„Directed subtilisin E evolution into a chaotolerant protease“

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

Master of Fermentation Engineering

Zhenwei Li

aus Wuxi, Jiangsu, China

Berichter: Universitätsprofessor Dr. Ulrich Schwaneberg
Universitätsprofessor Dr. Danilo Roccatano

Tag der mündlichen Prüfung: 07.12.2012

Acknowledgements

I couldn’t imagine my Ph.D. study is going to finish soon. Peering into the last three and a half years, I find I’ve really learnt a lot, not only in research but also in life as well. I know all these are blessed by God. Hereby I would like to give my gratitude to all the people helping me through my Ph.D. study in Germany.

Firstly, I would like to thank Prof. Dr. Ulrich Schwaneberg for giving me such great opportunity to work in his research group. As my supervisor, he was always eager to give me advices when I confronted some problems in research, even though I was still in Jacobs University Bremen for the first two years, while he was in RWTH Aachen University at that moment, which is around 400 km away. Moreover, he helped me a lot with different issues in my life too, particularly for the transfer from Jacobs University Bremen to RWTH Aachen University and my personal career development.

Here I acknowledge BMBF for my financial support and let me become one research member of Biokatalyse2021 cluster. I would like to thank our Dean Prof. Dr. Bernhard Kramer in Jacobs University Bremen for allowing me to start my study in Jacobs University Bremen and later transfer to RWTH Aachen University.

I would like to thank Prof. Dr. Danilo Roccatano, Prof. Dr. Lars Blank and Prof. Dr. Alan Slusarenko to serve in my thesis committee and spend their precious time to read my thesis. Additionally I would like to thank Prof. Dr. Danilo Roccatano for teaching and helping me with all my MD simulation setups and analyses. Without his help, I couldn’t accomplish them on my own.

I would like to give gratitude to my cooperative partner Dr. Helge Mühl from Molzym GmbH & Co.KG for sharing protocols, arranging product transfer and teaching me in person at the beginning of this project and Dr. Florian Kuhnen from Bremen Hochschule for sharing ideas for enzyme production.

Schwaneberg group is really a team-working family. I couldn’t achieve

anything without continuous help from my coworkers. I would like to thank Dr. Ran Tu, Dr. Mathias Klein and Dr. Hemanshu Mundhada for supervision and fruitful discussions in my project. I also would like to thank Dr. Ronny Martinez and Aamir Shehzad for helping me with protein purification. I also would like to thank my students Blessing Musunda and Alexandra Balaceanu for their help in library screening and simulations. Of course, I will never forget countless help from all other group members. There wouldn’t be such a nice research atmosphere without them.

I am really lucky to have so many kind and lovely Chinese coworkers and friends around me. I would like to thank Dr. Liqing Zhao, Dr. Leilei Zhu and Dr. Heifeng Liu to become real guides in my research and life. I would like to thank junior Ph.D. fellows Jing Zhao, Feng Cheng and Chi Zhang to become my real companions in my last year of Ph.D. study. I would like to thank all my Chinese brothers and sisters in Chinese Christian Communities both in Bremen and Aachen for always comforting and praying for me.

Last but not least, I would like sincerely thank my parents to support me to come to Germany for my Ph.D. study and continuous comfort and care about me regardless any costs. I also would like to thank my best German friend Ralf Erdmann to help me a lot to adapt to the life here and become a best companion all through these years.

Anyways, I have to end up this acknowledgment, though there might be many people not mentioned here. Please accept my sincere gratitude from the bottom of my heart. I will keep my scientific work in the future and make contributions to make this world better.

Zhenwei Li

Abstract

Proteases are industrially important enzymes and have niche applications in diagnostic kits that use cell lysis and thereby require high resistance towards chaotropic salts and detergents, such as guanidinium chloride (GdmCl) and sodium dodecylsulfate (SDS). Subtilisin E, a well-studied serine protease, was selected to be re-engineered by directed evolution into a “chaotolerant” protease that would be resistant towards GdmCl and SDS, for application in diagnostic kits. In three iterative rounds of directed evolution, the variant SeSaM1-5 (S62I/A153V/G166S/I205V), designated as M4, was generated, with improved activity (330 % on Suc-AAPF-AMC) and increased half life in 1 M GdmCl (<2 min to 4.7 h) or in 0.5 % SDS (<2 min to 2.7 h).

Saturation mutagenesis in the subtilisin E wild type gene of each site identified in directed subtilisin E evolution revealed that positions 62, 166, 218 and 224 were mainly responsible for increased activity and for GdmCl or SDS resistance. A double mutant, M2 (S62I/G166M), generated by combination of the best two single substitutions showed significantly improved kinetic constants; in 2 M GdmCl the K_m value decreased (29 fold) from 7.31 to 0.25 mM, and the k_{cat} values increased (4 fold) from 15 to 61 s^{-1}. The catalytic efficiency, k_{cat}/K_m, improved dramatically (GdmCl: 247 $mM^{-1} \cdot s^{-1}$ (118-fold); SDS, 179 $mM^{-1} \cdot s^{-1}$ (13-fold)). In addition, M4 showed longer half life in 2.0 % SDS when compared to the wild-type ($t_{1/2}$ 54.8 min (>27 fold)).

Two other variants, M5 (S62I/A153V/G166S/T224A/T240S) and M6 (S62I/A153V/G166S/I205V/N218S/T224A) were constructed by combination of additional beneficial substitutions for further improvement of activity and resistance towards GdmCl and SDS. Inactivation curves showed that the activity of M6 towards Suc-AAPF-pNA as substrate was much improved compared to M4 in the presence of up to 5 M GdmCl (IC_{50} (M4): 2.7 M; IC_{50} (M6): 4.6 M) and up to 4 % SDS (IC_{50} (M4): 1.5 % ; IC_{50} (M6): 4.0 %). Half life

value of M6 in the presence of GdmCl was also increased (3 M GdmCl: $t_{1/2}$ 68 min; 2 % SDS: $t_{1/2}$ 66 min). M5 was slightly more active in GdmCl but less active in SDS compared to M4. Additionally M5 was more active towards complex protein substrate (Azocasein) when compared to both M4 and M6. However, M5 was not stable in GdmCl or SDS in terms of its half life. Circular dichroism (CD) spectra analysis showed that secondary structures of subtilisin E wild type and three mutants (M4, M5 and M6) were almost intact at lower concentrations of GdmCl (1-2 M) and SDS (0.5-2.0 %), indicating inactivation precedes conformation changes in subtilisin E in the presence of chaotropic agents.

Finally, model building, molecular docking and molecular dynamics (MD) simulations of subtilisin E wild type showed that the guanidinium cation (Gdm^+) or the monomeric dodecylsulfate anion (DS^-) are likely to directly interact with the active site or substrate binding pocket residues, probably limiting the access of the substrate to the catalytic centre. Simulations of M4 indicated that the residues forming part of the active site or substrate binding pocket of the mutant are less affected or accessible by Gdm^+ or DS^-.

Abbreviation

Amp	Ampicillin
BSA	Bovine serum albumin
B. subtilis	*Bacillus subtilis*
CD	Circular dichroism
CV	Coefficient of variations
DMF	Dimethylformamide
DS^-	Dodecylsulfate anion
E. coli	*Escherichia coli*
epPCR	Error-prone PCR
GdmCl	Guanidinium chloride
Gdm^+	Guanidinium cation
MAP	Mutagenesis assistant program
MD	Molecular Dynamics
MTP	Microtiter plate
PLICing	Phosphorothioate-based ligase-independent gene cloning
RMSD	Root mean square deviations
RMSF	Root mean square fluctuations
SASA	Solvent accessible surface area
SDF	Spatial distribution functions
SDM	Site directed mutagenesis
SDS	Sodium dodecylsulfate
SeSaM	Sequence saturation mutagenesis
SSM	Site saturation mutagenesis
StEP	Staggered extension process
Suc-AAPF-AMC	Succinyl-Ala-Ala-Pro-Phe-7-amino-4-methylcoumarin
Suc-AAPF-pNA	Succinyl- Ala-Ala-Pro-Phe-p-nitroanilide
Tet	Tetracycline
WT	Wild type

Contents

ACKNOWLEDGEMENTS ..... **I**

ABSTRACT ..... **III**

ABBREVIATION ..... **V**

1. INTRODUCTION **1**

1.1 Proteases **1**
1.1.1 Brief introduction 1
1.1.2 Application of protease in industry 3

1.2 Protease engineering by directed evolution **5**
1.2.1 Diversity generation 6
1.2.2 Screening systems 9
1.2.3 Directed evolution of subtilisin E 10

1.3 Stability of enzymes in chaotropes **11**
1.3.1 Chaotropes 11
1.3.2 Chaotolerant proteases 16

1.4 Computational simulations of chaotropes **17**
1.4.1 Simulations on GdmCl 17
1.4.2 Simulations on SDS 21

1.5 Objective of the project **23**

2. MATERIALS AND METHODS **24**

2.1 Materials **24**
2.1.1 Chemicals 24
2.1.2 Strains 24
2.1.3 Plasmids 25
2.1.4 Oligonucleotides 26
2.1.5 Media and culture cultivation 28

2.2 Methods **29**
2.2.1 General methods 29
2.2.2 Construction of expression vector and strains 30
2.2.3 Generation of random mutagenesis libraries 31
2.2.4 Site-directed and saturation mutagenesis 32
2.2.5 SeSaM-based shuffling 32

2.2.6 Agar plate assay for detection of protease activity 33
2.2.7 MTP screening assays for activity and resistance towards chaotropes 33
2.2.8 Purification 34
2.2.9 Determination of kinetic constants 35
2.2.10 Inactivation curve determination 36
2.2.11 Stability in the presence of chaotropes 36
2.2.12 Activity towards complex protein substrate 36
2.2.13 Circular dichroism (CD) analysis 37
2.2.14 Computational analyses 38

3. RESULTS 43

3.1 Directed evolution of subtilisin E 43
3.1.1 Cloning and expression of subtilisin E in *B. subtilis* 43
3.1.2 Development of screening systems 44
3.1.3 Construction of random mutagenesis libraries 49
3.1.4 Screening of random mutagenesis libraries 51
3.1.5 Comparison of selected variants 53

3.2 Site saturation mutagenesis and recombination of beneficial substitutions 55
3.2.1 Construction and screening of SSM libraries 56
3.2.2 Recombination of beneficial mutations 60
3.2.3 Purification of subtilisin E and its variants 68
3.2.4 Characterization of subtilisin E and its variants 71
3.2.5 Stability of variants in the presence of GdmCl or SDS 75
3.2.6 Analysis of structural stability 78

3.3 Computational analysis for chaotrope resistance 80
3.3.1 Model building of subtilisin E-substrate complex 80
3.3.2 Docking of chaotrope in subtilisin E 82
3.3.3 MD simulations of GdmCl aqueous solutions 85
3.3.4 MD simulations of subtilisin E in the presence of SDS aqueous solutions 100

4. DISCUSSION 110

4.1 Directed evolution of proteases for chaotrope resistance 110

4.2 Effects of substitutions on chaotrope resistance 111

4.3 Inactivation and denaturation of enzymes 114

4.4 Conclusions and outlook 118

5. REFERENCES 121

APPENDIX I: SEQUENCE OF SUBTILISIN E SYNTHETIC GENE APRE (M14) 126

APPENDIX II: PUBLICATION 127

CURRICULUM VITAE 128

1. Introduction

1.1 Proteases

1.1.1 Brief introduction

1.1.1.1Protease

Proteases are an important family of hydrolases, which includes proteinases and peptidases, and defined as a class of enzymes which catalyzes the hydrolytic breakdown of proteins through the peptide bond. Proteases can be found in all forms of life on the earth, including prokaryotes, fungi, plants and animals [1]. Proteases can be divided into four subgroups: 1) serine proteases, 2) aspartic proteases, 3) cysteine proteases, and 4) metalloproteases based on their catalytic site and mechanism [2]. Almost one-third of all proteases can be classified as serine proteases (named from serine residue at the active site for nucleophilic attack of peptide bond). Mechanism of serine proteases normally consists of an Asp-His-Ser catalytic triad system. By the differences of structural contexts serine proteases can be further divided into six clans. The Asp-His-Ser catalytic triad can be found in four clans: chymotrypsin (SA), subtilisin (SB), carboxypeptidase Y (SC) and Clp protease (SK) [3]. Figure 1.1 shows the catalytic triad in a typical serine protease subtilisin [4].

1.1.1.2 Subtilisin E

Subtilisin is an important group within the serine protease family and a very important industrial enzyme. It has served as a model enzyme for protein engineering investigation. It has many advantages, such as enormous rate of enhancements, timely cloning of the gene, ease of expression, purification and availability of atomic resolution structures. This has be proven by the fact that over 50 % of the 275 amino

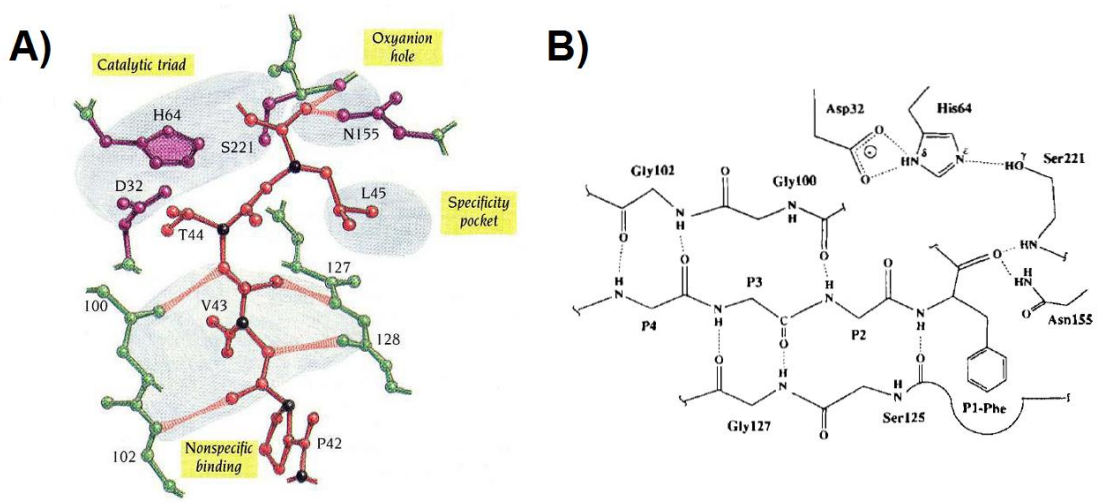

Figure 1.1 Schematic diagram of the active site of subtilisin. A) Active regions. A region of a bound polypeptide inhibitor is shown in red. The four essential features of the active site are highlighted in yellow [4]. B) The catalytic group interacting with polypeptide substrate binding to P1-P4 sites (Nomenclature for substrate amino acid residues is Pn) [5].

acids have been already mutated [6]. Though there are many types of subtilisins available, the most mutagenized are those secreted by *Bacillus amyloliquefaciens* (subtilisin BPN'), *subtilis* (subtilisin E) and *lentus* (savinase).

Subtilisin E is an alkaline serine protease that is produced by *Bacillus subtilis* 168 [7]. Like other subtilisin counterparts, it is encoded as a pre-pro-peptide. The N-terminal pre-sequence is a signal peptide responsible for secretion, while the pro-sequence is considered as a chaperone, allowing the correct folding of the protease, and then being cleaved by the mature protease itself [8]. Thus, unfolding of the mature protease is irreversible. As in most subtilisins, its catalytic triad consists of aspartic acid, histidine and serine (see Figure 1.2).

Up to now, site-directed mutagenesis (SDM) and directed evolution have been used in the study of subtilisin E. For instance, the introduction of an additional disulfide bond linkage between residues Cys 61 and Cys 98 using SDM method resulted in the increased thermostability [9]. SDM has also been employed in the construction of subtilisin variants showing improved storage and oxidation stabilities [10]. A mutant subtilisin E with enhanced thermostability at 60°C was generated using SDM, by forming a disulfide bridge between two subtilisin E molecules having the Ser236Cys substitution [11].

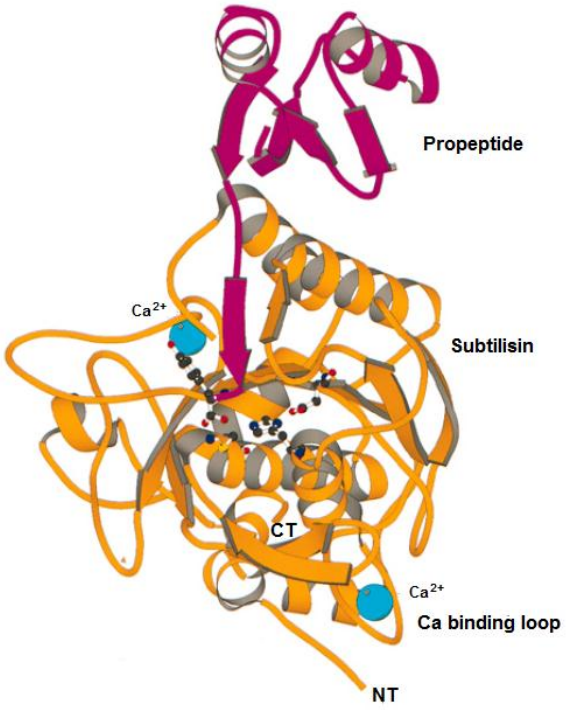

Figure 1.2 Structure of the subtilisin-propeptide complex. The subtilisin domain is shown in amber and the propeptide domain in magenta. Subtilisin's three catalytic residues of the active site are Asp32, His64 and Ser221 [12].

1.1.2 Application of protease in industry

The use of proteases has been widely applied to degradation of proteins in industrial process. Proteases account for approximately 40 % of the total enzyme sales in various industrial market sectors, such as detergent, food, pharmaceutical, leather, diagnostics, waste management and silver recovery [1]. Proteases for detergent occupy great section in the application of proteases in industry, so that great efforts are being made for novel proteases in this field [13]. In addition, proteases can be used in the removal of protein pollution as an additive in various cleaning and washing process related to environmental issues. In these decades, the stability of proteases under unfavorable industrial conditions is of great interest.

The term "stability" refers to a protein's resistance to adverse influences such as heat or denaturants, that is, to the persistence of its molecular integrity or biological function in the face of high temperatures or other deleterious influences [14].

Naturally available enzymes are usually not optimally suited for industrial applications due to their poor performance under industrial process conditions, such as high temperature, extreme pH and organic cosolvent. This can often be related to a stability of enzymes under those conditions. Thus, the use of stable enzymes under such conditions is often desirable in industry.

The stability of an enzyme is affected by many factors, such as temperature, pH, oxidative stress, solvent properties, the binding of metal ions or co-factors, and the presence of surfactants. Among these variables, the effect of temperature is the best studied.

Up until now, there are three biotechnological approaches to generate stable variants of a given protein: 1) isolating enzyme variants from organisms living in appropriate extreme environments [15]; 2) rationale-based mutagenesis [16] and 3) directed evolution [17]. Combinations of strategies (e.g. rational and evolutionary design) can reduce development times significantly and broaden our understanding of structure-function relationships.

Very interestingly, proteases were among the first enzymes to be targeted for directed evolution approach, mainly because of the need for stable and functional variants in detergent applications. One of the early studies on family shuffling of 26 subtilisins from *Bacillus* species for multiple properties: activity at 23°C and at three different pH values, thermostability and solvent stability (35 % DMF) [18]. It was reported that the best variants had up to three times more residual activity after heat treatment or up to 50 % greater residual activity in DMF. This revealed that new, valuable combinations of properties had been achieved, including considerable increases in stability.

The directed evolution application in protease stability is mainly on subtilisins, such as subtilisin E (see section 1.2) and its homologs. For instance, subtilisin S41 variant was 100-fold more thermostable than its wide type by using epPCR method [19].

Except the thermostability or stability in organic solvent, to our knowledge there are no studies in directed evolution applied to high resistance or stability of protease

in chaotropes (except some organic solvents) so far. Therefore, this project will make great contributions to broaden the research boundary concerning protease stability and deepen our understanding of structure-function relationship which governs chaotrope resistance of proteases.

1.2 Protease engineering by directed evolution

Directed evolution has been, during the past decades, a successful and promising approach for tailoring protein properties to industrial demands and advancing our understanding of structure-function relationships of biocatalysts.

The origin of directed evolution was derived from Spiegelman's concept "evolution in a test tube" [20]. Since then, the concept developed into a powerful algorithm as an increasing number of successful stories proved. Up to now, this method has been used in a wide range of applications, such as improving enantioselectivity [21], enzymes for bioremediation [22], and vaccines [23].

Though directed evolution comes from the idea of imitating Darwin's evolution process in laboratory rather than under natural evolutionary pressure, the former usually takes only few weeks or months whereas the later might takes thousands of years.

The directed evolution usually consists of three steps as illustrated in Figure 1.3. 1) Diversity generation; 2) Screening for improved variants; 3) Isolating the gene encoding for the improved protein variant. Iterative cycles are normally required until the desired property has been changed or improved significantly.

During the whole process, the first two steps are essential for a successful directed evolution experiments. However, it is still a challenge to generate an unbiased mutant library [24], and a challenge in setting up the screening systems. For instance, some high throughput screening techniques require further development to improve their accessibility.

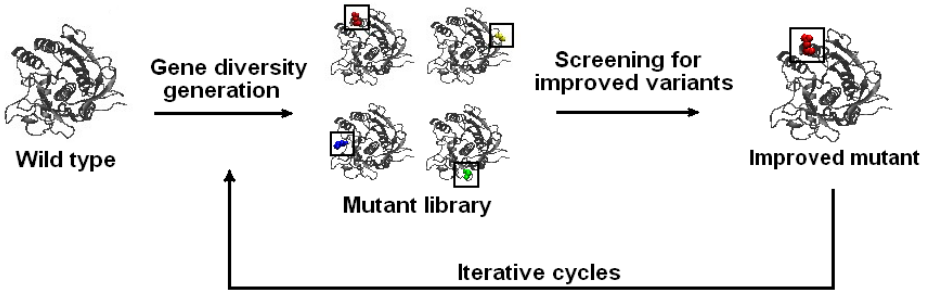

Figure 1.3 Overview of a typical directed evolution procedure. A gene library with diversity is generated firstly based on the wild type as template. An effective screening system is then applied to screen for the variants with improved properties as expected. The improved variant selected is considered as a new template for the next round of library generation. Iterative rounds of such process are usually performed until the desired variant is selected.

1.2.1 Diversity generation

During the past decades, more and more novel methods, besides DNA and genome shuffling for diversity generation have emerged. It is notable that a research begins focusing on the quality and comprehensiveness of library construction and analysis [25]. An increasing number of powerful new combinatorial techniques have joined rational design methods as effective tools for the manipulation and tailoring of biocatalysts, such as staggered extension process (StEP) [26], random chimeragenesis on transient templates (RACHITT) [27] and iterative truncation for the creation of hybrid enzymes (ITCHY) [28]. These advances are summarized in some recent literature [29]. Despite of this, some conventional methods, like error-prone PCR, due to its simplicity, are still widely used.

Error-prone PCR (epPCR)

The most commonly used random mutagenesis techniques are the PCR methods under low fidelity conditions. *Taq* DNA polymerase from *Thermus aquaticus* that lacks a 3'-5' proofreading exonuclease activity, is usually used [30]. According to its error-prone percentage, it gives around one mutation per 125,000 bp per PCR cycle which is insufficient for randomizing genes that seldom exceed 5,000 bp in length. There are many different ways of increasing the error rates of the polymerase, such

as by adding unbalancing nucleotide concentrations and/or chemical compounds such as manganese chloride, employing nucleotide analogs, increased extension times, etc [31].

Though epPCR based methods are quite simple, its disadvantages are obvious, like high number of transitions, lack of consecutive nucleotide exchange which makes it impossible to obtain some amino acid substitutions. There are also a high probability of stop codons and disruptive amino acids like Gly and Pro [32].

Sequence Saturation Mutagenesis (SeSaM)

In our group, a novel diversity generation method was developed and termed as Sequence Saturation Mutagenesis (SeSaM). SeSaM overcomes the current limitations caused by biased polymerases. Moreover, SeSaM can target a nucleotide species of the selected sequence and each nucleotide species can be exchanged in a controlled manner since SeSaM regulates the mutational spectra through a universal base [33].

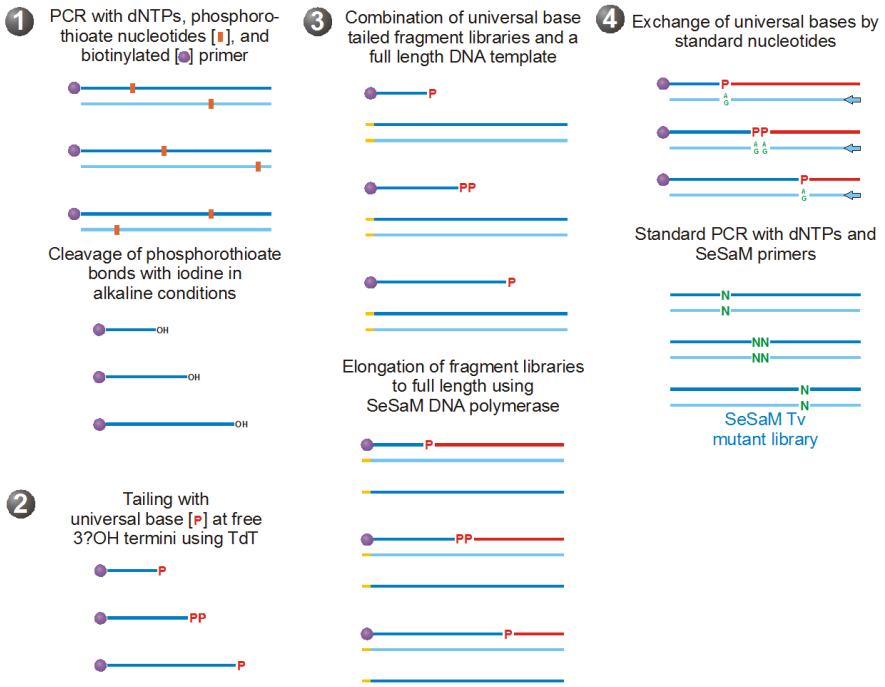

Figure 1.4 Overview of SeSaM method (http://www.sesam-biotech.com/sesamtech.html).

The SeSaM method comprises four steps (see Figure 1.4): Step 1, A PCR to

incorporate phosphorothioate nucleotides into the gene of interest followed by cleavage of the phosphorothioate bonds to create evenly distributed fragments of the gene; Step 2, Elongation of fragments in Step 1 with universal bases at 3'-OH terminal; Step 3, Fragments with the universal bases are elongated to their full length using the full length gene template; Step 4, Mutations are generated opposite to universal bases in a final amplification PCR.

SeSaM-based shuffling

A novel combinatorial mutagenesis method is designed based on SeSaM method and termed as "SeSaM-based shuffling". This is an alternative method for other widely used combinational methods such as gene shuffling or StEP [26]. The concept is gene fragmentation and recombination on SeSaM platform by using one wild type gene (WT) and one synthetic gene (MX) which includes all substitutions of interest. The procedure comprising four steps are shown in Figure 1.5: Step 1, gene template generation derived from SeSaM preparation step; Step 2, SeSaM PCRs (SeSaM biotin primers) and cleavage (iodine under alkaline conditions) derived from SeSaM step 1; Step 3, SeSaM full length synthesis derived from SeSaM step3; Step 4, Standard PCR amplification. Original SeSaM step 2 is not included since the addition of universal base is not required.

The key step is step 3 (see Figure 1.5), in which extension of PCR cycles can generate different recombination patterns. Besides, the ratio of amount of wild type and synthetic gene can also affect the recombination results. Thus these conditions should be optimized in order to obtain a combinatorial mutagenesis library with randomly distributed and combined mutations.

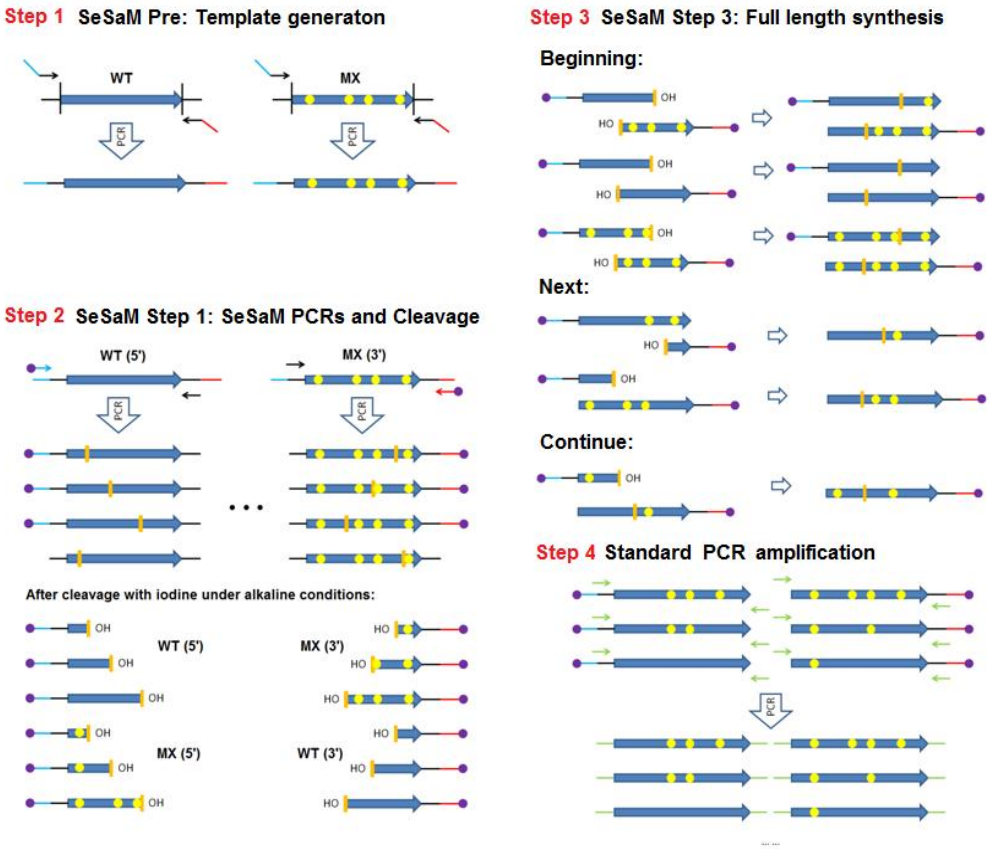

Figure 1.5 Overview of SeSaM-based shuffling procedure. The complete procedure includes four steps: Step 1, Template generation; Step 2, SeSaM PCRs (SeSaM biotin primers) and Cleavage (iodine); Step 3, SeSaM full length synthesis; Step 4, Standard PCR amplification. WT, wild type; MX, synthetic gene mutant (mutations highlighted in yellow).

1.2.2 Screening systems

Solid phase and microtiter plate based screening are still routinely and widely used in directed evolution, despite of the increasing number of novel screening methods available, such as *in vitro* compartmentalization [34] and cell surface display [35].

Solid phase, mainly agar plate screening, is usually based on color change, so that the activity difference could be simply distinguishable by observation. However, it lacks accuracy and therefore not suitable for quantitative activity measurements. In contrast, microtiter plate assays can be used to screen for up to 10^5 clones and are suitable for quantitative measurement, although practically this number of clones is seldom reached (see Figure 1.6).

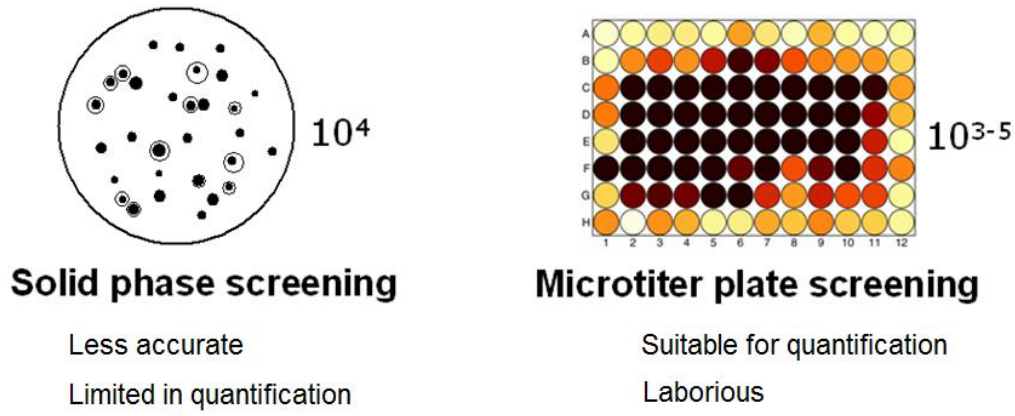

Figure 1.6 Comparison of commonly used screening systems for directed evolution. Solid phase screening and microtiter plate screening are based on color change or fluorescent change which follows enzymatic reaction, using either chromogenic substrate or coupled reaction of substrate.

1.2.3 Directed evolution of subtilisin E

Directed evolution can effectively alter protein properties despite any understanding of the targeted characteristic (random mutant libraries) in contrast to SDM (focused mutant libraries) which requires the identification of specific targets in a protein structure. Table 1.7 lists the representative successful directed evolution examples employing subtilisin E. The total activity and organic solvent stability of subtilisin E has been enhanced in aqueous dimethyl formamide (DMF) using a directed evolution approach [36]. To demonstrate the utility of this method, Staggered Extension Process (StEP) PCR was used to recombine a set of five thermostabilized subtilisin E variants identified during a single round of error-prone PCR mutagenesis and screening. Screening the StEP-recombined library yielded a subtilisin variant whose half-life was increased 50-fold at 65 °C, compared with wild type [37]. Shao used random-priming for generation of the mutant library, which could be also effective to elongate the half life [38]. In another report, the thermal stability of subtilisin E was increased by converting it into a thermitase using directed evolution [39].

To date, there are no reports available of subtilisin E regarding its resistance towards chaotropic agents, especially chaotropic salts or detergents. However, some

early references stated that subtilisins had modest tolerance towards chaotropic agents, like guanidinium chloride (GdmCl), as well as some detergents [40]. Thus, subtilisin E, though it is a mesophilic protease, is still regarded as a suitable protease candidate for evolution into a better chaotolerant protease for further applications.

1.3 Stability of enzymes in chaotropes

1.3.1 Chaotropes

A chaotrope (chaotropic agent) is a substance which disrupts and denatures the structure of macromolecules such as proteins and nucleic acids (e.g. DNA and RNA, not discussed in this dissertation).

Chaotropic solutes can interfere with intramolecular interactions within proteins mediated by non-covalent forces, like van der Waals forces, hydrogen bonding and hydrophobic forces, and thus increase the entropy of the system. The structure and function of proteins mostly depend on the net effect of these forces, therefore an increase of the entropy in chaotropic solutes in a biological system will denature proteins or reduce enzymatic activity. Tertiary protein folding mostly relies on hydrophobic forces from amino acids determined by the sequence of the protein. Conventionally it is assumed that chaotropic solutes decrease the net hydrophobic effect of hydrophobic regions because of the disordering of water molecules

Table 1.7 Overview of previous studies on directed evolution of subtilisin E.

Strategy	Change in property	Reference
Error-prone PCR	Activity increased up to 170-fold in dimethyl formamide (DMF)	[41]
Error-prone PCR	Activity increased up to 500-fold in dimethyl formamide (DMF)	[36]
StEP combination	$t_{1/2}$ increased up to 50-fold at 65°C	[37]
Random-priming	$t_{1/2}$ increase up to 8-fold at 65°C	[38]
Error-prone PCRDNA shuffling	Increased activity at different temperature and 17°C rise in Tm	[39]

surrounding the protein. This solubilization of the hydrophobic region results in the denaturation of the protein. It is reported such application in the hydrophobic region in lipid bilayers: when a critical concentration of a chaotropic solute is reached, membrane integrity will be compromised, and the cell will lyse [42]. Therefore it is the reason for its application for cell lysis in diagnostic industry.

Chaotropic agents include some organic solvents, like butanol, ethanol, propanol, some salts, like guanidinium chloride (GdmCl), magnesium chloride ($MgCl_2$), lithium perchlorate ($LiClO_4$); detergents, like sodium dodecyl sulfate (SDS) and other organic compound like urea ($(NH_2)_2CO$). This dissertation focuses on chaotropic salts or detergents, where GdmCl and SDS were selected as representatives for both classes due to their wide usage in academic and industrial applications.

1.3.1.1 Chaotropic salts

In general, salts can be either chaotropes or kosmotropes (structure-maker, opposite from chaotropes), depending on their specific ion effects, which are well known as “Hofmeister Series” [43]. In the past, the mechanism of protein denaturation by chaotropic salt like GdmCl was proposed that they perturbed the water’s bulk structure, destabilizing the protein fold. However, further studies imply that such denaturants exert their influence through direct interaction with the protein, not by restricting the bulk water, though it is still being debated [44]. The effects of ionic kosmotropes are usually related to the directed and polarized arrangements of surrounding water molecules. However, large single charged ions, with low charge density, such as SCN^-, $H_2PO_4^-$, HSO_4^-, HCO_3^-, I^-, Cl^-, NO_3^-, NH_4^+, Cs^+, K^+, $(NH_2)_3C^+$ (guanidinium) and $(CH_3)_4N^+$ (tetramethylammonium) ions, exhibiting weaker interactions with water than water with itself and therefore interfering little in the hydrogen bonding of the surrounding water molecules, are chaotropes. However, small or multiply-charged ions, with high charge density, are kosmotropes, such as SO_4^{2-}, HPO_4^{2-}, Mg^{2+}, Ca^{2+}, Li^+, Na^+, H^+, OH^- and HPO_4^{2-}, exhibiting stronger interactions with water molecules than water with itself and thus capable of breaking

water-water hydrogen bonds. Furthermore, neutron diffraction studies on a very effective chaotropic salt (guanidinium thiocyanate, GdmSCN) showed very poor hydration of both Gdm^+ and SCN^-, supporting the suggestion that they preferentially interact with the protein rather than the water [45].

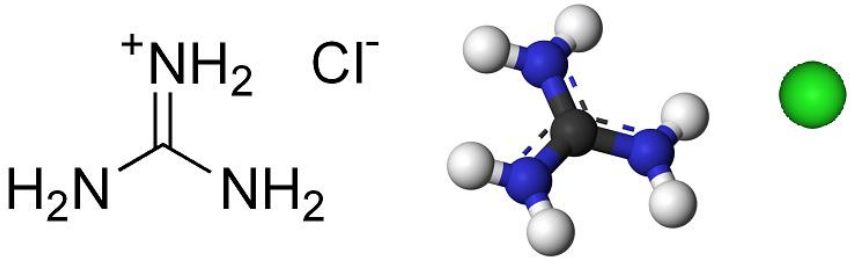

Figure 1.8 Structure of guanidinium chloride (GdmCl)

In case of GdmCl, guanidinium [46] (see Figure 1.8) is a planar ion that can form weak hydrogen bonds around its edge but may form strongly hydrogen-bonded ion pairs to protein carboxylate groups. Also, guanidinium possesses rather hydrophobic surfaces that may interact with similar protein surfaces to enable protein denaturation [47]. The hydration of the counterion is also important to the action of guanidinium with chloride ions, which is only weakly hydrated, allowing the water molecules to easily rearrange around the surface and allowing the denaturation process. However, sulfate ions trap their water molecules very strongly and prevent the water rearrangement and thus prevent the guanidinium-protein interactions and consequent denaturation [48].

1.3.1.2 Chaotropic detergents

Detergent usually refers to a surfactant or a mixture of surfactants with “cleaning properties”, which can classify ionic detergents (like SDS) and non-ionic detergents (like Tween series). This dissertation is only focused on ionic detergent SDS.

At high enough aqueous concentrations of surfactant (the critical micelle concentration or CMC), it becomes favorable for the surfactant molecules to associate via their hydrophobic chains to form micelles with a generally hydrophobic interior and

a hydrophilic water-exposed exterior (see Figure 1.9). The CMC is the most important characteristic of a surfactant, because proteins interact very differently with monomeric and micellar surfactants. The CMC value of certain surfactant can determined by many different methods [49] and is affected by ionic strength, since the increase in ionic strength could reduce the electrostatic repulsion between the ionic head groups. For instance, CMC of SDS in water is around 7 mM, but 1-2 mM in 25 mM Tris at pH 8 [50].

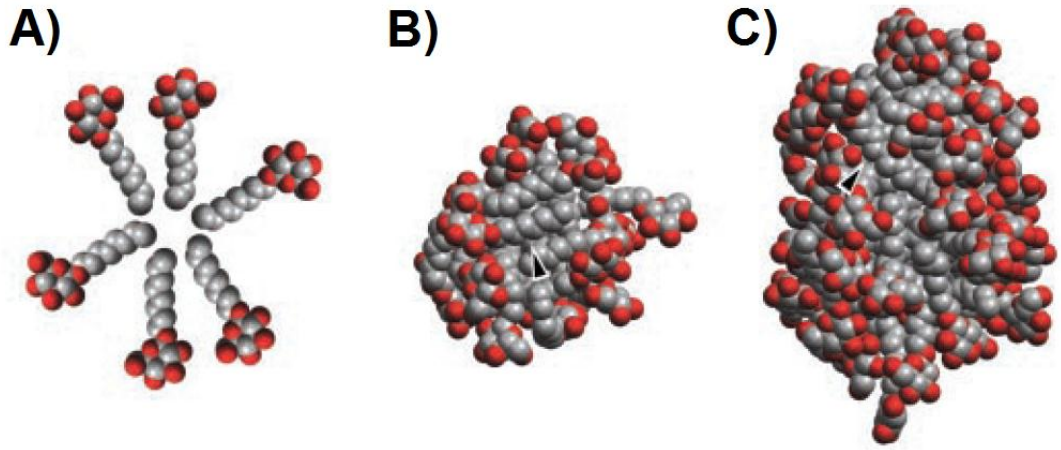

Figure 1.9 Space filling models of β-D-octyl glucoside micelles: A) classic representation based on 6-monomer; B) 20-monomer micelle; C) 50-monomer micelle. The micelles shown in B) and C) have nonspherical and nonuniform shapes. The polar portions of the detergents (oxygen atoms are red; carbon atoms, gray) do not cover completely the micelle surface. Substantial portions of the core are exposed to bulk solvent, including alkyl chains lying along the micelle surface (arrowheads) [51].

SDS is a very common used anionic surfactant in biochemistry. Its anion dodecyl sulfate is an amphiphilic molecule, consisting of a hydrophilic sulfate head group and a hydrophobic alkyl tail group. Different from other chaotropes such as GdmCl or urea, SDS is almost 1,000 fold more effective in protein denaturation, since it causes protein denaturation at a very low concentration in a millimolar range and often below its CMC, whereas a molar range of GdmCl or urea usually has such effect.

The protein-surfactant interaction has been studied for decades and many models were proposed, as summarized in a recent review [52]. A simple model for ionic surfactant like SDS is an “electrostatic-hydrophobic binding and denaturation” model, which means that hydrophilic head groups have electrostatic interactions with

charged residues (for DS^- with Arg, Lys) and hydrophobic tails have interactions with hydrophobic areas of a protein.

Regarding steps in protein-surfactant interactions can normally be split into two parts: below and above the CMC. Bhuyan concluded that tertiary structures unfold in the submicellar, and chain expands in the micellar range of SDS concentration. They are the two major and discrete events in the perturbation of protein structure [53]. However, in contrast to the random coil observed in GdmCl or urea, SDS-denatured proteins retain a large degree of ordered structure, albeit non-native stabilized by the SDS molecules.

Besides of its denaturation effects, surfactants like SDS can also activate or inhibit a protein. SDS is an activator of a β-glycosidase from *Sulfolobus solfataricus* [54]; this β-galactosidase is slightly activated by 35-70 mM SDS but inactivated at higher concentrations (350 mM) [55]. As an enzymatic inhibitor, there are also many examples, targeting α-chymotrypsin [56] as noncompetitive inhibitor, amino acid oxidase as mixed inhibitor [57], and also reducing enzymatic activity of the fatty acid synthase [58] and α-glucosidase [59]. These effects take place at concentrations well below those where conformational changes and loss of native structure occur, indicating that SDS binds as a ligand rather than a self-assembling and denaturing agent.

1.3.1.3 Application of Chaotropes

Chaotropic salts and detergents, such as GdmCl or SDS are widely used in routine biochemistry laboratory. For instance, 6 M GdmCl or 8 M urea is used as strong denaturant for studying protein unfolding. SDS is quite often used in SDS-PAGE for estimation of the protein molecular weight and for studying membrane associated proteins.

Besides, these denaturants also have niche markets in diagnostic application, such as nucleic acid purification. For instance, it is necessary during cell disruption to immediately inactivate released nucleolytic enzymes, like ubiquitous RNases. Thus, mainly strong chaotropic salts, such as guanidinium chloride (GdmCl) or guanidinium

thiocyanate (GdmSCN), are utilized in high concentration [60]. In addition, some detergents such as SDS and Tween 20 are also supplemented for cell lysis.

1.3.2 Chaotolerant proteases

Chaophiles *(*chaophilic microbes), were proposed not long ago as a new class of extremophilic microbes that prefer chaotropic conditions [61]. Hence “chaophilic” or “chaotolerant” proteases here are labeled as proteases that prefer or tolerate chaotropic conditions. Halophilic and thermophilic proteases are the natural first choice for applications with chaotropic conditions, especially in the case of chaotropic salts such as GdmCl. Halophilic proteases have been mainly reported from archaea [62] and less frequently from bacteria [63]. Halophilic proteases often preserve high activity, solubility, and stability in the presence of high salt concentrations (up to 4 M); however, they are often denatured by GdmCl at concentrations above 3 M [64]. One hyperthermophilic protease, the recently discovered Tk-subtilisin from *Thermococcus kodakaraensis* seems to be most chaotolerant protease to our knowledge, as it retains full activity at up to 6 M GdmCl and 5 % SDS (at 55^{o}C) [65]. However, it seems that it only retains its high activity and resistance towards these denaturants at its optimal high temperature and directed evolution has been applied for low temperature adaption resulting in a more rapid maturing mutant [66]. Besides, it is demonstrated that it requires Ca^{2+} for folding rather than propeptide, which are required for other subtilisin family members. However, Ca^{2+} binding in Tk-subtilisin does not seriously contribute to the stabilization in a native structure, and thus it is also not the same as other subtilisins [67]. Protease K was considered as another chaotropic protease, since activities assay showed stimulation by chaotropic agents like urea and GdmCl in a proportional manner [68], but studies showed apparent activation of proteinase K by urea and dodecylsulfate which was caused primarily by denaturation of the complex protein substrates. When a peptide substrate was applied, the residual activity of protease K was dramatically decreased (18 % residual activity remained in 0.5 % SDS on substrate Z-$(Ala)_2$-Nan) [69].

In application of nucleic acid purification by chaotropic salts and detergents (see 1.3.1.3), protease K has been supplemented to completely prevent degradation of nucleic acids through digestion of nucleolytic enzymes. Nevertheless the options are quite limited and performance is not satisfactory [70]. Additionally, understanding the principles that control the activity and stability of proteases [10] in chaotropic salts and detergents is a poorly explored research area, despite its significance in applications (e.g. diagnostic kits). Only one rational design study (cutinase resistance) against a specific anionic detergent (dioctyl sulfosuccinate sodium salt, AOT) has been reported; the authors postulated that surface substitutions are the main contributors to detergent tolerance in cutinase [71]. No protein engineering study has hitherto been published that used directed protease evolution with the chaotropes GdmCl and SDS.

1.4 Computational simulations of chaotropes

To date, a variety of molecular dynamics (MD) simulations of chaotropes, like urea, GdmCl or SDS have been carried out, and many different claims have been made by different research groups based on their analysis of a large variety of data obtained from these computational simulations, such as solvation of hydrophobic groups, dewetting effects, etc. This diversity in the conclusions might be consequence of the approximation of the force fields used to describe the interaction of the chaotropic solutes with water and proteins models. Moreover, the length of these simulation studies might be also a limitation of the proper simulation of process that might occur on larger time scale. Different observations and conclusions for chaotrope simulations were compared and reviewed recently [44].

1.4.1 Simulations on GdmCl

Table 1.10 summarizes recent MD simulations of peptides and proteins in the presence of GdmCl. These simulations are very often compared to the ones in urea,

Table 1.10 Overview of published molecular dynamics simulations on GdmCl

Year	Simulations	Force field	Temperature	Time	Concentration	Details	Reference
2004	GdmCl in water	CHARMM22	300K	800 ps	3 M	31.7894 Å box containing 48 GdmCl units in 882 TIP3P water molecules	[47]
2007	α-Helix (H1) in GdmCl	CHARMM22	300K	25-40 ns	3.05 M	102 Gdm^+ molecules, 100 Cl^- ions and 1441 TIP3P water molecules were used in a cubic box ~ 32 Å per side.	[72]
2007	Melittin in GdmCl	CHARMM22	300K	8 ns	3 M	44.7 Å cube containing 125 GdmCl units, one melittin, 6 Cl^- counterions, and 2319 TIP3P waters.	[73]
2008	Protein L in GdmCl	GROMOS96	300K 400K 480K	10 ns 20 ns 30 ns	5 M	~100 nm^3 dodecahedron box, SPC water model, one Protein L	[74]
2009	Isopropanol or pyridine in GdmCl	CHARMM22	300K	10 ns	3 M	36 Gdm^+, 36 Cl^- and 36 pyridine or isopropanol in a cubic box of length 34 Å. TIP3P water model was used.	[75]
2010	Polymer in GdmCl	Amber	300K	20x5 ns	4.9 M	A cubic box of length ~5 nm. SPC/E model of water, OPLS model for the guanidinium ion were used	[76]
2010	Magainin in GdmCl	GROMOS96	300K	40 ns	1, 2, 4, 6 M	Cubic box, SPC water model, one Magainin	[77]
2012	Nanoparticle C_{60} in GdmCl	Amber(ff03)	300K	10 ns	4 M	The topology of Gdm^+ was adopted from arginine force field. TIP3P water model was used.	[78]

another classic chaotrope, for which a greater number of simulation studies are available. Simulation studies of urea-induced unfolding of larger proteins (>100 aa), such as barnase [79], lysozyme [80] and β-lactoglobulin [81] have been performed on time scale up to microsecond, but, to the best of our knowledge, so far only the small protein L consisting of 62 amino acids [74] has been studied and for less than 50 nanosecond (see Table 1.10). Based on the combination of these theoretical simulations and experimental data, a 2-stage urea unfolding mechanism of large proteins has been proposed. [80]. In the first stage, the concentration of urea molecules increases in the vicinity of the protein surface. In this way, urea molecules solvate the protein surface by replacing water molecules. In the second stage, urea molecules start to penetrate the interior of the protein, forming hydrogen bonds with backbone atoms and breaking hydrophobic interactions. Differently from urea, the mechanism of unfolding of large protein by GdmCl to date is still unclear.

In 2004, Philip Mason and coworkers studied firstly the structure of aqueous guanidinium chloride solutions by neutron diffraction with isotopic substitution (NDIS) and MD simulations [47]. The force field for the guanidinium ion was based on the parameter used for the guanidinium group of arginine side chain in the CHARMM22 force field. In this way, the atomic partial charges assigned symmetrically (atom charged: C 0.64; N -0.80; H 0.46). The analysis of anisotropic distribution of water atoms around the Gdm^+ solute revealed a strong tendency for the water molecules to make linear hydrogen bonds as acceptors for the guanidinium hydrogen atoms and strongly constrained to remain in the plane of the Gdm^+ ions. Another observation revealed by simulation is that the significant tendency for the Gdm^+ ions to self-associate in a stacking fashion, while a considerable numbers of guanidinium ions assemble into dimers or tetramers during simulations. Additionally, this stacking mode of guanidinium ions was also observed in the interaction of arginine side chains from a simulation of a model helical peptide Melittin in 3 M GdmCl [73] and in simulations of a bulky nanoparticle C_{60} in 4 M GdmCl [78].

Another observation revealed by MD simulations on GdmCl solution is that Gdm^+ ion has preferential interactions with aromatic groups over aliphatic groups [75]. Philip

Mason investigated the structuring of isopropanol and pyridine in 3 M GdmCl and found that addition of 3 M GdmCl considerably reduced pyridine aggregation but had no effect on isopropanol aggregation, indicating that Gdm^+ interacts with the planar pyridine group through hydrophobic interaction to suppress pyridine-pyridine interactions in solution. Hydrophobic interactions involving the aliphatic groups of isopropanol were unaffected by GdmCl, indicating that the planar and weakly hydrated Gdm^+ cation cannot make productive interactions with the highly curved or "lumpy" aliphatic groups of this solute. Same observations in other simulations were also reported [73, 77] regardless of different force fields they applied.

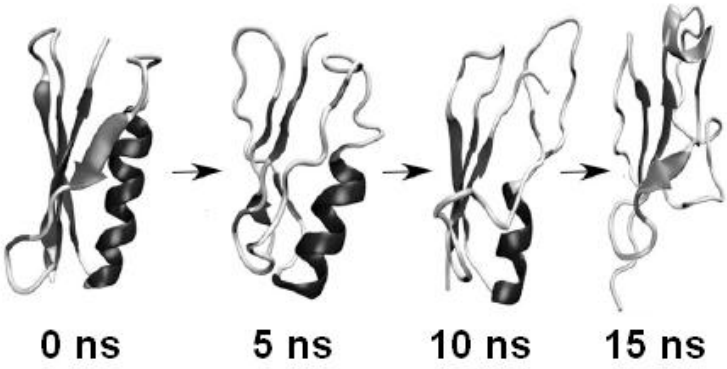

Figure 1.11 Snapshots of the unfolding of Protein L at 400K in 5 M GdmCl [74]. It is observed that the α-helix degrades first (within 5 ns) prior to β-sheet along 15 ns simulation.

The unfolding of peptides or proteins was also observed especially when the temperature is elevated. For instance, in the case of protein L (see Figure 1.11), the destabilization of α-helix prior to β-sheet in the presence of 5 M GdmCl at 400 K or 480 K were observed [74]. The unfolding of a peptide magainin at different concentrations of GdmCl revealed in lower concentrations of GdmCl (1 M and 2 M), that GdmCl molecules tend to engage in transient stacking interactions with aromatics and hydrophobic planar side chains. The latter can lead to displacement of water from the hydration surface and prevent proper solvation of the peptide at the beginning. In higher concentrations of GdmCl (>3 M), water solvates the peptide better than 1 M and 2 M GdmCl. Therefore, the results strongly suggest that hydrogen bonds between

water and the peptide are important factors in the destabilization of peptide in GdmCl solutions [77].

1.4.2 Simulations on SDS

Several molecular dynamics (MD) simulations have been performed to study the various aspects of SDS micelles, such as structure [82], behavior with the solvent water [83], free energy of micelle formation and size distribution [84], size effects on stability [85], surface structure [86], and self-assembly [87]. Besides of these atomistic simulations, a coarse-grained (CG) model of SDS micelles has been also proposed. This simplified model allow the simulation of significantly large systems (as the micelles) for long time scale (microseconds) [88].

These SDS models have been used to perform MD simulations on peptides or proteins in micelles at both atomistic and coarse-grained level. In these studies, the SDS micelle was considered as a simplified model of the biological membrane environment.

Rosemary and coworkers reported MD simulations of SDS micelle formation around dimeric glycophorin A (GpA) transmembrane helices to provide a detailed atomistic-level picture of conformation and dynamics of GpA-SDS system [89]. It was observed that the formation of a complete micelle around wild type GpA from an initially random place of SDS molecules in an aqueous solution. Additionally, the initial instability of GpA helices in water is reversed upon the coating of the helices by SDS.

Antimicrobial peptides (AMP), ovispirin-1 and its two analogs (G10 and T7) in SDS micelles were also studied by MD simulations by Khandelia and coworkers [90]. In the simulations, ovispirin, which was initially placed along the diameter of the SDS micelle, diffuses out of the micelle and keeps bound to the water-SDS interface. The final conformation and the structure of ovispirin are in a good agreement with experimental data in presence of lipid bilayers. Additionally, the simulations of two mutant peptides clarified the structural and binding properties which accounted for their less toxicity when compared to ovispirin.[90].

Simulation of one HIV-1 fusion peptide was also studied in SDS micelles [91]. At equilibrium, the peptide exhibited more flexible conformations at the N- and the C-terminus. The peptide with hydrophobic residues at the surface of the micelle could interact with the micelle's core. The primary hydrophobic interactions with the micelles are from the Leu and Phe residues, while the Ala and Gly residues interact favorably with water molecules. The results suggest that Phe-8, one of the highly conserved residues of the fusion peptide, plays a key role in the interaction of the peptide with membranes. This simulation provided atomistic details that explain their previous experimental findings such as NMR and fluorescent measurement.

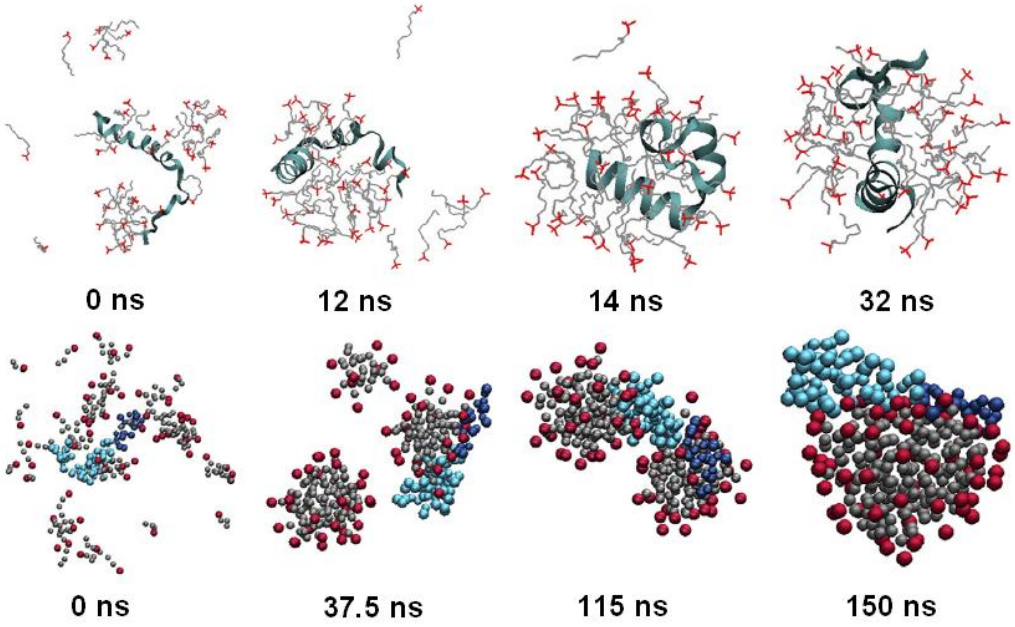

Figure 1.12 Formulation of the peptide-SDS complex. Snapshots from the atomistic simulation (top) and coarse-grained simulation (bottom) are shown at different time [92]. The main micelle (38 molecules) is formed in atomistic simulation. In the coarse-grained simulation, a complete micelle with 60 SDS molecules is formed around the peptide. The C-terminal (28-40) amino acids of the peptide are in dark blue.

Recently, MD simulations in both atomistic and coarse grain level of the Alzheimer's Aβ40 peptide in SDS micelles were reported [92]. The results showed that the experimental conformation of the peptide in SDS micelles is well reproduced using atomistic MD simulations, but the position of peptide relative to the micelle was shown to depend on the initial conditions most probably due to the limit of simulation time (50

ns). Coarse grained simulations in spite of its limitation in resolving atomistic details or secondary structure change were performed to clarify this discordance by extension of simulation time scale (500-1000 ns). Comparison of both atomistic and CG simulations showed that the used model is sufficient for protein-surfactant systems and could be used for more complex system. Figure 1.12 displayed the formation of the peptide-SDS complex in both atomistic and coarse-grained simulation in this report.

1.5 Objective of the project

The main objective of this project is to develop proteases which could tolerate high concentration of chaotropic salts and/or detergents by using directed evolution approach and understand the structure-function relationship of such chaotolerance with assistance of computational analyses.

The outcome of this project will be beneficial both in scientific research and industrial application. First of all, this project will provide one useful screening method for directed evolution of proteases towards chaotropic salts and detergents, which are web transferable to other proteases and chaotropes. Secondly, some promising variants with improved tolerance would be obtained, followed by intensive analyses on a structure-function relationship to shed the light on the protein-solvent interactions, especially under chaotropic conditions. Additionally, it could also elucidate the change of specific residues which govern protease chaotolerance. Computational analysis assists in first hypothesis to explain subtilisin E chaotolerance and last but not least, the promising protease variants will be applied into a new nucleic acid purification kit for commercial applications.

2. Materials and Methods

In this chapter, biological materials and methods related to this dissertation are listed. There are many molecular biology methods mentioned here as well. Some are conventional, such as error-prone PCR (epPCR), circular dichroism (CD) spectra, etc, but some are relatively novel, such as Sequence Saturation Mutagenesis (SeSaM) for generation of less-biased and transversion-enhanced mutant libraries and Phosphorothiated-based ligase-independent gene cloning (PLICing) for making generation of mutant libraries much easier. A novel combinatorial mutagenesis method “SeSaM-based shuffling” was also developed. These new methods were initially developed by author and/or coworkers, and the protocols were adjusted specifically to apply to this project.

2.1 Materials

2.1.1 Chemicals

All chemicals were of analytical reagent grade or higher quality and were purchased from Merck, Fluka, Sigma-Aldrich, and Applichem (Darmstadt, Germany), except the resins for purification (TOSOH, Stuttgart, Germany). All enzymes were purchased from New England Biolabs, Fermentas, and Sigma-Aldrich, unless otherwise stated. The fluorescent substrate Succinyl-Ala-Ala-Pro-Phe-7-amino-4-methylcoumarin (Suc-AAPF-AMC) was purchased from Bachem AG (Bubendorf, Switzerland), and Succinyl-Ala-Ala-Pro-Phe-p-nitroanilide (Suc-AAPF-pNA) and Azocasein were purchased from Sigma-Aldrich.

2.1.2 Strains

The bacteria strains used in this work are listed in Table 2.1.

Table 2.1 Summary of bacteria strains used in this work

Strain	Description	Reference
Escherichia coli DH5α	*F'/endA1 hsdR17(rK-mK+) supE44 thi-1 recA1 gyrA (Nalr) relA1 D(laclZYA-argF)U169 deoR (F80dlacD(lacZ)M15)*	Stratagene
Bacillus subtilis WB600	*nprE nprB aprE epr mpr bpr*	[93]

2.1.3 Plasmids

The plasmids used in this work are listed in Table 2.2.

Table 2.2 Summary of plasmids used in this work

Plasmid	Description	Reference
pHY300PLK	*E.coli* (Ap^r) and *B. sutbitlis* (Tc^r) shuttle vector, empty vector	TaKaRa
pBE3*	*E.coli* (Ap^r) and *B. sutbitlis* (Tc^r) shuttle vector with subE natural promoter and gene	[94]
pPCRscipt_aprE	vector (Ap^r) with synthetic subE gene (inner Nde I restriction site eliminated)	Synthetic
pMA-aprE(M14)	vector (Ap^r) with synthetic subE gene with all 14 mutations	Synthetic
pHY300PLK-aprE(epPCR1)	selected vectors with subE gene mutant after 1st round of epPCR	This work
pHY300PLK-aprE(epPCR2)	selected vectors with subE gene mutant after 2nd round of epPCR	This work
pHY300PLK-aprE(SeSaM1)	selected vectors with subE gene mutant after 1st round of SeSaM	This work
pHY300PLK-aprE(SSM1)	SSM of subE at position 1	This work
pHY300PLK-aprE(SSM62)	SSM of subE at position 62	This work
pHY300PLK-aprE(SSM97)	SSM of subE at position 97	This work
pHY300PLK-aprE(SSM103)	SSM of subE at position 103	This work
pHY300PLK-aprE(SSM149)	SSM of subE at position 149	This work
pHY300PLK-aprE(SSM153)	SSM of subE at position 153	This work
pHY300PLK-aprE(SSM166)	SSM of subE at position 166	This work
pHY300PLK-aprE(SSM181)	SSM of subE at position 181	This work
pHY300PLK-aprE(SSM184)	SSM of subE at position 184	This work
pHY300PLK-aprE(SSM205)	SSM of subE at position 205	This work
pHY300PLK-aprE(SSM218)	SSM of subE at position 218	This work

pHY300PLK-aprE(SSM224)	SSM of subE at position 224	This work
pHY300PLK-aprE(SSM232)	SSM of subE at position 232	This work
pHY300PLK-aprE(SSM240)	SSM of subE at position 240	This work
pHY300PLK-aprE(A153V)	subE A153V mutant	This work
pHY300PLK-aprE(I205V)	subE I205V mutant	This work
pHY300PLK-aprE(T240S)	subE T240S mutant	This work
pHY300PLK-aprE(M2)	subE mutant M2 (S62I/G166M)	This work
pHY300PLK-aprE(M4)	subE mutant SeSaM1-5 (M4, S62I/A153V/G166S/I205V)	This work
pHY300PLK-aprE(M5)	subE mutant M5 (S62I/A153V/G166S/T224A/T240S)	This work
pHY300PLK-aprE(M6)	subE mutant M6 (S62I/A153V/G166S/I205V/N218S/T224A)	This work

* This donated plasmid contains mutations in subtilisin E gene region.

2.1.4 Oligonucleotides

The primers used in this work are listed in Table 2.3.

Table 2.3 Summary of oligonucleotides used in this work

Name	Sequence (5' to 3')	Target
ZW_LIC_gFP	A*T*C*C*G*A*G*C*G*T*T*G*CATATGTGGAAGAAGATCATATTG CACATGAA (Nde I)	*aprE*
ZW_LIC_gRP	C*T*G*C*A*G*G*T*C*G*A*C*GGATCCTTATTGTGCAGCTGCTT GTACGTTGAT (BamH I)	*aprE*
ZW_LIC_pFP	C*A*A*C*G*C*T*C*G*G*A*T*CTTTTTTCAATTCTTTTACAGCT	pHY300P LK
ZW_LIC_pRP	G*T*C*G*A*C*C*T*G*C*A*G*ATCTCTAGAAGCTTGGGCAAAGC	pHY300P LK
SeSaM_fwd_ZW_gF	CACACTACCGCACTCCGTCGATCCGAGCGTTGCATATGTGGA AGAAG (Nde I)	*aprE*
SeSaM_rev_ZW_gR	GTGTGATGGCGTGAGGCAGCCTGCAGGTCGACGGATCCTTAT TGTG (BamH I)	*aprE*
SeSaM_up_ZW_gF	CGCCTGTCACATCCGAGCGTTGCATATGTGGAAGAAG (Nde I)	*aprE*
SeSaM_dn_ZW_gR	GCGGACAGTGCTGCAGGTCGACGGATCCTTATTGTG (BamH I)	*aprE*

ZW_FW_SSM1	5’-CCATACCAGGACGGCNNKTCTCACGGTACGCAT-3’	A1
ZW_RV_SSM1	5’-ATGCGTACCGTGAGAMNNGCCGTCCTGGTATGG-3’	A1
ZW_FW_SSM62	5’-CCATACCAGGACGGCNNKTCTCACGGTACGCAT-3’	S62
ZW_RV_SSM62	5’-ATGCGTACCGTGAGAMNNGCCGTCCTGGTATGG-3	S62
ZW_FW_SSM97	5’-GCAGTAAAAGTGCTTNNKTCAACAGGAAGCGGC-3’	D97
ZW_RV_SSM97	5’-GCCGCTTCCTGTTGAMNNAAGCACTTTTACTGC-3’	D97
ZW_FW_SSM103	5’-TCAACAGGAAGCGGCNNKTATAGCTGGATTATT-3’	Q103
ZW_RV_SSM103	5’-AATAATCCAGCTATAMNNGCCGCTTCCTGTTGA-3’	Q103
ZW_FW_SSM149	5’- TCCAGCGGTATCGTCNNKGCTGCCGCAGCCGGA -3’	V149
ZW_RV_SSM149	5’-TCCGGCTGCGGCAGCMNNGACGATACCGCTGGA-3’	V149
ZW_FW_SSM153	5’- GTCGTTGCTGCCGCANNKGGAAACGAAGGTTCA -3’	A153
ZW_RV_SSM153	5’- TGAACCTTCGTTTCCMNNTGCGGCAGCAACGAC -3’	A153
ZW_FW_SSM166	5’-AGCACAAGCACAGTCNNKTACCCTGCAAAATAT-3’	G166
ZW_RV_SSM166	5’-ATATTTTGCAGGGTAMNNGACTGTGCTTGTGCT-3’	G166
ZW_FW_SSM181	5’-GCAGTAGGTGCGGTANNKAGCAGCAACCAAAGA-3’	N181
ZW_RV_SSM181	5’-TCTTTGGTTGCTGCTMNNTACCGCACCTACTGC-3’	N181
ZW_FW_SSM184	5’-GCGGTAAACAGCAGCNNKCAAAGAGCTTCATTC-3’	N184
ZW_RV_SSM184	5’-GAATGAAGCTCTTTGMNNGCTGCTGTTTACCGC-3	N184
ZW_FW_SSM205	5’-GCTCCTGGCGTGTCCNNKCAAAGCACACTTCCT-3’	I205
ZW_RV_SSM205	5’-AGGAAGTGTGCTTTGMNNGGACACGCCAGGAGC-3’	I205
ZW_FW_SSM218	5’- ACTTACGGCGCTTATNNKGGAACGTCCATGGCG-3’	N218
ZW_RV_SSM218	5’- CGCCATGGACGTTCCMNNATAAGCGCCGTAAGT-3’	N218
ZW_FW_SSM224	5’- GGAACGTCCATGGCGNNKCCTCACGTTGCCGGA-3’	T224
ZW_RV_SSM224	5’- TCCGGCAACGTGAGGMNNCGCCATGGACGTTCC-3’	T224
ZW_FW_SSM232	5’- GTTGCCGGAGCAGCANNKTTAATTCTTTCTAAG -3’	A232

ZW_RV_SSM232	5'- CTTAGAAAGAATTAAMNNTGCTGCTCCGGCAAC -3	A232
ZW_FW_SSM240	5'-CTTTCTAAGCACCCGNNKTGGACAAACGCGCAA-3'	T240
ZW_RV_SSM240	5'-TTGCGCGTTTGTCCAMNNCGGGTGCTTAGAAAG-3'	T240
ZW_FW_A153V	5'- GTCGTTGCTGCCGCA**GTC**GGAAACGAAGGTTCA -3'	A153V
ZW_RV_A153V	5'- TGAACCTTCGTTTCC**GAC**TGCGGCAGCAACGAC -3'	A153V
ZW_FW_G166M	5'-AGCACAAGCACAGTC**ATG**TACCCTGCAAAATAT-3'	G166M
ZW_RV_G166M	5'-ATATTTTGCAGGGTA**CAT**GACTGTGCTTGTGCT-3'	G166M
ZW_FW_I205V	5'-GCTCCTGGCGTGTCC**GTC**CAAAGCACACTTCCT-3'	I205V
ZW_RV_I205V	5'-AGGAAGTGTGCTTTG**GAC**GGACACGCCAGGAGC-3'	I205V
ZW_FW_S62I	5'-CCATACCAGGACGGC**ATC**TCTCACGGTACGCAT-3'	S62I
ZW_RV_S62I	5'-ATGCGTACCGTGAGA**GAT**GCCGTCCTGGTATGG-3'	S62I
ZW_FW_S85A	5'- GGTGTTCTGGGCGTT**GCG**CCAAGCGCATCATTA-3'	S85A
ZW_RV_S85A	5'-TAATGATGCGCTTGG**CGC**AACGCCCAGAACACC-3	S85A
ZW_FW_D97A	5'-GCAGTAAAAGTGCTT**GCT**TCAACAGGAAGCGGC-3'	D97A
ZW_RV_D97A	5'-GCCGCTTCCTGTTGA**AGC**AAGCACTTTTACTGC-3'	D97A
ZW_FW_T240S	5'-CTTTCTAAGCACCCG**AGT**TGGACAAACGCGCAA-3'	T240S
ZW_RV_T240S	5'-TTGCGCGTTTGTCCA**ACT**CGGGTGCTTAGAAAG-3'	T240S

* indicates phosphorothioate (PTO) attached. Underlined sequences indicate the restriction sites, and letters in bold indicate the codons targeted for site directed mutagenesis (SDM).

2.1.5 Media and culture cultivation

E. coli cells were routinely cultivated in Luria broth (LB) medium supplemented with appropriate antibiotics (Amp: 100 μg·mL^{-1}) using a shaking incubator (Multitron II, Infors GmbH, Einsbach, Germany) at 37°C, 250 rpm for 12~16 h, while *B. subtilis* cells were cultivated in the same medium supplemented with appropriate antibiotics (Tet: 15 μg·mL^{-1}) under the same conditions for 24 h (test tubes) or 36 h (flasks).

Active colonies which showed halos when grown on LB_{tet} agar plates with skim

milk (1 %) as substrate were transferred by toothpicks into 96-well microtiter plates (V-bottomed, polystyrene plates; Greiner Bio-One GmbH, Frickenhausen, Germany) containing LB medium (150 µL) supplemented with tetracycline (15 $\mu g \cdot mL^{-1}$ per well). After 16 h of cultivation in a Multitron II microtiter plate shaker (Infors GmbH, Einsbach, Germany; 37°C, 900 rpm, 70 % humidity), each 96-well microtiter plate was replicated (replicator tool; Enzy-Screen BV, Leiden, Netherland) into a 96-well microtiter plate containing the same medium, and cultivated in LB_{tet} medium in a Multitron II microtiter plate shaker (24 h, 37°C, 900 rpm, 70 % humidity; Infors). Subtilisin E and variants were constitutively expressed and the cell cultures from microtiter plates were centrifuged (3,220 g, 4°C, 20 min). The supernatants were directly used for screening.

2.2 Methods

2.2.1 General methods

2.2.1.1 PCR reactions

A thermal cycler (Mastercycler gradient; Eppendorf) and thin-walled PCR tubes (0.2 mL Multi-ultra tubes; Carl Roth, Germany) were used in all PCR reactions. The PCR volume was always 50 mL except otherwise stated; larger volumes were prepared as multiple of 50 mL PCRs. The amount of DNA in cloning experiments was quantified by using a NanoDrop photometer (Thermo Scientific, Wilmington, USA).

Colony PCR was performed with a slight modification of standard PCR using colony from agar plate as template. An additional 5 min as denaturing step was applied at beginning for colony PCR.

2.2.1.2 Cloning

All gene manipulation steps carried out according to molecular cloning book [95], while the phosphorothioate-based ligase-independent gene cloning (PLICing) method was applied for all gene cloning [96]. Preparation of *E. coli* competent cells and

transformation were carried out as standard heat shock transformation [97] and electroporation [98]. Plasmids were verified by PCR, restriction enzyme digestion and sequence analysis. Sequence results were analyzed by Vector NTI 9 (Invitrogen). Plasmid isolation kit, PCR purification kit and rolling circle amplification kit (REPLI-g Mini Kit) were purchased from Qiagen (Hilden, Germany).

2.2.1.3 TCA precipitation and SDS-PAGE analysis

TCA precipitation and SDS-PAGE analysis of proteins were carried out according to described standard protocol [95].

2.2.1.4 Protein quantification

Protein concentration and purity were determined using automated electrophoresis by an Agilent 2100 Bioanalyzer (Agilent Protein 230 Kit; Agilent Technologies, Bölingen, Germany) or Experion microfluidic system (Bio-Rad, CA) and the BCA protein assay kit (no. 23225; Pierce Biotechnology, Rockford, IL) according to the manual's instruction. In addition BSA (200 $\mu g \cdot mL^{-1}$) was used as internal standard.

2.2.2 Construction of expression vector and strains

The subtilisin E gene (*aprE*) was ordered as a synthetic gene (Geneart, Regensburg, Germany), cloned into the commercial shuttle vector pHY300PLK (Takara Bio Inc, Japan) with restriction site EcoR I and BamH I, yielding pHY-*aprE* and transformed into *E. coli* DH5α for storage. The constructed vector pHY-*aprE* was transformed after replication (REPLI-g Mini Kit; Qiagen, Hilden, Germany) into *B. subtilis* WB600 as a host strain for expression.

2.2.3 Generation of random mutagenesis libraries

2.2.3.1 epPCR library

The Nde I and BamH I DNA fragment of pHY-*aprE* encodes the mature subtilisin E and was selected as the template for random mutagenesis. PLICing method was applied for all cloning procedures [96]. Two oligonucleotides, ZW_LIC_gFP and ZW_LIC_gRP (see 2.1.4) were used as forward and reverse primers to amplify the mature subtilisin E. The error prone PCR (epPCR) libraries were generated by using the template (pHY-*aprE*, 50 ng), dNTP mix (0.2 mM), dGTP (0.2 mM), $MnCl_2$ (0.05-0.5 mM), Taq DNA polymerase (5 U), forward and reverse primers (20 pmol) under the following PCR conditions: 94^{o}C for 2 min 1 cycle; 94 ^{o}C for 30 sec/60 ^{o}C for 1 min/72 ^{o}C for 1 min 29 cycles; 72 ^{o}C for 5 min 1 cycle.

The vector fragment pHY-*aprE*$^{-}$ was amplified by using a template (pHY-*aprE*, 50 ng), dNTP mix (0.2 mM), Betaine (1 M), Taq/Pfu DNA polymerase mix (4.5 U/0.072U) and primers ZW_LIC_pFP and ZW_LIC_pRP (see 2.1.4) under the following PCR conditions: 94^{o}C for 2 min 1 cycle; 94^{o}C for 30 sec/60^{o}C for 30 sec/68^{o}C for 9 min 35 cycles; 68^{o}C for 10 min 1 cycle. The resulting PCR products of pHY-*aprE*$^{-}$ were purified using a Qiagen PCR purification kit. The purified pHY-*aprE*$^{-}$ and epPCR products were digested by Dpn I (40 U; New England Biolabs) at 37^{o}C overnight, followed by further DNA purification (Qiagen quick PCR purification kit). Both fragments (0.1 pmol pHY-*aprE*$^{-}$ and 0.3 pmol epPCR products) were further digested by iodine mixture respectively at 70^{o}C for 10 min, and then mixed at room temperature for 30 min for ligation. The ligation mixture was afterwards transformed into *E. coli* DH5α and plated on LB_{Amp} agar plates. Thousands of *E. coli* DH5α variants were harvested by resuspending in double-distilled water and subsequent plasmid isolation. The resultant plasmid pool was then transformed into *B. subtilis* WB600. Ten randomly picked and sequenced variants of epPCR libraries confirmed within statistical errors in the expected mutational bias.

2.2.3.2 SeSaM library

The SeSaM [33] library was generated with the improved variants selected after two rounds of epPCR library screening as template. The generation of the SeSaM library was carried out according to the Handbook (SeSaM-Tv Classic version 0.16) provided by SeSaM-Biotech GmbH. PLICing method was again used for cloning. The PCR was carried out by using a SeSaM library as template (50 ng), dNTP mix (0.2 mM), Pfu DNA polymerase (2.5 U), together with forward and reverse primers (ZW_LIC_gFP and ZW_LIC_gRP, 20 pmol) under PCR conditions: 98^{o}C for 30 sec 1 cycle; 98^{o}C for 5 sec/60^{o}C for 15 sec/72^{o}C for 15 sec 15 cycles; 72^{o}C for 3 min 1 cycle. The further cloning was carried out according to the same procedure as previously described in this project.

2.2.4 Site-directed and saturation mutagenesis

Site directed mutagenesis (SDM) and saturation mutagenesis (SSM) of the subtilisin E wild type was employed to elucidate the role of each substitution identified in selected variants. Recombination of substitutions was performed based on different parents by SDM as well. In all focused mutagenesis experiments, a modified QuikChange method (Stratagene/Agilent Technologies) was utilized [99].

2.2.5 SeSaM-based shuffling

The SeSaM-based shuffling libraries were generated with both subtilsin E wild type and synthetic gene M14 (with 14 amino acid substitutions, sequence see Appendix I) as template. The generation of the SeSaM-based shuffling library was carried out according to the Handbook (SeSaM-Tv Classic version 0.21) provided by SeSaM-Biotech GmbH. However, Step 3 used two template ratios (WT/M14) of 1:1 and 1:2, and different PCR cycles (15, 30 and 45 cycles) were also applied for optimization of conditions for library generation. PLICing method was again used for cloning in the final step as described in section 2.2.3.

2.2.6 Agar plate assay for detection of protease activity

LB agar plates supplemented with appropriate antibiotics and 1 % Skim milk were used and protease activities were detected by halo formation.

2.2.7 MTP screening assays for activity and resistance towards chaotropes

Fluorimetric microtiter plate assay: The assay was carried out in 96-well microtiter plates (black flat bottomed, polystyrene plates; Greiner Bio-One GmbH, Frickenhausen, Germany) with Suc-AAPF-AMC in Tris-HCl buffer (0.1 M, pH 8.5, 37^{o}C) in the presence or absence of GdmCl (1 M, final concentration) or SDS (0.5 %, final concentration). The cell-free supernatants (10 µL) of the subtilisin E variants were supplemented in Tris-HCl buffer (40 µL, 0.1 M, pH 8.5), GdmCl solution (40 µL, 2.5 M GdmCl in 0.1 M Tris-HCl buffer, pH 8.5) and SDS solution (40 µL, 1.25 % SDS in 0.1 M Tris-HCl buffer, pH 8.5) respectively, incubated (25^{o}C, 30 min), and activities were determined by supplementing the substrate solution (50 µL, 200 µM Suc-AAPF-AMC in 0.1 M Tris-HCl buffer; pH 8.5). The final reaction volume was always 100 µL. Calcium was not supplemented to SDS reaction assays despite its stabilizing function [100], since it causes precipitation. The activity in the presence or absence of chaotropes was measured at 37°C by monitoring the increase of fluorescence intensity within 20 min using the Tecan SAFIRE (Excitation: 350 nm / Emission: 450 nm, Tecan Austria GmbH, Salzburg, Austria). One unit of enzymatic activity was defined as the amount of the enzyme that produced 1 pmol of 7-amino-4-methylcoumarin (AMC) per min. The specific activity was defined as the protease activity per milligram of protein.

Skim-milk microtiter plate assay: The assay was carried out in 96-well microtiter plates (flat-bottomed, polystyrene plates; Greiner Bio-One GmbH, Frickenhausen, Germany). The screening procedure was performed by applying 10 µL cell-free

supernatants of subtilisin E variants in 90 µL Tris-HCl buffer (0.1 M, pH 8.5, 37^{o}C), and activities were determined by supplementing the substrate solution (100 µL, 2.0 % Skim-milk in 0.1 M Tris-HCl buffer, pH 8.5). The final reaction volume is 200 µL. The activity was measured at 37^{o}C by monitoring the increase of absorbance at 580 nm within 15 min using Tecan Sunrise (Tecan Austria GmbH, Salzburg, Austria).

2.2.8 Purification

Shaking flask (1 L) containing LB medium (150 mL) supplemented with tetracycline (15 µg·mL^{-1}) were inoculated with 1:150 dilution of an overnight culture (*B. subtilis* WB600 harbouring pHY-*aprE*) grown in the same medium. After 36 h of expression, *B. subtilis* WB600 cells were harvested by centrifugation (Eppendorf 5810R, 4^{o}C, 3,220 g, 30 min) and the culture supernatant was cleared by filtration (0.2 µm Minisart sterile syringe filter; Sartorius, Goettingen, Germany). The filtered solution was subsequently concentrated through a 5 kDa cut-off membrane (Millipore, Billerica, MA, USA) by using a stirred Amicon ultrafitration cell (Millipore). Then two different methods of purification were applied for different variants:

Method 1: The concentrated supernatants were transferred into a Tris-HCl buffer (0.1 M; pH 7.5; 20 mL). Concentrated solutions of wild type subtilisin E and variants S62I, G166M, M2 (S62I/G166M), SeSaM1-5 (M4) were subsequently purified by ion exchange column chromatography: 1) Subtilisin E variants were loaded onto a SuperQ anion-exchange column (25 mL) equilibrated with the same buffer. Subtilisin E was eluted by linearly increasing the NaCl concentration from 0 to 1.0 M in the same buffer at a rate of 2 mL·min^{-1} (5 column volumes in total). 2) The obtained protein solution was loaded onto a SP650c cation-exchange column (25 mL) which was equilibrated with the same buffer. Subtilisin E was remained in the flow through while other proteins were bound. The SP650c cation-exchange column was regenerated via a NaCl gradient elution from 0 to 2.0 M (5 column volumes in total) in the same buffer at a rate of 3 mL· min^{-1}. Purified subtilisin E variants were

subsequently concentrated in an Amicon ultra centrifugal filter device (Millipore) with a 10 kDa cut-off membrane.

Method 2: The concentrated supernatants were transferred into a sodium phosphate buffer (10 mM; pH 6.2; 15 mL). Concentrated solutions of wild type subtilisin E and variants N218S, T224A, SeSaM1-5 (M4) were subsequently purified by ion exchange column chromatography: 1) Subtilisin E variants were loaded onto a CM cation-exchange column (25 mL) equilibrated with the same buffer. Subtilisin E was eluted by linearly increasing the NaCl concentration from 0 to 0.2 M in the same buffer at a rate of 2 $mL \cdot min^{-1}$ (5 column volumes in total). 2) The obtained protein solution was collected and dialysed in 1 L Tris-HCl buffer (0.05 M; pH 8.5) at $4^{o}C$ overnight. The dialysed solutions were then loaded onto a Super-Q anion-exchange column (25 mL) which was equilibrated with the same Tris-HCl buffer. Subtilisin E was remained in the flow through while other proteins were bound. The Super-Q anion-exchange column was regenerated via a NaCl gradient elution from 0 to 2.0 M (5 column volumes in total) in the same buffer at a rate of 2 $mL \cdot min^{-1}$. Purified subtilisin E variants were subsequently concentrated in an Amicon ultra centrifugal filter device (Millipore) with a 10 kDa cut-off membrane. Variant M5 and M6, due to their poor binding to CM column, were purified through SuperQ column twice instead.

2.2.9 Determination of kinetic constants

Measurements were carried out in 96-well microtiter plates (flat-bottomed, polystyrene plates; Greiner Bio-One GmbH, Frickenhausen, Germany) by employing 50 nM purified protease solution in the presence of 1 M, 2 M GdmCl, 0.5 % SDS or absence of chaotropes in Tris-HCl buffer (pH 8.5; 0.1 M) and varied Suc-AAPF-pNA concentrations (0.05 - 10.0 mM). The amount of p-nitroaniline (pNA) released from Suc-AAPF-pNA was determined by absorbance measurements (405 nm). One unit of protease activity was defined as the amount of protease that produced 1 µmol of p-nitroaniline per min. The specific activity was defined as the protease activity per milligram of protein. For kinetic analysis, k_{cat} and K_m values were determined from

initial velocity data measured as a function of substrate concentration with a linear correlation coefficient of >0.99.

2.2.10 Inactivation curve determination

The purified subtilisin E wild type and its variants (50 nM) were incubated at room temperature (25^{o}C) for 30 min in the presence of various concentrations of GdmCl (0 - 6.0 M) or SDS (0 - 4.0 %) in Tris-HCl buffer (pH 8.5; 0.1 M). The activities of purified subtilisin E wild type and its variants were determined by addition of Suc-AAPF-pNA (1.0 mM) as described in section 2.2.9. Relative activities were calculated as percentage of activities in the absence of chaotropes which was set as 100 %. Half life of inactivation of subtilisin E and its variants, designated as IC_{50}, was defined as the chaotrope concentration at which the activity of subtilisin E remained 50 % under conditions mentioned previously.

2.2.11 Stability in the presence of chaotropes

The purified subtilisin E wild type and its variants (50 nM) were incubated at room temperature (25^{o}C) in the presence of given concentrations of GdmCl or SDS. At different time points (0, 0.25, 0.5, 1, 2, 4, 6 and 9 h), the activities of purified subtilisin E wild type and its variants were determined using Suc-AAPF-pNA as described in section 2.2.9. Relative activities were calculated as percentage of initial activities in the presence and absence of chaotropes. The activity in the absence of a chaotrope was set as 100 %. Half life of stability of subtilisin E and its variants in terms of time, designated as $t_{1/2}$, was defined as the time point at which the activity of subtilisin E remained 50 % under conditions mentioned previously (see section 2.2.9).

2.2.12 Activity towards complex protein substrate

The enzymatic activities of purified subtilisin E and its variants towards complex protein substrate were determined by using Azocasein (Sigma) and an Eppendorf

Thermomixer comfort (Eppendorf AG, Hamburg, Germany; 1.5 ml Eppendorf tubes, 37^{o}C, 500 rpm). A modified method from Pulido et al. [101] was applied. The reaction mixture (300 µL) contained Tris-HCl (pH 8.5; 0.1 M) and 2.0 % Azocasein. The enzymatic reaction was initiated by adding the purified enzyme (0.5 µM) and terminated by adding trichloroacetic acid (200 µL; 15 %). The reaction time was usually 30 min. After centrifugation (15,000 g, 15 min), an aliquot of supernatant (80 µL) was withdrawn, mixed with NaOH (20 µL; 2 M), and measured for absorbance at 440 nm. One unit of enzymatic activity was defined as the amount of enzyme that increased the A_{440} value of the assay reaction mixture by 0.1 in 1 min.

2.2.13 Circular dichroism (CD) analysis

The total sample volume was 36 µL in each measurement using 0.1 mm cuvette. 24 µL of the sample was prepared by the respective amounts of Tris-HCl buffer (pH 8.5; 0.1 M) and chaotrope solutions (1.5 M, 3 M, 4.5 M, 7.5 M GdmCl, 0.75 %, 1.5 %, 3.0 % SDS in 0.1 M Tris-HCl buffer, pH 8.5) in 0.5 mL PCR tubes. 12 µL of appropriate concentrated purified subtilisin E and its variants (final concentration 200-300 µg· mL^{-1}; purified by *Method 2* in 2.2.7) were pipetted into the same PCR tube with pre-prepared buffer, GdmCl (final concentration of 1 M, 2 M, 3 M and 5 M) or SDS (final concentration of 0.5 %, 1.0 % and 2.0 %). The mixture of enzymes and chaotrope solutions were gently mixed by pipetting, and immediately analyzed at room temperature with an Olis SDM 17 CD (Olis, Bogart, USA).

For each sample, five scans from 190 to 240 nm were recorded and averaged. 210 nm was selected as a benchmarking wavelength for GdmCl since GdmCl, especially at high concentration, has an increased absorbance below 210 nm leading to high noise/signal ratios. The latter interferes with deconvolution of recorded CD spectra. The bandwidth was adjusted to 2 nm, the step width to 1 nm. As baseline, buffer and various concentrations of chaotrope solutions without subtilisin E (use buffer instead) were employed in all experiments to avoid deviations caused by different cosolvents. CD spectra were smoothed by the Savitzky-Golay filter (Olis

Global Works software package), and results were given as milli degrees and converted into mean residue ellipticity.

2.2.14 Computational analyses

2.2.14.1 Model building

A crystal structure of the Pro-subtilisin E complex is available at 0.2 nm resolution (Protein Data Bank accession code: 1SCJ) [12]. In order to generate the structure of subtilisin E–Suc-AAPF-pNA complex, the pro-peptide domain of crystal structure subtilisin E was removed, except the four amino acids (Ala-His-Glu-Tyr) chain at C-terminal which binds with active site residues as enzyme-substrate complex. The four amino acids chain was then mutated computationally to the sequence Ala-Ala-Pro-Phe (AAPF) and attached to both termini using two functional groups n-succinyl and p-nitroanilide to form the substrate Suc-AAPF-pNA. The favourable rotamer was selected in all cases. Finally, the active site Cys 221 of subtilisin E was mutated back to the catalytic Ser residue. All amino acid exchanges were performed using the *Model Build* tool of the program Fold X [102]. Two calcium ions (named Ca1, Ca2) which were present in the crystal structure were retained for modelling. The final structures of the protein-substrate complex was subsequently energy minimized in vacuum using the GROMOS96 force field [103]. For models without substrate Suc-AAPF-pNA, the same building procedure was carried out except the complete removal of Pro-peptide domain.

2.2.14.2 Docking

AutoDock4 [104], is currently widely used for *in silico* protein-ligand docking. The programs Yasara [105] with Autodock4 were used to predict the docking structure of subtilisin E and GdmCl (PubChem database CID:32838) or SDS (PubChem database CID:4329331). The docking preparation procedure steps for the ligands were (1) converting the structure from sdf to pdb followed by energy minimization (2) adding

hydrogen atoms, and (3) determining the docking grid (5 Å extended around the center of selected active site and/or substrate binding pocket residues). All these steps could be accomplished within Yasara program.

2.2.14.3 MD simulations of melittin

Gromacs engine (Version 3.3.3 and 4.0.7) was used for all molecular dynamics (MD) simulations (see Table 2.4). In these MD simulations, a neutral periodic cubic system was created containing a single helical melittin molecule and a number of independent guanidinium and chloride ions surrounded by explicit water molecules. The melittin starting coordinates were taken from the Protein Data Bank (accession code 2MLT), removed C-terminal NH_2 group and used in conjunction with the GROMOS96 43a1 force field parameters. Water molecules were represented using SPC model. The parameters of the force field concerning the GdmCl are taken from Camilloni and coworkers. [74].

Pre-generation of simulation boxes of various concentrations of GdmCl were performed to facilitate the setup of different MD simulations. For GdmCl, 3 M and 5 M boxes (gro files, cubic box with volume of 2.2 nm^3) were directly taken from Camilloni and coworkers [74], while 1 M box was generated in the same manner using the same topology file (itp file). Arbitrary starting coordinates were generated by randomly placing and orienting one melittin molecule, 234 Gdm^+/239 Cl^- (5 M), or 175 Gdm^+/180 Cl^- (3 M), or 51 Gdm^+/56 Cl^- (1 M) in a 5 nm^3 cubic box. Here the total charge of the box (+5) was neutralized by replacing five water molecules at the most positive electrostatic potential with 5 Cl^- ions.

Position restriction equilibration was preformed for 100 ps with velocity assigned as 0.002 ps. The simulation was then run for total 10.0 ns with no velocity reassignment. The first 1.0 ns of this was taken as annealing equilibration to 298 K, and rest 9 ns was used for analysis. MD simulation in pure water was also performed for comparison.

2.2.14.4 MD simulations of subtilisin E

Gromacs engine (Version 3.3.3 and 4.0.7) was also used for all MD simulations on subtilisin E (see Table 2.4). The starting structure of the subtilisin E for MD simulations was prepared as described in 2.2.14.1. In this case, the Pro-peptide domain present in the crystal structure was completely removed, since it is cleaved in the maturation process. The GROMOS96 43a1 force field parameters were used for all simulations [103]. The water was modeled using the Simple Point Charge model [106]. The GdmCl model was taken from Camilloni and coworkers. [74]. For SDS model, The initial coordinates for the surfactant monomers were generated from ProDRG [107] and partial charges on the sulfate head group were adopted from Bruce and coworkers. [82b].

Pre-generated simulation boxes of various concentrations of GdmCl were also performed in MD simulations on subltisin E (see 2.2.14.3), as well as SDS. For SDS, 1 % and 5 % boxes (gro files, cubic box with volume of 8.0 nm^3) were generated using the topology file from ProDRG as mentioned above followed by energy minimization using the Groms96 45a1 force field.

The MD simulations of subtilisin E and its variants were carried out in water and various concentrations of GdmCl and SDS at 298 K (see Table 2.4). The simulation boxes were built by centering the subtilisin E molecule in a cubic box (8 nm^3). Subtilisin E was solvated by stacking an equilibrated box of 216 water molecules to completely fill the simulation box. All solvent molecules with any atom within 0.15 nm from solute atoms were removed. 5 M GdmCl solution, for instance, was generated by placing 1139 Gdm^+ molecules and 1141 Cl^- ions in the cubic box with one subtilisin E molecule. 5 % SDS solution, for instance was generated by placing 32 dodecylsulfate ions (DS^-) and 32 Na^+ ions in the cubic box with one subtilisin E molecule. The water molecules were filled in the remaining space of the simulation box. The total charge of the box (+2) was neutralized by replacing two water molecules at the most positive electrostatic potential with two Cl^- ions. Subtilisin E variant M4 in GdmCl and SDS were performed in the same manner.

Simulation protocol: Systems described in the previous paragraph were energy minimized using the steepest descent method for at least 500 steps to remove clashes caused by generated solvent molecules. Subtilisin E was finally studied at 298 K. All simulations were performed by keeping the temperature at the reference values by weak coupling (coupling constant τ=0.1 ps) with an external bath [108]. The pressure of the system was kept constant at 1 bar by using the Berendsen barostat equation with a coupling constant of τ_p= 0.5 ps. The LINCS algorithm was used to constrain all bond lengths [109]. For the water molecules, the SETTLE algorithm was used [110]. A dielectric permittivity, ε_r= 1, and a time step of 2 fs were used. The non-bonded interactions (electrostatic and Leonard-Jones) were calculated using the Particle Mesh Ewalds (PME) method [111]. For the calculation of the long-range interactions, a grid spacing of 0.12 nm combined with a fourth-order B-spline interpolation were used to compute the potential and forces between grid points. A non-bonded pair-list cut-off of 0.9 nm was used and the pair-list was updated every 5 time steps. The MD simulations were started by assigning to all atoms initial velocities obtained from a Maxwellian distribution at the given temperature. The box density was equilibrated for 100 ps using position restraints on the protein heavy atoms. After the equilibration, the position restraints on the protein were removed and the system was gradually heated from 50 K to the simulation temperature during 50 ps of simulation and then hence equilibrated for 5 ns. The following production runs were 20 or 45 ns long.

Analysis of trajectory: The protein structure along the two trajectories was monitored using standard analysis tools to assess the amount of deviations from the crystallographic structure and its mobility. The distribution of the guanidinium ions or dodecylsulfate ions around subtilisin E was analysed using the spatial distribution functions (SDF). The calculation of SDF was carried out on the last 25 ns of the simulations using a g_spatial program of the Gromacs package (4.0). The resulting spatial distribution function (SDF) was visualized using the program VMD [112].

Besides, essential dynamics analysis [113] was also performed to clarify the subtle changes of protein mobility throughout the trajectory.

Table 2.4 Overview of all MD simulations in this work

Index	Peptide or Protein	Solvent	Duration (ns)
1	Melittin	Water	10
2	Melittin	1 M GdmCl	10
3	Melittin	3 M GdmCl	10
4	Melittin	5 M GdmCl	25
5	Subtilisin E WT	Water	50
6	Subtilisin E WT	5 M GdmCl	50
7	Subtilisin E M4	Water	25
8	Subtilisin E M4	5 M GdmCl	25
9*	Subtilisin E WT	Water	25
10*	Subtilisin E WT	1 M GdmCl	25
11*	Subtilisin E WT	3 M GdmCl	25
12*	Subtilisin E WT	5 M GdmCl	25
13	Subtilisin E WT	1 % SDS	25
14	Subtilisin E WT	5 % SDS	25
15	Subtilisin E M4	5 % SDS	25

* These simulations were run after 25 ns long equilibration of subtilisin E WT in pure water.

3. Results

In this chapter, main results of this dissertation are presented, including three different subjects: 1) Directed evolution of subtilisin E for improved GdmCl and SDS tolerance; 2) Systematic analysis of effects of substitutions identified from directed evolution using site directed, saturation mutagenesis and recombination; 3) Computational analysis of subtilisin E in the presence of GdmCl or SDS by docking and molecular dynamics (MD) simulations. These results revealed that several highly improved variants were identified after three rounds of directed evolution with respect to their resistance towards both GdmCl and SDS. Interestingly, most beneficial substitutions were located around the active site or substrate binding pockets, which are beyond our expectation and indicated of some important features in protein denaturation. Besides, computational analysis provided detailed information to understand such phenomena and strengthened the hypothesis in the denaturation process of proteins induced by GdmCl and SDS.

3.1 Directed evolution of subtilisin E

3.1.1 Cloning and expression of subtilisin E in *B. subtilis*

For cloning and expression of the subtilisin E gene (*aprE)*, vectors pBE3* and pHY300PLK, A synthetic subtilisin E gene (including pre and pro sequence) were used. The mature domain sequence of *aprE* was amplified by PCR and replaced by a corresponding sequence region of pBE3* into pBE3 (subtilisin E wild type). Then the EcoR I and BamH I digested fragment from pBE3 together with its natural promoter was cleaved off and cloned into a comparable small commercial shuttle vector pHY300PLK to construct the final expression plasmid pHY-*aprE* (see Figure 3.1), and then transformed into *E. coli* DH5α for storage and *B. subtilis* WB600 for expression (see Figure 3.2 A). The proteolytic activities could be easily detected from halos on 1 % Skim milk agar plate (see Figure 3.2 B).

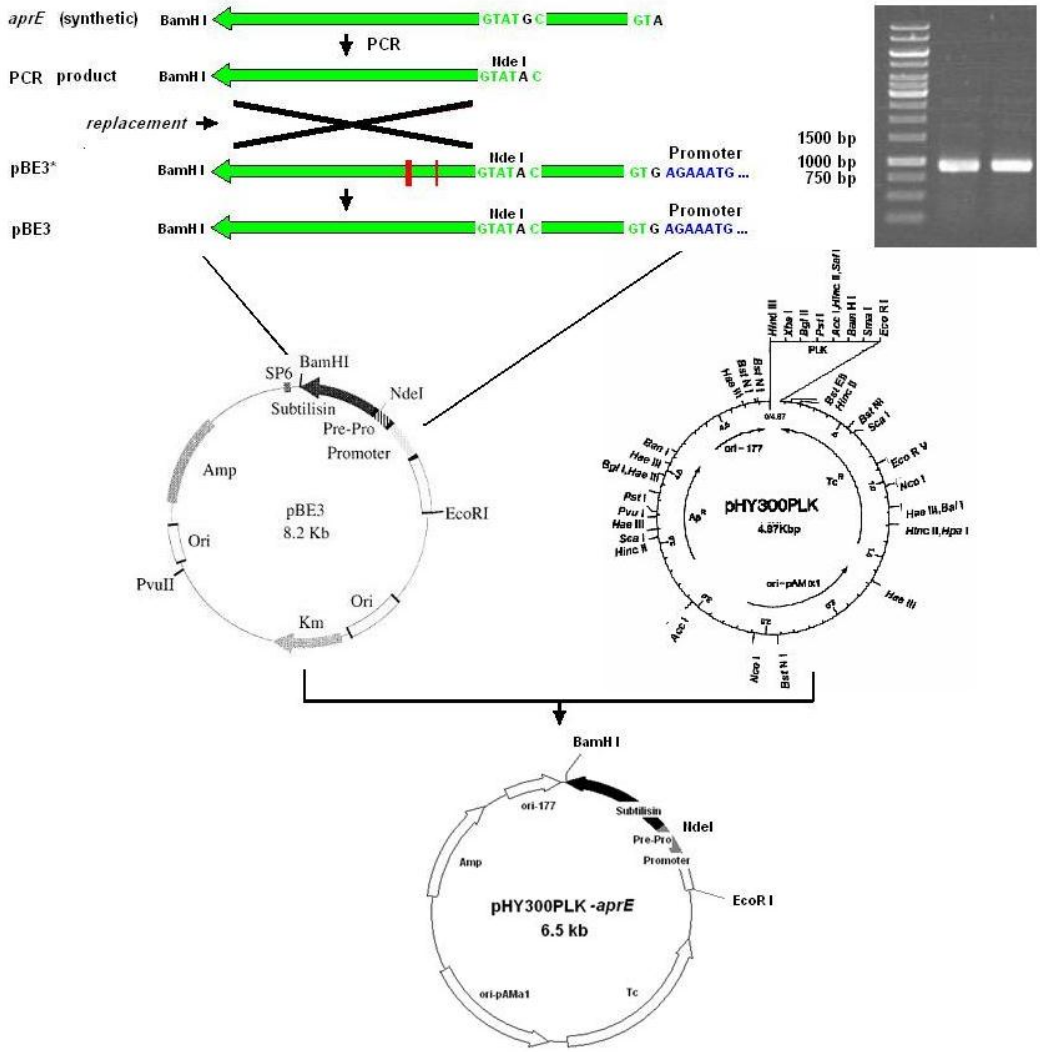

Figure 3.1 Construction of the expression plasmid pHY300PLK-*aprE*. A synthetic subtilsin E gene (*aprE*, wild type) was amplified by PCR and replaced Nde I and BamH I digested fragment of pBE3* to construct the correct pBE3. EcoR I and BamH I digested fragement of pBE3 including natural subtilisin E promoter was cleaved off and inserted into a commercial shuttle vector pHY300PLK to constructe the expression plasmid pHY300PLK-*aprE*.

3.1.2 Development of screening systems

Eight different substrates or methods were investigated for screening for chaotrope resistance (GdmCl or SDS) in microtiter plates (MTP). The comparison is summarized in Table 3.3. Two substrates, Skim-milk and Casein (F-C) (using Folin & Ciocalteu's Phenol Reagent to develop blue color) [114] are not suitable for chaotolerance screening due to cross-interactions. Standard casein assay, which doesn't have interaction problem, is still not recommended for further study, because of its less

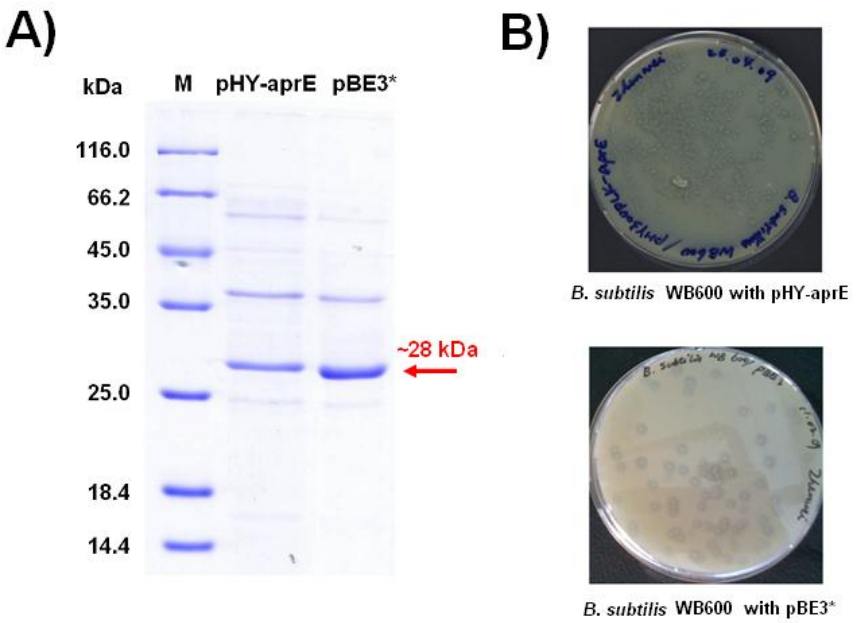

Figure 3.2 Expression of subtilisin E in *B. subtilis* WB600. A) SDS-PAGE analysis of expression of subtilisin E in *B. subtilis* WB600 harboring pHY-*aprE* or pBE3*. B) Detection of proteolytic activity of subtilisin E by visible halos on 1 % Skim milk agar plates.

sensitivity, tricky handling and high cost of materials (quartz MTPs are required). The fluorescent substrate Casein-FITC [115] is not suitable because of its high background. Another azurine-cross-linked substrate, Casein-AZCL (Megazyme International Ireland, Wicklow, Ireland), which could be an alternative agar-plate assay and has no interaction with denaturants, is not suitable for MTP screening, since it's difficult to keep a homogeneous solution in each well due to its insolubility in aqueous solutions. The assays using Suc-AAPF-pNA or Suc-AAPF-AMC, have a high sensitivity, low background and no cross-interaction with chaotropes, despite that they are specific for certain proteases. Azocasein gives a low sensitivity despite no cross-interaction with chaotropes.

Table 3.3 Comparison of different substrates or methods investigated for screening for tolerance towards chaotropes in microtiter plates

Substrate	Type	Advantage	Disadvantage
Skim-milk	Spectrophotometric	Very cheap, unspecific	Degraded by SDS or GdmCl
Casein	Spectrophotometric	Cheap, classic, unspecific	Less sensitive, tricky handling, quartz MTP required
Casein (F-C)	Colorimetric	Classic, unspecific	Interaction between F-C reagents and GdmCl
Casein-AZCL	Colorimetric	Sensitive, unspecific	Less homogeneous
Azocasein	Colorimetric	Unspecific	Less sensitive
Suc-AAPF-pNA	Colorimetric	Sensitive, standard for characterization	Expensive, specific
Suc-AAPF-AMC	Fluorescent	Very sensitive	Specific
Casein-FITC	Fluorescent	Sensitive, unspecific	High background

A screening strategy was finally proposed after consideration of all the Pros and Cons. Figure 3.4 shows the screening procedure from mutant library generation, pre-screening in agar plates, and microtiter plate based screening systems to identify subtilisin E variants with improved resistance (Suc-AAPF-AMC detection system) and preserved activity towards complex substrates (Skim-milk activity detection system).

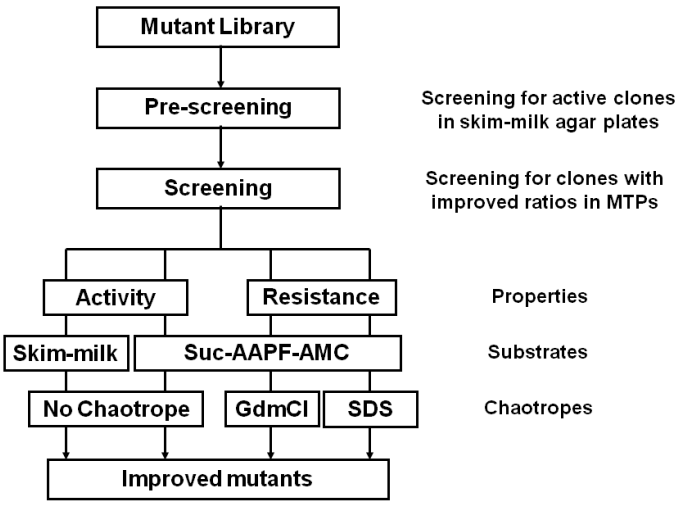

Figure 3.4 Screening procedure employed in each round of directed evolution. The screening procedure comprises a pre-screening on Skim-milk agar plates (1 % Skim milk) for active variant identification followed by two 96-well microtiter plate screening systems. The latter quantify GdmCl or SDS resistance of subtilisin E variants and investigate whether resistance improved subtilisin E variants still digest efficiently complex protein substrates.

The Skim milk agar plate assay was used for pre-screening and ruling out the inactive clones. Active clones were therefore selected and subjected to subsequent screening in 96-well MTPs in the absence and presence of GdmCl or SDS by using two substrates (see Figure 3.4). All screening conditions were optimized and discribed in details in Matericals and Methods (see section 2.2.7). Coefficient of variations (CV) in 96-well MTP based Suc-AAPF-AMC screening were determined for subtilisin E to be: 10.9 % (1 M GdmCl); 7.0 % (0.5 % SDS), and 6.5 % (no denaturant), and the coefficient of variations was determined to be 16.8 % for the Skim-milk detection system (see Figure 3.5), Coefficient of variations below 20 % in MTP has been commonly accepted for directed evolution experiments [116]. The resistance ratio (chaotrope: GdmCl or SDS) and activity ratio (substrate: Suc-AAPF-AMC or Skim-milk) were calculated as follows to identify improved subtilisin E variants:

$$\text{Resistance Ratio (chaotrope)} = (A_{(+)\text{chaotrope}} / A_{(-)\text{chaotrope}})_{\text{Variant}} / (A_{(+)\text{chaotrope}} / A_{(-)\text{chaotrope}})_{WT} \quad (1)$$

$$\text{Activity Ratio (substrate)} = A_{\text{Variant}} / A_{WT} \quad (2)$$

In Ratio (1), the activity A in the presence of chaotropes (marked as “+”) on Suc-AAPF-AMC as substrate was divided by the one in the absence (marked as “-“) to “eliminate” expression mutants. Ratio (1) shows improvements compared to the wild type (WT). Ratio (2) measures the ability of subtilisin E variants to still convert substrates, especially complex protein, efficiently. The latter is an important prerequisite for application in diagnostic kits. Both ratios were taken into account for selecting beneficial subtilisin E mutants in three rounds of directed evolution.

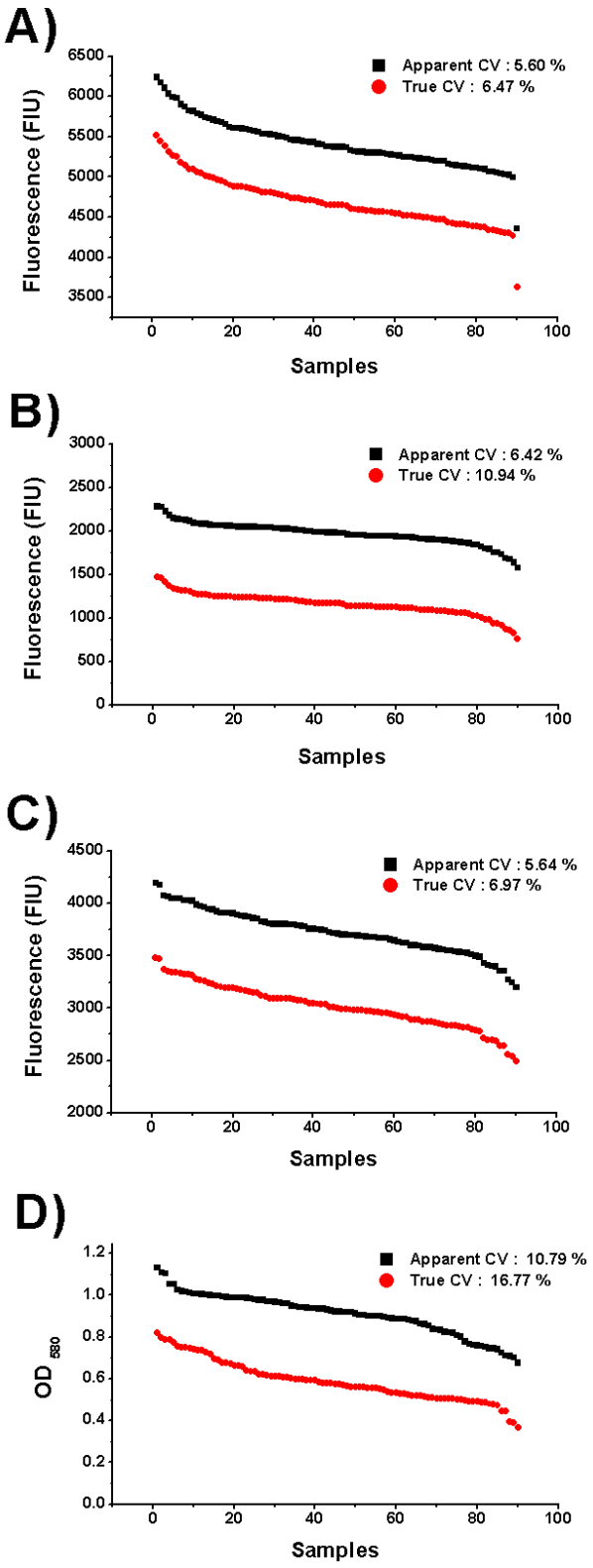

Figure 3.5 Evaluation of coefficient of variations (CV) in 96 well-MTP based screening assays. A) Suc-AAPF-AMC assay in the absence of chaotropes. B) Suc-AAPF-AMC assay in 1 M GdmCl. C) Suc-AAPF-AMC assay in 0.5 % SDS. D) Skim milk assay in the absence of chotropes.

3.1.3 Construction of random mutagenesis libraries

3.1.3.1 epPCR libraries

Three iterative rounds of directed evolution were carried out in which first two epPCR mutant libraries were screened by sampling per library ~1300 clones. Various concentrations of $MnCl_2$ coupled with unbalanced dNTP (dGTP>> dATP = dCTP = dTTP) were investigated for generating epPCR libraries (see Figure 3.6). Conditions mentioned above were selected after mutagenesis assistant program (MAP) analysis [24] to maximize charged substitutions. Charged substitutions were regarded as beneficial ones with the assumption that more charged residues on the surface improve interactions with cosolvents. A concentration of 0.10 mM $MnCl_2$ coupled with unbalanced dNTP (0.4 mM dGTP, 0.2 mM dATP = dCTP = dTTP), which generated ~50 % active mutants was selected finally for generation of both epPCR libraries (see Table 3.7).

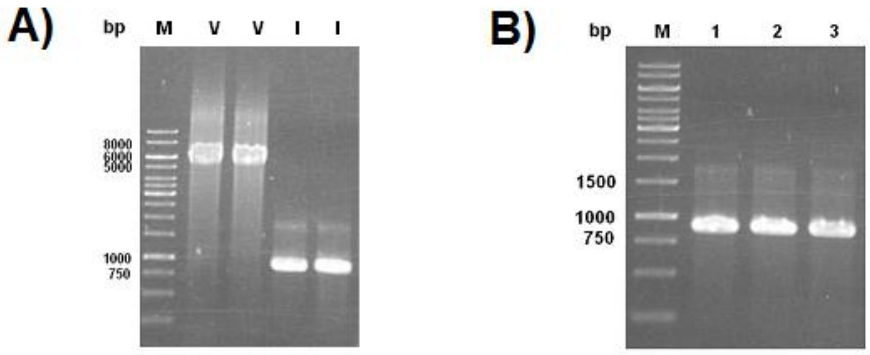

Figure 3.6 Gel electrophoresis analysis of epPCR products. A) PCR products of amplified vector and insert by PLICing methods [96]. M, DNA marker; V, vector (pHY-*aprE*$^-$); I, *aprE* (WT) insert. B) epPCR products of insertion by PLICing with addition of different concentrations of $MnCl_2$ and unbalanced dNTP (0.4 mM G, 0.2 mM A=C=T). M, DNA marker; Lane 1, 0.1 mM $MnCl_2$; Lane 2, 0.2 mM $MnCl_2$; Lane 3, 0.3 mM $MnCl_2$.

Table 3.7 Optimization of concentrations of $MnCl_2$ for generation of epPCR libraries

Conditions	Active Ratio (%)
Control[1]	~100
0.1 mM $MnCl_2$, G>>[2]	50~60
0.2 mM $MnCl_2$, G>>	15~20
0.3 mM $MnCl_2$, G>>	5~10

1: Control indicates the conditions without $MnCl_2$ and/or unbalanced dNTP.

2: G>> indicates 0.4 mM G, 0.2 mM A=C=T as unbalanced dNTP condition.

3.1.3.2 SeSaM library

Eight improved variants after screening of two iterative rounds of epPCR libraries were selected and mixed as the template for generation of the next round of SeSaM [33] library. The protocol was derived from SeSaM Tv-Handbook (classic) Revision 0.16 provided by SeSaM-Biotech GmbH.

Optimal phosphorothioate percentage was determined experimentally (see Figure 3.8): A_fwd library (dATPαS 22 %), G_fwd library (dGTPαS 18 %), A_rev library (dATPαS 32 %), G_rev library (dGTPαS 22 %).

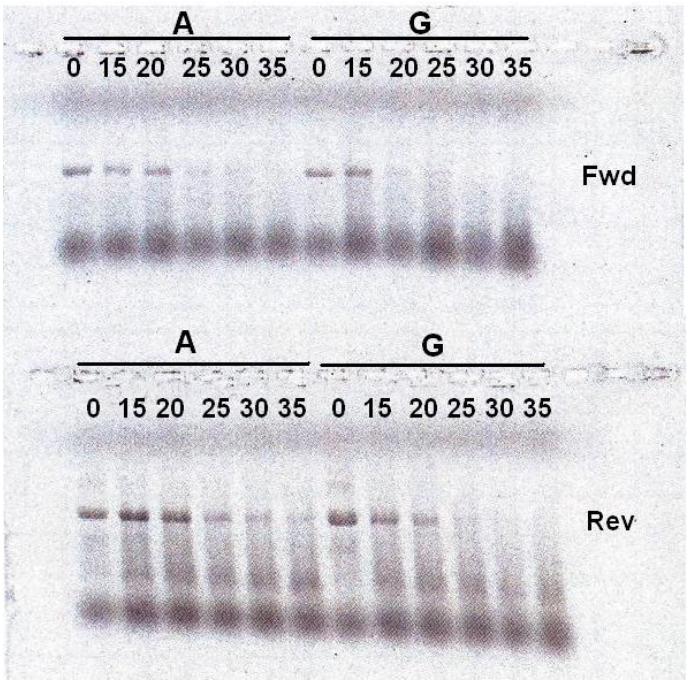

Figure 3.8 Fluorescent scanning of PCR products at different percentage of phosphorothioate nucleotides. A or G indicates library type; the number indicates the percentage (%) of phosphorothioate nucleotides.

The rest of protocols were performed according to the Handbook. The final libraries were mixed and amplified by the PLICing method (see Figure 3.9). The PCR products were further digested, ligased and transformed into *E. coli* DH5α and then into the host strain *B. subtilis* WB600. The active was ~33 % of variants for this round of SeSaM Library, and ~1300 clones were screened.

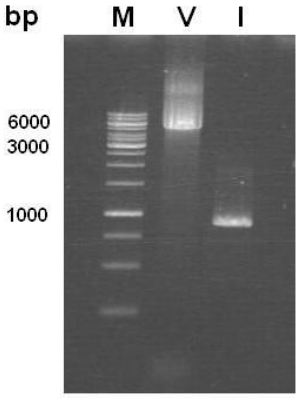

Figure 3.9 Gel electrophoresis analysis of PCR products for generation of SeSaM libraries. M, DNA marker. V, vector (pHY-*aprE*$^-$); I, *aprE* (SeSaM libraries) insert.

3.1.4 Screening of random mutagenesis libraries

Microtiter plate screening in the presence and absence of 1 M GdmCl or 0.5 % SDS and a Skim-milk activity screening were performed in parallel. Variants with a ratio (2) >0.3 (Skim-milk assay) were further screened with Suc-AAPF-AMC assays for improved resistance and variants with a ratio (1) >1.2 (1st round epPCR), >1.4 (2nd round epPCR), >2.5 (SeSaM) were used for diversity generation in the next round of directed evolution.

3.1.4.1 1st round of screening – epPCR1

In the first round, an epPCR library was screened (see section 3.1.3). MTP assays by using Suc-AAPF-AMC as substrate in the presence and absence of 1 M GdmCl or 0.5 % SDS, as well as Skim-milk assay were performed in parallel.

In all, 15 MTPs, including all 1350 active mutants were screened. Mutants, which showed no less than 30 % of residual activity towards Skim-milk and above 20 % of

residual activity enrichment either in 1 M GdmCl or 0.5 % SDS, were selected for rescreening to confirm their improvement.

Finally, 90 mutants of interest were selected and re-screened. 12 mutants of interest were thus selected and characterized (see Table 3.10). Their plasmids were isolated and mixed as the template for the next round of directed evolution.

Table 3.10 Overview of properties of selected variants from 1st round of epPCR library. The chaotolerance ratios (1 M GdmCl or 0.5 % SDS) or activity ratios (AAPF or Skim-milk) are calculated according to equations in section 3.1.2. Expression level of selected variants is also compared to the wild type as indicated.

Variants	1	2	3	4	5	6	7	8	9	10	11	12
1 M GdmCl	1.2	1.4	1.5	1.4	1.2	1.5	1.1	1.4	0.9	1.2	1.5	0.5
0.5 % SDS	1.1	1.2	1.2	1.2	1.0	1.2	1.2	1.2	0.9	1.1	1.2	1.4
AAPF	2.0	2.0	1.5	2.0	2.1	1.6	1.0	1.0	2.2	2.1	1.5	0.6
Skim-milk	2.2	0.7	1.0	0.9	2.1	1.0	0.7	1.0	1.3	2.0	1.1	0.8
Expression	↑	=	↓	↑	=	=	=	↓	↑	↑	=	=

3.1.4.2 2nd round of screening – epPCR2

The screening of second round of epPCR library (1260 clones) was carried out in the same way as the first round. Mutants which showed more than 50 % of residual activity towards Skim-milk and above 40 % of residual activity enhancement either in 1 M GdmCl or 0.5 % SDS were selected for rescreening.

Finally, 52 mutants of interest were selected and rescreened. 8 mutants of interest were thus selected and characterized (see Table 3.11). Their plasmids were isolated and mixed as the template for the next round of directed evolution.

Table 3.11 Overview of properties of selected mutants from 2nd round of epPCR library. The chaotolerance ratios (1 M GdmCl or 0.5 % SDS) or activity ratios (AAPF or Skim-milk) are calculated according to equations in section 3.1.2. Expression level of selected variants is also compared to the wild type as indicated.

Variants	1	2	3	4	5	6	7	8
1 M GdmCl	2.1	0.5	2.3	2.0	2.6	2.2	1.9	2.0
0.5 % SDS	1.4	1.5	1.4	1.4	1.4	1.4	1.4	1.4
AAPF	2.2	0.7	1.7	2.8	3.1	2.1	1.9	1.7
Skim-milk	0.9	0.7	0.7	1.0	0.6	0.9	0.8	0.6
Expression	↑	=	↓	↑	↑	↑	=	↑

3.1.4.3 3rd round of screening- SeSaM1

Screening of the SeSaM library was carried out in the same way as for epPCR libraries. Mutants which showed more than 30 % of residual activity towards Skim-milk and above 2.5 of resistance ratio in 1 M GdmCl or above 2.2 in 0.5 % SDS were selected for rescreening. Finally, 85 mutants of interest were selected and re-screened. 10 mutants of interest were selected and confirmed (see Table 3.12).

Table 3.12 Overview of properties of selected mutants from 1st round of SeSaM library. The chaotolerance ratios (1 M GdmCl or 0.5 % SDS) or activity ratios (AAPF or Skim-milk) are calculated according to equations in section 3.1.2. Expression level of selected variants is also compared to the wild type as indicated.

Variants	1	2	3	4	5	6*	7	8	9	10
1 M GdmCl	3.6	4.0	4.2	3.4	4.8	4.2	3.4	3.9	3.9	1.5
0.5 % SDS	2.5	2.9	2.9	4.2	4.9	2.8	2.4	1.4	2.6	2.7
AAPF	5.3	5.0	4.7	1.9	1.8	4.7	5.1	5.0	5.0	0.9
Skim-milk	0.7	0.5	0.4	0.7	0.3	0.6	2.0	0.6	0.6	0.5
Expression	↓	↓	=	↓	↓↓	-	=	↓	↓	↓↓

* No.6 mutant completely lost its activity and resistance during the isolation probably due to plasmid loss.

3.1.5 Comparison of selected variants

After three rounds of directed evolution, a few variants with highly improved tolerance towards GdmCl or SDS were identified. Substitutions of selected variants from all three rounds of directed evolution were summarized in Table 3.13. Variants designated as ep1-8, ep2-3 and SeSaM1-5, were considered as the “best”

Table 3.13 Overview of substitutions and properties of all selected variants from three rounds of directed evolution.

Variants	A1E[1]*	S62I	D97A	Q103P	V149A	A153V	G166S	N181D	N184H	I205V	N218S	T224A	A232V	T240S	GdmCl[2]	SDS[2]	AAPF	Skim-milk	Expression
ep1-4												+			1.4	1.2	2.0	0.9	↑
ep1-8**					+										1.4	1.2	1.0	1.0	↓
ep1-12				+											0.5	1.4	0.6	0.8	=
ep2-2				+											0.5	1.5	0.7	0.7	=
ep2-3						+	+			+					2.3	1.4	1.7	0.7	↓
ep2-5											+	+			2.6	1.4	3.1	0.6	↑
SeSaM1-1						+	+				+	+			3.6	2.5	5.3	0.7	↓
SeSaM1-2						+	+			+	+	+			4.0	2.9	5.0	0.5	↓
SeSaM1-3	+					+	+			+	+	+	+		4.2	2.9	4.7	0.4	=
SeSaM1-4			+			+	+	+		+					3.4	4.2	1.9	0.7	↓
SeSaM1-5		+				+	+			+					4.8	4.9	1.8	0.3	↓
SeSaM1-7						+	+					+		+	3.4	2.4	5.1	2.0	=
SeSaM1-8						+	+		+		+	+			3.9	1.4	5.0	0.6	↓
SeSaM1-9						+	+				+	+			3.9	2.6	5.0	0.6	↓

1. The order of number for amino acid substitution starts from mature domain of subtilisin E.
2. The concentration of GdmCl for screening is 1 M, while the concentration of SDS is 0.5 %.

* The underlined amino acid substitutions indicate the ones that have been reported in subtilisin E or its homologs.

** The variants highlighted in red indicate the “best” variant from corresponding round of directed evolution.

representative mutants in their corresponding round of directed evolution, and were subjected to activity assays in varied concentrations of GdmCl (1-5 M) or SDS (0.5-2.0%) for comparison (see Figure 3.14). Further comparison and validation are described in next section.

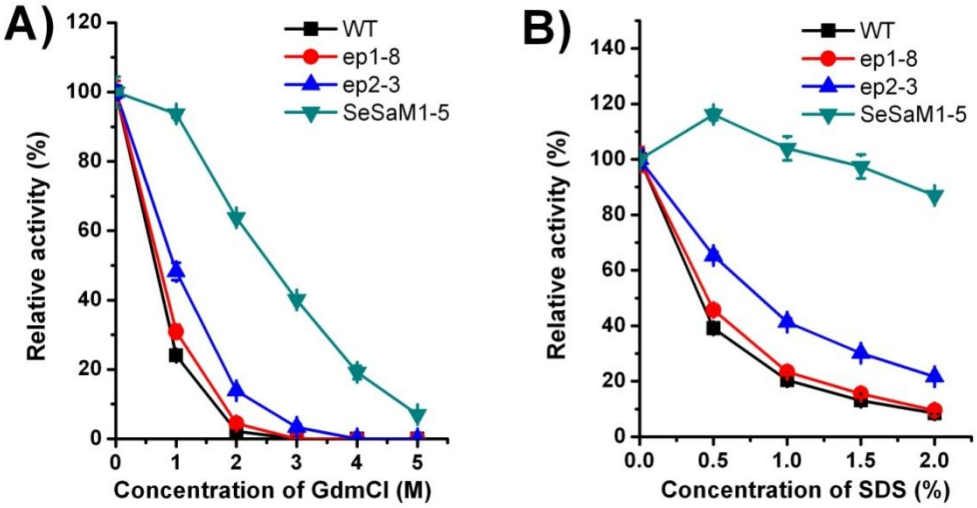

Figure 3.14 Comparison of GdmCl- and SDS-resistance of selected variants in cell-free supernatants from three rounds of directed evolution. A) Comparison of activities in the presence of different concentrations of GdmCl. B) Comparison of activities in the presence of different concentrations of SDS. The relative activities of cell-free supernatants were calculated by using Suc-AAPF-AMC as substrate in Tris-HCl buffer (0.1 M, pH 8.5) in the absence and presence of GdmCl or SDS, considering activity in the absence of GdmCl or SDS as 100 %.

3.2 Site saturation mutagenesis and recombination of beneficial substitutions

All 14 substitutions identified (see section 3.1.5) among selected variants in terms of their improved GdmCl- or SDS-resistance were investigated by site saturation mutagenesis (SSM). The study was conducted in two parts: one focused on 4 positions identified in the “best” mutant SeSaM1-5, designated as M4, and the other focused on the other 10 positions. The recombinatorial study of beneficial substitutions was followed after screening and confirmation, which details in Figure 3.15.

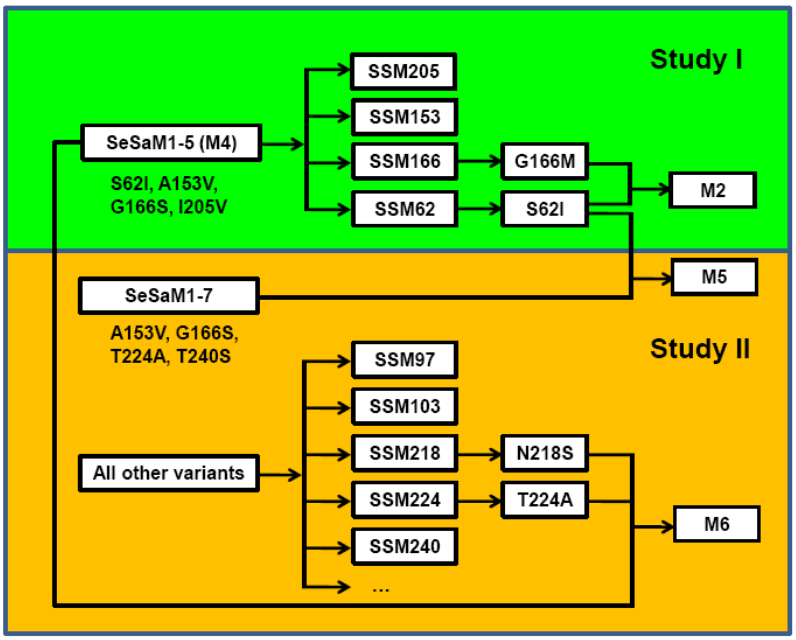

Figure 3.15 Overview of site saturation mutagenesis (SSM) and recombination of substitutions in subtilisin E identified in directed evolution. Study I focused on the substitutions in the "best" variant SeSaM1-5 (M4). Study II focused on the other substitutions. Single substitution in the box indicates the "best" single one at its positions in terms of GdmCl- or SDS-resistance. Recombinant variants M2, M5 and M6 were constructed as indicated.

3.2.1 Construction and screening of SSM libraries

The construction of SSM libraries at position 62, 166, 153, 205 (Study I) and at position 1, 97, 103, 149, 181, 184, 218, 224, 232, 240 (Study II) were carried out using the QuikChange method [99]. The SSM libraries were then transformed into *B. subtilis* WB600 for further screening.

180 clones for each SSM library were subjected to screening in duplicate under the same conditions of the directed evolution experimetns, and the results were detailed in Figure 3.16.

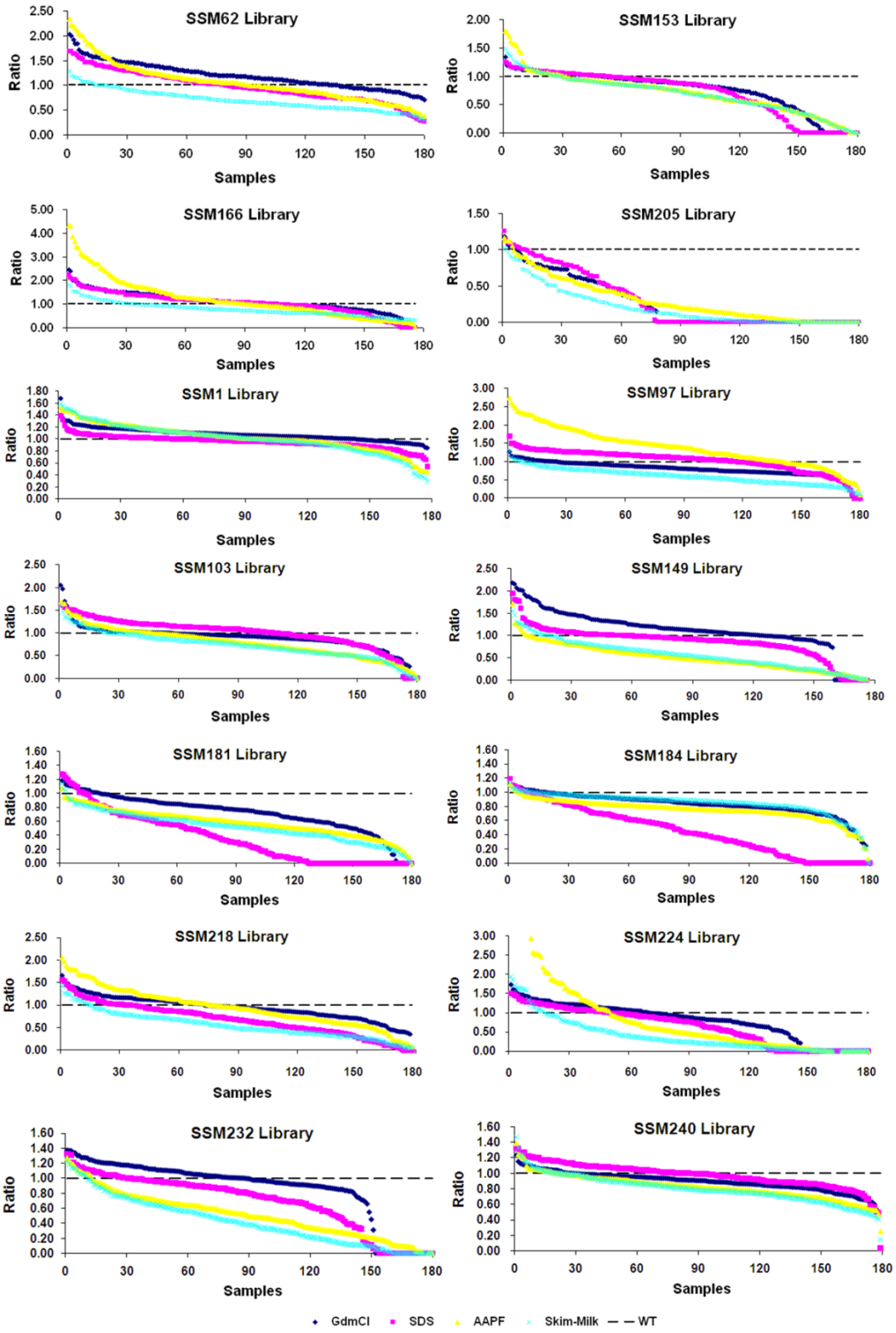

Figure 3.16 Overview of screening results of site saturation mutagenesis (SSM) libraries at positions 62, 153, 166 and 205 (Study I) and position 1, 97, 103, 149, 181, 184, 218, 224, 232 and 240 (Study II) identified from selected variants from directed evolution. 180 clones for each SSM library were screened in duplicate and their chaotolerance ratios (1 M GdmCl or 0.5 % SDS) and activity ratios (AAPF or Skim milk) were calculated and shown as indicated.

Further rescreening and confirmation proved that substitutions at position 62, 166 (Study I), 97, 103, 218 and 224 (Study II) have evident effects on GdmCl- and/or SDS-resistance. The improved variants were isolated and their substitutions were confirmed by sequencing (see Table 3.17).

Table 3.17 Overview of all beneficial substitutions for GdmCl- and/or SDS-resistance. The number of sequenced clones with beneficial substitutions is listed. The chaotolerance ratios (1 M GdmCl or 0.5 % SDS) or activity ratios (AAPF or Skim-milk) are calculated according to equations in section 3.1.2. Expression level of selected variants is also compared to the wild type as indicated.

Substitutions	Number (All)	Properties (Ratios)				
		GdmCl	SDS	AAPF	Skim-milk	Expression
S62T	3(10)	1.3	1.4	0.4	1.3	=
S62I[1,2]	2(10)	2.0	1.7	2.5	0.4	=
S62F	2(10)	1.6	1.5	2.7	0.5	=
S62L	1(10)	1.5	1.7	2.8	0.6	=
S62C	1(10)	1.5	0.6	1.5	0.2	=
S62K	1(10)	1.1	0.1	0.4	0.5	=
G166A	3(13)	2.2	1.5	3.5	0.9	=
G166F	3(13)	2.1	1.4	0.9	0.6	=
G166S	2(13)	1.8	1.1	1.7	1.0	=
G166C	2(13)	1.9	1.5	2.0	0.7	=
G166M	2(13)	2.5	2.0	3.3	0.9	=
G166Y	1(13)	1.6	1.2	0.7	0.8	=
D97A	1(1)	1.1	1.4	2.5	0.8	=
Q103D	1(6)	2.2	0.9	0.2	0.6	↓
Q103P	3(6)	0.6	1.5	0.6	1.0	↓
Q103E	2(6)	1.5	1.0	0.4	0.9	↓
N218S	4(4)	1.4	1.2	2.2	0.6	=
T224S	7(12)	1.6	1.2	3.4	0.9	=
T224C	3(12)	1.4	1.3	3.0	1.4	=
T224A	1(12)	1.4	1.1	3.7	1.3	=
T224G	1(12)	1.4	1.0	2.1	0.5	=

1. Underlined substitutions are also found in variants from directed evolution.
2. The variants highlighted in red indicate the best single variant from each SSM library.

The substitutions responsible for GdmCl- and/or SDS- tolerance have no specific patterns. However, all these positions are located either close to active site residues, such as S62 adjacent to H64 and position N218, T224 adjacent to S221, or close to

substrate binding pockets, such as position G166 in specific binding pocket P1 and position D97 or Q103 close or in unspecific substrate binding site P4 (see Figure 1.1 and Figure 3.18). Roughly a glimpse of substitutions at position 62 and 166, as well as 224, which have dominant impacts on GdmCl- and SDS-resistance, reveals that most are hydrophobic residues. Notabily, only single substitution is found responsible for tolerance at position 97 (D97A) and 218 (N218S), while substitution D97A only accounts for SDS tolerance. Interestingly, two substitutions, Q103D and Q103P were responsible specifically for GdmCl tolerance or SDS tolerance at position 103 respectively, but with dramatically reduced activity towards Suc-AAPF-AMC as substrate (see Table 3.17). More analysis for proposed mechanism of improved tolerance caused by these substitutions is detailed in the Discussion chapter.

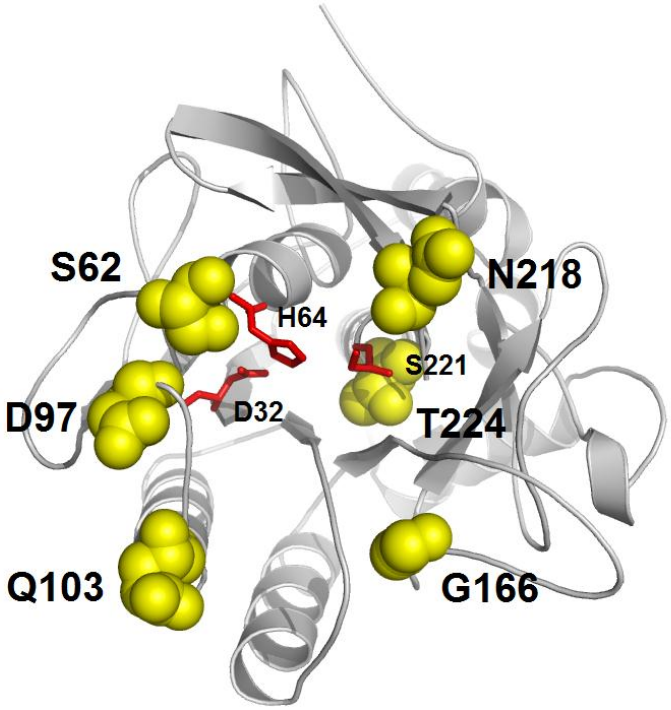

Figure 3.18 Schematic representation of subtilisin E wild type with identified amino acid substitutions. Six positions of amino acid substitutions (S62, G166, D97, Q103, N218 and T224 in yellow in sphere format) identified by site saturation mutagenesis (SSM) are labeled. Catalytic triad (D32, H64 and S221) is shown in red and labeled. The model was generated based on crystal structure complex of subtilisin E (PDB code: 1SCJ) by removing the propeptide and mutation of C221 to its catalytic residue S221.

3.2.2 Recombination of beneficial mutations

As seen in Figure 3.15, several recombinant variants were constructed based on the results of site saturation mutagenesis experiments: M2 (S62I/G166M) was generated by combination of two “best” single substitutions in Study I (S62I and G166M); M5 (S62I/A153V/G166S/T224A/T240S) was generated in Study II by additional substitution of S62I by site directed mutagenesis in variant SeSaM1-7, the only variant from directed evolution with enhanced activity towards complex substrate (see Table 3.13), to increase its tolerance but intentionally retain the activity towards complex substrate; M6 (S62I/A153V/G166S/I205V/N218S/T224A) was constructed in Study II by introduction of additional substitutions of N218S and T224A by site directed mutagenesis in variant SeSaM1-5 (M4) to further enhance its GdmCl- and/or SDS-resistance. Besides, some neutral single substitutions, revealed by site saturation mutagenesis, were also constructed by site directed mutagenesis for comparison, such as A153V, I205V (in SeSaM1-5) and T240S (in SeSaM1-7).

3.2.2.1 Study I: 1st round of SSM and recombination

The cell-free supernatants of selected variants in Study I were subjected to extensive activity assays in the presence of different concentrations of GdmCl or SDS. The results were showed in Figure 3.19 and more details were shown in Table 3.20. The results showed that two positions (62 and 166) contributed mainly to the improved resistance towards GdmCl and SDS. A153V and I205V, which were maintained in variant SeSaM1-5 (M4) had no superior resistance than the wild type (see Figure 3.19). As previously shown in section 3.2.1, SSM of position 62 yielded again the substitution S62I as most beneficial one. SSM of position 166 yielded as the “best” variant a different substitution (G166M instead of G166S) which is apparently more resistant. Finally a double mutant (M2: S62I and G166M) was generated by site directed mutagenesis. Comparison of the performance of M2 and SeSaM1-5 (M4) in Figure 3.19 shows a comparable activity on Skim-milk. In respect to GdmCl- and

SDS-resistance the SeSaM1-5 (M4) variant performs slightly better than M2 under screening conditions. The latter indicates a cooperative effect within the substitutions of SeSaM1-5 (M4) in the presence of Suc-AAPF-AMC.

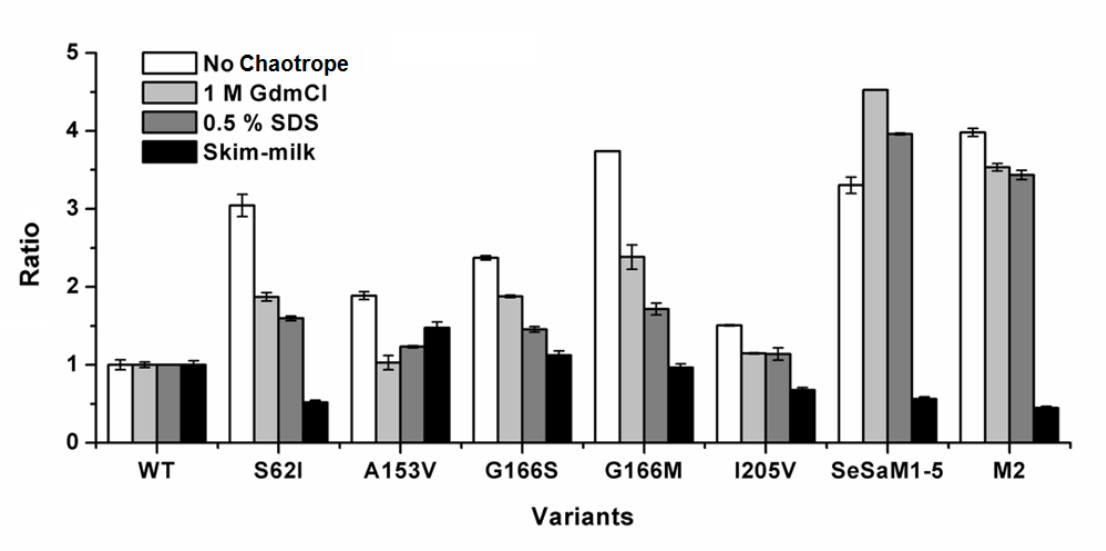

Figure 3.19 Activity and resistance ratios of subtilisin E variants (Study I) in the presence or absence of chaotropes. All ratios of subtilisin E variants in cell-free supernatants were calculated by using Suc-AAPF-AMC as substrate in Tris-HCl buffer (0.1 M, pH 8.5) in the absence and presence of 1 M GdmCl or 0.5 % SDS and normalized in U/mg using the Agilent Protein 230 Kit. Activity and resistance ratios were calculated from independent measurements in triplicate with standard deviation indicated. Activity ratios for both Suc-AAPF-AMC and Skim-milk as substrates without chaotropes were calculated according to equation for Ratio (2) and displayed as □: Suc-AAPF-AMC, ■: Skim-milk. Resistance ratios in the presence of 1 M GdmCl or 0.5 % SDS using Suc-AAPF-AMC as substrate were calculated according to equation for Ratio (1) and displayed as ■: 1 M GdmCl, ■: 0.5 % SDS.

3.2.2.2 Study II: 2nd round of SSM and recombination

The cell-free supernatants of selected variants in Study II were subjected to extensive activity assays in the presence of different concentrations of GdmCl or SDS. The results were showed in Figure 3.21 and more details were listed in Table 3.22. N218S and T224A have moderate effects on both GdmCl and SDS tolerance, and that is also why they were selected to introduce into M4 to generate M6 for further improvement at the first place. However, D97A and Q103P has only moderate effects on SDS tolerance, while Q103D has apparent effects on GdmCl tolerance, but loses 70 % activity on Suc-AAPF-AMC as substrate (see Figure 3.22). Thus they were not

Table 3.20 Specific activity of subtilisin E variants (Study I) and their ratios in the presence or absence of chaotropes.

Variants	No chaotropes	GdmCl					SDS				Skim-milk
		1 M	2 M	3 M	4 M	5 M	0.50%	1.00%	1.50%	2.00%	
Wild type	591.56 [1]	113.53	N.D.	N.D.	N.D.	N.D.	189.68	95.47	42.40	23.07	
	1.00 [2]	**1.00** [3]	**-**	**-**	**-**	**-**	**1.00** [3]	**1.00**	**1.00**	**1.00**	**1.00** [2]
SeSaM1-5	1954.10	1697.56	1129.21	549.10	168.69	N.D.	2482.60	2036.75	1751.45	1541.32	
	3.30	**4.53**	**-**	**-**	**-**	**-**	**3.96**	**6.46**	**12.51**	**20.22**	**0.56**
M2	2354.70	1596.55	1071.92	597.74	274.16	63.67	2594.62	1988.38	1670.03	1428.65	
	3.98	**3.53**	**-**	**-**	**-**	**-**	**3.44**	**5.23**	**9.90**	**15.56**	**0.44**
S62I	1800.86	646.64	156.96	N.D.	N.D.	N.D.	921.48	511.33	359.18	278.29	
	3.04	**1.87**	**-**	**-**	**-**	**-**	**1.60**	**1.76**	**2.78**	**3.96**	**0.52**
A153V	1116.59	219.93	N.D.	N.D.	N.D.	N.D.	441.39	235.97	137.08	58.04	
	1.89	**1.03**	**-**	**-**	**-**	**-**	**1.23**	**1.31**	**1.71**	**1.33**	**1.47**
G166S	1403.92	506.02	65.27	N.D.	N.D.	N.D.	654.68	352.37	222.83	140.52	
	2.37	**1.88**	**-**	**-**	**-**	**-**	**1.45**	**1.56**	**2.21**	**2.57**	**1.12**
G166M	2211.35	1010.93	275.20	34.59	N.D.	N.D.	1215.83	784.85	584.99	418.19	
	3.74	**2.38**	**-**	**-**	**-**	**-**	**1.71**	**2.20**	**3.69**	**4.85**	**0.96**
I205V	890.85	195.98	N.D.	N.D.	N.D.	N.D.	325.30	160.70	113.91	61.68	
	1.51	**1.15**	**-**	**-**	**-**	**-**	**1.14**	**1.12**	**1.78**	**1.78**	**0.68**

[1] All specific activities of subtilisin E variants in cell-free supernatants were calculated by using Suc-AAPF-AMC as substrate in Tris-HCl buffer (0.1 M, pH 8.5) in the absence and presence of given concentrations of GdmCl or SDS and normalized in U/mg using the Agilent Protein 230 Kit. Activity and resistance measurements were carried out in triplicate with <10 % standard deviation. N.D. indicates that no activity could be detected. [2] Activity ratios towards both Suc-AAPF-AMC and Skim-milk as substrates were calculated according to equation for Ratio (2), and highlighted in bold. [3] Resistance ratios towards given concentrations of GdmCl or SDS were calculated according to equation for Ratio (1) and highlighted in bold.

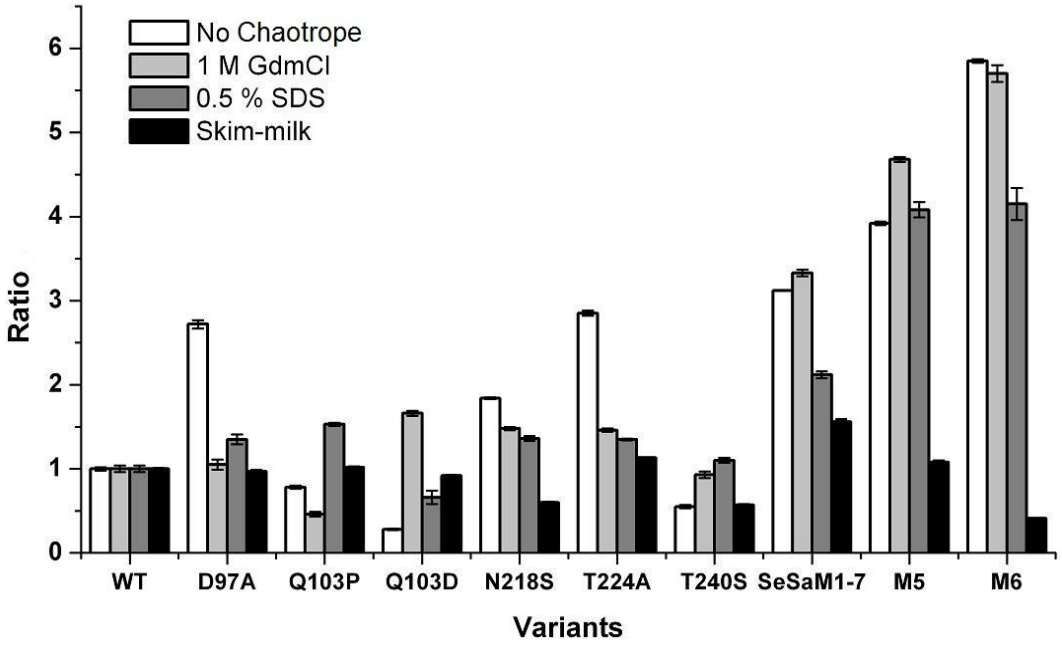

Figure 3.21 Activity and resistance ratios of subtilisin E variants (Study II) in the presence or absence of chaotropes. All ratios of subtilisin E variants in cell-free supernatants were calculated by using Suc-AAPF-AMC as substrate in Tris-HCl buffer (0.1 M, pH 8.5) in the absence and presence of 1 M GdmCl or 0.5 % SDS and normalized in U/mg using Experion microfluidic system. Activity and resistance ratios were calculated from independent measurements in triplicate with standard deviation indicated. Activity ratios for both Suc-AAPF-AMC and Skim-milk as substrates without chaotropes were calculated according to equation for Ratio (2) and displayed as □: Suc-AAPF-AMC, ■: Skim-milk. Resistance ratios in the presence of 1 M GdmCl or 0.5 % SDS using Suc-AAPF-AMC as substrate were calculated according to equation for Ratio (1) and displayed as ■: 1 M GdmCl, ■: 0.5 % SDS.

regarded as substitutions for combinations. M6, was constructed by introduction of N218S and T224A into SeSaM1-5 (M4), and showed the highest GdmCl- and SDS-resistance (1 M GdmCl: 5.7 fold; 0.5 % SDS: 4.2 fold) compared to SeSaM 1-5 (M4, 1 M GdmCl: 4.5 fold; 0.5 % SDS: 4.0 fold) or M2 (1 M GdmCl: 3.5 fold; 0.5 % SDS: 3.4 fold).

SeSaM1-7, as mentioned previously, is the only variant from directed evolution which has improved activity towards complex substrate (1.6 fold) and retained moderate GdmCl- and SDS-resistance (1 M GdmCl: 3.3 fold; 0.5 %: 2.1 fold). Additional substitution S62I was introduced into this variant to construct M5 (see Figure 3.15) to attempt to retain its activity towards complex substrate but increase

Table 3.22 Specific activity of subtilisin E variants (Study II) and their ratios in the presence or absence of chaotropes.

Variants	No chaotropes	GdmCl					SDS				Skim-milk
		1 M	2 M	3 M	4 M	5 M	0.50%	1.00%	1.50%	2.00%	
Wild type	570.39 [1]	119.04	N.D.	N.D.	N.D.	N.D.	191.52	95.16	43.06	25.59	
	1.00 [2]	**1.00** [3]	**-**	**-**	**-**	**-**	**1.00** [3]	**1.00**	**1.00**	**1.00**	**1.00** [2]
SeSaM1-7	1778.48	1234.92	513.20	106.46	1.48	N.D.	1265.60	933.53	625.51	453.46	
	3.12	**3.33**	**-**	**-**	**-**	**-**	**2.12**	**3.15**	**4.66**	**5.68**	**1.56**
M5	2237.64	2184.60	1742.60	1088.64	391.70	68.48	3066.16	2823.52	2540.43	2396.05	
	3.92	**4.68**	**-**	**-**	**-**	**-**	**4.08**	**7.56**	**15.04**	**23.85**	**1.08**
M6	3335.20	3968.31	3621.43	2967.85	1812.91	764.60	4650.90	4654.79	4401.85	4259.18	
	5.85	**5.70**	**-**	**-**	**-**	**-**	**4.15**	**8.37**	**17.48**	**28.44**	**0.41**
N218S	1047.94	322.98	39.22	N.D.	N.D.	N.D.	478.78	277.85	162.92	115.05	
	1.84	**1.48**	**-**	**-**	**-**	**-**	**1.36**	**1.59**	**2.06**	**2.45**	**0.60**
T224A	1625.21	496.13	95.14	N.D.	N.D.	N.D.	738.70	411.68	262.48	182.41	
	2.85	**1.46**	**-**	**-**	**-**	**-**	**1.35**	**1.52**	**2.14**	**2.50**	**1.13**
T240S	316.27	61.08	N.D.	N.D.	N.D.	N.D.	117.16	56.12	31.12	15.67	
	0.55	**0.93**	**-**	**-**	**-**	**-**	**1.10**	**1.06**	**1.30**	**1.10**	**0.57**
Q103P	446.43	42.78	N.D.	N.D.	N.D.	N.D.	229.69	86.04	34.93	7.38	
	0.78	**0.46**	**-**	**-**	**-**	**-**	**1.53**	**1.16**	**1.04**	**0.37**	**1.02**
Q103D	161.76	56.02	N.D.	N.D.	N.D.	N.D.	35.77	N.D.	N.D.	N.D.	
	0.28	**1.66**	**-**	**-**	**-**	**-**	**0.66**	**-**	**-**	**-**	**0.92**
D97A	1554.12	341.89	8.27	N.D.	N.D.	N.D.	703.88	300.92	174.10	73.68	
	2.72	**1.05**	**-**	**-**	**-**	**-**	**1.35**	**1.16**	**1.48**	**1.06**	**0.97**

[1] All specific activities of subtilisin E variants in cell-free supernatants were calculated by using Suc-AAPF-AMC as substrate in Tris-HCl buffer (0.1 M, pH 8.5) in the absence and presence of given concentrations of GdmCl or SDS and normalized in U/mg using the Experion microfluidic system. Activity and resistance measurements were carried out in triplicate with <10 % standard deviation. N.D. indicates that no activity could be detected. [2] Activity ratios towards both Suc-AAPF-AMC and Skim-milk as substrates were calculated according to equation for Ratio (2), and highlighted in bold. [3] Resistance ratios towards given concentrations of GdmCl or SDS were calculated according to equation for Ratio (1) and highlighted in bold.

its GdmCl- and SDS-resistance. The results in Figure 3.21 showed that the activity of M5 towards Skim-milk was equal to wild type activity (1.1 fold) but the GdmCl- and SDS-resistance was improved (1 M GdmCl: 4.7 fold; 0.5 % SDS: 4.0 fold) compared to SeSaM1-7, and was slightly better than SeSaM 1-5 (M4, 1 M GdmCl: 4.5 fold; 0.5 % SDS: 4.0 fold). The single substitution T240S, which was maintained in variant SeSaM1-7 had no superior resistance than wild type (see Figure 3.21). Its effect on SeSaM1-7 regarding its enhanced activity towards complex substrate is unknown. However, its position is at the back of the active site, and it could be assumed that this substitution might have effects on flexibility of active sites, that indicated by structure analysis of protease K, a active protease with very broad substrate specificity [117].

3.2.2.3 Recombination with SeSaM-based shuffling

Though sequencial addition of single beneficial substitution into wild type (M2 in Study I) or other variants (M5 or M6 in Study II) generated more chaotolerant variants (see section 3.2.2.1 and 3.2.2.2), this approach rules out single neutral or even deleterious substitution which might become beneficial with collaborative effects in muteins. A new method “SeSaM-based shuffling” is designed and developed for random combination of all mutations of interest to overcome such drawback.

Fragmentation of both wild type (WT) and synthetic gene with mutations (M14, see Appendix I) was performed according to SeSaM Tv-Handbook (classic) Revision 0.21 provided by SeSaM-Biotech GmbH. Optimal phosphorothioate percentage was determined experimentally for both WT and M14: G_fwd library (dGTPαS 40 %), G_rev library (dGTPαS 40 %).

Iodine digested fragments were then shuffled by PCR at different cycles to generate full length gene with different mutation recombinations. Different PCR cycles were thus investigated as well as gene template ratio WT:M14. Figure 3.23 showed the sequence coverage of 10 sequenced clones randomly picked from each SeSaM-based shuffling library. The statistical analysis indicates that the more M14 template amount and more shuffling PCR cycles are, the more randomly shuffled

genes can be generated (more S2 sequences). However, the coverage of template sequences (WT and M14) is still not ruled out.

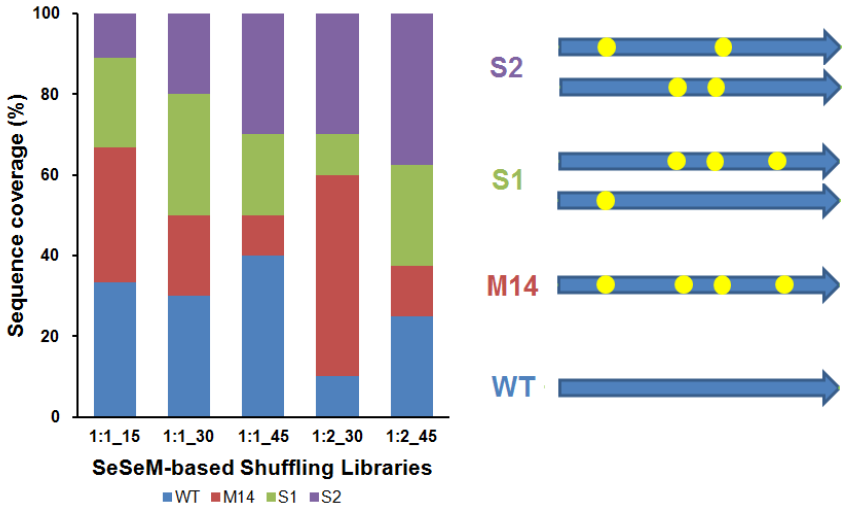

Figure 3.23. Sequence coverage of SeSaM-based shuffled genes with different template ratios and shuffling PCR cycles. WT, wild type sequence. M14, synthetic gene sequence (14 amino acid substitutions in yellow). S1, randomly shuffled gene sequence type 1 (without gaps). S2, randomly shuffled gene sequence type 2 (with gaps). Wild type and synthetic gene template ratios and shuffling PCR cycles are indicated for instance 1:1_15 with 1:1 for WT to M14 ratio and 15 for shuffling PCR cycles.

Screening of all five SeSaM-based shuffling libraries (~300 clones for each) resulted in dramatic loss of activity towards complex substrate Skim-milk (see Table 3.24). Additionally the chaotolerance ratios of selected “best” variant 1:2_45_3F11 with 10 substitutions were only comparable to SeSaM1-5 (M4) (see Table 3.13). It might be due to screening of the limited number of clones from each library (~300 clones for each), since combination of all 14 mutations requires 2^{14} (16,384) as the library size in theory. It is also observed that some substitutions were difficult to separate because they were very close to each other, such as position 218, 224, 232 and 240. Besides, the expression of all these improved variants decreased probably due to the high mutation load (see Table 3.24).

All above results show SeSaM-based shuffling is a feasible new method for random recombination of mutations of interest. However, more investigations are still required for further improvement.

Table 3.24 Overview of substitutions and properties of selected variants from SeSaM-based shuffling libraries.

Variants	A1E[1]	S62I	D97A	Q103P	V149A	A153V	G166S	N181D	N184H	I205V	N218S	T224A	A232V	T240S	GdmCl[2]	SDS[2]	AAPF	Skim-milk	Expression
1:2_30_1C8	+	+						+	+	+	+	+	+	+	3.3	3.2	7.8	0.1	↓
1:2_30_3C4		+					+	+	+	+	+	+	+	+	4.7	5.2	3.5	0.1	↓
1:2_30_3G10	+	+	+					+			+	+	+	+	3.5	4.5	3.3	0.0	↓
1:2_45_1F5	+	+								+	+	+	+	+	3.4	3.1	8.0	0.1	↓
1:2_45_1H4	+	+								+	+	+	+	+	3.3	3.1	7.9	0.1	↓
1:2_45_3F11	+	+					+	+	+	+	+	+	+	+	4.8	5.2	3.8	0.0	↓

1. The order of number for amino acid substitution starts from mature domain of subtilisin E.

2. The concentration of GdmCl for screening is 1 M, while the concentration of SDS is 0.5 %.

3.2.3 Purification of subtilisin E and its variants

3.2.3.1 Study I: 1st round of SSM and recombination

The purification of wild type, S62I, G166M, SeSaM1-5 (M4) and M2 was performed by using a two-step ion exchange chromatography procedure (see *Method 1* in 2.2.8). SDS-PAGE analysis revealed that this protocol was successful in protease purification (see Figure 3.25).

The purity of variant enyzmes was confirmed by Agilent Protein 230 Kit, and the contents of all purified variants were above 85 % (see Figure 3.26).

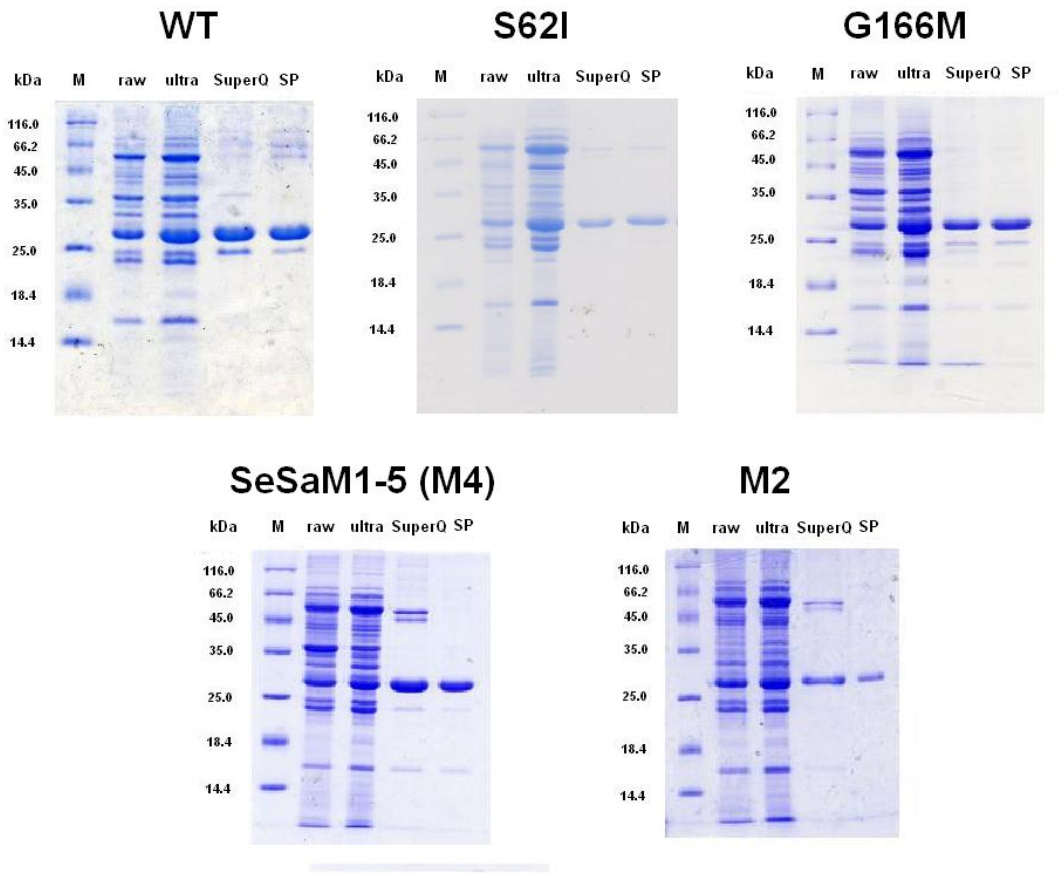

Figure 3.25 SDS-PAGE analysis of purification of subtilisin E and its variants (*Method 1*). M: Protein marker; raw, cell-free supernatant after filtration; ultra, ultra-filtrated enzyme solution; SuperQ, collected fractions after purification through anion exchange column SuperQ650; SP, collected fraction after purification through cation exchange column SP650c.

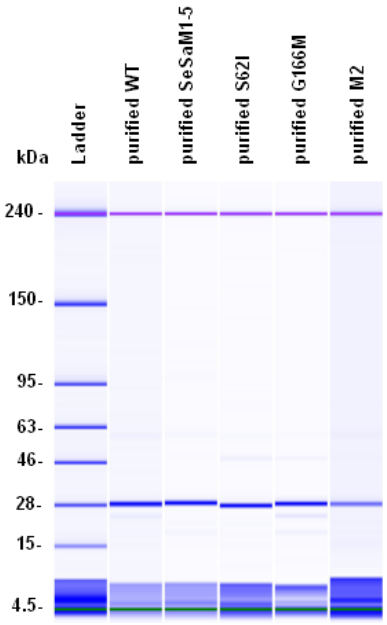

Figure 3.26 Agilent analysis of purified subtislin E and its variants (Study I). The contents of purified subtilisin E are all >85 %.

3.2.3.2 Study II: 2nd round of SSM and recombination

The purification of wild type, SeSaM1-4 (M4), N218S, T224A, M5 and M6 was performed by using a two-step ion exchange chromotrography procedure (see *Method 2* in 2.2.8). SDS-PAGE analysis revealed that this protocol was also successful (see Figure 3.27). Besides, the yield was also improved (data not shown).

The purity of these purified enzymes is confirmed by an Experion microfluidic system, and the contents of the purified wild type, M4, N218S and T224A were all around 90 %, while the M5 and M6 content was around 80 % (see Figure 3.28).

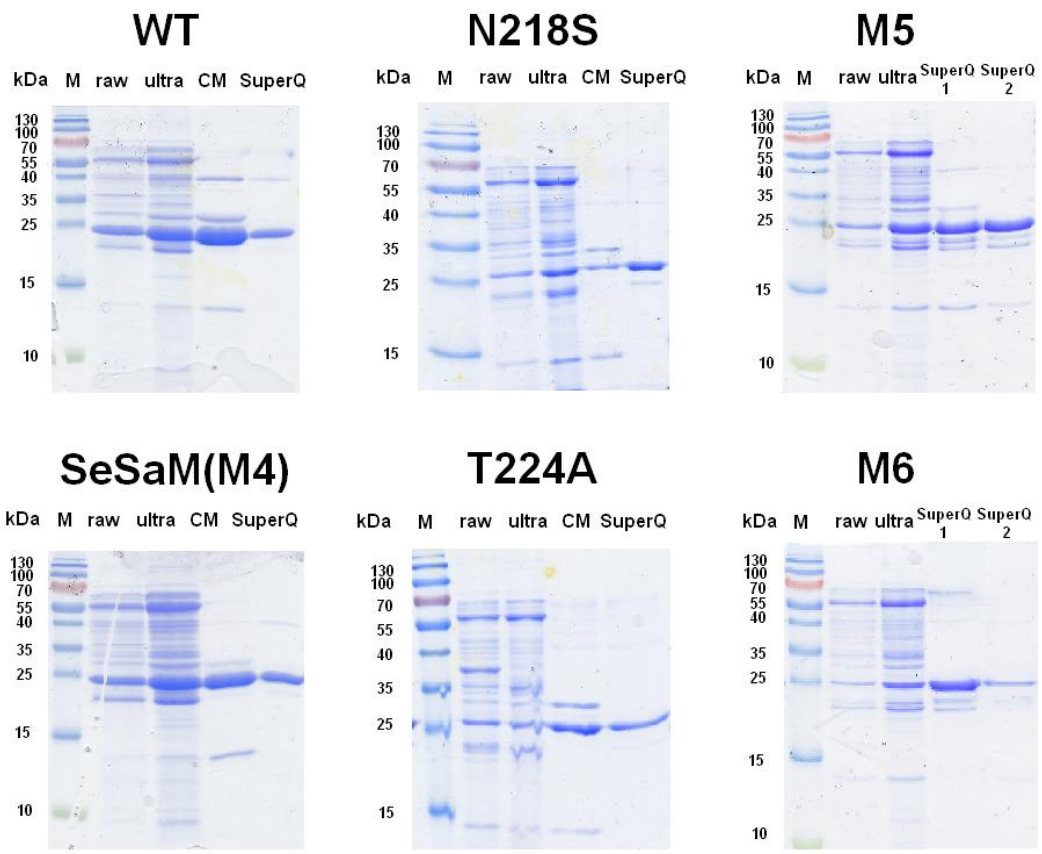

Figure 3.27 SDS-PAGE analysis of purification of subtilisin E and its variants (*Method 2*). M: Protein marker; raw, cell-free supernant after filtration; ultra, ultra-filtrated enzyme solution; CM: collected fractions after purification through cation exchange column CM650; SuperQ, collected fractions after purification through anion exchange column SuperQ650, 1&2 indicate the order of purification steps with SuperQ650 for M5 and M6.

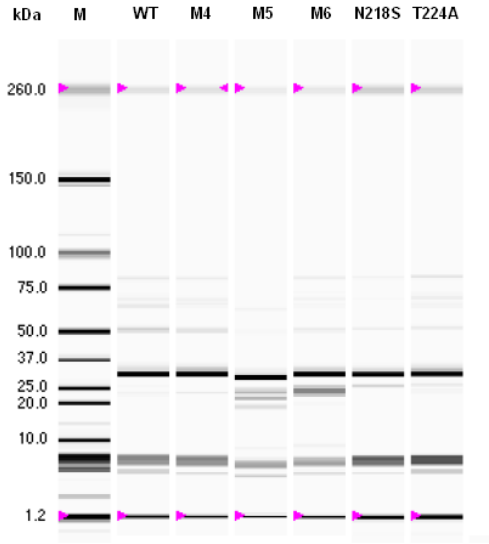

Figure 3.28 Experion analysis of purified subtilisin E and its variants (Study II). The contents of purified Subtilisin E are either around 90 % (WT, M4, N218S and T224A) or around 80 % (M5 and M6).

3.2.4 Characterization of subtilisin E and its variants

3.2.4.1 Determination of kinetic constants

Table 3.29 lists the kinetic data of the purified wild type subtilisin E and obtained subtilisin E variants (S62I, G166M, M2 (S62I/G166M), SeSaM1-5 (M4)) purified by *Method 1* in respect to k_{cat}, K_m and k_{cat}/K_m.

In the absence of GdmCl or SDS the k_{cat} value was improved from 50 s^{-1} (wild type) to 141 s^{-1} (G166M), 123 s^{-1} (M2), and 102 s^{-1} (S62I). SeSaM1-5 has however a lower k_{cat} value (42 s^{-1}). In the presence of 2 M GdmCl the specific activity from SeSaM1-5 is reduced to 34 s^{-1} and the wild type activity to 15 s^{-1}. M2's and G166M's specific activity are in absolute values higher than the one of SeSaM1-5 (62 s^{-1} & 70 s^{-1}; see Table 3.29). In SDS a similar trend can be observed with the following "resistance order" (35 s^{-1} wild type < 45 s^{-1} SeSaM1-5 < 65 s^{-1} S62I < 81 s^{-1} M2 < 104 s^{-1} G166M).

Interestingly, the K_m value for SeSaM1-5 (M4) was significantly reduced compared to the subtilisin E wild type in GdmCl (up to 8-fold (No chaotrope), 28-fold (1 M), and 46-fold (2 M)) and SDS (7 fold (0.5 %); see Table 3.29).

The catalytic efficiency improves in the absence of GdmCl or SDS in the following order (see Table 3.29): G166M (3 fold) < S62I (4 fold) ≈ M2 (4 fold) < SeSaM1-5 (6 fold). In presence of 2 M GdmCl: G166M (22 fold) < S62I (45 fold) < SeSaM1-5 (100 fold) < M2 (118 fold). In presence of 0.5 % SDS: G166M (7 fold) < SeSaM1-5 (9 fold) ≈ S62I (9 fold) < M2 (13 fold).

Table 3.29 Kinetic constants of substilisin E wild type and variants for the hydrolysis of Suc-AAPF-pNA

Variants	Chaotropes	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($mM^{-1} \cdot s^{-1}$)
Wild type	**No chaotrope**	1.03±0.17	50.3±3.0	48,8±0.2
	1 M GdmCl	1.93±0.11	22.6±0.4	11.8±0.1
	2 M GdmCl	7.31±0.58	15.2±0.6	2.1±0.1
	0.5 % SDS	2.45±0.18	35.0±0.9	14.3±0.1
SeSaM1-5	**No chaotrope**	0.13±0.01	42.1±0.6	315.9±0.1
(M4)	**1 M GdmCl**	0.07±0.01	38.4±1.0	539.6±0.2
	2 M GdmCl	0.16±0.01	33.7±0.7	211.0±0.1
	0.5 % SDS	0.35±0.03	44.5±0.9	125.8±0.1
S62I	**No chaotrope**	0.48±0.05	101.8±4.3	211.0±0.1
	1 M GdmCl	0.15±0.01	31.5±0.5	208.9±0.1
	2 M GdmCl	0.21±0.01	19.5±0.2	93.7±0.0
	0.5 % SDS	0.50±0.05	64.5±1.8	130.2±0.1
G166M	**No chaotrope**	0.83±0.04	140.9±2.9	169.7±0.1
	1 M GdmCl	0.67±0.02	96.2±0.8	144.0±0.0
	2 M GdmCl	1.54±0.07	69.8±1.3	45.4±0.1
	0.5 % SDS	0.97±0.04	103.5±1.4	106.8±0.1
M2	**No chaotrope**	0.57±0.06	123.0±4.8	216.9±0.1
	1 M GdmCl	0.19±0.01	81.7±1.3	428.7±0.1
	2 M GdmCl	0.25±0.02	61.5±1.4	246.9±0.1
	0.5 % SDS	0.45±0.05	80.5±2.4	179.3±0.1

Values are the means of three separate determinations and are shown ± the standard deviation of these means.

3.2.4.2 Activity in the presence of chaotropes

Different residual activities of purified subtilisin E variants (Study II) in the presence of various concentrations of GdmCl or SDS towards Suc-AAPF-pNA as substrate were measured. Thus, the inactivation curve was drawn to evaluate their activity in these chaotropic agents. IC_{50} values, the concentration at which 50 % of activity of these variants retained under measuring conditions (see Material and Methods 2.2.10), were determined for comparison of their capability of tolerance towards GdmCl or SDS (see Figure 3.30 and Table 3.31).

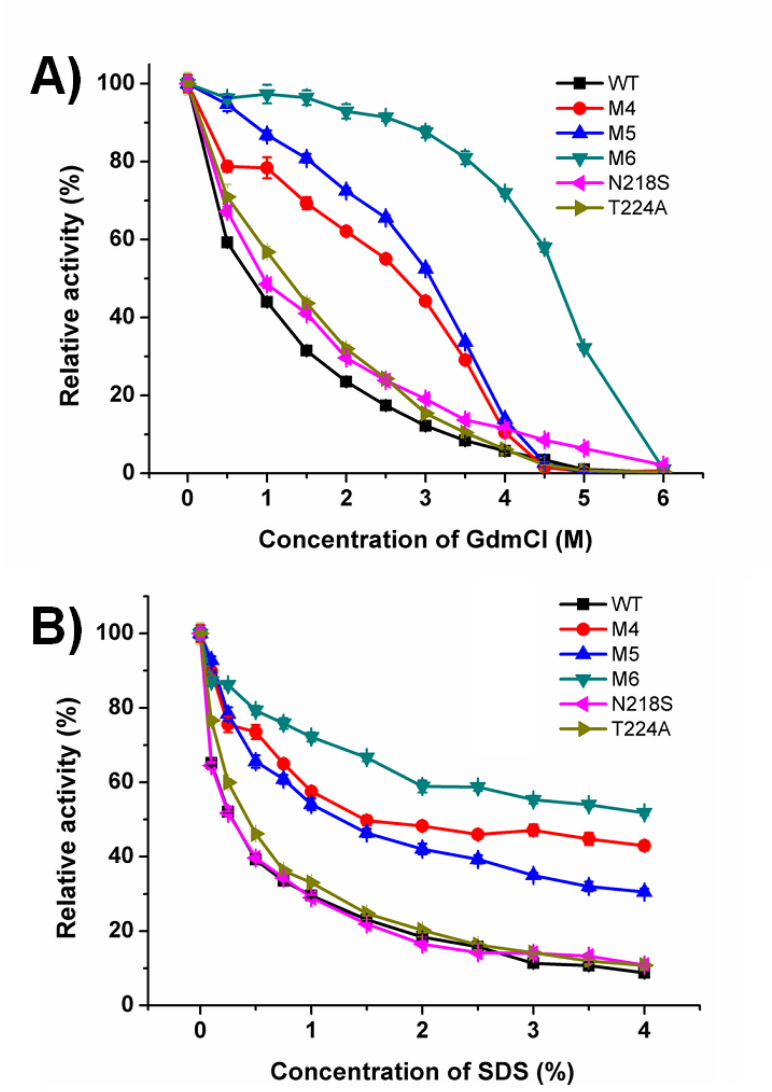

Figure 3.30 Inactivation curve of purified subtilisin E variants (50 nM) in the presence of chaotropes. A) Subtilisin E variants in the presence of GdmCl at different concentrations (0 – 6.0 M); B) Subtilisin E variants in the presence of SDS at different concentrations (0 – 4.0 %).

Both single mutants from Study II, N218S and T224A have slightly improved tolerance towards GdmCl [IC_{50} (N218S) 0.96 M; IC_{50} (T224A) 1.26 M] and SDS [IC_{50} (N218S) 0.29 %; IC_{50} (T224) 0.43 %] compared to WT [IC_{50} (GdmCl) 0.76 M; IC_{50} (SDS) 0.29 %]. But N218S seems more resistant towards GdmCl at higher

Table 3.31 Determination of IC_{50} of purified subtilisin E in the presence of GdmCl or SDS.

Variants		WT	M4	M5	M6	N218S	T224A
IC_{50}	GdmCl (M)	0.76	2.73	3.07	4.65	0.96	1.26
	SDS (%)	0.29	1.48	1.26	4.00	0.29	0.43

concentration (see Figure 3.30 A). For instance, at 5 M GdmCl, this single variant still retained slight activity, while other variants (except M6) lost their activities completely. Three variants with multiple substitutions, including M4, M5 and M6 have much improved IC_{50} values, indicating their tolerance towards GdmCl or SDS in terms of their activities were much improved. Especially M6, the most resistant variant towards chaotropes, showed almost 6 fold (WT: 0.76 M; M6: 4.65 M) and 14 fold (WT: 0.29 %; M6: 4.00 %) improved tolerance towards GdmCl or SDS, respectively. However, though M5 is slightly more tolerant than M4 towards GdmCl, but less tolerant towards SDS compared to M4.

3.2.4.3 Activities towards the complex substrate Azocasein

Azocasein was used as alternative to Skim-milk to quantify the proteolytic activities of subtilisin E variants. The activities of subtilisin E variants towards Suc-AAPF-pNA and Azocasein are listed in Table 3.32.

Table 3.32 Specific activities and ratios of purified subtilisin E variants towards Suc-AAPF-pNA and Azocasein as substrates.

Variants	Suc-AAPF-pNA		Azocasein	
	U/mg	Ratio	U/mg	Ratio
WT	24.6±0.3	1.00±0.01	486±24	1.00±0.04
M4	42.8±0.7	1.74±0.03	172±1	0.35±0.00
M5	111.4±2.0	4.53±0.08	257±3	0.53±0.01
M6	136.1±1.7	5.53±0.07	182±10	0.38±0.02
N218S	33.6±3	1.37±0.01	249±28	0.51±0.05
T224A	39.5±3	1.61±0.01	196±3	0.40±0.01

All purified subtilisin E variants in Table 3.32 lost more than half of their activities towards substrate Azocasein when compared to wild type. Activities of variants towards polypeptide substrate Suc-AAPF-pNA increased from 1.5 to 5.5 fold. However, the activities towards Azocasein were mostly even lower compared to ones towards Skim-milk as substrate (see Table 3.22), which can be attributed to the purification. Purified proteins are less stable than non-purified supernatants. In

addition, the Azocasein assay was performed at 37^{o}C normally for 30 min or 60 min, while Skim-milk assay was performed at 37^{o}C only for 15 min (see section 2.2.7).

3.2.5 Stability of variants in the presence of GdmCl or SDS

Stability of subtilisin E and its variants here in terms of their activities with time in the presence of various concentrations of GdmCl or SDS were investigated and half lives were determined to further evaluate their tolerance towards GdmCl and SDS.

3.2.5.1 Study I: 1st round of SSM and recombination

Figure 3.33 shows that all subtilisin E variants (Study I) are more stable in the presence of all concentrations of chaotropes when compared to the wild type. Subtilisin E wild type lost more than 50 % of its activity as soon as it was exposed to GdmCl or SDS (in less than 2 min), making a half life estimation challenging.

As seen in Table 3.34, in 3 M GdmCl the half lives are <2 min (wild type), <2 min (S62I), 5.7 min (G166M), 47.9 min (SeSaM1-5), 55.7 min (M2); Similar results are obtained for 2 M and 1 M GdmCl. M2 is especially at 3 M GdmCl significantly more stable than SeSaM1-5 (M4).

At 2 % SDS, the half lives are <2 min (wild type), 39.9 min (S62I), 14.5 min (G166M), 54.8 min (SeSaM1-5) and 44.4 min (M2). At 1 % SDS, the half lives are <2 min (wild type), 1.6 h (S62I), 2.5 h (G166M), 1.5 h (SeSaM1-5) and 1.5 h (M2). At 0.5 % SDS, the half lives are 2 min (wild type), 2.9 h (S62I), 2.0 h (G166M), 2.7 h (SeSaM1-5) and 2.8 h (M2). The order of resistance of the four mutants depends in contrast to GdmCl highly on the employed SDS concentration. At 2 % SDS SeSaM1-5 (M4) has the highest half life whereas M2 is more resistant than M4 in GdmCl (3 M).

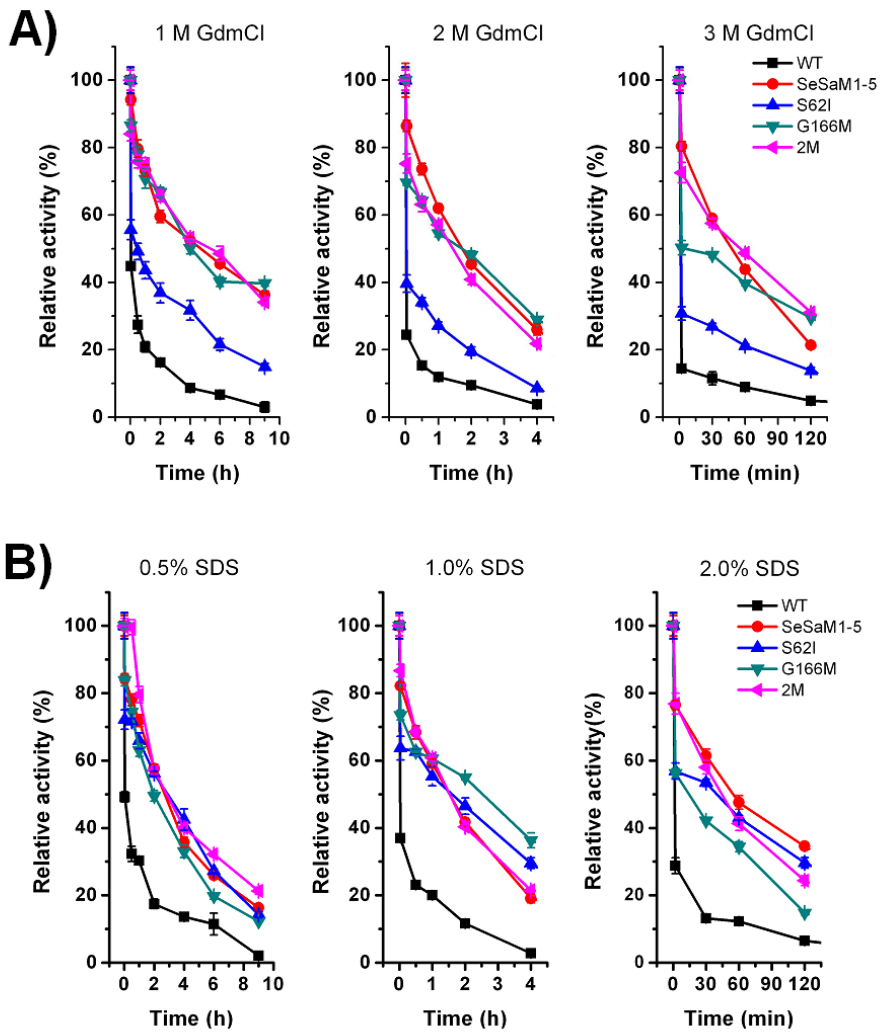

Figure 3.33 Stability of purified subtilisin E wild type and its variants (Study I) in the presence of various concentrations of GdmCl or SDS by determining residual proteolytic activity (Suc-AAPF-pNA assay, 25°C). Relative activities were measured in triplicate and averaged with < 10 % standard deviation considering the activities in the absence of chaotropic agents as 100 %. Final protease concentration was 50 nM.

Table 3.34 Half lives of purified subtilisin E and variants (Study I) in the presence of chaotropes.

Variants	GdmCl			SDS		
	1 M (h)	**2 M (h)**	**3 M (min)**	**0.5% (h)**	**1.0% (h)**	**2.0% (min)**
WT	<2 min	<2 min	<2 min	2 min	<2 min	<2 min
S62I	0.4	<2 min	<2 min	2.9	1.6	39.9
G166M	4.0	1.7	5.7	2.0	2.5	14.5
SeSaM1-5	4.7	1.7	47.9	2.7	1.5	54.8
M2	5.4	1.4	55.7	2.8	1.5	44.4

3.2.5.2 Study II: 2nd round of SSM and recombination

Figure 3.35 showed that, the same stability profile was observed for new variants constructed in Study II, except for M5 which was very unstable, although it still kept high resistance at beginning in the presence of chaotropes. In contrast, single mutant N218S, was extraordinarily stable in all cases, regardless its poor resistance towards chaotropes at beginning. However, M6 showed "highest" stability in GdmCl compared to SeSaM1-5 (M4), and almost the same stability in SDS in M4.

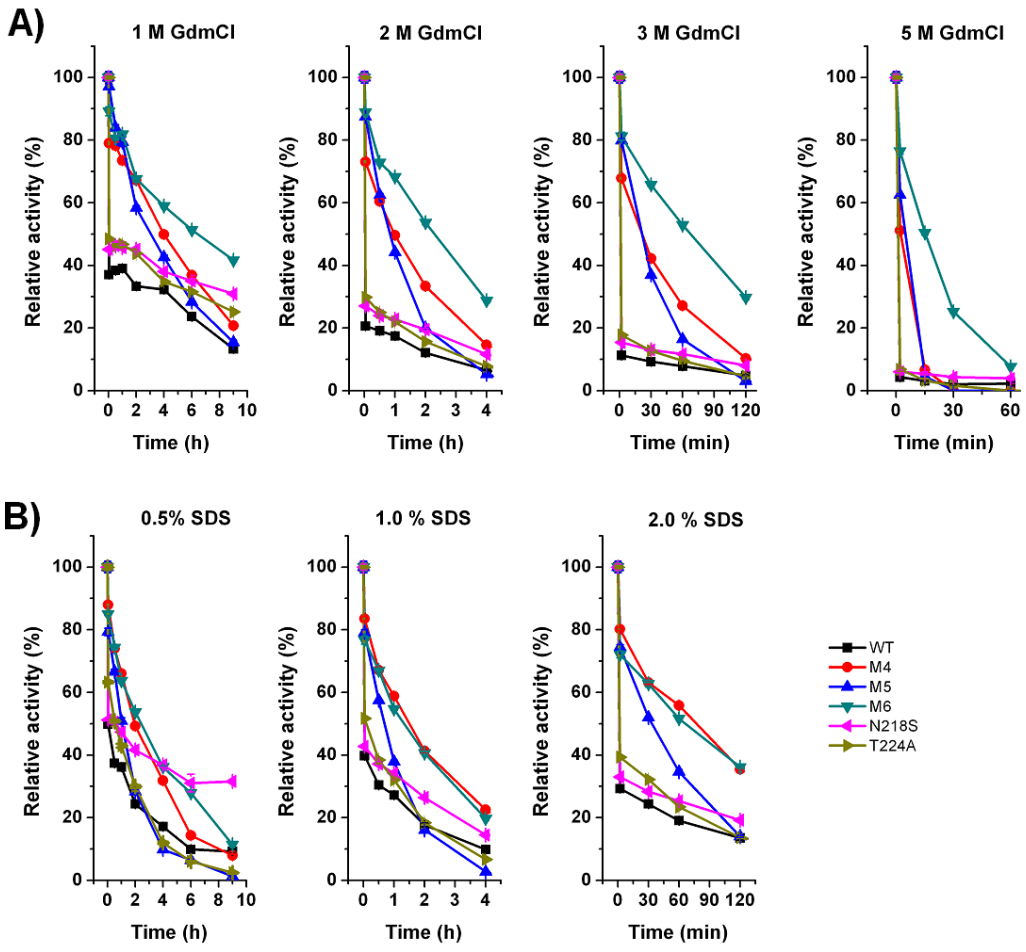

Figure 3.35 Stability of purified subtilisin E wild type and its variants (Study II) in the presence of various concentrations of GdmCl or SDS by determining residual proteolytic activity (Suc-AAPF-pNA assay, 25°C). Relative activities were measured in triplicate and averaged with <10 % standard deviation considering the activities in the absence of chaotropic agents as 100 %. Final protease concentration was 50 nM.

As seen in Table 3.36, in 5 M GdmCl, the half lives are <2 min (WT), <2 min (N218S), <2 min (T224A), 2.3 min (M4), 4.8 min (M5) and 15.2 min (M6). In 2.0 % SDS, the half lives are <2 min (WT), <2 min (N218S), <2 min (T224A), 77.3 min (M4), 33.4 min (M5), and 66.3 min (M6).

Table 3.36 Half lives of purified subtilisin E and variants (Study II) in the presence of chaotropes.

Variants	GdmCl				SDS		
	1 M (h)	2 M (h)	3 M (min)	5 M (min)	0.5% (h)	1.0% (h)	2.0% (min)
WT	<2 min	<2 min	<2 min	<2 min	2 min	<2 min	<2 min
N218S	<2 min	<2 min	<2 min	<2 min	2 min	<2 min	<2 min
T224A	<2 min	<2 min	<2 min	<2 min	0.6	0.1	<2 min
M4	4.0	1.0	17.0	2.3	2.0	1.5	77.0
M5	3.1	0.8	18.1	4.8	1.0	0.7	33.4
M6	6.4	2.3	67.6	15.2	2.4	1.3	66.3

3.2.6 Analysis of structural stability

Figure 3.37 showed the study of conformational changes of subtilisin E and its variants in GdmCl or SDS by analysis of circular dichroism (CD) spectra.

It is indicated that the structure of subtilisin E variants didn't change obviously in the presence of 1-2 M GdmCl (except M4 at 2 M GdmCl), since the curves were very similar to the one in the absence (see Figure 3.37, left column). M4 seemed more unstable in GdmCl solutions and started unfolding from concentration of 2 M GdmCl, and was completely unstructured in 5 M GdmCl, while other variants started unfolding from concentration of 3 M GdmCl, and still kept its structure partially in the presence of 5 M GdmCl. In general, the order of structural stability in the presence of GdmCl is roughly as M6 > WT > M5 > M4.

The study of protein structure in protein-SDS micelle complexes by CD has been reported [118], in which two helical proteins had a slight loss in α-helix content. However, the structure of all subtilisin E variants in this dissertation remained almost unaltered in concentrations up to 50 mM SDS. It has been reported that the structure of subtilisin Carlsberg ist not perturbated by up to 400 mM concentration of SDS [119].

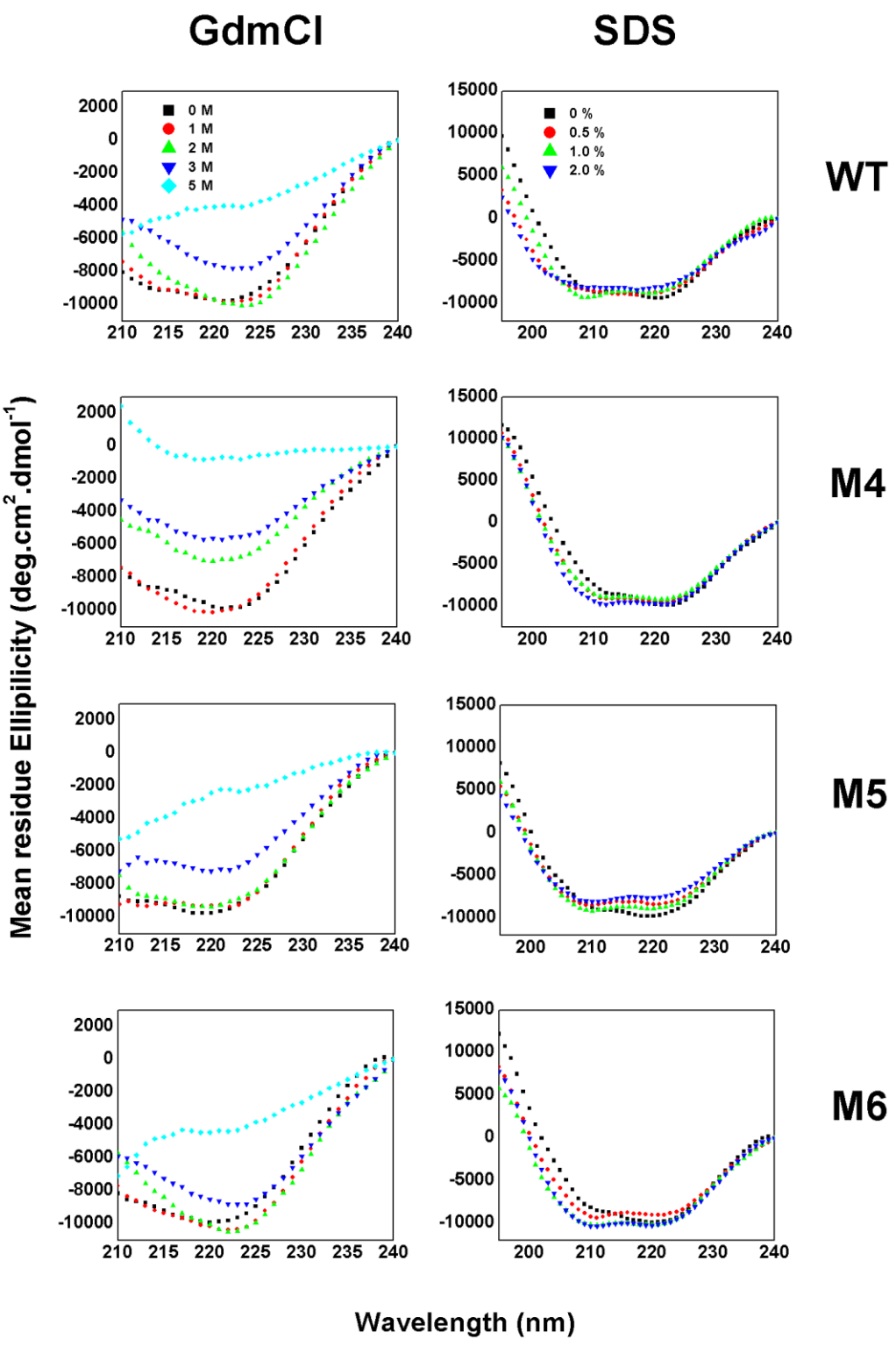

Figure 3.37 CD spectra analysis of subtilisin E and its variants in the presence of different concentrations of GdmCl or SDS. Purified subtilisin E WT, M4, M5 and M6 (300-500 ug/mL) were mixed with 0-5 M GdmCl or 0-2.0 % SDS (final concentration) and immediately measured at room temperature. The data were averaged by five scans.

The CD spectra results were not compatible with resistance of these variants in terms of IC_{50} and half life values. For instance, M4 was much more active in the presence of GdmCl compared with WT in 2 M, 3 M GdmCl (see Figure 3.30 and Figure 3.35), but CD spectra results showed that its secondary structure was unstable when compared to WT. This indicated the resistance had no direct correlation with their secondary structural stabilities. It could be speculated, at lower concentrations of GdmCl (1-2 M) or given concentrations of SDS, fast loss of activities was most likely not caused by a loss of secondary structure. However, the structural stability is still beneficial for improvement of resistance. M6, for instance, is more structurally stable in GdmCl than other variants, and showed much improved resistance (see Figure 3.30), which might be caused by the additional substitutions N218S and T224A.

3.3 Computational analysis for chaotrope resistance

This section focused on the study of the effects of GdmCl and SDS on subtilisin E by computational analysis, such as model building, molecular docking and molecular dynamics (MD) simulations. These analyses are very helpful to interpret or elucidate the mechanism indicated by the experimental results or propose some hypothesis which sheds lights on the profound mechanism of protein denaturation. However, protein denaturation is such a complex process, in which subtilisin E inactivation occurs prior to unfolding. The detailed analyses in this chapter also provide insights in how this inactivation occurs and why loss of activity could be compensated by those substitutions identified or generated in the last two sections.

3.3.1 Model building of subtilisin E-substrate complex

Different models, including wild type and other variants were built using FoldX [102]. The C-terminal of Pro-domain in the crystal structure was replaced and superimposed by the substrate Suc-AAPF-pNA in a protease-substrate manner, followed by energy minimization in vacuum (see section 2.2.14.1). This model building was an alternative of protease-substrate docking, since the attempts of docking of substrate Suc-AAPF-

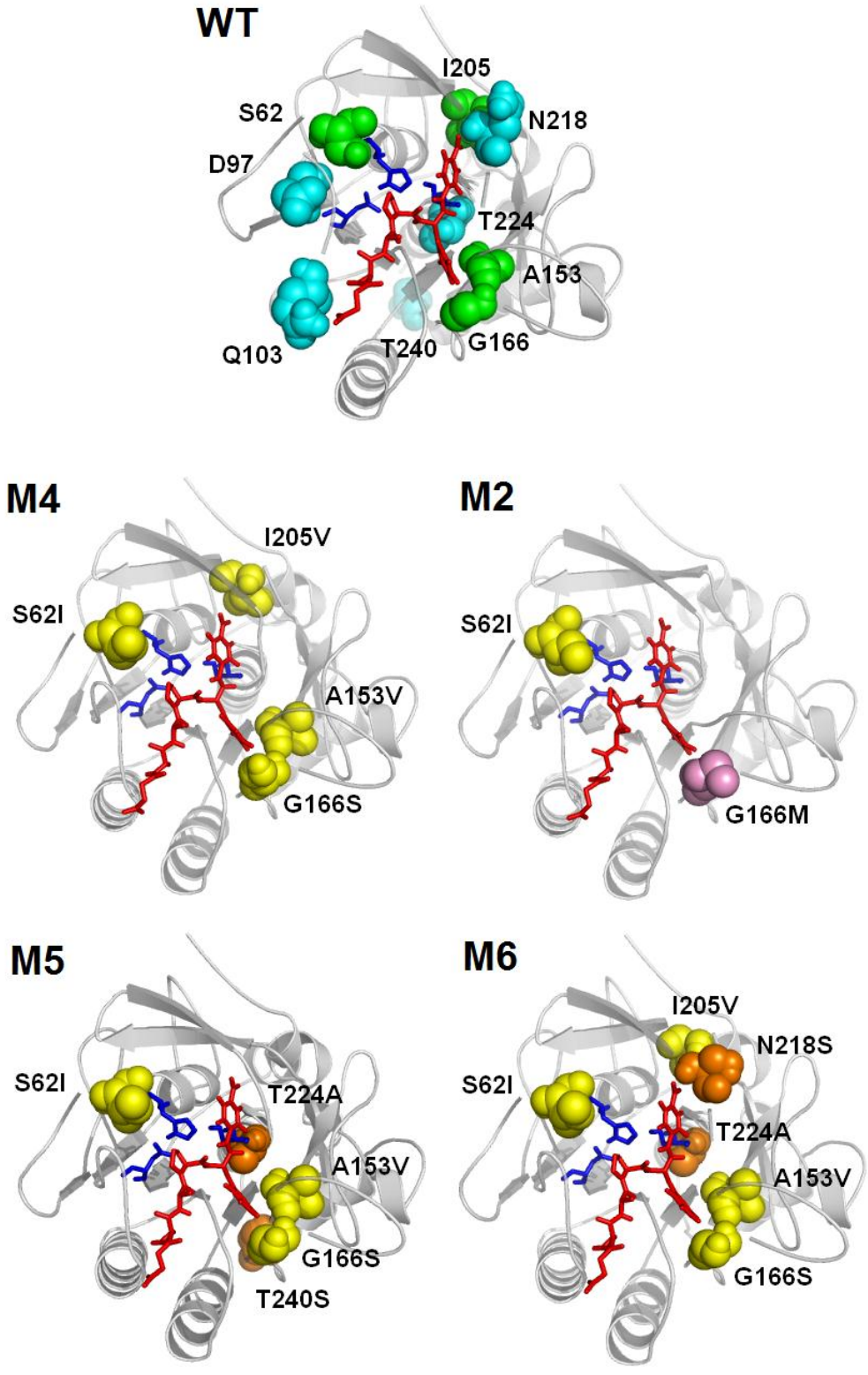

Figure 3.38 Structures of subtilisin E wild type and variants bound to chromogenic substrate Suc-AAPF-pNA. Different amino acid substitutions obtained by directed evolution are labeled with the catalytic triad (blue) and substrate (red). In model of wild type (WT), beneficial substitutions are labeled in green (Study I) or cyan (Study II). In model of variants, substitutions are labeled in yellow (derived from M4) or other colors (additional substitutions).

-pNA was not successful due to its bulky volume and flexibility. Besides, the mechanism of hydrolysis and substrate binding of such substrate was determined and well-known, which has been used to construct models [36].

Figure 3.38 displayed the models of subtilisin E-substrate complex for wild type and other variants, including SeSaM1-5 (M4), M2, M5 and M6. The conformation of substrate Suc-AAPF-pNA in all these models doesn't show much difference except for a slight rotation of the benzene ring of phenylalanine caused by a mutation at position 166. As mentioned previously, the beneficial substitutions for chaotolerance are mostly located in or close to the active site or substrate binding pockets, except T240. The substitutions at position 97 or 103 were excluded in these muteins.

3.3.2 Docking of chaotrope in subtilisin E

One molecule of chaotropic agent (guanidinium: Gdm^+ or dedecylsulfate: DS^-) was docked to the active site or substrate binding pockets of subtilisin E wild type and variants to investigate their most probable binding sites, in order to better understand their inactivation role as well as the difference between wild type and variants in respect to chaotolerance.

3.3.2.1 Docking of guanidinium in Subtilisin E

Docking of a single Gdm^+ in the region of the active site and binding pockets revealed that Gdm^+ was more prone to interact with essential residues in the active site or substrate binding pockets in wild type than other variants. The differences are most likely caused by substitutions in the variants.

Figure 3.39 showed two putative binding modes of Gdm^+ in subtilisin E wild type and variants which ranked at first and second place in terms of their binding energy. Only in wild type, as shown in Conformation 1 in WT (Figure 3.39), the position shift of the active site residue H64 was observed. In addition, Gdm^+ imbedded into the catalytic triad so that the active site was deformed. Conformation 2 in wild type showed the second possibility of binding site of Gdm^+ which is located at the

specificity binding pocket and imbeded into the cleft. Very interestingly, in the single mutant S62I, imbedding of Gdm^+ into the catalytic triad was not observed but in the specificity binding pocket cleft. However, in single mutant G166S, the opposite was observed. Additionally, both imbeddings were missing in muteins M4 and M6 which have both substitutions. These docking results highly indicate that most strong binding sites are either located in the catalytic triad or specificity binding pocket P1, and those substitutions at position 62 and 166 are beneficial for GdmCl tolerance. The latter is most likely caused by reducing Gdm^+ binding by structural extension of their side chains. These results are in accordance with experimental results. Docking of other variants, such as N218S and T224A, resulted in similar observations (data not shown), which indicates other substitutions are not as effective as those discussed above. This is also supported by experimental results.

3.3.2.2 Docking of dodecylsulfate in Subtilisin E

Dodecylsulfate (DS^-), as an amphiphilic anion of SDS, is supposed to have very different effects on proteins compared to Gdm^+, like electrostatic interaction of its head group with charged residues and hydrophobic interaction of its tail group with hydrophobic residues as expected. However, experimental results indicated that the most beneficial substitutions for SDS tolerance (S62I and G166S) were, interestingly, at the same positions as for GdmCl resistance, located around the active site or substrate binding pocket. Results of docking of DS^- on subtilisin E wild type and variants in this region showed less evident difference between wild type and variants than in the case of Gdm^+ docking. This is most probably due to large size of ligand molecule and the highly flexible hydrophobic tail, which makes docking of DS^- difficult to trace highly accurate conformation. However, it could be observed that the DS^- molecule is slightly more away from the active site in variants compared to wild type (see Figure 3.40 Conformation 1). MD simulations are in this case recommended to investigate more details of SDS interaction with subtilisin E.

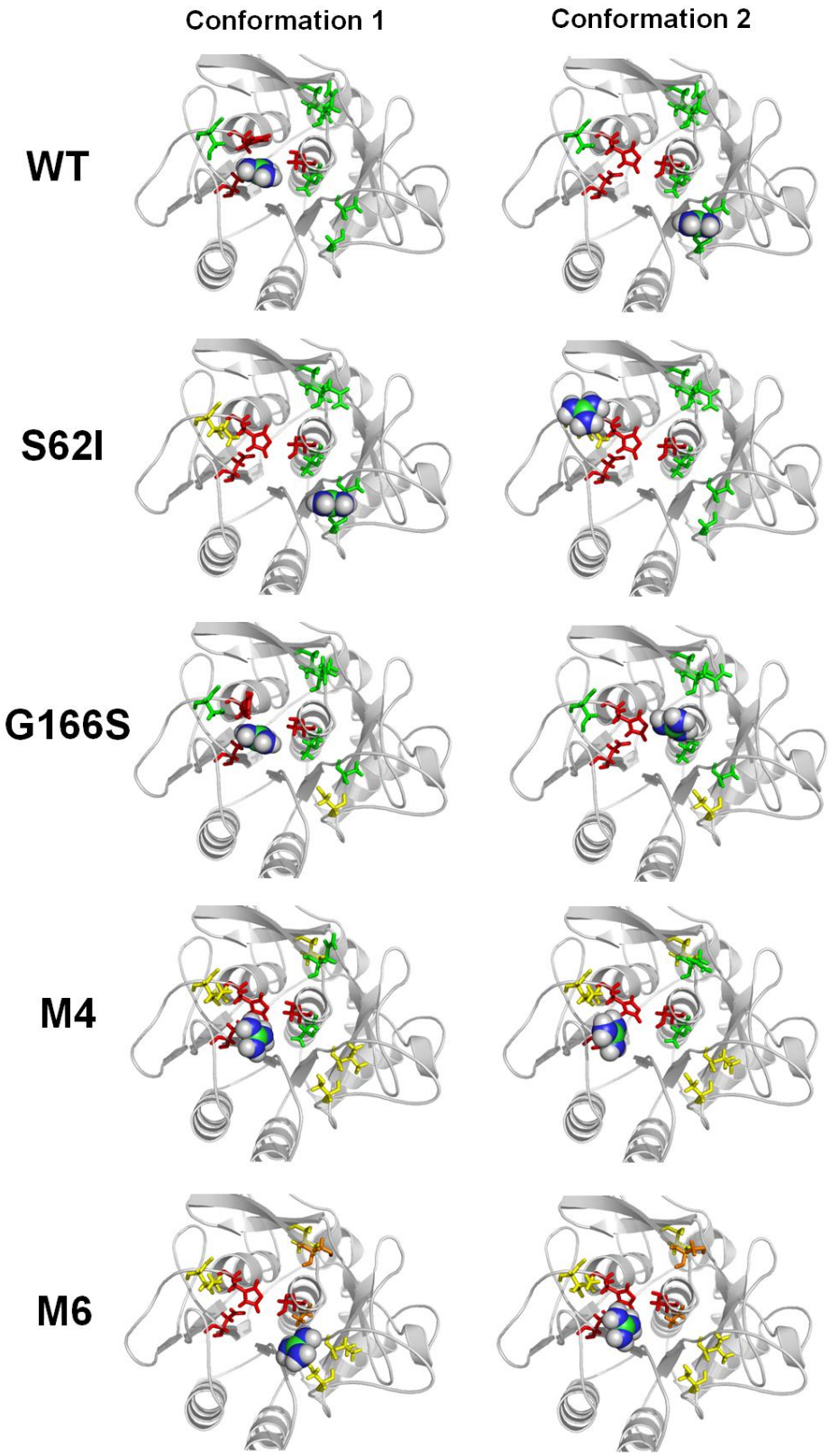

Figure 3.39 Conformation of Gdm^+ binding to subtillsin E and its variants in the active site or substrate binding pockets after docking using Yasara program. Conformation 1 & 2 are two most possible binding site of Gdm^+ (in sphere form) ranked by binding energy. Catalytic triad residues (D32, H64 and S221) are displayed in red. Substitutions are displayed in green (wild type), yellow (substitutions in study I) or orange (substitutions in study II).

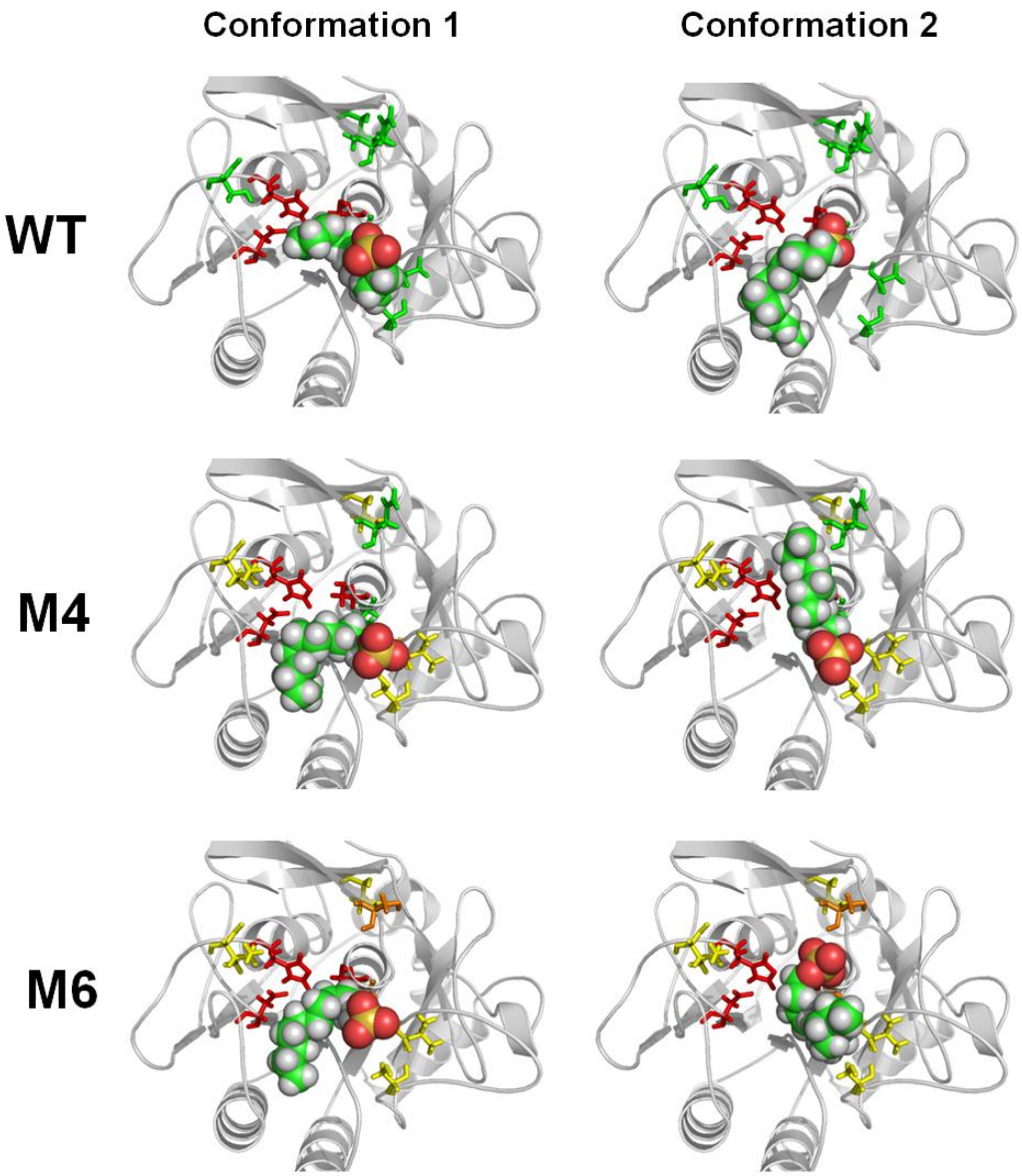

Figure 3.40 Conformation of DS^- binding to subtillsin E and its variants in the active site or substrate binding pockets after docking using Yasara program. Conformation 1 & 2 are two most possible binding site of dedecylsulfate (in sphere form) ranked by binding energy. Catalytic triad residues (D32, H64 and S221) are displayed in red. Substitutions are displayed in green (wild type), yellow (substitutions in study I) or orange (substitutions in study II).

3.3.3 MD simulations of GdmCl aqueous solutions

Molecular dynamics (MD) simulations were performed to investigate effects of chaotropic salts, like GdmCl on both one model peptide melittin and the protease subtilisin E in order to understand its mechanism of inactivation and comparison to experimental results.

3.3.3.1 Alpha-helical peptide melittin

A MD simulation of the model peptide melittin (26 amino acids) in 3 M GdmCl has been reported using CHARMM force field [73]. The simulation analysis indicated that Gdm^+ can interact with a number of planar amino-acid side chains (Arg, Trp, Gln) in a stacking manner, as well as more weakly with hydrophobic surfaces composed of aliphatic side chains, and that these interactions result in enhanced number densities of Gdm^+ at certain locations on the peptide surface.

Similar MD simulations of this model peptide melittin were run in different concentrations of GdmCl aqueous solution using the package for MD simulations Gromacs and GROMOS96 force field (see section 2.2.14.3) for comparison to the previous CHARMM simulation [73] and also to further provide insights into the effect of the concentration of Gdm^+ on this peptide at a molecular level.

Therefore, MD simulations of melittin in water and three different concentrations of GdmCl (1 M, 3 M and 5 M) were run for 10 ns including 1 ns of equilibration. Root mean square deviations (RMSD) analysis (see Figure 3.41) showed conformation of mellitin in GdmCl was less stable than in water (average RMSD in water 0.22 nm, 1 M GdmCl 0.30 nm, 3 M GdmCl 0.28 nm, 5 M GdmCl 0.27 nm) during simulations (last 5 ns). It is noticed that before 5 ns, RMSD was unexpectedly higher when melittin was in lower concentrations of GdmCl. After 5 ns, RMSD at all concentrations of GdmCl were very similar. In general, the conformation of melittin is less stable in molar concentrations of GdmCl solution than in water, but no rapid unfolding was observed (<0.5 nm).

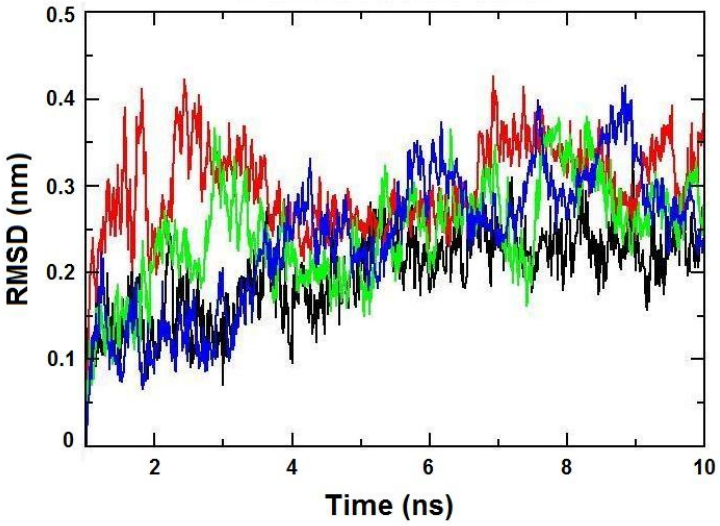

Figure 3.41 RMSD analysis of a model peptide melittin in the presence of water (black), 1 M GdmCl (red), 3 M GdmCl (green) and 5 M GdmCl (blue) with respect to starting structure (backbone). The simulations were run at 298 K for 10 ns, in which the first 1 ns was used for annealing and the remaining 9 ns for analysis.

Secondary structure analysis revealed that GdmCl could partially unfold the α-helix of melittin during simulations (see Figure 3.42). Moreover, unfolding effect is more evident at higher concentration, though the helix still remains. A similar unfolding effect was observed for the peptide, magainin (23 amino acids), in different concentrations of GdmCl solution during MD simulations (GROMOS96 force field) [77]. However, the concentration effects of GdmCl using CHARMM force fields were to date not reported.

Strong interactions between Gdm^+ and hydrophobic amino acids were observed throughout simulations at all concentrations of GdmCl, such as Trp19 (see Figure 3.43) with respect to their distance. However, stacking with guanidine group of Arg was not observed in all cases which were different from the report using CHARMM force field [74].

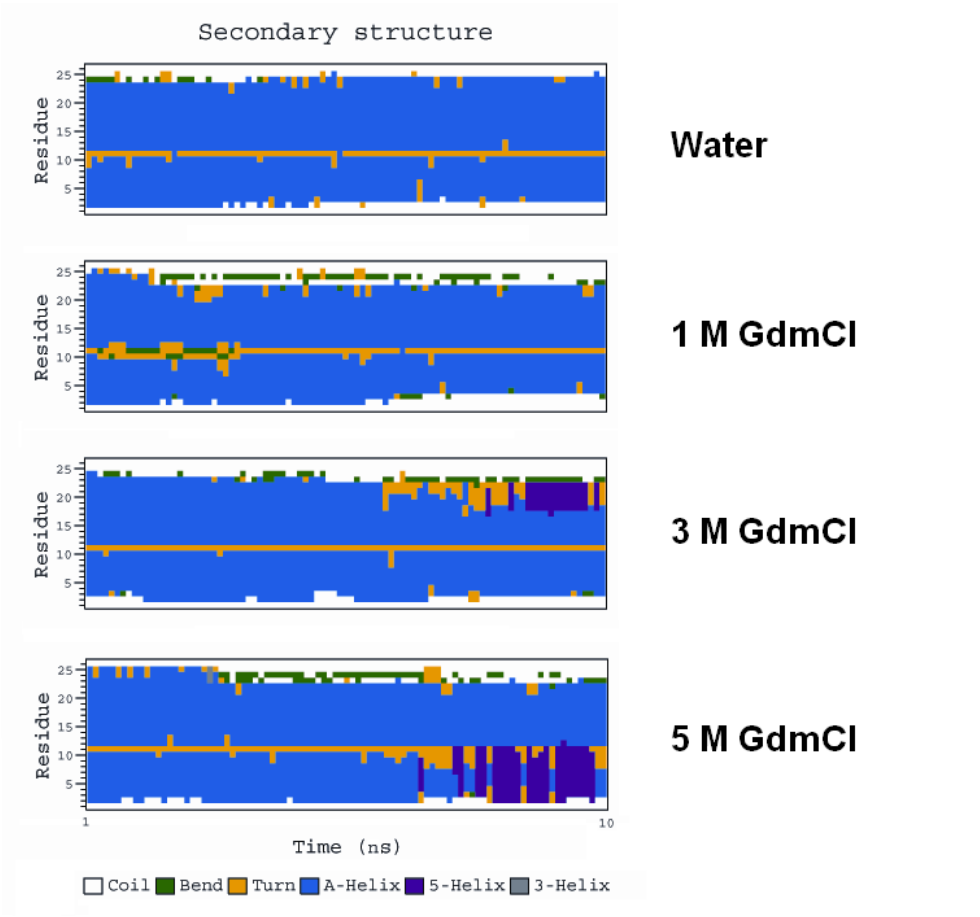

Figure 3.42 Secondary structure analysis of melittin in water and different concentrations of GdmCl during MD simulations at 298 K by DSSP program [120].

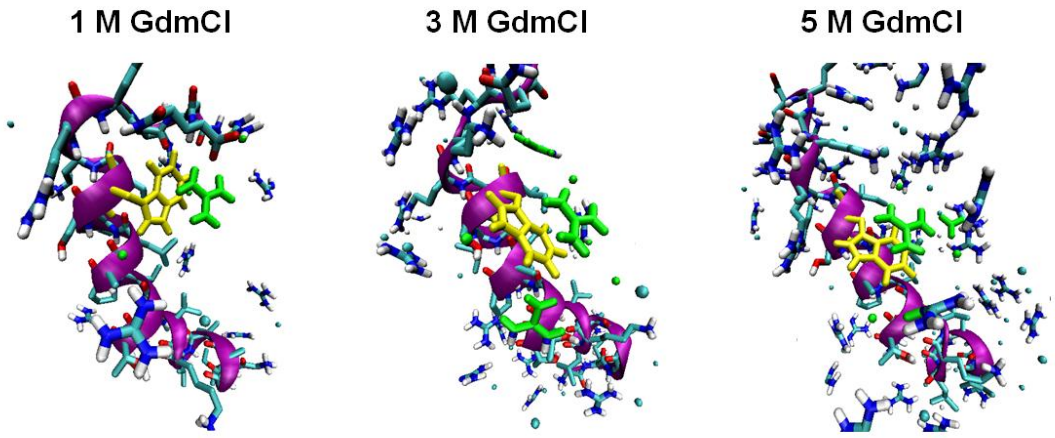

Figure 3.43 Snapshot of melittin structure throughout MD simulations in different concentrations of GdmCl with highlighted Trp19 (yellow) interaction with one Gdm^+ ion in a stacking manner. Molecules within 0.5 nm of residue Trp19 are highlighted in green.

3.3.3.2 Subtilisin E

A 50 ns long MD simulation running on subtilisin E in both pure water and 5 M GdmCl were performed in this work. In Figure 3.44 A, the backbone root mean square deviation (RMSD) values are reported in respect to the crystal structure for both simulations. In both simulations, RMSD curves reach an average plateau value of ~0.15 nm. Furthermore, the analysis of the RMS deviations and fluctuations per residue (see Figure 3.44 B&C) show that the protein residues undergo a similar average conformation change and have similar motilities in both water and GdmCl solution with the only exception for the loop region 20-26. The apparent stability of the protein in the GdmCl simulation can be explained considering the expected time scale of the unfolding process. In fact, in a recent theoretical study, the unfolding of a lysozyme mutant (less stable than the wild type) in 8 M urea required a 1000 ns long MD simulations at 310K [80]. Considering also the recent experimental evidences for a more direct and faster mechanism of urea-induced than GdmCl-induced unfolding [121], it is unlikely to observe a GdmCl mediated unfolding event in this 50 ns long simulations. Nevertheless, it was evidenced from experimental data on many enzymes that the disturbance of GdmCl at the active sites occurred prior to conformational changes of the enzyme as a whole [122]. Gdm^+ binding to subtilisin E in close proximity to the active sites can occur in time scales shorter than denaturation. Therefore our MD simulations can provide details for Gdm^+ binding on subtilisin E.

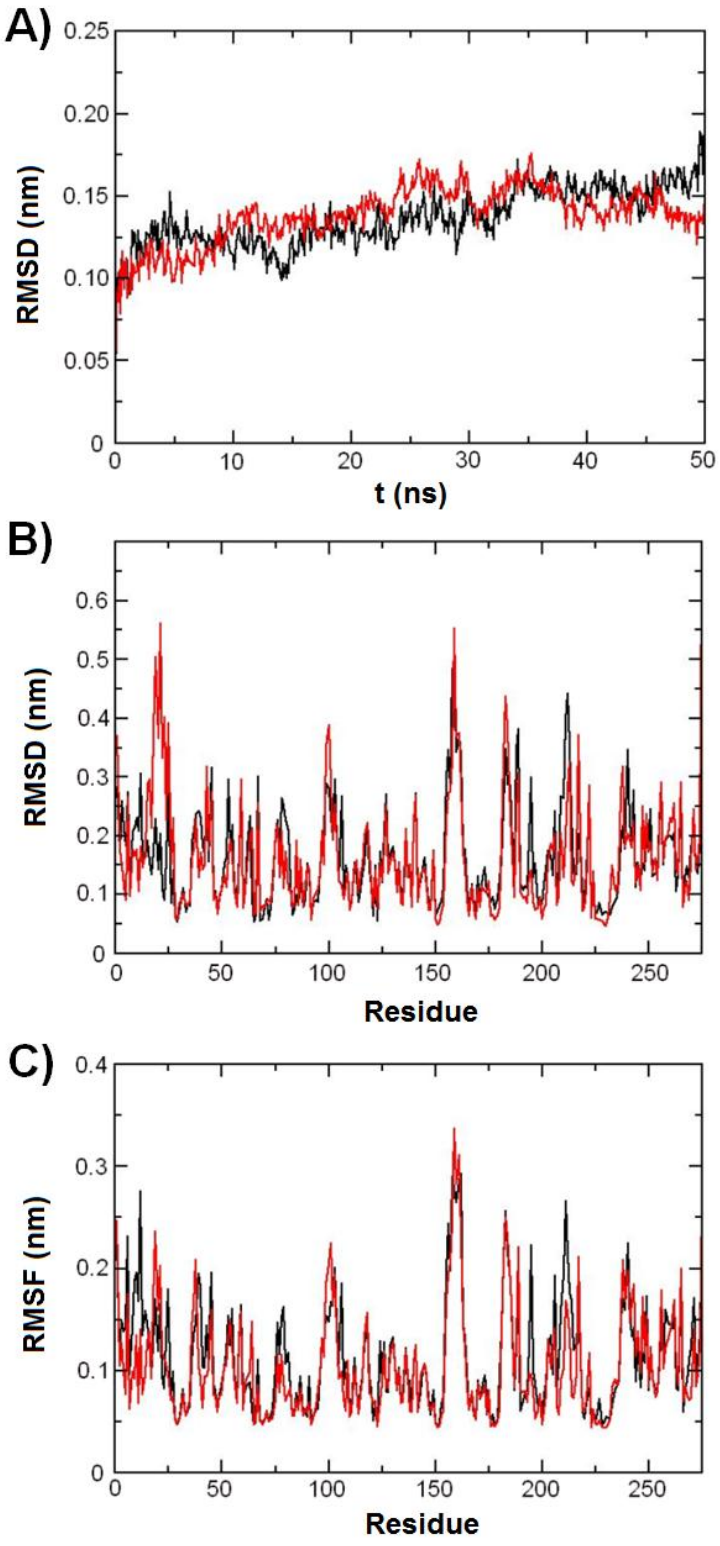

Figure 3.44 Analysis of MD simulations of subtilisin E wild type in pure water (black) and 5 M GdmCl (red) at 298K for 50 ns. A) The root mean square deviations (RMSDs) with respect to starting structure (backbone). B) RMSD analysis per residue. C) The root mean square fluctuation (RMSF) analysis per residue.

In Figure 3.45, the representation of spatial density function of Gdm^+ ions is reported. The blobs in Figure 3.45 A, represent isosurfaces with high Gdm^+ density and mainly localize in the active sites around the catalytic triad. The analysis of binding process during the simulation evidenced the following sequence of events. A first Gdm^+ ion diffuses into the active site (after 21.8 ns of simulation) and binds to the catalytic triad residue by facing the imidazolium ring towards His64 and by forming two strong hydrogen bonds, one with Ser221 and one with Asp32 (see Figure 3.45 B). As shown in the same figure, the presence of Gdm^+ in the active site weakens the hydrogen bonding between His64:Nε2 and Ser221:Oγ which is essential for proteolytic activity. Two additional Gdm^+ ions are observed to diffuse in the active site during the simulation time (50 ns; see Figure 3.45 B) surrounding the active site residue in a similar way as the first Gdm^+. Besides, Gdm^+ ions are also observed to accumulate in the substrate binding pocket P1, where residue Gly166 is located (see Figure 3.46). Gdm^+ ions might interact with Gly166 as well blocking the substrate specificity pocket of subtilisin E. Interestingly, all these observation are consistent with the results from docking (see section 3.3.2.1).

Furthermore, these results suggest that Gdm^+ ions are strongly bound to the active site residues weakening H-bond networks (see Figure 3.45 B) which might cause a very fast inactivation of subtilisin E and provides a possible explanation for increased tolerance of the variant S62I whose substitution is located next to His64, which is also discussed in previous sections (see section 3.3.1 and section 3.3.2).

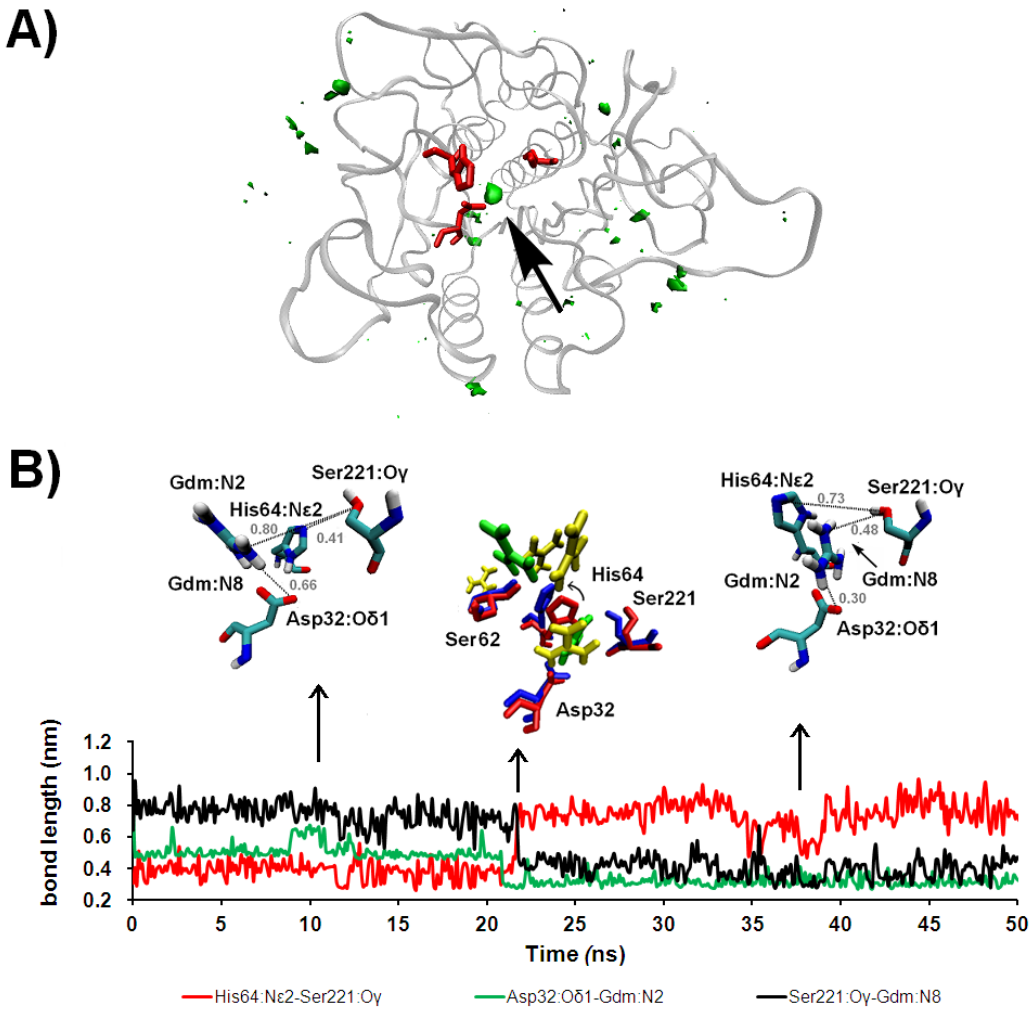

Figure 3.45 Analysis of Gdm^+ coordination in MD simulation of wild type subtilisin E (5 M GdmCl). A) Active site region of wild type subtlisin E with distribution of Gdm^+ ions during last 25 ns simulation time by using a g_spatial program. Catalytic triad residues (Asp32, His64 and Ser221) are labeled. The Gdm^+ counters in the active site of subtilisin E are highlighted (arrow). B) Changes in the lengths of hydrogen bonds between Gdm^+ (Nitrogens, N2 and N8) and Asp32 (Oδ1), Ser221 (Oγ), and between His64 (Nε2) and Ser221 (Oγ) during the 50 ns simulation. Three snapshots of catalytic triad with one Gdm^+ ion (left, 10.8 ns; right, 37.6 ns) show a transition state after 21.8 ns (middle) that results in Gdm^+ bond-flip to Asp32 and Ser221 at the active site. In the snapshot of transition state (middle), catalytic triad and residue 62 are labeled and displayed in red (at 21.8 ns, before transition) and blue (at 21.9 ns, after transition). Adjacent Gdm^+ ions are displayed in yellow (at 21.8 ns) and green (at 21.9 ns).The arrow points out the transition switch of His64.

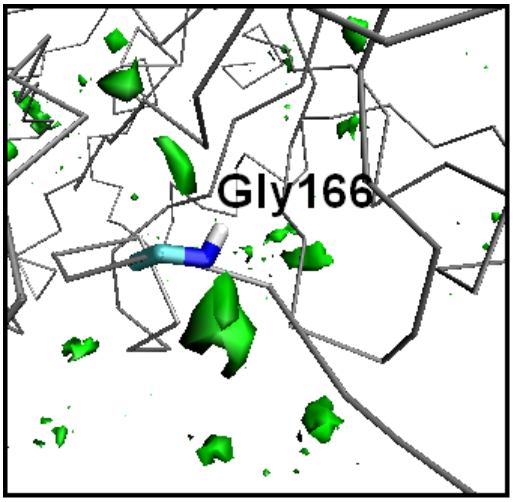

Figure 3.46 Binding pocket 1 region (P1) of wild type subtlisin E with distribution of Gdm^+ ions during last 25 ns simulation by using a g_spatial program. The subtilisin E backbone is displayed in gray. Residue Gly166 is displayed in licorice format and labeled. The Gdm^+ counters are displayed in solid surface format in green.

Finally, variation on the secondary structure distribution over simulation time was analyzed (see Figure 3.47) which showed no evident changes of α-helix and β-sheet during 50 ns simulations in water and 5 M GdmCl, which is different from the analysis reported for the model peptide melittin (see Figure 3.42). This might be due to the size of protein (melittin, 26 aa; subtilisin E, 275 aa) and limited simulation time scale (nanosecond scale). MD simulations at elevated temperature, such as 350K or 400K, could accelerate destabilization of secondary structure as reported for another small protein Protein L (62 aa) [74].

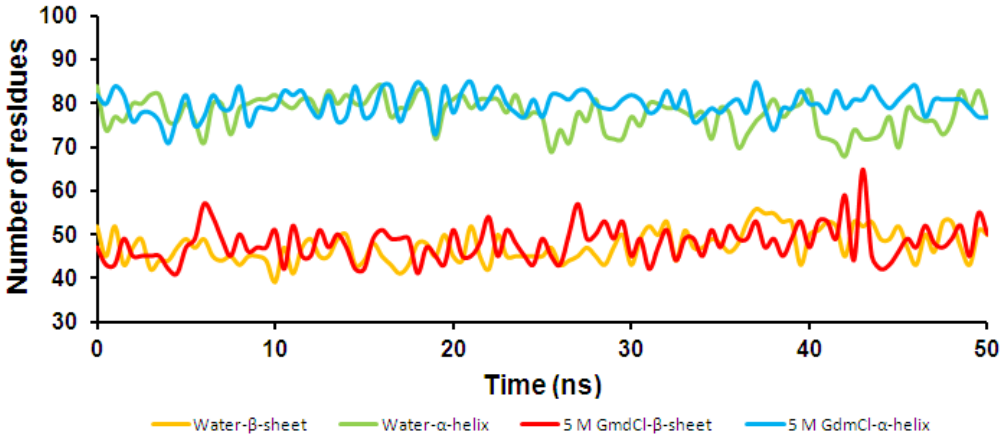

Figure 3.47 Secondary structure analysis of wild type subtilisin E in water and 5 M GdmCl during MD simulations at 298K for 50 ns by DSSP program.

MD simulations of subtilisin E wild type at three different concentrations were also performed to investigate their effects. Coordinate of subtilisin E were taken for setup of these simulations after equilibrium in pure water for 25 ns at 298K in order to exclude potential inappropriate structure conformation in the crystal structure, since this wild type model was generated from an inactive Pro-subtilisin E complex. Figure 3.48 A showed that the RMSD of subtilisin E wild type under all conditions converged at ~0.15 nm. However, the RMSD in 1 M GdmCl is very similar to the one in water, while RMSD in 3 M GdmCl keeps constant around ~0.12 nm after 7 ns. Different from 1 M and 3 M GdmCl simulations, RMSD gradually increases in the case of 5 M GdmCl. RMSF analysis shows that fluctuation of residues is almost the same in water and 1 M GdmCl, while fluctuation modes are similar in 3 M and 5 M GdmCl (see Figure 3.48 B). Additionally, the radius of gyration (Rg) slightly decreases in all cases, but decreases less slowly in higher concentrations of GdmCl (see Figure 3.48 C), indicating the compactness of conformation throughout MD simulations.

Spatial distribution function (SDF) analysis of Gdm^+ ions throughout all simulations of wild type subtilisin E in 1 M, 3 M and 5 M GdmCl showed a constant a blob around the catalytic triad residues regardless of GdmCl concentrations (see Figure 3.49 left column). The distance between such a blob and the center of active site is slightly shorter in higher concentration of GdmCl (see Figure 3.49 right column). This can be considered as a strong indication that strong binding of Gdm^+, even at a lower concentration. It could be the reason for an inactivation of subtilisin E, and the strong binding occurrs on initial phase which also reveals that such inactivation can be extremely fast. Besides, it is also observed that higher concentrations could evoke such binding earlier than lower concentrations (1 M: ~5.0 ns; 3 M: ~3.5 ns; 5 M: ~1.0 ns) indicated by monitoring the distance between a strong binding Gdm^+ and the center of catalytic triad (see Figure 3.50) during simulations. All these observations are supported by experimental results in section 3.2.5.

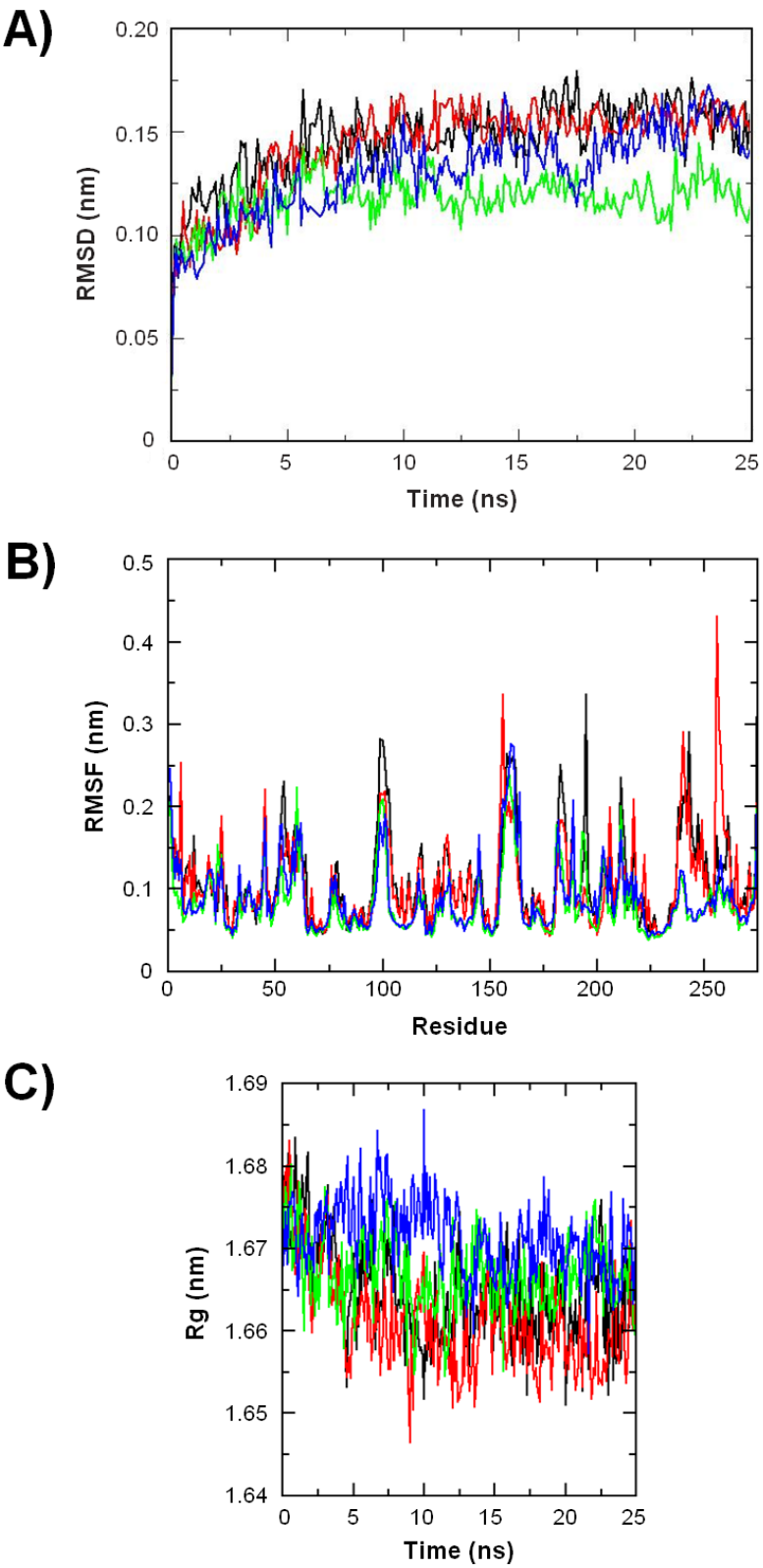

Figure 3.48 Analysis of MD simulations of subtilisin E wild type in pure water (black), 1 M GdmCl (red), 3 M GdmCl (green) and 5 M GdmCl (blue) at 298K for 25 ns. A) RMSD with respect to starting structure (backbone). B) RMSF analysis per residue. C) Radius of gyration (Rg) analysis.

It is noticed that strong bounded Gdm^+ could be also released after a certain time, as seen in Figure 3.50 for simulations in 1 M and 3 M GdmCl, which indicates a competitive inhibition effect at lower concentrations of GdmCl onsubtilisin E. However, imbedding of Gdm^+ into the active site did not reposite His64, as shown in Figure 3.45. These new simulations were performed at varied parameters in a shorter time (25 ns).

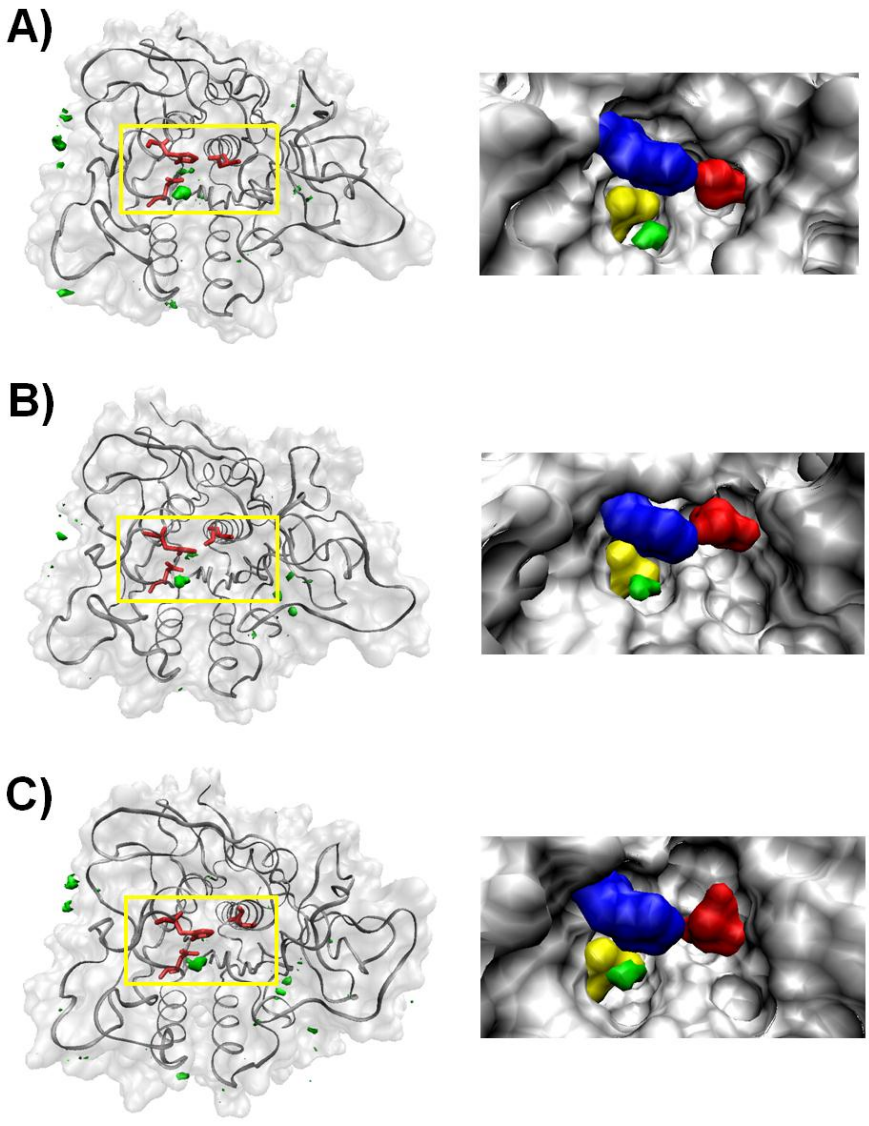

Figure 3.49 Analysis of Gdm^+ coordination in MD simulations of wild type subtilisin E in different concentrations of GdmCl. A) 1 M GdmCl; B) 3 M GdmCl; C) 5 M GdmCl. The left column displayed wild type subtlisin E (backbone in grey) with distribution of Gdm^+ ions during last 20 ns simulation time by using a g_spatial program. Catalytic triad residues are highlighted in red. The Gdm^+ counters in the active site of subtilisin E are highlighted in the yellow box and zoomed in right column, where catalytic triad residues are displayed in yellow (Asp32), blue (His64) and red (Ser221).

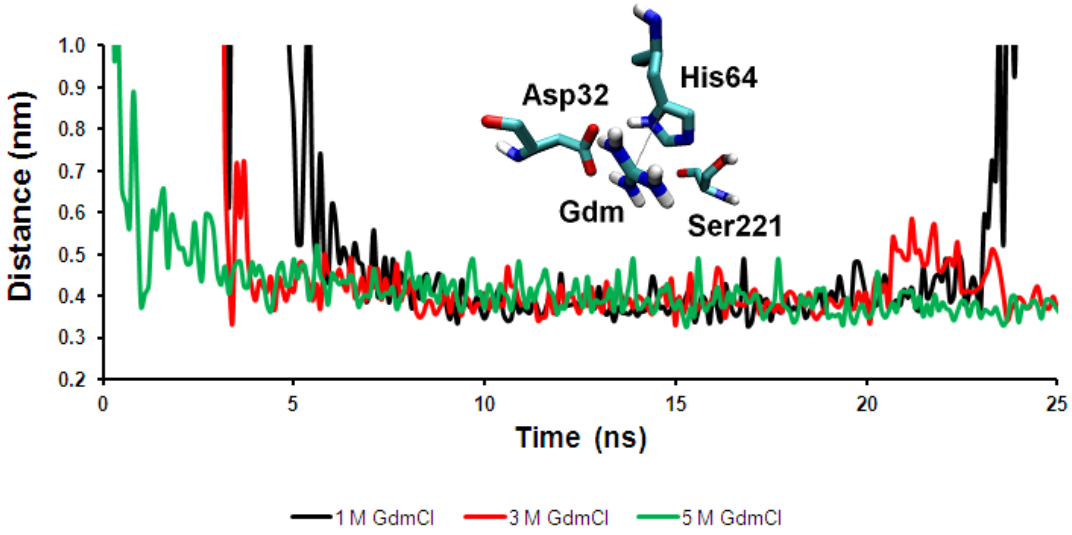

Figure 3.50 Distance between one strong binding Gdm^+ and center of catalytic triad during MD simulations of wild type subtilisin E in different concentrations of GdmCl. The distance was measured between carbon atoms of Gdm^+ and His64 as showed in a snapshot above the curve.

The tetramutant M4 was also considered for MD simulations in both pure water and 5 M GdmCl to investigate the effects of substitutions on subtilisin E tolerance towards GdmCl on the molecular level. RMSD analysis of M4 in water and 5 M GdmCl for 25 ns is compared to those of wild type (from 50 ns simulations). Figure 3.51 shows there are almost no differences in RMSD during 25 ns simulations in wild type or M4 in water and 5 M GdmCl, indicating that there would be no conformation differences between wild type and M4 in respect to their structure as a whole.

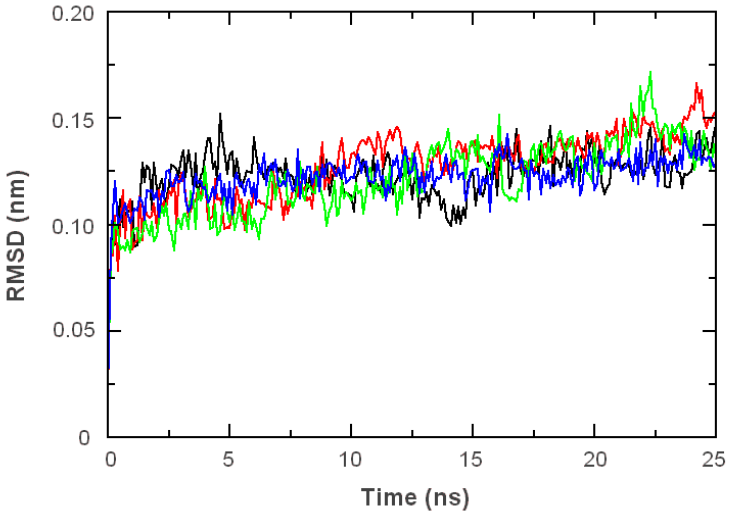

Figure 3.51 RMSD analysis of MD simulations of subtilisin E variants with respect to starting structure (backbone) in water (WT: black; M4: green) and 5 M GdmCl (WT: red; M4: blue).

RMSF analysis shows that fluctuation of both WT and M4 in water is almost the same, and M4 has less fluctuation in many regions, like Loop2, except Loop1 (see Figure 3.52 A). Conformations of both WT and M4 from each 2.5 ns were clustered as showed in Figure 3.52 B, which shows very obviously the difference of fluctuation in Loop1 and Loop2 in both variants along 25 ns simulations. It is supposed that such difference is most likely caused by mutations at position 166 in Loop2, as discussed previously by extension of side chain to expel Gdm^+ binding, but however also affect the flexibility of Loop1.

Furthermore, SDF analysis of Gdm^+ binding in Figure 3.53 showed a blob of Gdm^+ were also found around the catalytic triad. However, Gdm^+ binding seems to be blocked due to “shrinked” local surroundings so that Gdm^+ couldn’t further reposite residues in the active site. Besides, His64 becomes more rigid in M4 due to substitution S62I, while substitution G166S might have effects on Loop1 to indirectly interference the binding of Gdm^+ to the active site, as mentioned previously.

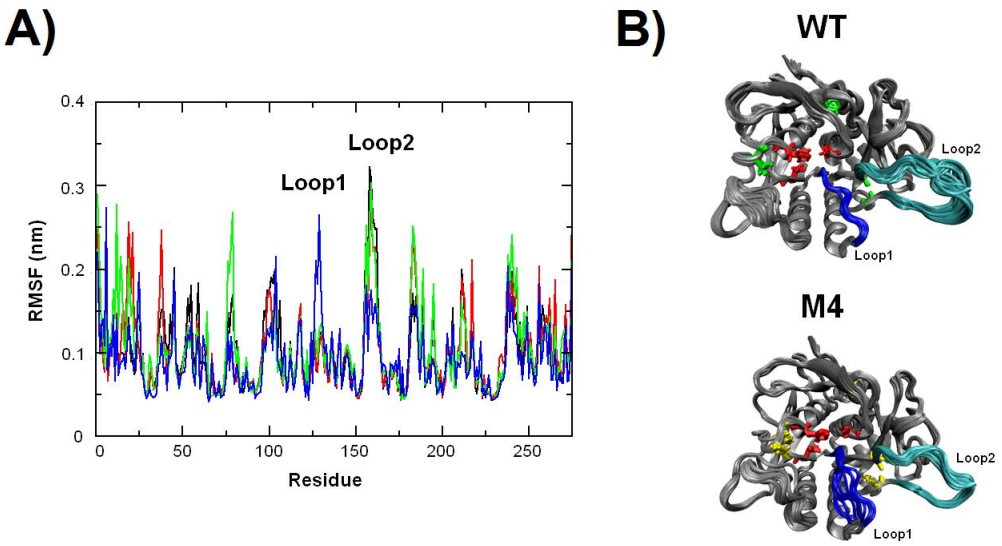

Figure 3.52 Fluctuation analysis of MD simulations of subtilisin E variants in water (WT: black; M4: green) and 5 M GdmCl (WT: red; M4: blue). A) RMSF. B) Cluster of structures of subtilisin E variants in 5 M GdmCl. Each 2.5 ns frame was aligned in a cluster. Loop1 (residue 125-130) and Loop2 (residue 155-165) are labeled. Catalytic triad residues are displayed in red, and positions of substitutions are highlighted in green (WT) or yellow (M4).

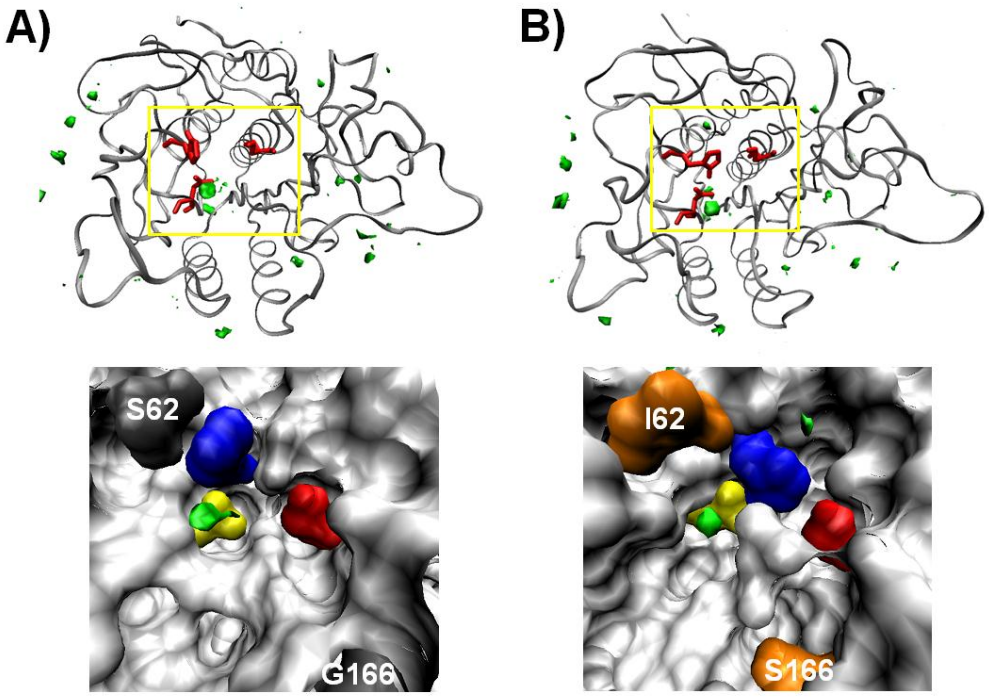

Figure 3.53 Analysis of Gdm^+ coordination in MD simulations of subtilisin E variants in 5 M GdmCl. A) WT; B) M4. Top row displayed wild type subtlisin E (backbone in grey) with distribution of Gdm^+ ions during 25 ns simulation by using a g_spatial program. Catalytic triad residues are highlighted in red. Gdm^+ counters in the active site of subtilisin E are highlighted in the yellow box and zoomed in the lower row, where catalytic triad residues are displayed in yellow (Asp32), blue (His64) and red (Ser221). The visible substitutions were also labeled.

3.3.4 MD simulations of subtilisin E in the presence of SDS aqueous solutions

MD simulations were performed on subtilisin E wild type in the presence of SDS molecules. Two different concentrations at 1 % (w/v) and 5 % (w/v) were taken into account in order to investigate SDS concentration effects on subtilisin E. As seen in Figure 3.54, the RMSD in water and two SDS solutions show a similar trend with the two curves from SDS slightly lower than the one from the water simulation. Interestingly, the same trend was observed in the case of GdmCl simulations regardless of different chaotrope type and their concentration (see section 3.3.3.2).

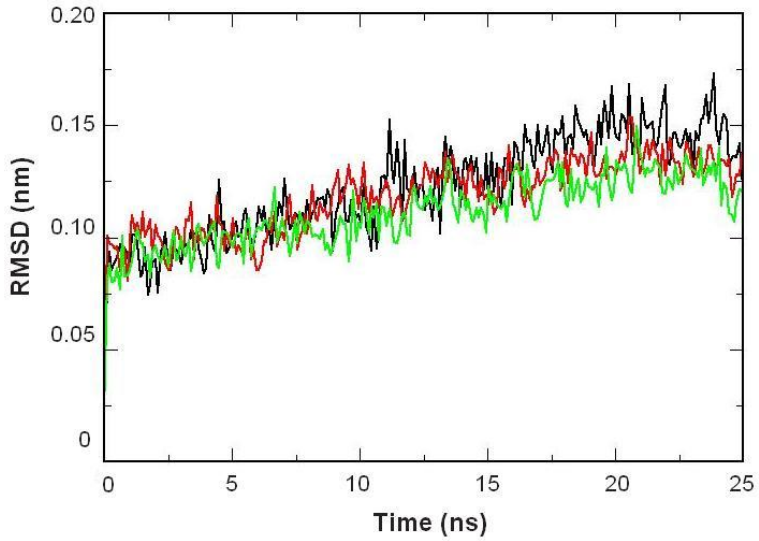

Figure 3.54 RMSD analysis of MD simulations of subtilisin E wild type with respect to starting structure (backbone) in water (black), 1 % SDS (red) and 5 % SDS (green).

The RMSF per residue (see Figure 3.55 A) reveal there are almost no changes in the residue flucturation in water, 1 % and 5 % SDS simulations. Essential dynamic analysis [113] was performed to further investigate the type of collective modes present in different proteins. As seen in Figure 3.55 B, three most representative motion modes (first 3 eigenvectors) were taken for comparison, which reveals in general that three essential motions are different in the three simulations.

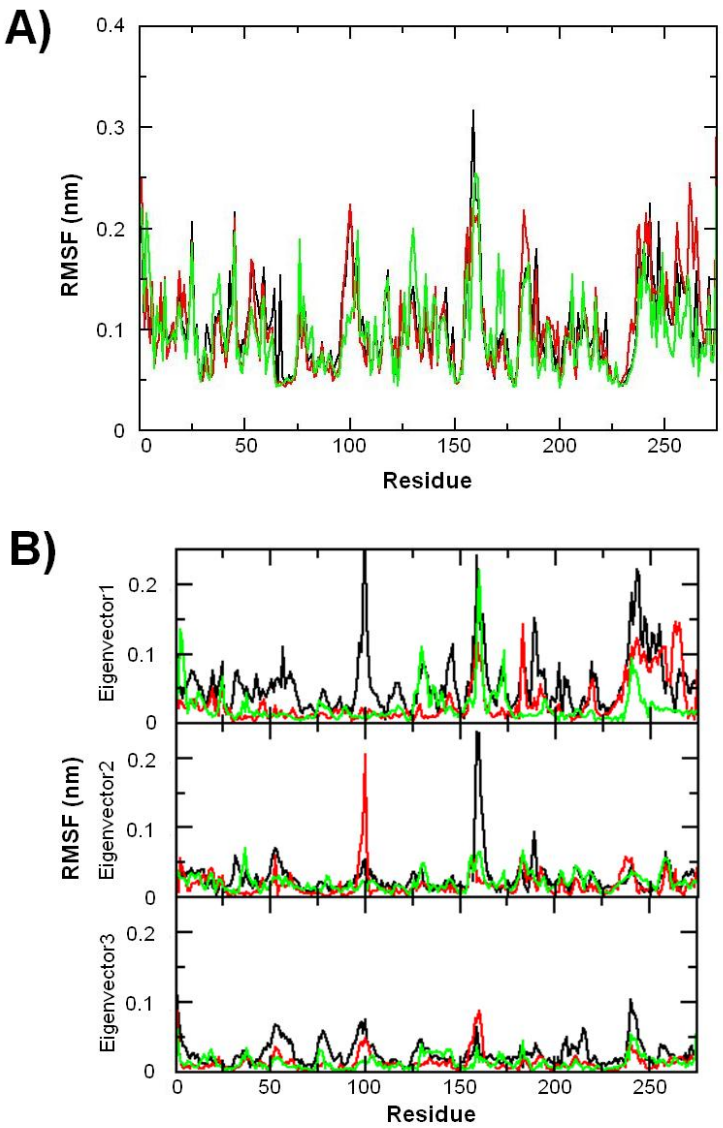

Figure 3.55 RMSF analysis of MD simulations of subtilisin E wild type in water (black), 1 % SDS (red) and 5 % SDS (green). A) RMSF per residue with respect to starting structure (protein). B) RMSF per residue obtained by projecting trajectories on the corresponding first 3 eigenvectors.

Secondary structure analysis shows that there are still no evident changes in simulations of subtilisin E wild type in water and SDS solutions (see Figure 3.56).

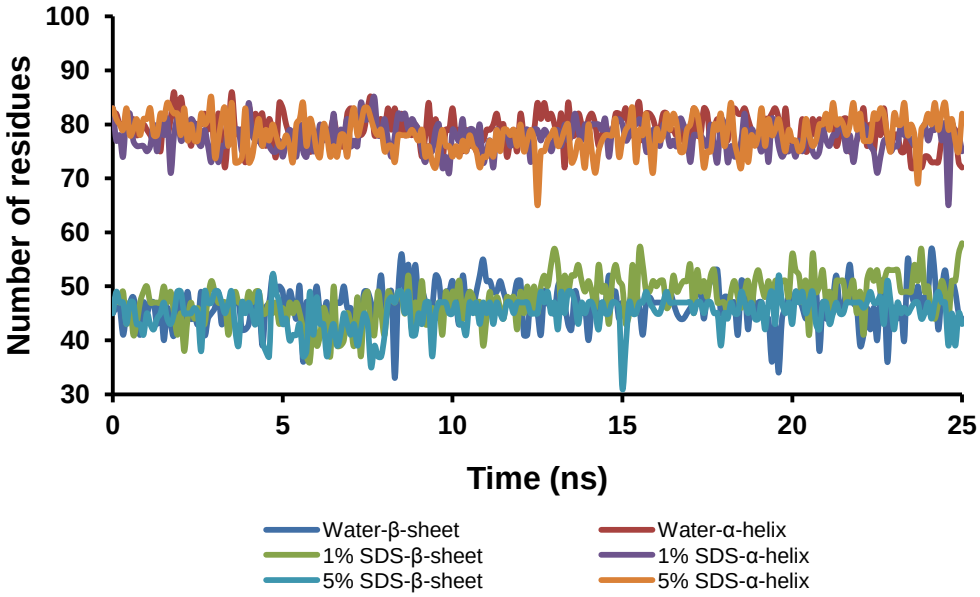

Figure 3.56 secondary structure analysis of wild type subtilisin E in water, 1 % SDS and 5 % SDS during MD simulations at 298K for 25 ns analysed by DSSP program.

When monitoring the minimum distances between all SDS molecules and protein, it is observed that in SDS simulations at both concentrations almost all SDS molecules were bound to the protein (< 0.6 nm) rapidly (in 5 ns). Additionally, all SDS molecules were strongly bound and didn't separate from the protein again, especially in 1 % SDS simulation (see Figure 3.57).

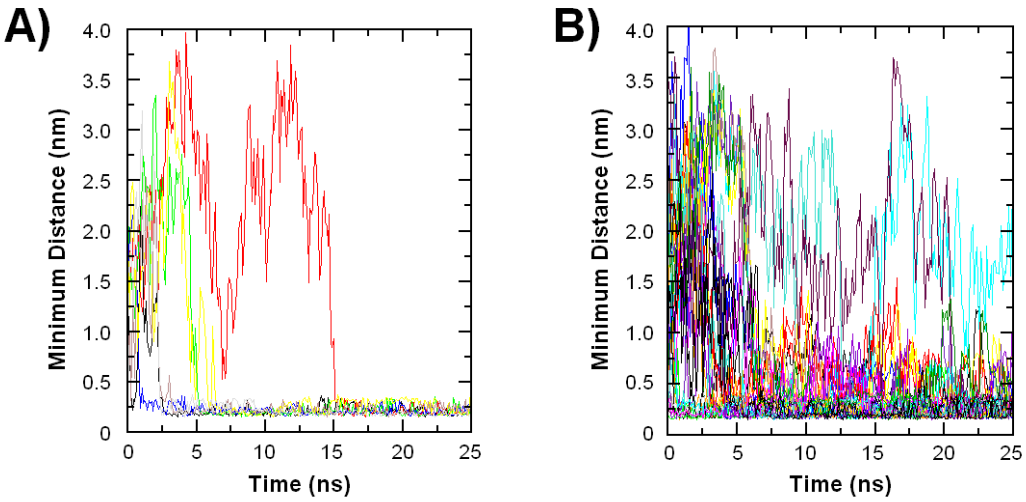

Figure 3.57 Minimum distances between all SDS molecules and subtilisin E wild type during 25 ns MD simulations in SDS. A) 1 % SDS (7 SDS molecules). B) 5 % SDS (32 SDS molecules). The minimum distances between selected groups were calculated using program g_mindist.

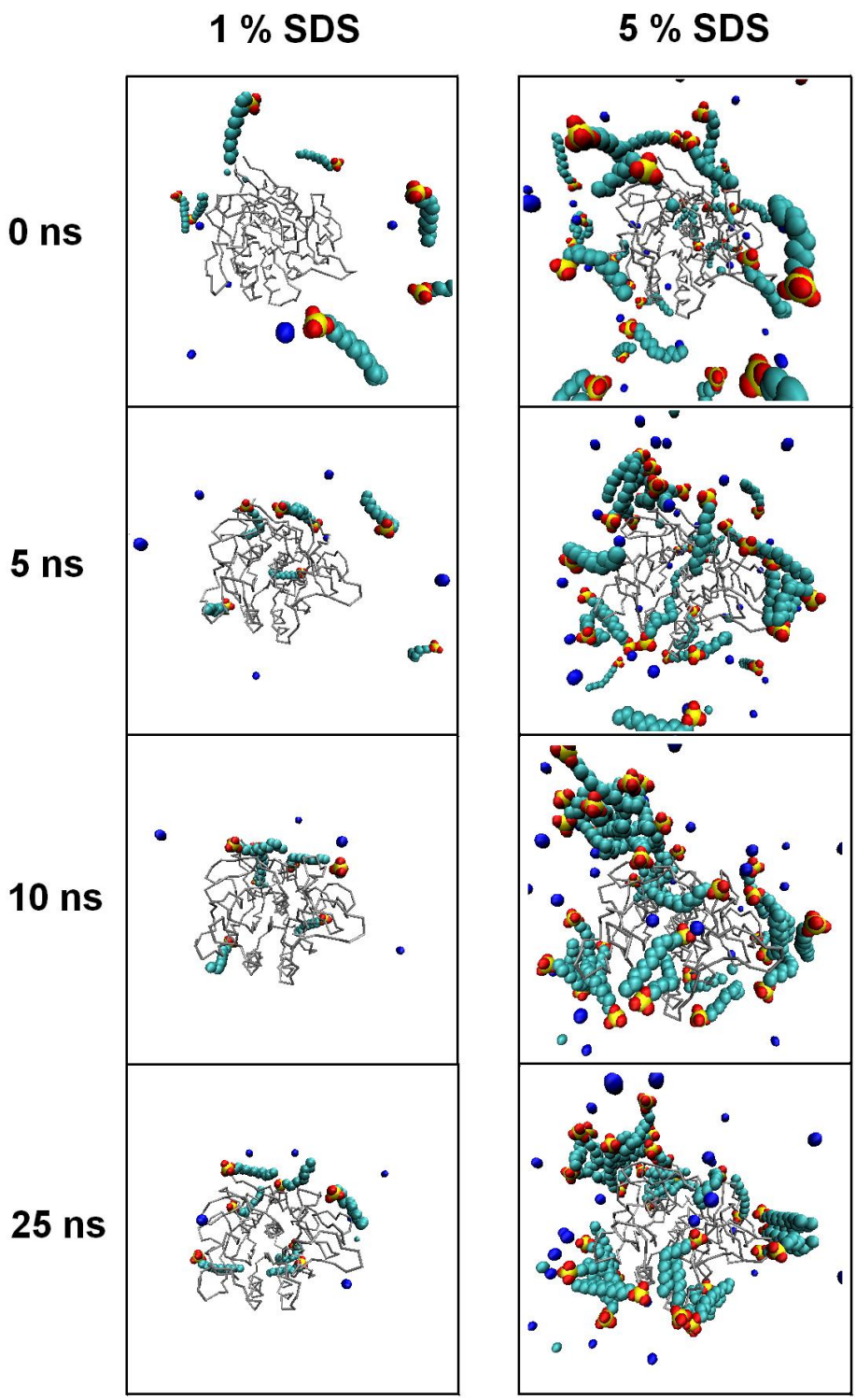

Figure 3.58 Snapshots of MD simulations of subtilisin E wild type in SDS solutions. Snapshots at 0 ns, 5 ns, 10 ns and 25 ns were displayed with subtilisin E in trace format and SDS molecules in VDW format. All water molecules in all snapshots are not shown for better visualization.

However, micelles were observed along simulation in 5 % SDS but not in 1 % SDS (see Figure 3.58). At 0 ns in 5 % SDS simulation, 32 SDS molecules were randomly

distributed around subtilisin E after equilibrium. The first three micelles were observed in less than 5 ns which consisted of 4, 5, 6 SDS molecules respectively, and one SDS molecule was attached to subtilisin E surface. After 5 ns, two micelles were attached to the protein surface. At 10 ns, 3 micelles were attached, while one of them consisted of as many as 11 SDS molecules. At 25 ns, 4 micelles were attached to the subtilisin E surface.

The mutant M4 was also set up for MD simulations in water and 5 % SDS, in order to investigate the effects of substitutions on its tolerance towards SDS, compared to wild type. RMSD analysis of M4 in 5 % SDS showed a slightly higher deviation than in water along 25 ns simulation, while WT showed a higher deviation in water (see Figure 3.59). However, in general, RMSD values are all less than 0.15 nm, indicating that there are no evident conformational changes in all cases.

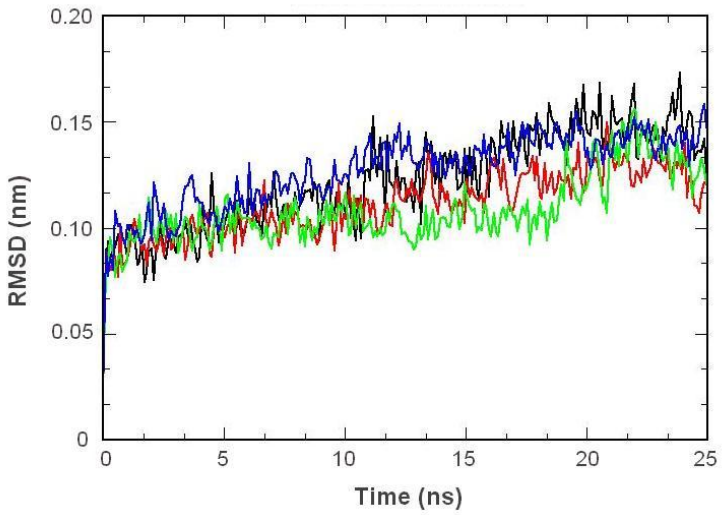

Figure 3.59 RMSD analysis of MD simulations of subtilisin E variants with respect to starting structure (backbone) in water (WT: black; M4: green) and 5 % SDS (WT: red; M4: blue).

Essential dynamic analysis was applied to better investigate the fluctuation patterns of both wild type and M4 in the presence of 5 % SDS, as showed in Figure 3.60. The results show that two motions in both cases are dominant: one has average fluctuations all over the structure, and the other has a high fluctuation only in one loop (residue 155-165), while the positions at substitutions in both cases are very similar in terms of their fluctuation (see Figure 3.60 A). However, in both modes, M4 seems to fluctuate more than wild type subtilisin E (see Figure 3.60 B).

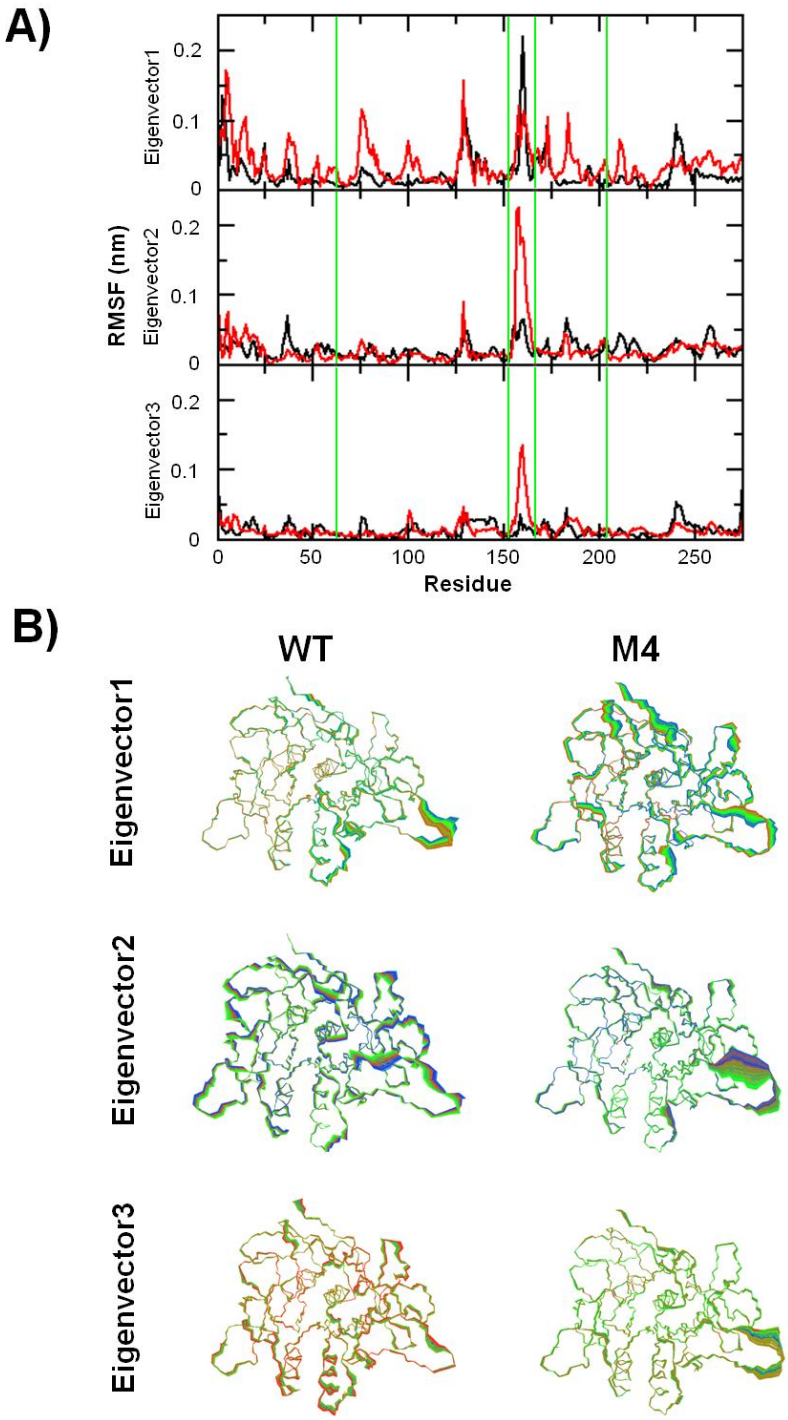

Figure 3.60 RMSF analysis of MD simulations of subtilisin E variants in 5 % SDS. A) RMSF per residue obtained by projecting trajectories on the corresponding first 3 eigenvectors. Four helical strip lines (green) indicated the positions of substitutions in M4. B) Structure clusters of first 3 eigenvectors.

Analysis of interaction of subtilisin E variants with cosolvents showed that the number of hydrogen bonds between protein and SDS increased faster in M4 compared to WT in the first 5 ns and kept constant as in the case of WT (see Figure

3.61 A). Analysis of the solvent accessible surface area (SASA) showed in both cases that the accessible areas decreased fast during the first 5 ns mostly due to shielding by SDS binding. However, the accessible areas are less in M4 than in WT when the number of SASA keeps constant (see Figure 3.61 B), which indicates that surface area is less accessible in M4 than in WT.

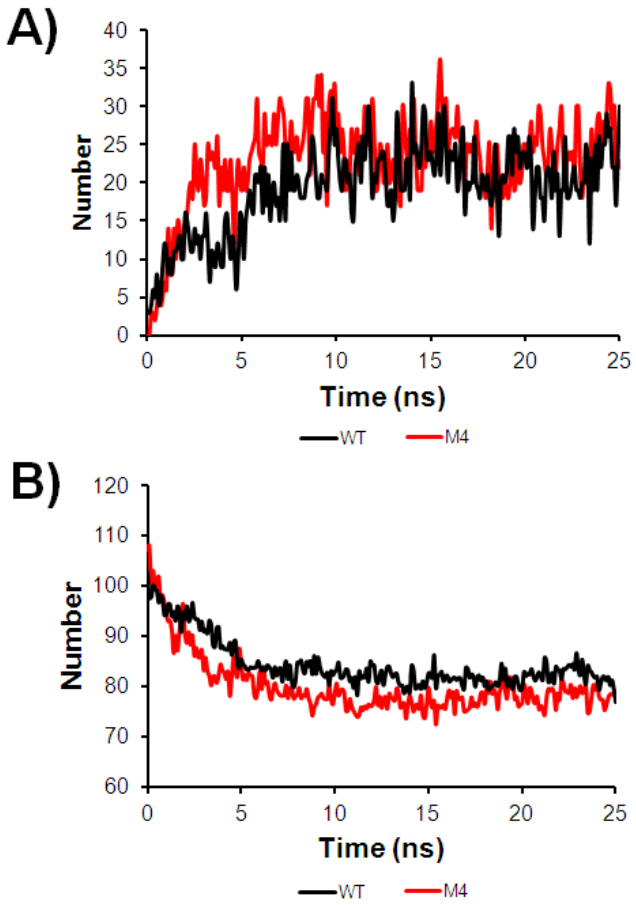

Figure 3.61 Analysis of subtilisin E variants (WT and M4) on interaction with cosolvents during 25 ns MD simulations in 5 % SDS. A) Hydrogen bonding of subtilisin E variants with SDS molecules. B) Solvent accessible surface area (SASA) of subtilisin E variants.

Spatial distribution of hydrophilic head and hydrophobic tail of dodecylsulfate was analyzed (see Figure 3.62). The results showed that the binding of SDS in WT and M4 was similar. The area of the active site and substrate binding pockets are highlighted and it is more likely, the SDS molecules are slightly positioned away from these areas when M4 is compared to WT, which might be due to the substitutions.

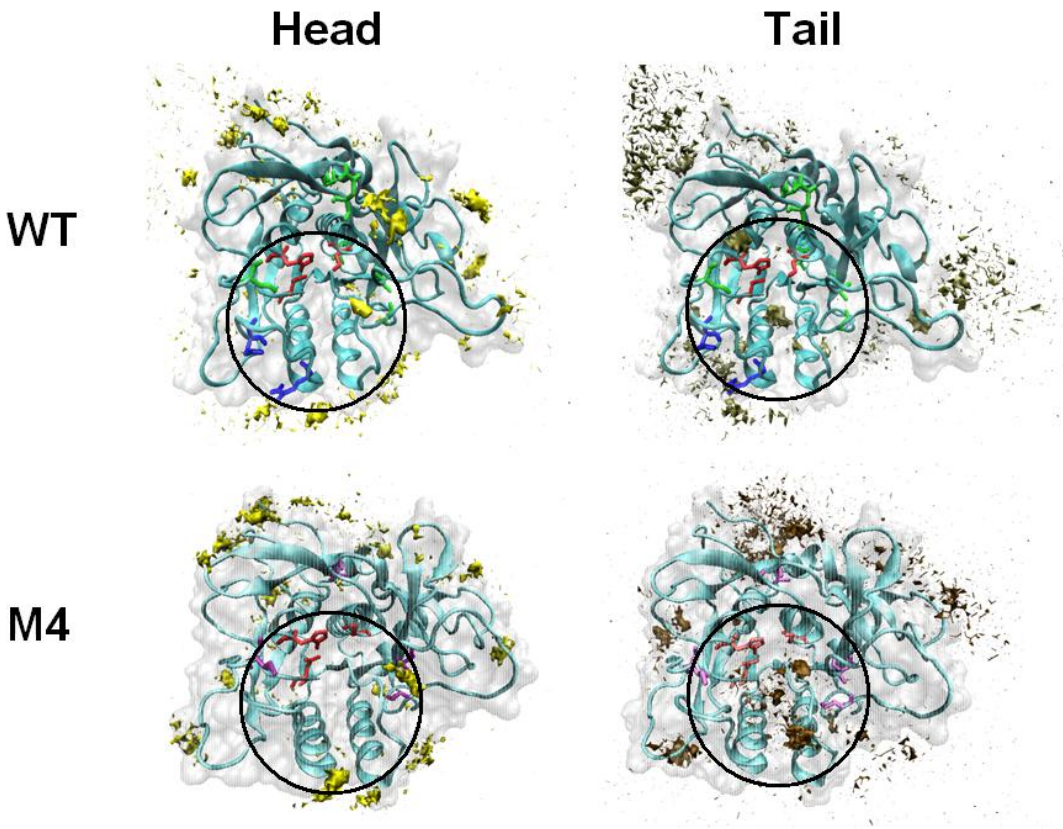

Figure 3.62 Analysis of SDS coordination in MD simulations of subtilisin E variants in 5 % SDS. Analysis of spatial distribution function (SDF) was performed into two groups: hydrophilic head group (left column) and hydrophobic tail (right column). The active site and substrate binding pockets are highlighted in the circle. Catalytic triad residues are displayed in red, and substitutions are also highlighted in green/blue (WT) or purple (M4).

Future examination along the whole simulations of both WT and M4 shows that, SDS molecules can bind to essential area for activity, such as the active site or the substrate binding pockets. As seen in Figure 3.63, two or three SDS molecules were stick to the active site or the substrate binding pocket in 25 ns simulations. For instance, in MD simulation of WT, hydrophilic head group of SDS1 was bound to Q103 in an unspecific binding pocket (Q103:Nε2-SDS1:OA), while hydrophobic tail was imbedded into the cleft. Besides, hydrophilic head group of SDS2 was strongly bound to N155 through hydrogen bonding (N155:Nδ2-SDS2:OA) which is essential for oxyanion hole stabilization. The hydrophobic tail of SDS3 was imbedded into a cavity between the catalytic triad and Y217 through mostly hydrophobic interactions. Nevertheless, SDS binding to Q103 and N155 was also observed in M4 simulations despite that the conformations of these SDS molecules were different.

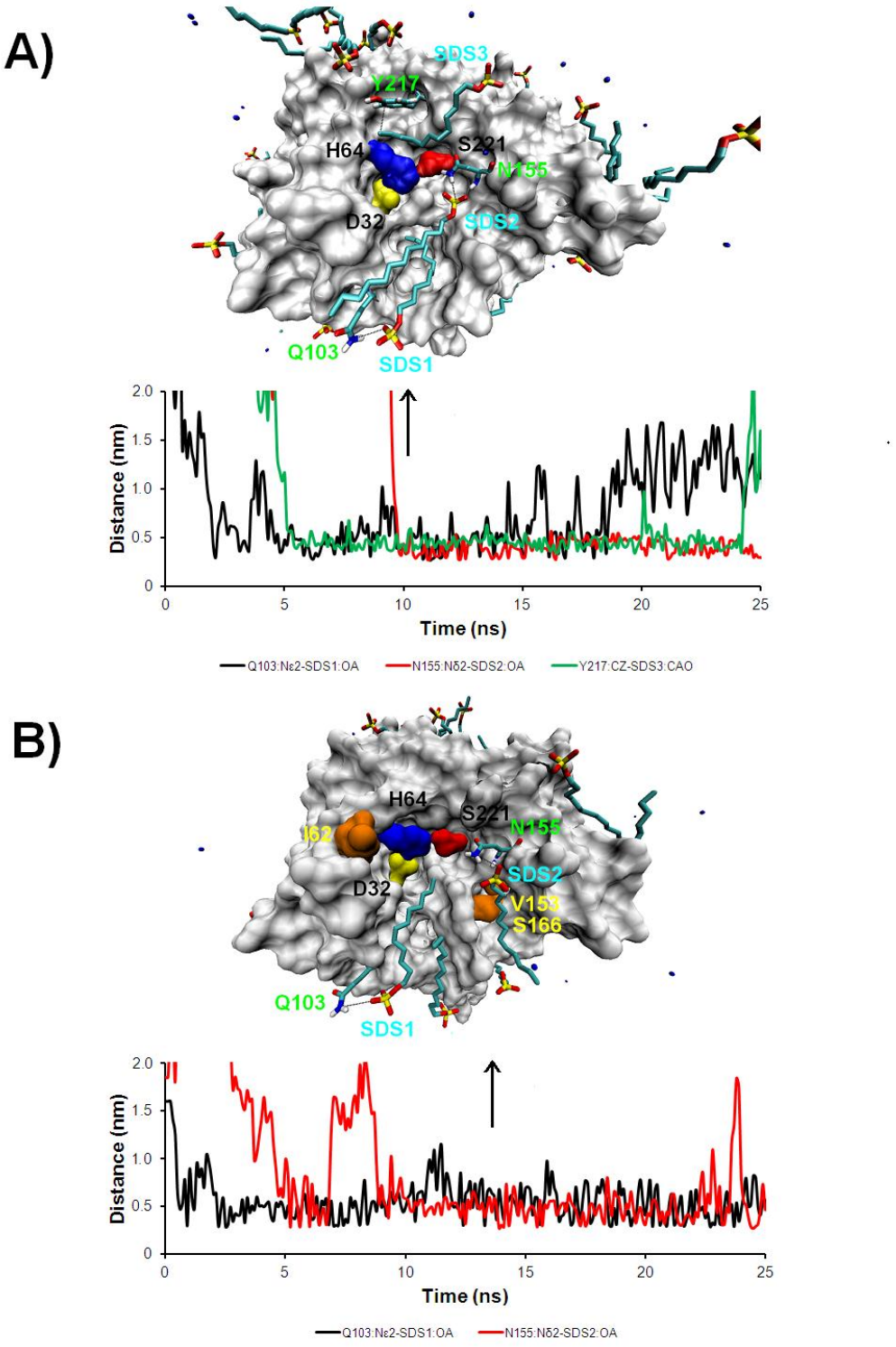

Figure 3.63 Changes in the interaction between SDS and residues at active sites or substrate binding pocket during the 25 ns simulation of subtilisin E WT (A) or M4 (B). Snapshot of SDS binding on WT (10.0 ns) or M4 (14.7 ns) show that there are several SDS molecules which are stick to residues either at the active site or the substrate binding pocket. The interaction between them is plotted in terms of distance of selected atoms. Catalytic triad residues (black), SDS molecules (cyan), SDS-interactive residues (green) and substitutions (yellow) are labeled.

However, Simulations (Figure 3.63) show that the active site are slightly away from interactions with SDS in M4 compared to WT, which might be due to “shrinkage” of the active site or substrate binding areas caused by substitutions (such as S62I, G166S). All these results are very similar to the observation in GdmCl simulations in which the local areas of the active site and substrate binding pockets are also supposed to be “shrinked” caused by the same substitutions to reduce Gdm^+ binding.

4. Discussion

This dissertation describes many aspects of inactivation/denaturation of protease in the presence of chaotropes (GdmCl or SDS in this case) and the feasibility to evolve proteases to tolerate chaotropes by protein engineering. However, the inactivation/denaturation of enzymes by the chaotropic molecules resulted in a multifaced mechanism and detailed picture of the whole process is still not completely clear. Here we focus on the initial inactivation of subtilisin E and try to explain or propose a hypothesis on governing principles for chaotrope resistance, effects of substitutions we identified towards chaotrope resistance and for the activation/denaturation of subtilisin E.

4.1 Directed evolution of proteases for chaotrope resistance

To our knowledge, there are very few publications about directed evolution of enzymes for chaotrope resistance. Especially this dissertation is the first directed evolution approach applied for increasing protease tolerance towards GdmCl by directed evolution. Besides, only two reports exit on improving detergent resistance of hydrolases by directed evolution or semi-rational design. In a first report several rounds of *in vivo* gene shuffling in yeast are applied to a *T. lanuginose* lipase to improve its stability in the presence of the commercial detergents. Several strongly improved variants were obtained [123]. In the second report Danielsen and coworkers selected nine amino acids in the hinge region of the lipase “Lipolase” to overcome inhibitory effects in the presence of detergent. Despite proving that the display systems yielded enriched populations no improved variant could be isolated [124]. However, there is a strong practical aspect to the understanding what drives protein inactivation or unfolding in the presence of a detergent due to its broad industrial application. Companies such as Novozymes and Genencor have built up an

enormous empirical database of mutations that affect enzyme stability und performance in the detergent industry, but very few were published [52].

4.2 Effects of substitutions on chaotrope resistance

Model analysis for the single mutants (see Figure 4.1) revealed that most substitutions could slightly change the local area of the active site or substrate binding pockets. For instance, substitution S62I affects the rotation of imidazole ring of catalytic triad residue His64 by extending the side chain; substitution G166S or G166M affect the specificity binding pocket P1; substitution Q103P or Q103D affects the unspecific binding pocket P4.

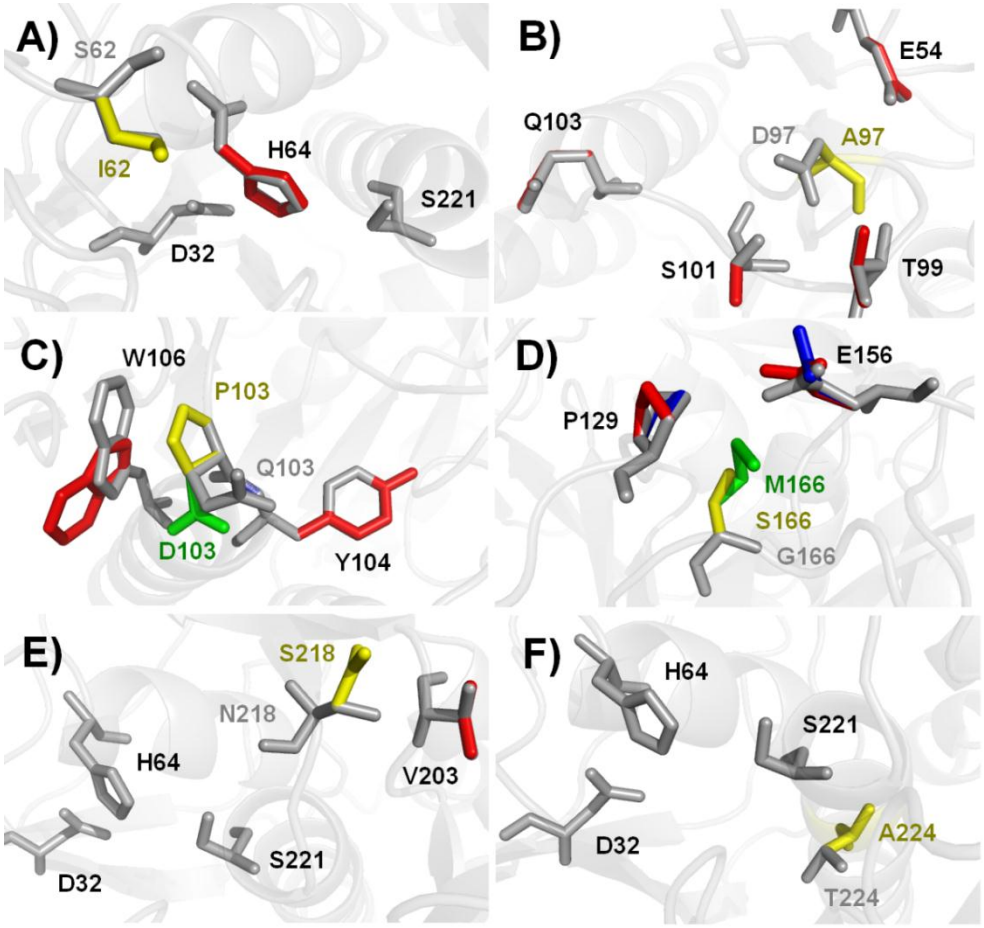

Figure 4.1 Substitutions of single mutants aligned with wild type. A) S62I; B) D97A; C) Q103P or Q103D; D) G166S or G166M; E) N218S; F) T224A. Wild type is displayed in grey. Single substitutions are displayed in yellow or green. Conformational changes of adjacent residues are displayed in red (substitutions in yellow) or blue (substitutions in green). All models of single mutant were generated based on wild type by Model Build module of FoldX.

To date, almost all beneficial residues in subtilisin E here (see Figure 4.1) have been reported in its homologs for different properties [6]. Substitution at position 62 which is located at the P2 site was responsible for substrate specificity. It's been reported that systematic shifts in substrate specificity (k_{cat}/K_m) towards basic residues and away from the natural preference for hydrophobic substrates were observed by mutating neutral residues to negative charged residues (Glu or Asp) at position 62 or 166 in subtilisin BPN'. The double mutant (N62D/G166D) had a larger than additive shift in specificity from hydrophobic to dibasic substrates [125]. However, different from other subtilisin homologs, Ser locates at position 62 in subtilisin E. Serine was mutated into Ile as "best" substitution in terms of GdmCl or SDS tolerance (see Table 3.17), and affects His64, as showed in Figure 4.1 A. It could be assumed this effect might play the main role in GdmCl or SDS tolerance rather than strongly interactions with the substrate to change its specificity (see Table 3.29).

Substitutions at position 166 have been extensively studied for its importance in substrate specificity. It is reported that replacement of conserved glycine by other 12 nonionic amino acids resulted in large changes in specificity towards substrates of increasing size and hydrophobicity. The catalytic efficiency (k_{cat}/K_m) towards small hydrophobic substrates was increased by hydrophobic substitutions, but dropped by enlarging either the substrate side chain or the side chain at position 166 due to steric hindrance [126]. Moreover, replacement of charged amino acids at this position resulted in increase of catalytic efficiency towards charged substrates [125-126]. Interestingly, substitution G166S in subtilisin BPN' proved to improve its thermostability, since the half life of thermal inactivation was doubled in double mutant N218S/G166S (740 % wild type at 65^{o}C) compared to single mutant N218S (350 % wild type at 65 ^{o}C) [127]. Thus that could be explained for subtilisin E M6 in this dissertation, which includes the same substitutions, and has significantly improved chaotrope tolerance (IC_{50} in GdmCl: WT, 0.76 M; M4: 2.73 M; M6: 4.65 M). Additionally, the same substitution was found in a hyperthermostable homolog TK-subtilisin as well, which is previously mentioned to retain full activity in chaotropic agents [128]. Another substitution G166M was considered as the "best" individual substitution in this dissertation regarding its

chaotrope tolerance (see Table 3.17), which was included in mutant M2.

Substitutions at position 218 have been investigated in subtilisin BPN' regarding that effect on thermostability. The same substitution from Asn to Ser proved to account for enhanced thermostability due to stronger hydrogen bonding between two β-sheets [129]. Interestingly, the stability of a double mutant N109S/N218S from aprA-subtilisin, besides its 7 degree higher transition temperature than wild type, was also improved in the presence of SDS as a "side effect" [130]. In the protein engineering of subtilisin E, N218S was found not only responsible for higher thermostability [94], and notably also proved for increased activity as well as stability in the presence of organic solvent dimethylformamide (DMF) [131]. In this work, such substitution was generated by random mutagenesis and proved later to be the only one at this position improving chaotrope tolerance (see Table 3.17), which indicates N218S could be a general beneficial substitution for increased stability of subtilisins.

The substitution at position 224 in a double mutant of subtilisin BPN' was reported for new activity as a subtiligase for ligation of peptides after screening of different libraries using phage display [132]. Thus it could be assumed that the substitution at this position might affect the catalytic triad residue Ser221 in proximity through surrounding hydrogen bonding networks. In this dissertation, some substitutions at position 224, such as T224A in subtilisin E showed improved chaotolerance (see Table 3.17) might work in the same manner. Interestingly, another beneficial substitution, T224S, was also found in other natural subtilisins, such as subtilisin BPN' and subtilisin Carlsberg. Nevertheless, more studies are required for further elucidation of its chaotolerance contribution.

Substitution D97G was reported in subtilisin E for improving activity in organic solvent DMF [41], while we reported that the substitution D97A was responsible for moderate improved tolerance towards SDS (see Table 3.22). From Figure 3.38 and 4.1 B, this substitution is supposed to exert effects on the loop region which binds the tetrapeptide substrate Suc-AAPF-pNA and improves activity towards SDS.

Position 103 on the surface of subtilisin E is very interesting for chaotrope tolerance in subtilisin E, because of its dual role for GdmCl and SDS tolerance (see

Table 3.22). Reports regarding substitution Q103R describe its role in improving activity in organic solvent DMF [41] due to a favorable electrostatic interactions between positive charged side chain of substituted Arg and negative charged protection succinyl group of substrate Suc-AAPF-pNA. However, another report stated that Q103R enhanced the specific activity of subtilisin E but decreased the thermal stability [133]. However, on one hand, it was found GdmCl tolerance was obviously improved by introduction of negative charged residues as in Q103D, Q103E mutant (see Table 3.17), which occur more often in halophilic counterparts [134]. It could be explained that excessive positive charged Gdm^+ ions were trapped at this site through electrostatic interaction so that inner essential active residues were shielded. However, activities of both single mutants (Q103D and Q103E) were dramatically reduced due to unfavorable electrostatic interactions with the succinyl group of the substrate Suc-AAPF-pNA. On the other hand, substitution Q103P was selected from epPCR libraries for improvement of SDS tolerance (see Table 3.17), which might be caused by an elimination of hydrogen bonding in this region which also has an effect on the local structure of substrate binding site (see Figure 4.1 C).

Some other individual substitutions in muteins are mostly neutral in terms of their tolerance, such as I205V, A153V or T240S (see Figure 3.19 and Figure 3.21). Additionally, single mutant A153V seems to recover the activity towards complex protein substrate, while I205V seems detrimental (see Figure 3.19). However, A153V and I205V together improved SDS tolerance in M4 compared to M2 (see Figure 3.19).

4.3 Inactivation and denaturation of enzymes

In the study of chaotolerance improvement of subtilisin E by directed evoultion in this dissertation there are several results which are beyond our expectations: 1) it is observed that subtilisin E wild type lost their activity dramatically and instantly when exposed to GdmCl solutions even at lower moler concentration like 1 M (see Figure 3.33), which came to the assumption that it might be caused by the direct disruption of conformation of this enzyme at lower concentration, especially the active site. 2) All

substitutions which are responsible for GdmCl- or SDS-resistance are located around the active site or substrate binding pockets of subtilisin E and these substitutions are mostly buried in the interior, not at the more accessible surface areas where they were normally reported for thermostability [39] or even tolerance in organic solvent like DMF [36]. Table 4.2 lists relative solvent accessibility (ASA) values of substitutions of subtilisin E variants after directed evolution under three different denaturant conditions, including temperature (thermostability), organic solvent like DMF and chaotropic agents like GdmCl or SDS (this project). Average ASA values are 55.7 % (Thermostability), 38.2 % (DMF tolerance) and 14.7 % (GdmCl/SDS tolerance), indicating that the substitutions for chaotolerance are least accessible and located in the interior. 3) It is unexpected that the several substitutions, such as S62I or G166S, could improve tolerance towards chaotropic salts (GdmCl) and detergent (SDS) simultaneously, which reveals that there might be some common pathways.

During the directed evolution of protease for its GdmCl- or SDS-resistance, comparatively lower concentrations of GdmCl (1 M) or SDS (0.5 %) were applied due to feasibility of screening system (higher concentration of denaturants induce complete loss of activity of WT). The effects of such low concentrations of GdmCl have been less studied since high concentrations (5-6 M) were mostly studied and used to denature proteins. GdmCl causes mostly conformational changes, as is also observed by circular dichroism results (see Figure 3.37). Tsou has reported in 1995 one review about a very important observation that inactivation precedes overall molecular conformation changes during enzyme denaturation [122b]. In this review, it is mentioned that such observation has been found for 18 different enzymes. It is noted that the initial rates of inactivation of a number of enzymes in GdmCl or urea are usually too fast to be measured by the conventional method of incubation of enzymes with the denaturants. With stopped flow measurements, the inactivation rates of creatine kinase during GdmCl or urea denaturation were over three orders of magnitude faster than the rates of conformational changes under the same denaturant concentrations [135]. This is consistent with the observation that

Table 4.2 Relative solvent accessibility (ASA) of substitutions of best subtilisin E variants for stability under different denaturing conditions after directed evolution

Residue Index	Relative ASA* value (%)		
	Themostability (5-3H5)[39]	**DMF tolerance (13M)[36]**	**GdmCl/SDS tolerance (M6)**
14	29.6	-	-
48	-	21.8	-
60	-	9.0	-
62	-	-	17.9
76	65.6	-	-
97	-	20.8	-
103	-	12.3	-
107	-	1.1	-
118	83.3	-	-
131	-	30.5	-
153	-	-	0.0
156	-	74.4	-
161	93.0	-	-
166	3.8	-	3.8
181	21.9	21.9	-
182	-	65.7	-
188	-	85.3	-
194	83.6	-	-
205	-	-	1.6
206	-	48.7	-
218	64.9	64.9	64.9
224	-	-	0.0
255	-	40.4	-
Average	55.7	38.2	14.7

*ASA values were calculated with ‘‘ASA view’’ (http://www.netasa.org/asaview/). The calculation is based on the crystal structure of subtilisin E (PDB code: 1SCJ)

conformation changes of subtilisin E wild type in GdmCl revealed by CD spectra analysis (see Figure 3. 37) were much slower than the activity loss measured for half life of subtilisin E muteins (see Figure 3.33).

Thus three possibilities were proposed for such observation: 1) an inhibition of the enzyme by the chaotropes; 2) dissociation of the dimeric enzyme into the inactive monomers and 3) conformation perturbation at a relative flexible enzyme active site. Since subtilisin E is a monomer, the possibility of the dissociation of the dimeric

enzyme could be ruled out. Therefore inactivation prior to unfolding is either caused by inhibition of those chaotropes or conformational disruption of the active site.

There are some reports which favored the inhibition mechanism. For instance,GdmCl inhibits RNase A [136] since the decrease in enzyme activity in dilute GdmCl solution is a reversible inhibition at the active site [137]. Wu and coworkers reported on α-glucosidase inactivation at lower concentrations of GdmCl or SDS directly considered both to act as inhibitors [59]. Table 4.3 lists the residual activities of wild type and variant M4 in 1-2 M GdmCl or 0.5 % SDS on different concentrations of peptide substrate Suc-AAPF-pNA to indicate competitive inhibition mode.

Tsou in the review argued that the possibility of inhibition could be excluded. Some reports claim that thermal inactivation precedes heat induced conformation change such as for adenylate kinase [138], D-glyceraldehyde-3-phosphate dehydrogenase [139] and β-galactosidase [140]. A number of monomeric enzymes, such as lysozyme and Thermolysin, during pressure denaturation showed inactivation prior to unfolding [141]. In these cases, there was no inhibitor involved so that the inhibition by an inhibitor being responsible for the loss of enzyme activity can be excluded. It is also pointed out in some literature that as the denaturant is usually diluted many folds with the enzyme in activity assays, some refolding and hence reactivation may have occurred with sharp decrease at denaturing concentrations [142]. Furthermore, it is also mentioned in the review by Tsou that the presence of substrates in the assay mixture can sometimes cause reactivation especially for partially inactivated enzymes as the initial stage of unfolding. Another method to directly demonstrate that conformation changes at the active site under mild denaturing conditions causing inactiviation is by probing the active site with fluorescent probes. One fluorescent isoindole derivative was introduced as a probe near the active site of creatine kinase [143] and the results suggested increased mobility of the fluorophore at the active site.

In this dissertation, the perturbation of the active site by Gdm^+ ions was also observed in one 50 ns MD simulations of subtilisin E wild type in 5 M GdmCl and the active site was disrupted by one “attacking” of one Gdm^+ ion which replaced the position of the His64 (see Figure 3.45). However, other simulations of GdmCl and

SDS showed only strong binding instead of a disruption of the active site. Hence these two different interpretations for the inactivation prior to unfolding in chaotropes might be both reasonable, while the experimental and computational analysis in this dissertation might be in more favor of inhibition interpretation.

Nevertheless, it has previously been proposed that Gdm^+ binding occurs at the active site of subtilisin E prior to molecular conformational changes, as indicated by MD simulations. Substitutions that dramatically decrease K_m (see Figure 3.29) often increase affinity to the peptide substrate (Suc-AAPF-pNA), and reduce the flexibility of the active site. Therefore, tight substrate binding assists in excluding GdmCl from the active site, thereby resulting in increased GdmCl resistance. However, the activity with complex substrates, such as Skim-milk, was often reduced (Figure 3.19) due to decreased flexibility of the active site. Similar stabilization against denaturation in dilute GdmCl was observed in muscle lactate dehydrogenase H4 (glutaraldehyde crosslinking or in concentrated ammonium sulfate solution), and reduced flexibility led to a decrease in activity.

Table 4.3 Relative activity of subtilisin E variants in GdmCl or SDS on different concentrations of substrate Suc-AAPF-pNA

Substrate concentration (mM)	Relative activity of WT (%)			Relative activity of M4 (%)		
	1 M GdmCl	2 M GdmCl	0.5 % SDS	1 M GdmCl	2 M GdmCl	0.5 % SDS
0.10	18	4	37	57	35	49
0.25	17	6	34	68	46	73
1.00	29	8	45	86	61	85
2.00	35	12	58	82	62	85
5.00	46	18	75	82	62	87

4.4 Conclusions and outlook

In summary, three iterative rounds of directed evolution and subsequent site saturation mutagenesis studies revealed that substitutions at positions 62, 97, 103, 166, 218 and 224 improved subtilisin E resistance to GdmCl and SDS inactivation.

For the first time subtilisin E was engineered into a chaotolerent protease. Mutant M6 showed highest chaotolerance towards GdmCl and SDS compared to wild type (half life 67.6 min in 3 M GdmCl for M6, <2 min for WT; half life 66.3 min for M6 in 2.0 % SDS; <2 min for WT). CD spectra results of subtilisin E in the presence of GdmCl or SDS indicates that inactivation precedes overall conformational changes at mild concentrations of these two chaotropes (1-2 M GdmCl; 0.5 -2.0 % SDS). MD simulations led to the hypothesis that Gdm^+ ions or DS^- might directly interact with residues at the active site or substrate binding pockets in an inhibition manner. The unusual finding that all substitutions surround the active site or substrate binding pocket residues in subtilisin E demonstrates the benefit of mutating amino acids close to these areas to stabilize subtilisin E against chaotropes like GdmCl or SDS. This approach might be of general interest for serine protease stabilization.

Activity of almost all GdmCl- or SDS- tolerant variants of subtilisin E (except SeSaM1-7) in the absence of those chaotropes was reduced towards complex protein substrate, like Skim-milk or Azocasein despite that their activities towards peptide substrate Suc-AAPF-AMC or Suc-AAPF-pNA was not reduced (see Table 3.13, 3.20, 3.22, 3.32). Activity loss of stability improved variant could be compensated again by protein engineering. For instance, total activity loss of subtilisin E variant with high tolerance towards DMF identified by directed evolution was recovered by additional rounds of directed evolution [36]. Another single mutation of Ser for Ala was introduced into the highly conserved 83-Gly-Val-Ala-85 region of aprE (subtilisin E) which resulted in the generation of a subtilisin mutant with an enhanced catalytic efficiency due to a larger k_{cat}. The position of the mutation was placed in the conserved N-terminal end of the loop that connects a β-sheet with the α-helix containing the catalytic residue His64 [144].

It should also be noted that it is unusual in directed evolution experiments, that all beneficial substitutions are located around the active site or substrate binding pockets, especially as only a few thousand variants were sampled [29, 145]. There are some substitutions observed outside this region which have neutral or even adverse effects on chaotolerance, such as position 181, 184 (see Figure 3.16) indicated by SSM

studies. However, the way of construction of mutant M2, M5 and M6 by addition of beneficial substitutions ruled out that these substitutions might have cooperative effects with other substitutions. The novel method SeSaM-based shuffling was developed as an alternative to above recombinatorial mutagenesis to overcome this drawback but ended up unfortunately with complete loss of activity towards complex substrate (see section 3.2.2.3). Therefore, it might be alternative semi-rational investigation on the role of amino acid positions in close proximity to the active site with multi-site saturation [146] in order to further stabilize subtilisin E towards GdmCl or SDS.

5. References

[1] R. Gupta, Q. K. Beg, P. Lorenz, *Appl Microbiol Biotechnol* **2002**, *59*, 15.

[2] M. B. Rao, A. M. Tanksale, M. S. Ghatge, V. V. Deshpande, *Microbiol Mol Biol Rev* **1998**, *62*, 597.

[3] L. Hedstrom, *Chem Rev* **2002**, *102*, 4501.

[4] C. Branden, J. Tooze, *Introduction to Protein Structure*, 2nd Edition ed., Garland Publishing, New York, **1999**.

[5] J. J. Perona, C. S. Craik, *Protein Sci* **1995**, *4*, 337.

[6] P. N. Bryan, *Biochim Biophys Acta* **2000**, *1543*, 203.

[7] F. Kunst, N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C. V. Bruschi, B. Caldwell, V. Capuano, N. M. Carter, S. K. Choi, J. J. Codani, I. F. Connerton, A. Danchin, et al., *Nature* **1997**, *390*, 249.

[8] Y. Y. Li, Z. X. Hu, F. Jordan, M. Inouye, *J Biol Chem* **1995**, *270*, 25127.

[9] H. Takagi, T. Takahashi, H. Momose, M. Inouye, Y. Maeda, H. Matsuzawa, T. Ohta, *J Biol Chem* **1990**, *265*, 6874.

[10] C. von der Osten, S. Branner, S. Hastrup, L. Hedegaard, M. D. Rasmussen, H. Bisgard-Frantzen, S. Carlsen, J. M. Mikkelsen, *J Biotechnol* **1993**, *28*, 55.

[11] Y. H. Yang, L. Jiang, S. L. Yang, L. Q. Zhu, Y. J. Wu, Z. J. Li, *World J Microbiol Biotechnol* **2000**, *16*, 249.

[12] S. C. Jain, U. Shinde, Y. Li, M. Inouye, H. M. Berman, *J Mol Biol* **1998**, *284*, 137.

[13] K. H. Maurer, *Curr Opin Biotechnol* **2004**, *15*, 330.

[14] C. O. Fagain, *Biochim Biophys Acta* **1995**, *1252*, 1.

[15] C. Schiraldi, M. De Rosa, *Trends Biotechnol* **2002**, *20*, 515.

[16] V. G. Eijsink, A. Bjork, S. Gaseidnes, R. Sirevag, B. Synstad, B. van den Burg, G. Vriend, *J Biotechnol* **2004**, *113*, 105.

[17] V. G. H. Eijsink, S. Gaseidnes, T. V. Borchert, B. van den Burg, *Biomol Eng* **2005**, *22*, 21.

[18] J. E. Ness, M. Welch, L. Giver, M. Bueno, J. R. Cherry, T. V. Borchert, W. P. Stemmer, J. Minshull, *Nat Biotechnol* **1999**, *17*, 893.

[19] K. Miyazaki, F. H. Arnold, *J Mol Evol* **1999**, *49*, 716.

[20] D. R. Mills, R. L. Peterson, S. Spiegelman, *Proc Natl Acad Sci U S A* **1967**, *58*, 217.

[21] K. E. Jaeger, T. Eggert, *Curr Opin Biotechnol* **2004**, *15*, 305.

[22] K. Furukawa, *Trends Biotechnol* **2003**, *21*, 187.

[23] C. P. Locher, N. W. Soong, R. G. Whalen, J. Punnonen, *Curr Opin Mol Ther* **2004**, *6*, 34.

[24] T. S. Wong, D. Roccatano, M. Zacharias, U. Schwaneberg, *J Mol Biol* **2006**, *355*, 858.

[25] S. Lutz, W. M. Patrick, *Curr Opin Biotechnol* **2004**, *15*, 291.

[26] A. M. Aguinaldo, F. H. Arnold, *Methods Mol Biol* **2003**, *231*, 105.

[27] W. M. Coco, *Methods Mol Biol* **2003**, *231*, 111.

[28] M. Ostermeier, J. H. Shim, S. J. Benkovic, *Nat Biotechnol* **1999**, *17*, 1205.

[29] A. V. Shivange, J. Marienhagen, H. Mundhada, A. Schenk, U. Schwaneberg, *Curr*

Opin Chem Biol **2009**, *13*, 19.
[30] P. C. Cirino, K. M. Mayer, D. Umeno, *Methods Mol Biol* **2003**, *231*, 3.
[31] R. C. Cadwell, G. F. Joyce, *PCR Methods Appl* **1992**, *2*, 28.
[32] T. S. Wong, D. Roccatano, U. Schwaneberg, *Biotechnol J* **2007**, *2*, 133.
[33] T. S. Wong, K. L. Tee, B. Hauer, U. Schwaneberg, *Nucleic Acids Res* **2004**, *32*, e26.
[34] K. Bernath, M. Hai, E. Mastrobattista, A. D. Griffiths, S. Magdassi, D. S. Tawfik, *Anal Biochem* **2004**, *325*, 151.
[35] K. D. Wittrup, *Curr Opin Biotechnol* **2001**, *12*, 395.
[36] L. You, F. H. Arnold, *Protein Eng* **1996**, *9*, 77.
[37] H. Zhao, L. Giver, Z. Shao, J. A. Affholter, F. H. Arnold, *Nat Biotechnol* **1998**, *16*, 258.
[38] Z. Shao, H. Zhao, L. Giver, F. H. Arnold, *Nucleic Acids Res* **1998**, *26*, 681.
[39] H. M. Zhao, F. H. Arnold, *Protein Eng* **1999**, *12*, 47.
[40] a)M. Ottesen, I. Svendsen, in *Methods in enzymology, Vol. 19* (Eds.: Perlmann GE, L. L), Academic Press, New York, **1970**, 199; b)L. O. Narhi, M. Zukowski, T. Arakawa, *Arch Biochem Biophys* **1988**, *261*, 161.
[41] K. Chen, F. H. Arnold, *Proc Natl Acad Sci U S A* **1993**, *90*, 5618.
[42] P. Bhaganna, R. J. Volkers, A. N. Bell, K. Kluge, D. J. Timson, J. W. McGrath, H. J. Ruijssenaars, J. E. Hallsworth, *Microb Biotechnol* **2010**, *3*, 701.
[43] R. L. Baldwin, *Biophys J* **1996**, *71*, 2056.
[44] J. L. England, G. Haran, *Annu Rev Phys Chem* **2011**, *62*, 257.
[45] P. E. Mason, G. W. Neilson, C. E. Dempsey, A. C. Barnes, J. M. Cruickshank, *Proc Natl Acad Sci U S A* **2003**, *100*, 4557.
[46] Y. Marcus, *J Chem Thermodyn* **2012**, *48*, 70.
[47] P. E. Mason, G. W. Neilson, J. E. Enderby, M. L. Saboungi, C. E. Dempsey, A. D. MacKerell, Jr., J. W. Brady, *J Am Chem Soc* **2004**, *126*, 11462.
[48] G. Graziano, *Phys Chem Chem Phys* **2011**, *13*, 12008.
[49] R. J. Hunter, *Foundations of Colloid Science*, Clarendon Press, Oxford, **1987**.
[50] D. E. Otzen, M. Oliveberg, *J Mol Biol* **2002**, *315*, 1231.
[51] R. M. Garavito, S. Ferguson-Miller, *J Biol Chem* **2001**, *276*, 32403.
[52] D. Otzen, *Biochim Biophys Acta* **2011**, *1814*, 562.
[53] A. K. Bhuyan, *Biopolymers* **2010**, *93*, 186.
[54] S. D'Auria, M. Rossi, R. Nucci, G. Irace, E. Bismuto, *Proteins* **1997**, *27*, 71.
[55] A. Muga, J. L. Arrondo, T. Bellon, J. Sancho, C. Bernabeu, *Arch Biochem Biophys* **1993**, *300*, 451.
[56] R. Schomaecker, B. H. Robinson, P. D. I. Fletcher, *J Chem Soc, Faraday Trans* **1988**, *84*, 4203.
[57] B. Shareghi, A. A. Moosavi-Movahedi, *Int J Biol Macromol* **1996**, *19*, 9.
[58] Y. Shi, W. Luo, W. X. Tian, T. Zhang, H. M. Zhou, *Int J Biochem Cell Biol* **1998**, *30*, 1319.
[59] X. Q. Wu, H. Xu, H. Yue, K. Q. Liu, X. Y. Wang, *Protein J* **2009**, *28*, 448.
[60] a)J. M. Chirgwin, A. E. Przybyla, R. J. MacDonald, W. J. Rutter, *Biochemistry* **1979**, *18*, 5294; b)D. D. Bowtell, *Anal Biochem* **1987**, *162*, 463; c)P. Chomczynski, N. Sacchi, *Anal Biochem* **1987**, *162*, 156; d)R. Boom, C. J. Sol, M. M. Salimans, C. L. Jansen, P. M. Wertheim-van Dillen, J. van der Noordaa, *J Clin Microbiol* **1990**, *28*, 495.

[61] J. P. Williams, J. E. Hallsworth, *Environ Microbiol* **2009**, *11*, 3292.
[62] R. E. De Castro, J. A. Maupin-Furlow, M. I. Gimenez, M. K. Herrera Seitz, J. J. Sanchez, *FEMS Microbiol Rev* **2006**, *30*, 17.
[63] C. Sanchez-Porro, E. Mellado, C. Bertoldo, G. Antranikian, A. Ventosa, *Extremophiles* **2003**, *7*, 221.
[64] K. Hecht, R. Jaenicke, *Biochemistry* **1989**, *28*, 4979.
[65] S. Kanaya, T. Foophow, S. Tanaka, Y. Koga, K. Takano, *Protein Eng Design & Selection* **2010**, *23*, 347.
[66] M. A. Pulido, Y. Koga, K. Takano, S. Kanaya, *Protein Eng Design & Selection* **2007**, *20*, 143.
[67] Y. Takeuchi, S. Tanaka, H. Matsumura, Y. Koga, K. Takano, S. Kanaya, *Biochemistry* **2009**, *48*, 10637.
[68] QIANGEN, *QIANGEN News* **1997**.
[69] H. Hilz, U. Wiegers, P. Adamietz, *Eur J Biochem* **1975**, *56*, 103.
[70] a)U. Wiegers, H. Hilz, *Biochem Biophys Res Commun* **1971**, *44*, 513; b)U. Wiegers, G. Kramer, H. Hilz, *Hoppe Seylers Z Physiol Chem* **1973**, *354*, 1576.
[71] V. Brissos, T. Eggert, J. M. Cabral, K. E. Jaeger, *Protein Eng Design & Selection* **2008**, *21*, 387.
[72] E. P. O'Brien, R. I. Dima, B. Brooks, D. Thirumalai, *J Am Chem Soc* **2007**, *129*, 7346.
[73] P. E. Mason, J. W. Brady, G. W. Neilson, C. E. Dempsey, *Biophys J* **2007**, *93*, L04.
[74] C. Camilloni, A. G. Rocco, I. Eberini, E. Gianazza, R. A. Broglia, G. Tiana, *Biophys J* **2008**, *94*, 4654.
[75] P. E. Mason, C. E. Dempsey, G. W. Neilson, S. R. Kline, J. W. Brady, *J Am Chem Soc* **2009**, *131*, 16689.
[76] R. Godawat, S. N. Jamadagni, S. Garde, *J Phys Chem B* **2010**, *114*, 2246.
[77] F. Mehrnejad, M. Khadem-Maaref, M. M. Ghahremanpour, F. Doustdar, *J Comput Aided Mol Des* **2010**, *24*, 829.
[78] W. Li, Y. Mu, *Nanoscale* **2012**, *4*, 1154.
[79] a)J. Tirado-Rives, M. Orozco, W. L. Jorgensen, *Biochemistry* **1997**, *36*, 7313; b)A. Caflisch, M. Karplus, *Structure* **1999**, *7*, 477.
[80] L. Hua, R. Zhou, D. Thirumalai, B. J. Berne, *Proc Natl Acad Sci U S A* **2008**, *105*, 16928.
[81] I. Eberini, A. Emerson, C. Sensi, L. Ragona, P. Ricchiuto, A. Pedretti, E. Gianazza, A. Tramontano, *J Mol Graph Model* **2011**, *30*, 24.
[82] a)A. D. Mackerell, *J Phys Chem* **1995**, *99*, 1846; b)C. D. Bruce, M. L. Berkowitz, L. Perera, M. D. E. Forbes, *J Phys Chem B* **2002**, *106*, 3788.
[83] C. D. Bruce, S. Senapati, M. L. Berkowitz, L. Perera, M. D. E. Forbes, *J Phys Chem B* **2002**, *106*, 10902.
[84] N. Yoshii, K. Iwahashi, S. Okazaki, *J Chem Phys* **2006**, *124*, 184901.
[85] N. Yoshii, S. Okazaki, *Chem Phys Lett* **2006**, *425*, 58.
[86] N. Yoshii, S. Okazaki, *Chem Phys Lett* **2006**, *426*, 66.
[87] M. Sammalkorpi, M. Karttunen, M. Haataja, *J Phys Chem B* **2007**, *111*, 11722.
[88] S. Jalili, M. Akhavan, *Colloids Surf A-Physicochem Eng Aspects* **2009**, *352*, 99.
[89] R. Braun, D. M. Engelman, K. Schulten, *Biophys J* **2004**, *87*, 754.

[90] H. Khandelia, Y. N. Kaznessis, *Peptides* **2005**, *26*, 2037.
[91] A. Langham, Y. Kaznessis, *J Pept Sci* **2005**, *11*, 215.
[92] S. Jalili, M. Akhavan, *Biophys Chem* **2011**, *153*, 179.
[93] X. C. Wu, W. Lee, L. Tran, S. L. Wong, *J Bacteriol* **1991**, *173*, 4952.
[94] H. Zhao, F. H. Arnold, *Proc Natl Acad Sci U S A* **1997**, *94*, 7997.
[95] J. Sambrook, D. W. Russell, *Molecular Cloning: A laboratory manual*, Cold Springer Harbor Laboratory Press, New York, **2001**.
[96] M. Blanusa, A. Schenk, H. Sadeghi, J. Marienhagen, U. Schwaneberg, *Anal Biochem* **2010**, *406*, 141.
[97] H. Inoue, H. Nojima, H. Okayama, *Gene* **1990**, *96*, 23.
[98] W. J. Dower, J. F. Miller, C. W. Ragsdale, *Nucleic Acids Res* **1988**, *16*, 6127.
[99] L. Zheng, U. Baumann, J. L. Reymond, *Nucleic Acids Res* **2004**, *32*, e115.
[100] P. A. Alexander, B. Ruan, P. N. Bryan, *Biochemistry* **2001**, *40*, 10634.
[101] M. Pulido, K. Saito, S. Tanaka, Y. Koga, M. Morikawa, K. Takano, S. Kanaya, *Appl Environ Microbiol* **2006**, *72*, 4154.
[102] J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, L. Serrano, *Nucleic Acids Res* **2005**, *33*, W382.
[103] W. F. van Gunsteren, W. R. P. Scott, P. H. Hunenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Kruger, *J Phys Chem A* **1999**, *103*, 3596.
[104] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, *J Comput Chem* **1998**, *19*, 1639.
[105] E. Krieger, G. Koraimann, G. Vriend, *Proteins* **2002**, *47*, 393.
[106] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans, in *Intermolecular Forces* (Ed.: B. Pullman), D. Reidel Publishing Company, Dordrecht, **1981**, 331.
[107] A. W. Schuttelkopf, D. M. van Aalten, *Acta Crystallogr D Biol Crystallogr* **2004**, *60*, 1355.
[108] H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola, J. R. Haak, *J Chem Phys* **1984**, *81*, 3684.
[109] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, *J Comput Chem* **1997**, *18*, 1463.
[110] S. Miyamoto, P. A. Kollman, *J Comput Chem* **1992**, *13*, 952.
[111] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J Chem Phys* **1995**, *103*, 8577.
[112] W. Humphrey, A. Dalke, K. Schulten, *J Mol Graph* **1996**, *14*, 33.
[113] A. Amadei, A. B. Linssen, H. J. Berendsen, *Proteins* **1993**, *17*, 412.
[114] C. Cupp-Enyard, *J. Vis. Exp.* **2008**, e899.
[115] U. Tisljar, H. W. Denker, *Anal Biochem* **1986**, *152*, 39.
[116] K. L. Tee, U. Schwaneberg, *Comb Chem High Throughput Screen* **2007**, *10*, 197.
[117] S. Q. Liu, Z. H. Meng, Y. X. Fu, K. Q. Zhang, *J Mol Model* **2009**.
[118] W. Parker, P. S. Song, *Biophys J* **1992**, *61*, 1435.
[119] A. K. Shaw, S. K. Pal, *J Photochem Photobiol B* **2007**, *86*, 199.
[120] W. Kabsch, C. Sander, *Biopolymers* **1983**, *22*, 2577.
[121] W. K. Lim, J. Rosgen, S. W. Englander, *Proc Natl Acad Sci U S A* **2009**, *106*, 2595.

[122] a)C. L. Tsou, *Science* **1993**, *262*, 380; b)C. L. Tsou, *Biochim Biophys Acta-Protein Struct Mol Enzymol* **1995**, *1253*, 151.

[123] J. S. Okkels, in *Enzyme Functionality: Design, Engineering and Screening* (Ed.: A. Svendsen), Marcel Dekker Inc., New York, **2004**, 413.

[124] S. Danielsen, M. Eklund, H. J. Deussen, T. Graslund, P. A. Nygren, T. V. Borchert, *Gene* **2001**, *272*, 267.

[125] M. D. Ballinger, J. Tom, J. A. Wells, *Biochemistry* **1995**, *34*, 13312.

[126] D. A. Estell, T. P. Graycar, J. V. Miller, D. B. Powers, J. A. Wells, J. P. Burnier, P. G. Ng, *Science* **1986**, *233*, 659.

[127] M. L. Rollence, D. Filpula, M. W. Pantoliano, P. N. Bryan, *Crit Rev Biotechnol* **1988**, *8*, 217.

[128] S. Tanaka, K. Saito, H. Chon, H. Matsumura, Y. Koga, K. Takano, S. Kanaya, *J Biol Chem* **2007**, *282*, 8246.

[129] P. N. Bryan, M. L. Rollence, M. W. Pantoliano, J. Wood, B. C. Finzel, G. L. Gilliland, A. J. Howard, T. L. Poulos, *Proteins* **1986**, *1*, 326.

[130] L. O. Narhi, Y. Stabinsky, M. Levitt, L. Miller, R. Sachdev, S. Finley, S. Park, C. Kolvenbach, T. Arakawa, M. Zukowski, *Biotechnol Appl Biochem* **1991**, *13*, 12.

[131] K. Q. Chen, A. C. Robinson, M. E. Van Dam, P. Martinez, C. Economou, F. H. Arnold, *Biotechnol Prog* **1991**, *7*, 125.

[132] S. Atwell, J. A. Wells, *Proc Natl Acad Sci U S A* **1999**, *96*, 9497.

[133] L. Zhu, Y. Ji, *Chin J Biotechnol* **1997**, *13*, 9.

[134] a)J. K. Rao, P. Argos, *Biochemistry* **1981**, *20*, 6536; b)M. Mevarech, F. Frolow, L. M. Gloss, *Biophys Chem* **2000**, *86*, 155.

[135] Q. Z. Yao, M. Tian, C. L. Tsou, *Biochemistry* **1984**, *23*, 2740.

[136] C. B. Anfinsen, W. F. Harrington, A. Hvidt, K. Linderstrom-Lang, M. Ottesen, J. Schellman, *Biochim Biophys Acta* **1955**, *17*, 141.

[137] T. E. Creighton, *Biochem J* **1990**, *270*, 1.

[138] Y. L. Zhang, J. M. Zhou, C. L. Tsou, *Biochim Biophys Acta* **1993**, *1164*, 61.

[139] Y. Z. Lin, S. J. Liang, J. M. Zhou, C. L. Tsou, P. Q. Wu, Z. K. Zhou, *Biochim Biophys Acta* **1990**, *1038*, 247.

[140] R. A. Edwards, A. L. Jacobson, R. E. Huber, *Biochemistry* **1990**, *29*, 11001.

[141] M. Gross, G. Auerbach, R. Jaenicke, *FEBS Lett* **1993**, *321*, 256.

[142] a)S. M. Kelly, N. C. Price, *Biochem J* **1991**, *275*, 745; b)O. S. Kwon, S. C. Lo, F. Kwok, J. E. Churchich, *J Biol Chem* **1993**, *268*, 7912.

[143] H. M. Zhou, X. H. Zhang, Y. Yin, C. L. Tsou, *Biochem J* **1993**, *291 (Pt 1)*, 103.

[144] Eliel R. Romero-García, Joel A. Esquivel-Naranjo, Norma Ramírez-Ramírez, Jesús García-Soto, M. Pedraza-Reyes, *Biotechnolgy* **2004**, *3*, 49.

[145] H. Jochens, M. Hesseler, K. Stiba, S. K. Padhi, R. J. Kazlauskas, U. T. Bornscheuer, *Chembiochem* **2011**, *12*, 1508.

[146] A. Dennig, A. V. Shivange, J. Marienhagen, U. Schwaneberg, *PLoS ONE* **2011**, *6*, e26222.

Appendix I: Sequence of subtilisin E synthetic gene aprE (M14)

CATATGTGGA AGAAGATCAT ATTGCACATG AATATGAGCA ATCTGTTCCT TATGGCATTT
CTCAAATTAA AGCGCCGGCT CTTCACTCTC AAGGCTACAC AGGCTCTAAC GTAAAAGTAG
CTGTTATCGA CAGCGGAATT GACTCTTCTC ATCCTGACTT AAACGTCAGA GGCGGAGCAA
GCTTCGTACC TTCTGAAACA AACCCATACC AGGACGGCAT CTCTCACGGT ACGCATGTAG
CCGGTACGAT TGCCGCTCTT AATAACTCAA TCGGTGTTCT GGGCGTTAGC CCAAGCGCAT
CATTATATGC AGTAAAAGTG CTTGCCTCAA CAGGAAGCGG CCCATATAGC TGGATTATTA
ACGGCATTGA GTGGGCCATT TCCAACAATA TGGATGTTAT CAACATGAGC CTTGGCGGAC
CTACTGGTTC TACAGCGCTG AAAACAGTCG TTGACAAAGC CGTTTCCAGC GGTATCGTCG
CTGCTGCCGC AGTCGGAAAC GAAGGTTCAT CCGGAAGCAC AAGCACAGTC AGCTACCCTG
CAAAATATCC TTCTACTATT GCAGTAGGTG CGGTAGACAG CAGCCACCAA AGAGCTTCAT
TCTCCAGCGC AGGTTCTGAG CTTGATGTGA TGGCTCCTGG CGTGTCCGTC CAAAGCACAC
TTCCTGGAGG CACTTACGGC GCTTATAGCG GAACGTCCAT GGCGGCTCCT CACGTTGCCG
GAGCAGCAGT GTTAATTCTT TCTAAGCACC CGAGTTGGAC AAACGCGCAA GTCCGTGATC
GTTTAGAAAG CACTGCAACA TATCTTGGAA ACTCTTTCTA CTATGGAAAA GGGTTAATCA
ACGTACAAGC AGCTGCACAA TAA

Overview of 14 amino acid substitutions

Mutation	Nucleotide	Amino acid
1	GCG 320 GAG	A1E
2	AGT 503/504 ATC	S62I
3	GAT 608/609 GCC	D97A
4	CAA 626 CCA	Q103P
5	GTT 764 GCT	V149A
6	GCC 776 GTC	A153V
7	GGC 814 AGC	G166S
8	AAC 859 GAC	N181D
9	AAC 868 CAC	N184H
10	ATC 931 GTC	I205V
11	AAC 971 AGC	N218S
12	ACT 988 GCT	T224A
13	GCG 1013 GTG	A232V
14	ACT 1037 AGT	T240S

Appendix II: Publication

Zhenwei Li, Danilo Roccatano, Michael Lorenz, Ulrich Schwaneberg. Directed evolution of subtilisin E into a highly active and guanidinium chloride- and sodium dodecylsulfate- tolerant protease. *ChemBioChem*. 2012. 13(5): 691-9

Curriculum Vitae

Personal data

Name: Zhenwei Li

Gender: Male

Date of Birth: 5 August, 1983

Place of Birth: Wuxi, Jiangsu, China

Marital Status: Single

Nationality: Chinese

Education

2011-2012 Ph.D fellow of Protein Engineering, RWTH Aachen University
Aachen, Germany

2009-2010 Ph.D fellow of Protein Engineering, Jacobs University Bremen
Bremen, Germany

2006-2008 M. Sc. in Fermentation Engineering, Jiangnan University
Wuxi, Jiangsu, China

2002-2006 B. Sc. in Biological engineering, Jiangnan University
Wuxi, Jiangsu, China