“Engineering of subtilisin proteases for detergent applications”

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

vorgelegt von

Diplom-Ingenieur (FH)
Bioingenieurwesen

Felix Jakob

aus Freiburg im Breisgau

Berichter: Universitätsprofessor Dr. Ulrich Schwaneberg
Universitätsprofessor Dr. Lars M. Blank

Tag der mündlichen Prüfung: 08.05.2013

TABLE OF CONTENTS

I. ACKNOWLEDGMENTS

I would like to thank Prof. Schwaneberg for giving me the opportunity to be part of his research group and for my scientific and personal development during the last three years. I thank Prof. Blank and Prof. Elling for being part of my examination committee.

I thank our collaboration partners at Henkel for the interesting challenges they gave us. Special thanks go to Dr. Hendrik Hellmuth, who was my contact person from day one onwards and who supported me in a perfect way.

A big thank you is for Ronny, who supervised and inspired me during my PhD. I felt lots of support and motivation. It was great to work with you!

I also would like to thank the whole Schwaneberg group for the nice and professional working atmosphere. Special thanks go to Marco for the help in computational analysis and to Andreas and his team for the good lab organization. My closest coworkers Ljubica, Dragana and Christian I would like to thank for sharing many hours with me and for their support. I thank my students Tim, Sugi and Sigi for their great job and hope they learned as much from me as I learned from them.

I cannot thank my parents and my brother enough for their endless support and trust during the last 3 years and the 26 years before.

Last but not least I thank Kristina for her endless support and for enriching my life.

II. ABBREVIATIONS

Ap	Ampicillin
AU	Absorbance unit
BgAP	*Bacillus gibsonii* alkaline protease
BLAP	*Bacillus lentus* alkaline protease
BMA	Benzylmalonic acid
BSA	Bovine serum albumin
B. subtilis	*Bacillus subtilis*
Da	Dalton
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
epPCR	Error-prone PCR
HTS	High Throughput Screening
MEGAWHOP	Megaprimer PCR of Whole Plasmid
MTP	Microtiter plate
OD	Optical Density
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PLICing	Phosphorothioate-based ligase-independent gene cloning
SDM	Site-Directed Mutagenesis
SeSaM	Sequence Saturarion Mutagenesis
SSM	Site-Saturation Mutagenesis
Suc-AAPF-*p*NA	N-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide
Tc	Tetracycline
TCA	Trichloroacetic acid
YASARA	Yet Another Scientific Artificial Realitiy Application

1 INTRODUCTION

1.1 Proteases

Enzymes are present in every living organism and catalyze countless of different chemical reactions. In biotechnology and biochemistry, the impressive repertoire of biocatalysts is utilized for chemical synthesis. Enzymes are often energy-efficient and environmentally friendly catalysts, as they are highly active under mild conditions and generate only few waste products (Arnold 2001).

Proteases catalyze the hydrolysis of peptide and ester bonds. Based upon the nature of the nucleophile in the hydrolysis reaction, they are broadly grouped into six catalytic types: serine, threonine, cystein, aspartic, glutamic and metallo-proteases (Rawlings et al. 2011). Approximately 40 % of the total enzyme sales account for proteases. Serine proteases are the most exploited industrially group within the protease family. Besides their important role in cellular metabolic processes these enzymes have become widely used in detergent industry since the introduction of pancreatic extracts in 1913 as detergent additive. Additionally, serine proteases are used in various industrial sectors such as food, pharmaceutical, leather, diagnostics and waste management (Gupta et al. 2002; Maurer 2004).

1.2 Proteases in detergent applications

Detergent formulations commonly contain enzymes like proteases to remove proteinaceous stains from textiles during the washing process. Additionally lipases, amylases and cellulases are supplemented to boost the washing performance. Nowadays detergent proteases from microbial sources are screened for hydrolytic activity on protein substrates and the potential candidates are isolated and tested under application conditions (Von der Osten et al. 1993). However, nature has evolved enzymes for millions of years towards the catalysis of specific reactions in distinct natural reaction environments. The identification and utilization of enzymes fulfilling all industrial requirements is therefore like looking for a needle in a haystack.

In detergent applications, the major challenges for an efficient implementation of proteases are producibility, stability, low substrate specificity, activity at low or high

pH, ionic strength behavior and resistance towards thermal treatment, surfactants and detergents (Farinas et al. 2001; Nascimento and Martins 2006). Subtilisins, an extracellular produced group of serine proteases, fulfill to a certain extent several requirements like resistance towards high temperature and surfactants as well as low substrate specificity and activity at alkaline pH.

With the change in consumer's preference towards liquid detergent formulations the working environment of detergent proteases has changed dramatically, therefore new proteases or the adaptation of detergent proteases to the new environment is a prerequisite for the effective removal of proteinaceous stains during the washing process. Among other characteristics the pH value of the distinct formulation plays an important role for the washing performance of the embedded proteases. The pH-value of powder formulations (pH ~10) is two units higher compared to liquid detergent formulations (pH ~8). The pH dependent activity of subtilisins or enzymes in general is not completely understood; therefore a rational strategy to shift the pH optimum of detergent proteases towards increased activity at pH values below pH 10, would enable the fast and effective adaptation of proteases used in powder formulations for their application in liquid detergents.

In powder and liquid detergent formulations the primary cause of inactivation differs. Proteases in powder formulation are granulated and therefore show low water activity or autoproteolytic degradation. The main cause for inactivation is oxidative inactivation during storage by bleaching agents like perborates and percarbonates. On the other hand, as liquid formulations don't include bleaching agents, the main cause of inactivation for proteases is autoproteolytic degradation. Additionally, proteolytic active proteases in liquid detergents lead to the degradation of other embedded enzymatic components like cellulases, amylases or lipases (Von der Osten et al. 1993).

Nowadays the inhibition of proteases in liquid detergent formulation is commonly accomplished by the supplementation of boric acid (Severson 1985). The stabilization and specifically the inhibition of proteases during storage became an intense challenge for the detergent industry, because the application of boric acid is due to its toxic potential controversially discussed and therefore the identification and application of alternative boron free inhibitors is of high interest.

Besides the new requirements for detergent proteases, which were caused by the introduction of liquid detergents to the market, the cost effective production of proteases for laundry detergents is an ongoing challenge and alternative straightforward and universal strategies to increase the protease production are of high importance for detergent industry.

1.3 Serine Proteases

The common reaction mechanism of serine proteases is shown in Figure 1. The catalytic triad consists of serine, histidine and aspartic acid. Once the substrate is bound, a nucleophilic attack of the serine oxygen to the peptide carbonyl group initiates the acylation reaction. The formed instable tetrahedral intermediate is stabilized by amide hydrogens coordinating the anionic oxygen. This protein area is known as the oxyanion hole. The tetrahedral intermediate decomposes to an acyl-enzyme by breaking the peptide bond and the amino group leaves the substrate-enzyme complex. The deacylation of the enzyme is initiated by a nucleophilic attack of a water molecule present at the catalytically active histidine. The histidine protonates while the resulting hydroxyl group forms a covalent bond to the carbonyl carbon. The resulting tetrahedral intermediate is again stabilized by the oxyanion hole. In the last step the tetrahedral intermediate decomposes by breaking the bond to serine hydroxyl group and the carboxylic acid product leaves the active site. The enzyme is again ready for the next reaction (Polgar 2005; Walker and Lynas 2001).

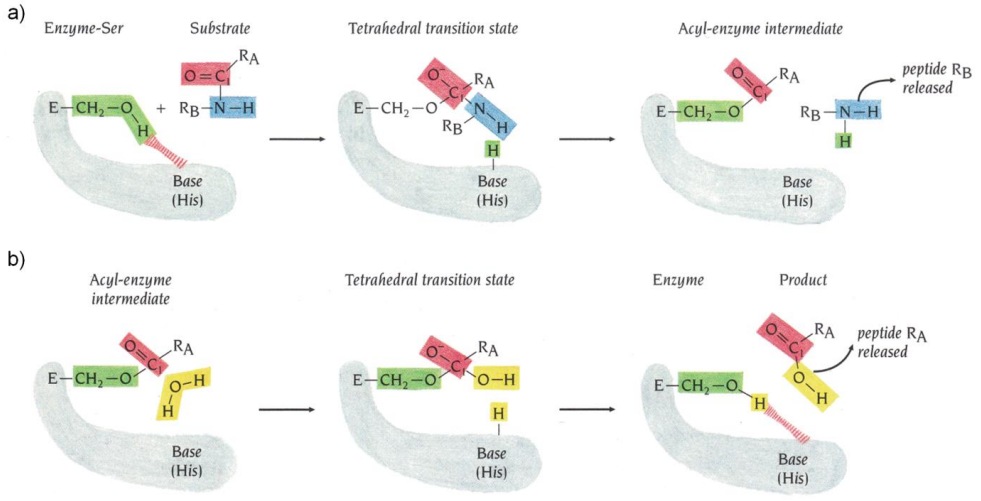


Figure 1: The hydrolysis reaction of serine proteases. **a)** The acylation reaction is the rate limiting step, defining the k_{cat}. **b)** The deacylation step is reported to be >33 faster than the acylation (Branden and Tooze 1999; Wells and Estell 1988).

1.4 Protein engineering of subtilisins

Enzymes are sophisticated catalysts which might consist of multiple subunit complexes and depend on cofactors or metal ions to carry out reactions. The single domain α/β-fold serine proteases, mainly secreted by *Bacillus* sp., are easily assayed and can be efficiently produced in gram-per-liter amounts by protease deficient strains such as *B. amyloliquefaciens or B. licheniformis*, achieving up to ~30 % of the total secreted protein (Wells and Estell 1988). By their catalytic mechanism subtilisins are defined as serine proteases (Figure 1). Based on the amino acid sequence and three-dimensional structure subtilisins can be clearly differentiated from other serine proteases, such as chymotrypsin, carboxypeptidase and peptidase A from *E. coli* (Branden and Tooze 1999; Maurer 2004). Enormous structural evidence of subtilisins is provided; in 2012 more than 250 entries for subtilisin structures are available in the protein data bank (http://www.rcsb.org; Figure 2). Success factors for subtilisins for the application in detergent industry are high stability and relatively low substrate specificity. Both features are common for extracellular produced proteases.

In the last decade, extensive protein engineering approaches -mainly accomplished by directed evolution- focused on altering the substrate specificity (Rheinnecker et al. 1993), general stability (Takagi et al. 2000) and proteolytic activity (Sternberg et al. 1987) of subtilisins. Amino acid substitutions in over 50 % of the 275 amino acids were performed and reported (Bryan 2000; Leisola and Turunen 2007). In the last years, the focus of protease engineering shifted towards engineering subtilisins and other proteases towards novel and precise amino acid sequence-specific cleavage for the use in analytical, biotechnological and therapeutic applications, since a single protease molecule can inactivate, due to its catalytic turnover, numerous target proteins (Craik et al. 2011; Pogson et al. 2009). For example, subtilisin BPN' was through iterative modeling, mutagenesis and kinetic analysis cycles, tailored to preferentially cleave phosphotyrosine peptides (2500-fold enhanced relative to the wild type; Knight et al. 2007). Amino acid substitutions responsible for the modified substrate specificity are mainly located in or around the S1 specificity pockets and the active site (Di Cera 2008; Figure 2). Recently, the protease subtilisin E was engineered for the application in DNA-sequence-based diagnostic kits requiring high activity and resistance towards chaotrophic salts and detergents (Li et al. 2012) and another subtilisin protease from *Bacillus gibsonii* was adapted for high thermo

stability and increased activity at low temperatures (15°C) for the use in liquid detergent formulations (Martinez et al. 2012).

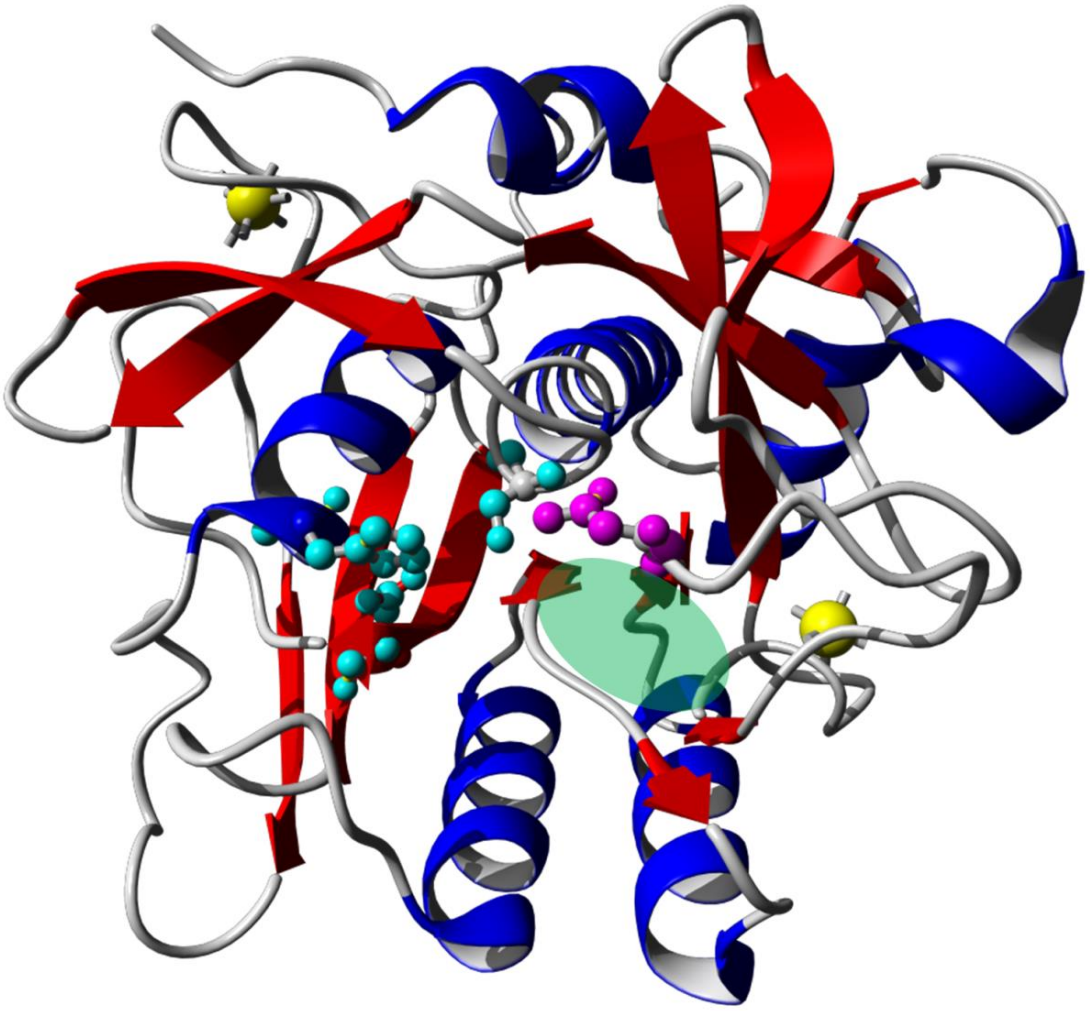


Figure 2: Crystal structure of *Bacillus lentus* alkaline protease (Goddette et al. 1992). Red: β-sheets. Blue: α-helices. Cyan: catalytic triad consisting of aspartate 32, histidine 62 and serine 215. Magenta: asparagine 153 (oxyanion hole). Yellow: calcium ions. Green: S1 pocket.

1.5 Protein engineering: directed evolution and rational design

Directed protein evolution represents the most versatile method for tailoring enzyme properties for industrial application requirements and for generating valuable information to understand structure-function relationships (Wong et al. 2006). Starting a directed evolution experiment does not require an understanding of structure-function relationships of the target protein and can be also applied to newly discovered enzymes for which no crystal structure has been solved and/or key amino

acid positions are unknown. The three major steps of directed evolution experiments are shown in Figure 3. The generation of a diverse mutant library includes the introduction of random mutations in the gene of interest and transformation of the mutated gene in an appropriate expression host (Step I). The generated mutant library is screened for improved variants using a suitable screening system (Step II). Genes encoding for improved variants are isolated and are usually used for iterative rounds of diversity generation and screening until a target improvement of the desired enzyme property is achieved (Wong et al. 2006).

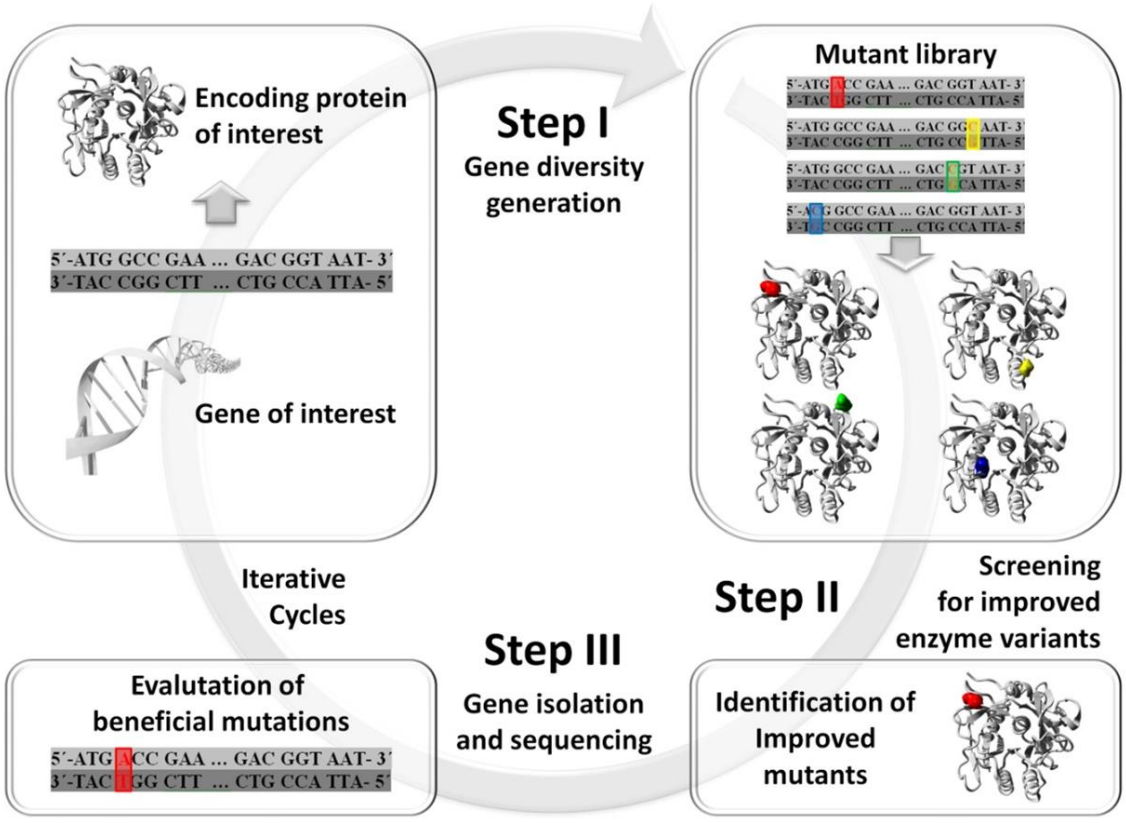


Figure 3: General overview of a directed evolution experiment. Scheme according to Güven et al. 2010.

The two main challenges in directed evolution campaigns are an unbiased diversity generation and development of specialized screening systems which enable the sampling of a meaningful mutant fraction in a reliable and timely way (Wong et al. 2006).

The most commonly used random mutagenesis techniques are enzyme based methods, introducing random mutations in the gene of interest under low fidelity conditions. The workhorse in directed protein evolution is *Taq* DNA polymerase from

Thermus aquaticus lacking 3'-5' proofreading exonuclease activity; however, its error-rate (1 mutation per 125.000 bp per PCR cycle) is not sufficient to generate mutant libraries. The most prominent methods to increase the error rates in a PCR reaction are the addition of $MnCl_2$, supplementing unbalanced nucleotide concentrations and increasing the number of amplification cycles (Cirino et al. 2003). Though versatile and simple, error-prone PCR (epPCR) methods are still significantly limited in their ability to create diversity on the gene level. An epPCR carried out under standard conditions can in average generate less than 40 % of all possible amino acid substitutions due to the biased mutational spectra of DNA polymerases, the lack of subsequent mutations due to low mutation frequencies and the redundant genetic code (Shivange et al. 2009). One of the latest developed random mutagenesis methods is the Sequence Saturation Mutagenesis (SeSaM), which overcomes limitations caused by biased DNA polymerases (Wong et al. 2005). SeSaM consists of four steps and is capable of saturating every single nucleotide position in the target sequence with all four standard nucleotides. A pool of DNA fragments with random length distribution is generated in a first PCR using α-phosphothioated nucleotides (Step 1). The generated DNA fragments are tailed at the 3'-termini with universal bases using terminal transferase (Step 2). The elongated DNA fragments are extended to full-length genes (Step 3). A concluding PCR is performed to replace the universal bases with standard nucleotides (Step 4; Figure 4).

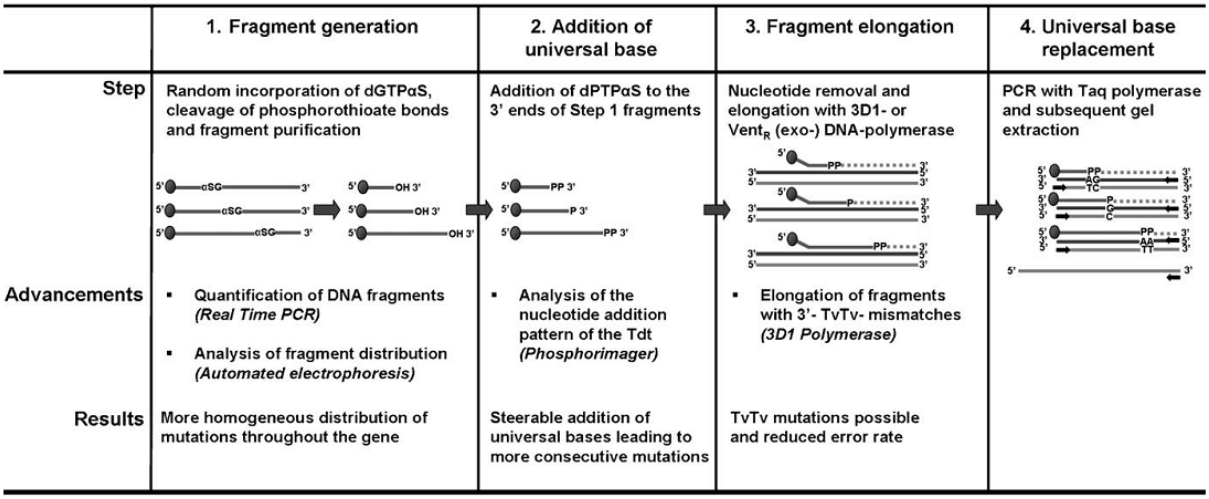


Figure 4: General scheme of SeSaM random mutagenesis: 1) Fragment generation, 2) Addition of universal base(s), 3) Fragment library elongation, and 4) Universal base replacements. The main improvements in steps 1, 2 and 3 are summarized in the corresponding columns (Mundhada et al. 2011).

The method of choice for mutant library generation depends highly on the enzyme property which is aimed to be engineered. Engineering a property like thermal stability usually requires high mutational load with amino acid substitutions towards chemically different amino acids, leading to SeSaM as a method of choice. The error-prone based methods lead to similar amino acid substitutions, which could be beneficial, for example, for modifying the enantioselectivity of an enzyme (Arnold 2001; Reetz and Jaeger 1999).

Another challenge in a directed evolution campaign is the availability of an appropriate screening system for the desired property. The screening procedure is a time consuming experimental work, and is the critical point that often determines if a directed evolution campaign succeeds or fails. The screening conditions have to reflect as close as possible the final application conditions, including substrate, temperature and environmental stress. The screening system has to be specially developed for each enzyme, should be compatible for medium and/or high throughput screening and sensitive over the final application range of the substrate concentration, enabling the identification of desired hits in the generated library (Aharoni et al. 2005). The most commonly used formats are microtiter plate screening systems (96-, 384- or 1536-wells) based on colorimetric or fluorimetric reactions. The microtiter plate systems are nowadays considered as medium throughput (10^3 to 10^5; Cohen et al. 2001). Solid phase screening systems, commonly based on agar plates, filter papers or membranes, are considered as high throughput (10^4 to 10^6; Shivange et al. 2012). Solid phase screening systems commonly lack in resolution, but allow qualitative differentiation between active and inactive variants. They are widely used as pre-screening systems, due to their high throughput, followed by microtiter plate based screening procedures (Li et al. 2012).

The introduction of flow cytometry-based screening systems to the field of directed evolution enabled ultra-high throughput screening of mutant libraries with more than 10^7 clones per hour (Farinas 2006). This system is based on the detection of fluorescent signals generated by enzymatic reactions. Flow cytometry screening systems further enable the analysis of *in vitro* generated enzyme libraries (Mastrobattista et al. 2005) and greatly accelerate the screening process in directed evolution campaigns, but are only available for a few enzyme classes, yet. Often only a qualitative differentiation between active and inactive enzyme variants from

high mutational load libraries is possible. The resulting enriched library is subsequently analyzed in depth.

A commonly used approach to tailor enzyme properties is rational design. In contrast to a directed evolution campaign, a successful rational design strategy requires detailed knowledge about the targeted protein, including structural information and the molecular basis of the targeted property in order to propose amino acid substitutions affecting the desired protein property. The method of choice for the rational protein design is site-directed mutagenesis (Toscano et al. 2007).

Semi-rational design bypasses limitations of directed protein evolution and rational design. Semi-rational design focuses on prior gained structural and functional knowledge targeting multiple relevant amino acid residues for site-saturation mutagenesis. The resulting 'smart' libraries are more likely to yield positive results. A comparison of different enzyme engineering approaches is given in Table 1 (Chica et al. 2005).

Table 1: Comparison of different enzyme engineering approaches (Chica et al. 2005).

	Rational design	Directed evolution (Random mutagenesis)	Semi-rational design
High-throughput screening or selection method	Not essential	Essential	Advantegeous but not essential
Structural and/or functional information	Both essential	Neither essential	Either sufficient
Sequence space exploration	Low	Moderate, random	Experimental: moderate, targeted Computational: vast, targeted
Probability of obtaining synergistic mutations	Moderate	Low	High

2 SURFACE CHARGE ENGINEERING OF *BACILLUS GIBSONII* ALKALINE PROTEASE

2.1 Declaration

Parts of this chapter were published by the author in the journal of Applied Microbiology and Biotechnology.

"Jakob F, Martinez R, Mandawe J, Hellmuth H, Siegert P, Maurer K-H, Schwaneberg U (2013) Surface charge engineering of a *Bacillus gibsonii* subtilisin protease. Appl Microbiol Biotechnol 97(15):6793-6802"

2.2 Objective

The objective was the development of a novel rational protein design strategy in order to alter the pH dependent activity and thermal resistance of the *Bacillus gibsonii* alkaline protease (BgAP) by surface charge engineering through rational single amino acid substitutions. The developed method was inspired by the natural post-translational deamidation reaction, wherein Asn and Gln residues are converted to Asp and Glu, respectively.

2.3 State of the art

2.3.1 Influence of protein surfaces charges on enzyme properties

The hydrophobic/hydrophilic balance and the charge distribution on the surface of proteins have significant effects on enzymatic performances like activity, thermal stability and pH optimum (Russell and Fersht 1987; Sakoda and Imanaka 1992; Sanchez-Ruiz and Makhatadze 2001; Grimsley et al. 1999). It was demonstrated that either the substitution of exposed hydrophobic amino acids against hydrophilic arginine or the addition of hydrophobic amino acids to the surface of a protein can increase protein stability (Strub et al. 2004; Van den Burg et al. 1998).

Amino acid composition studies in proteins from thermophilic and mesophilic bacteria showed that charged amino acids like glutamic acid (Glu), aspartic acid (Asp), lysine (Lys) and arginine (Arg) are comparatively overrepresented at the surface of proteins

from thermophilic organisms, whereas buried residues show no significant difference in their composition (Fukuchi and Nishikawa 2001).

The charge distribution and charge-charge interactions on protein surfaces have been widely studied, including ionic strength dependent activity, thermostability and pH-activity profile (De Kreij et al. 2002; Feller et al. 2010; Loladze et al. 1999). The pH-activity dependence on the protein surface charges has been reported for serine proteases from *B. amyloliquefaciens, B. lentus, B. pumilus* and *B. stearothermophilus* (De Kreij et al. 2002; DeSantis and Jones 1998; Jaouadi et al. 2010; Russell and Fersht 1987). Improvements in thermostability (up to 3-3.7-fold improved half-life times; Jaouadi et al. 2010) were achieved by single amino acid substitutions and by optimization of charge-charge interactions on the protein surface (Imanaka et al. 1986; Jaouadi et al. 2010; Yutani et al. 1977). Recently, the relationship between specific enzyme activity and protein charge of chymotrypsin-like serine protease from *Cellulomonas bogoriensis* in solution with different ionic strengths was studied. Variants with increased positive charge showed a higher reaction rate at increased ionic strength, evidencing a clear influence of the protein charge on specific activity (Feller et al. 2010).

2.3.2 The deamidation process and its influence on protein properties

The amino acid residues asparagine (Asn) and glutamine (Gln) can be unstable under physiological conditions, undergoing a spontaneous non-enzymatic deamidation, resulting in aspartic acid (Asp) and glutamic acid (Glu). As a result, the number of amino acids with a negative charge in a protein increases. Beside aspartic and glutamic acid, a third product, iso-asparagine, can be generated with a ratio of 3:1 (L-Asp to L-iso-Asp). The latter deamidation process has been described extensively and probable mechanisms were proposed (Capasso et al. 1989; Catak et al. 2009). Based on the deamidation process, a molecular clock hypothesis was proposed, in which the non-enzymatic deamidation (half-life time of asparagine and glutamine residues under physiological conditions) serves as a regulator or “timer” for biological processes (Robinson and Robinson 2004).

As alternative to the non-enzymatic deamidation, enzymatic deamidation of proteins has been studied in the case of proteases with amidolytic activity, like papain, pronase and chymotrypsin at high alkaline pH (Kato et al. 1987b). Enzymatic deamidation proved to be a useful method for altering functional properties such as

emulsifying and foaming properties for food processing (Kato et al. 1987a; Kato et al. 1987b).

First protein engineering studies by exchanging Asn to Asp at two positions in hen egg-white lysozyme were performed in the 1990's (Kato et al. 1992). The exchange of Asn to Asp shifted the pH optimum from a neutral to a more acidic pH-value (0.5 units shift). Thermostability of the deamidated variants was similar to the wild-type lysozyme and protease digestion studies suggested an increased accessibility and therefore a higher flexibility within the Asp enriched lysozyme variants (Kato et al. 1992).

Non-enzymatic deamidation of proteases by exposure to high ionic strength at neutral pH was for the first time reported in 1988, during crystallization of bovine trypsin (Kossiakoff 1988). Bovine trypsin was exposed to high ionic strength at neutral pH for more than one year. During the latter treatment, 3 out of 13 asparagine residues were deamidated, resulting in aspartic acid (Kossiakoff 1988). However, no systematic study has, to our best knowledge, been reported on property alterations (thermostability, pH-profile dependence, specific activity) caused by deamidation in proteases.

2.4 Material and Methods

All chemicals were of analytical-reagent grade or higher quality and were purchased from Sigma-Aldrich (Taufkirchen, Germany) and Applichem (Darmstadt, Germany). All enzymes were purchased from Fermentas (St. Leon-Rot, Germany) and New England Biolabs (Frankfurt, Germany). Thermal cycler (Eppendorf Mastercyler proS, Hamburg, Germany) and thin-wall PCR tubes (Multi-ultra tubes, 0.2 mL, Carl Roth, Karlsruhe, Germany) were used in all PCRs. The PCR volume was always 25 µL. The amount of DNA in cloning experiments was quantified by using a NanoDrop photometer (Thermo Scientific NanoDrop1000, Braunschweig, Germany).

2.4.1 *Bacillus gibsonii* alkaline protease (BgAP)

BgAP gene along with its pre-pro-sequence including the Bacillus promoter was inserted into the pHY300PLK shuttle vector (Takara Bio Inc, Shiga, Japan). The generated construct was named pHYBgAP. *BgAP* gene is translated into a

270 amino acid mature protease with a size of approximately 27 kDa. As an alkaline protease, BgAP has an optimum pH of 11 (Siegert et al. 2009).

2.4.2 Sequence alignment

The sequence alignment of 15 BgAP related proteases having an amino acid identity ranging from 79 % to 31 % according to BLASTP (Altschul et al. 1997; PDB code: 1MPT [79 %], 1SVN [79 %], 1AH2 [78 %], 1C3L [56 %], 1SIB [55 %], 1SCJ [54 %], 1GNS [51 %], 1DBI [47 %], 1THM [45 %], 2XRM [41 %], 2Z2X [40 %], 2IXT [40 %], 2GKO [39 %], 2ID8 [37 %], 1WMD [31 %]) was performed with the multiple sequence alignment function of Clone Manager 9 Professional Edition (Sci-Ed software, Cary, USA). Residues were defined as not conserved with an identity of less than 60 % in the alignment.

2.4.3 Site-directed mutagenesis

Site-directed mutagenesis of *BgAP* was performed as previously described (Wang and Malcolm 1999) at 18 selected sites on *BgAP* in the pHYBgAP construct. The site specific primers are listed in Table 2.

For the site-directed mutagenesis PCR (First stage: 98°C for 30 s, 1 cycle; 98°C 10 s/55°C, 30 s/72°C, 2 min, 4 cycles. Second stage: 98°C for 30 s, 1 cycle; 98°C, 10 s/55°C, 30 s/72°C 3 min, 12 cycles; 72°C for 3 min, 1 cycle), PhuS DNA Polymerase (2 U), 0.20 mM dNTP mix, 0.2 µM of each primer together with template (20 ng; pHYBgAP) were used. Following the PCR, *DpnI* (20 U; New England Biolabs) was supplemented for template digestion, and incubated overnight at 37°C. The *DpnI* digested PCR products were purified by using a NucleoSpin Extract II Purification Kit (Macherey-Nagel, Düren, Germany) and transformed into *E. coli* DH5α (Inoue et al. 1990). The *E. coli* DH5α mutants were cultivated (250 rpm, 37°C, 16 h) and plasmids were isolated using a NucleoSpin Plasmid Kit (Macherey-Nagel) followed by transformation into protease deficient *B. subtilis* DB104 strain (*nprR2 nprE18* and Δ*aprA3*; Kawamura and Doi 1984; Vojcic et al. 2012) for extracellular BgAP production.

2.4.4 Cell culture and expression

Expression of the generated protease variants in *B. subtilis* DB104 was performed in microtiter plates (Greiner, Frickenhausen, Germany). For expression buffered LB media (pH 7.6; 1 % (w/v) tryptone, 0.5 % (w/v) yeast extract and 1 % (w/v) sodium

chloride, 17 mM potassium dihydrogen phosphate, 72 mM dipotassium hydrogen phosphate) was used. The volume of 10 µL preculture (200 µL, 900 rpm, 37°C, 18 h, and 70 % humidity) was used to inoculate the mainculture (150 µL, 900 rpm, 37°C, 24 h, and 70 % humidity). The protease was separated from the cells by centrifugation (Eppendorf 5810R; 4°C, 3,220 *g*, 20 min) and the obtained protease containing supernatant was used for analysis.

Table 2: Primers used for site-directed mutagenesis of the *BgAP gene*.

Position	Sequence (5‘ to 3‘)
F_N60D	GTCAACAACGGCTGATCTAGATGGTCATGGTACTC
R_N60D	GAGTACCATGACCATCTAGATCAGCCGTTGTTGAC
F_N97D	TAAAGGTCCTAGGAGCAGATGGTAGAGGAAGCGTG
R_N97D	CACGCTTCCTCTACCATCTGCTCCTAGGACCTTTA
F_N115D	CTAGAGTGGGCTGCAACGGATAACATGCATATTGC
R_N115D	GCAATATGCATGTTATCCGTTGCAGCCCACTCTAG
F_N154D	GCGGCTACCGGTAACGATGGTACTGGTTCCA
R_N154D	TGGAACCAGTACCATCGTTACCGGTAGCCGC
F_N167D	GCTACCCAGCTCGTTATGCAGACGCAATGGC
R_N167D	GCCATTGCGTCTGCATAACGAGCTGGGTAGC
F_N205D	GAATACAAAGCACATACCTAGATAATAGCTATGCTAGTATGCCTG
R_N205D	CAGGCATACTAGCATAGCTATTATCTAGGTATGTGCTTTGTATTC
F_N236D	GTTAAACAAAAAAATCCATCTTGGGATGCGACTCAAATTCGTAATC
R_N236D	GATTACGAATTTGAGTCGCATCCCAAGATGGATTTTTTTGTTTAAC
F_N242D	GAATGCGACTCAAATTCGTGATCATTTGAAAAATACTGCGAC
R_N242D	GTCGCAGTATTTTTCAAATGATCACGAATTTGAGTCGCATTC
F_N250D	CATTTGAAAAATACTGCGACGGATCTAGGAAACTCATCTC
R_N250D	GAGATGAGTTTCCTAGATCCGTCGCAGTATTTTTCAAATG
F_N253D	CTGCGACGAATCTAGGAGACTCATCTCAATTTGGTAG
R_N253D	CTACCAAATTGAGATGAGTCTCCTAGATTCGTCGCAG
F_Q12E	CATGGGGTATTACACGTGTAGAGGCTCCCACTGTGCATAATC
R_Q12E	GATTATGCACAGTGGGAGCCTCTACACGTGTAATACCCCATG
F_Q37E	GCTATACTTGATACAGGTATAGCTGAGCATAGTGATTTAAC
R_Q37E	GTTAAATCACTATGCTCAGCTATACCTGTATCAAGTATAGC
F_Q107E	CGTGAGTGGGATTGCTGAGGGTCTAGAGTGGGC
R_Q107E	GCCCACTCTAGACCCTCAGCAATCCCACTCACG
F_Q176E	CTGTAGGAGCGACTGACGAGAACAACAGACGTGCGAG
R_Q176E	CTCGCACGTCTGTTGTTCTCGTCAGTCGCTCCTACAG
F_Q185E	GACGTGCGAGCTTTTCTGAGTATGGCACAGGAATTG
R_Q185E	CAATTCCTGTGCCATACTCAGAAAAGCTCGCACGTC
F_Q200E	GCACCTGGTGTTGGAATAGAGAGCACATACCTAAATAATAGC
R_Q200E	GCTATTATTTAGGTATGTGCTCTCTATTCCAACACCAGGTGC
F_Q230E	GGAGTAGCTGCGCTTGTTAAAGAGAAAAATCCATCTTGGAATGCG
R_Q230E	CGCATTCCAAGATGGATTTTTCTCTTTAACAAGCGCAGCTACTCC
F_Q256E	CTAGGAAACTCATCTGAGTTTGGTAGTGGACTAG
R_Q256E	CTAGTCCACTACCAAACTCAGATGAGTTTCCTAG

2.4.5 Skim milk detection system

The volume of 190 µL of substrate solution (2 % skim milk, 100 mM Tris/HCl, pH 8.6) was transferred in each well (flat bottom microtiter plates; Greiner, Frickenhausen, Germany) and 10 µL of supernatant were supplemented for quantification of proteolytic activity of secreted BgAP and BgAP variants. Protease activity was monitored by the decrease in absorbance with the a microtiter plate reader (650 nm, 23°C; Tecan Sunrise, Crailsheim, Germany). The decrease on absorbance at 650 nm across 5 min of incubation at 23°C was measured in Absorbance Units per min (AU/min).

2.4.6 Residual activity assay for thermal resistance

Thermal inactivation of protease variants in the cell culture supernatant was performed by incubating 25 µL of supernatant for 10 min at 60°C (Eppendorf Mastercyler proS). Residual activity was detected by using the skim milk detection system described in the previous paragraph. The residual activity (%) was defined as the activity (AU/min) of the thermal treated sample divided by the activity (AU/min) of the untreated sample, represented in percent.

2.4.7 Native polyacrylamid gel electrophoresis

Native gels (GE Healthcare, PhastGel Gradient, 8-25 % acrylamide, 112 mM acetate, 112 mM Tris, pH 6.4, Freiburg, Germany) and buffers (GE Healthcare, PhastGel Buffer Strips Native, 0.88 M L-alanine, 0.25 M Tris, pH 8.8) were used for protein electrophoresis to analyse the charge differences of BgAP and its variants. The separation was accomplished in the PhastSystem (GE Healthcare, PhastSystem Automated Electrophoresis Development Unit) at 15°C with reversed polarity of the electrodes. The protease bands were visualized using an activity overlay detection system with a gelatine photographic film (Paech et al. 1993).

2.4.8 Production and purification of BgAP and variants

BgAP and its variants were subcloned into a high-expression vector (provided by Henkel AG & Co. KGaA) and transformed into *B. subtilis* DB104. Production of BgAP and its variants was accomplished in shaking flasks (500 mL) using MLBSP media (100 mL, 180 rpm, 37°C, 48 h; Wilson et al. 1999). A clear supernatant was obtained after centrifugation (Eppendorf 5810R, 3,220 *g*, 40 min, 4°C). The clear supernatant was dialyzed in HEPES buffer (9 L, pH 7.8, 20 mM, 4°C, 14 h) using a dialysis

membrane with a 3.5 kDa cut off (Spectrum Spectra/Por 3, Breda, Netherlands). The conductivity was, if required, adjusted to 1.2 mS/cm with demineralized water.

The dialysed supernatant was loaded into an anion exchange chromatography column (GE Healthcare Q Sepharose Fast Flow; equilibrated with HEPES buffer (pH 7.8, 20 mM) which was connected in series to a cation exchange column (GE Healthcare SP Sepharose Fast Flow; equilibrated with HEPES buffer (pH 7.8, 20 mM). The flow rate and pressure were monitored by an ÄKTA*explorer* (GE Healthcare). The anion exchange column acts as a “negative purification” and BgAP or its variants bound to the cation exchange column at pH 7.8. For elution of BgAP the anion exchange column was disconnected and a linear gradient of sodium chloride up to 1 M in HEPES buffer (pH 7.8, 20 mM) was used to elute the absorbed proteins from the cation exchange column. The peak fractions were pooled together.

Protein concentration of purified BgAP variants was normalized by measuring the total protein concentration using the BCA protein assay kit (Thermo Scientific) and the Experion System (Experion Pro260 chip; Bio-Rad Laboratories Experion Automated Electrophoresis System, München, Germany).

2.4.9 pH-activity profile of purified BgAP and variants

Proteolytic activity on macromolecular substrate was assessed using azo-dye labelled collagen, as previously reported (Chavira et al. 1984). Azocoll (Merck/Calbiochem Azocoll^{TM} MESH >100, Darmstadt, Germany) was used in final concentrations of 1 % (w/v) in 100 mM glycin and a pH range of 8.5 to 12 with adjustments made using 2 M NaOH and HCl. The proteolytic assay was performed in V-bottom microtiter plates (Corning, Wiesbaden, Germany) with protease concentrations of 40 nM. After reaction (Eppendorf Thermomixer comfort; 700 rpm, 25 min, 40°C) the samples were incubated for 1 min on ice and centrifuged (Eppendorf 5810 R; 3,220 *g*, 20 min, 4°C). The absorbance of released azo-dye in 100 µL supernatant was detected at 520 nm in a Tecan Sunrise using flat bottom microtiter plates (Greiner).

The complete digestion of 1 mg substrate resulted in an absorbance value at 520 nm of 1,506. One unit was defined as the enzyme activity producing an increase in absorbance of 1.0 per minute at pH 9.5 in 100 mM glycine/NaOH buffer. One unit corresponded to the digestion of 0.664 mg substrate.

2.4.10 Thermal resistance of purified BgAP and variants

Thermal inactivation was monitored by incubating BgAP and its variants (3.5 µg/mL; 20 min) at various temperatures in a thermal cycler (Eppendorf Mastercyler proS, gradient 46 - 62°C) in 100 mM Glycin/NaOH buffer (pH 8.5 and pH 10.5, 25 µL) and 250 mg/mL bovine serum albumin as an additive to prevent self-digestion. Residual activity was detected by using the skim milk detection system. The enzymes incubated on ice were considered to be the control and represent 100 % activity.

2.4.11 Homology modelling and surface residue determination

The three-dimensional structure model for BgAP was generated using the homology model routine from YASARA (Krieger et al. 2002) based on the coordinates of a serine protease from *Bacillus lentus* (PDB ID: 1GCI, Kuhn et al. 1998) YASARA was used for visualization of the molecular structures. The surface residues of the generated BgAP model were determined using Discovery Studio 3.1 Visualizer Software (Accelrys, Inc., San Diego, USA) using the solvent accessibility dialog. A residue was defined as “exposed” or “buried” if the solvent accessibility was greater than 25 % or less than 10 % of the maximum solvent accessibility, respectively.

2.5 Results

The results section is divided in four parts: 2.5.1 Screening of proteolytic activity and thermal resistance, 2.5.2 Net charge difference analysis of BgAP and its variants, 2.5.3 pH-dependent activity profile of purified BgAP and its variants, and 2.5.4 Thermostability profiles of purified BgAP and its variants.

2.5.1 Screening of proteolytic activity and thermal resistance

All generated variants were constructed as described in chapter 2.4 Materials and Methods. For the selection of the targeted Asn and Gln residues the following selection criteria were defined: (1) surface exposed, (2) not conserved, and (3) neighboured by a glycine residue three amino acids up- or downstream from the targeted Asn or Gln residue. The third selection criteria rose empirically and is based on the deamidation promoting property of a glycin one amino acid sequence upstream of the target Asn or Gln (Kossiakoff 1988).

Figure 5 categorizes all Asn and Gln residues found in BgAP according to the three defined selection criteria. The underlined positions, a subset of in total 18 positions fulfilling at least two selection criteria, were subjected to site-directed mutagenesis and analyzed in microtiter plate format.

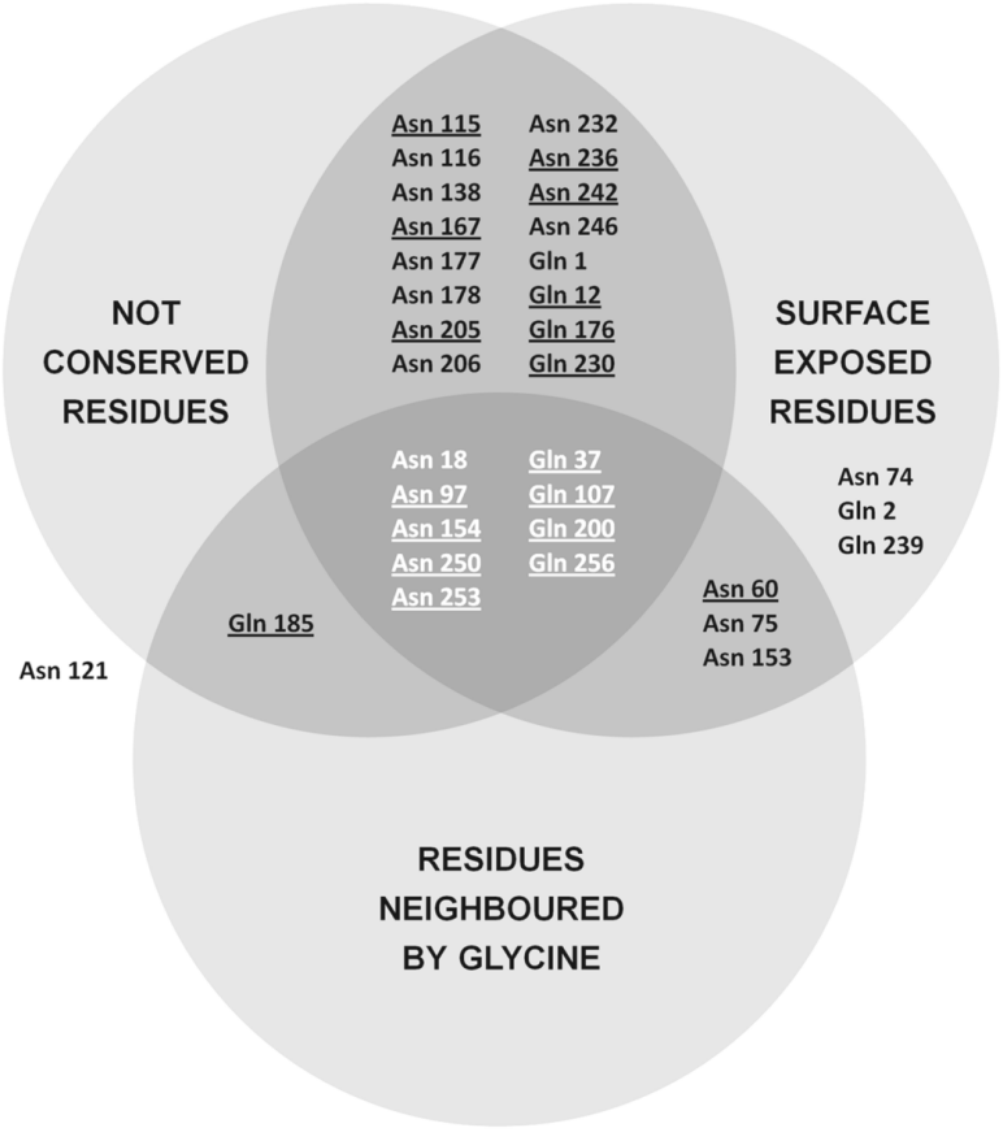


Figure 5: Representation of all asparagine and glutamine residues present in BgAP assigned to the three defined selection criteria (Not conserved residues, Surface exposed residues, and Residues neighboured by glycine). The selected and by site-directed mutagenesis mutated positions are underlined.

Variants having the substitutions N60D (surface exposed and neighbored by glycine) and Q185E (not conserved and neighbored by glycine) fulfill only two selection criteria and resulted in decreased proteolytic activity (N60D) or in an inactive variant

(Q185E). For both variants no residual activity was detected after thermal treatment (10 min, 60°C).

The selection criteria surface exposed (1) and not conserved (2) are fulfilled by 16 positions. A subset of eight positions was selected and substituted. Six variants (N205D, N236D, N242D, Q12E, Q176E and Q230) showed proteolytic activities in the range of wt-BgAP (±18 %). Variant N115D was inactive and variant N167D showed a decrease in proteolytic activity of 23 % compared to wt-BgAP. In case of thermal resistance, two variants (N205D and Q12E) showed a decrease in residual activity of more than 16 %; however, the majority of the generated variants (N167D, N236D, N242D, and Q176E) were in the range of wt-BgAP (±10 %). An exception for this group is variant Q230E, which showed an increased thermal resistance of 22 % compared to wt-BgAP.

Nine residues were identified which fulfill all three defined selection criteria (surface exposed, not conserved and neighbored by glycine). Out of these nine targetable residues, eight variants were generated and tested. Variant Q107E showed 47 % decreased proteolytic activity compared to wt-BgAP and variants N154D and N250D were proteolytically as active as wt-BgAP. The remaining five variants (N97D, N253D, Q37E, Q200E and Q256E) showed a more than 25 % increased proteolytic activity compared to wt-BgAP. The variants Q37E, N253D and N97D showed with 36 %, 35 % and 46 % the strongest increase in proteolytic activity, whereas variants Q200E and Q256E showed a moderate increase in proteolytic activity (25 % and 29 %). The aforementioned increase in activity was not observed in variants that did not fulfill all three defined selection criteria (Table 3).

Regarding thermal resistance, the majority of the generated variants (N97D, N250D, Q37E, Q107E and Q200E) had a comparable residual activity to wt-BgAP (±10 %), except variant N154D (12 % residual activity and less stable than wt-BgAP). Variants N253D and Q256E showed a moderately residual activity (84 % and 71 %) and a significantly higher thermal resistance than the wt-BgAP (Table 3).

Variants N253D and Q256E were finally selected for detailed characterization due to their simultaneous increase in proteolytic activity and thermal resistance when compared to wt-BgAP. In addition, the double mutant N253DQ256E was generated to study possible cooperative or additive effects.

Table 3: Proteolytic activity of BgAP and muteins before and after thermal treatment. Proteolytic activity was determined from the clear supernatant of B. subtilis DB104 cultures using the skim milk activity detection system (2 % (w/v) skim milk, 100 mM Tris/HCl pH 8.6). The residual proteolytic activity (%) was defined as the activity (AU/min) of the thermal treated sample (10 min, 60°C) divided by the activity (AU/min) of the untreated sample, represented in percent. The muteins are sorted by the conformance of the three selection criteria.

	Relative proteolytic activity [% of wt-BgAP]	**Residual proteolytic activity [% of initial activity]**
WT	100 ± 5 %	38 ± 5 %
Surface exposed and neighbored by glycine		
N60D	32 ± 14 %	not active*
Not conserved and neighbored by glycine		
Q185E	not active*	not active*
Surface exposed and not conserved		
N115D	not active*	not active*
N167D	77 ± 5 %	28 ± 6 %
N205D	113 ± 7 %	16 ± 3 %
N236D	115 ± 2 %	35 ± 6 %
N242D	118 ± 7 %	31 ± 5 %
Q12E	114 ± 8 %	13 ± 1 %
Q176E	97 ± 2 %	36 ± 2 %
Q230E	107 ± 8 %	50 ± 5 %
Surface exposed, not conserved and neighbored by glycine		
N97D	146 ± 3 %	44 ± 2 %
N154D	107 ± 4 %	12 ± 5 %
N250D	113 ± 8 %	45 ± 6 %
N253D	135 ± 4 %	84 ± 4 %
Q37E	136 ± 1 %	38 ± 5 %
Q107E	53 ± 8 %	32 ± 5 %
Q200E	125 ± 7 %	45 ± 6 %
Q256E	129 ± 4 %	71 ± 3 %

*no activity detected under the defined conditions

2.5.2 Net charge difference analysis of BgAP and variants

The decreased positive net charge of N253D, Q256E and N253DQ256E was confirmed by native polyacrylamide gel electrophoresis (Figure 6) in which proteolytic activity was visualized using a gelatin-film overlay (Paech et al. 1993).

Figure 6 shows distinct protein bands representing the different net charges of wt-BgAP, variants with Δ-1 (N253D or Q256E) and Δ-2 (N253DQ256E). The BgAP-

protease and variants are overall positively charged in the native gel at pH 6.4 and the migration distance is decreased with reduced positive net charge.

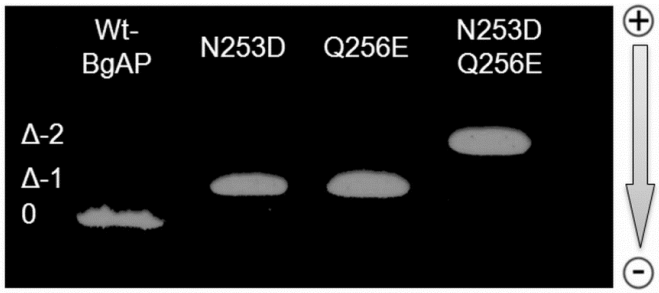


Figure 6: Native polyacrylamide gel electrophoresis and gelatin-film activity overlay of wt-BgAP and its variants (N253D, Q256E and N253DQ256E) which differ in the positive net charge.

2.5.3 pH-dependent activity profile of purified BgAP and variants

Figure 7a shows the pH dependence of the proteolytic activity profiles of purified wt-BgAP, N253D, Q256E and N253DQ256E which was determined using the Azocoll proteolytic assay (pH range 8.5 to 12; 100 mM Glycin/NaOH buffer).The Azocoll assay is reliable over a broad pH range, whereas skim milk spontaneously hydrolyzes at high alkaline conditions. BgAP shows an activity profile with a maximum at pH 11. The activity of BgAP at pH 8.5 and pH 12.0 was decreased to 50 % and 83 %, respectively. Variants N253D and Q256E showed equal pH dependence but differ from wt-BgAP since their pH-activity optimum is shifted by one unit to pH 10.0, without losing specific activity (Figure 7a,b). The activity of N253D and Q256E is reduced to 71 % and 72 % of the maximum at pH 8.5 and pH 12.0 respectively. The double mutant N253DQ256E shows also a shift on the pH-activity optimum from pH 11 to pH 10, with comparable specific activity to wt-BgAP at pH 11. The pH dependent activity of N253DQ256E is reduced to 90 % and 73 % at pH 8.5 and pH 12.0 respectively (Figure 7a,b). The pH-activity profiles are in general shifted with reduced positive net charge to lower pH values leading to higher activities at reduced pH and reduced activity at alkaline pH.

Figure 7c shows that, normalized on the wt-BgAP activity, the residual pH-activity dependence of the muteins N253D, Q256E and N253DQ256E at a pH range from 8.5 to 12.0. Variants N253D and Q256E show a higher activity compared to wt-BgAP at pH values lower than 10. The wt-BgAP has at alkaline conditions pH >10 a superior proteolytic activity than N253D and Q256E.

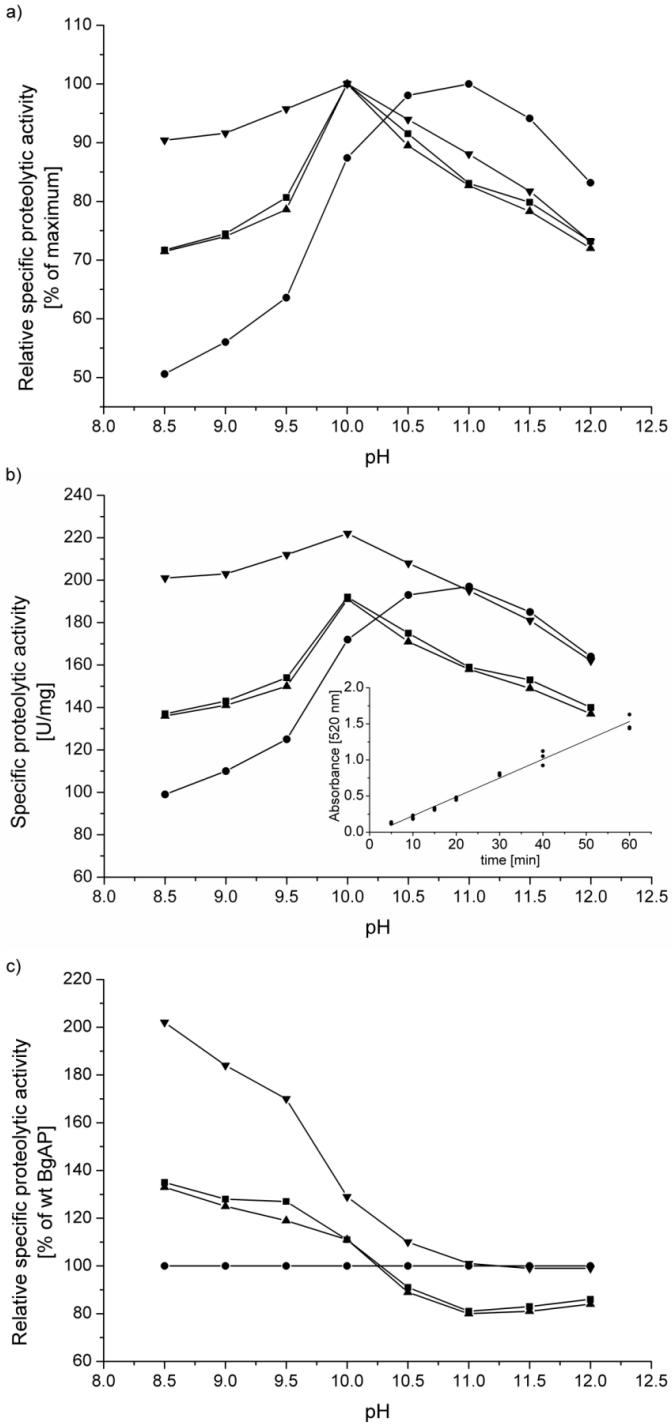


Figure 7: (● wt-BgAP, ■ N253D, ▲ Q256E, and ▼ N253DQ256E): **a)** pH-activity profile as percentage of maximum activity of each variant, **b)** Specific proteolytic activity determined with the Azocoll detection system (520 nm); inserted Figure shows the linear detection range of the Azocoll system, and **c)** Relative specific proteolytic activity as percentage of wt-BgAP activity at varied pH values. For activity measurements were employed 40 nM of wt-BgAP/muteins and 1 % (w/v) Azocoll in 100 mM glycin/NaOH buffer (pH 8.5 to 12). After 25 min incubation at 40°C and 20 min centrifugation at 3,220 *g* the absorbance of the supernatant was measured at 520 nm. Three measurements showed a standard deviation below 15 %.

The double mutant N253DQ256E shows at pH higher than 11 a nearly identical proteolytic activity compared to the single mutants. On the other hand, at lower pH

values the proteolytic activity exceeds significantly the values of the corresponding single mutants N253D and Q256E. At pH 8.5 N253DQ256E shows a ~2.0-fold improvement in specific activity compared to wt-BgAP activity.

2.5.4 Thermostability profiles of purified BgAP and variants

Figure 8a and b show the residual activity of purified wt-BgAP, N253D, Q256E and N253DQ256E after 20 min incubation at different temperatures and at two pH values (pH 8.5 and 10.5). The generated variants showed a residual activity improvement of ~28 % and ~37 % over wt-BgAP after being incubated at 53.2°C and 55.4°C (20 min; pH 8.5). When incubated at pH 10.5, the average residual activity of the BgAP muteins was improved by ~23 % and 18 % compared to wt-BgAP (51.1°C and 53.2°C; 20 min).

The T_{50} value is defined as the temperature at which the residual activity equals 50 % of the initial proteolytic activity after 20 min incubation at varied temperatures (gradient 46-62°C) in Glycin/NaOH buffer (100 mM, pH 8.5 and pH 10.5).

The thermostability constant T_{50} of wt-BgAP and the variants is lowered by 3-4°C at pH 10.5 when compared to pH 8.5. At pH 8.5, the T_{50}-value for N253D, Q256E and N253DQ256E was found to be increased by 2.3°C, 2.2°C and 2.5°C, in case of pH 10.5 the differences were determined to be 1.1°C, 1.6°C and 1.4°C respectively (Table 4).

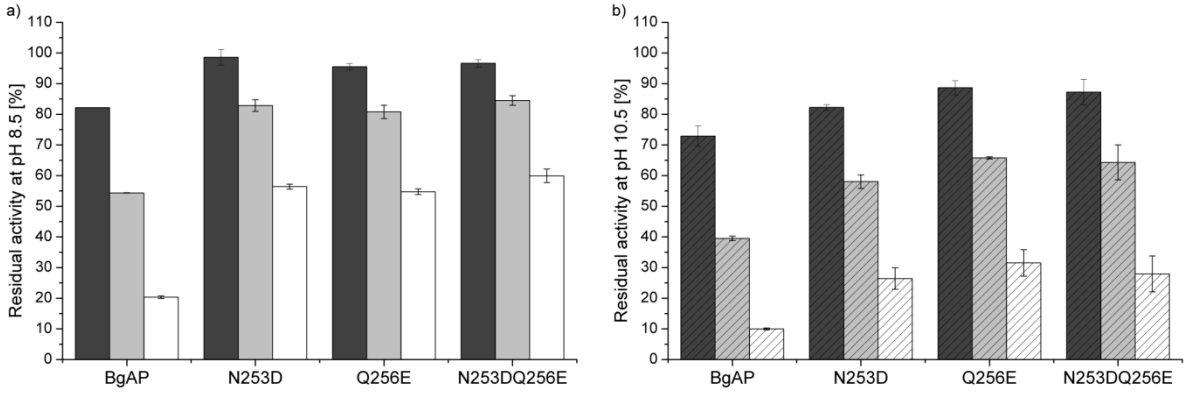


Figure 8: Residual enzyme activity of wt-BgAP, N253D, Q256E and N253DQ256E: a) after being incubated at 51.1°C (black), 53.2°C (grey) and 55.4°C (white) at pH 8.5 for 20 min, b) after being incubated at 49.2°C (black dashed), 51.1°C (grey dashed) and 53.2°C (white dashed) at pH 10.5 for 20 min. Heat incubation was performed with 4 µM of wt-BgAP and muteins (N253D, Q256E, N253DQ256E) for 20 min in 100 mM Glycin/NaOH buffer at pH 8.5 and 10.5 and 250 mg/mL bovine serum albumin. Residual activity was measured with the skim milk detection system (2 % (w/v) skim milk, 100 mM Tris/HCl, pH 8.6).

Table 4: The calculated T_{50} values of wt-BgAP, N253D, Q256E and N253DQ256E are defined as the temperature at which 50 % of the initial activity is retained after 20 min incubation (100 mM Glycin/NaOH, pH 8.5 or pH 10.5, no additional $CaCl_2$ supplemented). Average values of two independent measurements are shown and deviations are calculated from the fitted sigmoidal curves.

Mutant	pH 8.5		pH 10.5	
	T_{50} [°C]	Increase [°C]	T_{50} [°C]	Increase [°C]
Wt-BgAP	53.4 ± 0.3	-	50.5 ± 0.1	-
N253D	55.6 ± 0.3	2.2	51.6 ± 0.2	1.1
Q256E	55.6 ± 0.1	2.2	52.0 ± 0.1	1.5
N253DQ256E	55.8 ± 0.2	2.4	51.8 ± 0.2	1.3

2.6 Discussion

Enzymatic and non-enzymatic deamidation has proven to have a significant influence on enzyme properties like activity, thermal resistance and pH optimum (Capasso et al. 1989; Catak et al. 2009; Kato et al. 1987a; Kato et al. 1987b; Kato et al. 1992; Robinson and Robinson 2004). In all these reports enzyme properties were altered by an increasing number of negatively charged surface residues. The three selection criteria to target surface residues represent a straight forward and empirical sequence alignment-based strategy to identify Asn and Gln residues which can preferentially be targeted for deamidation by site-directed mutagenesis to improve activity, thermal resistance or to shift pH optima. The intention was not to simulate or to study factors which affect the naturally occurring deamidation reaction. The aim was to study the effect of the introduced negative charges to properties of the BgAP protease. The substitutions of the deamidation type (Asn → Asp, Gln → Glu) have accordingly to the BLOSUM62 matrix a score of one and two, respectively (Henikoff and Henikoff 1992) so that the generated variants harboring substitutions of the deamidation type should remain active. The three proposed selection criteria dramatically reduce the potentially targetable residues to introduce negative charges. The first selection criterion of excluding Asn and Gln residues which are not exposed to the surface will further reduce the number of inactive enzyme variants, as demonstrated with variant Q185E which was not surface exposed and inactive (Table 3).

In a second step non-conserved Asn and Gln residues are determined through sequence alignment with related subtilisin proteases to exclude structurally important positions. In the third step Asn and Gln residues are preferred which are neighbored

by a deamidation promoting amino acid within three amino acids up- or downstream on the sequence (like glycine; Bodanszky and Kwei 1978; Kossiakoff 1988; Teshima et al. 1991). The third selection criterion rose empirically and increases the chance to identify protease variants with an altered pH-profile or increased thermal resistance.

The protease BgAP consists of 270 amino acids out of which 113 amino acids are surface exposed. The surface amino acids include 31 Asn and Gln residues, wherein 26 residues are not conserved. Only 9 not conserved residues are neighbored by a glycine (three amino acids up- or downstream). By applying our developed selection criteria and the “deamidation type” substitution pattern (Asn → Asp, Gln → Glu) only nine variants have to be generated and studied in order to shift and improve pH dependent activity and thermal resistance of BgAP (compare Figure 5 and Table 3).

Out of these nine variants eight were generated and investigated. Five variants had an increased proteolytic activity at pH 8.6. Two of the generated variants (N253D, Q256E) fulfill all three selection criteria and show a strong increase in thermal resistance when compared to wt-BgAP (Table 3).

Out of the 18 Asn and Gln residues that don’t fulfill the third criterion (neighboured by glycine) 10 corresponding variants were generated, resulting in no variant with increased proteolytic activity when compared to wt-BgAP. Only one variant (Q230E) had an increased residual activity after thermal treatment when compared to wt-BgAP. The two variants N60D (conserved) and Q185E (not surface exposed) showed as expected decreased or no proteolytic activity when compared to wt-BgAP.

A success rate of five out of eight modifications, resulting in an increased proteolytic BgAP activity, proves that sequence alignment-based deamidation of rational selected positions by site-directed mutagenesis represents a successful approach to improve activity by surface charge engineering.

Due to the extent of improvements, positions N253D, Q256E and combination N253DQ256E were finally selected for detailed characterization in terms of: a) activity increases (Figure 7b,c,) influences on pH-optima (Figure 7a), and c) thermal resistance improvements (Table 4, Figure 8). Substitutions N253D and Q256E are located on the same loop and N253DQ256E was generated to study possible cooperative or additive effects.

Deamidated variants showed up to 2.4°C increased T_{50} for N253DQ256E at pH 8.5. No additive or cooperative effect could be observed when compared to the individual single mutant Q256E and N253D (Table 4, Figure 8). Wt-BgAP and the three deamidated variants showed at pH 10.5 a lower increase in thermal resistance compared to pH 8.5, but still significant. The determined residual activity is an indicator for thermal stability since the mature BgAP protease is not capable to refold correctly in absence of the pro-peptide (Siezen and Leunissen 1997).

The pH activity profiles were determined using an azo-dye labeled collagen as a substrate. Azo-dye labeled collagen is a complex protease substrate and close to application conditions in the detergent industry and therefore a better choice than short chain peptides with chromogenic or flourogenic cleaving site. The pH activity profiles for deamidated variants (N253D, Q256E, and N253DQ256E) were significantly shifted with a general trend towards higher proteolytic activity at lowered pH values resulting in a pH-optimum shift from 11 to 10. The single muteins (N253D and Q256E) show similar pH profile shifts. The combined variant N253DQ256E loses at pH 8.5 only ~10 % of its activity whereas as the single mutants have a ~30 % and the wt-BgAP has a 50 % reduced activity (Figure 7). Figure 7c shows that the relative improvement in specific proteolytic activity for the double mutant exceeds the additive effects of the two single mutants. The latter result suggests that negative charges on surface could cause an increase in proteolytic activity through additive and/or cooperative effects.

The effects of modifying surface charges on the catalytic activity of a thermolysin-like protease (pH-optimum: ~6.8) by inserting or removing charges on the protein surface were studied previously. Significant changes were observed in the activity of the thermolysin-like protease even when amino acid substitutions were as much as 25 Å apart from the active site (De Kreij et al. 2002).

In our case, the finally studied amino acid substitutions (N253D, Q256E, N253DQ256E) are located in a loop which is not close to the active site, however activity as well as thermal resistance were increased. A similar observation has been reported for other proteases by Miyazaki *et al.*, where the thermostability and low temperature activity of a psychrophilic subtilisin S41 was enhanced by directed evolution. The introduced mutations were, similar to our findings, far from the active site, and located in the outer loops of the protease (Miyazaki et al. 2000). The

substitutions generated in BgAP will likely cause electrostatic long range interactions which are challenging to be addressed by computational modeling since structural changes that are related to increased protein stability are commonly described to cause reinforcement on the overall structural rigidity, which most likely goes along with a decrease in catalytic performance. The challenge for a molecular understanding lies in the build-up of a model in which the flexibility, as prerequisite for high activity, has to be maintained or improved around the active site whereas global motions have to be reduced in regions at which unfolding occurs (Martinez et al. 2011).

In essence, a novel protein engineering approach was developed, independent or additive to other methods, to optimize protein properties like activity, thermal resistance and pH optima through surface charge engineering and “simulated” deamidation by site-directed mutagenesis.

3 DIRECTED PROTEASE EVOLUTION TOWARDS INCREASED INHIBITION BY BENZYLMALONIC ACID

3.1 Objective

The aim of this project was the identification of *Bacillus lentus* alkaline protease (BLAP) variants showing an increased inhibition towards benzylmalonic acid compared to the BLAP parent. Subsequently, the structure-function relationships were investigated to identify inhibitor-protease interactions, which could be responsible for increased inhibition. The target was to increase the inhibition of BLAP by benzylmalonic acid in order to match the performance of the industrial standard boric acid.

3.2 State of the art

3.2.1 Inhibitors and subtilisins used in liquid detergent applications

The use of enzymes in laundry detergents as extremely effective catalysts represents the largest field of industrial enzyme applications. Among others, proteases, lipases, amylases and cellulases are supplemented into detergent formulations for the efficient removal of specific stains (Kirk et al. 2002). In 2004, less than 15 different protease molecules were used in detergents worldwide, being all subtilisin proteases originated from *Bacillus* sp. (e.g. *Bacillus licheniformis, Bacillus clausii, Bacillus lentus, Bacillus alkaloophilus,* and *Bacillus halodurans*). Among these, proteases originated from *Bacillus lentus* (e.g. BLAP S and BLAP X) represented one third of the industrially used subtilisins (Maurer 2004). These enzymes are stable towards high temperature, high pH and surfactants used in liquid detergents. Proteases represent the major enzymatic component in detergents implying unspecific autolysis and proteolysis of enzymatic components during storage (Stoner et al. 2004). The optimization of stabilization and inhibition of the supplemented proteases is nowadays an interesting research field (Lund et al. 2011; Daugherty et al. 2012; Smoum et al. 2012) and is commercially accomplished by the addition of boron containing inhibitors e.g. boric acid (Severson 1985) or 4-formylphenylboronic acid (Nielsen and Deane-Wray 1999). The importance of inhibitors in detergents to

stabilize enzymes is illustrated by the fact that 6,400 tons of boric acid were used in the European market in 2004 explicitly for detergent applications (Human and Environmental Risk Assessment on ingredients of Household Cleaning Products (HERA); http://www.heraproject.com).

The major drawback of boric acid is its toxic potential, which was reported in long-term studies showing critical effects on male reproductive toxicity and developmental toxicity, when consumed in high doses (a 60 kg person would need to consume daily 3.3 g boric acid; Hubbard 1998). However, exposure to boric acid in detergent products under normal handling and use, by consumers' are very low (0.0453 µg boric acid/kg bw/day; HERA). The elimination of boron from liquid laundry detergents is nevertheless an interesting industrial and scientific objective.

In therapeutic studies, the identification of novel inhibitors for target enzymes is accomplished by the screening of enormous drug-like molecule libraries. As an example, out of 50,000 drug-like small molecules, five novel molecules were identified which efficiently inhibit the main proteinase of SARS-CoV $3CL^{pro}$ using a fully automated screening system (Blanchard et al. 2004).

Engineering recombinant target enzymes towards increased inhibition is an alternative to screening of immense libraries to inhibit detergent proteases. The favored tool to engineer novel protein properties with an unsolved mechanism is directed protein evolution. In 2004 the half of the commercially used detergent subtilisin proteases were protein engineered variants (Maurer 2004). A typical experiment comprises three major steps: (1) diversity generation, (2) screening to identify improved mutants out of a large pool of variants, and (3) isolating the gene encoding for the improved protein variants. These genes are subsequently used in iterative cycles of diversity generation and screening until the targeted improvement has hopefully been reached (Tee and Schwaneberg 2007; Figure 3).

3.3 Material and Methods

All chemicals were of analytical-reagent grade or higher quality and were purchased from Carl Roth GmbH (Karlsruhe, Germany), Sigma-Aldrich (Hamburg, Germany), Bachem (Bubendorf, Switzerland) and AppliChem (Darmstadt, Germany).

Oligonucleotides were purchased from Eurofins MWG Operon (Ebersberg, Germany) in salt-free form. Enzymes were purchased from New England BioLabs (Frankfurt am Main, Germany). Plasmid extraction and PCR purification kits were purchased from Macherey-Nagel (Düren, Germany). Microtiter plates (Greiner Bio-One GmbH, Frickenhausen, Germany) were incubated in a Microtron Infors shaker (Infors AG, Bottmingen, Switzerland). DNA concentrations were quantified using a NanoDrop photometer (Thermo Scientific NanoDrop1000, Braunschweig, Germany). A Mastercycler gradient (Eppendorf, Hamburg, Germany) and thin-wall PCR tubes (Multi ultra-tubes; 0.2 mL; Carl Roth, Germany) were used in all PCRs.

3.3.1 *Bacillus lentus* alkaline protease (BLAP)

The used *Bacillus lentus* alkaline protease (BLAP) gene codes for a 269 amino acid mature protease (BLAP, PDB: 1ST3A) and has a molecular weight of ~27 kDa. The BLAP parent contained two amino acid substitutions at positions G1A and R99E compared to the literature described BLAP (PDB ID: 1ST3; Goddette et al. 1992). The *BLAP* gene was contained in an industrial expression vector (pHKL) suitable for high level extracellular protease production in *Bacillus subtilis* DB104 (Kawamura and Doi 1984). The pHKLBLAP construct was kindly provided by Henkel AG & Co. KGaA (Düsseldorf, Germany).

3.3.2 Generation of Sequence Saturation Mutagenesis (SeSaM) libraries

Random mutagenesis libraries and DNA recombination of *BLAP* gene was performed using the Sequence Saturation Mutagenesis method (SeSaM; SeSaM-Biotech, Bremen, Germany) followed by the insertion of the modified genes in the pHKL vector using PLICing (Blanusa et al. 2010). For randomization, the mature sequence of the protease was chosen and the SeSaM template was generated by PCR using specially designed primers: SeSaM_FW_BLAP (5'-CACACTACCGCACTCCGTCGC TGACGGCCTTCATGCTC-3') and SeSaM_RV_BLAP (5'-GTGTGATGGCGTGAGGC AGCCAACGCACCTTTCAGCAAC-3'). SeSaM library generation was performed as previously described (Wong et al. 2008).

The resulting library was cloned into the pHKL vector using the PLICing method with specially designed primers for the amplification of the vector backbone PLIC_pHKL_RV (5'-CGCCATTGTCGTTACTTCTGC-3'), PLIC_pHKL_FW (5'-GACT TGTCAATGCAGAAGCGGC-3') and of the mature *BLAP* sequence PLIC_BLAP_RV (5'-CATTGACAAGTCCGCTTCCATACAAGTTC-3'), PLIC_BLAP_FW (5'-CGACAAT

GGCGCAATCAGTGC-3'). The phosphorothioated nucleotides are underlined. After the vector-backbone amplification, differing from the published method, a subsequent gel-extraction was performed to eliminate the PCR template. The iodine cleavage and hybridization reactions were performed as previously described (Blanusa et al. 2010). Subsequently, 20 µL of PLICing reaction were transformed into *B. subtilis* DB104 by protoplast transformation as previously described (Chang and Cohen 1979).

3.3.3 Generation of site-saturation and site-directed mutagenesis libraries

Site-saturation mutagenesis of the BLAP variants at positions 22, 37, 99, 192, 210 and 218 and site-directed mutagenesis at position 160 and 183 were performed according to a standard mutagenesis protocol (Wang and Malcolm 1999). The used primers are listed in Table 5. The PCR templates were methylated with *dam* methyltransferase (New England Biolabs) according to the manual, to guarantee complete *DpnI* (New England Biolabs) digestion of the used templates.

Table 5: Oligonucleotides used for site-directed mutagenesis and generation of SSM libraries.

Primer/Position	Sequence (5' to 3')
F_E99_SSM	GTTTAGGAGCCGACGGTNNKGGTGCAATCAGCTC
R_E99_SSM	GAGCTGATTGCACCMNNACCGTCGGCTCCTAAAC
F_I192_SSM	GCAGGGCTTGACNNKGTCGCACCAGGGGTAAAC
R_192_SSM	GTTTACCCCTGGTGCGACMNNGTCAAGCCCTGC
F_T22_SSM	GCCCATAACCGTGGATTGNNKGGTTCTGGTGTAAAAG
R_T22_SSM	CTTTTACACCAGAACCMNNCAATCCACGGTTATGGGC
F_T37_SSM	CTCGATACAGGTATTTCCNNKCATCCAGACTTAAATATTCGTG
R_T37_SSM	CACGAATATTTAAGTCTGGATGMNNGGAAATACCTGTATCGAG
F_S210_SSM	ACCCAGGTTCAACGTATGCCNNKTTAAACGGTACATCGATGGC
R_S210_SSM	GCCATCGATGTACCGTTTAAMNNGGCATACGTTGAACCTGGGT
F_S218_SSM	CTTAAACGGTACATCGATGGCTNNKCCTCATGTTGCAGGTGCAG
R_S218_SSM	CTGCACCTGCAACATGAGGMNNAGCCATCGATGTACCGTTTAAG
F_S160G_SDM	GTGCAAGCTCAATCGGCTATCCGGCCCG
R_S160G_SDM	CGGGCCGGATAGCCGATTGAGCTTGCAC
F_F183R_SDM	CAACCGCGCCAGCCGTTCACAGTATGGC
R_F183R_SDM	GCCATACTGTGAACGGCTGGCGCGGTTG

3.3.4 Simultaneous saturation mutagenesis of three sites (3xSSM)

The 3xSSM was performed simultaneously at position 181, 183 and 185. The generation was accomplished by using NDT codon degeneracy for position 181 and 183, involving 12 codons, reducing the number of amino acids to twelve (Arg, Asn,

Asp, Cys, Gly, His, Ile, Leu, Phe, Ser, Tyr, Val). These twelve amino acids represent a balanced mix of aliphatic, aromatic, polar and non-polar, negatively and positively charged amino acids (Reetz et al. 2008). Position 185 was saturated simultaneously using NDK codon degeneracy, resulting in 24 codons for 17 amino acids. In addition to the NDT combination, five more amino acids (Gln, Glu, Lys, Met, Trp) are included. The three degenerated codons were located on the primer 3xSSM_BLAP_PLIC_RV (5'-GCCCTGCGCCATAMHNTGAAHNGCTAHNGCGGTTGTTGTTTTGGTCAG-3'), which was used in combination with primer FW_BLAP_PLIC to amplify partly the mature gene sequence. For the amplification of the vector-backbone and the residual mature gene sequence primer 3xSSM_pHKL_PLIC_FW (5'-GGCGCAGGGCTT GACATTG-3') in combination with primer RV_pHKL_PLIC was used. Transformation was performed as described in the generation of SeSaM libraries.

3.3.5 Indicator plates for pre-screening

Skim milk was used as a substrate for proteolytic activity detection (Sokol et al., 1979). LB agar plates supplemented with 1 % (w/v) skim milk and 50 µg/mL kanamycin were used as pre-screening plates.

3.3.6 Production of BLAP and variants

Single colonies of *B. subtilis* DB104 which showed proteolytic activity in pre-screening plates were transferred into flat-bottom microtiter plates (pre culture plates) containing 200 µl of buffered LB medium (pH 7.6; 1 % (w/v) tryptone, 0.5 % (w/v) yeast extract and 1 % (w/v) sodium chloride, 17 mM potassium dihydrogen phosphate, 72 mM dipotassium hydrogen phosphate, 50 µg/mL kanamycin) and were incubated (37°C, 900 rpm, 18 h, 70 % humidity). The volume of 10 µL pre culture was used to inoculate the main culture (150 µL buffered LB, 37°C, 900 rpm, 24 h, 70 % humidity). After expression, the microtiter plates were centrifuged (3,220 *g*, 20 min, 4°C) to separate the cells from the secreted protease, which was obtained in the supernatant. The supernatant was transferred to a new microtiter plate for further analysis.

3.3.7 Screening for increased inhibition by reversible inhibitors

Screening without supplemented inhibitor was performed by adding 20 µL supernatant from each clone to 100 µL of 5 - 50 % Henkel detergent model matrix pH 8.6 containing 1.1 mM suc-AAPF-*p*NA (Bachem). The reaction was carried out in flat bottom microtiter plates (Greiner Bio-One GmbH) at room temperature (23°C).

The amount of *p*-nitroanilide released was measured at 410 nm for 5 min using a molar absorption coefficient of 8,480/M x cm. One unit is defined as the proteolytic activity necessary for the release of 1 µmol of p-nitroanilide per minute.

The abovementioned setup was used to screen for increased inhibition by benzylmalonic acid. The final benzylmalonic acid (Sigma) concentration was 30 mM (pH adjustment to 8.6 with NaOH required) corresponding to the necessary concentration to reduce the BLAP wild-type activity to ~80 % of its initial activity (without supplemented inhibitor).

For screening in the presence of boric acid, the same described conditions were used, where benzylmalonic acid was exchanged by boric acid (Sigma) in a final concentration of 14 mM.

3.3.8 Production and purification of BLAP and variants

BLAP and its variants were transformed into an industrial *Bacillus* production strain (provided by Henkel AG & Co. KGaA). Expression of BLAP and its variants was accomplished by fermentation (1.6 L) using industrial production media in a fed-batch process. The purification of BLAP and variants was performed as described in chapter 2.4.8 “Production and Purification of BgAP and its variants”.

Protein concentration of purified BLAP variants was normalized by measuring the total protein concentration using the BCA protein assay kit (Thermo Scientific) and the homogeneity was analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) using a 5 % stacking gel and a 15 % separating gel (Laemmli 1970).

3.3.9 Kinetic characterization of BLAP and variants

The purified BLAP variants were characterized using the synthetic tetrapeptide suc-AAPF-*p*NA (0.13 to 6.03 mM) in 5 % detergent model matrix (pH 8.6, ~50 mS/cm) and in Tris/HCl buffer (100 mM, pH 8.6). The proteolytic reaction was initiated by the addition of the suc-AAPF-*p*NA substrate solution (0.13 to 6.03 mM) to the purified protease variant (wild-type: 0.30 µM; M1: 8.62 µM; M2: 3.74 µM). The amount of *p*-nitroanilide released was measured at 410 nm for 5 min using a molar absorption coefficient of 8,480/M x cm. One unit is defined as the proteolityc activity necessary for the release of 1 µmol of *p*-nitroanilide per minute. The initial velocity data was

fitted using GraphPad Prism software (GraphPad software, San Diego, CA, USA; hyperbolic fitting).

3.3.10 Determination of IC_{50} values of BLAP and variants

The interaction between BLAP variants and benzylmalonic acid (Sigma) was compared by the specific inhibitor concentration required to reduce the protease activity to 50 % of the initial value. The proteolytic activity was determined using 10 µL purified enzyme (wild-type: 0,30 µM; M1: 8,62 µM; M2: 3,74 µM) and 100 µL 5 % detergent model matrix containing 3.3 mM suc-AAPF-*p*NA (Bachem) with adjusted pH 8.6 (NaOH) and conductivity to ~50 mS/cm (NaCl), which was the final conductivity reached by the addition of the maximum benzylmalonic acid concentration, with subsequent pH adjustment. Benzylmalonic acid concentration varied from 0.5 to 2.9 mM final concentration. Data analysis was performed using GraphPad Prism software (GraphPad software, San Diego, CA, USA; symmetrical sigmoidal shape fitting). The IC_{50} value represents the concentration that provokes a response half way between the maximal response and the maximally inhibited response.

3.3.11 Computational analysis

The three-dimensional structures of BLAP parent and its variants were generated using the homology model routine from YASARA (Krieger et al. 2002) based on the coordinates of the *Bacillus lentus* alkaline protease (PDB ID: 1ST3; Goddette et al. 1992). The obtained homology models were energy minimized and equilibrated using YASARA (Krieger et al. 2002). Simulations were performed using AMBER03 force field (Duan et al. 2003; Krieger et al. 2004). The TIP3P water box (Miyamoto and Kollman 1992) was neutralized (pH 8.6, 0.9 % NaCl; Krieger et al. 2006) and the MD was performed at 298K and 1 bar using periodic boundary conditions and particle mesh Ewald method for long range electrostatics (Essmann et al. 1995) for a duration of 2000 picoseconds .

Docking of benzylmalonic acid in the double deprotonated form to the protease variants was performed using AutoDock 4.2 with default docking parameters (Morris et al. 1998) and point charges assigned according to the AMBER03 (Duan et al. 2003) for the protein and GAFF (Wang et al. 2004) force field for the ligand using AM1/BCC point charges (Jakalian et al. 2002). The setup was done with the YASARA molecular modeling program (Krieger et al. 2002). The docking was

performed by fixing the residues in the catalytic triad (Ser215, His62, Asp32) and the removal of water molecules from the active site and the S1 binding pocket. The simulation cell was defined with 15 Å around the C α-atom of residue Asn153.

Multiple structural alignments of BLAP and its variants were performed using the MUSTANG algorithm (Konagurthu et al. 2006).

3.4 Results

The results section is divided in six parts: 3.4.1 Development of a detection system for altered BLAP inhibition, 3.4.2 Library generation and screening results, 3.4.3 Investigation of single amino acid substitution in variants V8 and V9, 3.4.4 Kinetic characterization of BLAP and variants, 3.4.5 Dose response analysis for BLAP and its variants tested with benzylmalonic acid and boric acid, and 3.4.6 Docking analysis.

3.4.1 Development of a detection system for altered BLAP inhibition

In order to identify BLAP variants with increased inhibition for benzylmalonic acid and/or boric acid, the widely used colorimetric suc-AAPF-*p*NA assay for proteolytic activity detection (Degering et al. 2010; Jaouadi et al. 2010; Stoner et al. 2004), was modified. In order to screen for variants with increased inhibition towards reversible inhibitors, the inhibitor concentrations necessary to reduce the initial proteolytic activity (without supplemented inhibitor) to 80 % was determined. The aim was to see a clear difference between the non-inhibited and the inhibited system allowing a sufficient range to detect improvements by an increased inhibited variant. An inhibitor concentration of 30 mM for benzylmalonic acid and 14 mM for boric acid resulted in a determined residual activity of ~80 %. The coefficient of variation is an important performance criterion of a screening system to determine its accuracy (Tee and Schwaneberg 2006). The developed assay conditions resulted in a true coefficient of variation of 6.8 % determined in 50 % detergent model matrix. The matrix concentration was varied from 5-50 % without influencing the performance of the inhibitor on BLAP and its variants.

3.4.2 Library generation and screening results

The aim of the directed evolution campaign was to generate and identify *Bacillus lentus* alkaline protease (BLAP) variants having increased inhibition towards benzylmalonic acid, compared to the BLAP parent. Two consecutive rounds of SeSaM (Wong et al. 2008) were performed yielding in the identification of four key residues (R1: 181, R2: 183, R3: 185 and R4: 160). Residue 181, 183 and 185 were finally saturated simultaneously on the parent S160G, with the subsequent inversion of the amino acid substitution to G160S, in order to see cooperative effects. The selected variants V8 (S160G, Q185R) and V9 (A181D, F183R) were purified and characterized in comparison with the wild-type. Figure 11 shows the genealogic tree of the directed BLAP evolution towards increased inhibition by benzylmalonic.

3.4.2.1 Sequence Saturation Mutagenesis (SeSaM)

The first round of SeSaM was constructed using BLAP wild-type as the parent, after prescreening of ~2000 clones with skim milk indicator plates 1800 proteolytically active clones (90 %) were analyzed with the microtiter plate based screening system for increased inhibition by benzylmalonic acid and boric acid, developed in this study. The identified variants (V1 [I102V]; V2 [A181D]; V3 [F183S]; V4 [Q185R]) were used as parents for the second round of diversity generation and showed with position 181, 183 and 185 already a mutational “hot-spot” (Table 6).

In the second round of SeSaM, from ~2800 prescreened clones, 1800 remained proteolytically active (73 %), which were finally screened for increased inhibition by benzylmalonic acid and boric acid in microtiter plates, resulting in four variants with altered inhibition behavior (V5 [E99G, I102V, F183S, I192V]; V6 [I102V, F183S]; V7 [T22A, T37N, Q185R, S210K, T218A]; V8 [S160G, Q185R]; Table 7). Variants V5 to V8 carried all at least one amino acid substitution from the first round with up to four additional mutations generated in the second SeSaM experiment.

Table 6: Summary of the generated and screened variants in SeSaM I and II.

	SeSaM I	SeSaM II
Parent:	BLAP	BLAP • V1 [I102V] • V2 [A181D] • V3 [F183S] • V4 [Q185R]
Pre-screening:	~2000 variants 10 % inacive variants	~2800 variants 27 % inactive variants
MTP-Screening:	1800 active variants 50 % detergent matrix	1800 active variants 25 % detergent matrix
Resulting variants:	BLAP • V1 [I102V] • V2 [A181D] • V3 [F183S] • V4 [Q185R]	BLAP • V5 [E99G I102V F183S I192V] • V6 [I102V F183S] • V7 [T22A T37N Q185R S210K T218A] • V8 [S160G Q185R]

3.4.2.2 Site-saturation mutagenesis

The additional amino acid substitutions on variants V5-V7 identified in the second round of SeSaM were analyzed using site-saturation mutagenesis (300 clones per position) yielding in no further improved variant. Residue 102 was excluded from further analysis since the increased inhibition from the single variant turned out to be not significant and the improvements of variants V5 and V6, were on a comparable level like variant V3 carrying the mutation F183S (Table 7).

3.4.2.3 Simultaneous saturation of three “hot-spots”

The identified “hot-spot” positions 181, 183 and 185 were saturated simultaneously (3xSSM) with a chosen set of degenerated codons (181: NDT, 183: NDT, 185: NDK) on the BLAP parent harboring glycin at position 160, which showed in variant V8 a dramatically cooperative effect with the introduced arginine at position 185 (Table 7). Further analysis of amino acid 160 exchanged to a set of 9 different amino acids resulted in equal benzylmalonic acid inhibition compared to the BLAP parent (Table 8).

In the 3xSSM experiment, 7150 clones were analyzed with the skim milk prescreening system and the remaining 1520 proteolytically active colonies (21 %) were screened for increased inhibition with the two inhibitors. The screening of the

3xSSM essentially resulted in variant V9 [A181D, F183R] having increased inhibition for benzylmalonic acid and boric acid (Table 7).

Variants V8 and V9 were finally selected for detailed characterization due to their increase in inhibition with benzylmalonic acid compared to BLAP parent.

Table 7: Summary of the identified variants after directed BLAP evolution and site-saturation mutagenesis analysis towards increased inhibition by benzylmalonic acid. The assay conditions resulted in a true coefficient of variation 6.8 %. The reported values are the average of three independent measurements.

Variant		Residual activity [%]	
		14 mM boric acid	30 mM benzylmalonic acid
	BLAP (parent)	79 ± 5 %	81 ± 6 %
V1	[I102V]	71 ± 5 %	79 ± 5 %
V2	[A181D]	47 ± 3 %	79 ± 5 %
V3	[F183S]	59 ± 4 %	70 ± 5 %
V4	[Q185R]	76 ± 5 %	63 ± 4 %
V5	[E99G I102V F183S I192V]	56 ± 4 %	56 ± 4 %
V6	[I102V F183S]	57 ± 4 %	70 ± 5 %
V7	[T22A T37N Q185R S210K T218A]	89 ± 6 %	58 ± 4 %
V8	[S160G Q185R]	68 ± 5 %	35 ± 2 %
V9	[A181D F183R]	38 ± 3 %	43 ± 3 %
V10	[S160G]	71 ± 5 %	80 ± 5 %
V11	[F183R]	57 ± 4 %	79 ± 5 %

Table 8: Inhibition analysis of position S160. Given values represent the reduction in proteolytic activity caused by the supplementation of 30 mM benzylmalonic acid.

Variant	Residual activity [%]
BLAP (parent)	79 ± 4 %
S160T	82 ± 6 %
S160A	75 ± 6 %
S160G	83 ± 8 %
S160E	82 ± 7 %
S160D	83 ± 10 %
S160I	94 ± 10 %
S160V	80 ± 3 %
S160F	80 ± 4 %
S160Y	81 ± 7 %

3.4.3 Investigation of single amino acid substitutions in variants V8 and V9

The missing variants carrying the single amino acid substitution of variants V8 and V9 were generated by site-directed mutagenesis and analyzed with the developed microtiter plate screening system (Table 7). In variant 10 the amino acid substitution S160G showed no altered inhibition compared to BLAP parent, but amino acid substitution Q185R decreased the residual activity caused by 30 mM benzylmalonic acid by a difference of 18%. The combination in the final variant 8 showed a cooparative effect of the two introduced mutations resulting in a decreased residual activity by a difference of 46% compared to BLAP parent. In variants 2 and 11 neither amino acid substitution A181D nor F183R showed an effect on the residual activity caused by 30 mM benzylmalonic acid. The combination of the two mutations in variant 9 causes a synergistic effect and decreased the residual activity by a difference of 38% compared to BLAP parent.

3.4.4 Kinetic characterization of BLAP and variants

Table 9 summarizes the kinetic data (k_{cat} and K_m) of purified BLAP parent and the obtained subtilisin variants 8 and 9.

All generated variants showed in presence and absence of the detergent matrix decreased k_{cat} values compared to BLAP parent, for hydrolysis of suc-AAPF-*p*NA. The detergent matrix decreased the k_{cat} values for V8, V9 and BLAP parent around 4.2, 3.5, and 1.6 times, respectively. The K_m value for BLAP parent is consistent in presence and absence of the detergent matrix, in contrast to V8 and V9, which showed decreased K_m values in presence of detergent matrix, meeting those of the BLAP parent.

Table 9: Kinetic parameters of BLAP parent ant its variants determined with varying suc-AAPF-*p*NA concentrations in given buffer systems; normalized to 1 µg (0.037 µmol) of protease. The reported values are the average of three measurements.

	5 % detergent matrix		**100 mM Tris/HCl pH 8.6**	
	K_m [mM]	**k_{cat} [s^{-1}]**	**K_m [mM]**	**k_{cat} [s^{-1}]**
BLAP (parent)	3.36 ± 0.37	45.55 ± 2.66	3.54 ± 0.37	72.80 ± 4.00
V8 [S160G Q185R]	3.96 ± 0.46	1.80 ± 0.12	6.20 ± 0.46	7.55 ± 0.36
V9 [A181D F183R]	2.67 ± 0.42	4.83 ± 0.36	5.88 ± 0.47	16.86 ± 0.84

3.4.5 Dose response analysis of BLAP and variants

Figure 9 shows the dose response curves of the tested variants with benzylmalonic acid and boric acid as inhibitor. In Table 10 the corresponding IC_{50} values are given. The IC_{50} value represents the required inhibitor concentration to reduce the initial activity (without inhibitor) to 50 %.

BLAP parent shows a ~9-times higher affinity towards boric acid (IC_{50}: 25.1 mM) than towards benzylmalonic acid (IC_{50}: 221.3 mM). Variant V8 shows with 19.4 mM for boric acid an equal IC_{50} value like its parent but for benzylmalonic acid a significantly ~9-times lowered IC_{50} value of 24.1 mM compared to its parent was determined. The benzylmalonic acid concentration is equal to the boric acid concentration required to reduce the initial parent activity to 50 %. Variant V9 shows a ~3-times lower IC_{50} value for boric acid and a ~3-times lowered IC_{50} value for benzylmalonic acid compared to its parent.

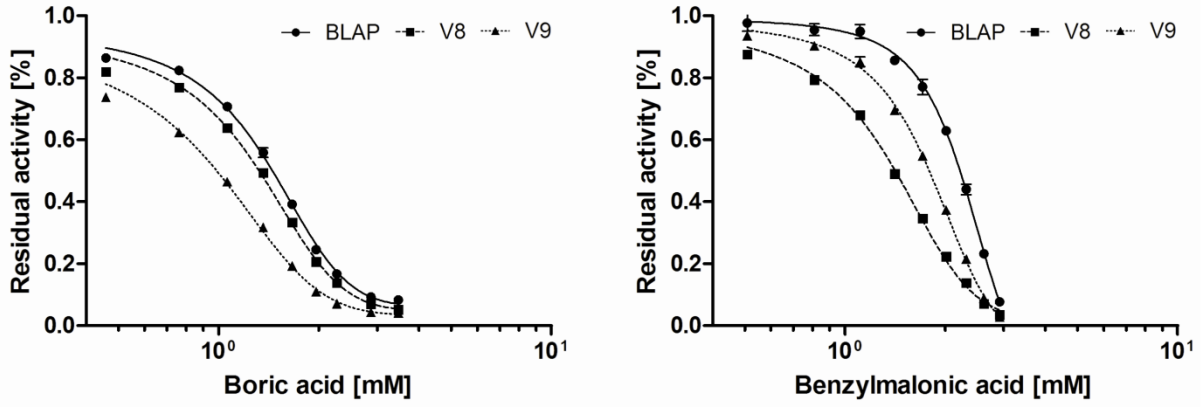


Figure 9: Dose-response curves for the BLAP parent, variant 8 [S160G Q185R] and variant 9 [A181D F183R] with **a)** boric acid and **b)** benzylmalonic acid as inhibitor.

Table 10: The IC_{50} value represents the required inhibitor concentration to reduce the initial activity (without inhibitor) to 50 %. The reported values are the average of three measurements.

Variant	Boric acid [mM]	Benzylmalonic acid [mM]
BLAP (parent)	25.1 ± 2.7	221.3 ± 36.1
V8 [S160G Q185R]	19.4 ± 2.3	24.1 ± 2.1
V9 [A181D F183R]	8.5 ± 1.1	68.5 ± 5.3

3.4.6 Docking analysis

Docking experiments were performed to identify the different possible binding conformations between benzylmalonic acid (BMA) and BLAP parent or the generated variants. Binding conformations with the lowest calculated docked energies belonging to the top-ranked cluster and showing interaction with the introduced amino acid substitutions were selected for each protease variant. The main focus of the docking analysis was the identification of possible ligand-receptor interactions between benzylmalonic acid and the protease molecule, and to suggest how this interaction changes with the substituted amino acids, possibly leading to binding conformations responsible for the increased inhibition. The docking scoring function provided an estimation of the free binding energy, which resulted for the top-ranked clusters in values in the range of the standard error of the method (2 kcal/Mol). The top-ranked docking poses were used for the identification of new ligand-receptor interaction in the variants compared with BLAP. In Figure 10a,b the prominent binding modes of benzylmalonic acid and BLAP parent are shown. Additionally identified prominent binding modes for benzylmalonic acid and the introduced amino acids of variants V8 and V9 are shown in Figure 10c-f.

In the case of BLAP parent, benzylmalonic acid showed two major orientations occupying the catalytic triad or S1 binding pocket (compare Figure 2), both essential for the hydrolysis reaction (Figure 10a,b). Benzylmalonic acid shows hydrogen bond interactions with Gln 185 and the backbone amine group of Asn 153 (Figure 10a). In the second shown orientation benzylmalonic acid blocks the active site by hydrogen bond interactions with the catalytic Ser 215 and Asn 153 (Figure 10b).

In variant V8 [S160G Q185R] the introduced amino acid Arg 185 and the backbone amine groups of Asn 153 and Ser 154 (not shown) form hydrogen bonds with the inhibitor benzylmalonic acid, which is occupying the S1 binding pocket (Figure 10c). In the second encountered binding pose the inhibitor forms hydrogen bonds with Arg 185 and a surface Arg 164 at the end of the S1 binding pocket (Figure 10d).

In the docking analysis of variant V9 [A181D F183R], the inserted Arg 182 and the Asn 153 form hydrogen bonds with benzylmalonic acid. Benzylmalonic acid is blocking residue Asn 153, which is essential for the hydrolysis reaction (Figure 10e). In the second novel binding pose the inhibitor builds two hydrogen bonds with residue Arg 183 (Figure 10f). Interestingly in both docking conformations a hydrogen

bond between the two residues Asp 181 and Arg 183 is formed (Figure 10e,f), which could stabilize or support the orientation of Arg 183 in a favored benzylmalonic acid binding conformation.

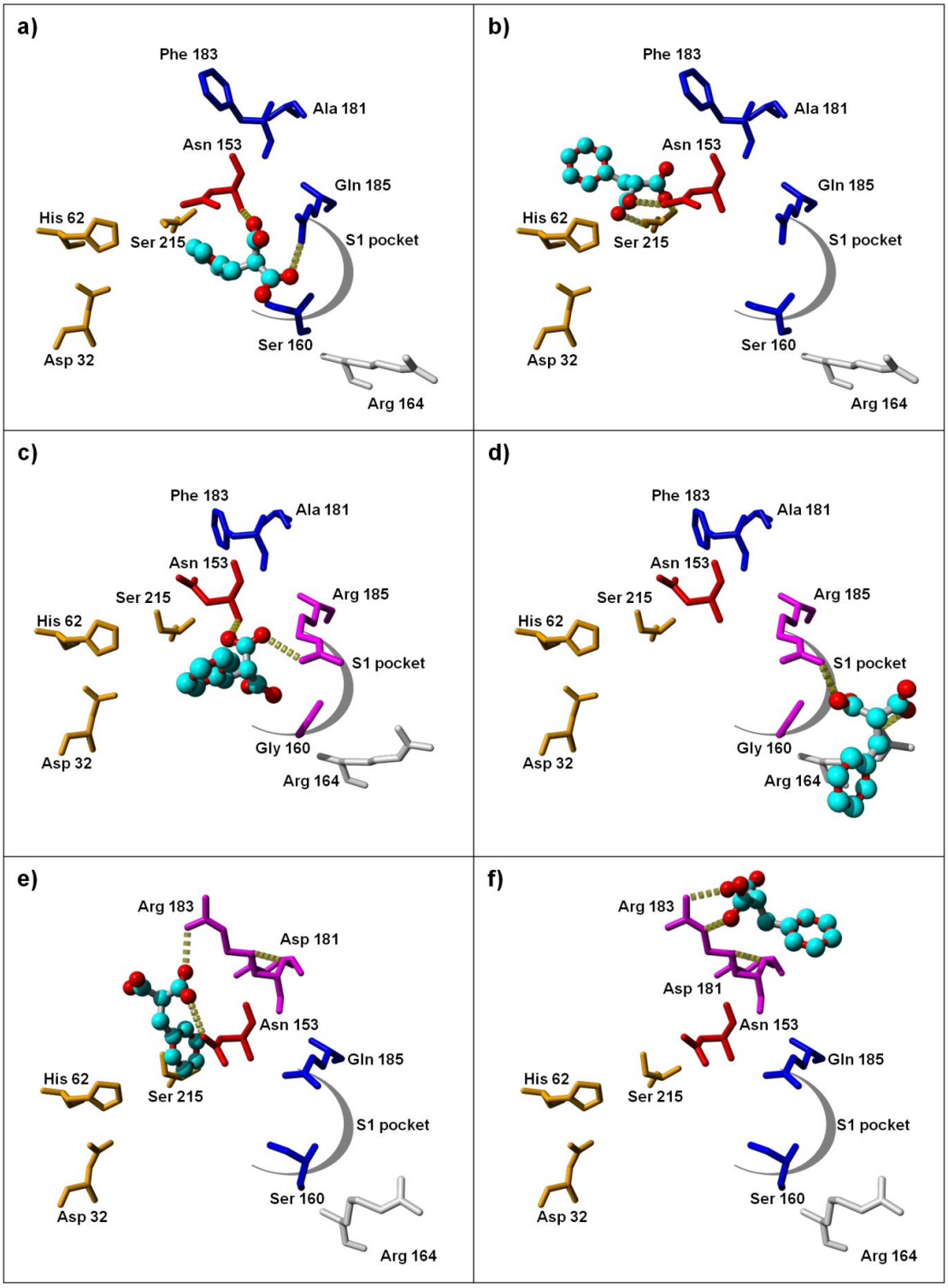


Figure 10: Prominent binding modes of benzylmalonic acid and **a,b)** BLAP parent **c,d)** variant V8 [S160G Q185R] and **e,f)** variant V9 [A181D F183R]. Orange: catalytic triad (Asp 32, His 63 and Ser 215). Red: oxyanion hole (Asn 215). Blue and Grey: residues of BLAP parent. Magenta: introduced amino acids.

3.5 Discussion

The study described in this chapter aimed to engineer a protease towards increased inhibition by benzylmalonic acid. Benzylmalonic acid is an alternative boron-free inhibitor with no indications for safety and toxicity risks in liquid detergent formulations. Industrial standards are currently boric acid or boric acid derivates (Smoum et al. 2012). Currently, the major drawback of benzylmalonic acid is the ~9-times lower inhibition efficiency compared to boric acid (Table 10). Crystal structures of boric acid adducts complexed with a subtilisin were solved and used as transition state analogs for serine proteases since the boron atom is tetrahedral and the oxygen atom was positioned in the oxyanion hole (Matthews et al. 1975; Menard and Storer 1992).

Conversion of boric acid-protease interactions to benzylmalonic acid-protease interactions seemed not plausible since no tetrahedral atom is present in benzylmalonic acid and different binding poses are expected to take place. The missing knowledge of the specific interaction conformations of benzylmalonic acid with the subtilisin protease made directed evolution to the method of choice to alter the inhibition efficiency.

Directed evolution allows tailoring all enzyme properties that can be reflected in a screening system (Zhu et al. 2010; Zhu et al. 2006) and lead us to the identification of amino acid residues interacting with benzylmalonic acid.

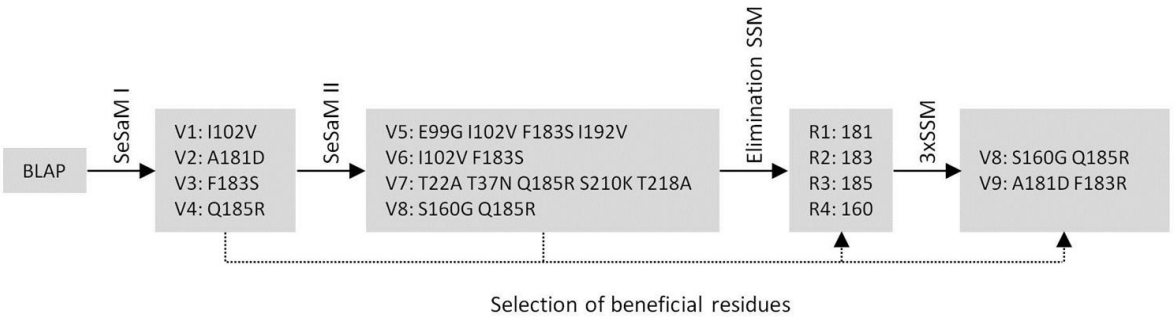


Figure 11: Genealogic tree of directed BLAP evolution towards increased inhibition by benzylmalonic acid.

Interestingly, three (A181, F183 and Q185) out of four (S160, A181, F183 and Q185) key residues for inhibition were identified in the first round of SeSaM. It is very unusual for directed evolution experiments, to identify three key residues within five

amino acids on the sequence, especially as only 2000 variants were sampled in the first SeSaM round (Shivange et al. 2009). Recently, a similar finding was reported for subtilisin E where four beneficial amino acid residues for GdmCl stability surrounding the active site were identified, where SeSaM was used as well as the diversity generation method (Li et al. 2012).

The fourth key residue (S160) was identified in the second round of SeSaM, yielding the characterized variant V8 [S160G Q185R]. The subsequent simultaneous saturation of three amino acids (A181, F183 and Q185) resulted in variant 9 with the beneficial amino acid substitutions A181D and F183R (Figure 11). After purification, the kinetic data showed significantly decreased k_{cat} values for the variants V8 and V9 in buffer and detergent matrix compared to BLAP parent. Further the k_{cat} values measured in the detergent matrix are reduced compared to the k_{cat} values determined in buffer. A similar behavior of a subtilisin protease from *Bacillus lentus* in heavy-duty liquid detergent formulations was reported (Stoner et al. 2004). However the K_m values for variants V8 and V9 are in buffer (100 mM Tris/HCl pH 8.6) approximately two times increased compared to BLAP parent, but in the matrix the K_m of the variants decreases to the same level of BLAP parent. The altered specific activity for suc-AAPF-*p*NA variant V8 could result from an altered substrate specificity caused by an altered size and charge of the S1 binding pocket resulting from the introduced amino acid substitutions S160G and Q185R at the end of the S1 binding pocket. The S1 binding pocket is known to play a crucial role for substrate specificity in subtilisin proteases (Despotovic et al. 2012; Knight et al. 2007; Pogson et al. 2009; Varadarajan et al. 2008). In variant V9, amino acid substitution F183R resulted in a more positive charged environment and could directly interact with the oxyanion hole (N153) since the distance between Cα N153 and Cα R183 is 4.94 Å. This altered oxyanion hole environment is very likely to be the reason for the changed suc-AAPF-*p*NA specificity of variant V9 compared to BLAP parent. Amino acids 181, 183 and 185 are to our best knowledge the first time identified in a directed evolution experiment. However, position 160 is well known and described to be a key residue in altering substrate specificity for substilisins (Despotovic et al. 2012; Li et al. 2012).

For detergent applications the k_{cat} values for the artificial substrate suc-AAPF-*p*NA are less important than the kept BLAP parent washing performance tested with different egg stains at Henkel AG & Co. KGaA (Data not shown; Henkel AG & Co. KGaA internal information). The results of the washing test indicate that the catalytic

efficiency determined for the artificial substrate suc-AAPF-*p*NA of the generated variants is not the rate limiting step for washing applications. The major characteristic of a detergent protease, like BLAP, is the low substrate specificity to remove efficiently protein stains (Maurer 2004), which is represented in the Henkel washing tests.

During the protein engineering campaign we introduced, beside others, two arginine residues into variants V8 and V9 leading to the assumption, that the negatively charged acid group of benzylmalonic interacts with the positive guanidinium group of the arginine. During the docking analysis we observed novel binding poses of benzylmalonic acid with the introduced arginines for both characterized variants (Figure 10c,d,e,f).

In variant V8, the key position for inhibitor interaction was 185 since the amino acid substitution S160G alone did not result in altered inhibition properties (Table 7). The combination of S160G and Q185R further increased inhibition with benzylmalonic acid. The introduced arginine at position 185 is more bulky than the replaced glutamine. For position 160, the introduced glycine is less bulky than the replaced serine, leading to a combinatorial effect in benzylmalonic acid inhibition for amino acid substitutions S160G and Q185R. The cooperative effect in terms of inhibition could result from the generated space by amino acid substitution S160G, which could allow a preferred orientation of the benzylmalonic acid inhibitor next to Arg 185 for a stronger interaction (Figure 10c). This novel observed binding poses of benzylmalonic acid with variant V8 could be responsible for the ~9-times decreased IC_{50} determined for variant V8 compared to BLAP parent.

The docking poses observed for variant V9 showed again preferred interactions of benzylmalonic acid with the introduced arginine at position 183. Benzylmalonic acid binds to R183 and the oxyanion hole residue N153, which is most likely preventing hydrolysis. The second inhibitor-ligand complex is not directly interacting with catalytically important residues, since benzylmalonic acid interacts in the second binding pose exclusively with the introduced arginine at position 183. However, this could affect inhibition since it may result in an increased local inhibitor concentration close to the catalytic triad and the oxyanion hole. The analysis of single amino acid substitution A181D and F183R reveal that both mutations on their own show no effect on the inhibition compared to BLAP parent (Table 7) exclusively the

combination reduces the IC_{50} of variant V9 ~3 times compared to BLAP parent. This result indicates a crucial role of both mutations (A181D and F183R) in the inhibition process. Docking analysis suggested a direct interaction of residue R183 with benzylmalonic acid and a novel hydrogen bond interaction of the side chain of the introduced aspartate at position 181 with the backbone amine group of R183, which may stabilize the orientation of R183 for favored inhibitor interactions (Figure 10e,f). Variant V9 showed in addition to the increased inhibition by benzylmalonic acid, an ~3-times increased inhibition for boric acid compared to BLAP parent. This could result from the more positive charge distribution of the surrounding of the oxyanion hole (N153) since the distance is 4.94 Å and the negatively charged boric acid is, as previously reported, directly interacting with the asparagines residue of the oxyanion hole (Matthews et al. 1975).

In summary, we report for the first time the successful engineering of a subtilisin protease towards increased inhibition. Two iterative rounds of directed evolution and subsequent simultaneous site-saturation mutagenesis revealed that the identified and studied amino acid substitutions in variant V8 [S160G Q185R] as well as V9 [A181D F185R] are responsible for the improved inhibition by benzylmalonic acid. The importance of the introduced arginines was demonstrated in the docking analysis since in all novel docking poses of the generated variants with benzylmalonic acid (compared to BLAP parent with benzylmalonic) an introduced arginine was directly interacting with benzylmalonic acid. The beneficial amino acids for increased inhibition are all close to the active-site and the oxyanion hole. This finding is similar to the observations made in substrate specificity engineering campaigns where the mutated residues are located close to the active site.

4 INCREASING PROTEASE PRODUCTION IN *BACILLUS SUBTILIS* BY DIRECTED VECTOR BACKBONE EVOLUTION

4.1 Declaration

Parts of this chapter were published by the author in the journal AMB Express.

"Jakob F, Lehmann C, Martinez R, Schwaneberg U (2013) Increasing protein production by directed vector backbone evolution. AMB Express 3(1):39"

4.2 Objective

This project was inspired by the previously described project (chapter 3.2.1), in which a high expression level of the secreted protease was required to determine proteolytic activity in the detergent matrix. This was accomplished by the use of the industrial expression vector pHKL (provided by Henkel AG & Co. KGaA). In this project the shuttle-vector pHY300PLK was engineered by directed evolution towards increased protease production in *B. subtilis* DB104, without changing the sequence of the expressed enzyme, its promoter-, terminator-, pre- or pro-sequence.

4.3 State of the art

4.3.1 Advances in heterologous protein production

Development of cloning and expression technologies to produce heterologous proteins in prokaryotic and eukaryotic organisms was a key technology for the rapid development of biotechnology in the last three decades, enabling the production of numerous enzymes for diagnostic and industrial applications (food, leather, textile or detergent industry; Rai and Padh 2001; Kirk et al. 2002). Recombinant production of proteins in microbial hosts can be increased by extensive optimization of multiple expression factors, such as various temperatures, transcription rates, and media composition. In addition, there are several strategies to increase heterologous protein expression: protein fusion (Cabrita et al. 2006; Shih et al. 2002; Sorensen and Mortensen 2005), codon-optimization of the target gene sequence (Gustafsson et al. 2004; Li et al. 2003) and metabolic engineering (Heyland et al. 2010; Kabisch et al. 2012). Protein production levels can be increased by modifying the corresponding

promoter, signal and/or secretion sequence (Alper et al. 2005; Xue et al. 1997). Caspers et al. and Degering et al. demonstrated the importance of the signal peptide for protein secretion in *Bacillus subtilis* and *Bacillus licheniformis*. Saturation mutagenesis of the N-domain of the α-amylase (AmyE) signal peptide yielded in a three-fold increased cutinase activity and protein amounts in the supernatant (Caspers et al. 2010). The substitution of a protease signal peptide by a library of approximately 400 heterologous signal peptides from *B. subtilis* and *B. licheniformis*, yielded in the identification of a signal peptide variant leading in a 9-fold increased protease secretion (Degering et al. 2010).

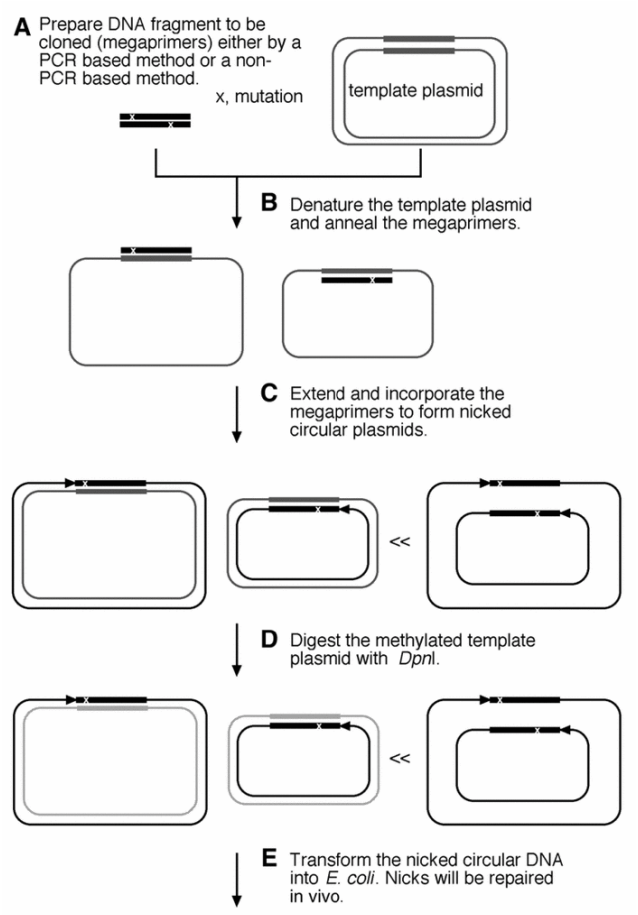


Figure 12: Outline of the MEGAWHOP principle: **A)** mutated DNA fragments (megaprimers) are prepared by random mutagenesis. **B)** megaprimers are annealed to the template plasmid (propagated in *E. coli dam*$^+$ strain). **C)** whole plasmid PCR is carried out. **D)** The methylated template is *DpnI* digested. **E)** The nicked circular DNA is transformed into *E. coli* cells and nicks are repaired *in vivo* (Miyazaki 2011).

Moreover, expression variants (increased activity due to higher amounts of enzyme, but no mutation in the enzyme gene sequence) were identified in the initial rounds of directed enzyme evolution campaigns (Tee and Schwaneberg 2007). In this project

the random mutagenesis was used to increase protease production levels by introducing mutations in the vector backbone of pHY300PLK. The developed strategy is based on the MEGAWHOP method (megaprimer PCR of whole plasmid), generally used for cloning random gene libraries into a desired expression vector harboring the wild type gene (Miyazaki 2003, 2011; Figure 12).

To increase protein production levels, the mutations were introduced into the vector backbone of pHY300PLK by its amplification under error-prone conditions (epMEGAWHOP). Subsequently, the generated mutant library was screened for increased activity (Sokol et al. 1979; Lawrence et al. 1967; Lehmann et al. 2012; Figure 13).

4.4 Material and methods

All chemicals were of analytical-reagent grade or higher quality and were purchased from Carl Roth GmbH (Karlsruhe, Germany), Sigma-Aldrich (Hamburg, Germany) and AppliChem (Darmstadt, Germany). Enzymes were purchased from New England Biolabs (Beverly, USA) and Fermentas (St. Leon-Rot, Germany). Oligonucleotides were purchased from Eurofins MWG Operon (Ebersberg, Germany) in salt-free form. Plasmid extraction and PCR purification kits were purchased from Macherey-Nagel (Düren, Germany). Microtiter plates (Greiner Bio-One GmbH, Frickenhausen, Germany) were incubated in a Multitron II Infors shaker (Infors AG, Bottmingen, Switzerland). DNA concentrations were quantified using a NanoDrop photometer (Thermo Scientific NanoDrop1000, Braunschweig, Germany). A Mastercycler gradient (Eppendorf, Hamburg, Germany) and thin-wall PCR tubes (Multi ultra-tubes; 0.2 mL; Carl Roth, Germany) were used in all PCRs.

4.4.1 *Bacillus subtilis* DB104 and pHY300PLK

B. subtilis DB104 (Kawamura and Doi 1984) and *E. coli* DH5α (purchased from Agilent Technologies; Santa Clara, USA) were used as hosts for DNA manipulation and recombinant protein expression. The shuttle vector pHY300PLK (Takara Bio Inc., Shiga, Japan) containing the *subtilisin Carlsberg* gene was used.

Chemically competent *E. coli* DH5α cells were prepared in-house using the rubidium chloride technique (Hanahan 1983). Transformation of *B. subtilis* DB104 was

performed using a natural competence based method (Vojcic et al. 2012).

4.4.2 Gene cloning, construction of pHYscarlsberg, and sequencing

After double digestion of *subtilisin Carlsberg* (Jacobs et al. 1985), including promoter and pre-pro-sequence) with *Bam*HI and *Xma*I, the fragment was subcloned into pHY300PLK using T4 DNA ligase. The resulting recombinant plasmid, named pHYscarlsberg was transformed into *E. coli* DH5α and sequenced using F_scarlsberg and R_scarlsberg primers (Table 11). DNA sequencing of the identified expression vector variants was conducted at Eurofins MWG Operon (Ebersberg, Germany) with primers listed in Table 11 and Clone Manager 9 Professional Edition (Sci-Ed software, Cary, USA) was used for all sequence alignments.

Table 11: Primers for sequencing analysis of the whole pHYscarlsberg construct.

Name	Sequence (5‘ to 3‘)
Fw_scarlsberg	CAGATTTCGTGATGCTTGTCAGG
Rev_scarlsberg	CGTTAAGGGATCAACTTTGGGAG
pHY_1	TCCTCATGTAGCGGGAG
pHY_2	GTATCCCACCAATGTAGCCG
pHY_3	CCATTACGAGTGCTGGAAATGC
pHY_4	CGCCTTCTTCTGTGTCATCAAG
pHY_5	GCCATGTTTCATTGCTCTCCTC
pHY_6	GGCTTTTAAGCGTCTGTACGTTC
pHY_7	CGGAGCTGAATGAAGCCATAC
pHY_8	CGAAGGTTCTCTGAGCTACCAACTC

4.4.3 Generation of error-prone MEGAWHOP libraries

Megaprimers for the *scarlsberg* gene were generated by PCR under standard conditions. The amplification of the *subtilisin Carlsberg* gene including promoter-, terminator-, pre- and pro-sequence was performed using unmodified DNA primers 5’-CAGATTTCGTGATGCTTGTCAGG-3’ and 5’-CGTTAAGGGATCAACTTTGGG AG-3’ (5 µM each). For the PCR (98°C for 30 sec , one cycle; 98°C, 15 sec/63°C, 30 sec/68°C, 25 cycles; 68°C for 10 min, one cycle), PfuS DNA polymerase (2.5 U), dNTP mix (10 mM), template pHYscarlsberg: 30 ng/µL was used. The PCR product (megaprimer) was purified using a PCR purification kit.

For the epMEGAWHOP PCR (72°C for 5 min, one cycle; 94°C for 1:30 min, one cycle; 94°C, 45 sec/60°C, 45 sec/72°C, 5 min, 25 cycles; 72°C for 10 min, one cycle),

Taq DNA polymerase (2.5 U), dNTP mix (10 mM) together with plasmid template (pHYscarlsberg: 120 ng), megaprimer (scarlsberg: 550 ng) and 0.05 mM Mn^{2+} were used. A subsequent DpnI digestion (20 U) of the template was performed overnight at 37°C. The epMEGAWHOP PCR product was transformed into *E. coli* DH5α and the obtained colonies from the agar plates were pooled and used for plasmid isolation. Next step was the transformation into the expression host *B. subtilis* DB104 using a natural competence based method (Vojcic et al. 2012).

4.4.4 Indicator plates for prescreening

For pre-screening, LB agar plates supplemented with 1 % (w/v) skim milk and 15 µg/mL tetracycline were used for proteolytic activity detection (Sokol et al. 1979).

4.4.5 Production of subtilisin Carlsberg in microtiter plates

Single colonies of *B. subtilis* DB104 from LB skim milk agar plate, that show proteolytic activity, were transferred into flat-bottom microtiter plates (preculture plates) containing 200 µl of buffered LB medium (pH 7.6; 1 % (w/v) tryptone, 0.5 % (w/v) yeast extract and 1 % (w/v) sodium chloride, 17 mM potassium dihydrogen phosphate, 72 mM dipotassium hydrogen phosphate, 15 µg/mL tetracycline) and incubated (37°C, 900 rpm, 18 h, 70 % humidity). The volume of 10 µL preculture was used to inoculate the main culture (150 µL buffered LB, 37°C, 900 rpm, 24 h, 70 % humidity). After expression, the microtiter plates were centrifuged (3,220 *g*, 20 min, 4°C) to separate the cells from the secreted protease in the supernatant. The supernatant was transferred to a new microtiter plate for detailed analysis.

4.4.6 Colorimetric assay for proteolytic activity of subtilisin Carlsberg

Proteolytic activity was determined using the synthetic peptide substrate succinyl-L-Ala-L-Ala-L-Pro-L-Phe-*p*-nitroanilide (suc-AAPF-*p*NA) by quantification of the release of free *p*-nitroaniline (*p*NA) at 410 nm (DelMar et al. 1979). For the screening in microtiter plate the proteolytic reaction was initiated by supplementing 10 µl of the 1:20 diluted supernatant with 1.1 mM suc-AAPF*p*NA substrate into 100 µL Tris/HCl buffer (100 mM, pH 8.6). The release of *p*-nitroanilide was continuously monitored at 410 nm in a microtiter plate reader (Tecan Sunrise) and was used for protease quantification.

4.4.7 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

The protein content from subtilisin Carlsberg supernatants was precipitated with 30 % trichloroacetic acid (TCA; 15 min on ice) and resuspended after washing with acetone in Tris/HCl (100 mM; pH 8.0). Equal volumes of the concentrated and precipitated supernatants were separated by SDS-PAGE (5 % stacking gel and a 15 % separating gel; Laemmli 1970).

4.5 Results

The results section is divided in three parts: 4.5.1 Library generation and screening, 4.5.2 Activity and expression analysis and 4.5.3 Sequencing analysis.

4.5.1 Library generation and screening

Figure 13 shows the five steps of the developed epMEGAWHOP protocol. In step I the generated DNA fragments (megaprimers) are prepared by PCR and subsequently annealed to the template plasmid. Step II comprises the amplification of the whole plasmid under error-prone conditions and the methylated template is *DpnI* digested. In step III the nicked circular DNA is transformed into *E. coli* cells and nicks are repaired *in vivo.* The plasmids are isolated from *E. coli* and transformed into the expression host *B. subtilis* DB104. The pre-screening on skim milk indicator plates is performed in step IV and the colonies showing the biggest diameter of the generated halos are transferred into microtiter plates for quantification of the secreted protease by activity detection (step V). Vector backbone variants causing increased activity were isolated and the wild-type gene sequence of subtilisin Carlsberg was subcloned into the mutated backbone as described under Materials and Methods to confirm that the improved protein expression can be attributed to the vector backbone modification and is not caused by altered enzyme activity.

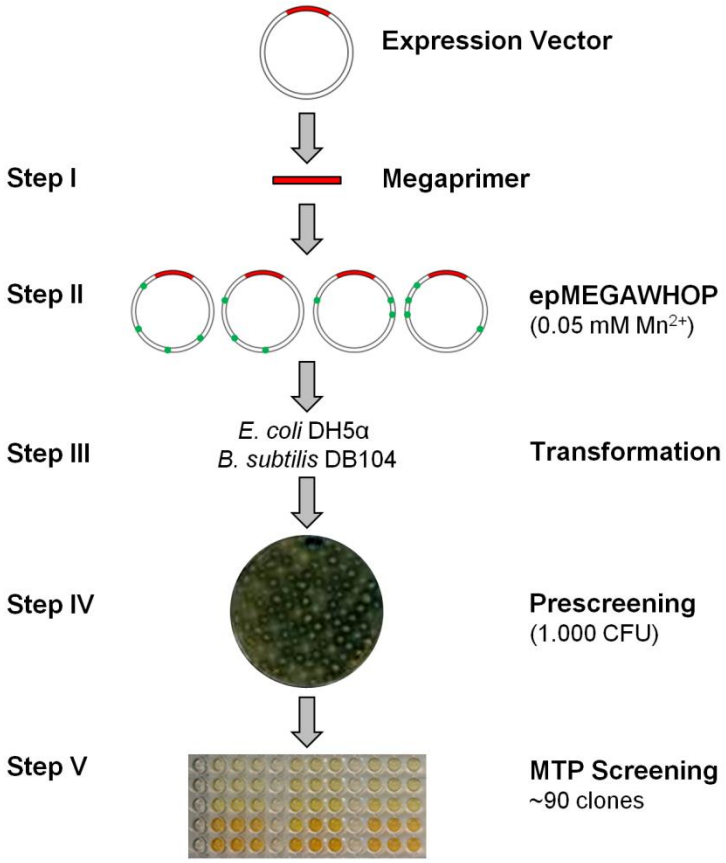


Figure 13: The five key steps for increasing protein production by directed vector backbone evolution. **Step I**: megaprimer generation. **Step II**: amplification of the vector backbone under error prone conditions (epMEGAWHOP). **Step III**: Transformation of the expression host. **Step IV**: Agar plate-based prescreening. **Step V**: Screening in microtiter plate format to quantify expression levels.

4.5.2 Activity and expression analysis

The proteolytic activity was determined in the cell free supernatants obtained from *B. subtilis* DB104/pHY-WT-scarlsberg and *B. subtilis* DB104/pHY-M1-scarlsberg containing the secreted protease subtilisin Carlsberg. The supernatant activity of pHY-WT-scarlsberg and the identified variant pHY-M1-scarlsberg is shown in Figure 14a. The obtained supernatant activity of the mutated vector backbone pHY-M1-scarlsberg shows a ~4-times increased activity compared to the supernatant of the initial construct pHY-WT-scarlsberg under the given conditions. In Figure 14b,c a SDS-Page of the TCA precipitated cell free supernatants shows a distinct band of subtilisin Carlsberg at ~27kDa, except for the empty vector control (no subtilisin Carlsberg secreted). The intensity of the corresponding subtilisin Carlsberg band obtained from pHY-M1-scarlsberg is compared to its wild-type significantly increased (Figure 14b).

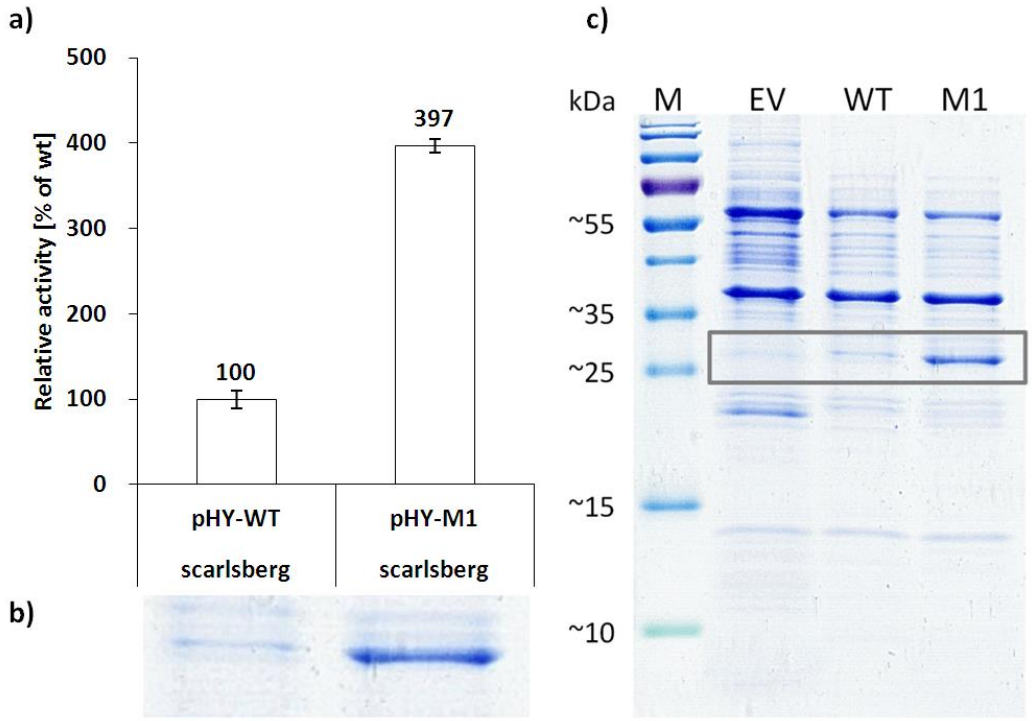
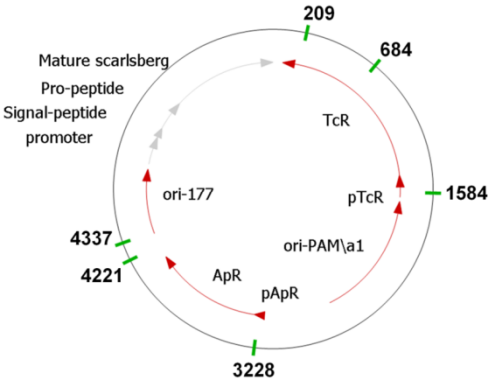


Figure 14: a) Comparisson of the proteolytic activity of pHYscarlsberg and variant pHY-M1-scarlsberg determined in the cell free supernatant; **b)** Corresponding band of subtilisin Carlsberg in the cell free supernatant visualized by SDS-PAGE; **c)** SDS-PAGE of the cell supernatant for the following constructs: EV (DB104/pHY300PLK), WT (DB104/pHYscarlsberg) and M1 (DB104/pHY-M1-scarlsberg); gray box: corresponding band of subtilisin Carlsberg protease.

4.5.3 Sequencing analysis

The generated mutant pHY-M1-scarlsberg showed no mutation in the coding gene-sequences for subtilisin Carlsberg, but the vector backbone (pHY-M1) carried six mutations. The allocation of the individual mutations found on the vector backbones is shown in Figure 15b. The identified mutation T6007C in the pHY-M1 vector is located in a non-coding region of the plasmid. All other 5 mutations could be assigned to functional regions. Two mutations are located in the tetracycline resistance (Tc^R) region, encoding a 458 amino acid 50-kDa protein (TET). The translation of the Tc^R sequence into amino acids identified the substitution of isoleucine by threonine at position 256 caused by the nucleotide exchange (A to G) at position 684.

a)



b)

Nucleotide	Substitution wt	-	M1	Region
209	A	-	G	Tc^R
684	A	-	G	Tc^R
1584	T	-	C	pTc^R
3238	T	-	C	pAp^R
4221	T	-	C	na
4337	A	-	-	ori-177

Figure 15: a) Vector map of pHYscarlsberg. The mutations identified for pHY-M1-scarlsberg are numbered and marked in green. Nucleotide 1 is regarded as the base after the stop codon (ATT) of the mature *subtilisin Carlsberg* gene. **b)** Location of the introduced mutations in the vector backbone. The positions of the mutations in the corresponding vector backbone are based on the sequenced vector backbone prior error-prone amplification. Nucleotide 1 is regarded as the base after the stop codon (ATT) of the mature *subtilisin Carlsberg* gene. (na: no corresponding function of the gene sequence could be assigned; pTc^R/pAp^R: promoter region Tetracyclin/Ampicilin resistance gen, identified with the promoter prediction tool (Reese 2001).

4.6 Discussion

In white biotechnology, recombinant protein expression is an ongoing challenge. Many methods and strategies, such as codon-optimization of the target gene sequence (Gustafsson et al. 2004; Li et al. 2003), metabolic engineering (Heyland et al. 2010; Kabisch et al. 2012) and/or modifying the corresponding promoter, signal and/or secretion sequence (Alper et al. 2005; Xue et al. 1997), are reported to increase the level of recombinant produced proteins.

The MEGAWHOP method is commonly used to insert a mutated target-gene sequence into an expression vector. The here developed epMEGAWHOP method inverts this strategy and allows the insertion of a not mutated gene into a randomly mutated expression vector. This developed strategy is a rapid and straightforward way to increase protein expression of any gene for which an activity screening system is available.

Koyanagi et al. reported a method to construct randomly mutagenized plasmid libraries free from ligation by using a modified QuikChange site-directed mutagenesis protocol. Unspecific mutations were introduced in the gene and the vector-backbone during the amplification of the whole vector under error- prone conditions. However, the influence of the vector backbone mutations was not analyzed (Koyanagi et al.

2008). A similar principle was followed by Matsumura et al. (2005) here the whole vector-gene construct was amplified under error-prone conditions with the simultaneous introduction of a restriction site, which was afterwards utilized to perform an intra-molecular self-ligation step. The possibility that the introduced vector backbone mutations could affect the performance of a plasmid was stated, but not further discussed since the focus was the improvement of the target enzyme beta-galactosidase (Matsumura and Rowe 2005). All these protocols had the common target to alter the sequence of the expressed target-gene. In the epMEGAWHOP approach the aim was to randomly mutagenize the vector backbone of pHY300PLK, not the target-gene sequence, to increase the production level of subtilisin Carlsberg by *Bacillus subtilis* DB104.

The production level of subtilisin Carlsberg was significantly enhanced (up to 4-fold), by the random introduction of five mutations and one deletion in the vector backbone of pHY300PLK (Figure 14). Sequence analysis confirmed that the *subtilisin Carlsberg gene*, promoter-, terminator-, pre- and pro-sequence was not modified (Figure 15). Out of the six introduced mutations, two could be assigned to promoter sequences of tetracycline and ampicillin resistance regions. It is reported that mutations in promoter regions can cause an altered expression level of the specific gene and/or influence the regulatory properties of a given promoter (Alper et al. 2005; Nevoigt et al. 2007). Furthermore the mutation A684G resulted in an amino acid exchange in the TET protein (Ile256Thr) possibly resulting in an altered repressor behavior beneficial for protease secretion. The identified deletion in the origin of replication could increase or decrease the copy number of the plasmids in the cells leading to an increased or more stable gene expression. Single nucleotide substitutions in the vector pBR322 are reported to influence the plasmid copy significantly (Boros et al. 1984). However, the protein production level is not directly proportional to the plasmid copy numbers, but can be affect to a certain extend (Nacken et al. 1996). To determine the distinct effect of each mutation, single mutants have to be generated and analyzed in detail in case of protease production level, plasmid copy number and/or growth behavior with altered tetracycline concentrations.

The generality of the epMEGAWHOP approach was already demonstrated by applying the method to *E. coli* systems with intracellular Cellulase and periplasmatic Lipase production in a group-internal joint project. The observed increase in protease

production is not necessarily the limit of the *B. subtilis* DB104 secretion machinery, and therefore the main advantage of epMEGAWHOP is the possibility to optimize the vector backbones in iterative rounds. Furthermore, the random mutagenesis of vector backbones can be combined with any already existing strategy to further enhance protein production in prokaryotes (e.g. signal peptide optimization, fusion protein or codon optimization of the target enzyme; Gustafsson et al. 2004; Caspers et al. 2010; Cabrita et al. 2006). Especially, the addition of the signal- and promoter- sequence for random mutagenesis seems to be a promising and straight forward approach, since these regions were already reported for *Bacillus* sp. having significant effects on the production levels of different target enzyme (Caspers et al. 2010; Degering et al. 2010).

5 SUMMARY AND CONCLUSION

The goal of this work was to shift the pH activity profile, the thermal resistance of the *Bacillus gibsonii* protease (BgAP) and to increase the inhibition efficiency of the *Bacillus lentus* alkaline protease (BLAP) towards benzylmalonic acid in order to make subtilisin proteases BgAP and BLAP applicable for laundry applications. Furthermore, a novel method was developed to increase the protease production by vector backbone mutagenesis in a directed evolution experiment employing pHY300PLK and *B. subtilis* DB104.

In the first part we developed a novel rational design strategy to reengineer the surface charge of BgAP in order to adapt the pH dependent activity and thermal resistance. Out of 113 possible surface amino acids 18 (11 Asn and 7 Gln) residues of BgAP were selected and evaluated based on three proposed criteria: (1) The Asn or Gln residues should not be conserved, (2) should be surface exposed, and (3) neighboured by glycine. Amino acid substitutions of the “deamidation type” (Asn to Asp and Gln to Glu) in five out of eight amino acids meeting all criteria resulted in up to ~1.5-times increased proteolytic activity. In addition, pH activity profiles of the purified variants N253D, Q256E and the combined variant N253DQ256E were dramatically shifted towards higher activity at lower pH (range of 8.5-10). Variant N253DQ256E showed twice the specific activity of wild type BgAP and its thermal resistance was increased by 2.4°C at pH 8.5.

Surface charge engineering by amino acid substitutions following the “deamidation type” is a novel and straightforward sequence alignment based protein engineering strategy for modulating pH dependent activity and thermal resistance of subtilisins and possibly other enzyme families. It allows the rapid adaptation of pH dependent activity for industrial or medical protease applications, such as the adaptation of the proteolytic activity to the required pH of detergent formulations (pH powder formulations > pH liquid detergent formulations). Compared to the current time intensive directed evolution experiments to adapt pH activity or thermal resistance, the deamidation strategy is a promising and time-efficient alternative.

In the second project the interactions of the reversible inhibitor benzylmalonic acid with the *Bacillus lentus* alkaline proteases (BLAP) were investigated; so far protein structure-function relationships of benzylmalonic acid and subtilisin proteases have

not been explored. Two rounds of directed BLAP evolution towards increased inhibition by benzylmalonic acid resulted in variant V8 [S160G Q185R] having a ~9-times increased inhibition compared to BLAP parent. A subsequently performed simultaneous site-saturation mutagenesis at the three identified key residues (A181, F183 and Q185) resulted in variant V9 [A181D F183R] showing a ~3-times increased inhibition in presence of benzylmalonic acid. Docking analysis of benzylmalonic acid with the variants V8, V9 and their parent BLAP, lead to the identification of novel inhibitor interactions with the introduced amino acid substitutions. The results provide a molecular explanation for the observed increased inhibition of the identified variants V8 and V9. The lowered inhibitior concentrations (~9-times) for the generated variant V8, enables the replacement of the environmentally problematic inhibitor boric acid and derivates by benzylmalonic acid in liquid detergent formulations.

For the first time a subtilisin protease was engineered towards increased inhibition. The boron-free inhibitor benzylmalonic acid combined with the engineered variant V8 is, to our knowledge, the first known inhibition system capable to replace the industrial standard boric acid and derivates from liquid detergent formulations. Computational docking analyses provided novel molecular explanations for the increased inhibition of BLAP towards benzylmalonic acid. The generated knowledge enables the direct transfer of the identified amino acid substitutions to related subtilins in order to increase the inhibition towards benzylmalonic acid. This project demonstrates the immense potential of directed evolution combined with computational analysis to explore and understand novel and/or unknown structure-function relationships.

With the random mutagenesis of vector backbones a further promising application of directed evolution was developed (epMEGAWHOP). The epMEGAWHOP method with subsequent screening for variants with increased protease production offers a rapid and straight forward strategy to increase recombinant protein production levels. In particular a ~4-times increased production of subtilisin Carlsberg was achieved, resulting from a randomly mutated vector backbone library of pHY300PLK. The generality of this approach was already shown by applying the epMEGAWHOP method to *E. coli* systems with intracellular and periplasmatic protein production. A main advantage is the general application by optimizing expression vectors in

iterative rounds of epMEGAWHOP, even when the expression and/or secretion pathways of the host organisms are not explored.

In essence, we developed a novel rational design approach to alter pH dependent activity and thermal resistance of the subtilisin BgAP; we reengineered the subtilisin BLAP for increased inhibition with benzylmalonic acid matching the performance for use in industrial liquid detergent formulations and developed and validated the novel epMEGAWHOP method for increased protease production in *Bacillus subtilis* DB104 by random mutagenesis of the vector backbone. During the thesis two invention reports were filed in a patent application together with Henkel AG & Co. KGaA.

6 REFERENCES

Aharoni A, Griffiths AD, Tawfik DS (2005) High-throughput screens and selections of enzyme-encoding genes. Curr Opin Chem Biol 9:210-216

Alper H, Fischer C, Nevoigt E, Stephanopoulos G (2005) Tuning genetic control through promoter engineering. Proc Natl Acad Sci U S A 102:12678-12683

Arnold FH (2001) Combinatorial and computational challenges for biocatalyst design. Nature 409:253-257

Blanchard JE, Elowe NH, Huitema C, Fortin PD, Cechetto JD, Eltis LD, Brown ED (2004) High-throughput screening identifies inhibitors of the SARS coronavirus main proteinase. Chem Biol 11:1445-1453

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

Bodanszky M, Kwei JZ (1978) Side reactions in peptide synthesis. VII. Sequence dependence in the formation of aminosuccinyl derivatives from beta-benzyl-aspartyl peptides. Int J Pept Protein Res 12:69-74

Boros I, Pósfai G, Venetianer P (1984) High-copy-number derivatives of the plasmid cloning vector pBR322. Gene 30:257-260

Branden C, Tooze J (1999) Introduction to Protein Structure. Second Edition. Garland Publishing, Inc.

Bryan PN (2000) Protein engineering of subtilisin. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology 1543:203-222

Cabrita LD, Dai W, Bottomley SP (2006) A family of *E. coli* expression vectors for laboratory scale and high throughput soluble protein production. BMC Biotechnol 6:12

Capasso S, Mazzarella L, Sica F, Zagari A (1989) Deamidation via cyclic imide in asparaginyl peptides. Pept Res 2:195-200

Caspers M, Brockmeier U, Degering C, Eggert T, Freudl R (2010) Improvement of Sec-dependent secretion of a heterologous model protein in *Bacillus subtilis* by saturation mutagenesis of the N-domain of the AmyE signal peptide. Appl Microbiol Biotechnol 86:1877-1885

Catak S, Monard G, Aviyente V, Ruiz-Lopez MF (2009) Deamidation of asparagine residues: direct hydrolysis versus succinimide-mediated deamidation mechanisms. J Phys Chem A 113:1111-1120

Chang S, Cohen SN (1979) High frequency transformation of *Bacillus subtilis* protoplasts by plasmid DNA. Mol Gen Genet 168:111-115

Chavira R, Jr, Burnett TJ, Hageman JH (1984) Assaying proteinases with azocoll. Anal Biochem 136:446-450

Chica RA, Doucet N, Pelletier JN (2005) Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. Curr Opin Biotechnol 16:378-384

Cirino PC, Mayer KM, Umeno D (2003) Generating mutant libraries using error-prone PCR. Methods Mol Biol 231:3-9

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

Craik CS, Page MJ, Madison EL (2011) Proteases as therapeutics. Biochemical Journal 435:1-16

Daugherty AB, Muthu P, Lutz S (2012) Novel Protease Inhibitors via Computational Redesign of Subtilisin BPN' Propeptide. Biochemistry 51(41):8247-8255

De Kreij A, Van den Burg B, Venema G, Vriend G, Eijsink VGH, Nielsen JE (2002) The effects of modifying the surface charge on the catalytic activity of a thermolysin-like protease. J Biol Chem 277:15432-15438

Degering C, Eggert T, Puls M, Bongaerts J, Evers S, Maurer K-H, Jaeger KE (2010) Optimization of protease secretion in *Bacillus subtilis* and *Bacillus licheniformis* by screening of homologous and heterologous signal peptides. Appl Environ Microbiol 76:6370-6376

DelMar EG, Largman C, Brodrick JW, Geokas MC (1979) A sensitive new substrate for chymotrypsin. Anal Biochem 99:316-320

DeSantis G, Jones JB (1998) Chemical modifications at a single site can induce significant shifts in the pH profiles of a serine protease. J Am Chem Soc 120:8582-8586

Despotovic D, Vojcic L, Prodanovic R, Martinez R, Maurer K-H, Schwaneberg U (2012) Fluorescent assay for directed evolution of perhydrolases. J Biomol Screen 17:796-805

Di Cera E (2008) Engineering protease specificity made simple, but not simpler. Nat Chem Biol 4:270-271

Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 24:1999-2012

Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. J Chem Phys 103:8577-8593

Farinas ET (2006) Fluorescence activated cell sorting for enzymatic activity. Comb Chem High Throughput Screen 9:321-328

Farinas ET, Bulter T, Arnold FH (2001) Directed enzyme evolution. Curr Opin Biotechnol 12:545-551

Feller BE, Kellis JT, Jr., Cascao-Pereira LG, Robertson CR, Frank CW (2010) The role of electrostatic interactions in protease surface diffusion and the consequence for interfacial biocatalysis. Langmuir 26:18916-18925

Fukuchi S, Nishikawa K (2001) Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria. J Mol Biol 309:835-843

Goddette DW, Paech C, Yang SS, Mielenz JR, Bystroff C, Wilke ME, Fletterick RJ (1992) The crystal structure of the *Bacillus lentus* alkaline protease, subtilisin BL, at 1.4 A resolution. Journal of molecular biology 228:580-595

Grimsley GR, Shaw KL, Fee LR, Alston RW, Huyghues-Despointes BM, Thurlkill RL, Scholtz JM, Pace CN (1999) Increasing protein stability by altering long-range coulombic interactions. Protein Sci 8:1843-1849

Gupta R, Beg QK, Lorenz P (2002) Bacterial alkaline proteases: molecular approaches and industrial applications. Appl Microbiol Biotechnol 59:15-32

Gustafsson C, Govindarajan S, Minshull J (2004) Codon bias and heterologous protein expression. Trends Biotechnol 22:346-353

Güven G, Prodanovic R, Schwaneberg U (2010) Protein Engineering - An Option for Enzymatic Biofuel Cell Design. Electroanalysis 22:765-775

Hanahan D (1983) Studies on transformation of *Escherichia coli* with plasmids. J Mol Biol 166:557-580

Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A 89:10915-10919

Heyland J, Blank LM, Schmid A (2010) Quantification of metabolic limitations during recombinant protein production in *Escherichia coli*. J Biotechnol 155:178-184

Hubbard SA (1998) Comparative toxicology of borates. Biol Trace Elem Res 66:343-357

Imanaka T, Shibazaki M, Takagi M (1986) A new way of enhancing the thermostability of proteases. Nature 324:695-697

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

Jacobs M, Eliasson M, Uhlen M, Flock JI (1985) Cloning, sequencing and expression of subtilisin Carlsberg from *Bacillus licheniformis*. Nucleic Acids Res 13:8913-8926

Jakalian A, Jack DB, Bayly CI (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. J Comput Chem 23:1623-1641

Jaouadi B, Aghajari N, Haser R, Bejar S (2010) Enhancement of the thermostability and the catalytic efficiency of *Bacillus pumilus* CBS protease by site-directed mutagenesis. Biochimie 92:360-369

Kabisch J, Thurmer A, Hubel T, Popper L, Daniel R, Schweder T (2012) Characterization and optimization of *Bacillus subtilis* ATCC 6051 as an expression host. J Biotechnol 163(2):97-104

Kato A, Tanaka A, Lee Y, Matsudomi N, Kobayashi K (1987a) Effects of deamidation with chymotrypsin at pH 10 on the functional properties of proteins. J Agric Food Chem 35:285-288

Kato A, Tanaka A, Matsudomi N, Kobayashi K (1987b) Deamidation of food proteins by protease in alkaline pH. J Agric Food Chem 35:224-227

Kato A, Tanimoto S, Muraki Y, Kobayashi K, Kumagai I (1992) Structural and functional properties of hen egg-white lysozyme deamidated by protein engineering. Biosci Biotech Biochem 56:1424-1428

Kawamura F, Doi RH (1984) Construction of a *Bacillus subtilis* double mutant deficient in extracellular alkaline and neutral proteases. J Bacteriol 160:442-444

Kirk O, Borchert TV, Fuglsang CC (2002) Industrial enzyme applications. Curr Opin Biotechnol 13:345-351

Knight ZA, Garrison JL, Chan K, King DS, Shokat KM (2007) A Remodelled Protease That Cleaves Phosphotyrosine Substrates. Journal of the American Chemical Society 129:11672-11673

Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM (2006) MUSTANG: a multiple structural alignment algorithm. Proteins 64:559-574

Kossiakoff AA (1988) Tertiary structure is a principal determinant to protein deamidation. Science 240:191-194

Koyanagi T, Yoshida E, Minami H, Katayama T, Kumagai H (2008) A rapid, simple, and effective method of constructing a randomly mutagenized plasmid library free from ligation. Biosci Biotechnol Biochem 72:1134-1137

Krieger E, Darden T, Nabuurs SB, Finkelstein A, Vriend G (2004) Making optimal use of empirical energy functions: force-field parameterization in crystal space. Proteins 57:678-683

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

Krieger E, Nielsen JE, Spronk CA, Vriend G (2006) Fast empirical pKa prediction by Ewald summation. J Mol Graph Model 25:481-486

Kuhn P, Knapp M, Soltis SM, Ganshaw G, Thoene M, Bott R (1998) The 0.78 A structure of a serine protease: *Bacillus lentus* subtilisin. Biochemistry 37:13446-13452

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680-685

Lawrence RC, Fryer TF, Reiter B (1967) Rapid method for the quantitative estimation of microbial lipases. Nature 213:1264-1265

Lehmann C, Sibilla F, Maugeri Z, Streit WR, Domínguez de María P, Martinez R, Schwaneberg U (2012) Reengineering CelA2 cellulase for hydrolysis in aqueous solutions of deep eutectic solvents and concentrated seawater. Green Chem 14:2719-2726

Leisola M, Turunen O (2007) Protein engineering: opportunities and challenges. Appl Microbiol Biotechnol 75:1225-1232

Li A, Kato Z, Ohnishi H, Hashimoto K, Matsukuma E, Omoya K, Yamamoto Y, Kondo N (2003) Optimized gene synthesis and high expression of human interleukin-18. Protein Expr Purif 32:110-118

Li Z, Roccatano D, Lorenz M, Schwaneberg U (2012) Directed evolution of subtilisin E into a highly active and guanidinium chloride- and sodium dodecylsulfate-tolerant protease. Chembiochem 13:691-699

Loladze VV, Ibarra-Molero B, Sanchez-Ruiz JM, Makhatadze GI (1999) Engineering a thermostable protein via optimization of charge-charge interactions on the protein surface. Biochemistry 38:16419-16423

Lund H, Kaasgaard S, Skagerlind P, Jorgensen L, Jorgensen C, van de Weert M (2011) Correlation Between Enzyme Activity and Stability of a Protease, an Alpha-Amylase and a Lipase in a Simplified Liquid Laundry Detergent System, Determined by Differential Scanning Calorimetry. Journal of Surfactants and Detergents 15:9-21

Martinez R, Jakob F, Tu R, Siegert P, Maurer K-H, Schwaneberg U (2013) Increasing activity and thermal resistance of *Bacillus gibsonii* alkaline protease (BgAP) by directed evolution. Biotechnol Bioeng 110(3):711-720

Martinez R, Schwaneberg U, Roccatano D (2011) Temperature effects on structure and dynamics of the psychrophilic protease subtilisin S41 and its thermostable mutants in solution. Protein Eng Des Sel 24:533-544

Mastrobattista E, Taly V, Chanudet E, Treacy P, Kelly BT, Griffiths AD (2005) High-throughput screening of enzyme libraries: in vitro evolution of a beta-galactosidase by fluorescence-activated sorting of double emulsions. Chem Biol 12:1291-1300

Matsumura I, Rowe LA (2005) Whole plasmid mutagenic PCR for directed protein evolution. Biomol Eng 22:73-79

Matthews DA, Alden RA, Birktoft JJ, Freer ST, Kraut J (1975) X-ray crystallographic study of boronic acid adducts with subtilisin BPN' (Novo). A model for the catalytic transition state. J Biol Chem 250:7120-7126

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

Menard R, Storer AC (1992) Oxyanion hole interactions in serine and cysteine proteases. Biol Chem Hoppe Seyler 373:393-400

Miyamoto S, Kollman PA (1992) Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comput Chem 13:952-962

Miyazaki K (2003) Creating random mutagenesis libraries by megaprimer PCR of whole plasmid (MEGAWHOP). Methods Mol Biol 231:23-28

Miyazaki K (2011) MEGAWHOP cloning: a method of creating random mutagenesis libraries via megaprimer PCR of whole plasmids. Methods Enzymol 498:399-406

Miyazaki K, Wintrode PL, Grayling RA, Rubingh DN, Arnold FH (2000) Directed evolution study of temperature adaptation in a psychrophilic enzyme. J Mol Biol 297:1015-1026

Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 19:1639-1662

Mundhada H, Marienhagen J, Scacioc A, Schenk A, Roccatano D, Schwaneberg U (2011) SeSaM-Tv-II generates a protein sequence space that is unobtainable by epPCR. Chembiochem 12:1595-1601

Nacken Vr, Achstetter T, Degryse E (1996) Probing the limits of expression levels by varying promoter strength and plasmid copy number in Saccharomyces cerevisiae. Gene 175:253-260

Nascimento WCA, Martins MLL (2006) Studies on the stability of protease from *Bacillus* sp. and its compatibility with commercial detergent. Braz J Microbiol 37: 307–311

Nevoigt E, Fischer C, Mucha O, Matthäus F, Stahl U, Stephanopoulos G (2007) Engineering promoter regulation. Biotechnol Bioeng 96:550-558

Nielsen LK, Deane-Wray A (1999) 4-Substituted-Phenyl-Boronic acids as enzyme stabilizers. US Patent 5,972,873

Von der Osten C, Branner S, Hastrup S, Hedegaard L, Rasmussen MD, Bisgård-Frantzen H, Carlsen S, Mikkelesen JM (1993) Protein engineering of subtilisins to improve stability in detergent formulations. J Biotechnol 28:55–68

Paech C, Christianson T, Maurer K-H (1993) Zymogram of proteases made with developed film from nondenaturing polyacrylamide gels after electrophoresis. Anal Biochem 208:249-254

Pogson M, Georgiou G, Iverson BL (2009) Engineering next generation proteases. Current Opinion in Biotechnology 20:390-397

Polgar L (2005) The catalytic triad of serine peptidases. Cell Mol Life Sci 62:2161-2172

Rai M, Padh H (2001) Expression systems for production of heterologous proteins. Current Science 80:1121-1128

Rawlings ND, Barrett AJ, Bateman A (2011) Asparagine peptide lyases: a seventh catalytic type of proteolytic enzymes. J Biol Chem 286:38321-38328

Reese MG (2001) Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. Computers & Chemistry 26:51-56

Reetz M, Jaeger K-E (1999) Superior Biocatalysts by Directed Evolution. In: Biocatalysis - From Discovery to Application, vol 200. Topics in Current Chemistry. Springer Berlin / Heidelberg, pp 31-57

Reetz MT, Kahakeaw D, Lohmer R (2008) Addressing the numbers problem in directed evolution. Chembiochem 9:1797-1804

Rheinnecker M, Baker G, Eder J, Fersht AR (1993) Engineering a novel specificity in subtilisin BPN'. Biochemistry 32:1199-1203

Robinson NE, Robinson AB (2004) Molecular clocks: deamidation of asparaginyl and glutaminyl residues in peptides and proteins. Althouse Press, Cave Junction, OR

Severson RG Jr (1985) Liquid detergents containing boric acid to stabilize enzymes. US Patent 4,537,706

Russell AJ, Fersht AR (1987) Rational modification of enzyme catalysis by engineering surface charge. Nature 328:496-500

Sakoda H, Imanaka T (1992) Cloning and sequencing of the gene coding for alcohol dehydrogenase of *Bacillus stearothermophilus* and rational shift of the optimum pH. J Bacteriol 174:1397-1402

Sanchez-Ruiz JM, Makhatadze GI (2001) To charge or not to charge? Trends Biotechnol 19:132-135

Shih YP, Kung WM, Chen JC, Yeh CH, Wang AH, Wang TF (2002) High-throughput screening of soluble recombinant proteins. Protein Sci 11:1714-1719

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

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

Siegert P, Wieland S, Engelskirchen J, Merkel M, Maurer K-H, Bessler C (2009) Novel alkaline protease from *Bacillus gibsonii* and washing and cleaning agents containing said novel alkaline protease. US Patent US 2009/0275493 A1

Siezen RJ, Leunissen JA (1997) Subtilases: the superfamily of subtilisin-like serine proteases. Protein Sci 6:501-523

Smoum R, Rubinstein A, Dembitsky VM, Srebnik M (2012) Boron containing compounds as protease inhibitors. Chem Rev 112:4156-4220

Sokol PA, Ohman DE, Iglewski BH (1979) A more sensitive plate assay for detection of protease production by *Pseudomanas aeruginosa*. J Clin Microbiol 9:538-540

Sorensen HP, Mortensen KK (2005) Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. Microb Cell Fact 4:1.

Sternberg MJ, Hayes FR, Russell AJ, Thomas PG, Fersht AR (1987) Prediction of electrostatic effects of engineering of protein charges. Nature 330:86-88

Stoner MR, Dale DA, Gualfetti PJ, Becker T, Manning MC, Carpenter JF, Randolph TW (2004) Protease autolysis in heavy-duty liquid detergent formulations: effects of thermodynamic stabilizers and protease inhibitors. Enzyme and Microbial Technology 34:114-125

Strub C, Alies C, Lougarre A, Ladurantie C, Czaplicki J, Fournier D (2004) Mutation of exposed hydrophobic amino acids to arginine to increase protein stability. BMC Biochem 5:9

Takagi H, Hirai K, Maeda Y, Matsuzawa H, Nakamori S (2000) Engineering subtilisin E for enhanced stability and activity in polar organic solvents. J Biochem 127:617-625

Tee KL, Schwaneberg U (2006) A screening system for the directed evolution of epoxygenases: importance of position 184 in P450 BM3 for stereoselective styrene epoxidation. Angew Chem Int Ed Engl 45:5380-5383

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

Teshima G, Porter J, Yim K, Ling V, Guzzetta A (1991) Deamidation of soluble CD4 at asparagine-52 results in reduced binding capacity for the HIV-1 envelope glycoprotein gp120. Biochemistry 30:3916-3922

Toscano MD, Woycechowsky KJ, Hilvert D (2007) Minimalist active-site redesign: teaching old enzymes new tricks. Angew Chem Int Ed Engl 46:3212-3236

Van den Burg B, Vriend G, Veltman OR, Venema G, Eijsink VG (1998) Engineering an enzyme to resist boiling. Proc Natl Acad Sci U S A 95:2056-2060

Varadarajan N, Rodriguez S, Hwang BY, Georgiou G, Iverson BL (2008) Highly active and selective endopeptidases with programmed substrate specificities. Nat Chem Biol 4:290-294

Vojcic L, Despotovic D, Martinez R, Maurer K-H, Schwaneberg U (2012) An efficient transformation method for Bacillus subtilis DB104. Appl Microbiol Biotechnol 94:487-493

Walker B, Lynas JF (2001) Strategies for the inhibition of serine proteases. Cell Mol Life Sci 58:596-624

Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25:1157-1174

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

Wells JA, Estell DA (1988) Subtilisin-an enzyme designed to be engineered. Trends Biochem Sci 13:291-297

Wilson CR, Tang MR, Christianson T, Maurer K-H, Weiss A (1999) Expression systems for commercial production of cellulase and xylanase in *Bacillus subtilis* and *Bacillus licheniformis*. US Patent 5,888,800

Wong TS, Roccatano D, Loakes D, Tee KL, Schenk A, Hauer B, Schwaneberg U (2008) Transversion-enriched sequence saturation mutagenesis (SeSaM-Tv+): a random mutagenesis method with consecutive nucleotide exchanges that complements the bias of error-prone PCR. Biotechnol J 3:74-82

Wong TS, Tee KL, Hauer B, Schwaneberg U (2005) Sequence saturation mutagenesis with tunable mutation frequencies. Anal Biochem 341:187-189

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

Xue GP, Johnson JS, Bransgrove KL, Gregg K, Beard CE, Dalrymple BP, Gobius KS, Aylward JH (1997) Improvement of expression and secretion of a fungal xylanase in the rumen bacterium *Butyrivibrio fibrisolvens* OB156 by manipulation of promoter and signal sequences. J Biotechnol 54:139-148

Yutani K, Ogasahara K, Sugino Y, Matsushiro A (1977) Effect of a single amino acid substitution on stability of conformation of a protein. Nature 267:274-275

Zhu L, Verma R, Roccatano D, Ni Y, Sun ZH, Schwaneberg U (2010) A potential antitumor drug (arginine deiminase) reengineered for efficient operation under physiological conditions. Chembiochem 11:2294-2301

Zhu Z, Momeu C, Zakhartsev M, Schwaneberg U (2006) Making glucose oxidase fit for biofuel cell applications by directed protein evolution. Biosens Bioelectron 21:2046-2051

7 PUBLICATIONS

Jakob F, Martinez R, Mandawe J, Hellmuth H, Siegert P, Maurer K-H, Schwaneberg U (2013) Surface charge engineering of a Bacillus gibsonii subtilisin protease. Appl Microbiol Biotechnol 97(15):6793-6802

Martinez R, Jakob F, Tu R, Siegert P, Maurer K-H, Schwaneberg U (2013) Increasing activity and thermal resistance of Bacillus gibsonii alkaline protease (BgAP) by directed evolution. Biotechnol Bioeng 110(3):711-720

Siegert P, Evers S, Merkel M, Mußmann N, Hellmuth H, O'Connell T, Maurer K-H, Schwaneberg U, Jakob F, Martinez R, Laufs B, Aydemir A, Weber T (2012) Method for adapting a hydrolytic enzyme to a component that stabilizes the hydrolytic enzyme. Henkel AG & Co KGaA WO/2013/037609

Jakob F, Lehmann C, Martinez R, Schwaneberg U (2013) Increasing protein production by directed vector backbone evolution. AMB Express 3(1):39

8 DECLARATION

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

Felix Jakob, Aachen

9 CURRICULUM VITAE

Name	Felix Jakob
Birthday	28th October 1983
Place of birth	Freiburg im Breisgau, Germany

Education

2009 - 2013	**Ph.D. fellow at the Institute of Biotechnology, RWTH Aachen University** Thesis: “Engineering of subtilisin proteases for detergent applications”
2004 - 2009	**Diploma in Bioengineering, FH Aachen University of Applied Sciences** Thesis: „Einfluss von chromosomal codierten Proteasen auf die Proteaseproduktion mit *Bacillus licheniformis* unter industriellen Bedingungen“
1997 - 2003	Geschwister-Scholl-Gymnasium, Lüdenscheid
1994 - 1997	Martin-Schongauer-Gymnasium, Breisach

Felix Jakob, Aachen