JACOBS UNIVERSITY

Flow cytometry based screening systems for directed evolution of proteases

by

Ran Tu

A thesis submitted in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy

in

Biochemical Engineering

Approved, Thesis Committee

_____

Prof. Dr. Ulrich Schwaneberg, RWTH Aachen University

_____

Prof. Dr. Danilo Roccatano, Jacobs University Bremen

_____

Prof. Dr. Anett Kirschner, RWTH Aachen University

Date of Defense: July 26, 2010

School of Engineering and Science

Acknowledgement

Time always flies faster than I expect and now it has already been three years since I came to Jacobs University Bremen. I do enjoy the life here and did learn a lot from past years. During these three years life here, I have got a lot of help from so many people.

Firstly I would like to give my sincerely thanks to Prof. Dr. Ulrich Schwaneberg, my supervisor, for giving me the opportunity to work in his lab. I admire his thoughts to science and his dedication to his work. I thank him for his help and support not only in science research but also in my personal career. I always think I am very lucky to be here and find out the interesting research I want to do and further decide the future work I will do.

Here I acknowledge BMBF for financial support and thank our Dean Prof. Dr. Bernhard Kramer for allowing me to start and accomplish my study at JUB. I would also like to thank Prof. Dr. Danilo Roccatano and Prof. Dr. Anett Kirschner for serving in my thesis committee. I appreciate their time to read this thesis and Prof. Kirschner for making her trip from Aachen to Bremen.

During this work I got a lot of support from Dr. Jürgen Eck and Dr. Frank Niehaus who are from our collaborator B.R.A.I.N. Company, especially Dr. Frank Niehaus for his advices and friendly help with materials and methods used in the project. I have to also thank Dr. Danny Köhler from Partec Company for technical support of CyFlow space flow cytometry, Dr. Bernhard Fuchs and Jörg Wulf for their help to use flow cytometry in MPI.

I would like to thank our HTS subgroup colleagues: Dr. Prodanovic, Milan and Raluca. I really got a lot help from them in this challenge project. I thank them for their valuable discussion. I also want to thank Ronny for his help for protease related work and modeling work. I thank Kanglan and Liqing for their help at the beginning of the project, especially Kanglan for her time to show me how to start directed evolution experiments and discuss work with me later on. I have had the pleasure to work with our lab rotation students Sinan and Marvin; they are bright and diligent to their work.

I would like to thank our team assistant Marina and Daniela for their outstanding work to ensure and help us with lab work. Special thanks will also go to my colleagues in Schwaneberg group, too many names to list here. They make a very nice communication atmosphere for the research and life in the lab. I will remember all the good time we had and thank them for all the help. Here I would like to especially thank Mathias for his kindly help to review and revise my publication and thesis. Thank him for his time and valuable suggestions.

I am happy to know lots of friends here from my country-China. They certainly make my life much easier and comfortable to live here. I especially want to thank my roommate and friend Yuan. I enjoy the time with her and sometimes having a short trip with her in European do help me to relax myself out of some stress work.

I would like to express my deepest thanks to my parents who always support me over years without any complaints. Bear with me for all of these years for my curiosity of science and endless learning. Without their support and love, I couldn't be here and find out the research work I like. For them, I couldn't ever repay their love and care and the only thing what I can do is to keep going and to be a good person who worth the love.

At last, I would like to end up this acknowledgement with my sincere gratitude to all the people who helped me during the study. Thank you all for your support and help. I will stay curiously for science and never stop improving myself whenever and wherever I am for all of you.

Abstract

Proteases are industrially important enzymes and advances in screening technologies such as flow cytometry based *in vitro* compartmentalization (IVC) offer novel opportunities to tailor protease properties by directed evolution. In the project two flow cytometry based screening systems (cell-free based ProCF-FCIVC and whole-cell based Pro-FCIVC) with a high throughput (>7500 s^{-1}) have been developed for directed evolution of proteases. The development was performed in three aspects. Firstly, system related technologies such as generation of w/o/w double emulsions, selection of a suitable substrate and calibration of the sorting performance of flow cytometry were investigated and optimized. Results showed that an ABIL EM90/CMC system could provide homogenous and stable emulsions, and therefore was used as a means for emulsification. By analyzing enzymatic reactions in emulsions, R110 based substrates were proved to be suitable for the double emulsion IVC system. Sorting calibration of flow cytometry was set up in a routine work by three steps starting from calibration beads to double emulsion samples. Then a cell-free based ProCF-FCIVC was developed using the model protease M57, a serine protease isolated from a metagenome library. Different parameters like *in vitro* expression using various vector sources and expression reaction mixtures, enzymatic reactions in emulsions with altered pH and increased substrate concentration as well as recovery of sorted DNA variants were investigated. Sorting and subsequent activity analysis of the M57 reference library demonstrated a proof of concept of the developed ProCF-FCIVC system. For the whole-cell based Pro-FCIVC system, an extracellular protease deficient *Bacillus* strain together with a Subtilisin Carlsberg (SC) secretion system were chosen due to its low background activity of other non-target proteases. Several main challenges such as background of natural occurring proteases, high diversity library generation of *Bacillus*, sorting strategies and growth competition effects on sorting enrichment were studied and optimized. The developed whole-cell Pro-FCIVC system was then verified by screening an epPCR library with a high mutational load (~5 mutations per 1 kb) for inhibition resistance of SC towards the antipain dihydrochloride inhibitor. One variant (SCm2) with a 1.6-fold relative to wild-type SC improved resistance was isolated from a small population of active variants (315) validating the developed protease flow cytometry screening technology for increased inhibitor resistance. The reported two flow cytometry IVC screening systems can due to the sensitivity likely be modified for other proteases and applied for additional properties for directed evolution of proteases.

Abbreviation

AAPF	Ala-Ala-Pro-Phe peptides
Ap	Ampicillin
AU	Absolute unit
BSA	Bovine serum albumin
B. subtilis	*Bacillus subtilis*
Cm	Chloramphenicol
DMSO	Dimethyl sulfoxide
E. coli	*Escherichia coli*
epPCR	Error-prone PCR
FC	Flow cytometry
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IVC	*In vitro* compartmentalization
IVT	*In vitro* translation
IVTT	*In vitro* translation in tube
IVTE	*In vitro* translation in emulsion
LP	Linear polyacrylamide
Km	Kanamycin
MTP	Microtiter plate
Nm	Neomycin
OD	Optical density
PBS	Phosphate buffered saline
pI	Isoelectric point
R110	Rhodamine 110
RCA	Rolling circle amplification
SC	Subtilisin Carlsberg
SDM	Site directed mutagenesis
TEA	Triethylamine
w/o/w	Water-in-oil-in-water

Contents

1 Introduction

1.1 Proteases and their applications

Protease including proteinases and peptidases belong to a class of enzyme which breaks peptide bonds and catalyzes the hydrolytic breakdown of proteins. They are essential constituents of all forms of life on earth, including prokaryotes, fungi, plants and animals (Gupta, Beg et al. 2002). Proteases could be classified by where they cleave a peptide bond within their substrates or by the mechanism of their catalysis. For example, an endopeptidase cleaves peptide bonds within a peptide chain while an exopeptidase cleaves peptide bonds at the end of a peptide chain. Based on their catalytic mechanism proteases can be divided into four subgroups: 1) serine proteases, 2) aspartic proteases, 3) cysteine proteases, and 4) metallo proteases (Rao, Tanksale et al. 1998). Almost one-third of all proteases can be classified as serine proteases (named from serine residue at active site for nucleophilic attack of peptide bond). Mechanism of serine proteases normally consists of an Asp-His-Ser catalytic triad system. By the differences of structural contexts serine proteases can further be subdivided into about six clans. The Asp-His-Ser could be found in four clans: chymotrypsin (SA), subtilisin (SB), carboxypeptidase Y (SC) and Clp protease (SK) (Hedstrom 2002). Figure 1.1 shows the Asp-His-Ser “catalytic triad” in a typical serine protease subtilisin (Branden and Tooze 1998). Protease candidates, M46, M57 and Subtilisin Carlsberg, used in this project were obtained from collaborator B.R.A.I.N. AG and belong to the family of serine proteases.

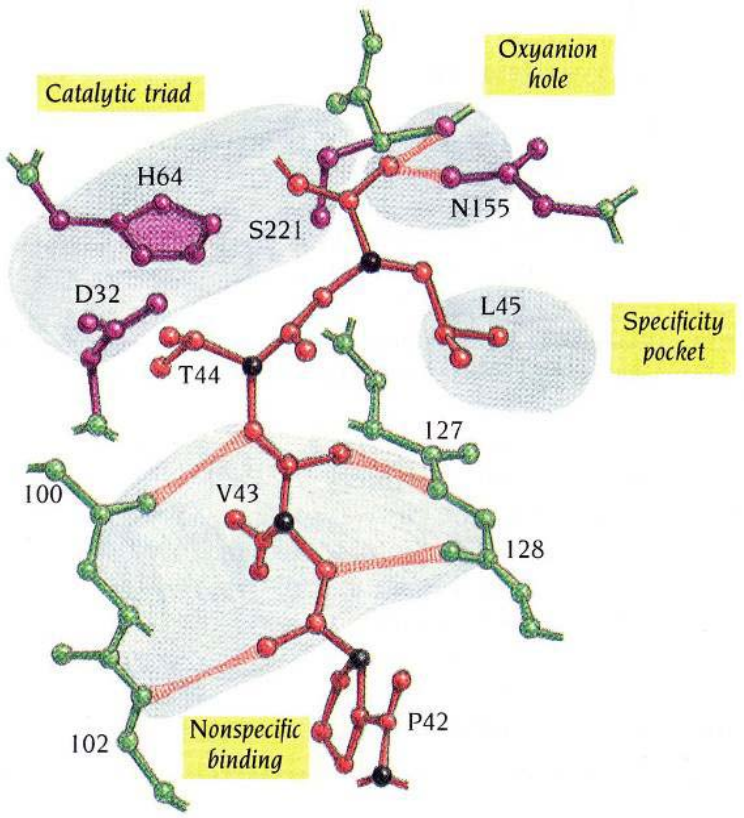


Figure 1.1 Schematic diagram of the active site of subtilisin. A region of a bound polypeptide inhibitor is shown in red. The four essential features of the active site are highlighted in yellow (Branden and Tooze 1998).

In serine proteases the largest subgroups are serine alkaline proteases and subtilisins. Serine alkaline proteases usually have an optimal pH around 10, a molecular weight in a range of 15 to 30 kDa and an isoelectric point (pI) near 9 (Gupta, Beg et al. 2002). Although serine alkaline proteases are produced by several bacteria such as *Arthrobacter*, *Streptomyces*, and *Flavobacterium* spp., the best known ones are subtilisins produced by *Bacillus* species such as *amyloliquefaciens* (BPN´), *subtilis* (subtilisin E), *lentus* (savinase) and *licheniformis* (carlsberg) (Rao, Tanksale et al. 1998). Due to their high activities and stabilities in alkaline pH range, serine alkaline proteases are widely used in detergent industry (Gupta, Beg et al. 2002). In total enzyme market proteases account for approximately 40 % of enzyme sales and the largest distribution is dominated by detergent alkaline proteases (Cherry and Fidantsef 2003). Other applications of serine alkaline proteases including meat and soy processing, dairy goods processing, leather

treatment, wastewater treatment and organic synthesis have been summarized in reviews (Rao, Tanksale et al. 1998; Gupta, Beg et al. 2002). Because of their importance to the industry as well as models for understanding the enormous rate enhancements affected by enzymes, serine alkaline proteases especially subtilisins have became model systems for protein engineering studies and have been well described in several reviews (Bryan 2000; Gupta, Beg et al. 2002; Hedstrom 2002; Cherry and Fidantsef 2003; Maurer 2004; Saeki, Ozaki et al. 2007).

1.2 Directed evolution

Many industrial enzymes for improving enzyme properties often need to be tuned to suit desired properties for commercial applications. Directed evolution, also called molecular evolution, is a particularly powerful algorithm to achieve these goals. As shown in Figure 1.2, directed evolution comprises two main steps. The first step is genetic diversity generation for mutant libraries, and the second important step is isolation of the desired mutants from the libraries using a proper screening assay system.

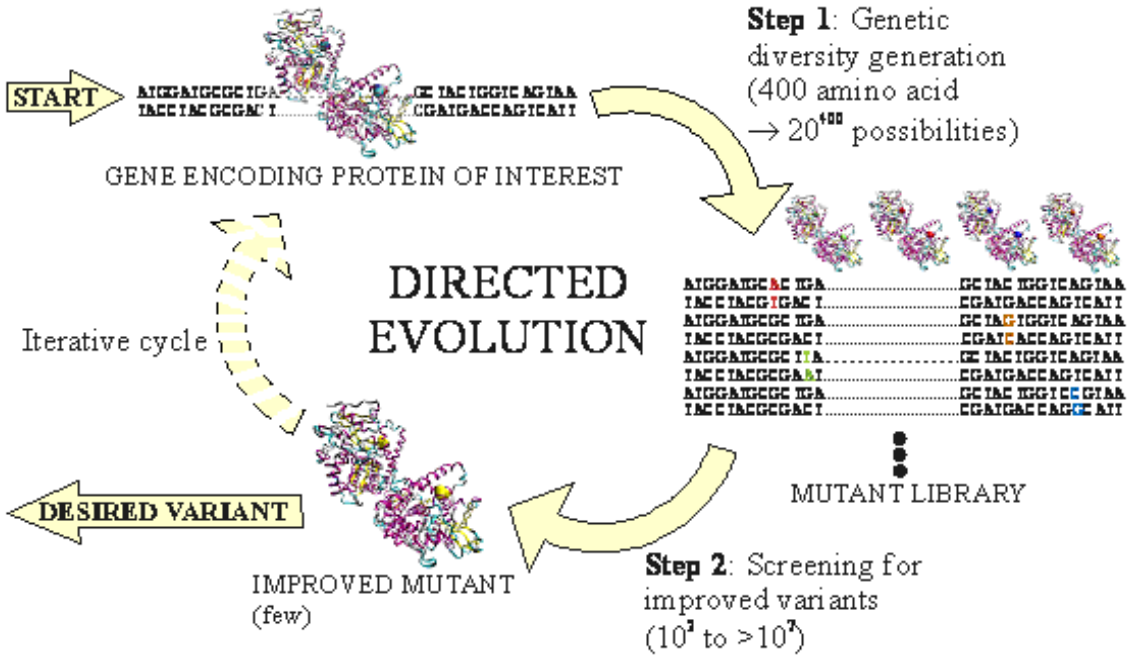


Figure 1.2 The main steps in a directed evolution experiment (Tee and Schwaneberg 2007).

For diversity generation of mutant libraries, mutations are usually introduced in the gene level and these nucleotide exchanges further results on the protein level via amino acid substitutions. Figure 1.3 shows four common used random mutagenesis strategies to for library generation including oligonucleotide-directed mutagenesis (Hayashi, Welschof et al. 1994), error-prone polymerase chain reaction (epPCR) (Cadwell and Joyce 1994), DNA shuffling (Stemmer 1994), and Sequence Saturation Mutagenesis (SeSaM) (Wong, Tee et al. 2004). Although having a high transition favored demerit, epPCR, a method based on reducing fidelity of employed DNA polymerases, is still the most commonly used one due to its technically simple and robust as well as tunable mutation frequency by simply varying concentration of manganese ions (Tee and Schwaneberg 2007). Nineteen various random mutagenesis methods were summarized and comprehensively discussed for their genetic diversity pros and cons in a review (Wong, Zhurina et al. 2006).

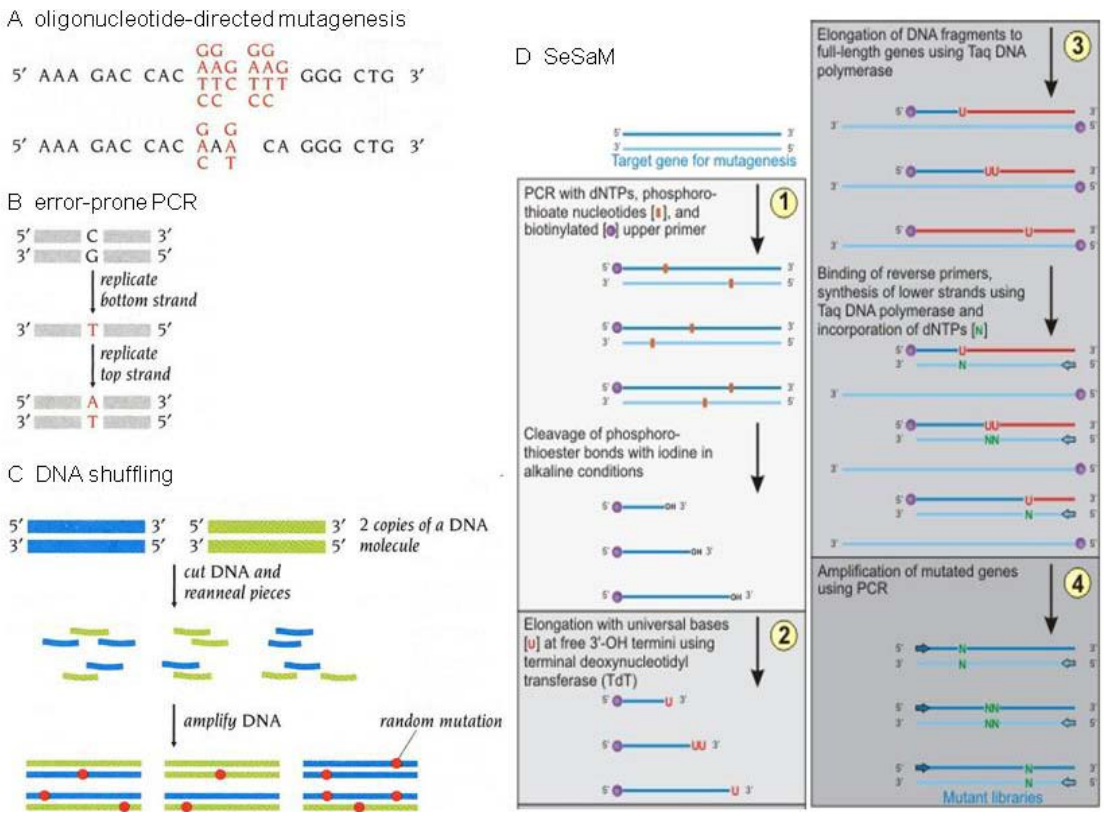


Figure 1.3 Methods of random mutagenesis. A: Oligonucleotide-directed mutagenesis. Oligonucleotides can be synthesized to contain mixtures of nucleotides at specific codons, B: error-prone polymerase chain reaction (epPCR), C: DNA shuffling, D: Sequence Saturation Mutagenesis (SeSaM) (Branden and Tooze 1998; Wong, Tee et al. 2004; Wong, Zhurina et al. 2006).

Besides random mutagenesis on a single gene level, metagenome library could be another alternative source. Since up to 99.8 % of the microbes present in many environments cannot be cultured with currently available technologies, metagenome employed library shows an import source to find 'novel' candidates from nature with desired functions (Jaeger 2004; Streit, Daniel et al. 2004; Lorenz and Eck 2005).

In many cases the second step in directed evolution, screening for the desired mutants, is the most challenging step for finding improved variants. An effective screening strategy usually is the key to accomplish directed evolution task. There are several principles to judge a good screening system. The best one should have a high throughput to increase the probability of finding desired clones and at the mean time be sufficiently sensitive to allow the isolation of lower activity clones showed in early stage of evolution process (Arnold and Georgiou 2003).

A well known traditional enzyme screening method is phenotype evolved selection systems which is based on complementation of auxotroph or resistance to cytotoxic agents. Phenotypic selection provides a powerful means for screening large libraries and isolating interesting protein mutants. Phenotype selection based system has been used for decades in a diverse cases to obtain useful genetic information such as protein function and protein-protein relationships (Oue, Okamoto et al. 1999). The principles used for phenotypic selection is to set a suitable complementation system. However, in most cases the isolated individual cannot be directly linked to a selectable phenotype because the selective pressure often results some mutations which do not affect the catalytic activity. For example instead of mutating the target enzyme, selections often yield mutations with a increased expression level of a poor catalyst, a different bypass pathway which is unrelated to the enzyme of interest, or a general stress response that is capable of conferring growth under stringent selective conditions (Olsen, Iverson et al. 2000).

Another high throughput protein library screening tool is phage display. Since 1985 when George Smith first demonstrated that peptide-phage fusions could be selected through the binding of the peptide to an antibody immobilized on a plate, the phage display has been used in a wide field such as improvement of affinity and specificity of antibodies and

antigens, interactions between proteins and small molecules (Smith 1985; Fernandez-Gacio, Uguen et al. 2003). However, in diverse phage display and variations, libraries are screened on the basis of binding properties and not on activity assays.

In contrast in microtiter plate assays individual clones/candidates could be arrayed spatially in the plate and their enzymatic reactions or characterizations could be subsequently carried out to screen for the variants with enzyme activities related properties. Despite of its lower throughput compared to the 'smart' phenotype selection systems and phage display based screening systems, microtiter plate is a widely used format in directed evolution of enzymes especially in a complex reaction case. Figure 1.4 illustrates a schematic diagram of microtiter plate assay.

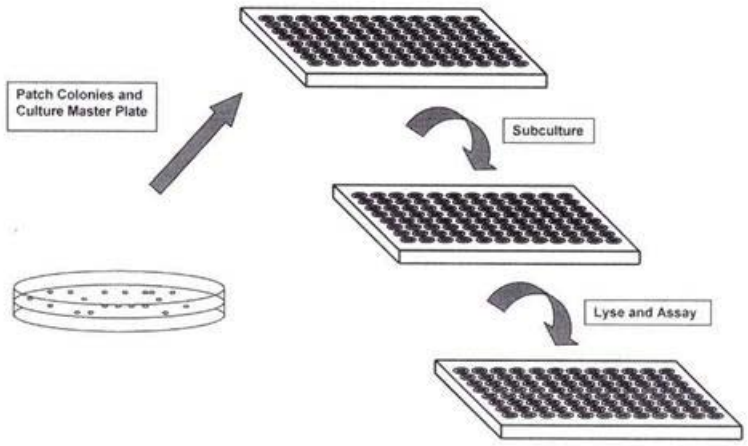


Figure 1.4 Schematic diagram of a typical microtiter plate assay (Arnold and Georgiou 2003).

Although microtiter plate assays usually are relatively time-consuming and often not very high throughput, they still represent the most effective means for evolving enzymes of industrial interest. Microtiter plate based assays and strategies for adapting several important classes of enzymes for the engineered physical properties were comprehensively described by Arnold and coworkers in details (Arnold and Georgiou 2003). With the combination of robotic automation, liquid handling systems, and plate reader technologies, throughput of microtiter plate screening systems can be increased substantially. However a significant cost of resources and energy is also risen up. In contrast the latest developments in flow cytometry techniques allow screening of large

libraries in high throughput with a rather low cost of screening assay (Yang and Withers 2009).

1.3 Flow cytometry based high throughput screening strategies

Flow cytometry is a technology that simultaneously measures and analyzes multiple physical characteristics of single particles by passing them in flow chamber and analyzing them by electronic detectors. The properties measured include a particles's relative size, relative granularity or internal complexity and relative fluorescence intensity. These characteristics are determined using an optical to electronic coupling system that records how the cell or particle scatters incident laser light and emits fluorescence (Manual from BD Biosciences Company).

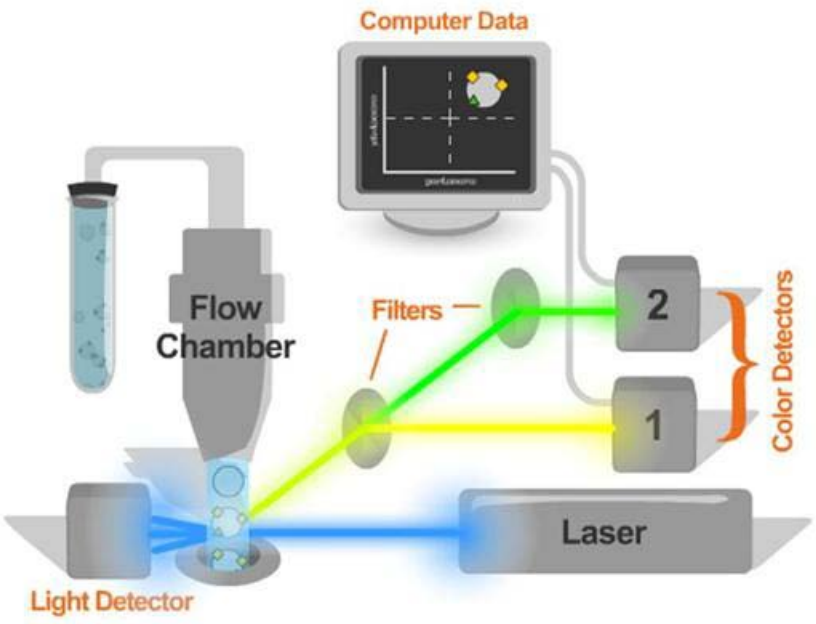


Figure 1.5 Scheme of work principle of flow cytometry flow cytometry (http://oregonstate.edu).

As shown in Figure 1.5, a flow cytometer is made up of three main systems fluidics, optics and electronics. The principle is well described by BD Biosciences Company flow cytometer Manual as cited as follows. The fluidics system transports particles in a stream to the laser beam for interrogation. The optics system consists of lasers to illuminate the

particles in the sample stream and optical filters to direct the resulting light signals to the appropriate detectors. The electronics system converts the detected light signals into electronic signals that can be processed by the computer. When particles pass through the laser intercept, they scatter laser light. If the particles have fluorophore with comparable wavelengths of laser light, the particles will be excited and present fluorescence. The scattered and fluorescent light signals are explained as follows. Forward-scattered light (FSC) is proportional to cell surface area or size. FSC measures diffracted light and it is detected off axis of the incident laser beam in the forward direction. Side-scattered light (SSC) is proportional to cell granularity or internal complexity. SSC measures refracted and reflected light that occurs at any interface within the cell where there is a change in refractive index and it is collected at approximately 90 degrees to the laser beam by a collection lens and then redirected by a beam splitter to the appropriate detector. For fluorescence signal (FL), the amount of fluorescence signal detected is proportional to the number of fluorochrome molecules on the molecules on the particle. The collected parameters (FSC, SSC, FL) can be displayed in several formats such as a single parameter histogram, where the horizontal axis represents the parameter's signal value in channel numbers and the vertical axis represents the number of events per channel number, while two parameters can be displayed by a dot plot. To sort interested particles, firstly a sorting region needs to be gated to identify the particles of interest from others. The gating regions can be set either in one parameter histogram or two parameter dot plot, displayed as RN or R in CyFlow® space respectively. Then the sort gate identifies particles and sorts them out of the stream using different sorting modules for example piezo activator sorting module of CyFlow® space flow cytometer used in this study.

Compared with other screening technologies used in directed evolution such as GC/LC-MS, NMR, HPLC (10^2-10^4), microtiter plate (10^3-10^5), solid phase (10^4-10^6) and phage display (>10^7) per day (Tee and Schwaneberg 2007), flow cytometry based screening systems can analyze and sort up to 10^7 variants per hour (Mastrobattista, Taly et al. 2005). This powerful throughput offers flow cytometry technology various applications for directed evolution of enzymes. Flow cytometry based screening systems have been

used to screen libraries of proteins displayed on bacterial, yeast, or mammalian cells and have yield highly useful tailor-made proteins (Mattanovich and Borth 2006).

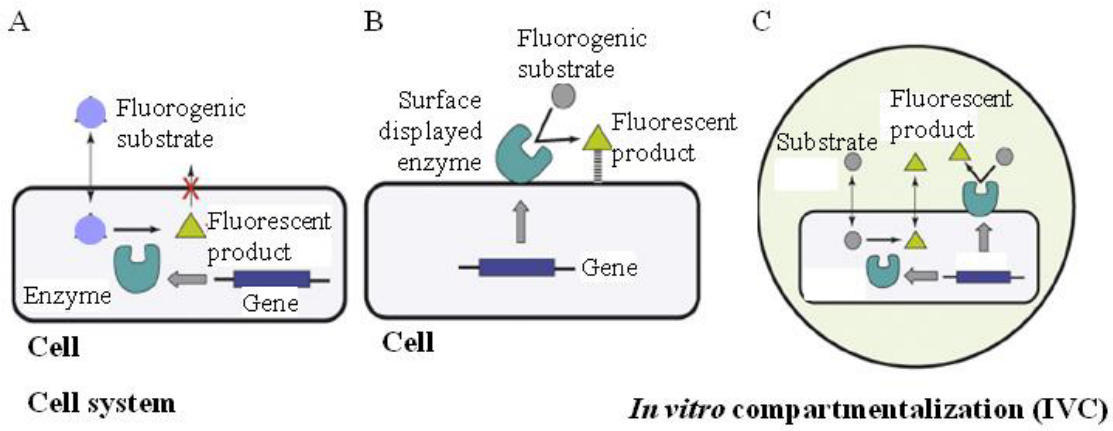


Figure 1.6 Screening enzyme libraries by flow cytometry (Bershtein and Tawfik 2008).

Flow cytometry based systems has been adapted in several ways which can be divided into three screening modes as shown in Figure 1.6: a) entrapment of product within the enzyme expressing cells (Griswold, Kawarasaki et al. 2005; Aharoni, Thieme et al. 2006; Griswold, Aiyappan et al. 2006), b) entrapment of product onto the surface of cells displaying target enzymes (Becker, Schmoldt et al. 2004; Varadarajan, Gam et al. 2005; Varadarajan, Rodriguez et al. 2008), and c) entrapment of fluorescent product in an emulsion droplet *in vitro* compartmentalization carrying the enzyme and its encoding gene (Griffiths and Tawfik 2003; Aharoni, Amitai et al. 2005; Mastrobattista, Taly et al. 2005). Applications of various flow cytometry based enzyme screening systems have been recently reviewed (Yang and Withers 2009).

As a principle in all evolutionary systems, a link between genotype (gene of target enzyme that can be replicated) and phenotype (a functional trait such as binding or catalytic activity) is also required in flow cytometry based screening systems. In nature the link is performed in cellular environment. In the laboratory evolution a variety of techniques provided a physical (mainly depending of binding ability) of genotype and phenotype have been developed including diverse display technologies such as phage display, plasmid display, ribosome display, CIS display, mRNA-peptide fusion as well as cell surface display (Taly, Kelly et al. 2007). Despite of successes of generating binding

proteins, such methods have the following limitations: 1) it relies on living cells to compartmentalize the gene library and display the selected proteins; 2) selection is primarily through binding interactions and couldn't efficiently meet the requirement of activity selection of enzyme.

In contrast man-made cell-like compartments, named *in vitro* compartmentalization (IVC) can be used as a man-made cell compartmentalization in which enzymatic reaction can be performed individually without cross-talking between other compartments. Furthermore the size of compartmentalization can be easily adjusted as small as a bacterium with a diameter around 2-3 μm resulting a dispersal of approximately 10^{10} individual microreactors by using only 50 μL of reaction mixture (Miller, Bernath et al. 2006). This allows screening performs in femtoliter volume scale which can significantly reduce the consumable cost. Since the first demonstration of using emulsion droplets to link genotype and phenotype for selection of large gene repertories, many applications of IVC in molecular evolution have been developed for selection of large gene libraries ($>10^9$), isolation of proteins or RNAs with binding, enzymatic or regulatory functions purposes (Tawfik and Griffiths 1998; Griffiths and Tawfik 2006). In these investigations processes, various means of modifications of IVC emulsions were also developed. Figure 1.7 shows 1) temperature can be altered from 4°C up to 94°C; 2) hydrophobic molecules (such as substrates or ligands) can be delivered through hydrophobic oil phase to the inner aqueous phase; 3) water-soluble components can be delivered into inner droplets by nanodroplets or swollen micelles; 4) pH of emulsions can be altered for example by delivering acetic acid; and 5) Photocaged and irradiated of substrates or ligands can be carried out in emulsions (Griffiths and Tawfik 2006).

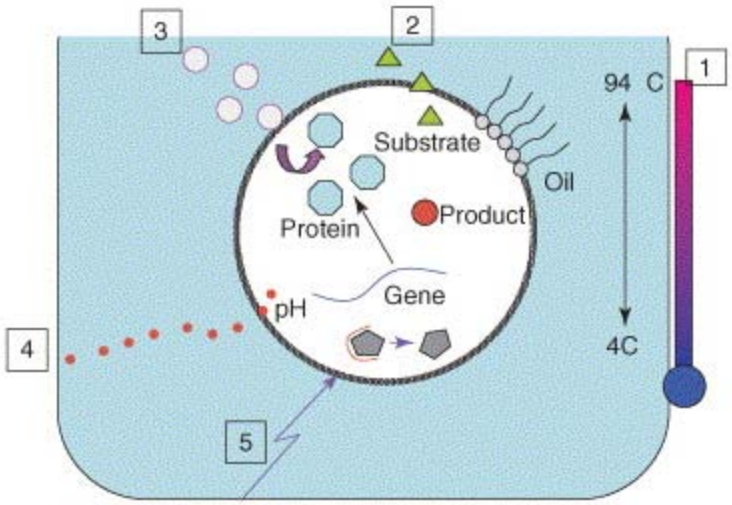


Figure 1.7 Physical and chemical manipulation of the content of emulsion droplets (Griffiths and Tawfik 2006).

IVC is based on water-in-oil (w/o) emulsions where the water phase (inner aqueous droplet) is dispersed in the oil phase to form a w/o single emulsion (Aharoni, Griffiths et al. 2005). However the w/o single emulsion with a continuous external oil phase is not compatible with flow cytometry while w/o/w double emulsions allow to generate an external aqueous phase without alternation of the inner aqueous droplets imbedded in the primary w/o emulsion and make it amenable to be analyzed by flow cytometry (Bernath, Magdassi et al. 2004). Principle of IVC based flow cytometry screening systems employs a fluorescence detection of fluorescent product from fluorogenic substrate (non fluorescence before the conversation of substrate to product). The general procedure for flow cytometry based IVC screening systems can be performed as follows: 1) preparation of gene library for screening, 2) compartmentalization of target enzyme (expressed using cells or cell-free *in vitro* transcription/translation solution) with fluorogenic substrate in w/o primary emulsion, 3) preparation of secondary aqueous phase to generate w/o/w double emulsions and incubation of double emulsions to carry out enzymatic reactions (or incubation first to perform enzymatic reactions and then prepare secondary w/o/w emulsions), 4) flow cytometry sorting of fluorescent droplets which represent a accomplished reaction by active enzyme, 5) recovery of sorted variants for further experiments or iterative sorting rounds. Due to small volume scale of inner aqueous

phase of the w/o/w emulsion droplets, compartmentalized enzyme can be presented in a very high concentration. Therefore IVC based flow cytometry system could be a very sensitive screening means due to a sensitive activity detection of low amount of enzymes and a strong intensity of fluorescence signal. Progress in w/o single emulsion or w/o/w double emulsion IVC technologies have been summarized well in reviews (Griffiths and Tawfik 2006; Bershtein and Tawfik 2008; Yang and Withers 2009).

Figure 1.8 shows two applications of w/o/w double emulsion IVC employed flow cytometry screening systems for directed evolution of paraoxonase (Aharoni, Amitai et al. 2005) and beta-galactosidase (Mastrobattista, Taly et al. 2005).

In whole-cell case (serum paraoxonase PON1, Figure 1.8 A) two different sorting strategies were applied in the sorting. One used a high stringency and single round of selection. Another used a lower stringency and several rounds of sorting. For both sorting strategies mutants exhibiting 20-fold higher activity rates relative to wild-tpye PNO1 were obtained (Aharoni, Amitai et al. 2005).

For cell-free system (Figure 1.8 B) the demonstration was done by evolving Ebg, a protein of unknown function with a possibility to be evolved for β-galactorsidase showed in previous studies. Two mutant libraries of Ebg with high mutational loads (9-12 mutations per gene) were screened and selected. Positive events from each round of sorting were collected and the DNA was recovered by PCR amplification for next round of sorting or analysis. One best variant showing a kcat/Km value 300-fold higher than wild-type Ebg was isolated in the research (Mastrobattista, Taly et al. 2005).

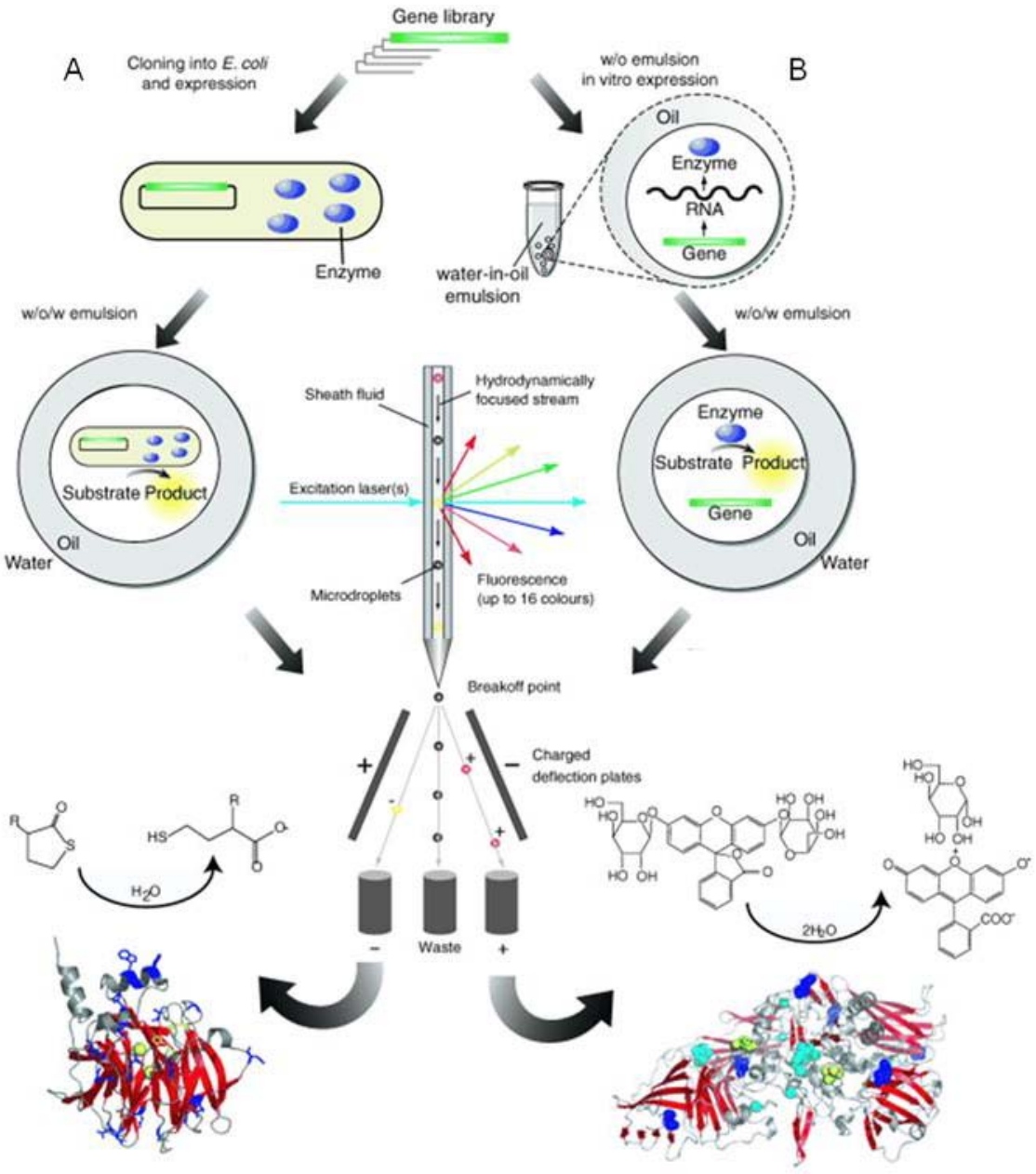


Figure 1.8 Selection by flow cytometry based screening system using double emulsion droplets (Taly, Kelly et al. 2007).

1.4 Protease screening assays

Numerous protease assays have been described and classified by detection sensitive to either agar plate based qualitative assays or microtiter plate based quantitative assays (Gupta, Beg et al. 2002). The basic principle underlying both types of methods involves

measurement of the products of protein hydrolysis or the residual protein itself. Plate assays normally rely on the formation of a clear zone of proteolysis on agar plates to indicate proteases activities and usually apply for a prescreening to identify the proteolytic activities. The most commonly used substrates for proteases agar plate assays are skim milk, casein, gelatin, bovine serum albumin and keratin (Gupta, Beg et al. 2002). Assay conditions on agar plate can be adjusted to neutral or alkaline by different buffers. Quantitative assays are kinds of spectrophotometric assays which can be carried out with hydrolysis of natural or synthetic substrates and a subsequent detection by spectrophotometry, fluorimetry, radiometry or enzyme-linked immunosorbent assay-based (ELISA). Due to simple performance and inexpensive cost the skim milk agar assay and casein-based spectrophotometric assays are most commonly used protease assays for routine laboratory analysis.

Even protein engineering of proteases commenced for a long term, today's directed protease evolution campaigns, especially for detergent alkaline proteases, still mainly use traditional screening formats such as agar plate (halo formation) or microtiter plate based detection systems with a medium throughput (10^3-10^5). The latter screening methods are usually associated with low mutational loads (1-3 amino acid changes per mutated gene) so that a medium throughput is sufficient to find improved variants (Wong, Zhurina et al. 2006). A notable example for a high throughput screening of protease variants is the directed evolution of the surface membrane protease OmpT to alter its substrate specificities by employing an *E. coli* OmpT deficient strain (UT5600) (Varadarajan, Rodriguez et al. 2008). Improved variants were identified using flow cytometry in aqueous solution without employing any compartmentalization technology. Additionally methods based on selection systems are always extremely powerful as the survival of an organism is coupled to the desired protease function reducing the necessary workload by far. Such selections systems have been developed for screening trypsin using minimal media supplemented with arginyl naphthylamide (Evnin, Vasquez et al. 1990). The mentioned selection systems are limited in their general applicability as suitable host strains have to be generated for growth selection.

Using directed evolution proteases have been evolved for newer protease preparations with improved catalytic efficiency such as cold adaption (Siddiqui and Cavicchioli 2006; Pulido, Koga et al. 2007; Yang, Zhu et al. 2008; Zhong, Song et al. 2009), increased thermostability (Zhao and Arnold 1999; Wintrode, Miyazaki et al. 2001; Haki and Rakshit 2003), higher resistance towards oxidizing agents (Weng, Zheng et al. 2009), activities in the absence of calcium-ions (Strausberg, Ruan et al. 2005; Romero-Garcia, Tellez-Valencia et al. 2009), organic solvents resistance (You and Arnold 1996) and substrate specificity (Varadarajan, Rodriguez et al. 2008; Pogson, Georgiou et al. 2009). Several previous literatures have excellently reviewed protein engineering of subtilisins (Bryan 2000; Gupta, Beg et al. 2002; Maurer 2004) and some recently approaches of directed subtilisin/subtilisin-like protease evolution are further summarized in Table 1.1.

Table 1.1 Directed evolution of subtilisins and subtilases.

Proteases	**Libraries employed**	**Screening strategies**	**Change in properties**	**References**
Subtilase WF146	epPCR and StEP	MTP	Six-fold increased activity at 15-25°C	(Zhong, Song et al. 2009)
Subtilisin nattokinase	SDM	Agar plate and MTP	10-fold increased oxidative stability compared with wild-type enzyme	(Weng, Zheng et al. 2009)
Subtilisin E	SDM	Agar plate	Demonstrated structure function of the loop that connects β-sheet e3 with α-helix on enzyme activity; Tm value of mutant was 4°C higher than wild-type	(Romero-Garcia, Tellez-Valencia et al. 2009)
TK-subtilisin	epPCR	Agar plate	Evolution of low temperature adaption	(Pulido, Koga et al. 2007)
Savinase-S39 hydrid	SDM	MTP	Broader substrate specificity at RT, increased activity at low temperature, increased binding region and global flexibility	(Tindbaek, Svendsen et al. 2004)

1.5 Aim of the project

Proteases are important hydrolytic enzymes that are widespread in nature. They do not only play an important role in physiology, but also have a commercial value. As industrially important enzymes, proteases often have to be tailored for their catalytic

efficiencies and stabilities to allow their efficient application. Directed evolution is a powerful algorithm for protease reengineering for these approaches. However, directed protease evolution mainly uses traditional screening formats with a medium throughput (10^3-10^5 per day) which only allow to screen mutant libraries with low mutational loads (1-3 amino acid changers per mutated gene) and need several iterative rounds to find improved variants. In order to be able to screen large libraries with high mutational loads (~5 mutations per kb) in high throughput for faster improvements and investigation of cooperative effects between multiple mutations flow cytometry and IVC based high throughput screening systems (~10^8 events per day) should be developed. Thereby cell-free as well as whole-cell systems were supposed to be investigated. Several aspects should be addressed to achieve the goal: 1) set-up of the flow cytometry based screening system by generation of emulsion compartmentalization, selection of a fluorogenic substrate and calibration of sorting condition of flow cytometry, 2) optimization of enzymatic reactions and fluorescence signal for sorting, 3) verification of the developed systems by sorting a reference library for an increased population or screening mutant libraries for improved inhibitor resistance. Identified variants were supposed to be characterized to validate the developed screening platforms for their applicability in directed protease evolution.

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

All chemicals used were of analytical reagent grade or higher quality and purchased from Sigma-Aldrich (Taufkirchen, Germany), Applichem (Darmstadt, Germany) or Carl Roth (Karlsruhe, Germany). All enzymes were purchased from New England Biolabs (Frankfurt, Germany), Fermentas GmbH (St. Leon-Rot, Germany), Sigma-Aldrich or Codexis (Redwood city, CA, USA). Fluorogenic substrate and fluorescent reference dye are listed in Table 2.1. Protease reversible inhibitor antipain dihydrochloride (antipain-HCl) was purchased from Sigma-Aldrich.

Table 2.1 Summary of fluorogenic substrates/fluorescent dye used.

Fluorogenic substrates/fluorescent dyes	**Ex/Em [nm]**	**Sources**
Succinyl-Ala-Ala-Pro-Phe-amino-7-methyl coumarin (Suc-AAPF-AMC)	375/455	Bachem (Bubendorf, Switzerland)
7-hydroxycoumarin-3-carboxylic acid (AMC-acid)	375/455	Sigma-Aldrich (Taufkirchen, Germany
bis-(*N*-CBZ-Val-Asp-Val-Ala-Asp)-rhodamine 110 (bis, Z-VDVAD)	488/520	Invitrogen (Darmstadt, Germany)
bis-(*N*-CBZ-Arg)-rhodamine 110 (BZAR)	488/520	Invitrogen (Darmstadt, Germany)
rhodamine 110 reference standard (R110)	488/520	AnaSpec Inc. (San Jose, CA, USA)
bis, Succinyl-Ala-Ala-Pro-Phe-rhodamine 110 (bis, Suc-AAPF-R110)	488/520	AnaSpec Inc. (San Jose, CA, USA)

2.1.2 Strains

The bacteria strains used in this work are listed in Table 2.2.

Table 2.2 Summary of the bacterial strains used.

Strains	Description	References
Escherichia coli	-	-
DH5α	*F'/endA1 hsdR17(rK-mK+) supE44 thi-1 recA1 gyrA (Nalr) relA1 D(laclZYA-argF)U169 deoR (F80dlacD(lacZ)M15)*	Stratagene
DH10B	E-Shot™ DH10B™ T1^{R} Electrocomp™ Cells	Invitrogen
Bacillus subtilis	-	-
WB600	*nprE nprB aprE epr mpr bpr*	(Wu, Lee et al. 1991)
WB800N	*nprE aprE epr bpr mpr::ble nprB::bsr Δvpr wprA::hyg cm::neo*	(Nguyen 2006)

2.1.3 Plasmids

The plasmids used during this work are listed in Table 2.3.

Table 2.3 Summary of plasmids used. SC: Subtilisin Carlsberg.

Plasmids	Description	References
pET-28a(+)	T7 promoter, Km^{r}	Novagen
pET-42b(+)	High level expression vector T7 promoter, Km^{r}	Novagen
Cosmid-M57	Cosmid Expand III carrying M57 protease, Ap^{r}	B.R.A.I.N. AG
pM46-N7	pUC carrying M46 protease, Ap^{r}	B.R.A.I.N. AG
pET-M46	pET-28a(+) carrying M46 protease, Km^{r}	This work
pET-M57	pET-42b(+) carrying M57 protease, Km^{r}	This work
pIX-M57	pIX3.0 carrying M57 protease, Ap^{r}	This work
pIVEX-M57	pIVEX2.3d carrying M57 protease, Ap^{r}	This work
pHCMC04	*E. coli* and *B. subtilis* shuttle vector for expression, *xylA* promoter, xylose inducible cassette, Cm^{r}	(Nguyen, Nguyen et al. 2005)
pHCMC05	*E. coli* and *B. subtilis* shuttle vector for expression, P*spac* promoter, IPTG inducible cassette, Cm^{r}	(Nguyen, Nguyen et al. 2005)

Plasmids	Description	References
pHCMC04SC	pHCMC04 carrying SC, Cm^r	This work
pHCMC05SC	pHCMC05 carrying SC, Cm^r	This work
pBCS-PvegII	*E. coli* and *B. subtilis* shuttle vector for expression, Cm^r	B.R.A.I.N. AG
pBCSSC	pBCS-PvegII carrying SC, Cm^r	B.R.A.I.N. AG
pBCSSCm2	pBCS-PvegII carrying SC with six mutations, Cm^r	This work
pBCSSCm13	pBCS-PvegII carrying SC with seven mutations, Cm^r	This work
pm2SDM1	SDM of SC at position 1 of pBCS-SCm2, Cm^r	This work
pm2SDM2	SDM of SC at position 2 of pBCS-SCm2, Cm^r	This work
pm2SDM3	SDM of SC at position 3 of pBCS-SCm2, Cm^r	This work
pm2SDM4	SDM of SC at position 4 of pBCS-SCm2, Cm^r	This work
pm2SDM5	SDM of SC at position 5 of pBCS-SCm2, Cm^r	This work
pm2SDM6	SDM of SC at position 6 of pBCS-SCm2, Cm^r	This work

2.1.4 Oligonucleotides

The oligonucleotides used in this work are listed in Table 2.4.

Table 2.4 Summary of oligonucleotides used. Underline indicated restriction sites or the substituted site for SDM of variant SCm2.

Name	Sequence (5´ to 3´)	Description
M46_FNc	TAATCCATGGATGCGATGCCC	Generation of pET28a-M46 plasmid
M46_RHi	TACAAGCTTTTACGGCATGCTATCAG	Generation of pET28a-M46 plasmid
M57_FNd	GGAATTCCATATGTTTGAACAAGCGAGTTTTTC	Generation of pET42b-M57 plasmid
M57_RHi	TACAAGCTTTCATTGGGGCGCTTGTG	Generation of pET42b-M57 plasmid
M57_MEGA_F	ATGTTTGAACAAGCGAGTTTTTC	Generation of M57 mutagenesis library
M57_MEGA_R	ATTGGGGCGCTTGTGCTAAG	Generation of M57 mutagenesis library
SCf1	ATGATGAGGAAAAAGAGTTTT	Generation of SC mutagenesis library

Name	Sequence (5´ to 3´)	Description
SCr1	AGCGGCAGCTTCGACATTG	Generation of SC mutagenesis library
NTagF_M57	AGAAGGAGATAAACAATGTTTGAACAAGCGAGTTTTTC	Generation of *in vitro* translation linear template
NTagR_M57	CTTGGTTAGTTAGTTATCATTGGGGCGCTTGTGCTAAG	Generation of *in vitro* translation linear template
BCS_Fseq	TTTTCTACGGGGTCTGACGC	Sequence primer for SDM
BCS_Rseq	AACTCATTCCCTGATCTCG	Sequence primer for SDM
K127Rf	CAGGCTCAAGGCTTTAGGGGAGCGAATGTAAAA	Generation of pm2SDM1 plasmid
K127Rr	TTTTACATTCGCTCCCCTAAAGCCTTGAGCCTG	Generation of pm2SDM1 plasmid
T237Pf	GGAGCATCAGGCTCGCCAGCGATGAAACAGGCA	Generation of pm2SDM2 plasmid
T237Pr	TGCCTGTTTCATCGCTGGCGAGCCTGATGCTCC	Generation of pm2SDM2 plasmid
M239If	TCAGGCTCGACAGCGATAAAACAGGCAGTCGAC	Generation of pm2SDM3 plasmid
M239Ir	GTCGACTGCCTGTTTTATCGCTGTCGAGCCTGA	Generation of pm2SDM3 plasmid
I269Vf	GGAAACACGAATACAGTTGGCTATCCTGCGAAA	Generation of pm2SDM4 plasmid
I269Vr	TTTCGCAGGATAGCCAACTGTATTCGTGTTTCC	Generation of pm2SDM4 plasmid
Y310Ff	CCTGGCGCAGGCGTATTCAGCACTTACCCAACG	Generation of pm2SDM5 plasmid
Y310Fr	CGTTGGGTAAGTGCTGAATACGCCTGCGCCAGG	Generation of pm2SDM5 plasmid
I372Vf	TATGGGAAAGGTCTGGTCAATGTCGAAGCTGCC	Generation of pm2SDM6 plasmid
I372Vr	GGCAGCTTCGACATTGACCAGACCTTT CCCATA	Generation of pm2SD61 plasmid

2.1.5 Media and culture cultivation

Cells were routinely cultivated in Luria broth (LB) supplemented with appropriate antibiotics using a shaking incubator (Multitron II, Infors GmbH, Einsbach, Germany) at 37°C and 250 rpm for 20 h. Antibiotics used in the experiments are listed in Table 2.5.

LB medium: 1 % (w/v) typton, 0.5 % (w/v) yeast extract, 1 % (w/v) NaCl

LBM medium: LB medium containing 1 % (w/v) skim-milk

LB or LBM agar plates contain 1.5 % (w/v) agar

Table 2.5 Summary of antibiotic solutions used.

Antibiotics	**Stock concentration (mg/mL)**	**Solvent used**	**Working concentration (μg/mL)**
Ampicillin	100	Water	100
Kanamycin	50	Water	50
Tetracycline	15	70 % ethanol	10
Chloramphenicol	20	70 % ethanol	12.5
Neomycin	10	Water	10

2.2 Methods

2.2.1 General methods

2.2.1.1 PCR and colony PCR

All PCRs were performed using a thermal cycler (Mastercycler gradient, Eppendorf AG, Hamburg, Germany). The amount of DNA in the experiments was quantified by using a NanoDrop photometer (ND-1000, NanoDrop Technologies, Wilmington, DE, USA).

Colony PCR was performed with a slight modification of standard PCR using colony from agar plate as template. An additional 8 min at 95°C of denaturing step at beginning was applied for colony PCR.

2.2.1.2 Cloning

All gene cloning and manipulation steps carried out according to molecular cloning book (Sambrook and Russell 2001). Preparation of *E. coli* competent cells and transformation were carried out as standard heat shock transformation (Inoue, Nojima et al. 1990) and electroporation (Dower, Miller et al. 1988). *B. subtilis* competent cells were supplied by B.R.A.I.N. AG (Zwingenberg, Germany). Plasmids were verified by PCR, restriction enzyme digestion and sequence analysis. Sequence results were analyzed by Vector NTI 9 (Invitrogen). Plasmid isolation kit, PCR purification kit and rolling circle amplification kit (RCA) were purchased from Qiagen (Hilden, Germany).

2.2.1.3 Fluorescence detection

Fluorescence was measured by a fluorescence microtiter plate reader (Tecan SAFIRE, Tecan Austria GmbH, Salzburg, Austria) or a flow cytometer (CyFlow® space, Partec GmbH, Muenster, Germany).

2.2.1.4 Assays for the detection of protease activity

Agar plate assay: LBM agar plates supplemented with appropriate antibiotics were used and proteases activities were detected by halo formation.

Fluorometric microtiter plate assay: Activity assay was carried out in 96-well microtiter plate (FIA-PLATE black flat 96 well no. 655076, Greiner Bio-One GmbH, Frickenhausen, Germany) or 384-well microtiter plate (Polystyrene black flat 384 well no. 3710, Corning Incorporated, Corning, NY, USA). Reaction was performed Tris-HCl buffer (0.1 M, pH8.6) containing suitable fluorogenic substrate (100 µM of Suc-AAPF-AMC, 100 µM of bis, Z-VDVAD-R110 or 5 µM of bis, Suc-AAPF-R110) and protease (1 µL of cell supernatant, 100 nM of purified SC protease) in a final volume of 100 µL (96-well MTP) or 50 µL (384-well MTP). The enzyme activity was measured at 37°C by monitoring the increase of fluorescence intensity using Tecan SAFIRE.

2.2.2 *In vitro* expression of M57 protease

In vitro expression was carried out according to manual supplied by manufacturer. Reactions were carried out in eppendorf tube, 384-format microtiter plate or double emulsion.

2.2.3 Emulsification of M57 reference library and sorting

The w/o/w double emulsions were prepared as described previously (Miller, Bernath et al. 2006). Fifty microliter of *in vitro* expression mix including 1 µg of M57 reference (1 % in molecule ratio of active plasmid pET-M57 to inactive plasmid pIX-delM57) and bis, Z-VDVAD-R110 (100 µM) were emulsified in w/o primary emulsion and incubated at 37°C for 4-5 h. After *in vitro* expression triethylamine was delivered through oil phase to inner aqueous phase and incubated at 37°C for 1 h. Then the second water phase layer was emulsified. Double emulsion samples were diluted 100-fold in PBS buffer and analyzed by CyFlow® space. Sorted positive events (DNA) were recovered by ethanol precipitation with linear polyacrylamide co-precipitant as described previously (Gaillard and Strauss 1990).

2.2.4 TCA precipitation and SDS-PAGE analysis

TCA precipitation and SDS-PAGE analysis protein were carried out according to standard protocol described (Sambrook and Russell 2001).

2.2.5 Generation of epPCR mutagenesis library of Subtilisin Carlsberg

The random mutagenesis library was constructed by the standard epPCR (Cadwell and Joyce 1994) and megaprimer PCR of whole plasmid (MEGAWHOP) (Miyazaki and Takenouchi 2002) as described. The epPCR library was generated by PCR (95°C for 5 min, one cycle; 94°C for 30 s/ 55°C for 30 s/ 72°C for 90 s, 30 cycles; 72°C for 7 min, one cycle) using dNTPs mix (0.2 mM), primers (SCf1: 5´-ATG ATG AGG AAA AAG AGT TTT-3´; SCr1: 5´-AGC GGC AGC TTC GAC ATT G-3´; 0.4 µM each), plasmid DNA template (pBCSSC, 10 ng/50 µL), *Taq* polymerase (2.5 U/50 µL) and $MnCl_2$ (0.4 mM). The MEGAWHOP PCR (68°C for 5 min, one cycle; 95°C for 2 min, one cycle;

95°C for 30 s/ 55°C for 30 s/ 68°C for 12 min, 18 cycles; 68°C for 30 min, one cycle) was performed in a final volume of 50 µL using dNTPs mix (0.2 mM), megaprimer (purified epPCR amplification product, 400 ng), plasmid DNA template (pBCSSC, 50 ng) and *Pfu* polymerase (2.5 U). The amplified MEGAWHOP product was digested by DpnI (20 U) at 37°C for 2 hour, purified using Qiagen kit and transformed into *E. coli* strains. Plasmids were then pooled, isolated and after rolling circle amplification (RCA) (REPLI-g, Qiagen) transformed into *B. subtilis* strains for expression and screening.

2.2.6 Emulsification of *Bacillus* cells

The w/o/w double emulsions were prepared as described previously (Miller, Bernath et al. 2006). *Bacillus* cultures were harvested using a bench centrifuge (Eppendorf centrifuge 5414D, Eppendorf AG, Hamburg, Germany) at 8000 g for 3 min and rinsed three times using ice-cold phosphate buffered saline (PBS). Approximately $5x10^9$ cells were suspended in 1 mL of Tris-HCl buffer (0.1 M, pH 8.6) and passed through a 5 µm syringe filter membrane (PC MB 25 mm 8.0 µm, Whatman GmbH, Dassel, Germany) prior to emulsification.

For *B. subtilis* host cells selection, the cell suspension (~$2x10^8$ cells) together with fluorogenic substrate bis, Suc-AAPF-R110 (5 µM), internal dye 7-hydroxycoumarin-3-carboxylic acid (no. 55155, Sigma-Alrich; 50 µM) was emulsified into w/o/w double emulsions and incubated (BD 115, BINDER, Tuttlingen, Germany) at 37°C for 2-3 h.

For epPCR mutant library screening, the emulsification condition was used as above in presence of protease inhibition (antipain-HCl, 1 µM)

2.2.7 Flow cytometry based screening and sorting of *Bacillus* mutant library

Double emulsions were diluted 100-fold using PBS buffer and analyzed by CyFlow® space (Partec GmbH; sheath fluid was composed of 0.9 % (w/w) of NaCl and 0.01 % (v/w) of Triton X-100). Fluorescence intensities of rhodamine 110 (R110) products (green fluorescence) and internal dye 7-hydroxycoumarin-3-carboxylic acid (AMC-acid) (blue fluorescence) were recorded and the sort gates were set to collect events with control blue fluorescence (internal dye AMC-acid) and increased green fluorescence

(R110 products). The sorting rate was controlled to 5-10 events/s using 8000 events/s analysis speed and around 10^4 events was sorted out in 50 mL solution. The sorted samples were then recovered by filtration (no. 7182-009, Whatman GmbH) followed by incubation on LB plates with membrane upwards placed. Grown cells were further scraped from membrane and subjected to the next round screening.

2.2.8 Single amino acid substitution and computational analysis of SC variants

Single amino acid site-directed mutagenesis (SDM) of SCm2 variant was performed according to the published method (Wang and Malcolm 1999). The oligonucleotides for six SDMs of SCm2 were listed as in Table 2.4. The procedure consists of two stages. In stage one, two extension reactions are performed in separate tubes; one containing the forward primer and the other containing the reverse primer. Subsequently, the two reactions were mixed and the stage two was carried out. For the mutagenic PCR (First stage: 95°C for 2 min, one cycle; 95°C for 30 s/ 55°C for 1 min/ 72°C for 10 min, three cycles. Second stage: 95°C for 2 min, one cycle; 95°C for 30 s/ 55°C for 1 min/ 68°C for 10 min, 15 cycles; 68°C for 30 min, one cycle), *Pfu* DNA polymerase (1 U), dNTP mix (0.2 mM), each primer (0.4 µM) together with plasmid template (50 ng) were used in 50 µL reaction volume. Following the PCR, DpnI (20 U) was added, and the mixture was incubated at 37°C for 4 h, purified using Qiagen kit and transformed into *E. coli* strains and verified by sequencing, and further transformed into *B. subtilis* cells for expression and analysis. Molecular modeling structures were obtained from SWISS-MODEL SIB Service (http://swissmodel.expasy.org/) and analyzed by using Visual Molecular Dynamics software (http://www.ks.uiuc.edu/Research/vmd/).

2.2.9 Purification and kinetic characterization of SC variants

Purification using ultra centrifugal filters: Cultivated supernatant was purified and exchanged buffer by using Amicon Ultra Centrifugal Filters (Ultracel-10 K, Millipore Corporation, Billerica, MA, USA). The target mature protein quantification was determined by BCA™ protein assay kit (no. 232225, Pierce Biotechnology, Rockford,

IL, USA) and Agilent 2100 Bioanalyzer (Agilent Protein 230 Kit, Agilent technologies, Waldbronn, Germany).

Purification using column: SC variants were purified by a two-step ion exchange chromatography procedure. The 250 mL of supernatant (37°C and 250 rpm for 24 h) was concentrated into Tris-HCl buffer (50 mM, pH 8.6) using an ultrafiltration system (amicon® 8200, Millipore Corporation, Bedford, MA, USA; ultrafiltration membrane NMWL 5000) and purified by a cation exchange (Toyopearl SP-650C, Tosoh Bioscience LLC, PA, USA; running buffer: 50 mM Tris-HCl pH 8.6, elution buffer: 50 mM Tris-HCl pH 8.6 with 2 M NaCl) followed by an anion exchange (Toyopearl SuperQ-650C, Tosoh Bioscience LLC; running buffer: 50 mM Tris-HCl pH 8.6, elution buffer: 50 mM Tris-HCl pH 8.6 with 2 M NaCl). The total protein concentration was determined by BCA™ protein assay kit (no. 23225). Kinetic characterization and inhibition resistance analysis were performed as described above in activity assay part. K_m and k_{cat} data were plotted and fitted using Prism 4 (GraphPad Software, Inc., La Jolla, CA, USA) and Origin 7.0 (OriginLab Corporation, Northampton, MA, USA).

Normalization of protein concentrations for wild-type SC and variants: Protein concentrations in fluorescence assays were normalized by using the Agilent 2100 Bioanalyzer (Agilent Protein 230 Kit) and the BCA™ protein assay kit (no. 232225) according to the manual's instructions.

3 Results

In this work, two flow cytometry based screening systems (cell-free and whole-cell based systems) for directed evolution of proteases have been developed. The experimental results are shown in sections as follows: section 3.1 shows general information of preparing w/o/w double emulsion *in vitro* compartmentalization (IVC) and calibrating flow cytometry for sorting which are required in both screening systems; section 3.2 shows the development of cell-free based screening using M57 protease as a model protease (isolated from metagenomic library and provided by B.R.A.I.N. AG); and section 3.3 shows the development of whole-cell based screening system employing Subtilisin Carlsberg and *B. subtilis* and an epPCR mutant library with high mutational loads.

3.1 Set-up of flow cytometry based screening system

Several aspects have to be addressed and developed for a successful flow cytometry IVC screening system: 1) a suitable emulsion system where a single candidate (DNA or cell) with other necessary components for enzymatic reaction are encapsulated in the IVC emulsion, 2) a detectable fluorescence signal for flow cytometry analysis and sorting, and 3) a well done calibration and optimization of sorting performance of flow cytometry regarding to a diverse of target variants to ensure an effectual sorting.

3.1.1 Preparation of emulsions

A suitable w/o/w emulsion IVC system has the following properties: 1) external phase is biochemically inert and internal phase is biochemically active to eliminate their negative effect on complex biochemical processes such as protein *in vitro* translation and enzymatic reactions; 2) generation of emulsions should be simple and the emulsion size can be adjusted; 3) emulsions should be stable during experiments and be able to recover target events entrapped afterwards (Bernath, Magdassi et al. 2004). As double emulsion IVC is used as a man-made cell-like environment to link genotype to phenotype, an ideal emulsion employs events dispersed individually in different compartments. Table 3.1

summarizs currently used emulsion systems made by two means and three emulsification systems Span/Tween, Decane/Tween and ABIL EM 90/CMC to generate w/o/w double emulsion. In this work, all of them were tested and generated emulsions were analyzed by fluorescence microscopy and CyFlow® space flow cytometry for their appearances, stabilities, manipulations.

Table 3.1 Summary of emulsification methods to generate w/o/w double emulsion IVC.

Phase I (water 1 phase)	**Phase II (oil phase)**	**Phase III (water 2 phase)**	**Generating Methods**	**References**
Model testing system 4.4 wt % FITC-BSA in 4.87 wt % Tween 80-PBS solution	4.5 wt % Span 80 in light mineral oil	2 wt % Tween 20 aqueous solution or 2 wt % Tween 20 in PBS buffer	Homogenizer	(Hai, Bernath et al. 2004)
Model testing system Model genes with FITC-BSA internal marker in 4.8 wt % Tween 80-PBS solution	4.5 wt % Span 80 in light mineral oil	2 wt % Tween 20 in PBS buffer	Homogenizer	(Bernath, Hai et al. 2004)
Cell-free system Beta-galactosidase library in *in vitro* translation reaction mix	1 wt % Span 60 and 1 wt % cholesterol in decane	0.5 wt % Tween 80 in PBS buffer	Extrusion	(Mastrobattista, Taly et al. 2005)
Whole-cell system Thiolactonase Cell suspensons in activity buffer (50 mM Tris-HCl pH 8.0, 1 mM $CaCl_2$, 100 mM NaCl) using GFP as a internal marker	2.9 wt % ABIL EM 90 in light mineral oil	1.5 wt % carboxy methyl cellulose and 1 wt % Triton X 102 in activity buffer	Homogenizer	(Aharoni, Amitai et al. 2005)

Figure 3.1 shows the procedure to generate w/o/w double emulsions by homogenizer and extrusion. For homogenizer method (Figure 3.1 A), 1 to 10-fold ratio of hydrophilic phase w1 to lipophilic oil phase was homogenized to form primary emulsion w/o then second hydrophilic phase w2 (same volume as w/o primary emulsion) was added and the mixture was emulsified to form final w/o/w double emulsions. For extrusion method (Figure 3.1 B), 1 to 4-fold ratio of hydrophilic phase w1 to lipophilic oil phase was extruded (using 12 μm pore size polycarbonate membrane) to primary emulsion w/o then second hydrophilic phase w2 (3-fold volume compared to w/o primary emulsion) was

extruded (using 8 µm pore size polycarbonate membrane) to generate final double emulsions. Integrated picture in Figure 3.1 shows the configurations of emulsions made by ABIL EM 90/CMC or Decane/Tween systems.

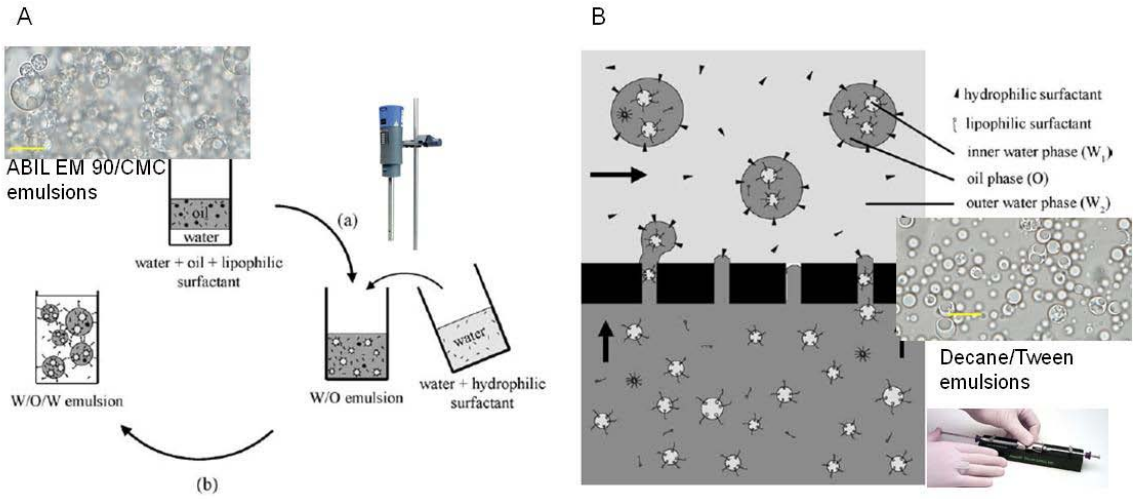


Figure 3.1 Preparation of a w/o/w double emulsion by a ultra turrax homogenizer (A) or extrusion (B). A: preparation of the w/o/w double emulsion in two steps (a) a high-shear emulsification step with lipophilic surfactants for the w/o emulsion and (b) a low shear emulsification step with hydrophilic surfactants for the w/o/w emulsion; B: preparation of a double emulsion w1/o/w2 by membrane emulsification with a sample emulsion as dispersed phase. The arrows represent the direction of the fluid flow. Embedded emulsion pictures were taken by florescent microcopy and yellow bar represents 10 µm. Picture was modified from van der Graaf (van der Graaf, Schroen et al. 2005).

Fluorescence microscopy analysis showed that double emulsions made by Decane/Tween system (Figure 3.2 B1) has better emulsified compartments compared to ABIL EM 90/CMC system (Figure 3.2 A1). However the tested Decane/Tween double emuliosns were not stable. An apparent precipitation could be detected at temperature lower than 37°C resulting in a release of entrapped fluorescence signal (Figure 3.2 B2). A separation of two-layer phase (aqueous and dacane) could also be easily observed during a long-term storage (more than 1 day at 4°C) indicating of a fusion of Decane/Tween double emulsions. The same phenomenon was also observed in Span/Tween system (data not shown). In contrast double emulsions made by ABIL EM 90/CMC sytem could stabilize for hours at 37°C, days at room temperature, and up to several months at 4°C.

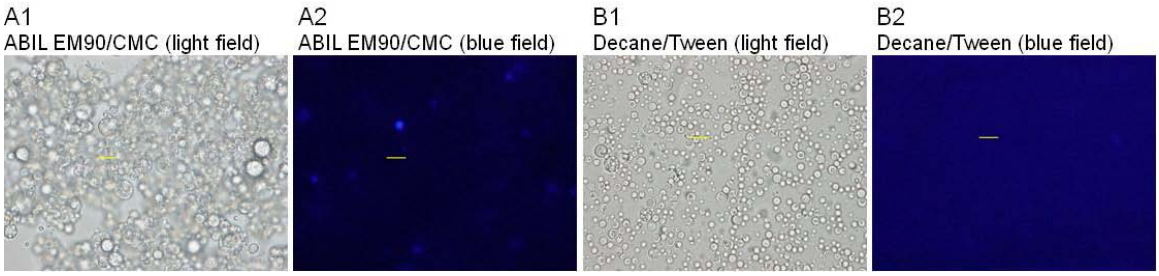


Figure 3.2 Fluorescence microscopy of w/o/w double emulsion made by different systems. ABIL EM 90/CMC system (A1, A2) and Decane/Tween system (B1, B2) using 7-hydroxycoumarin-3-carboxylic acid coumarin internal dye (50 µM). A1, B1 were taken at light field and A2, B2 were taken at blue fluorescence field. Yellow marker represents 10 µm.

Figure 3.3 shows ABIL EM90/CMC system yielded homogeneous double emulsions. By using cells carrying green fluorescent protein (GFP) emulsions can be easily optimized to emulsify one event per droplet. With internal dye control (AMC-acid displaying blue fluorescence) the 'negative' emulsions without an inner water phase for biochemical process like enzymatic reactions can be distinguished (indicated by yellow arrows in Figure 3.3) due to an absence of blue fluorescence.

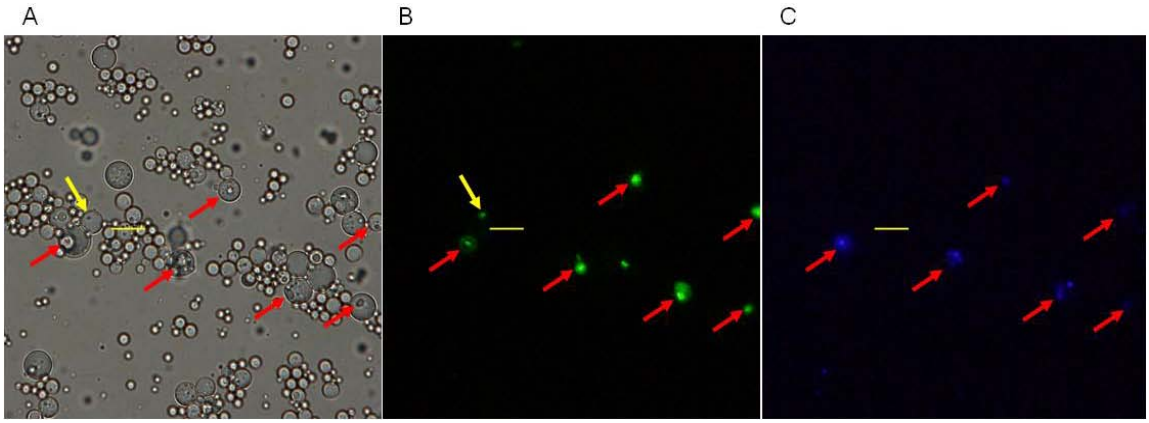


Figure 3.3 Fluorescence microscopy of ABIL EM90/CMC employed double emulsions with GFP cells entrapped inside emulsions. A: light field, B: green fluorescence field, C: blue fluorescence field. Yellow maker bars represent 10 µm. Red arrows indicate the w/o/w double emulsions with GFP and internal blue dye inside. Yellow arrows indicate the 'negative' emulsions without inner water phase.

Regarding to the performances of emulsions on configuration, stability and emulsification means, ABIL EM 90/CMC system was chosen for emulsion generations in later experiments.

3.1.2 Selection of suitable substrate

An existense of retainable fluorescence signal in IVC is a key parameter for flow cytometry analysis. Therefore to determine single IVC droplet with an accomplished enzymatic reaction, fluorescent product should be remained inside emulsion compartments without cross-talk between compartments. To perform enzymatic reactions inside individual w/o/w double emulsion compartmentalization, substrate should get into inner aqueous phase to react with enzyme for the product conversion. In this work three commercial fluorogenic substrates were tested due to either low price or high fluorescent intensity including one 7-amnio-4-methylcoumarin (AMC) based substrate (Suc-AAPF-AMC) and two rhodamine 110 (R110) based substrates (BZAR and bis, Z-VDVAD-R110). Conversion of non-fluorescent substrate to fluorescent product AMC or R110 are shown in Figure 3.4 and Figure 3.5. Both substrates have a compatible excitation/emission wavelength to CyFlow® space flow cytometry analysis.

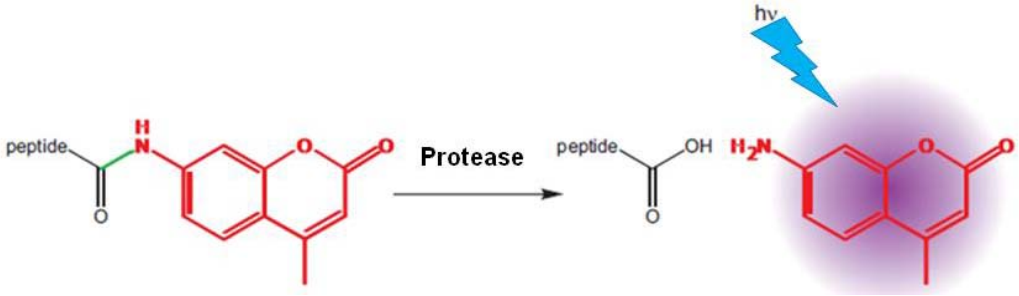


Figure 3.4 Reaction scheme of AMC based substrate assay (Ex 342 nm, Em 441 nm). Enzyme substrate based on AMC fluorophore is typically linked to the peptide through formation of an amide bond between the coumarin amine and the carboxyl group of the C-terminal amino-acid residue. Proteolysis of this amide bond liberates free AMC, resulting in a large increase in fluorescence (EMD Biosciences).

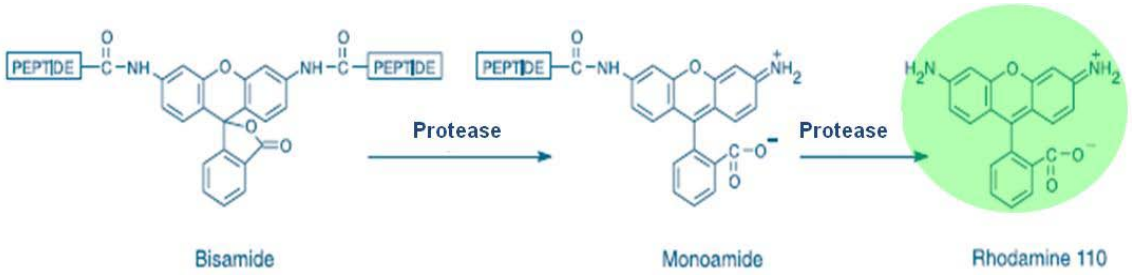


Figure 3.5 Reaction scheme of rhodamine 110 assay (Ex 496 nm, Em 520 nm). R110-based substrate usually comprise two identical amino acides or peptides attached to a single fluorophore. Upon enzymatic cleavage, the nonfluorscent bisamide substrate is converted in a two-steop process first to the fluorescent monoamide and then to the even more fluorescent R110 (Invitrogen).

Selection of suitable fluorogenic substrate was performed in three steps. Step 1: perform enzymatic reactions using substrate inside microtiter plate and analysis fluorescence signal by Tecan SAFIRE to determine fluorescent product intensity and background noise; Step 2: emulsify product mixture (from step 1) in w/o/w emulsion to determine fluorescent product retainable inside emulsion; Step 3: perform enzymatic reactions inside compartmentalized emulsions to test whether fluorogenic substrate could reach enzymes inside emulsion to carry out reaction and give signal for analysis. Experiments were performed using fluorogenic substrates with protease solution (M57 protease) and results are shown in Table 3.2.

Table 3.2 Summary of different fluorogenic substrates tested. MTP: microtiter plate, FC: flow cytometry.

Substrates	**Applications**		**Performance**
	MTP	**FC**	
Suc-AAPF-AMC	Yes	No	Substrate couldn't remain inside droplets of double emulsion
bis, Z-VDVAD-R110	Yes	Yes	Substrate and relative product could remain inside
BZAR	No	No	Background noise from negative control was too high

Results showed in step 1 AMC based substrate (Suc-AAPF-AMC) gave more fluorescence signal in microtiter plate than R110 based substrate (bis, Z-VDVAD-R110) indicating its favorite to serine protease (M57 protease). However step 2 investigation revealed that AMC fluorescent product diffused quickly out of emulsions as non-fluorescence signals could be determined by flow cytometry. In contrast R110 product could remain inside emulsion compartments. In step 3 enzymatic reaction was performed in emulsions and results were shown in Figure 3.6 by fluorescence microscopy analysis of an existence of fluorescent product converted from fluorogenic substrate bis, Z-VDVAD-R110 by enzymatic reaction in emulsions. Figure 3.6 A1 shows ABIL EM 90 (2.9 wt % in light mineral oil) yielded a homogeneous single emulsion with an average compartment size of 2-3 µm. Figure 3.6 A2 shows rhodamine 110 fluorophore was retained in w/o emulsions in a well distribution. Figure 3.6 B also demonstrated the R110 can remain in w/o/w double emulsions. These results indicated that bis, Z-VDVAD-R110

is suitable for protease enzymatic reaction in ABIL EM90/CMC based double emulsions condition.

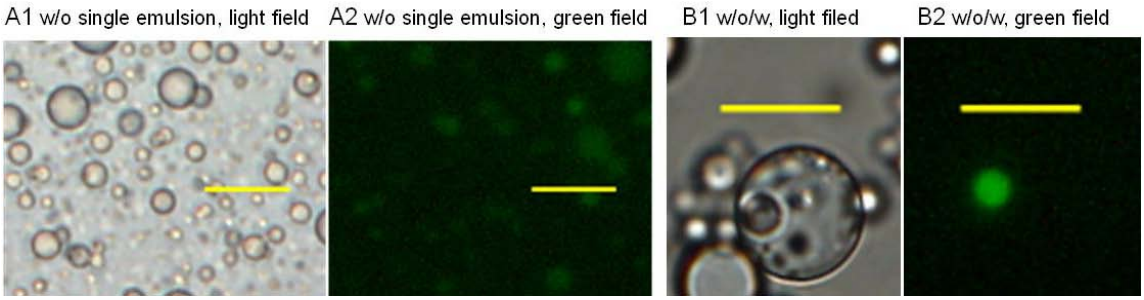


Figure 3.6 Fluorescence microscopy recording of w/o single emulsion (A1, A2) and w/o/w double emulsion (B1, B2). Enzymatic reactions were carried out by encapsulateing R110 based substrate (bis, Z-VDVAD-R110, 100 µM), internal control coumarin fluorophore dye (AMC-acid, 50 µM) and protease solution (M57 protease). Pictures were taken after enzymatic reaction at 37°C for 4 h at bright light field (A1, B1), blue fluorescence (A2) and green fluorescence field (B2). Marker bars represent 10 µm.

Further experiments showed R110 fluorophore exhibited an intensive green fluorescence at 488 nm excitation. The fluorescence signal was very stable and showed non-influence by working pH (from 7.0 to 9.0) or temperature (25°C, 37°C). The correlation between the fluorescence signal and the amount of R110 product concentrations also gave a perfect linear detection range (r^2=1) below 2.5 µM of R110 which was sufficient for detection assay. The readout value of R110 product could be adjusted from one to fifty thousand with different modulated amplifier stage gains (60 to 100). Figure 3.7 shows the correlation between R110 product concentrations and the intensity of the detected fluorescence signal in microtiter plate conditions (Figure 3.7 A) and flow cytometry analysis (Figure 3.7 B).

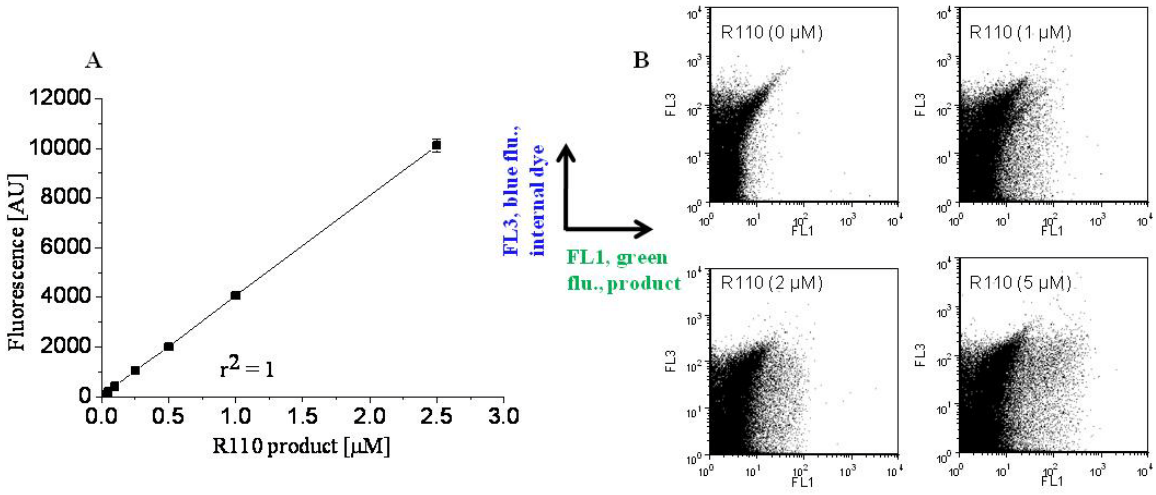


Figure 3.7 Correlation between the fluorescence signal and the amount of R110 concentration. The fluorescence intensity was determined using Tecan SAFIRE (Tecan Austria GmbH; excitation: 488 nm, emission: 520 nm, gain 70) in Tris-HCl (0.1 M, pH 8.6) at 37°C (Figure 3.8 A) or CyFlow® space analysis employed different concentrations of R110 and 100 µM of internal dye 7-hydroxycoumarin-3-carboxylic acid (Figure 3.9 B). Data in microtiter plate were measured from triplicate experiments.

3.1.3 Calibration and optimization of flow cytometry for sorting

For flow cytometry based screening system a calibration of sorting performance is necessary before any sorting procedure. A three-step calibration was carried out to calibrate the Partec CyFlow® space flow cytometry. First, the CyFlow® space was calibrated by using fluorescent microbeads (purchased from Partec). Two types of calibration beads, beads green and beads 3 µm, were chosen due to their sorting parameter could be gated by FL1 (green fluorescence channel) which is related to the R110 product excitation wavelength used.

As shown in Figure 3.10, a reference library with approximately 2 % of calibration beads green (CBg) in excess of calibration beads 3 µm (CB3) was analyzed and sorted by flow cytometry. Two types of sorting gate were used either by one parameter in histogram based on green fluorescence signal (RN1) or by two parameters in dotplot using polygon region gate (R1).

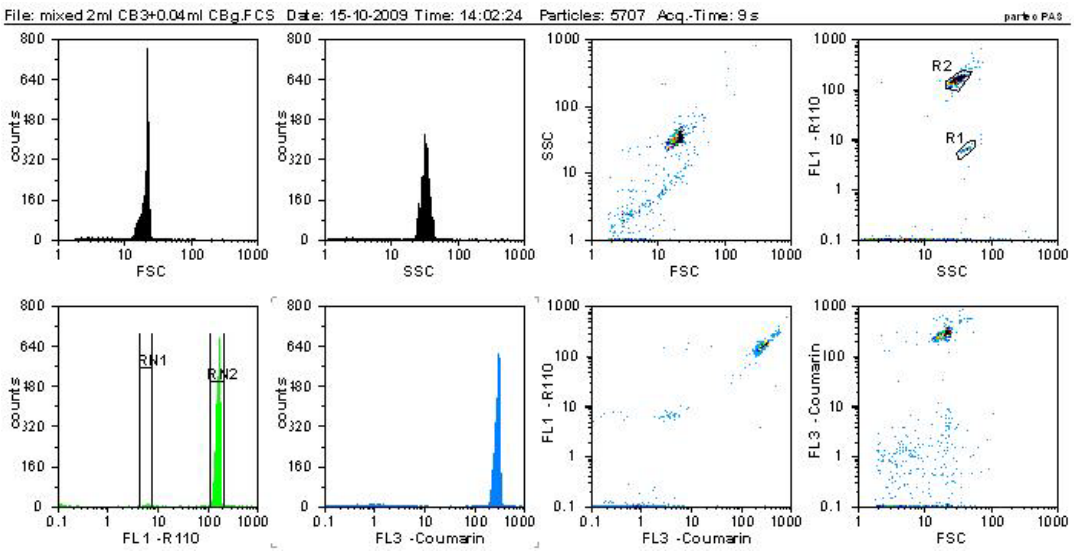


Figure 3.10 Flow cytometry recording of sorting calibration beads. Two types of calibration beads were used i.e. calibration beads green (CBg) and calibration beads 3 µm (CB3). FSC (forward scatter) represents particles size, SSC (side scatter) represents gradient of particles, FL1 represents green fluorescence density, FL3 represents blue fluorescent density. RN1 and R1 gate were chosen as sorting region to enrich CBg.

The sorting results (Table 3.3) showed Cbg could be efficiently enriched in both types of sorting parameter. In R1 applied sorting process 2 % of CBg could be sorted and enriched to 96 % in one round of sorting verifying a ready use of CyFlow® space. The enrichment from R1 gate was higher than RN1 indicating a better performance by using two parameters involved polygon gate R1 for sorting.

Table 3.3 Verifying sorting function of CyFlow® space by using calibration beads. RN1 and R1 gate represents CBg events in mixed microbeads. RN2 and R2 gate represents excess CB3 beads in the mix. RN1/(RN1+RN2) ratio represents the CBg percentage in total mixed microbeads library.

2 % CBg	**RN1**	**RN2**	**R1**	**R2**	**RN1/(RN1+RN2)**	**R1/(R1+R2)**	**Enrichment**
Before sorting	92	4818	92	4830	1.8 %	1.9 %	
After sorting (sort RN1)	61	35	52	33	63.5 %		35-fold
After sorting (sort R1)	63	3	60	3		95.5 %	50-fold

Secondly, because of emulsions size also need to be counted in efficient sorting procedure, calibration of sorting performance according to size gate was carried out. Two

types of well distinct cells (yeast cells and *E. coli* cells carrying GFP) were used to verify the sorting function of different size of particles by CyFlow® space. Flow cytometry recording of signals from Yeast cells, GFP cells, and two mixed reference libraries before sorting or after sorting are shown in Figure 3.11.

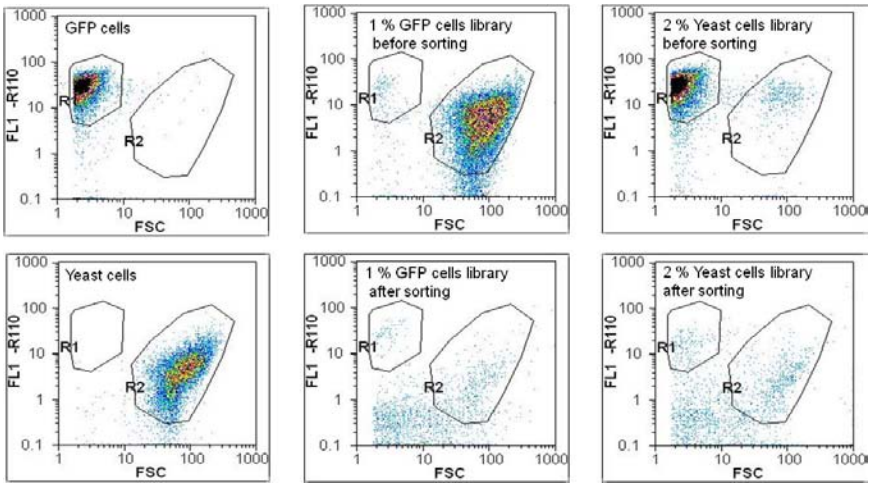


Figure 3.11 Flow cytometry recording of sorting two types of cells. Yeast cells and GFP cells mix was prepared by simple mixing of two cells. R1 gate represents the target for GFP cells sorting while R2 gate represents the target events for Yeast cells sorting.

Sorting results (Table 3.4) demonstrated that a more than 24-fold enrichment could be obtained from both sorting revealing an efficient sorting performance of CyFlow® space towards different particle sizes.

Table 3.4 Verifying sorting function of CyFlow® space by using Yeast and GFP cells. R1 gate represents GFP cells. R2 gate represents Yeast cells.

Library	**R1**	**R2**	**R1 % or R2 %**	**Enrichment**
1 % GFP library	**R1**	**R2**	**R1/(R1+R2) %**	**Enrichment**
Before sorting	139	23656	0.6 %	/
After sorting	72	425	14.5 %	24-fold
2 % Yeast library	**R1**	**R2**	**R2/(R1+R2) %**	**Enrichment**
Before sorting	20020	410	2 %	/
After sorting	163	592	78 %	39-fold

In the last step, calibration was done by using double emulsion samples. GFP labeled cells expressing a green fluorescent protein was chosen for sorting test and optimization because emulsion entrapped GFP gives same ex/em wavelength parameter as R110 product. Therefore in this model system CyFlow® space sorting performance towards green fluorescence signal under double emulsions condition could be optimized without considering enzymatic reaction effect on fluorescence signal.

Reference library employed 1 % GFP labeled cells in excess of blank *E. coli* cells (non fluorescence) was emulsified into w/o/w double emulsions and applied for sorting and enriching GFP labeled cells. Different relative parameters such as trigger parameter for lower noise by omitting events with less level of parameter value, and time delay which optimizes the time difference of analyzed particles through detecting point and sorting point were optimized. Table 3.5 shows the sorting performance with different parameter tested.

Table 3.5 Optimization of sorting performance towards green fluorescence signal under emulsions condition. FC100: double emulsions were diluted 100-fold by PBS buffer for flow cytometry analysis.

Samples	**Sorted volume [mL]**	**200 events in X mL for plate**	**Non-selection plate***	**Selection plate***	**GFP percentage**	**Enrichment**
2 % GFP FC100 emulsion (before sorting)	/	/	**916**	**14**	1.1 %	/
Sort 1 (Trigger FL1, Time delay 2, HV 4.40)	1.3	0.52	**110**	**23**	20.9 %	19
Sort 2 (Trigger FL1, Time delay 2, HV 4.60)	1.1	0.44	**143**	**35**	24.5 %	22
Sort 3 (Trigger FL1, Time delay 2, HV 4.80)	1.4	0.56	**111**	**22**	21.6 %	20
Sort 4 (Trigger FL1, Time delay 2, HV 5.00)	1.35	0.54	**101**	**26**	25.7 %	23
Sort 5 (Trigger FL1, Time delay 2, HV 5.20)	**1.3**	**0.52**	**80**	**37**	**46.3 %**	**42**
Sort 6 (Trigger FL1, Time delay 2.5, HV 4.60)	**1.4**	**0.56**	**88**	**35**	**39.8 %**	**36**

Samples	Sorted volume [mL]	200 events in X mL for plate	Non-selection plate*	Selection plate*	GFP percentage	Enrichment
Sort 7 (Trigger FL1, Time delay 1.5, HV 4.60)	1.4	0.56	**121**	**29**	24.0 %	22
Sort 8 (Trigger FL1, Time delay 1, HV 4.60)	1.6	0.64	**115**	**16**	13.9 %	13
Sort 9 (Trigger FSC, Time delay 2, HV 4.60)	**2.7**	**1.35**	**81**	**34**	**42.0 %**	**38**

Table 3.5 shows at optimized conditions (Sort 5, Sort 6 and Sort 9), the CyFlow® space can enrich GFP emulsion sample from around 1 % up to 40 % in one round of sorting. The sort 5 and 6 setup parameter was recommended here to avoid a dilution problem due to piezoelectric activation sorter module of CyFlow® space. In piezoelectric activation sorter module sorting channel is a side branch of main sheath fluid path channel. Therefore an unavoidable continuous fluid leakage from the sorting channel is unavoidable and always results in 'extra sorted' negative events which influence the sorting performance with a lower enrichment factor.

By using cells as a checking reference, the leakage level was determined by using cells as a reference. Amount of cells were counted from agar plate. As shown in Table 3.6, the results reveal: 1) continuous fluid from sorting channel is approximately 2.3 mL/min; 2) unexpected leakage cells from analyzed samples is 0.6 cells/s using FC100 sample (100-fold diluted w/o/w double emulsions in PBS buffer) at an analysis rate ~10,000 events/s.

Therefore by using FC100 samples, the minimum sorting rate is 5-10 events/s at analysis rate ~8000 events/s is required to control the leakage of unexpected events in a low range. Otherwise, the enrichment will be highly decreased by leaked cells.

Table 3.6 Determination of leakage level of CyFlow® space.

Samples	Viable cells for CyFlow® [1/s]	Leakage cells [1/mL in 25 s] & [1/s]		Leakage ratio to analyzed events [%]	Analysis rate [1/s]	Theoretic analyzed events [1/s]	Theoretic cells [1/s]
	A	B	B/25	(B/25)/A	-	-	C
FC1000**	22	1	-	0.18 %	5,300	10^4	100
FC200**	59	3	-	0.20 %	9,800	$5x10^4$	500
FC100**	242	16	0.6	0.26 %	11,000	10^5	1000
0.01 OD*	1,120	225	9	0.8 %	350		10^4
0.1 OD*	8,700	2240	90	1.0 %	2,700		10^5
0.5 OD*	53,500	8000	320	0.6 %	6,800		$5x10^5$
1.0 OD*	86,000	~20000	800	1.0 %	8,700		10^6

*1 OD unit is assumed as 1E9 cells; **FC samples were prepared by diluting emulsions into PBS buffer at different folds. In FC sample, 1 OD unit is rectified as $2x10^8$ cells.

3.2 Cell-free screening system for proteases (ProCF-FCIVC)

Based on the results described above a cell-free based flow cytometry screening system using w/o/w double emulsion technology was developed, named as ProCF-FCIVC (screening system for Protease using cell-free, flow cytomety and *in vitro* compartmentalization). The procedure includes: 1) selection of a suitable substrate for enzymatic reactions in emulsion compartments, 2) optimization of expression of protease using *in vitro* translation system for sorting and 3) sorting library to verify the ProCF-FCIVC system.

Selection of suitable substrate was carried out as described in chapter 3.1.2. bis, Z-VDVAD-R110 was chosen as enzymatic fluorogenic substrate in this work. The rest two main steps are addressed as below. Two reference proteases M46 and M57 (isolated from metagenomic library, provided by B.R.A.I.N. AG) were used.

3.2.1 Optimization of protease expression and enzymatic reaction

As high fluorescence signal is a key parameter for sorting, efficient protease expression is necessary to ensure conversion rates of substrate to fluorescent product which provide a detectable signal for screening and sorting. In this work optimization of protease expression and enzymatic reaction in cell-free IVC system was performed by two steps. Optimization was first done outside emulsion then inside emulsion.

As mentioned above to avoid any effect on protein expression caused by emulsion reagents, protease expression using *in vitro* translation (IVT) system was firstly optimized in non emulsification conditions such as eppendorf PCR tube or microtiter plate.

According to IVT expression kit manual, a variety of DNA templates containing a T7 or a strong *E. coli* promoter upstream from the coding sequence and a ribosome binding site can be used for expression. Therefore at the beginning the two reference proteases M46 and M57 were subcloned into pET expression vector which has a T7 promoter to fit the expression requirement. Expression ability was analyzed by activity determination. M57 protease using pET system was selected due to a high activity and a high fluorescence signal which might result in a improved distinct positive gate region for sorting. Table 3.7 shows a comparison of *in vitro* expression of at different conditions. So in following experiment M57 protease was chosen for the work.

Table 3.7 IVT of two reference protease. IVT was carried out using 100 ng of DNA template in 10 µL of IVT reaction mix. n.d.: not detected.

Activities [AU]	**Blank control**	**GFP control**	**pET-42b(+) control**	**pET-M57 protease**	**pET-28a(+) control**	**pET-M46 protease**
Suc-AAPF-AMC[1]	21	21	21	766	21	19
bis, Z-VDVAD-R110[2]	n.d.	n.d.	12	958	n.d.	n.d.

[1] Assay condition: Tris-HCl buffer (0.5 mM, pH 8.5), Suc-AAPF-AMC (100 µM), 2.0 µL IVT mix (expressed at 37°C for 1 h) at 37°C for 1 h. [2]Assay condition: as above using bis,Z-VDVAD-R110 (100 µM) at room temperature for overnight.

The protein expression levels using different DNA template sources were analyzed. Among them expression buffer (without DNA template) was used as a negative control for blank background and pET-M57 was used as a positive control for protease expression. Linear-M57 was analyzed due to its convenience for iterative rounds of sorting, and Cosmid-M57 served as an original vector system from metagenome library. The results (Figure 3.12) showed pET system exhibited the highest activity in all samples indicating its highest expression level. SDS-PAGE analysis results (Figure 3.12) showed that hydrolysis varied with protease activity which was consistent with the results found in activity analysis. Hereby the activity of protease can be also roughly estimated according to the degradation level of IVT mix.

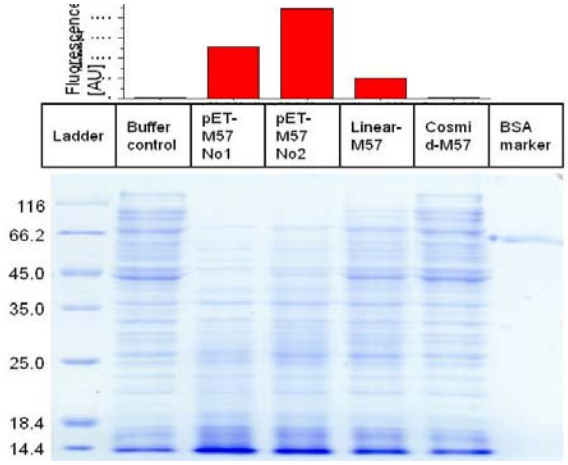


Figure 3.12 Determination of IVT expression of M57 protease using different DNA templates. IVT was carried out using 100 ng of DNA template in 10 µL of IVT reaction mix. Assay reaction was carried out using 2.5 µL of IVT expression mix (at 37°C for 4 h) and Suc-AAPF-AMC (100 µM) in Tris-HCl buffer (0.1 mM, pH 8.5). 2 µL of IVT expression mix was loaded to 12 % SDS-PAGE to check the expression.

As shown in Figure 3.12 a degradation of whole proteins in IVT mix was detected, this hydrolysis caused by expressed protease might also degrade other function proteins used for *in vitro* transcription/translation. To protect other functional proteins in IVT mix from the degradation by expressed proteases, bovine serum albumin (BSA) bulk protein additive was added into the IVT system for analysis. Other factors such as substrate dissolvent DMSO and cleavage effect of expressed protease to IVT reaction mix were also investigated using GFP reference protein. GFP without additives was used as reference control; GFP with other additives like DMSO and M57 DNA template were

used to detect for the influence of organic solvent and expressed proteases; GFP with M57 protease and BSA was used to analyze BSA effect on the degradation. All these different IVT expression conditions are shown in Figure 3.13. A clear decreasing expression yield of GFP protein could be observed in additional coexpressed M57 protease samples.

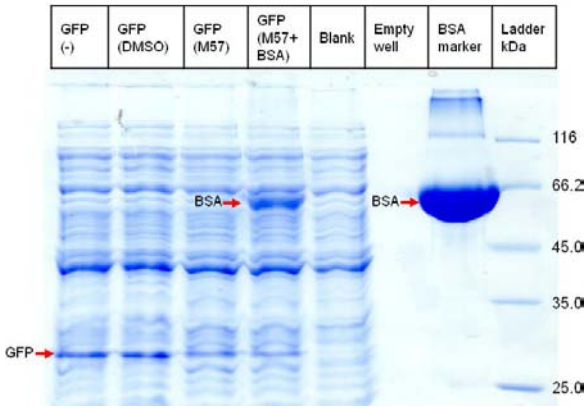


Figure 3.13 Investigation of factors effect on IVT expression system. IVT was carried out using 100 ng of DNA template in 10 µL of IVT reaction mix. Additional solvent DMSO (0.2 µL including Suc-AAPF-AMC 50 µM), M57 DNA template (100 ng) and BSA protein (0.5 µg) were added into different samples. 3 µL of IVT expression mix was loaded to 12 % SDS-PAGE to check the expression.

Further green fluorescence determination showed around 40 % yield of GFP protein was depressed by organic solvent DMSO and expressed proteases. Adding BSA bulk protein couldn't prevent GFP from the degradation by coexpressed M57 protease but it could slightly increase 10 % yield of M57 protease expression (Table 3.8).

Table 3.8 Investigation of factors effect on IVT expression system. IVT was carried out using 100 ng of DNA template in 10 µL of IVT reaction mix. Activity assay reaction was carried out using 2.5 µL of IVT expression mix (at 37°C for 4 h) and Suc-AAPF-AMC (100 µM) in Tris-HCl buffer (0.1 mM, pH 8.5). Green fluorescence intensity from expressed GFP protein was directly measured out.

Parameter	**Determination**	**GFP**	**GFP+ DMSO**	**GFP+ M57**	**GFP+ M57+BSA**	**Buffer control**	**Water blank**
488 nm/520 nm (Gain 80)	Expression samples	302	186	199	197	20	14
375 nm/455 nm (Gain 60)	Expression samples	1	321	7	7	6	8
375 nm/455 nm (Gain 60)	Activity assay	2	295	3813	4167	8	3

In order to obtain an even better yield of expression, besides pET system (5390 bp, from EMD company) another two well optimized commercial expression vectors pIVEX2.3d (2560 bp, for Roche Rapid Translation System) and pIX3 (2700 bp, for Qiagen EasyXpress Protease Synthesis), with a much smaller size compared to pET system which can help to generate a library, were examined.

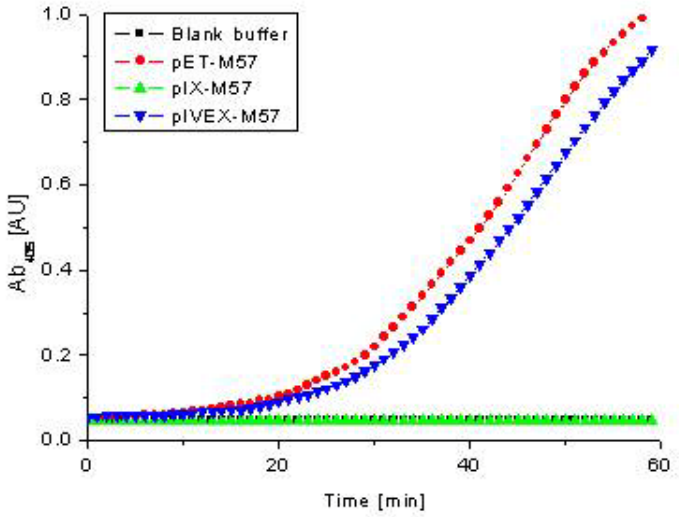


Figure 3.14 Determination of M57 protease expression in different IVTT vector systems. IVT was carried out using 100 ng of DNA template in 10 µL of IVT reaction mix. Activity assay reaction was carried out using 2.5 µL of IVTT expression mix (at 37°C for 4 h) and Suc-AAPF-pNA (1 mM) in Tris-HCl buffer (0.1 mM, pH 8.5).

As shown in Figure 3.15 experiment results revealed when M57 protease was expressed in the tube and then encapsulated in emulsion (using small amount of IVT mix) for enzymatic reaction in Tirs-HCl buffer, there was a very high fluorescence signal. However when M57 protease was directly expressed in emulsion and afterwards performed enzymatic reaction using original IVT mix, there was no visible fluorescence signal.

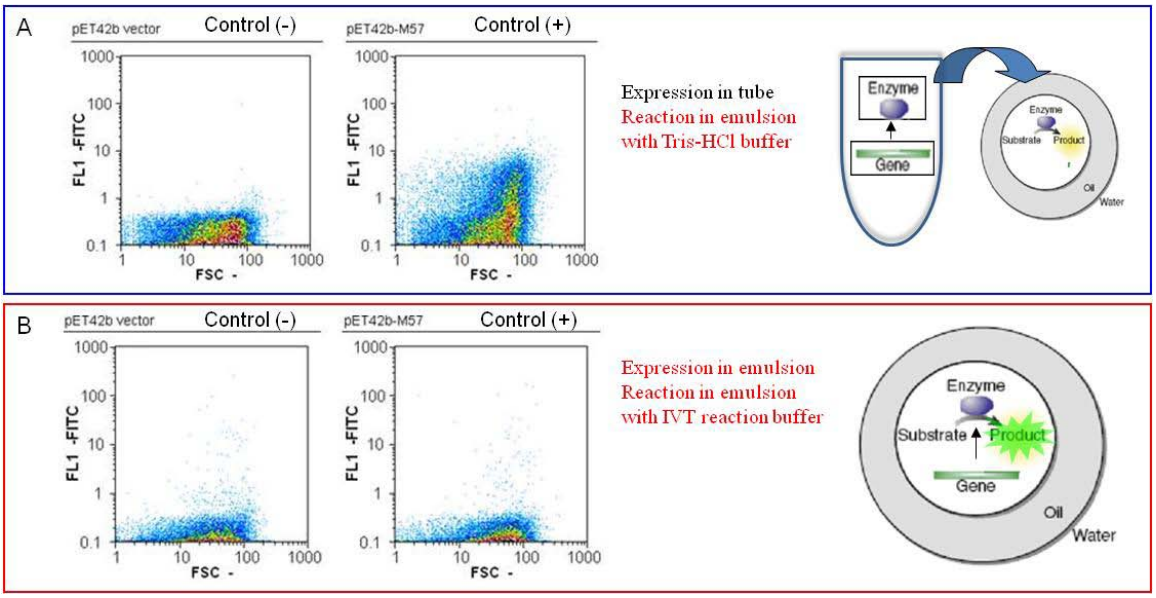


Figure 3.15 Comparison of enzymatic reaction signal with different procedures. In Acase, M57 protease was first expressed in tube and expressed protease was diluted into Tris-HCl buffer and then encapsulated into emulsion to perform enzymatic reaction in. In B case, M57 protease was expressed directly inside emulsion and enzymatic reaction was subsequently used to perform in original IVT reaction mix.

The possible reasons are 1) protease could not be expressed inside w/o/w emulsion compartments; 2) protease could be expressed in emulsion but didn't have activity. In order to figure out the actual reason why there was no detectable fluorescence signal coming from enzymatic reactions when M57 protease was directly expressed in emulsion compartmentalization, the following experiments were carried out: 1) GFP reference protein was expressed in emulsion to test whether protein could be expressed using IVT in w/o/w emulsion, 2) IVT reaction mix was checked for its effect on protease enzymatic reaction at different parameters like pH and substrate concentration.

Expression abilities under non-emulsification or emulsification conditions were analyzed using GFP protein as a reference. In the experiment GFP protein was expressed using *in vitro* expression system in either tube (IVTT) or emulsion (IVTE) conditions. The IVTT expressed blank and GFP samples were emulsified after expression and subsequently used for flow cytometry analysis. As shown in Figure 3.16 an increased green fluorescence signal (gate RN1) was observed indicting the expression of GFP protein.

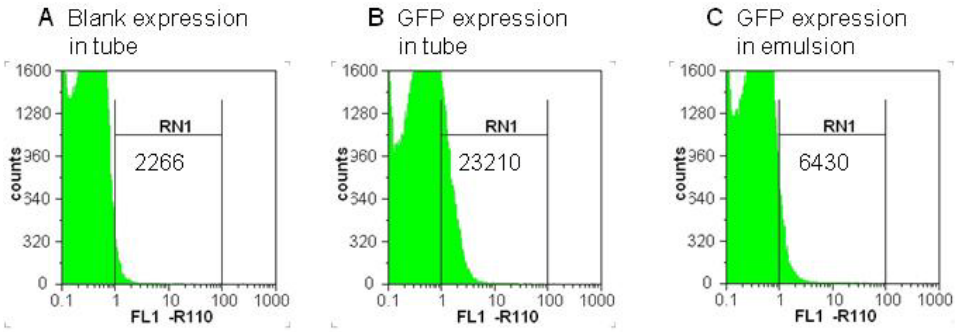


Figure 3.16 Comparison of *in vitro* expression ability under tube (IVTT) or emulsion (IVTE) conditions using GFP reference protein. IVT was carried out using 1 µg of DNA template in 50 µL of IVT reaction mix.

The relative fluorescence signal was also measured by Tecan SAFIRE (488 nm/520 nm, Gain 80, 384-well using 25 µL volume of IVTT or IVTE mix) and the fluorescence value of each emulsion sample was 2, 47 and 9 which was consistent with the results obtained from flow cytometry analysis. These results revealed emulsion did affect the expression level of IVT but protein still could be expressed inside emulsion compartmentalization and the expression level was around 16 % in IVTE compared to IVTT condition.

pH measurement of IVT reaction mix revealed IVT reaction buffer (Buffer Q) was neutral. As M57 protease belongs to serine protease family, a higher pH would increase the enzymatic rate. pH effect on protease activity was determined using purified M57 protease in microtiter plate condition. Figure 3.17 demonstrates purified M57 protease has an improved activity in high pH (Tris-HCl buffer, pH 8.6). Comparison of activities obtained in only Tris-HCl buffer or Buffer Q involved Tris-HCl buffer, a less activities were observed in Buffer Q involved conditions indicating there might be other effects (probably protease inhibitors) in Buffer Q system which depress the enzymatic activity of expressed proteases.

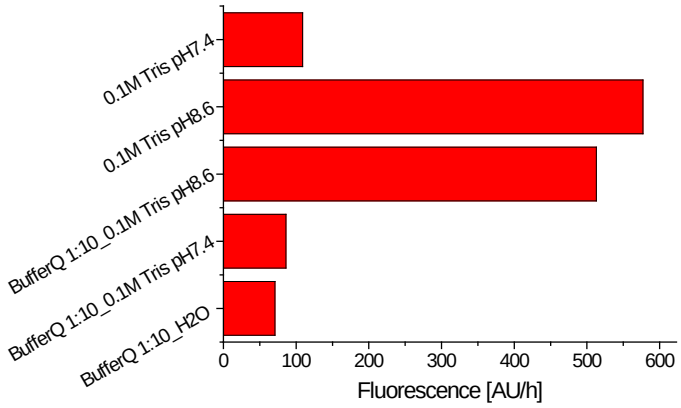


Figure 3.17 pH effect on enzymatic reaction using purified M57. Reactions were carried out using bis, Z-VDVAD-R110 (100 µM) and 300 nM purified protease in 50 µL volume at different Tris-HCl buffer and diluted IVT reaction buffer (Buffer Q).

pH effect on enzymatic reaction of M57 was further determined in w/o/w emulsion condition (as shown in Figure 3.18). The result was consistent with data found in microtiter plate as above. A higher fluorescence signal (green fluorescence intensity indicated at FL1 x-axis) was found at a higher pH (Figure 3.18 C, Tris-HCl buffer pH 8.5).

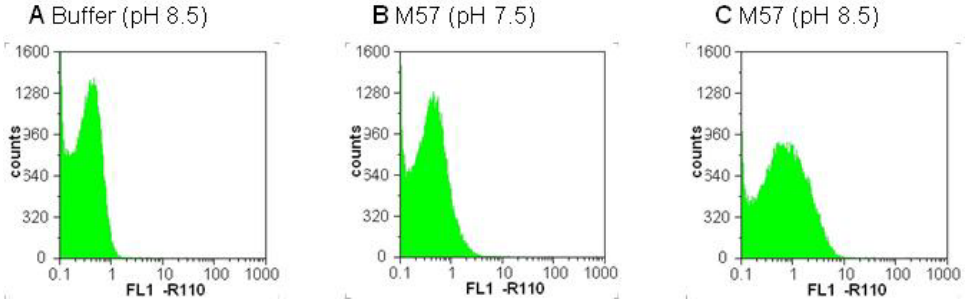


Figure 3.18 pH effect on enzymatic reaction of purified M57 in emulsion condition. Parameter FL1 represents green fluorescence intensity from R110 product and FSC represents particle size. Enzymatic reactions were carried out using purified M57 and bis, Z-VDVAD-R110 (50 µM) in various Tris-HCl buffer conditions (0.1 M, pH 7.5 or pH 8.5).

To elevate enzymatic activity in emulsion condition, several chemicals were examined to shift pH of emulsion. The results revealed that triethylamine (TEA) could efficiently alter the pH of emulsion. As shown in Figure 3.19 purified M57 activity was increased

according to increased pH of emulsion. Fluorescence signal (green fluorescence indicated by FL1 parameter) was significantly increased by increasing the pH of emulsions. The pH of emulsions was measured by pH paper measurement and the results demonstrated successful pH adjustment by using TEA.

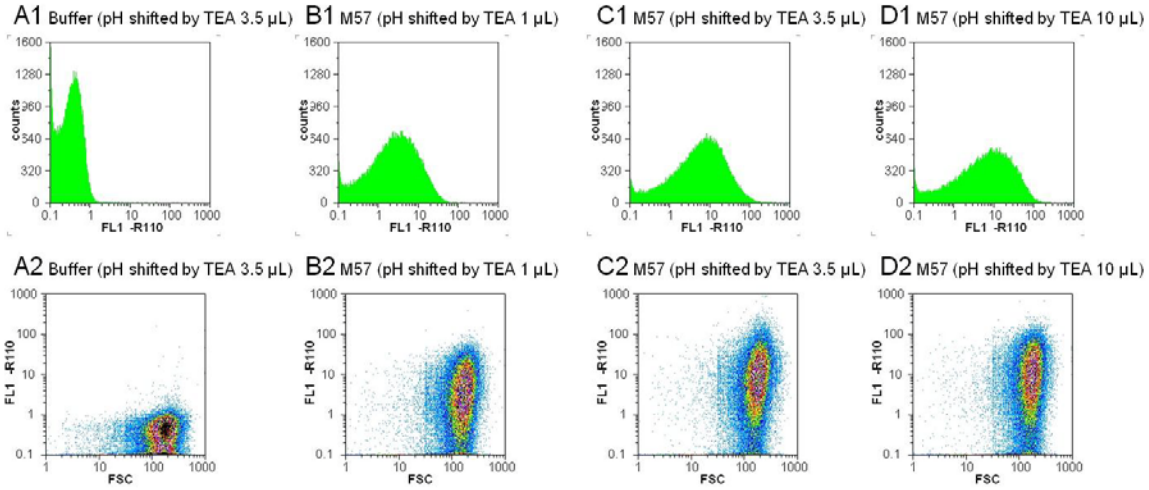


Figure 3.19 Altering pH by triethylamine (TEA). Enzymatic reactions were carried out using purified M57 and bis, Z-VDVAD-R110 (50 µM) in Tris-HCl buffer (0.1 M, pH 7.4). TEA was added through primary w/o emulsion to alter the pH and perform enzymatic reaction at 37°C for 4 h and then second w phase was added and emulsified to double emulsion for CyFlow® space measurement. Parameter FL1 represents green fluorescence intensity from R110 product and FSC represents particle size.

A determination was also done by using IVTT expressed M57 protease (Figure 3.20). Flow cytometry recording data disclosed the pH of emulsion could be adjusted from neutral to alkaline (pH 8.5) by adding 3.5 µL of TEA.

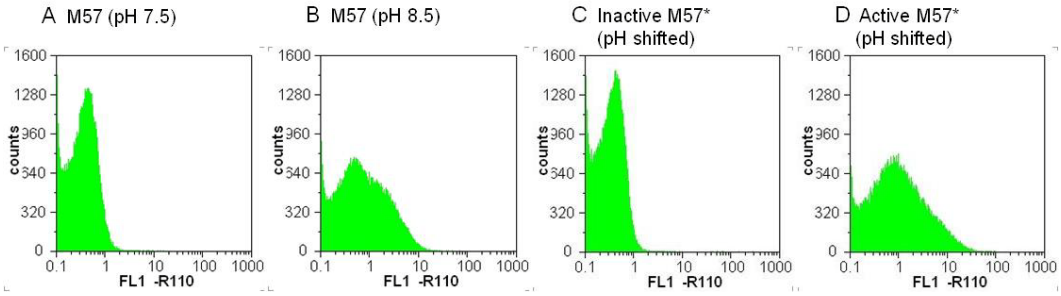


Figure 3.20 CyFlow® space flow cytometer recording of emulsions using purified protease in Tris-HCl buffer at pH7.5 (A), pH 8.5 (B) and *in vitro* expressed (in eppendorf tube) inactive protease (C) or active protease (D) with emulsion pH shifted from neutral to alkaline.

Based on the optimization results shown above, expression of M57 protease and enzymatic reactions were finally carried out in w/o/w double emulsions. Figure 3.21 A shows fluorescence signals from expressed protein samples where inactive/active M57 proteases was expressed in w/o primary emulsion firstly and then its enzymatic reaction condition (pH) was altered by adding TEA. Figure 3.21 B shows additional improvement of target green fluorescence signal (represented at FL1 parameter) by increasing 5-fold more substrate to enzymatic reactions. The difference between negative control (inactive protease sample) and positive control (active protease sample) was increased from 3-fold in Figure 3.21 A to 6.2-fold in Figure 3.21 B.

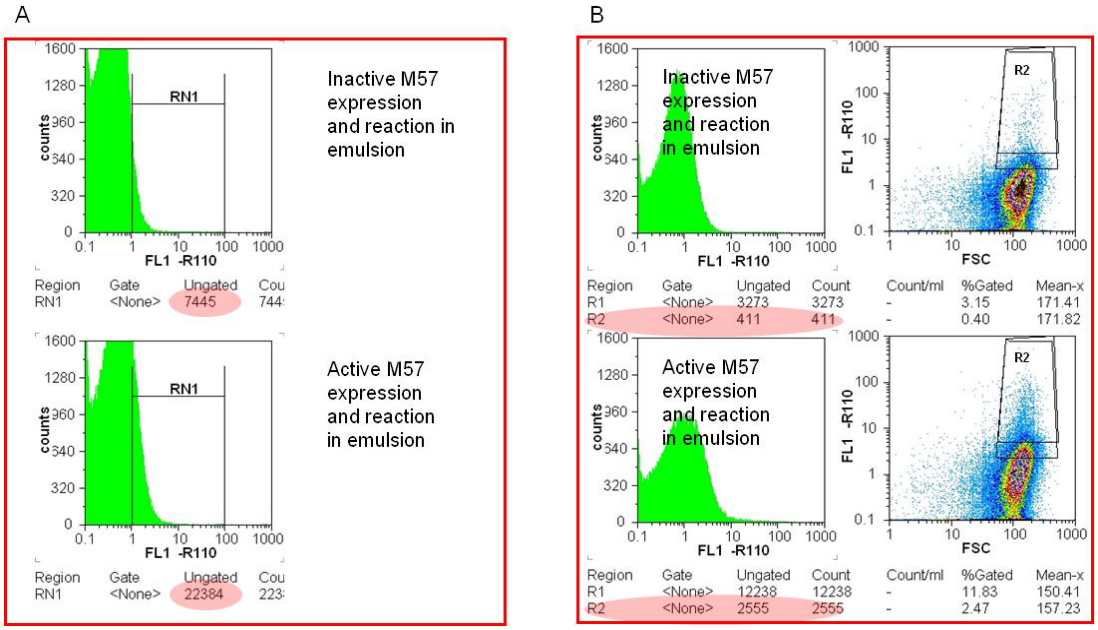


Figure 3.21 CyFlow® space recording of fluorescence signal of optimized ProCF-FCIVC system. A shows pH effect on enzymatic reaction by altering pH with additive TEA; B shows pH effect and inhibition effect on enzymatic reaction by increasing substrate concentration. IVT was carried out using 500 ng of DNA template and substrate bis, Z-VDVAD-R110 (100 µM in A, and 500 µM in B) in emulsions. Enzymatic reactions were carried out inside emulsions. TEA was added through primary w/o emulsion after IVT expression at 37°C for 4 h to alter the pH and incubated at 37°C for another 2 h to perform enzymatic reaction and then second w phase was added and emulsified to double emulsion for CyFlow® space measurement. Parameter FL1 represents green fluorescence intensity from R110 product and FSC represents particle size. Red ellipse highlight represents target fluorescence signal (R110 product).

To prevent protease degradation of expressed protein candidates, IVT kits usually contain protease inhibitors. The only surely non inhibitor contained IVT kit, NEB PURExpress In Vitro Protein Synthesis kit in which all functional proteins for expression were added

from purified proteins, was tested. However the expressed M57 protease using PURExpress kit showed a very low activity due to a low yield. Two other *E. coli* extract made IVT kits (Roche Rapid Translation System, Qiagen EasyXpress Protease Synthesis kit) were then investigated for their inhibitor inhibition effect on the target expressed M57 protease. M57 protease was firstly expressed and subsequently added into original IVT expression buffer or 10-fold diluted expression buffer to perform activity determination. As shown in Figure 3.22 Qiagen kit showed a better activity performance than Roche kit.

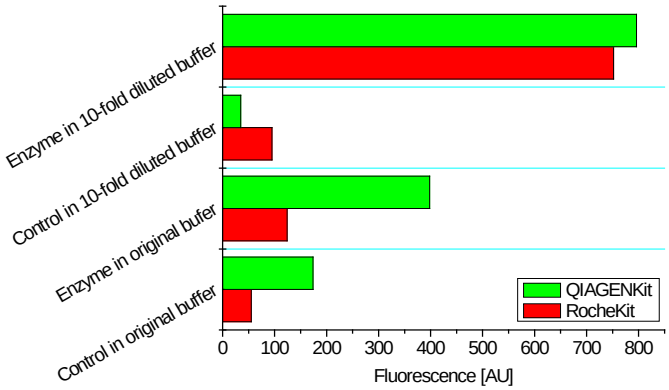


Figure 3.22 Inhibitor effect on protease activity in *in vitro* translation reaction. 25 µL of *in vitro* translation reaction mixtures without DNA (Blank control sample) or with plasmid DNA (protease enzyme sample) were incubated at 30°C (Roche RTS kit) or 37°C (Qiagen EasyXpress kit) for 4 h. Then 2.5 µL of accomplished *in vitro* translation mix was added into 22.5 L of original IVT reaction buffer or 10-fold water diluted IVT reaction buffer to measure enzymatic activity by Tecan SAFIRE.

Above results showed a higher enzymatic activity was detected in water diluted IVT expression buffer. Therefore a dilution of IVT mixture effect on enzymatic reaction was investigated. M57 protease was firstly expressed using Qiagen IVT kit with non diluted, 2-flod diluted or 4-fold diluted expression buffer conditions. Then small amount of the expressed M57 proteases from different expression conditions were added into Tris-HCl buffer for activity analysis to determine expression level (Figure 3.23 A). The rest of expressed protease samples were used to enzymatic activity assay at original diluted expression buffer condition (Figure 3.23 B). Regarding to its expression level and enzymatic reaction rate, 2-fold diluted IVT reaction mix was chosen for expression.

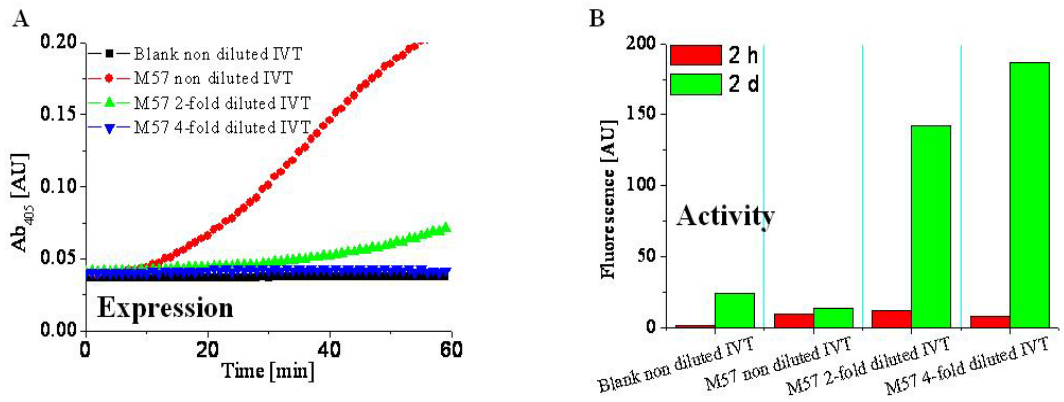


Figure 3.23 Inhibitor effect on protease activity in Qiagen IVT system. *In vitro* translation of protease was performed with initial IVT reaction buffer, 2-fold diluted reaction buffer, or 4-fold diluted reaction buffer in PCR tube. In figure 2A, small amount of IVT translation mix with substrate was added into Tris-HCl buffer (0.1 M, pH 8.6). In figure 2B, fluorogenic substrate was directly added into initial IVT mix and activity was measured at the time points indicated (2h and 2 days).

After all optimization experiments mentioned above, Figure 3.24 shows fluorescence signal of ProCF-FCIVC system with the optimized conditions such as altered pH, increased substrate concentration and diluted IVT reaction mix. Flow cytometry recording data clearly displayed a distinct positive region for sorting. A successful PCR amplification using six days incubated emulsions sample indicated the possibility to sort and recover DNA from a prolonged reaction.

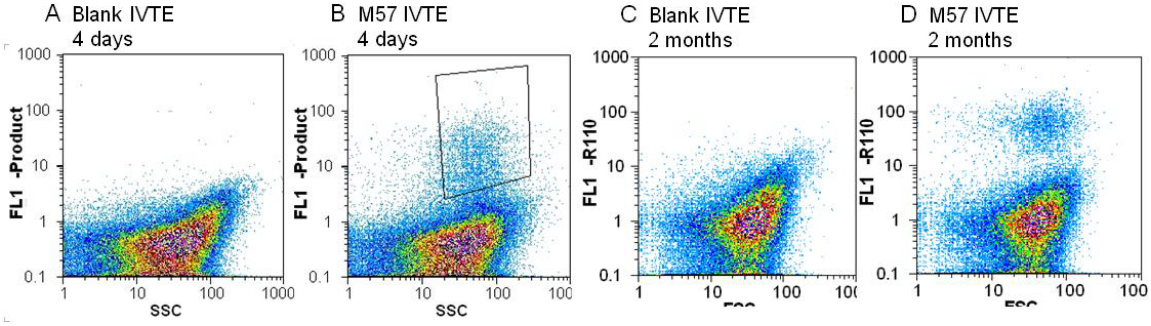


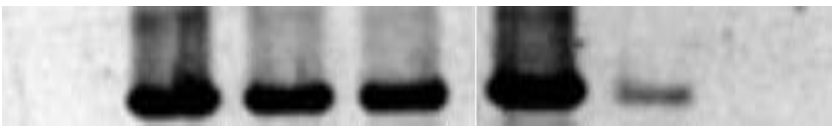
Figure 3.24 CyFlow® space flow cytometer recording of R110 product signal of samples (expression and enzymatic reaction inside emulsion) after 4 days (A, B) or 2 months (C, D) reaction time at 4°C. IVT was carried out using 500 ng of DNA template and substrate (bis, Z-VDVAD-R110, 500 µM) in emulsion. TEA was added through primary w/o emulsion after IVT expression at 37°C for 4 h and incubated at 37°C for another 0.5 h to alter pH and then second w phase was added and then emulsified to double emulsion for CyFlow® space measurement. Parameter FL1 represents green fluorescence intensity from R110 product and SSC represents particle granularity.

3.2.2 Sorting library using ProCF-FCIVC

As described in chapter 3.1.3 due to piezoelectric activation sorter module, a continuous fluid comes out from sorting channel resulting in a dilution problem for sorted solution. Therefore recovery of target DNA from an excess of sorting solution was explored.

To recover low concentration of target DNA from diluted sorting solution, different carriers (or coprecipitants) were examined in this work. The results showed using carriers (linear ployacrylmide and pellet paint) 50 pg of DNA could be recovered from 10 mL sheath fluid and the recovery efficiency was around 20 % (Table 3.9).

Table 3.9 Recovery of DNA using different carriers. LP: linear polyacrylamide, PP: Pellet paint®, PPNF: Pellet paint non fluorescence.

Samples	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
DNA concentration	50 pg/ 10 mL	50 pg/ 10 mL	50 pg/ 10 mL	50 pg/ 10 mL	10 pg/ mL	1 pg/ mL	0	50 pg/ 0.5 mL
Carriers	Non carrier	LP	PP	PPNF	/	/	/	LP
Transformation	0	9	9	4	51	1	/	21
Recovery efficiency	**0**	**18 %**	**18 %**	**8 %**	100 %	2 %		42 %
	-	+	+	+	+	+	-	
PCR amplification								

Plasmid DNA was diluted into sheath fluid at different concentration and recovered by ethanol precipitation in presence of or absence of carriers as shown in table. Transformation was carried out using 1 µL of recovered DNA (assumed as 10 pg of original DNA) or various amount of control DNA (10 pg, 1 pg and blank). PCR amplification was carried out using 1 µL of recovered DNA in 25 µL reaction volume.

To verify the ProCF-FCIVC system, reference libraries with different percentage of active DNA template were generated. The verification was supposed to be performed by sorting target active DNA and determining the active DNA precentage by DNA gel. Figure 3.25 shows the determination results of preparation of reference libraries and detection sensitivities of this system. DNA precentage could be distinguished by PCR

and the minimal amount of DNA for PCR amplification was about 0.1 pg (~10^5 molecular copies of M57 plasmid DNA).

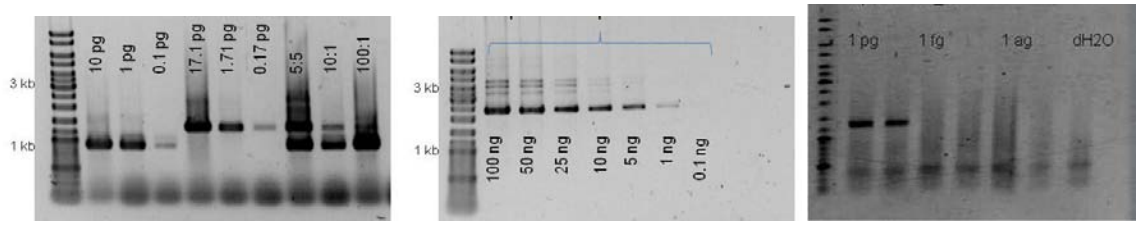


Figure 3.25 Preparation of reference libraries for verifying ProCF-FCIVC system. Figure (left one) shows reference libraries with 50 %, 10 % and 1 % of active DNA in total. Figure (middle one) shows detection sensitive of DNA gel by SYBR Gold dye. Figure (right one) indicates detection level of minimal amount of DNA template for PCR amplification.

In sorting experiment the 1 % active DNA reference library was expressed in emulsion using conditions optimized above. Figure 3.26 shows the target signal region (R1) for sorting. Aound 1 mL (< 10^4 events, around 10^5 copies DNA) of sorted solution was precipitated together linear polyacrylamide and then tranformated into *E. coli* electroporation competent cells. Two clones were grown on LB milk detection agar plate showing haloes while non clones grown on negative control plate indicating the proof of concept of ProCF-FCIVC.

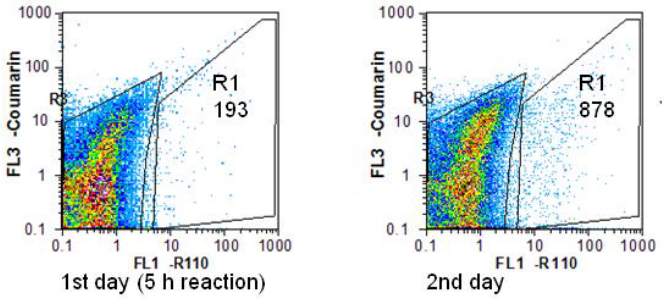


Figure 3.26 CyFlow® space recording of fluorescence signal of reference library. IVT was carried out using 500 ng of DNA template reference library (pET-M57 and pIX-delM57) and 500 µM of substrate (bis, Z-VDVAD-R110) in emulsion. TEA was added through primary w/o emulsion after IVT expression at 37°C for 4 h and incubated at 37°C for another 0.5 h to alter pH and then second w phase was added and then emulsified to double emulsions and subsequently incubated at 37°C for another 5 h to perform enzymatic reaction. R1 represents target events for sorting. Events indicated in R1 region was counted from first 10^5 analyzed events.

Due to dilution limitation of CyFlow® space the ProCF-FCIVC system was not verified by sorting a mutant library. In this case verification was recommended to be performed using a flow cytometry machine without dilution demerit.

3.3 Whole-cell based protease screening system (Pro-FCIVC)

As cells can be easily recovered from sorting solution a development of whole-cell based screening system was then explored in this work. The whole-cell based screening system was named as Pro-FCIVC (Protease flow cytometry and IVC screening system). The development of Pro-FCIVC screening system comprised of same elements described as in ProCF-FCIVC system including a) selection of substrate, b) optimization of fluorescence signal for sorting, c) verification of Pro-FCIVC by sorting a mutant library and subsequent experiment d) characterization of protease variants.

3.3.1 Pro-FCIVC using *E. coli* cells

Due to high transformation efficiencies, *E. coli* cells based whole-cell screening system was explored first in this work using the two model proteases M46 and M57 expressed in pET systems.

As describe in chapter 3.1.2 fluorogenic substrate bis, Z-VDVAD-R110 was examined for the whole-cell based flow cytometry screening system. The results showed bis, Z-VDVAD-R110 gave an evidently high difference of enzymatic activities between negative control (cultivation supernatant from cells carrying empty vector) and positive control (cultivation supernatant from cells carrying proteases) in microtiter plate format. To determine the performance of enzymatic reaction and the signal difference afterwards under emulsion conditions, enzymatic reactions were carried out in emulsions using ultrasonication treated cell lysate containing intracellular expressed proteases. Figure 3.27 shows a distinct fluorescence signal between negative control (Figure 3.27 A) and positive control at various reaction times (Figure 3.27 B, C) could be observed.

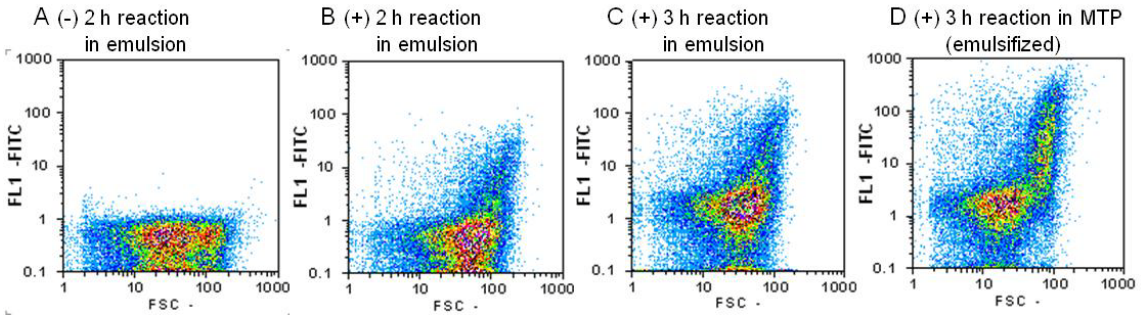


Figure 3.27 Determination of enzymatic reaction of protease using bis, Z-VDVAD-R110 and *E. coli* expression system. Enzymatic reaction was carried out in emulsion using 100 µM substrate and cell lysate. A represents signal from negative sample (lysate from cells carrying empty vector), B and C represents signal from positive samples (lyaste from cells carrying M46 protease) at different reaction times. D represents signal from positive sample (as B and C) which had its enzymatic reaction firstly in microtiter plate and then the accomplished reaction mixture was emulsified for flow cytometry measurement.

For whole-cell based screening system, cells have to be intact to prevent a diffusion of proteases. Activity analysis of cultivated supernatant was carried out to analyze the cell integrity. Results showed cells carrying M57 protease tended to autolysis in few hours after IPTG induction. SDS-PAGE analysis (Figure 3.28) of total cell proteins confirmed the hydrolysis of M57 protease cells, thereby M46 protease was chosen in further optimization.

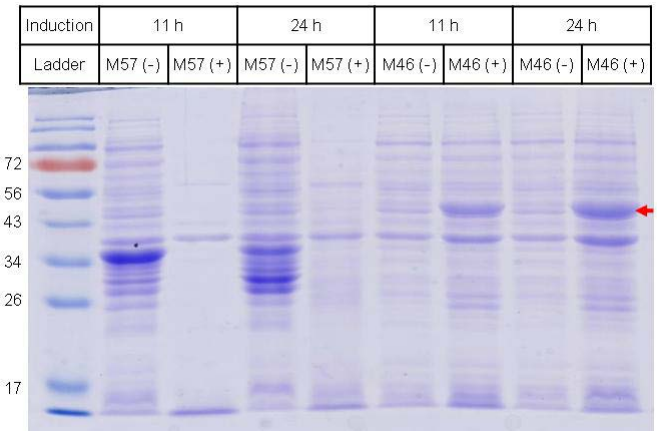


Figure 3.28 Expression of M57 and M46 protease in pET system. 20 µL of expressed cell culture was loaded in 12 % SDS-PAGE to determine protease expression. Red arrow indicates expressed M46 protease.

In order to get high fluorescence signal, further optimization such as cell membrane penetrated reagent polymyxin B sulfate (PB), reaction temperature and reaction pH were

explored for M46 protease variant (data not shown). The optimized condition were performed in Tris-HCl buffer (50 mM, pH 8.6) with PB (144 µM) and bis, Z-VDVAD-R110 (100 µM) at 37°C for 2-4 h. Figure 3.29 shows after optimization a detectable increase of fluorescence signal could be observed in positive and negative samples.

Due to the high background noise which might be caused by other proteases existing naturally inside cells, *E. coli* cells employed an intercellular heterologous protease expression system is not a suitable for flow cytometry (a high sensitive detection system) based screening system for proteases. To minimize the background caused by non target intercellular proteases, a ready to use secretion system employed extracellular proteases deficient strain (*Bacillus* cells) and Subtilisin Carlsberg was chosen for the development of Pro-FCIVC screening system afterwards.

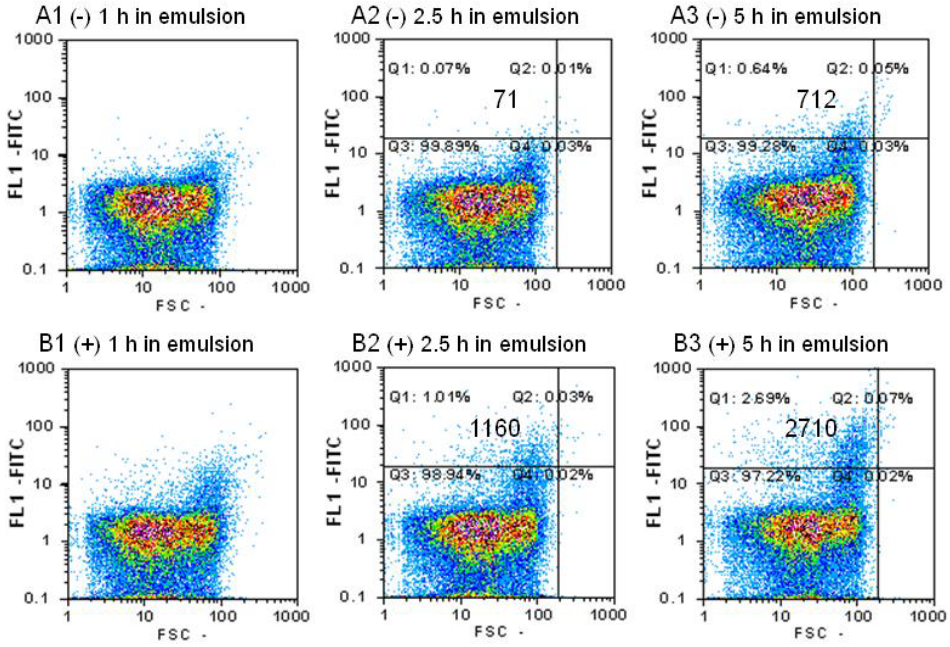


Figure 3.29 Flow cytometry recording to fluorescence signal of optimized screening system based on M46 protease in E. coli cells. Enzymatic reactions employed $2x10^9$ cells, bis, Z-VDVAD-R110 (100 µM) and Tris-HCl reaction buffer (50 mM, pH 8.6). Reactions were carried out at 37°C for certain time. Q1 region represents the events with high green fluorescence signal.

3.3.2 Development of Pro-FCIVC using *Bacillus* cells

In this section several key factors for the development of *Bacillus* cells based Pro-FCIVC screening system, especially like improvement of fluorescence signal of system, recovery

of sorted positive events and verification of system by screening epPCR mutant library are addressed.

3.3.2.1 Optimization of enzymatic reaction and sorting condition

For *Bacillus* cells based system, besides bis, Z-VDVAD-R110, another R110 based serine protease specific substrate bis, Suc-AAPF-R110 was also tested. As shown in Figure 3.30 fluorescence microscopy detection revealed fluorescent product (R110) is solely located within the inner aqueous phase of the emulsion (w1) illustrating that the conversion of substrate to product reaction only takes place inside the generated compartments. No considerable diffusion into the surrounding phase could be observed disclosing serine protease specific substrate (bis, Suc-AAPF-R110) was suitable for the emulsion enzymatic reaction condition.

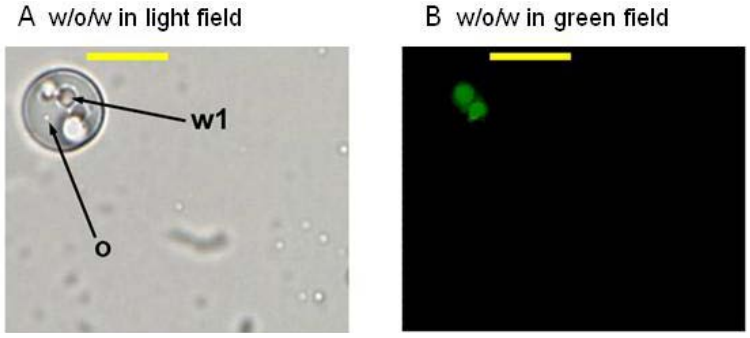


Figure 3.30 Fluorescence microscopy of double emulsions at light field (A) and green fluorescence field (B) after enzymatic conversion of the fluorogenic substrate bis, Suc-AAPF-R110 by Subtilisin Carlsberg (SC) at 37°C for 4 h. w1 and o indicate inner aqueous phase and oil phase, respectively. The marker bar represents a length of 10 µm. Enzymatic reactions were carried out in the aqueous phase (w1) of the w/o/w double emulsions.

Fluorescence signal optimization experiments were carried out firstly in microtiter plate format then in emulsions condition. *B. subtilis* WB600, a six extracellular proteases deficient strain, was used as a host due to its non detectable background activity in activity assay (MTP condition). Results showed both R110 based substrates could provide a high distinct fluorescence signal between positive and negative samples and the enzymatic rate using bis, Suc-AAPF-R110 was more than 60-fold higher than bis, Z-VDVAD-R110 substrate. In second optimization step, the enzymatic reaction inside w/o/w double emulsion compartments was optimized for varied reaction buffers, pH,

temperature and reaction time. The results showed alkaline pH and high temperature increased the activity of SC. Testing of two substrates showed bis, Z-VDVAD-R110 gave a remarkably increased fluorescence signal when triethylamine (TEA) was added while bis, Suc-AAPF-R110 only slightly improved protease activity by adding TEA. However TEA caused an approximately 10 % remained cell viability compared to non TEA condition. To keep cell viable after enzymatic reaction and sorting, Tris-HCl buffer with a moderate pH 8.6 and temperature 37°C were chosen because they only led slight effects on cell viability under several hours reaction time whereas they resulted in a detectable fluorescence signal. In this case bis, Suc-AAPF-R110 performed a better signal than bis, Z-VDVAD so it was used for subsequent experiments.

During exploring experiments of sorting with the selected substrate (bis, Suc-AAPF-R110) found that an elimination of residual protease activity is an important prerequisite for an IVC screening technology for proteases when using whole *Bacillus* cells as host. Therefore background activity of a deficient *B. subtilis* WB800N, which has two more extracellular proteases, was further investigated. Compared to *B. subtilis* WB600 (*nprE nprB aprE epr mpr bpr*), *B. subtilis* WB800N has two additional extracellular proteases (*vpr wprA*) knocked out. Figure 3.31 shows the corresponding fluorescence signals of the released R110 product. Despite the knockouts fluorescence signals could be clearly detected proving the superb sensitivity of the Pro-FCIVC detection system. In comparison *B. subtilis* WB800N revealed a 6.7-fold lower background compared to *B. subtilis* WB600. The difference of signal intensities between protease expressing cells and the empty vector control additionally proved to be by far more distinct for *B. subtilis* WB800N (26-fold) than for *B. subtilis* WB600 (3-fold). Figure 3.31 C displays a visual illustration of the difference by overlaying samples in one FL1 histogram where the M1 marker represents the district for positive events. Due to the observed signal properties *B. subtilis* WB800N proved to be a potentially suitable host for directed protease evolution although exponential growth of this strain is ~2-fold slower than that of *B. subtilis* WB600.

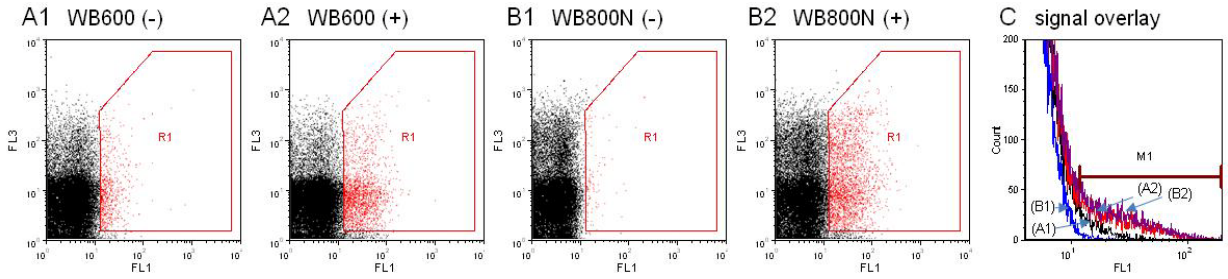


Figure 3.31 Flow cytometer recordings of fluorescence signals of protease deficient *B. subtilis* strains entrapped in w/o/w double emulsions. Strains harbouring pBCSSC express Subtilisin Carlsberg (+) while strains harbouring pBCS served as negative controls (-). A1: *B. subtilis* WB600 [pBCS], A2: *B. subtilis* WB600 [pBCSSC], B1: *B. subtilis* WB800N [pBCS], and B2: *B. subtilis* WB800N [pBCSSC]. FL1 represents green fluorescence of the released R110 product, FL3 represents blue fluorescence of the internal dye 7-hydroxycoumarin-3-carboxylic acid. The positive events in the R1 gate region are 647, 1972, 97, and 2548 out of 100,000 total events respectively. C: FL1 histogram as overlay of the signals from all 4 samples (A1, A2, B1, B2) used to determine the background activities of the tested host strains. The M1 marker represents the positive events region.

During Pro-FCIVC sorting a growth competition was observed when cells were directly subjected to liquid cultivation after the sorting process. To depress the competition problem caused by heterologously expressed protease, inducible vector system was investigated in the work. The results showed the expression levels of SC using IPTG or xylose inducible systems in *B. subtilis* were reduced a lot compared to the constitutive expression system (Figure 3.32). SDS-PAGE analysis revealed SC protease yield in inducible vector systems were much lower than in constitutional expression system using pBCS vector (Figure 3.33).

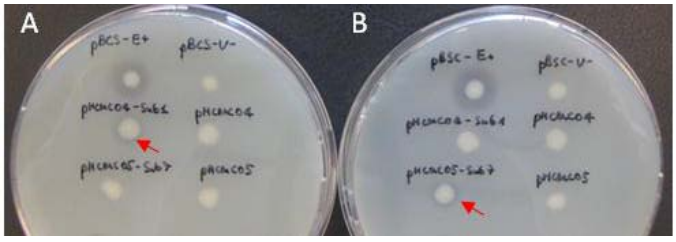


Figure 3.32 Visualization of protease activities in inducible plasmids based on pHCMC series. *B. Subtilis* strains were cultivated for 16 h on 1 % skim milk indicator plates containing either 0.2 % xylose (A) or 0.2 mM IPTG (B). Activities of induced proteases from pHCMC04-sub1 and pHCMC05-sub7 were indicated by red arrows.

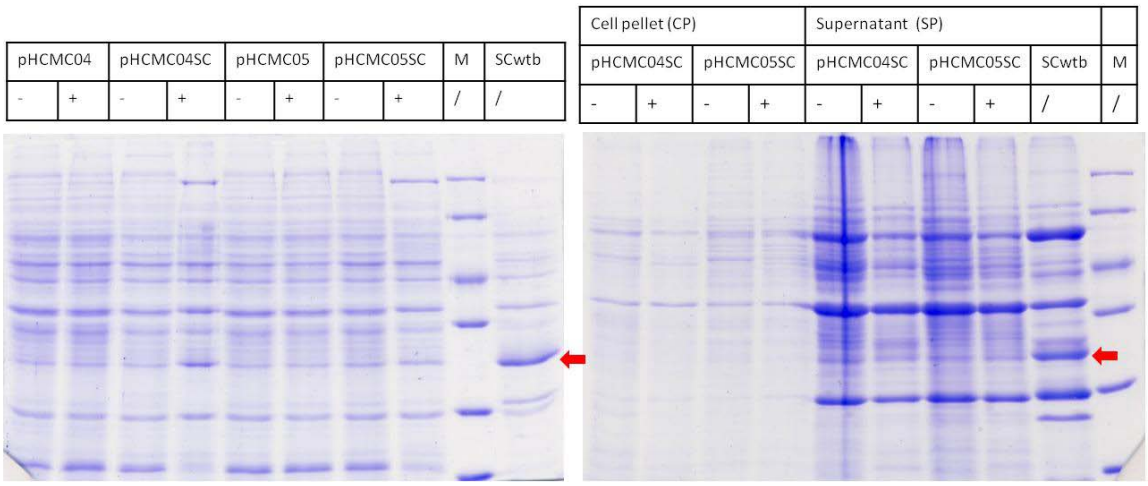


Figure 3.33 Expression of SC using different expression systems. *B. Subtilisin* WB600 strains carrying different plasmids were grown in LB medium with appropriate antibiotic at 37°C to mid-exponential growth phase and divided into individual tubes for further grown without induction (indicated as -) or with induction (0.5 % xylose or 1 mM IPTG, indicated as +). Samples were taken after 3 h (left SDS gel) or 21 h (right SDS gel) after induction. Mature SC is indicated by red arrow.

Due to a dilution in sorting process, sorted cells need to be recovered prior to next round of sorting. Experiment results showed membrane filtration was more efficient (around 3-fold) than centrifugation using diluted emulsion solution (~10^4 cells encapsulated). During sorting condition optimization experiment, a decrease of enrichment (active cells population of sorted samples) was observed. Active cells ratio was dropped down 3-5 folds after inoculation of a small amount of cells directly in liquid media. As shown in Figure 3.34 a rich inoculum (>10^5 cells) is a prerequisite for preserving enrichment factors achieved in previous rounds of sorting. Due to this limitation sorted variants (usually ~10^4 variants) were recovered firstly by vacuum filtration with a sterile membrane and subsequently grown on LB agar plates prior to liquid cultivation for the next round of sorting.

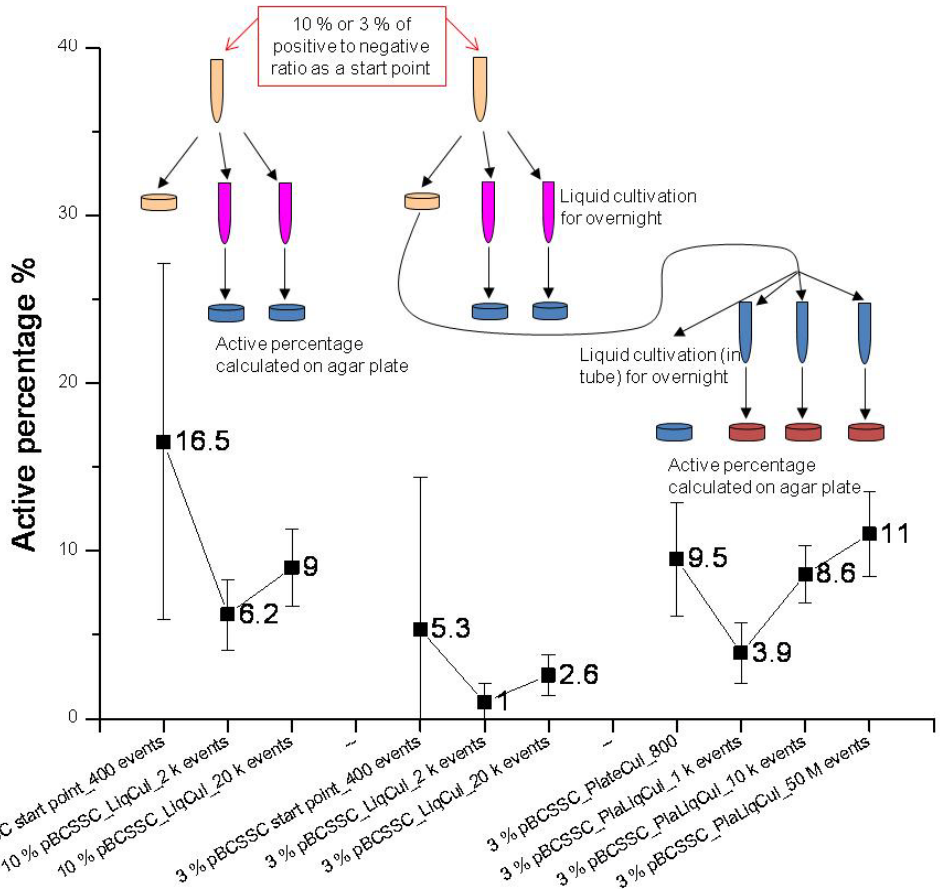


Figure 3.34 Investigation of cultivation condition effect on active cells ratio in inoculated culture. *Bacillus* cells were cultivated on agar plate or liquid media to analysis active cells ratio after cultivation. Cultivation was carried out in 3 mL LB appropriate antibiotics in 15 mL of glass tube. Active cells percentage was calculated on LB milk agar plate by halo formation.

3.3.2.2 Verification of Pro-FCIVC by sorting an epPCR mutant library

After optimization the Pro-FCIVC screening system was verified by screening an epPCR library in which the mutational load was adjusted to 2-3 % of active clones with high mutational loads.

In the experiment three mutant libraries of Subtilisin Carlsberg protease (SC) were generated by epPCR using by varying $MnCl_2$ concentrations (0.1 mM, 0.2 mM and 0.4 mM) and their determined active ratios were around 85 %, 50 % and 2.3 %, respectively. The high mutational loads library (0.4 mM $MnCl_2$ generating ~2.3 % of active clones) was finally selected to generate the mutant library for screening. *Bacillus* strains usually

display transformation rates <10^5 per µg DNA. To provide a library with an adequate size a two-step transformation procedure as described in materials and methods was performed using *E. coli* for efficient subcloning. Individual colonies harbouring protease mutants were pooled and the corresponding plasmids were isolated and after rolling circle amplification transformed into *B. subtilis* WB800N. The size of the generated mutant library was approximately 10^5-10^6 and Subtilisin Carlsberg variants were finally screened towards increased inhibitor resistance (antipain-HCl) to validate the developed Pro-FCIVC screening system.

As described in chapter 3.1.3, the primary investigation experiments revealed an existence of 'escapee' negative events present in the sorting channel. A minimal sorting rate of 5-10 events/s at an analysis rate of ~8,000 events/s using 100-fold diluted w/o/w double emulsions in PBS buffer was required to control this 'leakage' in a low range. For each round of sorting in total ~10^7 cells were screened using an input speed of 1-2 µL/s. After three iterative rounds of screening and sorting the library could be enriched approximately 12-fold (active ratio was increased from ~2.5 % to 30 %). Figure 3.35 shows the fluorescence signals of different sorting stages illustrating the achieved enrichment.

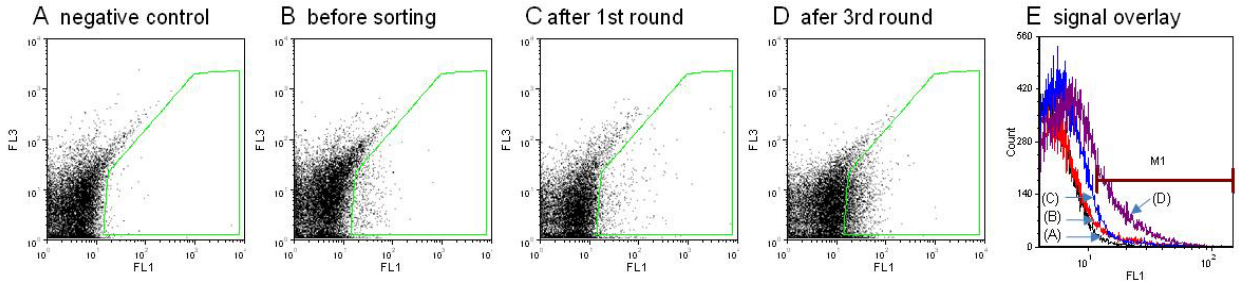


Figure 3.35 Flow cytometer recording of fluorescence signals of the compartmentalized Subtilisin Carlsberg mutant library in B. subtilis during different stages of sorting. A: empty vector control, B: before sorting, C: after 1st round of sorting and D: after 3rd round of sorting. The M1 marker represents droplets with a positive fluorescence signal for sorting. FL1 represents green fluorescence of the released R110 product, FL3 represents blue fluorescence of the internal dye 7-hydroxycoumarin-3-carboxylic acid. The positive events in the polygon gate region are 33, 368, 375 and 1664 out of 100,000 total events, respectively; e: FL1 histogram as overlay of the signals from all 4 samples (A, B, C, D).

After the last round of sorting a small fraction of sorted cells was plated on LB skim milk agar plates and 15 active clones with the largest ratios of halo/colony size were selected

from in total 315 active colonies for rescreening. From the tested clones, 4 exhibited either increased inhibitor resistance (SCm2, SCm13) or a higher activity (SCm3, SCm5) than the original wild-type and therefore were chosen for sequence analysis. Sequencing results revealed that variants SCm3 and SCm5 with increased activity were wild-type while mutants SCm2 and SCm13, which displayed an increased residual activity in the presence of antipain-HCl (~2.1 and 1.2-fold, respectively), contained multiple mutations leading to amino acid substitutions. Additionally several silent mutations occurred on DNA level of SCm2 and SCm13. Both variants showed a decreased overall activity compared to wild-type SC in absence of the inhibitor (Table 3.10).

Table 3.10 Inhibitor resistance and amino acid changes of Pro-FCIVC sorted variants of Subtilisin Carlsberg (SC) compared to the wild-type enzyme.

Variant	**Activity in absence of inhibitor [nM/min]**	**Activity in presence of inhibitor [nM/min]**	**[a]Residual activity of variants [%]**	**[b]Relative inhibitor resistance compared to wild-type [%]**	**Relative activity compared to wild-type [%]**	**[c]Sequence analysis**
SCwtb	33.45 ± 0.75	4.84 ± 0.10	14.5 ± 0.1	100	100	wild-type
SCm2	12.97 ± 0.36	4.00 ± 0.18	30.8 ± 0.7	213 ± 4	39 ± 2	K127R, T237P, M239I, I269V, Y310F, I372V
SCm13	25.63 ± 0.30	4.61 ± 0.14	18.0 ± 0.4	124 ± 3	77 ± 3	F15L, V18A, T50A, K64R, E82G, Q124R, M223V

[a]Residual activity was calculated as ratio of activity in presence of antipain-HCl to activity in absence of antipain-HCl. [b]The relative inhibitor resistance was then calculated as ratio of the residual activity of the variant to the residual activity of the wild-type. Reactions were carried out in Tris-HCl (0.1 M, pH 8.6), fluorogenic substrate bis, Suc-AAPF-R110 (5 µM) and 1 µL of unnormalized crude supernatant of *B. subtils* WB800N in the presence or absence of antipain-HCl (1 µM). Standard deviations are obtained from triplicate measurements. [c]Amino acid numbering begins at the start codon.

3.3.2.3 Characterization of protease variants

As shown in Table 3.10, inhibitor resistance ability was calculated by the ratio of residual activity ratio in presence of to absence of antipain-HC of variants' compared to wild-type' (Mutant $_{\text{residual activity}}$/Wild-type $_{\text{residual activity}}$). Variant SCm2, which had 2.1-fold

increased inhibition resistance to antipian-HCl reversible inhibitor, was further purified together SCwtb for kinetic and resistance characterization.

The purification was performed by using a two-step exchange procedure (cation exchange followed by anion exchange). SDS-PAGE analysis (Figure 3.36) showed purification was successful and product (purified SC mature protease) had a high purity.

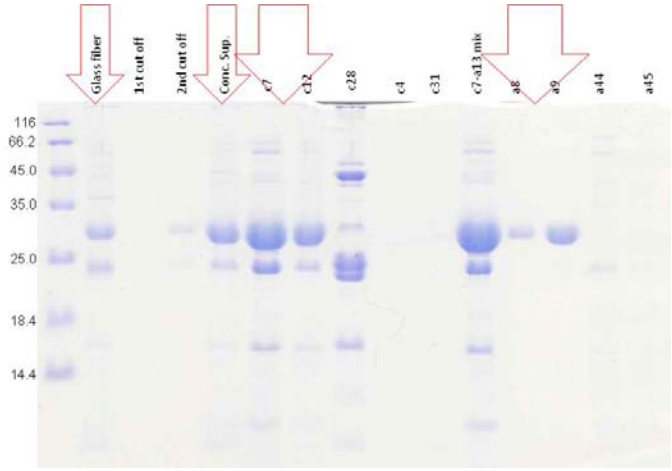


Figure 3.36 SDS-PAGE (15 % gel) analysis of protein samples from steps. For each samples, 250 µL of supernatant/purified solution was prepared by TCA precipitation and dissolved in 20 µL of Tris-HCl (0.1 M, pH 8.0) and 5 µL loading dye. 2.5 µL pre-heated (95°C for 5 min) samples were loaded to SDS-PAGE for analysis. Red unfilled arrow represents the one containing mature Subtilisin Carlsberg protease. Letter c and c represent samples from cation and anion column respectively.

Activity analysis of purified mature SC protease from two ion exchange columns indicated the mature SC protease couldn't bind to both of columns (data shown in appendix). According to previous research (Vitale and Gamulin 1975) and protein amino acid analysis mature SC protease has an isoelectric point (pI) around 6.56 indicating SC mature protease should unbind to cation exchange column but bind to anion exchange column using Tris-HCl bffer pH 8.6 as running buffer. However the purification experiments showed the SC mature protease flowed through anion exchange column and didn't bind at all. a44, a45 eluted protein samples (data shown in appendix) by salt gradient washing (50 mM Tris-HCl pH 8.6 with 2 M NaCl) was yellow color and turned out to be impurity protein. SDS-PAGE analysis (Figure 3.36) confirmed the results.

To find out running buffer effect on purification (Subtilisin Carlsberg binding ability to anion column), five different running buffers were examined in small eppendorf tube scale. Four groups of samples were analyzed by protease activity assay. Group 1 (running buffer) and group 2 (elution buffer) were employed as control assay condition to determine protease activity in applied buffer/salt conditions. Group 3 showed activities of protease samples passed through anion resin and group 4 showed activities of protease samples eluted from anion resin. The results showed in most of tested running buffers (except 5 mM Tris-HCl pH 8.6 running) SC mature protease could not bind to anion resin.

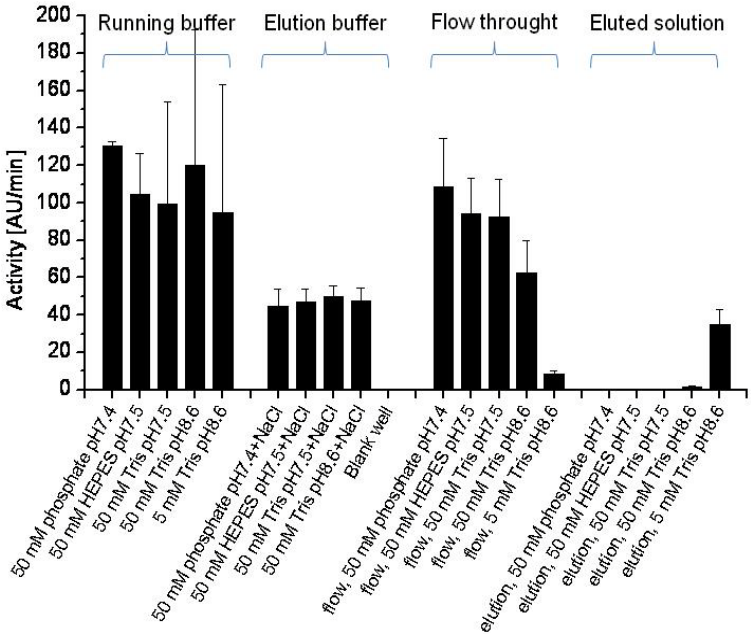


Figure 3.37 Running buffer effect on purification of Substrate Carlsberg using anion exchange resin (Toyopearl SuperQ-650C). Assay was carried out in Tris-HCl buffer (0.1 M, pH 8.6) using protease samples from four groups. Group 1 (running buffer) was carried out at different running buffer, group 2 (elution buffer) was carried out at different elution buffer i.e. running buffer added 2 M NaCl, group 3 (flow through) was carried out using proteases flowed through resin and group 4 (eluted solution) was carried out using proteases eluted out from resin.

To get a 'normal' purified Subtilisin Carlsberg which could bind to anion column and be eluted out using high salt buffer, the two-step purification were performed again but using Tris-HCl (5 mM, pH 8.6) as running buffer for the anion column. The purification results showed target SC still flowed through the anion exchanger (ion exchange

chromatogram shown in appendix). SDS-PAGE (Figure 3.38) and Agilent analysis (Figure 3.39) showed purified SC variants also had a high purity.

Figure 3.38 SDS-PAGE (15 %) analysis of purification results by a two-step ion exchange purification of Subtlisin Carlsberg variants. For each samples, 250 µL of supernatant/purified solution was prepared by TCA precipitation and dissolved in 20 µL of Tris-HCl (0.1 M, pH 8.0) and 5 µL loading dye. 2.5 µL pre-heated (95°C for 5 min) samples were loaded to SDS-PAGE for analysis. SC variants were SCm13, SCm2 and SCwtb.

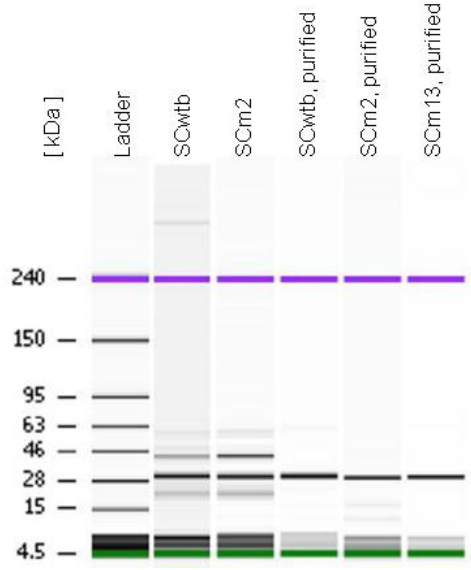


Figure 3.39 Agilent analysis of Subtislin Carlsberg proteases before purification and after purification. The percentages of purified Subtilisin Carlsberg variants in total proteins were 95.4 %, 90.3 % and 96.2 % respectively.

After purification kinetic characterization was performed using wild-type SC (SCwtb) and inhibitor resistance improved mutant variant SCm2. Variant SCm2 showed a slightly decreased Km and Kmobs compared to SCwtb (Table 3.11). SCm2 also had a higher calculated dissociation constant (Ki) for binding inhibitor to enzyme, which was consistent with resistance analysis by plotting residual activity of varying concentrations of antipain-HCl inhibitor using 5 µM of bis, Suc-AAPF-R110, 5 µM (see Figure 3.40).

Table 3.11 Kinetic characterization of wild-type SCwtb and mutant SCm2 variants.

Variants	**Km [µM][a]**	**Vm [µM/min][a]**	**Kmobs [µM][b]**	**Vmobs [µM/min][b]**	**Ki [µM][c]**
SCwtb	16.46 ± 3.74	1.30 ± 0.10	83.60 ± 21.99	0.37 ± 0.03	0.2452
SCm2	15.30 ± 3.55	0.19 ± 0.01	76.98 ± 20.32	0.07 ± 0.01	0.2480

[a]Km and Vm were characterized in Tris-HCl (pH 8.6, 0.1 M), substrate bis, Suc-AAPF-rhodamine 110 and 100 nM of purified SC variants in COS384fb microtiter plate at 37°C. [b]Kmobs and Vmobs represents observed Km and Vm in presence of antipain-HCl inhibitor (1 µM). [c]Ki was calculated by formula Ki=[inhibitor]/(Kmobs/Km-1).

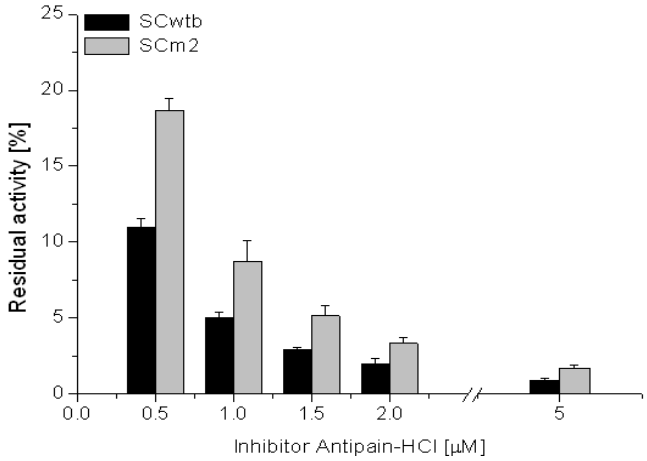


Figure 3.40 Comparison of antipain-HCl (protease inhibitor) resistance by wild-type SCwtb and mutant SCm2 at different concentrations of inhibitor. Reactions were carried out in Tris-HCl (pH8.6, 0.1 M), fluorogenic substrate bis, Suc-AAPF-R110 (5 µM) and 100 nM of purified SC variants in COS384fb microtiter plate at 37°C. Residual activity was calculated from the activity of in the presence of antipian-HCl to the activity in the absence of antipain-HCl.

Figure 3.41 shows the positions of the obtained amino acid substitutions of variant SCm2. The corresponding 6 residues in the wild-type enzyme were further substituted

individually by site-directed mutagenesis (SDM) to explore a contribution of each mutated position to the observed resistance property.

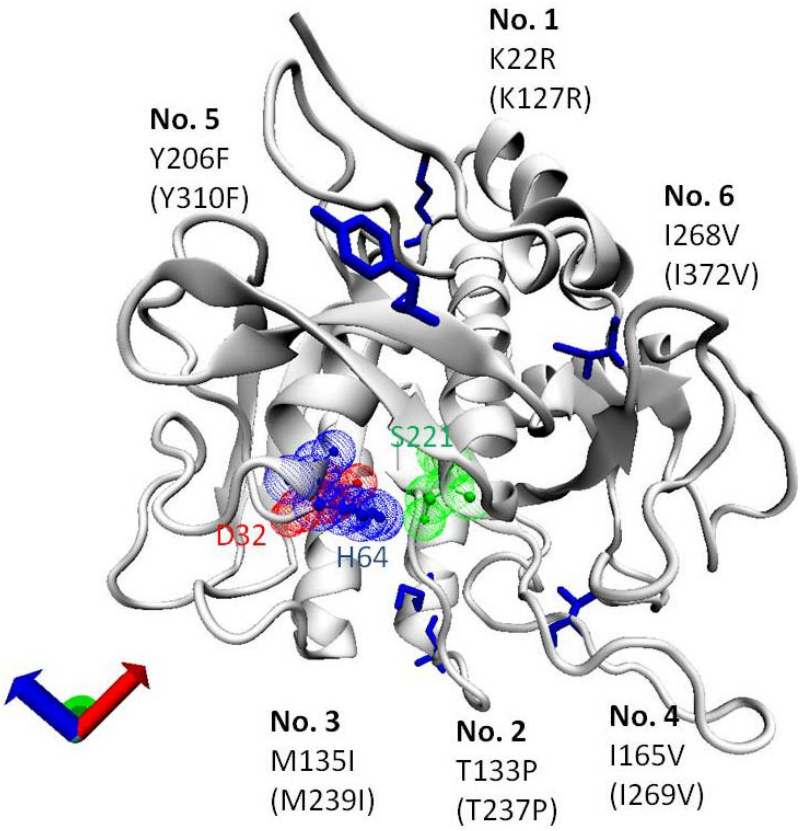


Figure 3.41 Positions of amino acid substitutions in mature Subtilisin Carlsberg variant SCm2 displaying an increased protease inhibitor resistance. Substitutions are indicated by arrows and numbers in brackets specify the corresponding position counting from the start codon. D32, H64 and S221 form the catalytic triade in the active site of the protein. Discrepancies in numbering of labeled mutation positions (No.2-No.6) was caused by numeration of used model (1c3lA) where all amino acids starting from position 56 are shifted afterwards by one position (missing position 56).

In all assays SC concentrations were normalized to avoid false results due to differential expression of the analyzed variants. Activity assay results showed, after normalization the SCm2 variant (6 mutations) still proved to display a higher (1.6-fold compared to the wild-type) inhibitor resistance than the mutants with single amino acid substitutions. Except for variant SCm2SDM3 (M239I) which exhibited a 1.2-fold increased inhibitor resistance compared to wild-type enzyme, all other single mutants did not show any improvements indicating there was no additive effect on the single mutations.

4 Discussion

4.1 Set-up of flow cytometry based screening systems

Due to a high sensitivity and throughput (10^8 variants per day), flow cytometry has recently emerged to be a powerful tool for directed evolution of enzymes. As other evolution systems a link between genotype (gene of target enzyme that can be replicated) and phenotype (a functional trait such as binding or catalytic activity) is also required in a flow cytometry based screening system. Based on the way to link genotype of phenotype of employed variant, flow cytometry based screening systems could be classified into three modes: 1) cell surface display of active enzymes, 2) entrapment of product within cells, and 3) *in vitro* compartmentalization (IVC). Because of autolysis of proteases, cell surface display is not recommended except the proteases are naturally surface displayed. A notable example for a high throughput screening of protease variants is the directed evolution of the surface membrane protease OmpT to alter its substrate specificities by employing an E. coli OmpT deficient strain (UT5600). Improved variants were identified using flow cytometry in aqueous solution without employing any compartmentalization technology (Pogson, Georgiou et al. 2009). Considering background caused by other proteases existing naturally in cells, entrapment of product within cells was also not suitable. As a contrast IVC (a man-made 'cell environment') can performed enzymatic reactions in independent compartments therefore supplies a way to link genes and their encoded proteins (Taly, Kelly et al. 2007). In this work IVC is prepared by water-in-oil-in-water (w/o/w) double emulsion. The second water phase of double emulsion enables a compatible form for flow cytometry analysis and sorting.

The w/o/w double emulsions were generated as described previously (Miller, Bernath et al. 2006). Three different emulsification systems, Span/Tween, Decane/Tween and ABIL EM 90/CMC, were explored. Although a successful approach was obtained in other research work (Mastrobattista, Taly et al. 2005), a precipitation phenomenon and diffusion of coumarin fluorescence dye was observed from the emulsions made by Decane/Tween system. In contrast ABIL EM 90/CMC employed w/o/w double emulsions were very stable over a prolonged period at 4°C and 37°C. Previous research

showed the silicone-based surfactant ABIL EM 90 prepared w/o emulsion is compatible with a coupled rabbit reticulocyte *in vitro* expression system (Ghadessy and Holliger 2004). Therefore ABIL EM 90/CMC system is also supposed to be compatible with a coupled *in vitro* transcription/translation *E. coli* extract expression system and should be tested for cell-free system based protease screening system.

The fluorescence signal in IVC compartments is an essential element to analyze and sort events by flow cytometry. The fluorescence signal can be obtained from enzymatic conversion of fluorogenic substrate to fluorescent product. The retention of fluorescent product inside w/o/w emulsion is required to IVC system. Additionally the retention of product can avoid a cross talk between individual compartments or a reduction of the fluorescence signal by diffusion of substrates and/or products into the surrounding medium resulting in a successful IVC screening system. In a w/o/w double emulsion based IVC system, as the inner aqueous phase is surrounded by a hydrophobic oil phase, substrate/product should be hydrophilic enough to remain inside droplets (Link, Jeong et al. 2007). Results showed that fluorophore compound rhodamine 110 and its bisamide derivatives bis, Z-VDVD-R110 and bis, Suc-AAPF-R110 turned to be suitable to the study. Investigation results demonstrated that R110 had a high intense fluorescence signal and the fluorescence intensity of R110 was constant in a pH range from pH 7-9 at 25°C to 37°C with a perfect detection linear range (r^2=1) when R110 concentration was below 2.5 µM.

To investigate and optimize sorting function of flow cytometer (Partec CyFlow® space) used in the project a three-step calibration procedure was set up. First a fast check of machine condition was performed using standard references material calibration beads. Then sorting condition calibration was carried out by cell size difference. In the last sorting performance using w/o/w double emulsion samples were investigated and optimized. The whole setup was specially designed to green fluorescence related sorting process which is compatible with the target product R110 from enzymatic reaction.

In previous research an internal reference dye (7-hydroxycoumarin-3-carboxylic acid) was co-compartmentalized within enzymatic reaction system so that the fluorescence of internal dye and product could be displayed as dot-plots. Only double emulsions, which

contained internal coumarin dye displaying a blue fluorescence, were analyzed (Mastrobattista, Taly et al. 2005). Use internal dye as an analysis trigger (parameter to control events to be analyzed and displayed on graph) can increase analysis rate and improve sorting performance. Because of the constitution of CyFlow® space only forward scatter, side scatter and green fluorescence signal could be used as the analysis trigger.

4.2 Cell-free screening system for proteases

Considering a convenient way to generate a library with high diversity as well as low background caused by other non-target proteases, the cell-free *in vitro* translation based screening system (named ProCF-FCIVC) was optimized in this work. Totally five different template systems (pET, pIX, pIVEX, Cosmid and Linear template) and three commercial *in vitro* expression kits (Qiagen, Roche, NEB) were investigated. The NEB kit is the most pure kit containing only synthesized essential proteins for expression, but the best expression was obtained by using pET template system and Qiagen *in vitro* expression kit.

A big challenge regarding to proteases *in vitro* expression is how to prevent hydrolytic activities caused by *in vitro* expressed proteases during expression time and later on to improve enzymatic activities after expression for a good analysis signal. Protease for example subtilase protease M57 used in this work has a low activity at natural pH (the pH of *in vitro* translation kit) and a high activity at alkaline pH range can prevent the degradation of other functional proteins during M57 protease expression in *in vitro* expression mix. After expression, M57 protease activity can be activated by altering pH of emulsion. Experimental investigation showed triethylaime could successfully alter pH of inner aqueous phase through primary w/o emulsion and significantly increase M57 protease activity. This increase of protease activity by adding triethylamine was presumed as a effect on catalytic reaction in acceleration of deacylation step (Yamamoto and Kise 1993). Another noticeable challenge is inhibition of proteases activities caused by inhibitors present in the *in vitro* expression kit. The results shown by increasing substrate concentration and using diluted *in vitro* translation solution (in which the

inhibitor concentration was diluted), an increased fluorescence signal could be observed. To further improve the fluorescence signal, preparation and optimization of an adjustable *in vitro* expression system with flexible reaction buffer conditions is still worth to try in future.

CyFlow® space flow cytometer has a piezoelectric activation sorter module. This sorter module has a continuous fluid in sorting channel resulting in a dilution of sorted samples in sheath buffer. To recover sorted DNA molecule from bulk of sheath buffer, different inert carriers which can increase the recovery of low concentration of nucleic acids by ethanol precipitation such as linear polyacrylamide, Pellet paint®, Pellet paint_NF and glycogen were explored (Gaillard and Strauss 1990; Sambrook and Russell 2001). Linear ploy acrylmide turned to be the best one which could coprecipitated an extremely diluted DNA sample (5 fg/µL) with 20 % recovery efficiency. However for the optimized sorting experiment, approximately 10^4-10^5 events (0.01-0.1 pg in the ideal condition as one DNA molecular per compartmentalization) would be sorted out in 50 mL sheath solution. The amount of sorted DNA was still a bit lower than the minimum required amount for recovery. Therefore the DNA recovery need to be further developed. Adding other non-target DNA can increase the total amount of DNA and this might help the precipitation of target DNA which has a low concentration.

4.3 Whole-cell based protease screening system

Compared to cell-free system, whole-cell based screening system has advantages like it is easy to recover sorted cells (for example by centrifuge or membrane filtration) and protein expression level is higher using cells. Whole-cell based protease screening system firstly was developed by *E. coli* cells based intercellular expression system concerning a high transformation efficiency of *E. coli* cells. However a high background activity caused by other proteases inside *E. coli* cells was observed. To eliminate intracellular protease activities from other non-target proteases, a secretion system should be used. *E. coli* has secretion systems and in a case subtilisin E was expressed in periplasm for evolving of its esterase activity (Sroga and Dordick 2001; Choi and Lee 2004;

Mergulhao, Summers et al. 2005). But in this study a secretion system employed extracellular proteases deficient *Bacillus cells* and secretion protease Subtilisin Carlsberg was used due to a ready to use condition. The whole-cell based screening system was named as Pro-FCIVC. As described in cell-free based screening system, the main development includes selection of substrate, optimization of fluorescence signal for sorting and finally verification of developed screening system. Figure 4.1 shows an overview of the developed Pro-FCIVC screening technology including the main screening procedure and solutions to occurred challenges.

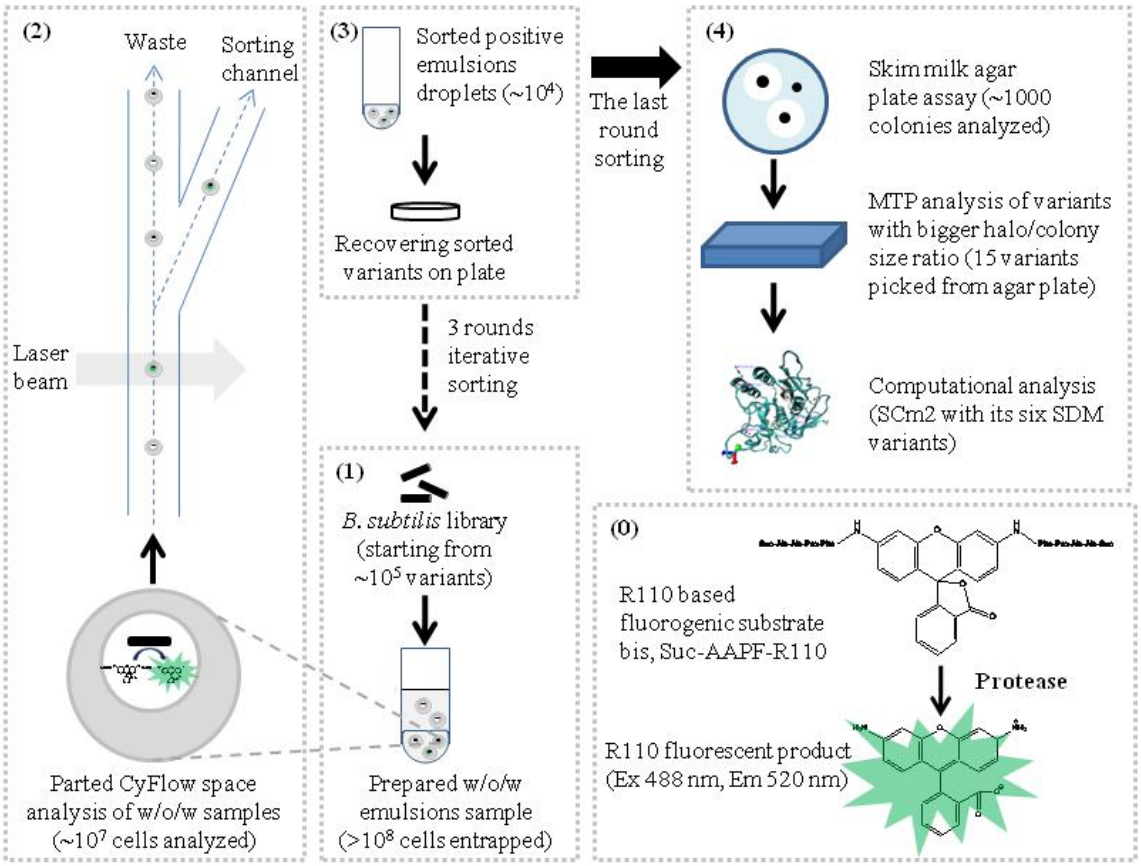


Figure 4.1 Scheme of the developed flow cytometry based whole cell screening system for proteases (Pro-FCIVC). (0) Generation of a fluorescence signal through protease catalyzed cleavage of the fluorogenic substrate bis, Suc-AAPF-R110. (1) Encapsulation of protease expressing *B. subtilis* cells together with the fluorogenic substrate into w/o/w double emulsions. (2) Analysis and sorting of w/o/w double emulsions by flow cytometry. (3) Recovery of sorted positive events by vacuum filtration and propagation on agar plates for the next round inoculation, emulsification and sorting. (4) Positive events from last round of sorting were plated on skim milk agar plates and active variants with the largest ratios of halo/colony sizes were picked for microtiter plate rescreening and further characterization.

As discussed in section 4.1, for Pro-FCIVC screening system, bis, Suc-AAPF-R110 (Figure 4.1 (0)) was selected as substrate for the model protease Subtilisin Carlsberg (SC) as the charges of its succinyl-goup and its fluorescent hydrolysis product rhodamine 110 (R110) allow the retention of substrate and product inside the compartments. Additionally the chosen substrate is not membrane permeable avoiding additional background activities from intracellular proteases.

In the work extracellular proteases deficient *Bacillus cells* and secreted Subtilisin Carlsberg (SC) protease were selected to decrease the background caused by other naturally presented proteases from host strain. *B. subtilis* strains also have the merit such as its naturally high secretion capacity as well as its safe and practical handling properties (Westers, Westers et al. 2004). Our investigation of two extracellular protease-deficient *B. subtilis* strains showed that a strain where eight genes coding for extracellular proteases had been knocked out (*B. subtilis* WB800N) (Nguyen 2006) was optimal for the intended application as it displayed the lowest background activities combined with high product fluorescence signals. In contrast, due to lower signal intensities and higher background *B. subtilis* WB600 deficient for only six extracellular proteases proved to be less suitable.

During optimization of sorting conditions, growth competition was observed when cells were directly subjected to liquid cultivation after the sorting process. The biased growth led to a loss of enrichments particularly after the first round of sorting when the number of active cells in the inoculum was still low. Investigation experiments showed that agar plate propagation prior to liquid cultivation could limit the effect on growth competition and preserve the accomplished enrichments. According to a previous review, a tight regulation of protein expression within the host cell is an essential feature to mask the growth competition (Daugherty, Iverson et al. 2000). Therefore two inducible expression systems (IPTG and xylose) in *B. subtilis* (Nguyen, Nguyen et al. 2005) were investigated in the work. However, due to low expression levels, the tested inducible expression

systems were not considered for Pro-FCIVC and the additional plate cultivation step proved to be optimal to preserve enrichments in sorted libraries.

Finally the developed Pro-FCIVC screening system was validated by screening an epPCR mutant library of SC with high mutational loads (diversity of ~10^5; ~5 mutations per 1 kb) for variants with improved resistance towards the protease inhibitor antipain-HCl. Two variants (SCm2 and SCm13) with 1.6-fold and 1.2-fold improved resistance were isolated from a small population of active variants (315) validating the developed protease flow cytometry screening technology for increased inhibitor resistance. In this work transformation efficiency of *B. subtilis* restricted the size of the screened epPCR mutant library. However, by employing advanced transformation technologies to overcome the transformation limitation, the developed ProFC-IVC screening system can be used to analyze large libraries in a high throughput.

Further kinetic characterizations of improved variants were performed using purified enzymes. Although Subtilisin Carlsberg has been studied several decades ago and it is commercially available, purification protocols are not well updated and published. Purification of other subtilisins mainly used affinity chromatography employing sepharose based absorbent (Fujiwara, Osue et al. 1975; Prescott, Peek et al. 1993; Ibrahim-Granet and Bertrand 1996) or gel filtration with a low loading capacity (Ibrahim-Granet and Bertrand 1996; Srimathi, Jayaraman et al. 2006). Herein a two-step ion exchange procedure (cation exchange followed by anion exchange) was carried out. Despite of an unusual flow-through phenomenon observed with cation and anion exchange columns, purified Subtilisin Carlsberg active proteases had a high purity (approximately 96 % purity in total proteins) and could directly apply for characterization. Running buffer of ion exchange purification investigations showed the unbinding phenomenon was not really affected by various buffer conditions at pH lower than 8.6. In previous research the isoelectric point (pI) value of Subtilisin Carlsberg was characterized and showed a value of 6.7 and 8.6 from two main fractions (Vitale and Gamulin 1975) with enzymatic activities which was disputed to the value of 9.4 from Ottesen and Spector (Ottesen and Spector 1960). Hence an isoelectric point analysis of purified SC variant should be examined to find out the unbinding reason in this study.

At last the variant SCm2 with 6 amino acid substitutions could be identified exhibiting a 1.6-fold increased resistance to the inhibitor compared to the wild-type at a wide range of antipain-HCl inhibitor concentration 0.5-5 µM. The mutated gene contained several silent mutations possibly contributing to increased expression levels of this variant. The amino acid substitution investigations showed no additive effect of the single mutations. Hence, double and triple mutations would have to be generated to investigate the likely cooperative effects of the identified amino acid changes to explain the improved inhibitor resistance.

4.4 Conclusion and outlook

In summary two first flow cytometry and IVC based screening systems (ProCF-FCIVC and Pro-FCIVC) with a high throughput (>7500 s^{-1}) have been developed for directed protease evolution in this work. The cell-free based ProCF-FCIVC system was demonstrated at a stage of proof of concept by screening and sorting a M57 reference library. The developed whole-cell Pro-FCIVC system was verified by screening an epPCR mutant library with a high mutational load (~5 mutations per 1 kb) for inhibition resistance of Subtilisin Carlsberg (SC) towards the antipain dihydrochloride inhibitor. Variant SCm2 with a 1.6-fold relative to wild-type SC improved resistance was isolated from a small population of active variants (315) validating the developed protease flow cytometry screening technology for increased inhibitor resistance.

The described whole-cell based Pro-FCIVC screening platform presents a powerful screening tool for engineering proteases by directed evolution. Nevertheless, its successful application demands the consideration of several limiting factors such as substrate selection, background activities of the employed host cells, transformation efficiency and cell recovery (Miller, Bernath et al. 2006). Most of the above mentioned limitations might be overcome by utilization of a well optimized efficient cell-free *in vitro* expression based system using an air-in-jet sorting module flow cytometer machine to avoid the dilution demerit. This development would allow harnessing fully the throughput capabilities of the flow cytometry screening technology, and further be

modified for other proteases and applied for additional properties such as high stability/activity at low or high temperature, altered pH profile or ionic strength resistance.

In general the high throughput capacities of flow cytometry based screening systems (>10^8 events per day) offer an opportunity for alternative directed evolution strategies. In contrast to conventional microtiter plate or agar plate assays where small libraries with low mutational loads can be screened for modest improvements and repeated rounds of mutation and screening, the high throughput of flow cytometry based methods allows screening of large libraries with high mutational loads for a faster improvement and the investigation of cooperative effects between multiple mutations especially in terms of structure function illumination. These research efforts will expand the range of proteases that can be selected from mutant libraries by high mutational loads; and lead to a deeper understanding of the factors influencing the structure-catalytic mechanism of proteases and an improved knowledge of functional genomics relevant to production.

5 References

Aharoni, A., G. Amitai, et al. (2005). "High-throughput screening of enzyme libraries: thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments." Chem Biol **12**(12): 1281-1289.

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

Aharoni, A., K. Thieme, et al. (2006). "High-throughput screening methodology for the directed evolution of glycosyltransferases." Nat Methods **3**(8): 609-614.

Arnold, F. H. and G. Georgiou (2003). Directed enzyme evolution. New Jersey, Humana Press.

Becker, S., H. U. Schmoldt, et al. (2004). "Ultra-high-throughput screening based on cell-surface display and fluorescence-activated cell sorting for the identification of novel biocatalysts." Curr Opin Biotechnol **15**(4): 323-329.

Bernath, K., M. Hai, et al. (2004). "*In vitro* compartmentalization by double emulsions: sorting and gene enrichment by fluorescence activated cell sorting." Anal Biochem **325**(1): 151-157.

Bernath, K., S. Magdassi, et al. (2004). "*In vitro* compartmentalization (IVC): A high-througput screening technology using emulsions and FACS." Discovery Medicine **4**(20): 49-53.

Bershtein, S. and D. S. Tawfik (2008). "Advances in laboratory evolution of enzymes." Curr Opin Chem Biol **12**(2): 151-158.

Branden, O. R. C. and J. Tooze (1998). Introduction to protein structure. New York, Garland Publishing.

Bryan, P. N. (2000). "Protein engineering of subtilisin." Biochim Biophys Acta **1543**(2): 203-222.

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

Cherry, J. R. and A. L. Fidantsef (2003). "Directed evolution of industrial enzymes: an update." Curr Opin Biotechnol **14**(4): 438-443.

Choi, J. H. and S. Y. Lee (2004). "Secretory and extracellular production of recombinant proteins using Escherichia coli." Appl Microbiol Biotechnol **64**(5): 625-635.

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

Dower, W. J., J. F. Miller, et al. (1988). "High efficiency transformation of *E. coli* by high voltage electroporation." Nucleic Acids Res **16**(13): 6127-6145.

Evnin, L. B., J. R. Vasquez, et al. (1990). "Substrate specificity of trypsin investigated by using a genetic selection." Proc Natl Acad Sci USA **87**(17): 6659-6663.

Fernandez-Gacio, A., M. Uguen, et al. (2003). "Phage display as a tool for the directed evolution of enzymes." Trends Biotechnol **21**(9): 408-414.

Fujiwara, K., K. Osue, et al. (1975). "Affinity chromatography of alpha-chymotrypsin, subtilism, and metalloendopeptidases on carbobenzoxy-L-phenylalanyl-triehtylenetetraminyl-sepharose." J Biochem **77**(4): 739-743.

Gaillard, C. and F. Strauss (1990). "Ethanol precipitation of DNA with linear polyacrylamide as carrier." Nucleic Acids Res **18**(2): 378.

Ghadessy, F. J. and P. Holliger (2004). "A novel emulsion mixture for in vitro compartmentalization of transcription and translation in the rabbit reticulocyte system." Protein Eng Des Sel **17**(3): 201-204.

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

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

Griswold, K. E., N. S. Aiyappan, et al. (2006). "The evolution of catalytic efficiency and substrate promiscuity in human theta class 1-1 glutathione transferase." J Mol Biol **364**(3): 400-410.

Griswold, K. E., Y. Kawarasaki, et al. (2005). "Evolution of highly active enzymes by homology-independent recombination." Proc Natl Acad Sci USA **102**(29): 10082-10087.

Gupta, R., Q. K. Beg, et al. (2002). "An overview on fermentation, downstream processing and properties of microbial alkaline proteases." Appl Microbiol Biotechnol **60**(4): 381-395.

Gupta, R., Q. K. Beg, et al. (2002). "Bacterial alkaline proteases: molecular approaches and industrial applications." Appl Microbiol Biotechnol **59**(1): 15-32.

Hai, M., K. Bernath, et al. (2004). "Flow cytometry: a new method to investigate the properties of water-in-oil-in-water emulsions." Langmuir **20**(6): 2081-2085.

Haki, G. D. and S. K. Rakshit (2003). "Developments in industrially important thermostable enzymes: a review." Bioresour Technol **89**(1): 17-34.

Hayashi, N., M. Welschof, et al. (1994). "Simultaneous mutagenesis of antibody CDR regions by overlap extension and PCR." Biotechniques **17**(2): 310, 312, 314-315.

Hedstrom, L. (2002). "Serine protease mechanism and specificity." Chem Rev **102**(12): 4501-4524.

Ibrahim-Granet, O. and O. Bertrand (1996). "Separation of proteases: old and new approaches." J Chromatogr B Biomed Appl **684**(1-2): 239-263.

Inoue, H., H. Nojima, et al. (1990). "High efficiency transformation of Escherichia coli with plasmids." Gene **96**(1): 23-28.

Jaeger, K. E. (2004). "Protein technologies and commercial enzymes: White is the hype–biocatalysts on the move." Curr Opin Biotechnol **15**: 269-271.

Link, A. J., K. J. Jeong, et al. (2007). "Beyond toothpicks: new methods for isolating mutant bacteria." Nat Rev Microbiol **5**(9): 680-688.

Lorenz, P. and J. Eck (2005). "Metagenomics and industrial applications." Nat Rev Microbiol **3**(6): 510-516.

Mastrobattista, E., V. Taly, et al. (2005). "High-throughput screening of enzyme libraries: in vitro evolution of a beta-galactosidase by fluorescence-activated sorting of double emulsions." Chem Biol **12**(12): 1291-1300.

Mattanovich, D. and N. Borth (2006). "Applications of cell sorting in biotechnology." Microb Cell Fact **5**: 12.

Maurer, K. H. (2004). "Detergent proteases." Curr Opin Biotechnol **15**(4): 330-334.

Mergulhao, F. J. M., D. K. Summers, et al. (2005). "Recombinant protein secretion in Escherichia coli." Biotech Adv **23**(3): 177-202.

Miller, O. J., K. Bernath, et al. (2006). "Directed evolution by in vitro compartmentalization." Nat Methods **3**(7): 561-570.

Miyazaki, K. and M. Takenouchi (2002). "Creating random mutagenesis libraries using megaprimer PCR of whole plasmid." Biotechniques **33**(5): 1033-1038.

Nguyen, H. D. (2006). Construction of plasmid-based expression and secretion vectors and study of the immobilization of proteins on the surface of *Bacillus subtilis* cells, University Bayreuth. **Dr. Rer. Nat.:** 114.

Nguyen, H. D., Q. A. Nguyen, et al. (2005). "Construction of plasmid-based expression vectors for *Bacillus subtilis* exhibiting full structural stability." Plasmid **54**(3): 241-248.

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

Ottesen, M. and A. Spector (1960). "A comparison of two proteinases from Bacillus subtilis." C R Trav Lab Carlsberg **32**(-): 63-74.

Oue, S., A. Okamoto, et al. (1999). "Redesigning the substrate specificity of an enzyme by cumulative effects of the mutations of non-active site residues." J Biol Chem **274**(4): 2344-2349.

Pogson, M., G. Georgiou, et al. (2009). "Engineering next generation proteases." Curr Opin Biotechnol **20**(4): 390-397.

Prescott, M., K. Peek, et al. (1993). "The use of phenyl-Sepharose for the affinity purification of proteinases." J Biochem Biophys Methods **26**(1): 51-60.

Pulido, M. A., Y. Koga, et al. (2007). "Directed evolution of Tk-subtilisin from a hyperthermophilic archaeon: identification of a single amino acid substitution responsible for low-temperature adaptation." Protein Eng Des Sel **20**(3): 143-153.

Rao, M. B., A. M. Tanksale, et al. (1998). "Molecular and biotechnological aspects of microbial proteases." Microbiol Mol Biol Rev **62**(3): 597-635.

Romero-Garcia, E. R., A. Tellez-Valencia, et al. (2009). "Engineering and directed evolution of a Ca2+ binding site A-deficient AprE mutant reveal an essential contribution of the loop Leu75-Leu82 to enzyme activity." J Biomed Biotechnol **2009**: 201075.

Saeki, K., K. Ozaki, et al. (2007). "Detergent alkaline proteases: enzymatic properties, genes, and crystal structures." J Biosci Bioeng **103**(6): 501-508.

Sambrook, J. and D. W. Russell (2001). Molecular Cloning: A laboratory manual, Cold Springer Harbor Laboratory Press.

Siddiqui, K. S. and R. Cavicchioli (2006). "Cold-adapted enzymes." Annu Rev Biochem **75**(-): 403-433.

Smith, G. P. (1985). "Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface." Science **228**(4705): 1315-1317.

Srimathi, S., G. Jayaraman, et al. (2006). "Improved thermodynamic stability of subtilisin Carlsberg by covalent modification." Enzyme and Microbial Technology **39**(2): 301-307.

Sroga, G. E. and J. S. Dordick (2001). "Generation of a broad esterolytic subtilisin using combined molecular evolution and periplasmic expression." Protein Eng **14**(11): 929-937.

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

Strausberg, S. L., B. Ruan, et al. (2005). "Directed coevolution of stability and catalytic activity in calcium-free subtilisin." Biochemistry **44**(9): 3272-3279.

Streit, W. R., R. Daniel, et al. (2004). "Prospecting for biocatalysts and drugs in the genomes of non-cultured microorganisms." Curr Opin Biotechnol **15**(4): 285-290.

Taly, V., B. T. Kelly, et al. (2007). "Droplets as microreactors for high-throughput biology." Chembiochem **8**(3): 263-272.

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

Tee, K. L. and U. Schwaneberg (2007). "Directed evolution of oxygenases: screening systems, success stories and challenges." Comb Chem High Throughput Screen **10**(3): 197-217.

Tindbaek, N., A. Svendsen, et al. (2004). "Engineering a substrate-specific cold-adapted subtilisin." Protein Eng Des Sel **17**(2): 149-156.

van der Graaf, S., C. G. P. H. Schroen, et al. (2005). "Preparation of double emulsions by membrane emulsification--a review." J Memb Sci **251**(1-2): 7-15.

Varadarajan, N., J. Gam, et al. (2005). "Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity." Proc Natl Acad Sci USA **102**(19): 6855-6860.

Varadarajan, N., S. Rodriguez, et al. (2008). "Highly active and selective endopeptidases with programmed substrate specificities." Nat Chem Biol **4**(5): 290-294.

Vitale, L. F. and S. Gamulin (1975). "Characterization of subtilisins by isoelectric focusing and polyacrylamide gel electrophoresis." Int J Biochem **6**(3): 165-171.

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

Weng, M., Z. Zheng, et al. (2009). "Enhancement of oxidative stability of the subtilisin nattokinase by site-directed mutagenesis expressed in Escherichia coli." Biochim Biophys Acta **1794**(11): 1566-1572.

Westers, L., H. Westers, et al. (2004). "*Bacillus subtilis* as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism." Biochim Biophys Acta **1694**(1-3): 299-310.

Wintrode, P. L., K. Miyazaki, et al. (2001). "Patterns of adaptation in a laboratory evolved thermophilic enzyme." Biochim Biophys Acta **1549**(1): 1-8.

Wong, T. S., K. L. Tee, et al. (2004). "Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution." Nucleic Acids Res **32**(3): e26.

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

Wu, X. C., W. Lee, et al. (1991). "Engineering a Bacillus subtilis expression-secretion system with a strain deficient in six extracellular proteases." J Bacteriol **173**(16): 4952-4958.

Yamamoto, Y. and H. Kise (1993). "Unusual enhancement of protease activity in organic solvents by amines." Chemistry Letters **22**(11): 1821-1824.

Yang, G. and S. G. Withers (2009). "Ultrahigh-throughput FACS-based screening for directed enzyme evolution." Chembiochem **10**(17): 2704-2715.

Yang, Y. R., H. Zhu, et al. (2008). "Cold-adapted maturation of thermophilic WF146 protease by mimicking the propeptide binding interactions of psychrophilic subtilisin S41." FEBS Lett **582**(17): 2620-2626.

You, L. and F. H. Arnold (1996). "Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide." Protein Eng **9**(1): 77-83.

Zhao, H. and F. H. Arnold (1999). "Directed evolution converts subtilisin E into a functional equivalent of thermitase." Protein Eng **12**(1): 47-53.

Zhong, C. Q., S. Song, et al. (2009). "Improvement of low-temperature caseinolytic activity of a thermophilic subtilase by directed evolution and site-directed mutagenesis." Biotechnol Bioeng **104**(5): 862-870.

Appendix

Table 0.1 Procedure of ion exchange column purification of Subtilisin Carlsberg. Activity assay was carried out using 1 µL of supernatant/purified solution and bis, Suc-AAPF-R110 (5 µM) in Tris-HCl buffer (0.1 M, pH 8.6).

Purification step	**sample**	**Volume [mL]**	**Activity [AU/min]**	**Total activity [AU/min]**	**Purification efficiency**
Filtration	Glass fiber	250	274.0	68500	100 %
Amicon filtration	Concentrated supernatant	14	754.8	10567.2	15.43 %
Cation exchange column	c7-c13 mix	14	661.0	9254.0	13.51 %
	c7	2	667.4		
	c12	2	584.6		
	c28	2	2.6		
Anion exchange column	a6-a12 mix	/	/	7671.2	11.20 %
	a6	2	336.0		
	a7	2	604.0		
	a8	2	671.4		
	a9	2	626.6		
	a10	2	582.8		
	a11	2	565.8		
	a12	2	449.0		

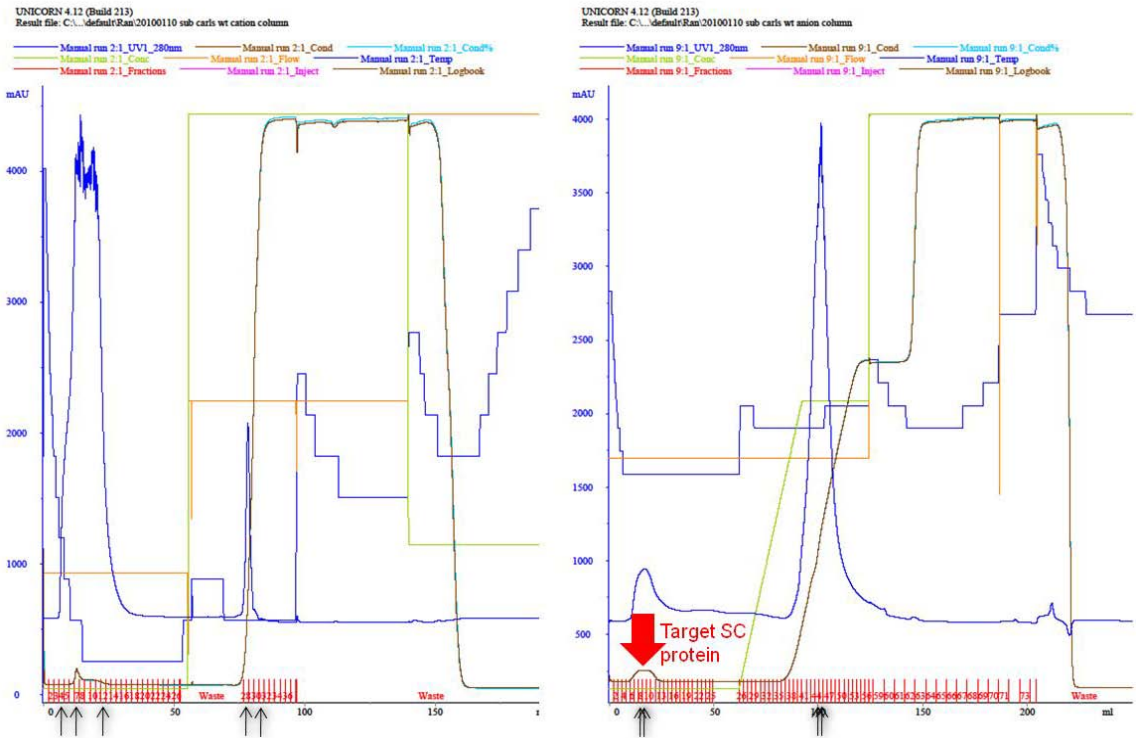
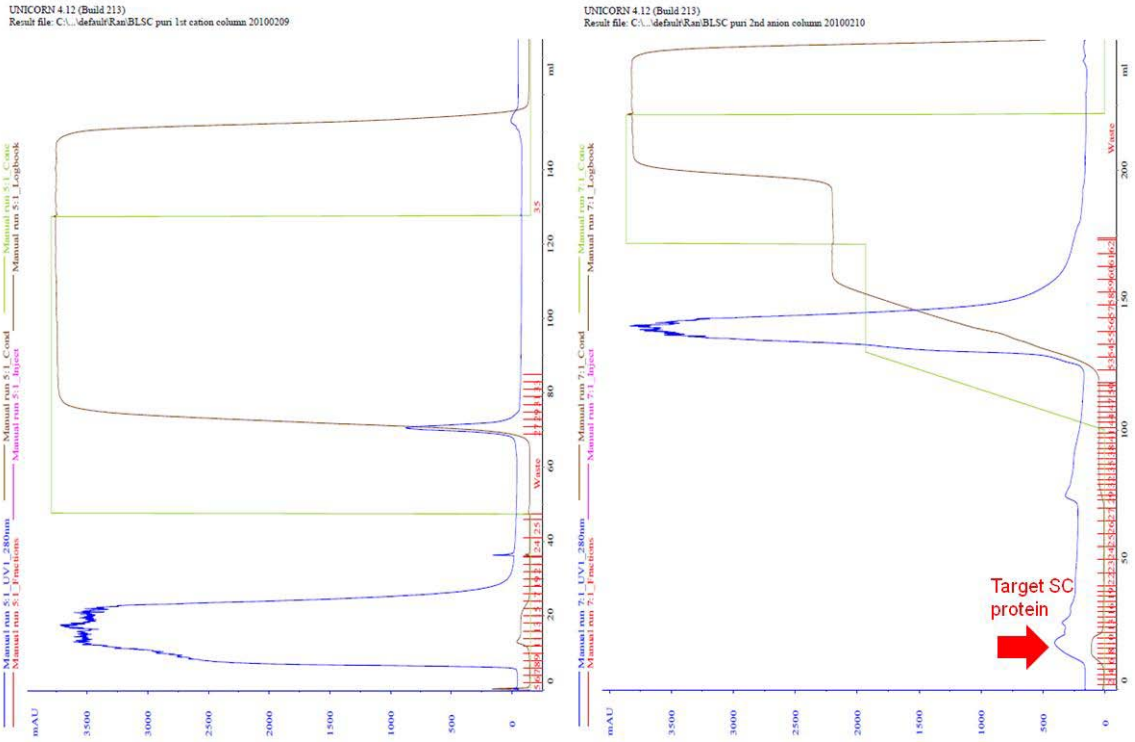
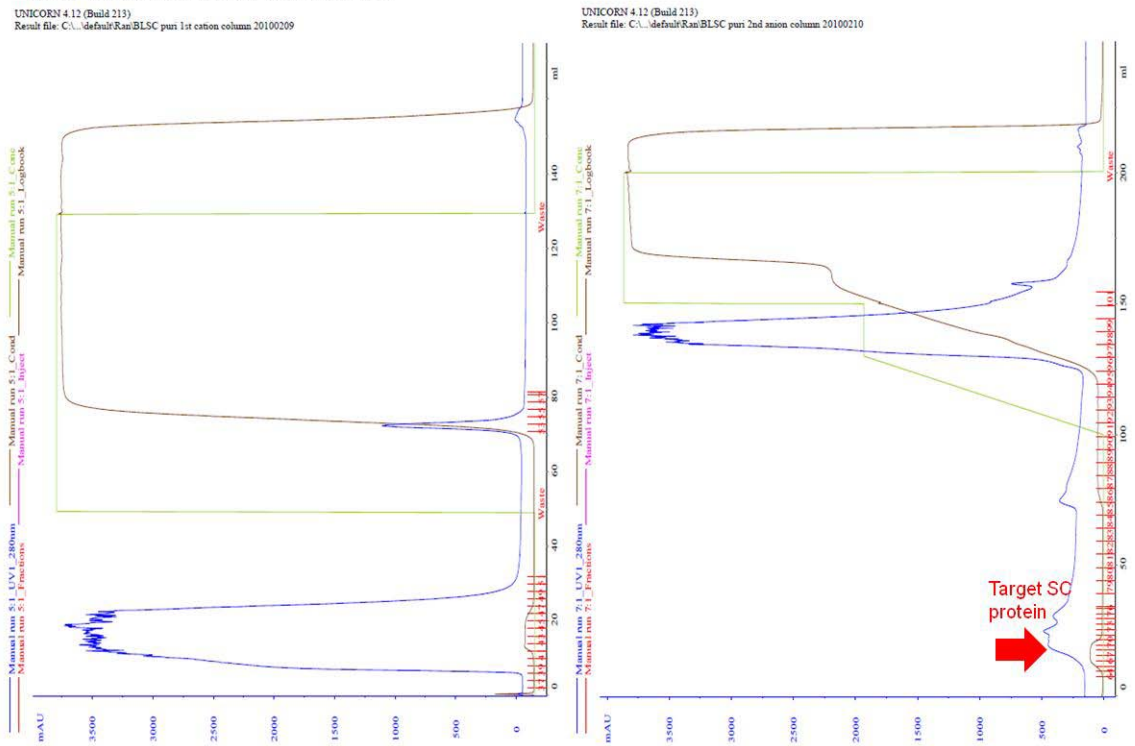


Figure 0.1 Two-step ion exchange purification of Subtilisin Carlsberg (SC) protease. 250 mL of supernatant was concentrated into 14 mL Tris-HCl buffer (50 mM, pH 8.6) and loaded to cation exchange (Toyopearl SP-650C; running buffer: 50 mM Tris-HCl pH 8.6, elution buffer: 50 mM Tris-HCl pH 8.6 with 2 M NaCl) followed by anion exchange (Toyopearl SuperQ-650C; running buffer: 50 mM Tris-HCl pH 8.6, elution buffer: 50 mM Tris-HCl pH 8.6 with 2 M NaCl). Blank arrows indicate the samples collected for activity analysis and SDS-PAGE analysis.

SCm13 1st cation column, 2nd anion column



SCm2 1st cation column, 2nd anion column



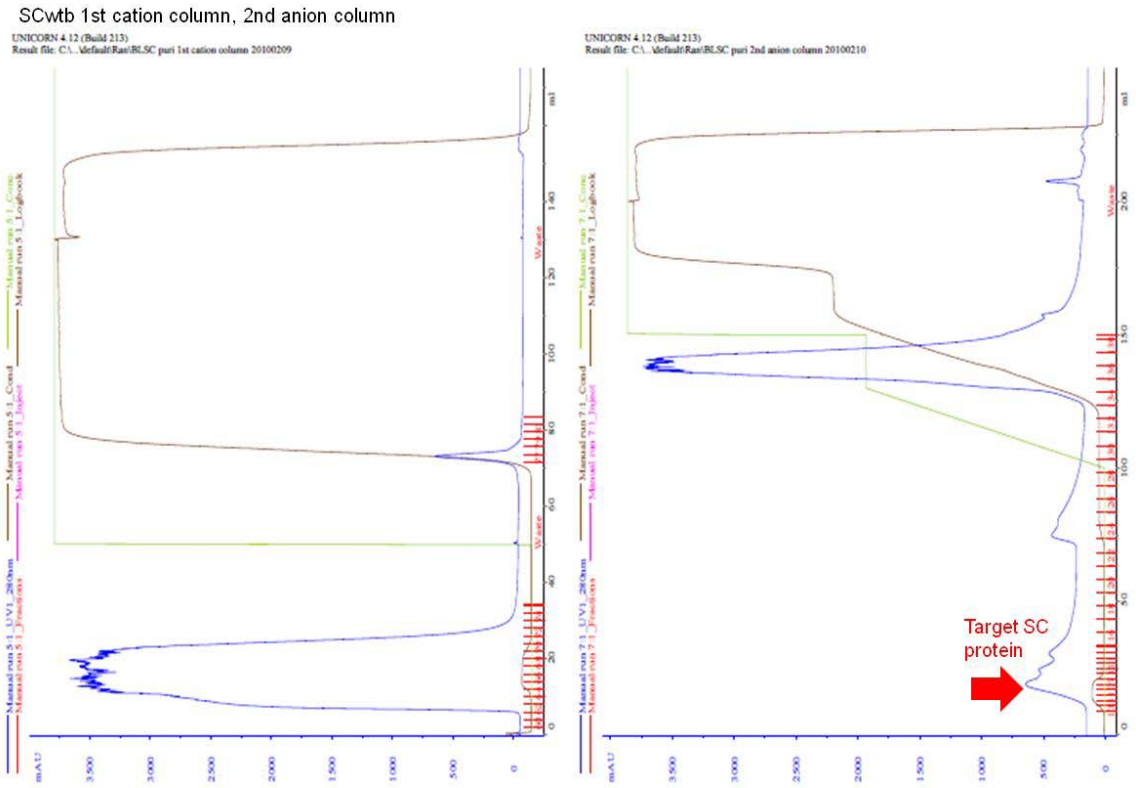


Figure 0.2 Two-step ion exchange purification of Subtilisin Carlsberg variants. Purification was carried out as above with one modification that 5 mM Tris-HCl (pH 8.6) was used as running buffer for 2nd column (anion column). Red arrow represents the target SC proteases.

Declaration

I hereby declare that this thesis was written by myself. The scientific data presented in this thesis were my own research findings.

Ran Tu

June 25, 2010

Bremen, Germany