. Investigating and engineering enzymes by genetic selection. *Angew. Chem. Int. Ed. Engl.* 40, 3310-3335 (2001).
140. Lynd, L.R., Laser, M.S., Bransby, D., Dale, B.E., Davison, B., Hamilton, R., Himmel, M.E., Keller, M., McMillan, J.D., Sheehan, J. et al. How biotech can transform biofuels. *Nat. Biotechnol.* 26, 169-172 (2008).

141. O'Connor, J.E. & al., e. The relevance of flow cytometry for biochemical analysis. *IUBMB Life* 51, 231-239 (2001).
142. Zengler, K., Toledo, G., Rappe, M., Elkins, J., Mathur, E.J., Short, J.M. & Keller, M. Cultivating the uncultured. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15681-15686 (2002).
143. Nir, R., Yisraeli, Y., Lamed, R. & Sahar, E. Cytometry sorting of viable bacteria and yeasts according to beta-galactosidase activity. *App. Env. Microbiol.* 56, 3861-3856 (1990).
144. Powell, K.T. & Weaver, J.C. Gel microdroplets and flow cytometry: Rapid determination of antibody secretion by individual cells within a cell population. *Biotechnology* 8, 333-337 (1990).
145. Haas, A., Brehm, K., Kreft, J. & Goebel, W. Cloning, characterization, and expression in *Escherichia coli* of a gene encoding *Listeria seeligeri* catalase, a bacterial enzyme highly homologous to mammalian catalases. *J. Bacteriol.* 173, 5159-5167 (1991).
146. Switala, J. & Loewen, P.C. Diversity of properties among catalases. *Arch. Biochem. Biophys.* 401, 145-154 (2002).
147. Neuenschwander, M., Butz, M., Heintz, C., Kast, P. & Hilvert, D. A simple selection strategy for evolving highly efficient enzymes. *Nat. Biotechnol.* 25, 1145-1147 (2007).
148. Dykxhoorn, D.M., St. Pierre, R. & Linn, T. A set of compatible tac promoter expression vectors. *Gene* 177, 133-136 (1996).
149. Blonk, J.C.G., van Eendenburg, J., Koning, M.M.G., Weisenborn, P.C.M. & Winkel, C. A new CSLM-based method for determination of the phase behaviour of aqueous mixtures of biopolymers. *Carbohydr. Polym.* 28, 287-295 (1995).
150. Voss, H., Schwager, C., Wiemann, S., Zimmermann, J., Stegemann, J., Erfle, H., Voie, A.-M., Drzonek, H. & Ansorge, W. Efficient low redundancy large-scale DNA sequencing at EMBL. *J. Biotechnol.* 41, 121-129 (1995).
151. Metzker, M.L. Emerging technologies in DNA sequencing. *Genome Res.* 15, 1767-1776 (2005).
152. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A. et al. The sequence of the human genome. *Science* 291, 1304-1351 (2001).

153. Dalin, E. & Detter, C. Doe Joint Genome Institute; 3 Kb WG library creation protocol. [http://www.jgi.doe.gov/sequencing/protocols/protos\\_production.html](http://www.jgi.doe.gov/sequencing/protocols/protos_production.html) Version 7.8, 1-12 (2009).
154. DoE DoE Joint Genome Institute; Sequencing statistics; <http://www.jgi.doe.gov/sequencing/statistics.html>. (2010).
155. Walser, M., Pellaux, R., Meyer, A., Bechtold, M., Vanderschuren, H., Reinhardt, R., Magyar, J., Panke, S. & Held, M. Novel method for high-throughput colony PCR screening in nanoliter-reactors. *Nucleic Acids Res.* 37, e57 (2009).
156. Wadowsky, R.M., Laus, S., Libert, T., States, S.J. & Ehrlich, G.D. Inhibition of PCR-based assay for *Bordetella pertussis* by using calcium alginate fiber and aluminum shaft components of a nasopharyngeal swab. *J. Clin. Microbiol.* 32, 1054-1057 (1994).
157. Pluen, A., Netti, P.A., Jain, R.K. & Berk, D.A. Diffusion of macromolecules in agarose gels: comparison of linear and globular configurations. *Biophys. J.* 77, 542-552 (1999).
158. Mitra, R. & Church, G. In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res.* 27, e34- (1999).
159. Samatov, T.R., Chetverina, H.V. & Chetverin, A.B. Expressible molecular colonies. *Nucleic Acids Res.* 33, e145- (2005).



## Curriculum Vitae

### Marcel Walser

Born: 14. November 1976, in Zürich Switzerland  
Citizen: Rehetobel (AR), Switzerland  
Contact: marcel@segelklinik.ch

### Education

1984-1990 Primary school, Samstagern, Switzerland  
1990-1997 Gymnasium Stiftschule Einsiedeln, Switzerland  
1997-2003 Studies in Biology, ETH Zurich, Switzerland (MSc)  
2002-2003 Diploma thesis, Plant Biotechnology, ETH Zurich, Switzerland  
2003-2004 Scientific collaborator, Plant pathology group, ETH Zurich, Switzerland  
2004-2010 PhD student in the group of Prof. Dr. Sven Panke, ETH Zurich, Switzerland

### Publications – first author

1. Patocchi, A., Walser, M., Tartarini, S., Broggin, G.A.L., Gennari, F., Sansavini, S. and Gessler, C. (2005) Identification by genome scanning approach (GSA) of a microsatellite tightly associated with the apple scab resistance gene *Vm*. *Genome*, **48**, 630-636.
2. Walser, M., Leibundgut, R.M., Pellaux, R., Panke, S. and Held, M. (2008) Isolation of monoclonal microcarriers colonized by fluorescent *E. coli*. *Cytometry Part A*, **73A**, 788-798.
3. Walser, M., Pellaux, P., Meyer, A., Bechtold, M., Vanderschuren, H., Reinhardt, R., Magyar, J., Panke, S. and Held, M. (2009) Novel method for high-throughput colony PCR screening in nanoliter-reactors. *Nucl. Acids Res.*, gkp160.

### Publications – co-author

1. Hennig, L., Taranto, P., Walser, M., Schonrock, N. and Grissem, W. (2003) Arabidopsis MSI1 is required for epigenetic maintenance of reproductive development. *Development*, **130**, 2555-2565.
2. Silfverberg-Dilworth, E., Matasci, C., Van de Weg, W., Van Kaauwen, M., Walser, M., Kodde, L., Soglio, V., Gianfranceschi, L., Durel, C., Costa, F. *et al.* (2006) Microsatellite markers spanning the apple (*Malus x domestica* Borkh.) genome. *Tree Genetics & Genomes*, **2**, 202-224.

## **Personal Acknowledgements to**

Martin, for all your endurance, passion and inventing spirit during all the projects from the beginning till the end and beyond – really admirable!

Sven, for all your help in writing and fighting problems during the entire thesis whenever I needed it!

Rene and Andy, not only for your never lasting technical support but as well your personal input!

Sonja, for endless talking about science and “the life outside” to boost my motivation!

Herve, Matthias Be. and Reto, for your support in individual chapters of this thesis!

Richard and Willi, for expertises in individual chapters and for being co-referees during my PhD exams

Fabian, for all your input to potential use of developed platforms in combination with yeast!

Mareike, for extraordinary transformations late at night ;)

The entire BPL-group, the D-MAVT and the D-BSSE! I do not want to name them all in fear that I might forget somebody!

All my family, Linda, Dario, Claus, Markus, Simon and Caspar for being with me at really all situations in my life – without you, this would not have happened – love you all sooo much!!!