

# **Directed Evolution of DNA Polymerases for Advancement of the SeSaM Mutagenesis Method and Biotransformations with P450 BM3 Monooxygenase**

**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 Science (MSc)  
Molecular Life Science

**Tsvetan Dinkov Kardashliev**

aus Stara Zagora, Bulgarien

**Berichter:** Universitätsprofessor Dr. rer. nat. Ulrich Schwaneberg  
Universitätsprofessor Dr. rer. nat. Lothar Elling

**Tag der mündlichen Prüfung:** 29. Januar 2015

Diese Dissertation ist auf den Internetseiten der Hochschulbibliothek online verfügbar



## ABSTRACT

Directed evolution is a powerful algorithm to tailor proteins to needs and requirements in industry. Error-prone PCR (epPCR) based methods are the “golden standard” in random mutagenesis due to their robustness and simplicity. Despite their wide use in protein engineering experiments, epPCR methods are limited in their ability to generate highly diverse mutant libraries. This is due to 4 reasons – 1) the redundancy of the genetic code, 2) the low mutagenic frequency and lack of subsequent nucleotide exchanges in a standard epPCR library, 3) the tendency of polymerases to introduce mutations preferentially in certain DNA sequence contexts and 4) the innate transition bias of DNA polymerases leading to conservative amino acid exchanges. Sequence Saturation Mutagenesis (SeSaM) is a random mutagenesis method that has been developed to overcome the aforementioned limitations. SeSaM complements the mutagenic spectrum of epPCR by introducing transversions and consecutive nucleotide substitutions. Indeed, SeSaM libraries enriched in consecutive nucleotide substitutions and transversion mutations were reported; however the frequency of occurrence of such mutations remains low. In particular, the fraction of consecutive transversion mutations accounted for a meager 4.6 %. The increase of the fraction of consecutive transversion mutations is critical in order to advance the SeSaM technology and access previously unattainable sequence space. This can be achieved by the use of DNA polymerases adapted to the requirements of the SeSaM method. The biggest challenge is posed in SeSaM step 3 in which the employed DNA polymerase must be able to elongate consecutive transversion mismatches formed between degenerate base analogs in the primer strand and standard nucleobases in the template strand. The most feasible solution to this problem is to engineer a DNA polymerase capable of efficient consecutive transversion mismatch elongation. The latter has been the main objective pursued in CHAPTER I of this doctoral thesis. The work towards fulfilling this goal included identification from genetic databases of 4 potential candidates from the Y-family of DNA polymerases (exclusively involved in translesion DNA synthesis), followed by expression and preliminary biochemical characterization of the putative polymerases, and, finally, selection of one candidate (Dpo4 from *Sulfolobus solfataricus*) for directed evolution. The protein engineering work comprised development of a novel high-throughput screening system for non-processive DNA polymerases, screening and identification of Dpo4 variants capable of

consecutive mismatch elongation. Finally, the most promising polymerase mutant was used in the preparation of a model SeSaM library. Direct comparison to data generated using an earlier version of the SeSaM protocol indicated a marked improvement in frequency of consecutive transversion mutations (relative increase of 40 %) in the libraries prepared with the identified Dpo4 polymerase mutant. The identified polymerase variant enabled a significant advancement in consecutive transversion generation and, consequently, of the SeSaM random mutagenesis method.

In CHAPTER II of this dissertation, P450 BM3 monooxygenase was studied in the context of regioselective biotransformations of benzenes with isolated enzyme (P450 Project I) and as a whole cell catalyst (P450 Project II). P450 BM3 is a promising biocatalyst which enables challenging chemical reactions, e.g., insertion of an oxygen atom into a non-activated C-H bond. In P450 Project I, P450 BM3 in purified form was employed in the synthesis of mono- and di-hydroxylated products from six monosubstituted benzene substrates. A P450 BM3 mutant (P450 BM3 M2 (R47S/Y51W/A330F) proved to be promiscuous and highly regioselective (95 % - 99 %) hydroxylation catalyst achieving good overall yields (up to 50 % with benzenes and  $\geq 90$  % with phenols). The developed process showed promising potential for sustainable synthesis on a semi-preparative scale of valuable chemical precursors, i.e., phenols and hydroquinones, even prior to bioprocess optimizations. Nevertheless, productivity, especially with non-halogenated substrates, would need to be further improved. Especially, whole cell cofactor regeneration systems have to be developed in order to bring this attractive synthesis route closer to industrial exploitation. The issue of cofactor regeneration served as a stimulus to initiate follow-up project dealing with whole cell catalysis with P450 BM3. The requirement for expensive reduced cofactor (NAD(P)H) in equimolar amounts is a major disadvantage of P450 BM3 preventing its wider use in organic synthesis. The NAD(P)H dependency of P450 BM3 necessitates the use of whole cells in preparative synthesis in order to achieve cost-efficient cofactor regeneration. However, the semi-permeable nature of the outer membrane of industrially relevant bacteria such as *E. coli* limits their use as whole cell systems. The issue of substrate permeability in whole-cell biocatalysts has been addressed in this thesis by co-expressing a large passive diffusion channel, FhuA  $\Delta 1-160$ , in the outer membrane of *E. coli*. The influence of the channel protein on P450 BM3-catalyzed conversions of 2 monosubstituted benzenes has been investigated in P450 Project II. Preliminary experiments indicated that the co-expression in the outer *E. coli*

membrane of FhuA  $\Delta$ 1-160 had a positive effect, possibly of global nature, on the mass transfer in whole cell biotransformations (up to 10-fold relative increase of product titers). The presented drawback of whole-cell biocatalysis has not been extensively addressed to date, and with these preliminary and promising first results, we hope to entice further interest in the topic.

In summary, this doctoral thesis addresses three questions from the related fields of protein engineering and biocatalysis:

- 1) Advancement of diversity generation methods in directed protein evolution;
- 2) Establishing biocatalytic process for oxy-functionalization of generally unreactive aromatic C-H bonds;
- 3) Developing a strategy to improve the mass transfer across the outer membrane of *E. coli* in whole cell oxy-functionalization of aromatic compounds.

## PUBLICATIONS

### Published:

- 1) Kardashliev T., Ruff A. J., Zhao J., Schwaneberg U., "A High-Throughput Screening Method to Reengineer DNA Polymerases for Random Mutagenesis", 2014, ***Molecular Biotechnology***, **56(3):274-83**
- 2) Ruff A. J., Kardashliev T., Dennig A., Schwaneberg U. "The Sequence Saturation Mutagenesis (SeSaM) method", 2014, ***Methods in Molecular Biology Vol. 1179: Directed Evolution Library Creation 2nd ed: 45-68***
- 3) Zhao J., Kardashliev T., Ruff A. J., Bocola M., Schwaneberg U., "Lessons from diversity of directed evolution experiments by an analysis of 3,000 mutations", 2014, ***Biotechnology and Bioengineering***, **111: 2380-2389**
- 4) Cheng F., Kardashliev T., Pitzler C., Shehzad A., Lue H, Bernhagen J., Zhu L., Schwaneberg U., "A competitive flow cytometry screening system for arginine-metabolizing enzyme in cancer treatment", 2015, ***ACS Synthetic Biology***, **DOI: 10.1021/sb500343g**

### Presented as posters and oral presentations:

- 5) Kardashliev T., Ruff A. J., Schwaneberg U., "Engineering of DNA polymerases for applications in random mutagenesis", Biotrans 2013 International Conference, Manchester, UK, ***poster presentation***.
- 6) Kardashliev T., Ruff A. J., Schwaneberg U., "A high-throughput screening system for engineering of DNA polymerases", IBN 2013 International Conference, Hamburg, Germany, ***poster presentation***.
- 7) Kardashliev T., "Advancement of Sequence Saturation Mutagenesis (SeSaM) method for genetic diversity generation", Biokatalyse2013 Scientific Cluster Conference, Hamburg, Germany, ***oral presentation***.

### Planned for publication:

- 8) Kardashliev T., Dennig A., Halmschlag B., Ruff AJ., Schwaneberg U. "One step synthesis of hydroquinones from benzenes with a single cytochrome P450 BM3 variant", 2015, submitted to ***Advanced Synthesis & Catalysis (in Feb 2015)***.
- 9) Kardashliev T., Dennig A., Arlt M., Ruff AJ., Schwaneberg U. "Improving the mass transfer in whole-cell biocatalysis by expression of a passive diffusion channel in the outer membrane of *E. coli*", 2015

# TABLE OF CONTENTS

ABSTRACT .....	1
PUBLICATIONS .....	4
TABLE OF CONTENTS .....	5
CHAPTER I. Engineering of DNA Polymerases for Application in Directed Protein Evolution.....	8
1. Introduction to protein engineering .....	9
1.1. Laboratory evolution .....	9
1.2. Challenges in directed protein evolution .....	11
1.3. Strategies for laboratory evolution .....	12
1.4. The Sequence Saturation Mutagenesis (SeSaM) method .....	22
1.5. DNA polymerase-based technologies and high-throughput screening systems for DNA polymerases .....	31
1.6. DNA polymerases in diversity generation for directed evolution .....	33
1.7. A concise overview of Y-family of DNA polymerases .....	33
2. Materials and methods .....	38
2.1. Chemicals.....	38
2.2. Bacterial strains, plasmids and list of primers.....	38
2.3. Cloning procedures.....	42
2.4. Pilot expression and purification of putative Y-polymerase genes.....	42
2.5. Thermostability valuation in crude cell extract.....	43
2.6. <i>In vitro</i> expression of Dpo4 .....	43
2.7. Forward primer extension assay .....	44
2.8. Compartmentalized self-replication.....	44
2.9. PicoGreen® assay.....	46
2.10. 96-well microtiter plate-format Scorpion probe fluorescent screening assay for DNA polymerase activity .....	47
2.11. Construction and screening of epPCR libraries of Dpo4 .....	48
2.12. Construction of saturation mutagenesis libraries of Dpo4 .....	49
2.13. Purification of Dpo4 variants for biochemical characterization.....	50
2.14. Detailed SeSaM-Tv protocol.....	51
3. Results and discussion .....	71
3.1. Identification and preliminary characterization of Y-family polymerases with suitability for SeSaM .....	71

3.2. Development of a screening system for high-throughput screening of Y-family polymerases .....	81
3.3. Engineering of the Y-family polymerase Dpo4 for application in directed evolution.....	91
4. Summary and conclusions: Engineering of DNA polymerase for Application in directed protein evolution.....	104
CHAPTER II. Regioselective Synthesis of Mono- and Di-hydroxy Products from Benzenes using P450 BM3 Monooxygenase in Purified Form and as Whole Cell Catalyst .....	107
6. Introduction.....	108
6.1. Industrial biotechnology.....	108
6.2. Cytochrome P450 Monooxygenases.....	110
6.3. Cytochrome P450 (P450 BM3) .....	112
6.4. P450 BM3 catalyzed hydroxylation of aromatic hydrocarbons .....	113
6.5. The “cofactor challenge” in catalysis with P450 BM3 .....	114
6.6. Whole cell catalysis .....	115
6.7. The outer membrane (OM) of gram-negative bacteria .....	115
6.8. Outer membrane permeability issues in whole cell catalysis .....	116
6.9. Ferric-hydroxamate uptake protein component A (FhuA).....	118
6.10. Aim and objectives .....	120
7. Materials and methods .....	122
7.1. Chemicals.....	122
7.2. Cloning of FhuA WT and FhuA $\Delta$ 1-160 in pALXtreme-1a-P450 BM3 M2 .....	122
7.3. Shake flask expression of P450 BM3 wild type and M2.....	123
7.4. Shake flask co-expression of P450 BM3 and outer membrane protein FhuA $\Delta$ 1-160.....	123
7.5. Purification of P450 BM3.....	123
7.6. BCCE activity assay .....	124
7.7. Flow cytometry analysis of cell populations expressing P450 BM3 and co-expressing P450 BM3 and FhuA channel protein variants.....	124
7.8. Long term conversion of toluene and anisole by whole cells expressing P450 BM3 or co-expressing P450 BM3 and FhuA channel protein .....	125
7.9. Long term conversion of benzenes or phenols to hydroquinone with purified P450 BM3 .	125
7.10. Carbon monoxide differential spectroscopy .....	126
7.11. Determination of NADPH turnover frequency, coupling efficiency and initial turnover rate .....	127
7.12. HPLC measurement of phenols and hydroquinones.....	128
7.13. GC analysis.....	128
8. Results and discussion.....	130



8.1. P450 project I: One step enzymatic synthesis of hydroquinones from monosubstituted benzenes.....	130
8.2. P450 project II: Improving the rate of P450-mediated whole cell biotransformations by co-expressing a passive diffusion channel in the outer membrane of <i>E. coli</i> .....	145
9. References.....	153
FURTHER SCIENTIFIC CONTRIBUTIONS .....	167
FINAL SUMMARY AND CONCLUSIVE REMARKS .....	168
APPENDIX .....	171
ACKNOWLEDGEMENTS .....	179
STATEMENT .....	180
LEBENS LAUF .....	181

## **CHAPTER I. Engineering of DNA Polymerases for Application in Directed Protein Evolution**

*Parts of this chapter have been published in the book "Directed Evolution Library Creation: Methods and Protocols", 2nd Edition (Eds.: D. Ackerley, J. Copp,, E. Gillam) Methods in Molecular Biology Vol 1179, Humana Press, Totowa, 2014 and in the article "A High-Throughput Screening Method to Reengineer DNA Polymerases for Random Mutagenesis", Molecular Biotechnology, 2014, Volume 56, Issue 3, pp 274-283 .*

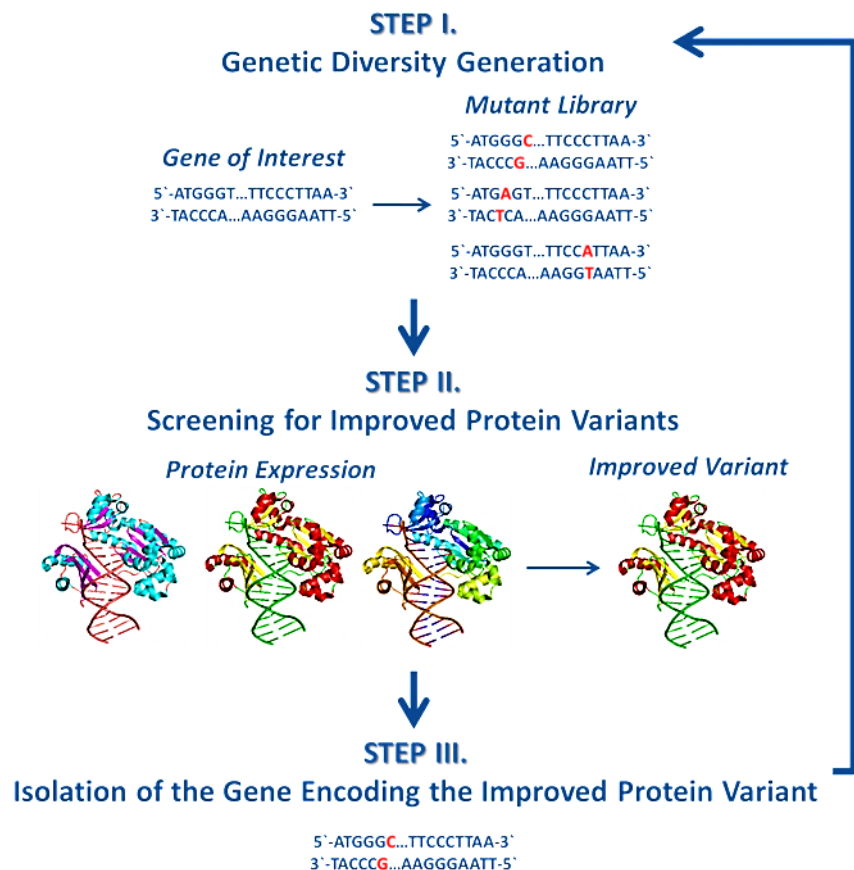
## 1. Introduction to protein engineering

In the past 25 years, biocatalysts have increasingly been used in industrial processes, e.g., the production of specialty and fine chemicals, pharmaceuticals, vitamins, detergents, cleaning agents, biofuels, and in the refinement of textiles, leather and paper, among others (Erickson et al., 2012; Philp et al., 2013; Soetaert and Vandamme, 2010). Naturally occurring enzymes have evolved to suit the needs of the organisms they are derived from and are usually not fit for direct application in industrial processes. Nevertheless, biocatalysts become increasingly important for industrial production of chemicals. Their full potential can be realized by adapting promising enzymes to the needs of industry by means of protein and bioprocess engineering (Bommarius et al., 2011; Bornscheuer et al., 2012; Lan Tee and Schwaneberg, 2007). Laboratory protein evolution relies heavily on methodologies for generation and screening of mutants libraries (Ruff et al., 2013a; Shivange et al., 2009; Wong et al., 2006a). In line with some of the main objectives of this thesis, methods for genetic diversity generation including approaches to mutant library creation, drawbacks, advantages and remaining challenges are discussed in the following sections.

### 1.1. Laboratory evolution

The understanding of how life on Earth developed has been vastly redefined since the 19<sup>th</sup> century. Despite early criticism (Hull, 1973), Charles Darwin's theory of evolution from 1859 is now a widely accepted idea that not only helps to explain how the tremendous diversity of life forms on Earth has developed but also to inspire numerous new scientific concepts (Bowler, 1996). For instance, natural evolution has served as a pivotal point for the establishment of the modern field of laboratory protein evolution (Bornscheuer and Pohl, 2001; Jaeger et al., 2001). The idea that biological molecules can be tailored to one's needs by iterative cycles of mutation and selection has been postulated as early as the 1960s (Mills et al., 1967) and since the early 1990s, with the advancement of methods for *in vitro* DNA synthesis and amplification, cloning and recombinant protein expression (Bornscheuer et al., 2012), laboratory evolution really took off. Until the present moment, protein engineering has been applied on numerous occasions to adapt biocatalysts for industrial and medical applications (Arango Gutierrez et al., 2013; Bommarius et al., 2011; Gomes et al., 2012; Zhu et al., 2010).

A typical directed evolution experiment comprises three main steps (Fig. 1). Every laboratory evolution campaign starts with generation of genetic diversity of a target sequence by random mutagenesis, focused mutagenesis or recombination of highly homologous sequences (Ruff et al., 2013b). The functionally expressed genetic diversity is then screened on protein level using a validated high-throughput screening or selection system. Finally, the genes encoding improved variants are isolated, and, if required, subjected to further rounds of diversity generation and screening until the desired improvement of the target property is achieved. A successful evolution campaign should have its final goal set such that the desired function is physically, biologically and evolutionally plausible (Lan Tee and Schwaneberg, 2007). When it comes to technical considerations, the successful outcome is largely determined by two factors - the quality of the mutant libraries and the robustness, throughput and similarity of the employed screening system to the “real-life” application.



**Figure 1.** Schematic illustrating the main steps of a directed evolution campaign: 1) genetic diversity generation; 2) screening of a pool of protein variants; 3) isolation of gene(s) encoding mutants with improvement in the desired characteristics. The outlined procedure is usually performed in an iterative manner.

## 1.2. Challenges in directed protein evolution

The limitations in protein engineering can be classified in two categories - challenges of diversity generation and challenges arising from the limited screening capabilities of commonly used screening platforms.

### 1.2.1. The importance of diversity generation

The diversity challenge of directed protein evolution can be illustrated through a simple example in the context of random mutagenesis. In order to thoroughly randomize a short polypeptide (say a 50-mer) at every position and to all other 19 amino acids, a protein sequence space of  $2 \times 10^{50}$  unique variants is generated. A library of such magnitude covers the theoretically attainable genetic complexity. Consequently, this diversity provides the maximal number of possible “adaptive pathways” and provides the possibility to circumvent local fitness minima (Wong et al., 2006a). In contrast, mutant libraries that do not cover the complete genetic diversity (which is the realistic output of the majority of random mutagenesis methods (Verma et al., 2012)) will only yield phenotypes with small or medium improvement of fitness. This largely disallows reaching an absolute optimum (Wong et al., 2006a). Thus, mutagenesis methods that can render libraries with broad diversity are necessary to identify rare and beneficial variants in laboratory evolution experiments. At present most methods, especially those for random mutagenesis, have limited potential to cover the theoretical genetic sequence space (Zhao et al., 2014).

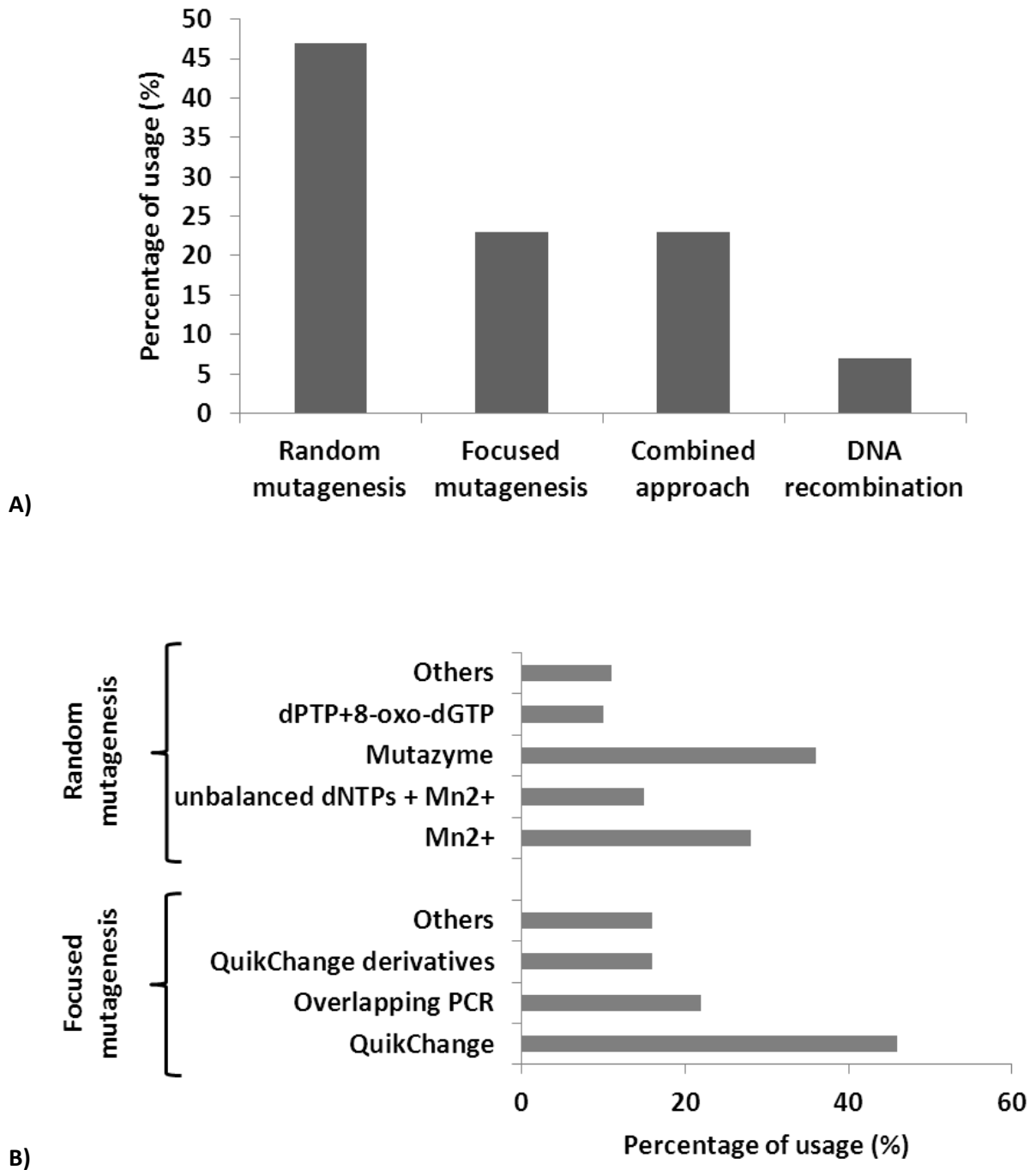
### 1.2.1. Challenges and requirements for screening of mutant libraries

On the protein level, the screening of even a fraction of the theoretically attainable genetic diversity is understandably a resource- and labor-intensive process. Even the most advanced screening and selection systems (e.g., yeast/ *E. coli* hybrid systems (Stynen et al., 2012), molecular display (Bratkovič, 2010) and FACS (Agresti et al., 2010; Ruff et al., 2012a)) are limited in their throughput. A screening rate of  $10^7$  events per hour can be seen as the upper limit of screening systems (Tolia and Joshua-Tor, 2006). Irrespective of its throughput, every screening or selection system needs to fulfill a number of criteria in order to be considered suitable for application in laboratory evolution. Firstly, these methodologies should be reproducing the “real-life” application conditions as close as possible in order to ensure the

identification of protein variants with desired properties. For example, the use of a dummy substrate in laboratory evolution, e.g., employing a fluorescently labelled disaccharide that holds resemblance to the more complex target polysaccharide, often results in a quick identification of protein mutants with improved kinetic parameters for the molecule used in screening. However, similar or decreased performance with the complex substrate is often observed (Leemhuis et al., 2009). Secondly, the screening system must ensure maintenance of the linkage between genotype and phenotype during the entire screening procedure. If this connection is lost, it is impossible to identify the amino acid exchanges in the target protein contributing to the improvement of its fitness. Thirdly, a validated screening system should give reproducible results and confer a low coefficient of variation to the overall process. Additional desired characteristics of a screening system are an ultra high-throughput, possibility for automation and low costs in terms of time and resources invested in the process. Current state-of-the-art screening technologies fulfill these criteria but remain largely limited to sampling only a fraction of the size of diverse mutant libraries (Leemhuis et al., 2009; Ruff et al., 2013a; Lan Tee and Schwaneberg, 2007).

### **1.3. Strategies for laboratory evolution**

Laboratory evolution of proteins is generally conducted by one of three general strategies (Bommarius et al., 2011; Hilvert, 2013; Ruff et al., 2013a; Tee and Wong, 2013) – 1) rational protein design (employing focused mutagenesis and, in rare cases, *de novo* protein design), 2) directed evolution (employing random mutagenesis) and 3) shuffling of homologous sequences (employing DNA recombination). It should be noted that, these strategies for protein engineering differ in the way genetic diversity is generated; however they are similar in other aspects such as the requirement for a high-throughput screening system and the follow-up steps after identification of protein variants with improved properties. Statistical data showing the popularity of different mutagenesis strategies for the 5-year period between 2008 and 2013 is presented in Fig. 2.



**Figure 2.** Mutagenesis methods embraced in 100 randomly selected papers on protein engineering in the period 2011 - 2013. **A)** Distribution according to diversity generation strategy, also reflecting used approach (random vs rational); **B)** Percentage of usage of the most frequently utilized methods classified as random mutagenesis or focused mutagenesis methods ( adapted from Tee and Wong, 2013).

### 1.3.1. Rational mutagenesis

Rational protein design is a knowledge-driven mutagenesis approach that heavily relies on available information of the target protein. Empirical data from structural, biochemical and biophysical characterizations, previous mutagenesis studies, high-resolution crystal structures and computational predictive algorithms are used in order to identify potential regions within the protein chain for localized mutagenic studies. Such regions can be partially or thoroughly randomized in a site-specific manner to generate focused libraries with a finite size that can be screened using a validated screening method. This approach is particularly helpful for elucidation of structure/function relationships in proteins but has been proven useful in generation of mutants with improvement in desired properties as well (Bornscheuer and Pohl, 2001; Shivange et al., 2009). From the viewpoint of the experimentalist, rational design translates into reduced effort and increased efficiency of identifying improved variants (Lutz, 2010). This approach has the potential to largely eliminate the need for screening assays with ultra-high-throughput, to provide a framework to predict and rationalize experimental findings and, consequently, to alter the field of protein engineering from discovery-based to hypothesis-driven (Lutz, 2010). Despite its enormous potential, the use of this approach to engineer proteins remains limited because of the number of pre-requisites it requires (e.g., detailed structural and functional knowledge of the target protein) and the vast computational power needed to predict the effect of structural changes on the functional performance of complex biopolymers. This is reflected by the fact that rational design mutagenesis is the methodology of choice in less than a quarter of recent protein engineering studies (Fig. 2A). A few of the most prominent examples of methods for focused mutagenesis are discussed in the next paragraphs.

#### 1.3.1.1. QuikChange® PCR

The diversity generation technique in more than half of all rational design studies (Fig. 2A) is a whole plasmid PCR amplification method termed QuikChange® PCR (Wang and Malcolm, 1999). This method enables a thorough or partial exchange of a single residue in a protein sequence to any other amino acid or a subset of amino acids in a single step. It employs complementary (or partially complementary) primer pair that harbors a sense and antisense mutant codon. A high-fidelity DNA polymerase is used to amplify the entire plasmid carrying



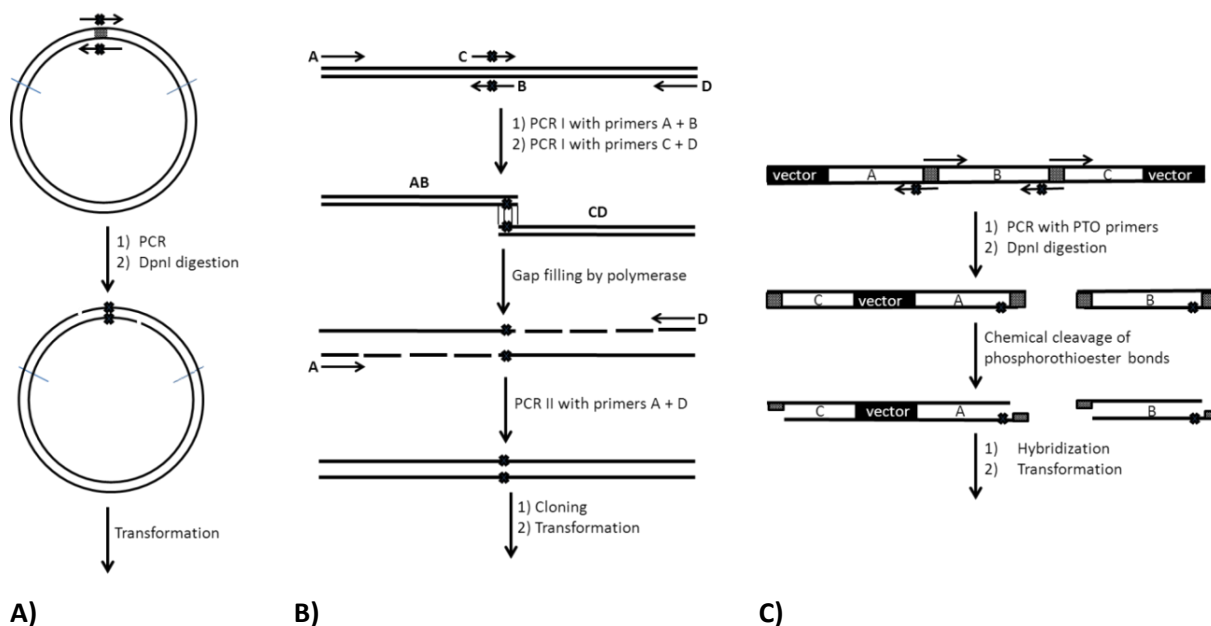
the target sequence. The resulting “nicked” double stranded DNA, i.e., containing two non-overlapping DNA breaks, can be directly transformed into competent cells after enzymatic digestion of the template plasmid (using DpnI endonuclease) and the non-overlapping breaks are repaired *in vivo* in order to obtain a circular mutated plasmid (Fig 3A). Derivatives of this classic method have been developed which enable mutli site-directed mutagenesis (of up to 5 sites) often at the cost of decreased mutagenic efficiency (Tee and Wong, 2013).

### 1.3.1.2. Overlap-extension PCR (oePCR)

oePCR (Fig. 2B) is the second most common approach for diversity generation in rational design studies (Urban et al., 1997). This method employs two primer pairs whereas one primer from each pair carries the mutagenic codon. The four primers are used in PCRs to obtain double stranded DNA products that are mixed and hybridized to form two heteroduplexes, each of which contains the desired mutagenic codon. The overlapping 3'- and 5'-ends of each heteroduplex are filled by application of a polymerase then a second PCR reaction with the non-mutated primer pair is performed to amplify the mutated DNA.

### 1.3.1.3. OmniChange

A promising alternative to QuikChange® and oePCR is the method OmniChange (Dennig et al., 2011). OmniChange employs sets of primers containing phosphorothioated (PTO) nucleotides at their 5`-ends and harbor the randomization codon in the unmodified part to amplify fragments that constitute the entire vector and target gene sequence (Fig. 3C). Due to the specific primer design these fragments have partially identical sequences at their ends which, upon cleavage of the PTO fragments, become complementary and hybridize to each other. The DNA hybrids can be directly transformed in competent cells where the DNA nicks are repaired. Circular plasmid harboring one or more modified codons (depending on experimental design) is eventually rendered. Using this method, up to five sites have been randomized with excellent mutagenic efficiency (100% reported) and coverage of the theoretical sequence space (NNK coverage of 65 to 85% from analysis of 48 clones is reported).



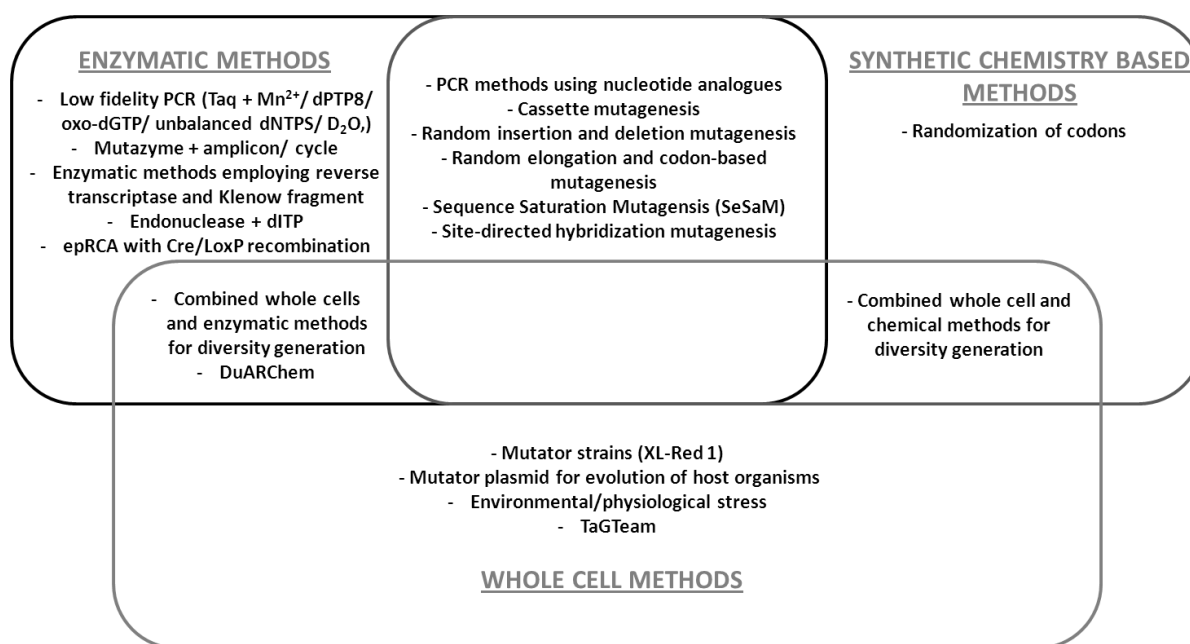
**Figure 3.** Graphic representation of selected methods for diversity generation in rational protein design. **A)** QuikChange®; **B)** Overlap-extension PCR; **C)** OmniChange.

### 1.3.2. Directed protein evolution

Directed evolution is a widespread and well-established approach to modify biomolecules to meet application criteria in chemistry and biology (Jaeger et al., 2001; Obeid et al., 2011; Lan Tee and Schwaneberg, 2007). In contrast to rational protein engineering approach which is largely knowledge-driven, directed protein evolution aims at mimicking the natural evolution cycle and relies on random insertion of mutations in the target sequence to generate large number of target protein variations. Therefore, structural or functional information about the target molecule are not a pre-requisite for carrying our directed evolution. In random mutagenesis, libraries are produced and subjected to screening using a high-throughput screening system in order to identify variants with improvement in the desired performance parameters. Iterative cycles of random mutagenesis and screening are usually necessary to identify biomolecules that meet application criteria.

Random mutagenesis libraries can be generated using a variety of methods. These can be generally grouped in 4 categories (Wong et al., 2007a) - enzyme-based methods, chemical methods, whole cell methods and hybrid methods (Fig. 4). Enzymatic methods most frequently employ DNA polymerases and to a lesser extend other DNA modulating enzymes

in order to introduce mutations in DNA strands. The techniques that are classified in the first category are the most versatile and commonly used in random mutagenesis. This is due to their robustness and simplicity. The “golden standard” in this category is the error-prone PCR (epPCR) techniques. In chemical methods, mutations are usually introduced through DNA-modifying agents or degenerate nucleotide analogues and whole cell methods rely on engineered “mutator strains” to introduce nucleotide substitutions *in vivo*. Most of the aforementioned methods fail to generate truly diverse libraries on the protein level (Verma et al., 2012) which in turn necessitates the advancement of existing technologies and the development of completely new strategies for random mutagenesis. A few examples of random mutagenesis methods are discussed in more details in the following sections.



**Figure 4.** Classification of random mutagenesis methods (adapted from Tee and Wong, 2013; Wong et al., 2006a)

### 1.3.2.1. Error-prone PCR (epPCR)

epPCR introduces mutations by altering the fidelity of the DNA polymerase (Pritchard et al., 2005). The fidelity of polymerases can be affected by modulating the PCR mastermix composition, e.g., by addition of divalent ions such as Mn<sup>2+</sup> in the or by disbalancing dNTP concentrations (Wang et al., 2006a). Sub-optimal reaction conditions can lead to

misincorporation of nucleotides during the reaction and yield randomly mutated products. In either of the epPCR methods, the frequency of incorporation of nucleotide analogs and, consequently, the number of errors introduced into the sequence can be tuned by the composition of the mastermix. In a standard directed evolution experiment by epPCR, the mutational frequency is normally set to 1 - 3 base pair substitutions per kilobase of DNA (Zhao et al., 2014). The relative simplicity of use and versatility makes epPCR the most widely applied mutagenesis method. However, epPCR has its drawbacks. Limitations and drawbacks of methods for diversity generation are discussed later in the introductory section.

### **1.3.2.2. DNA shuffling and staggered extension process (StEP)**

DNA shuffling is an enzyme-based recombination method in which genes with high degree of homology are randomly cleaved then reassembled in a PCR whereas DNA fragments serve to prime each other's elongation (Stemmer, 1994). Recombination occurs when a fragment derived from one gene anneals to fragments from another gene, causing a crossover event. This method is generally robust, flexible and tends to remove non-essential mutations by back-crossing to parent, however it suffers from low crossover rate, bias to crossovers in highly homology regions and high percentage of remaining unmutated DNA which limits its use.

A variation of the DNA shuffling which requires no fragmentation of parental genes is termed StEP (Zhao et al., 1998). This method uses flanking primers for partial replication of target sequence with very short extension times. Short extensions generate truncated DNA fragments which are separated from template strand and re-anneal to a different template copy acting as primers. The growing DNA strand binds to different templates multiple times and accumulates sequences from different parent genes. Two different low-fidelity polymerases are routinely used in order to reduce mutational bias, such as, Taq and Mutazyme, which has partially complementary mutational spectra. Even though this method has certain advantages over DNA shuffling, it largely shares the drawbacks of the classical homologous recombination method.

### 1.3.2.3. Sequence Saturation Mutagenesis (SeSaM)

SeSaM is a four-step chemo-enzymatic mutagenesis method that is capable of targeting all nucleotide positions in a target sequence with equal probability and independently of polymerase bias and with the possibility of introducing subsequent nucleotide exchanges (Wong et al., 2004, 2008b). This method is of special significance for the objectives of this thesis; therefore it is extensively discussed in a dedicated section (section 1.4).

### 1.3.3. Limitations of random mutagenesis methods

Three reasons for the limited diversity of most random mutagenesis methods can be pinpointed.

Firstly, the obtainable genetic diversity by methods that sporadically introduce transition (i.e., purine-to-purine or pyrimidine-to-pyrimidine) nucleotide exchanges in a sequence (the case with most existing methods (Wong et al., 2006a)) leads to preservation of chemical functionality on the protein level. This is a consequence of the organization of the genetic code (Wong et al., 2006b, 2007a). A transition biased library has been shown to contain half the diversity on protein level that could be theoretically obtained in a transversion (i.e., pyrimidine-to-purine nucleotide exchange or vice versa) biased library. In quantitative terms, a transition biased library contains only 11% of the effective protein space. In a transversion biased library, the same accounts for 21.5% of the effective protein space (Wong et al., 2007b). In addition, a transition biased library contains 34.9% silent mutations while in a transversion biased library there are only 15.3% silent mutations (Wong et al., 2007b). The average number of unique amino acid substitutions per protein position in a transition biased library is 2.2 versus 4.7 for the transversion biased library (Wong et al., 2007b). Secondly, random mutagenesis methods seldom introduce subsequent nucleotide exchanges. This causes a major limitation in terms of obtainable genetic diversity as only 40% of the theoretical diversity can be explored by single point mutations (Wong et al., 2006c, 2007a). The contribution of subsequent nucleotide exchanges is well illustrated in Fig. 5 on the example of the codon TTA (encoding Leu). A maximum of 9 different codons can be obtained by targeting one nucleotide in this codon; this number increases by a factor of 5 (45 possible amino acid substitutions) with two nucleotide exchanges in the same codon. Thirdly, methods that employ DNA polymerases result in uneven distribution of mutations

over the gene sequence and generation of mutational “hotspots” (Rogozin and Pavlov, 2003a).

Efforts to alleviate the limitations of epPCR have resulted in the development of the Sequence Saturation Mutagenesis (SeSaM) method (Wong et al., 2004). In the context of random mutant library generation, the Sequence Saturation Mutagenesis (SeSaM) represents the most advanced random mutagenesis method. By means of SeSaM, libraries enriched in transversion nucleotide exchanges can be generated. Transversion mutations translate into much higher diversity on the protein level (Wong et al., 2007b). An additional beneficial feature of this mutagenesis method is its ability to introduce subsequent nucleotide substitutions which further increases the obtainable diversity on DNA and subsequently on protein level (Wong et al., 2007c).

The decisive factor in the further advancement of SeSaM is the identification of a DNA polymerase specifically adapted to the requirements of this mutagenesis method. Fulfilling this objective would enable faster, efficient and more cost-effective mutagenesis services for industrial biotechnology in general. The advancement has been a major goal of this thesis. Before presenting the achievements made in this direction, a comprehensive description of SeSaM and DNA polymerases is given in the following sections of this chapter.



## 1.4. The Sequence Saturation Mutagenesis (SeSaM) method

Sequence Saturation Mutagenesis (SeSaM) is a random mutagenesis method specifically developed to overcome the limitations of existing error-prone PCR (epPCR) protocols. SeSaM is advantageous with respect to 1) elimination of mutagenic “hotspots”, 2) increase in frequency of subsequent nucleotide substitutions, 3) control over the mutational bias through the utilization of degenerate base analogs and, consequently, 4) the prospect of generating transversion enriched mutant libraries. These advanced features lead to chemically diverse mutant libraries on the protein level, essentially making SeSaM a complementary technology to transition biased epPCR mutagenesis methods.

The advancement of the SeSaM technology, especially with regard to increasing the fraction of consecutive mutations, was a main objective of this thesis. In order to fully grasp the essence of the method, a comprehensive overview of the SeSaM protocol at its state prior to the beginning of this thesis as well as the remaining challenges are deliberated in this section.

### 1.4.1. Overview

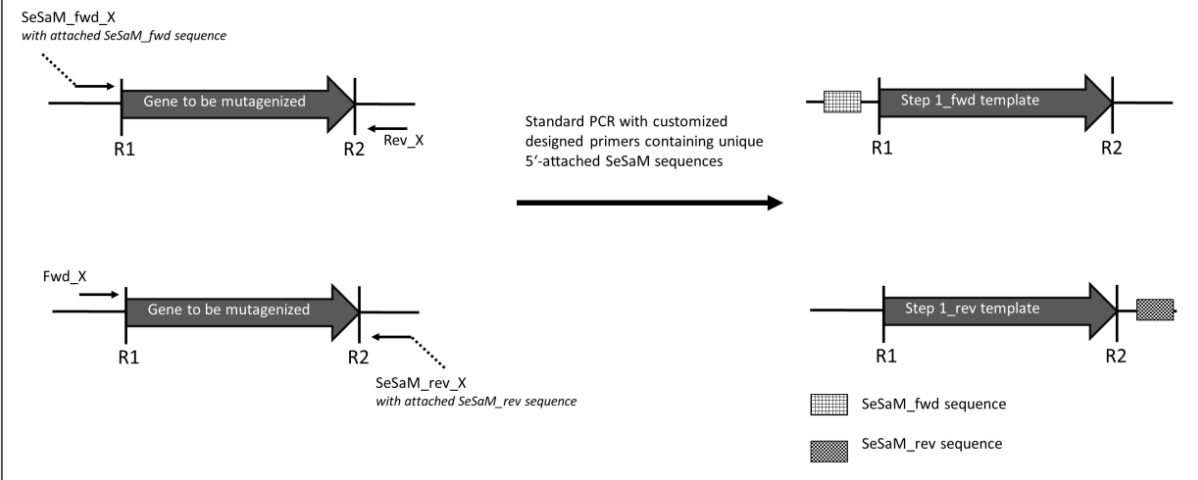
Directed evolution is a powerful algorithm for tailoring proteins to practical needs and requirements. Error-prone PCR (epPCR) based methods are the standard for random mutagenesis due to their robustness and simplicity (Rasila et al., 2009). Despite some impressive success stories (Glieder et al., 2002; Zhu et al., 2010), these methods are generally limited in their ability to generate highly diverse mutant libraries for the 4 reasons: the redundancy of the genetic code; commonly low mutagenic frequency in epPCR libraries which disallows subsequent nucleotide exchanges (Ruff et al., 2013a); the innate transitional bias of DNA polymerases used in epPCR (transition mutations typically lead to conservative amino acid substitutions, for example a hydrophobic to hydrophobic amino acid exchange) (Shivange et al., 2009; Wong et al., 2006a); mutagenic “hot spots”, due to the propensity for polymerases to introduce mutations preferentially in certain DNA sequence contexts (Rogozin and Pavlov, 2003b). Taken together, these factors significantly limit the chemical diversity of random mutagenesis methods (Shivange et al., 2009; Wong et al., 2007a). A good indication for the latter is the observation that on average only 7 of the 19 possible



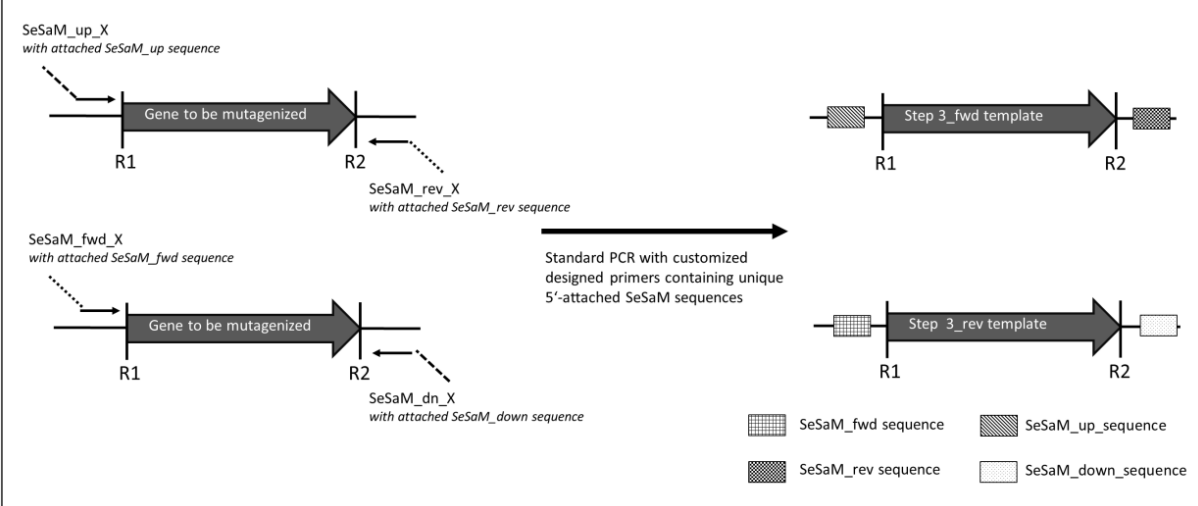
amino acid substitutions are typically achieved by epPCR (Rasila et al., 2009). To evaluate the limitations and facilitate the selection of a suitable random mutagenesis method, computational tools such as the mutagenesis assistant program (MAP) (Wong et al., 2006b) and its improved version, MAP3D, have been developed. MAP3D is implemented on a server (<http://map.jacobs-university.de/map3d.html>) (Verma et al., 2012) and can be used as a benchmarking system for random mutagenesis methods on the protein level by comparing amino acid substitution patterns.

Sequence Saturation Mutagenesis (SeSaM) is a chemo-enzymatic mutagenesis method that aims at eliminating most limitations of current state-of-the-art random mutagenesis techniques. SeSaM is a four step process that employs selective chemical fragmentation of DNA, phosphorothioate and degenerate nucleotide analogs as well as several enzymatic steps in order to introduce nucleotide exchanges with a desirable mutational bias, distribution and frequency. In step 1 of SeSaM, a pool of single-stranded DNA (ssDNA) fragments with various gene lengths is generated in a PCR that employs a biotin-labeled forward primer and a pre-defined mixture of standard and cleavable phosphorothioate nucleotides. Phosphorothioester bonds within the resulting PCR products can be cleaved selectively in the presence of iodine, under alkaline conditions and at elevated temperature. The ssDNA fragments are subsequently isolated from non-biotinylated DNA strands using streptavidin-coated magnetic beads. In step 2, the purified ssDNA fragments are “tailed” with a degenerate base of choice in a reaction catalyzed by terminal deoxynucleotidyl transferase (TdT). In step 3, the “tailed” ssDNA fragments are elongated to full gene length. Step 3 amplicons serve as templates in step 4 where the incorporated nucleotide analogs are replaced by standard nucleotides in a final PCR. The two preliminary experiments necessary for SeSaM template generation and determination of the optimal concentration of phosphorothioate nucleotides (Fig. 6) as well as SeSaM steps 1 to 4 (Fig. 7) are shown below.

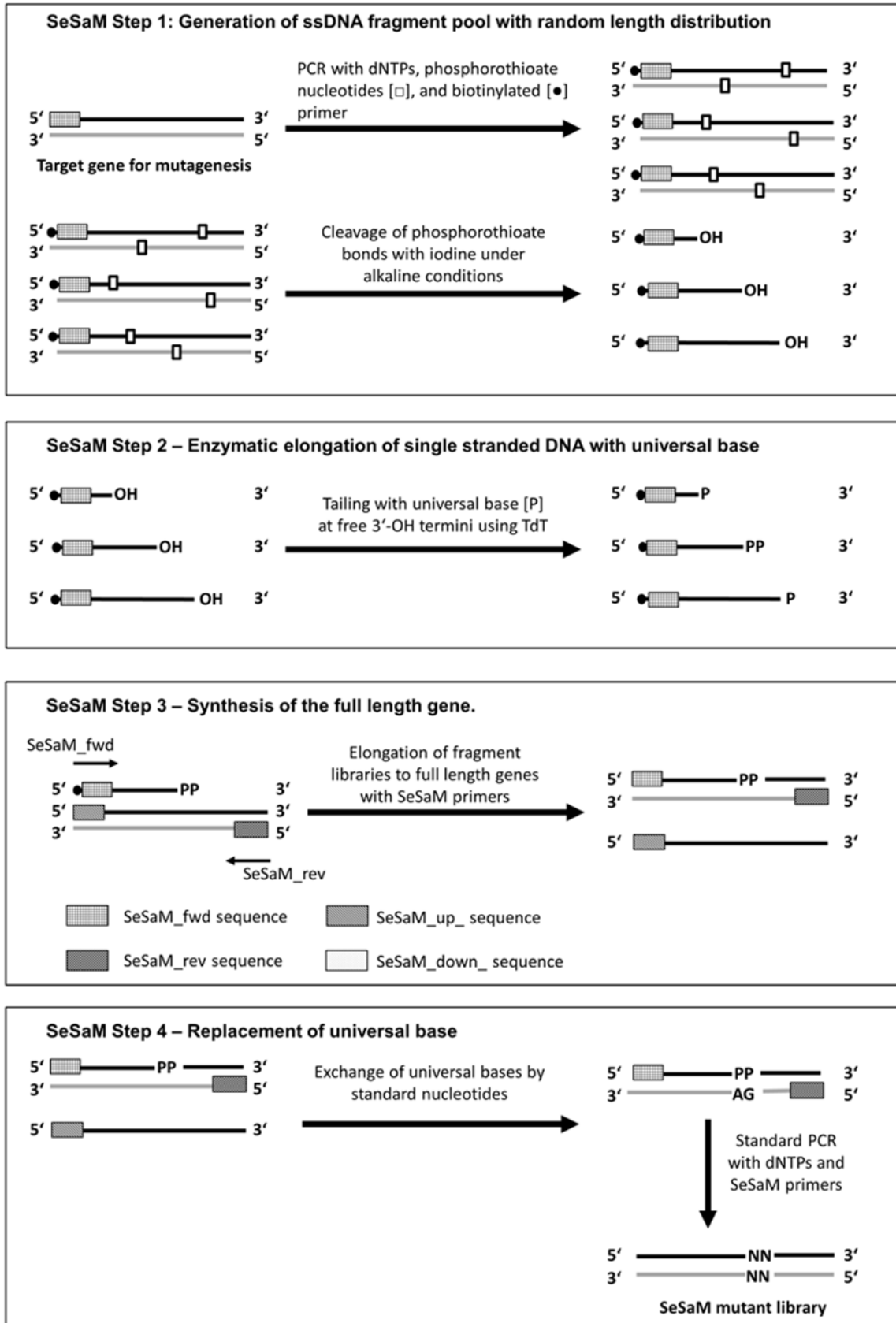
### 3.1.1 SeSaM Preliminary Experiment I: SeSaM step 1 and step 3 template generation



### 3.1.2 SeSaM Preliminary Experiment II: determination of the optimal concentration of phosphorothioate nucleotides for Step 1 of SeSaM



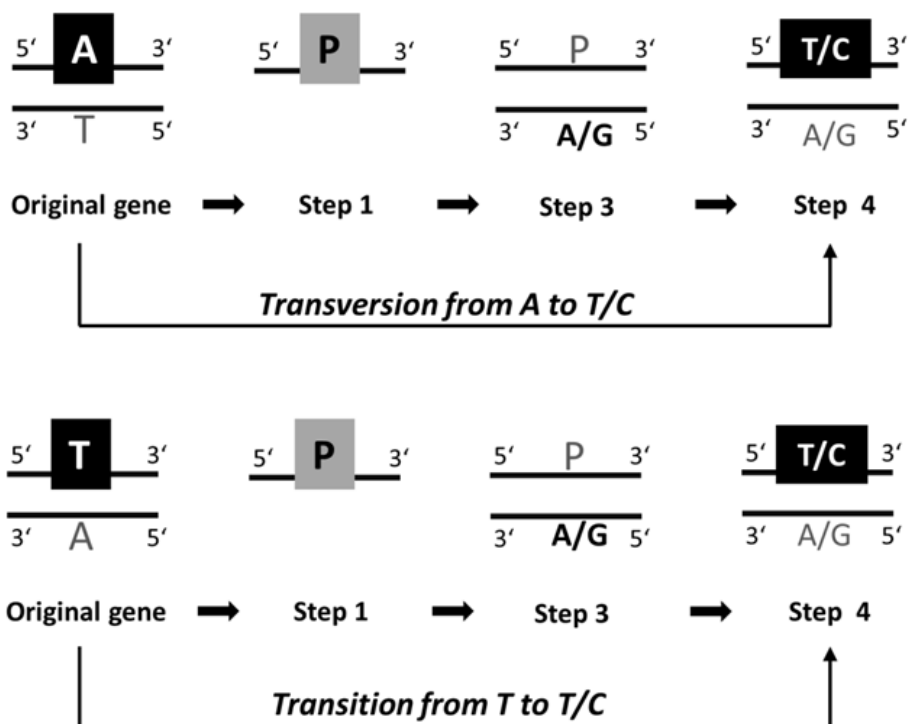
**Figure 6.** Illustration of the two preliminary experiments performed prior to SeSaM step 1 to 4. In “Preliminary Experiment I”, linear DNA templates used step 1 and step 3 of SeSaM are generated. These templates harbor artificial flanking sequences that come into action in SeSaM step 1 and SeSaM step 4. In “Preliminary experiment II”, the optimal concentration of phosphorothioate nucleotides to use in step 1 is determined. This ensures that fragments with uniform size distribution spanning the entire gene sequence are generated in the step 1 of SeSaM. (adapted from Ruff et al., 2014)



**Figure 7.** Schematic of the SeSaM method comprising four steps. SeSaM step 1: Generation of ssDNA fragment pool with random length distribution; SeSaM step 2: Enzymatic elongation of single stranded DNA “tailed” with degenerate base; SeSaM step 3: Synthesis of the full length gene; SeSaM step 4: Replacement of degenerate base (adapted from Ruff et al., 2014)

A SeSaM library can be completed in two to three days. Despite the increased workload in comparison to epPCR methods, SeSaM is a relatively economical and straightforward to perform. Notably, through SeSaM three of the four limitations of most random mutagenesis methods are eliminated, i.e. each codon in a given gene is targeted with equal probability, independently of polymerase bias and with a high probability of consecutive nucleotide exchanges. In SeSaM libraries, up to 30% of all substitutions are consecutive, which leads to expansion of the obtainable amino acid substitution pattern of epPCR (Mundhada et al., 2011a). The significance of SeSaM is proven by several success stories (Patents: WO2012119955A1; WO2012028709A3; WO2012017008A1) and reports on evolved enzymes such as proteases (Li et al., 2012; Martinez et al., 2012) and a phytase (Shivange et al., 2012).

The mutational bias introduced by SeSaM is typically governed by the choice of nucleotide analog employed in step 2. Several degenerate bases (Mundhada et al., 2011a; Ruff et al., 2012b; Wong et al., 2008b) have been used to control the ratio of transition to transversion mutations introduced in SeSaM libraries, with dPTP (6-(2-deoxy- $\beta$ -D-ribofuranosyl)-3,4-dihydro-8H-pyrimido-[4,5-C][1,2]oxazin-7-one) being the most frequently used one. dPTP is a pyrimidine base analog used to target adenine (A) and guanine (G) in DNA strands. The substitution of an A- or G-site with dPTP ultimately results in a transversion nucleotide exchange to thymine (T) or cytosine (C) (Fig. 8).



**Figure 8.** The obtainable mutational spectrum when dPTP is introduced at an A- (top) or a T- (bottom) site. In SeSaM-Tv-classic, dPTP is used to substitute A (or G, not shown in the figure) in order to obtain transversion-enriched libraries. This nucleotide analog is not used at T-sites (or C-sites) as this would result in either nucleotide preservation or a transition that generates conserved amino acid substitution pattern and could be achieved by epPCR. (Ruff et al., 2012b). (taken from Ruff et al., 2014)

More recently, dRTP (1-β-D-ribofuranosyl-1,2,4-triazole-3-carboxamide), a purine analog, has been successfully employed in SeSaM to target T- and C-sites in DNA sequences and render exchanges to A and G (Ruff et al., 2012b). Thus, all four nucleotides in a DNA strand can be methodically exchanged by random mutagenesis. Coding and non-coding DNA strands can be independently targeted in step 1 and 2 to generate forward and reverse libraries. These can be used separately in step 3 or recombined in order to increase the mutational frequency of SeSaM libraries. In step 3, it is also crucial to utilize a DNA polymerase capable of elongation of mismatching primer-template DNA. Polymerases such as Vent (exo-), 3D1 (a chimera of three genes from the genus *Thermus*) have been used in this step. 3D1 polymerase, for instance, achieves up to 30% subsequent mutations when dPTP is used in SeSaM and can even “read” through a triple consecutive mismatch in *in vitro* assays (Agresti et al., 2010) and, in rare occasions, under application conditions (Mundhada et al., 2011a). In the last step of SeSaM, *Taq* polymerase is used in a nested PCR to

specifically amplify mutated DNA as well as to exchange the previously introduced degenerate base analogs with standard nucleotides. The PCR products of step 4 can be directly cloned in an expression vector by restriction cloning, although ligase-independent cloning methods such as PLICing (Blanusa et al., 2010a) are recommended in order to achieve higher cloning efficiency.

#### **1.4.2. Advancements of SeSaM in retrospect**

The development of the SeSaM protocol to its present status has been a decade-long, cumulative effort of a number of researchers. In the first publication of the series (Wong et al., 2004), the overall mutagenic strategy comprising 4 steps (Step 1: Generation of a pool of DNA fragments, Step 2: Enzymatic elongation of DNA fragments with degenerate bases, Step 3: Full-length gene synthesis, and Step 4: Degenerate base replacement) was postulated. Validation by preparation of a model mutant library in which adenines were targeted and the base analog deoxyinosine (dITP, which pairs with cytosine) provided the first proof-of-concept and gave hints for its remaining unrealized potential. This first study proved that SeSaM is a mutagenesis method that is completely independent of the mutational bias of DNA polymerases, and that it is capable of saturating or randomizing every targeted nucleotide position in a sequence. The authors reported that 49.0% of all mutations appeared at adenine positions (as anticipated) and, interestingly, 24.5% of the mutations occur at one or two nucleotides downstream of an adenine position. This observation implied that an additional feature of SeSaM could be, after further optimizations, the generation of subsequent nucleotide substitutions. This property had not been reported for any other random mutagenesis method. This interesting feature was addressed in a later study; however expansion of the mutational bias of SeSaM through employment of other degenerate or degenerate bases with different base-pairing preferences was pursued next. In 2008, the first transversion (Tv) enhanced random mutagenesis protocol (SeSaM-Tv) was reported (Wong et al., 2008b). Three degenerate nucleotides (dPTP, dKTP and dITP) were investigated in the context of SeSaM for preparation of mutant libraries with increased fraction of transversion nucleotide substitutions. While all three nucleotides were proven to be accepted by terminal transferase (SeSaM step 2) and confer the expected “read” (SeSaM step 3) and “write” (SeSaM step 4) properties, dPTP (6-(2-deoxy- $\beta$ -D-ribofuranosyl)-3,4-dihydro-8H-pyrimido-[4,5-C][1,2]oxazin-7-one) was finally selected by the authors to

illustrate the propensity for SeSaM to generate transversion mutations. As anticipated, in the model SeSaM library in which Gs were specifically targeted with dPTP, the number of transversion mutations was enriched whereas G→T accounted for 16.22 - 22.58% and G→C for 6.38 - 9.69% of the observed mutations and even contained a fraction of subsequent nucleotide substitutions. The degree to which mutations unobtainable by other epPCR methods were generated was largely dependent on the combination of DNA polymerases used in the last two steps of SeSaM.

SeSaM-Tv, a later version of the SeSaM protocol, rendered improvements especially in terms of frequency of consecutive nucleotide substitutions but also with regard to achieving a homogeneous fragment distribution pattern and improving the incorporation of multiple degenerate bases in SeSaM step 2 (Mundhada et al., 2011a). The increase of subsequent mutations was achieved by employing 3D1 polymerase in SeSaM step 3. The use of 3D1 polymerase increased the overall fraction of consecutive mutations a factor of 1.4 and made possible for the first time the generation of two consecutive transversion nucleotide substitutions. On the protein level, this translated to 40% increase of hardly obtainable or unobtainable amino acid substitutions compared to epPCR methods. In quantitative terms the beneficial effect on diversity of transversion bias alone was ~2-fold, i.e. a transition biased library lead to half the diversity attainable by a transversions biased library (Wong et al., 2007b). The latter result indicated that a further increase in the fraction of consecutive (double and even triple) mutations is a highly desirable feature in SeSaM. Even though, in the SeSaM-Tv protocol (Mundhada et al., 2011a), consecutive mutations were enriched and even subsequent consecutive transversion mutations were achieved for the first time, the frequency of these remained suboptimal. This was illustrated by the fraction of consecutive transversion mutations that accounted for a meager 4.6% of all clones.

The evolution of SeSaM from a mutagenesis method that eliminates the drawbacks associated with the polymerase bias to a technique capable of generating mutant libraries with tunable mutational bias and containing consecutive nucleotide substitutions was instrumental as both of these features ultimately resulted in previously unexplored protein diversity and opened new pathways to climb up the “fitness landscape” of protein optimization and adaptation. In particular, the effect of consecutive mutations and transversions, even though often overlooked by protein engineers, played a dominant role

for the increased diversity on protein level found in mutant libraries prepared by SeSaM. However, the low frequency of consecutive transversions which even though conferred a spectrum unattainable by conventional random mutagenesis also reinforced the significance of DNA polymerases for the successful outcome of SeSaM. Especially in SeSaM step 3 the employed polymerase played the leading role in determining the frequency of consecutive nucleotide substitution as well as the type of substitution obtained.

### 1.4.3. The remaining challenge: SeSaM step 3

SeSaM is a mutagenesis method that relies on the promiscuous base pairing ability of degenerate nucleotides to generate mutations. The nucleotide analogs are added to the 3'-end of single stranded truncated gene copies (SeSaM step 1 and 2) which are subsequently elongated to full length (SeSaM step 3). In the elongations step, the 'tailed' ends of ssDNA form mismatches with the supplied template. The magnitude of mismatch increases in relation to the number of consecutive degenerate bases and type of mismatch formed. The transversion mismatch is most difficult to elongate by DNA polymerases (Huang et al., 1992) but notably offers a large number of diverse substitutions (Wong et al., 2007b); therefore, transversion mismatch elongation is highly desirable in SeSaM mutant libraries. The only feasible solution to further increase the frequency of occurrence of consecutive transversion mutations in SeSaM libraries is to employ a polymerase which can elongate at least two consecutive transversion mismatches in SeSaM step 3.

The employed polymerase in the aforementioned step of SeSaM has to match a number of criteria in order to generate mutant libraries enriched in transversion mutations as well as consecutive nucleotide substitutions. In the first place, the employed DNA polymerase has to be able to extend (or "read" through) primers with a non-canonical nucleotides at the 3'-termini that mispair with the supplied DNA template. Secondly, the polymerase should not possess a strong 5'→3' exonuclease activity so that the degenerate bases at 3'-ends of primers remain intact. At the same time, sufficient fidelity in order to avoid the introduction of secondary mutation is desired. Lastly, the elongation of multiple mismatches at the 3'-end of primers is also highly desired in order to be able to target more than one nucleotide in a codon and consequently increase the diversity of SeSaM libraries.



Consequently, the remaining challenge in SeSaM in order for the method to generate libraries with close to theoretically attainable diversity is to improve the efficiency of subsequent mismatch elongation in SeSaM step 3. In terms of strategies to achieve this, the only plausible solution is the identification or engineering of a DNA polymerase. The latter has been a main objective in this doctoral work. For the identification of suitable polymerase candidates, our efforts were focused on a special family of polymerases involved in the *in vivo* elongation of non-canonically primed DNA amplification and translesion DNA synthesis.

## **1.5. DNA polymerase-based technologies and high-throughput screening systems for DNA polymerases**

### **1.5.1. DNA polymerase-based technologies**

DNA polymerases have been routinely employed in asserted methodologies such as *in vitro* DNA replication and molecular cloning, DNA sequencing, mutagenesis, and diagnostics (Blanusa et al., 2010b; Dennig et al., 2011; Erlich, 2013; Shendure and Ji, 2008; Thelwell et al., 2000; Wong et al., 2004). However, from a practical perspective, naturally occurring polymerases often suffer from a number of drawbacks that limit their biotechnological applications. Depending on the envisioned application, these can include insufficient processivity and fidelity, inhibition by sample impurities and inability to efficiently utilize non-canonical nucleotides.

### **1.5.2. Screening systems for DNA polymerases**

Platforms for polynucleotide polymerase engineering comprise selection and screening systems. One example of the former type is the compartmentalized self-replication (CSR) method which allows the enrichment of improved polymerase mutants among more than  $10^8$  individual variants in a single run (Ghadessy et al., 2001). Other selection systems with similar throughput are the platform for functional complementation of polymerase activity in polymerase deficient strains (Camps and Loeb, 2007) and the duet plasmid-based selection system (Brakmann and Grzeszik, 2001). The aforementioned strategies require that polymerases possess sufficiently high processivity, fidelity and/or thermostability to be able to replicate long fragments of DNA. These selection systems are, therefore, not suited for distributive polymerases, e.g., members of the Y-polymerase family. Alternative detection

systems for polymerase activity rely on the continuous detection of pyrophosphate release during polymerisation of nucleoside triphosphates. The produced diphosphate is quantified through a coupled enzymatic cascade (employing ATP-sulfurylase and firefly luciferase) (Nyrén, 1987), through pyrophosphate sensors (Fabbrizzi et al., 2002) or changes in current during dNTPs polymerisation (Credo et al., 2012). None of these activity assays is designed for use in combination with crude cell extracts of *E. coli*, thus not easily applicable in directed evolution of polymerases.

Quantitative polymerase chain reaction (qPCR) is used for qualitative and quantitative analysis of nucleic acids (Orlando et al., 1998). One type of reporters in qPCR are fluorescent DNA dyes that become fluorescent upon intercalation with double stranded DNA. An alternative approach for detection employs sequence-specific DNA probes of various „molecular architecture“ that become fluorescent after hybridization to or release from a target DNA sequence. One example of the latter kind, the Scorpion probe, comprise a primer with a covalently attached fluorophore. Unelongated Scorpion probes are not fluorescent due to hybridization with a separate quenching oligo that is supplied in the reaction mix. Elongated Scorpion probes contain the labeled probe element and target sequence on the same strand and hybridize preferentially with themselves and not the quenching oligo, thus an increase of fluorescence signal is generated. The kinetics of such an intramolecular rearrangement is very rapid and ensures reliability of probing as well as high signal-to-noise ratio (Thelwell et al., 2000; Whitcombe et al., 1999). The stoichiometry of the reaction (one amplicon leads to one fluorescent signal) and concentration independency of probing provides reliable quantification and enhanced overall sensitivity (Carters et al., 2008). Additionally, unlike most qPCR reporters (e.g., TaqMan probes), Scorpion probes rely solely on polymerases' primary enzymatic activity (Heid et al., 1996; Huggett et al., 2005). Thus, Scorpion probes can be used in combination with polymerases that lack secondary exonuclease or displacement activities. The adaptation of a Scorpion probe for screening of polymerase mutant libraries is described in the “Results and Discussion” section of this chapter.

## 1.6. DNA polymerases in diversity generation for directed evolution

In the context of directed evolution, polymerases which are naturally optimized for error-free DNA replication *in vivo* are often limited in their ability to generate diverse mutant libraries due to their high fidelity and specificity for natural nucleotides. In contrast, DNA polymerases involved in DNA repair which are found in all polymerase families, including the most recently reported Y-polymerase family (Ohmori et al., 2001), are generally characterized by low fidelity, moderate-to-low processivity as well as augmented ability to utilize a range of non-canonical primer-template DNA (Chandani et al., 2010; McDonald et al., 2006; Sale et al., 2012; Yang and Woodgate, 2007). Polymerases from the Y-family have not been widely employed in random mutagenesis methods despite their promising potential.

## 1.7. A concise overview of Y-family of DNA polymerases

Polymerase-mediated replication and repair of DNA are of fundamental importance for the survival of all living organisms. Based on sequence homology, DNA polymerases are classified in 7 families, A-, B-, C-, D-, X-, Y- and RT-family (Filée et al., 2002). Replicative polymerases are found all families while those involved in DNA repair mechanisms belong to all but the C- and D-family. The Y-polymerase family (Ohmori et al., 2001) is the most recently identified class which encompasses polymerases from prokaryotes and eukaryotes alike which are primarily involved in translesion DNA synthesis (TLS). In the context of this thesis, Y-polymerases have been considered a good source of candidates to realize the objectives related to increasing the efficiency of mismatch elongation, a feature that is highly desirable for the generation of transversion-enriched SeSaM libraries. Their intrinsic property to 'read through' damaged DNA indicates that some Y-family members may already be able to fulfil the requirements for implementation in SeSaM or at least provide a good starting point for optimization by protein engineering. In addition, a good deal of information on the structure-function relationship for some members has been generated through crystallography and mutational studies. This section gives a concise overview of some structural and mechanistic features of Y-polymerases.

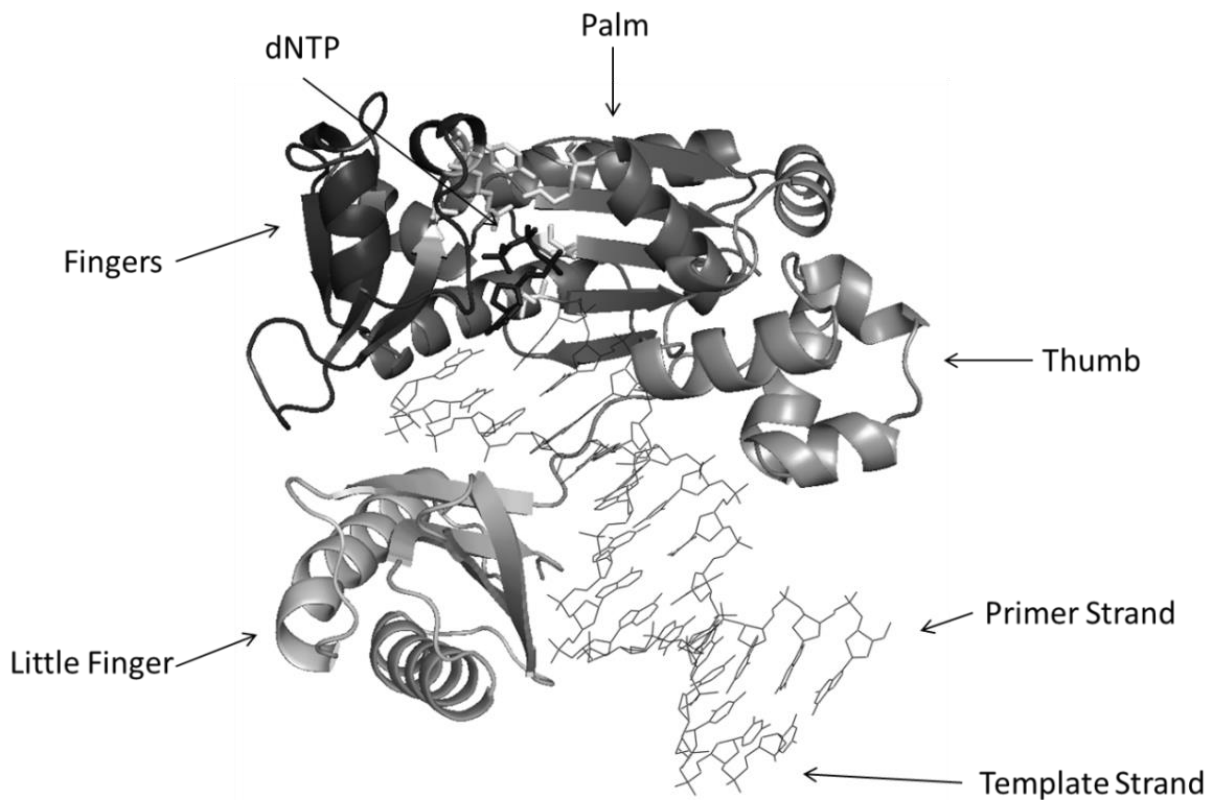
### 1.6.1. Y-family polymerases

DNA lesions caused by endogenous or exogenous agents represent a common thread to the viability of all living cells. Prevention and repair mechanisms exist to revert damage (Friedberg, 2003; Friedberg et al., 1995; Kelman and White, 2005; Takata et al., 1998), still a portion of damaged DNA can escape this process and stall the DNA replication machinery. To tackle this challenge, cells have developed the ability to replicate through lesions. Translesion DNA synthesis (TLS) is catalyzed by specialized DNA polymerases referred to as Y-family polymerases (Ohmori et al., 2001). Y-family members are characterized by high error rates (100- to 1000-fold decrease in accuracy compared to polymerases involved in replication of DNA) as well as low processivity (McCulloch and Kunkel, 2008). Different Y-family members also exhibit different lesion bypass capabilities as well as different mutational spectra (Johnson, 2010; Nelson et al., 1996). These enzymes are universally conserved in all living organisms but the highest number of crystal structures are those of archaeal DinB homologues – Dbh and Dpo4 (Zhou et al., 2001a); however, yeast Pol $\eta$  (Trincao et al., 2001), human Pol $\iota$ , Pol $\kappa$  and Rev1 (Nair et al., 2004; Swan et al., 2009; Uljon et al., 2004) structures have also been solved.

### 1.6.2. Structural features of Y-family polymerases

Y-family polymerases share with other polymerase families a conserved right-handed polymerase core of a thumb, palm and finger domain (Fig. 9). Additionally, a specific 'wrist'/little finger (LF), sometimes also termed polymerase associated domain (PAD) domain (Goodman, 2002; Ling et al., 2001; Silvan et al., 2001) is found only in members of this family and characterizes the structural appearance of Y-polymerases. By means of multiple sequence alignments, at least two specific domains connected by a relatively unstructured region can be identified (Pata, 2010). The N-terminal region contains the catalytic core (thumb, palm, finger domain) and is functionally well preserved among polymerase families. Unlike replicative polymerases, Y-polymerases have a rather small finger and thumb domain which confers an open and solvent accessible active site. The C-terminal region contains the unique LF domain and may contain additional features for localization and interaction with other proteins (Yang and Woodgate, 2007). The LF has a conserved tertiary structure of four beta sheets and two alpha helices. However, there are no primary sequence motifs that are

similar among members of the Y-family. The LF appears to be the most mobile region in Y-family members (Eoff et al., 2007). The LF domain determines the catalytic efficiency and mutation spectra of each polymerase by influencing enzyme–substrate interactions. This is suggested by the observation that swapping the LF domains between two archaeal DinB homologs, Dbh and Dpo4, which share 50% sequence identity, confers drastically different catalytic efficiencies and mutation spectra to these polymerases (Boudsocq et al., 2004).



**Figure 9.** Structural features of the archetypal Y-family polymerase Dpo4 from *Sulfolobus solfataricus*. Ribbon diagram of a ternary complex of Dpo4 with the different structural features are highlighted in shades of gray. The important residues in the active site of the protein and discussed in section 3 of the chapter are shown as sticks in light grey, the incoming dNTP is shown in black sticks and the template and primer DNA stand are presented in light grey sticks.

### 1.6.3. Catalytic Mechanism

Similar to polymerases from other families, Y-family members show a well-conserved two-metal ion catalytic mechanism (Trincao et al., 2001; Uljon et al., 2004; Zhou et al., 2001b). Two  $Mg^{2+}$  ions (coordinated by carboxylates of two conserved aspartate residues (D7 and D105 in case of the archetypal Dpo4 from *Sulfolobus solfataricus*) stabilize the resulting penta-coordinated transition state. One metal ion activates the primer's 3'-OH for attack on the  $\alpha$ -phosphate of the dNTP. The other metal ion plays a role of stabilizing the negative charge that builds up on the leaving oxygen and chelates the  $\beta$  and  $\gamma$  phosphates. Nucleophilic attack of 3'-OH group results in phosphodiester bond formation. In addition to participating in coordination with one of the metal ions, the phosphates of the incoming nucleotide are bound by conserved residues in the finger (Y48 and R51 in Dpo4) and the palm (K159 in Dpo4) domains. The sugar moiety of the incoming nucleotide is positioned by stacking on top of a "steric gate" residue (Y12 in Dpo4; Y or F in other Y-family polymerases) and by the 3'-OH's formation of a hydrogen bond with the backbone nitrogen of the same residue (Pata, 2010). The steric gate residue selects against the incorporation of ribonucleotides by the polymerase (DeLucia et al., 2003). Other interesting features in the vicinity of the active site are the "roof-amino acid" (A45 in Dpo4), which is the amino acid that lies above the nucleobase of the deoxynucleotide triphosphate (dNTP) and plays a role in dNTP insertion efficiency, and a cluster of three amino acids, including the roof-amino acid, which anchors the base of a loop, whose detailed structure dictates important mechanistic function, i.e., controls of the size of a 'chimney' (opening in the protein surface) which in turn can favor the insertion of one type of base over the other (Seo et al., 2009).

### 1.6.4. Implications of Y-polymerase structure on DNA binding and catalysis

The implications on the catalytic performance of Y-polymerases' open and solvent exposed active site, especially in comparison to, e.g., Taq polymerase, is illustrated through the example of the well-characterized Y-family polymerase, Dpo4 from *Sulfolobus solfataricus*. In Dpo4, the interaction between protein and substrates is mainly through the DNA backbone of template and primer, and the sugar-phosphate moiety of an incoming deoxynucleotide (Pata, 2010). There is minimal contact between the enzyme's active site and the replicating base pair or preceding DNA duplex in the major or minor groove. This underlies the less

stringent distinction of a perfect (Watson-Crick type) or mismatching base pairing in the active site. In the case of replicative polymerases where non-Watson-Crick base pairing is usually a disallowed conformation, base pairing always provides a flat and smooth minor groove with distinct hydrogen bond patterns. The minor groove of mismatched base pairs is uneven and presents different patterns of hydrogen-bond donors and acceptors. This effectively prevents further catalysis in case of replicative polymerases. The lack of a complementary interface between enzyme and replicating base pair in Dpo4 and Y-family polymerases in general provides a foundation for the high-error-rate and low-fidelity DNA synthesis (Wong et al., 2008a).

The high fidelity of replicative DNA polymerases also depends on an 'induced-fit' - a conformational change to discriminate against a wrong incoming nucleotide. A correct incoming deoxynucleoside triphosphate makes a Watson-Crick base pair with the opposite base which induces structural rearrangement of the finger domain and secludes the replicating base pair in a closed active site (Doubl   et al., 1999). An incorrect incoming nucleotide or damaged template base hinders this conformational change and reduces the rate of polymerization. In contrast to the "induced-fit" mechanism of replicative polymerases, the active site of Dpo4 is pre-formed regardless of whether an incoming nucleotide is incorrect or the template base is damaged or even absent. Dpo4 is always ready to catalyse the nucleotidyl transfer reaction. A preformed active site is a common feature among the Y-family polymerases (Yang, 2005).

## 2. Materials and methods

### 2.1. Chemicals

All chemicals were of analytical grade or higher quality and purchased from Sigma-Aldrich Chemie (Steinheim, Germany), AppliChem (Darmstadt, Germany) and Carl Roth (Karlsruhe, Germany) unless specified. DNA probes and other oligonucleotides in salt-free, lyophilized form were purchased from Biomer.net (Ulm, Germany) and Eurofins MWG Operon (Ebersberg, Germany). Degenerate base dPTP was provided by Biolog Life Science Institute (Bremen, Germany). The primer sequences used in the project are shown in Table 2. Commercial enzymes were obtained from New England Biolabs (New England Biolabs, Frankfurt am Main, Germany).

### 2.2. Bacterial strains, plasmids and list of primers

**Table 1.** List of bacterial strains used in this study.

Bacterial strain	Genotype	Comments
<i>E. coli</i> DH5 $\alpha$	[(supE44 $\Delta$ lacU169 $\Phi$ 80 lacZ $\Delta$ M15) hsdR17 recA1 gyrA96 thi-1 relA1]]	Life Technonologies, Darmstadt, Germany
<i>E. coli</i> Top 10	[F- mcrA $\Delta$ (mrr-hsdRMS-mcrBC) $\phi$ 80lacZ $\Delta$ M15 $\Delta$ lacX74 nupG recA1 araD139 $\Delta$ (ara-leu)7697 galE15 galK16 rpsL(Str <sup>R</sup> ) endA1 $\lambda$ ]	Life Technonologies, Darmstadt, Germany
<i>E. coli</i> BL21 (DE3)	[F- ompT gal dcm lon hsdSB (rB- mB-) $\lambda$ (DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])]	Novagen/ Merck KGaA, Darmstadt Germany
<i>E. coli</i> BL21 (DE3) lacI <sup>q1</sup>	[F- ompT gal dcm lon hsdSB (rB- mB-) $\lambda$ (DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])]	(Blanusa et al., 2010b)

**Table 2.** List of plasmids used in this study.

Plasmid name	Features	Comments
pET28a(+)	T7 promoter, His•Tag, lacI, pBR322, Kan, f1 origin	Merck KGaA, Darmstadt Germany
pALXTREME-1a	T7 promoter, His•Tag, pBR322, Kan, f1 origin	(Blanusa et al., 2010b)
pALXTREME-5b	T7 promoter, pBR322, Amp, f1 origin	(Blanusa et al., 2010b)
pBAD-N	Ara promoter, pBR322, Amp, AraC, f1 origin	(Guzman et al., 1995)



**Table 3.** List of primers, DNA probes and other oligos used in the study, and primers used in CSR. (SNN = randomization codon; S = cytosine or thymine; N = cytosine or thymine or adenine or guanine; [FITC] = fluorescein isothiocyanate, fluorescent dye; [FAM] = 6-carboxy fluorescein, fluorescent dye; ~~~~ = Spacer18, hexaethylene glycol; [BHQ] = black hole quencher-3, quencher molecule; [Biotin] = 5`biotin; ctaagaaagcccg = phosphothioated oligo; P = degenerate pyrimidine analog, dPTP; codes for A or G in ratio 1.6:1; Y = cytosine or thymine)

	PRIMER NAME	SEQUENCE (5`-->3`)	APPLICATION
P1	SSM_Fwd	CGGTTCTGGCCTTTTGCTGG	SSM library generation
P2	SSM_RV_Y10	TCAACCTGCGCGTAGAASNNATCAAGTCTACGAACAGC	SSM library generation
P3	SSM_RV_V32	CGAAACGACCGGAGAAASNNGCACACAACAACTGGTTT	SSM library generation
P4	SSM_RV_A34	GCTGTCTTCGAAACGACCSNNGAAAACGCACACAAC	SSM library generation
P5	SSM_RV_A44	TGCTTCGTAGTTAGCGGTSNNAACCGCACCGCTGT	SSM library generation
P6	SSM_RV_V62	GGCAGGATTTTCTTAGCTCSNNGATCGGGATACCAGCTT	SSM library generation
P7	SSM_RV_M76	GTTGTTAAACTTCTTTGCGSNNAGGACGGTAGACGGC	SSM library generation
P8	SSM_RV_K78	GCTCACTTGGTAAACTCSNNGCGCATAGGCAGGTAG	SSM library generation
P9	SSM_RV_D156	CCGTTCCGGTTTAGCCATSNNCGCCGCGATCTTC	SSM library generation
P10	SSM_RV_A181	CCAATGCCTGGAACGTCSNNAATATCCAGTTCGCGG	SSM library generation
P11	SSM_RV_P184	GTGATGTTGCCAATGCCSNNAACGTCAGCAATATCCAG	SSM library generation
P12	SSM_RV_N188	TCAGTTTTTCAGCGGTGATSNNGCCAATGCCTGGAAC	SSM library generation
P13	SSM_RV_A191	CCCAGTTTCTTCAGTTTTTCSNNGTGATGTTGCCAATGC	SSM library generation
P14	SSM_RV_R242	TGCGACCGATGCTCTTSNNCACACGCGTGCGAA	SSM library generation
P15	SSM_RV_R247	ATTACGTTTCATGGTCACAATSNNACCGATGCTCTTGCG	SSM library generation
P16	SSM_RV_K275	GGGATGCGTTTATCCAGSNNATAGTAAGATTCTTCAATAGCG	SSM library generation
P17	SSM_RV_V289	GATGTCCAGATCCTCCGTSNNAGCAACCACGTGGATAG	SSM library generation
P18	SSM_RV_L293	CCGGGACACGATGTCSNNATCCTCCGTAACAGCAAC	SSM library generation
P19	SSM_RV_R331	CACGCCGATACGSNNGATCTTGCGCTC	SSM library generation
P20	SSM_RV_R332	GCGCACGCCGATSNNCGGATCTTGCG	SSM library generation
P21	SSM_RV_R336	GCCTCGATGAATTTAGAGAASNNCACGCCGATACGGC	SSM library generation
P22	6-FAM	[FAM]CGCTGAGCAATAACTAGCATAACCGAC~~~~GTGTGATGGC GTGAGGCAGC	Scorpion probe primer

CHAPTER I. Engineering of DNA Polymerases for Application in Directed Protein Evolution  
Materials and Methods

P23	BHQ-1	GTCGGTTATGCTAGTTATTGCTCAGCG[BHQ1]	Scorpion probe quencher
P24	Scorpion template match	CGCTGAGCAATAACTAGCATAACCGACGATAAATCACAATAGGTT CCTTGTAG <b>G</b> CTGCCTCACGCCATCACAC	Scorpion probe template match
P25	Scorpion template 1 mismatch	CGCTGAGCAATAACTAGCATAACCGACGATAAATCACAATAGGTT CCTTGTAG <b>C</b> CTGCCTCACGCCATCACAC	Scorpion probe template 1 mismatch
P26	Scorpion template 2 mismatches	CGCTGAGCAATAACTAGCATAACCGACGATAAATCACAATAGGTT CCTTGTAG <b>CG</b> TCCTCACGCCATCACAC	Scorpion probe template 2 mismatches
P27	F1	CGACTCACTATAGGGGAATTGTGAGCGGA	SeSaM library generation
P28	R3	CGGGCTTTGTTAGCAGCCGGATCTCAG	SeSaM library generation
P29	SeSaMF1	CACACTACCGCACTCCGTCGCGACTCACTATAGGGGAATTGTGAG CGGA	SeSaM library generation
P30	SeSaMR3	GTGTGATGGCGTGAGGCAGCCGGCTTTGTTAGCAGCCGGATCTC AG	SeSaM library generation
P31	F1_up	CGCCTGTCACCGACTCACTATAGGGGAATTGTGAGCGGA	SeSaM library generation
P32	R3_dn	GCGGACAGTGCGGGCTTTGTTAGCAGCCGGATCTCAG	SeSaM library generation
P33	Bio_Fw	[Biotin]CACACTACCGCACTCCGTCG	SeSaM library generation
P34	Bio_Rv	[Biotin]GTGTGATGGCGTGAGGCAGC	SeSaM library generation
P35	SeSaM_Fw	CACACTACCGCACTCCGTCG	SeSaM library generation
P36	SeSaM_Rv	GTGTGATGGCGTGAGGCAGC	SeSaM library generation
P37	Vector_PTO_Fw	<u>ctaacaagcccg</u> AAAGGAAGCTGAGTTG	Cloning of SeSaM libraries
P38	Vector_PTO_Rv	<u>attgtgtcagc</u> GATATCCATGGCCATC	Cloning of SeSaM libraries
P39	Insert_PTO_Fw	<u>gctgaacacaat</u> CCAGTCGTTATGGTTC	Cloning of SeSaM libraries

P40	Insert_PTO_Rv	<u>cgggctttgtag</u> CAGCCGGATCTCAG	Cloning of SeSaM libraries
P41	F1_short	CGACTCACTATAGGGGAATTGTGAGC	Compartmentalized Self-replication
P42	R3_short	CGGGCTTTGTTAGCAGCCGGATCTC	Compartmentalized Self-replication
P43	F1_PP	CGACTCACTATAGGGGAATTGTGAGCG*P*P	Compartmentalized Self-replication
P44	R3_PP	CGGGCTTTGTTAGCAGCCGGATCTC*P*P	Compartmentalized Self-replication
P45	F1_YY	CGACTCACTATAGGGGAATTGTGAGCGYY	Compartmentalized Self-replication
P46	R3_YY	CGGGCTTTGTTAGCAGCCGGATCTCY	Compartmentalized Self-replication
P47	FITC-PP	[FITC]-GCAATACGCAAACAGTCTCTCTTPP	Forward primer elongation assay
P48	FITC_template match	GCAATACGCAAACAGTCTCTCTTGGGCCCG	Forward primer elongation assay
P49	FITC_template 1_mismatch	GCAATACGCAAACAGTCTCTCTTGC GCCCG	Forward primer elongation assay
P50	FITC_template 2_mism.	GCAATACGCAAACAGTCTCTCTTTGCCCG	Forward primer elongation assay
P51	FITC_match complementary	CGGGGCCCAAGAGAGACTGTTTGC GTATTGC	Forward primer elongation assay
P52	FITC_1_mism. complementary	CGGGGCGCAAGAGAGACTGTTTGC GTATTGC	Forward primer elongation assay
P53	FITC_2_mism. complementary	CGGGGCGAAAGAGAGACTGTTTGC GTATTGC	Forward primer elongation assay

## 2.3. Cloning procedures

The genes which encode putative Y-family polymerases have been codon optimized for expression in *E. coli* and ordered as synthetic constructs (Mr. Gene Geneart, Darmstadt, Germany). The synthetic DNA was cloned in pET28a(+) (EMD Bioscience, Darmstadt, Germany), pALXtreme-1a (Blanusa et al., 2010b) in frame with N-terminal poly His-tag and in pALXtreme-5b (Blanusa et al., 2010b) and pBADN (Life Technologies, Darmstadt, Germany) using standard cloning protocol and NdeI/XhoI restriction enzymes (New England Biolabs, Frankfurt am Main, Germany). The resulting constructs were checked for the presence of the correct insert by analytical restriction digestion and sequencing, and transformed into expression host *E. coli* BL21 Gold (DE3) lacI<sup>q1</sup> (Blanusa et al., 2010b).

## 2.4. Pilot expression and purification of putative Y-polymerase genes

### 2.4.1. Pilot expression studies

Flask expression in 500 mL Erlenmeyer flasks was carried out in 50 mL TB<sub>Kan</sub>. The main culture was inoculated with 1% (v/v) overnight pre-culture of *E. coli* BL21 Gold (DE3) lacI<sup>q1</sup> pET28a(+) or pALXtreme-1a harboring the gene of interests, prepared in LB<sub>Kan</sub> and cultivated at 37 °C, 250 rpm for 12 -14 h. . Induction of Dpo4, Mse and Tps was carried out at OD<sub>600</sub> of 0.6 with 0.1 mM IPTG and cell cultures were incubated for 12 - 14 hours at 37 °C. Induction of Pto was achieved by addition of IPTG to a final concentration of 0.1 mM when OD<sub>600</sub> reached 0.6 and cells were grown for 4 hours at 37 °C. Expression cultures were harvested by centrifugation (4 °C, 3 220 × g, 20 min, Eppendorf 5810R centrifuge, Eppendorf AG, Hamburg, Germany), washed with phosphate-buffered saline (PBS, pH 7.4) and used immediately or stored at -20 °C.

### 2.4.2. Pilot purification studies

Cell pellets were thoroughly re-suspended in 1× LEW washing buffer (Macherey–Nagel, Düren, Germany) such that a 40% (w/v) cell suspension was prepared. Cells were lysed by sonication (10 × 1 min, 40% sonication amplitude) using Bandelin M73 sonotrode (or where specified, by B-PER lysis reagent (Thermo Fischer Scientific, Waltham, USA) according to manufacturer's instructions), followed by heat treatment (60 °C or 80 °C, 10 min) to denature

*E. coli* proteins and separation by centrifugation (4 °C, 16 000 × g, 20 min, Eppendorf 5804, Eppendorf, Hamburg, Germany). The crude cell extracts were further clarified by filtration through a low-protein-binding filter (0.45 µm; Minisart RC 25 disposable syringe filter; Sartorius, Hamburg, Germany) before being subjected to His-tag purification (pre-packed Protino® Ni-IDA columns; Macherey–Nagel, Düren, Germany). After extensive washing with 1× LEW buffer (5 × 5 mL), the His-tagged protein was eluted in 3 mL 1× Elution Buffer (Macherey–Nagel, Düren, Germany). Elution fractions of 500 µL were collected and the protein content of each was estimated using NanoDrop spectrophotometer (NanoDrop Technologies Inc., Wilmington, USA). Fractions containing more than 0.3 mg mL<sup>-1</sup> of protein were pooled and concentrated using Amicon centrifugal filters (Millipore Inc., Billerica, USA), if necessary. Samples were collected after each purification step for SDS-PAGE analysis. Purified proteins were dialysed using appropriate dialysis membrane (Spectra/Por® dialysis membrane 10 - 12 kDA, Spectrum Laboratories Inc, Rancho Dominguez, USA) to exchange the elution buffer which contained imidazole with buffer suited for storage of polymerases (10 mM Tris, 100 mM KCl, 1 mM DTT and 10 mM EDTA). On the following day, 50% glycerol was added to the purified fractions to prepare the proteins for long term storage at -20 °C.

## 2.5. Thermostability valuation in crude cell extract

Thermostability tests were performed with crude cell extracts of cell cultures expressing Y-polymerases. Cells were lysed by sonication (40 % (w/v) cell suspension, 10 x 1 min sonication and 40% sonication amplitude) and clarified cell extracts were subjected to heat treatment at temperatures ranging from 40 °C to 90 °C for 10 min in a thermo block (Eppendorf ThermoMixer® C, Eppendorf, Hamburg, Germany). Precipitated and soluble proteins were separated by centrifugation (4 °C, 16 000 × g, 20 min, Eppendorf 5804, Eppendorf, Hamburg, Germany) and separately loaded on SDS acrylamide gel for analysis.

## 2.6. *In vitro* expression of Dpo4

*In vitro* expression of Dpo4 was carried out for 1.5 hours from a linear DNA template using EasyXpress Protein Synthesis Kit (Qiagen, Hilden, Germany) according to the manufacturer's recommendations. Emulsification of *in vitro* expression mix was carried out as outlined in section 2.8.1.

## 2.7. Forward primer extension assay

A fluorescently-labelled (FITC-labeled) primer harboring two pyrimidine base analogues (dPTP) at its 5'-end was extended by Dpo4 polymerase on a short, single-stranded DNA template (Table 3; P47 - P50) in the presence of dNTPs and ThermoPol activity buffer (NEB, Frankfurt am Main, Germany). The primer extension reaction with Y-polymerases was carried out in ThermoPol buffer (New England Biolabs, Frankfurt am Main, Germany) using final concentrations of up to 550 nM protein, 2  $\mu$ M template, 1  $\mu$ M FITC-labelled primer and 0.2 mM dNTPs in a total volume of 25  $\mu$ L. Primer and template were annealed to each other by heating the reaction mix to 95 °C for 2 min, then slowly cooling it down to 4 °C prior to enzyme addition. Samples were incubated at 37 °C/ 55 °C or run using thermocycling conditions (e.g., 80 °C, 30 sec; 55 °C, 1 min) and collected at different time intervals/thermocycles and immediately treated with EDTA (20 mM, final) to terminate the polymerization reaction. Competitor DNA (Table 3; P51 - P53) was added in 100-fold excess and the samples were heated up to 95 °C for 5 min. In this way, elongated primer was detached from the template and reaction products alone could be separated on 25% native acrylamide gel and visualized (ex: 494 nm; em: 518 nm) on a FLA 3000 PhosphorImager (FujiFilm, Düsseldorf, Germany).

## 2.8. Compartmentalized self-replication

### 2.8.1 Oil Phase preparation

The following components were mixed together in the listed order to prepare the oil phase:

---

95.05% Light Mineral Oil (Cat. No. M5904, Sigma)  
4.5% Span 80 (Cat. No. 85548, Fluka)  
0.4% Tween 80 (Cat. No. P8074, Fluka)  
0.05% Triton X-100 (Cat. No. A4975, AppliChem)

---

Components were thoroughly mixed by stirring in 5 mL cryogenic tube (Cat. No. 60.58.001, Sarstedt, Nümbrecht, Germany). Tips with cut edges or specialized tips for viscous liquids were used for pipetting. The oil phase was stored in the dark at room temperature for up to 2 weeks.

### 2.8.2. Cell sample preparation

Expressed clones (in flasks) harboring polymerase genes and empty vector were harvested by centrifugation. The cell pellets were once washed with phosphate buffer saline (PBS) and re-suspended in PBS such that OD<sub>600</sub> of the suspension was 5. Cell suspensions were kept on ice until needed but always used it within 30 min after their preparation

### 2.8.3. Preparation of aqueous phase, emulsification and PCR reaction

The mastermix was prepared on ice according as outlined below. All components were thawed on ice prior to use.

	1x	4x
<b>10x ThermoPol buffer (NEB)</b>	5 µL	20 µL
<b>dNTPs (10 mM)</b>	1 µL	4 µL
<b>F1/F1PP (5 µM)</b>	4 µL/10 µL	16 µL/ 40 µL
<b>R3/R3PP (5 µM)</b>	4 µL/10 µL	16 µL/ 40 µL
<b>ddH2O</b>	35 µL/ 23 µL	128 µL/ 96 µL
<b>Cell suspension, OD600 = 5</b>	1 µL	4 µL
<b>&lt;Taq Pol (5U/µL)&gt;</b>	<0.25 µL>	<1 µL>
<b>Total Volume</b>	50 µL	200

All components were mixed on ice. Cell suspension was added at last just before emulsification *Taq* polymerase (1.25 U in 50 µL) was on some occasions added to the mastermix when F1PP and R3PP primers (Table 3; P43 - P44) were used. (NOTE: *Taq* polymerase cannot elongate the latter primers but it can take over the elongation once a mutant polymerase has overcome the mismatch. This can be particularly helpful when mutant libraries of low processive polymerases such as Y-family polymerases are screened.)

For emulsification, 150 µL aqueous phase was added drop-wise to 300 µL oil phase in 5 mL cryogenic vial (Corning, Kaiserslautern, Germany) under constant stirring (700 rpm, IKAMAG REO, IKA-Werke, Staufen, Germany) with magnetic stirring bar (8 x 3 mm with pivot ring) on ice. Stirring for 5 min after addition of the last drop of aqueous phase was done. The resulting emulsion appeared creamy white and viscous. The emulsion was distributed into

PCR tubes (50  $\mu\text{L}$ / tube) with tips with cut edges to avoid damaging of the formed compartments. Thermocycling program (25 cycles; volume: 25  $\mu\text{L}$ ; 105  $^{\circ}\text{C}$  heated lid) : 2 min 90  $^{\circ}\text{C}$  (1x); 30 sec 90  $^{\circ}\text{C}$ , 30 sec 51  $^{\circ}\text{C}$ , 3 min 65  $^{\circ}\text{C}$  (24x); Hold 8  $^{\circ}\text{C}$ ; annealing temperature for F1PP and R3PP is 51  $^{\circ}\text{C}$ ; or 60  $^{\circ}\text{C}$  if F1 and R3 primers were employed (Table 3; P27 - P28). After thermocycling the emulsion remained white and creamy and was overlaid by a thin layer of clear oil phase

#### **2.8.4. DNA recovery and re-amplification**

After thermocycling, all samples were pooled together in 1.5 mL tube and centrifuged for 2 min at 13 200 x g. Three distinct phases were formed after centrifugation. The clear oil phase on top was pipetted out and 1 mL butanol was added thoroughly mixed in by vortexing. The sample was centrifuged again for 3 min at 13 200 x g. A small pellet appeared at the bottom of the tube. The liquid is removed as much as possible by pipetting and the samples were further dried at 70  $^{\circ}\text{C}$  to remove any traces of butanol. The pellet was then re-suspended in 15 - 30  $\mu\text{L}$  ddH<sub>2</sub>O. Re-amplification of emulsion PCR products was usually necessary, especially when selection pressure was applied. Re-amplification PCR was carried out with F1 short and R3 short (Table 3; P41 and P42) and *Taq* polymerase after DpnI digestion of emulsion PCR products to remove methylate plasmid DNA. Alternatively, F1YY and R3YY primers (Table 3; P45 and P46) and *Taq* polymerase were used which in theory specifically bind to emulsion PCR products and form mismatching pairs with plasmid DNA.

### **2.9. PicoGreen<sup>®</sup> assay**

#### **2.9.1. Protein expression in 96-well microtiter plates (MTP)**

Master microtiter plates containing Dpo4 WT cloned in pALXtreme-5b vector and transformed in *E. coli* BL21 Gold (DE3) *lacI*<sup>q1</sup> were duplicated using a 96-well pin replicator (constructed in Werkstatt der Biologie, RWTH Aachen University, Aachen, Germany) to pre-culture flat bottom microtiter plates (Greiner Bio-One, Frickenhausen, Germany) containing 200  $\mu\text{L}$  LB<sub>Amp</sub> per well. The latter were cultivated for 12 h in a microtiter plate shaker (Multitron II, Infors GmbH, Einsbach, Germany; 37  $^{\circ}\text{C}$ , 900 rpm, 70% relative humidity) before being replicated to V-bottom microtiter plates (transparent polystyrene plate, Corning, Kaiserslautern, Germany) with 150  $\mu\text{L}$  TYM-5052<sub>Amp</sub> auto-induction media per well



(Studier, 2005). Expression was carried out at 37 °C for 12 h in a microtiter plate shaker (Multitron II, Infors, Einsbach, Germany; 900 rpm, 70% relative humidity), the expression cultures were harvested by centrifugation (4 °C, 3 220 × g, 20 min) using Eppendorf 5810R centrifuge (Eppendorf, Hamburg, Germany) stored at -20 °C for up to one week. ). Cell lysis was performed by addition of 50 µL lysozyme solution (2 mg mL<sup>-1</sup>, 50 mM Tris, pH 7.4), followed by incubation (37 °C, 1 h) and centrifugation (4 °C, 3 220 × g, 20 min).

### **2.9.2. PicoGreen® assay in 96-well microtiter plates**

Crude cell extract prepared and used for elongation of a short PCR product from pALXtreme-1a empty vector using primers P27 and P28 (Table 3). For the PCR (90 °C 2 min (1x), 90 °C 30 sec, 56 °C 30 sec, 65 °C 3 min, (25x ); 65 °C for 10 min (1x)), 1 – 4 µL crude cell extract, 0.20 mM of dNTP mix, 50 ng of plasmid template (pALXtreme-1a empty), and 10 pmol of each primer were used. PicoGreen® (Promega, Madison, USA) based quantification of dsDNA in MTP was performed according to manufacturer's recommendation.

## **2.10. 96-well microtiter plate-format Scorpion probe fluorescent screening assay for DNA polymerase activity**

### **2.10.1. Dpo4 expression in 96-well MTP**

Master microtiter plates containing Dpo4 mutant libraries cloned in pALXtreme-5b vector and transformed in *E. coli* BL21 Gold (DE3) *lacI*<sup>q1</sup> were duplicated using a 96-well pin replicator (constructed in Werkstatt der Biologie, RWTH Aachen University, Aachen, Germany) to pre-culture flat bottom microtiter plates (Greiner Bio-One, Frickenhausen, Germany) containing 200 µL LB<sub>Amp</sub> per well. The latter were cultivated for 12 h in a microtiter plate shaker (Multitron II, Infors, Einsbach, Germany; 37 °C, 900 rpm, 70% relative humidity) before being replicated to V-bottom microtiter plates (transparent polystyrene plate, Corning GmbH, Kaiserslautern, Germany) with 150 µL TYM-5052<sub>Amp</sub> auto-induction media per well (Studier, 2005). Expression was carried out at 37 °C for 18 h in a microtiter plate shaker (Multitron II, Infors, Einsbach, Germany; 900 rpm, 70% relative humidity), the expression cultures were harvested by centrifugation (4 °C, 3 220 × g, 20 min) using Eppendorf 5810R centrifuge (Eppendorf, Hamburg, Germany) stored at -20 °C for up to one week.

### 2.10.2. Scorpion probe assay in 96-well microtiter plates

The 96-well microtiter plates containing different Dpo4 variants were taken out of the freezer and kept at room temperature for 5 min. Cell lysates were prepared by imbibing each cell pellet for 5 min in 100  $\mu$ L Tris-HCl buffer (50 mM Tris-HCl, pH 7.4) and re-suspending by shaking (1 200 rpm for 5 min) using microplate shaker TiMix 2 (Edmund Bühler, Hechingen, Germany). Cell lysis was performed by addition of 100  $\mu$ L lysozyme solution (2 mg mL<sup>-1</sup>, 50 mM Tris, pH 7.4), followed by incubation (37 °C, 1 h) and centrifugation (4 °C, 3 220  $\times$  g, 20 min).

The cell lysate (10  $\mu$ L) was mixed with 15  $\mu$ L Scorpion probe reaction mix containing 1x ThermoPol buffer (New England Biolabs, Frankfurt am Main, Germany), 0.2 mM dNTP mix, 0.25  $\mu$ M 6-FAM-labeled Scorpion probe (Biomers.net, Ulm, Germany), 0.533  $\mu$ M BHQ-1 quencher (Biomers.net, Ulm, Germany) and 0.016  $\mu$ M single-stranded DNA template in a 96-well white PCR plates (Flat deck Thermo-Fast 96 detection plate, Thermo Scientific, Bonn, Germany) sealed with transparent film (MicroAmp<sup>TM</sup> Optical Adhesive Film, AB Biosciences, Allston, USA). Thermocycling (80 °C 30 sec, 51 °C, 30 sec, 65 °C, 1 min (25x)) and fluorescence reads during step 3 (65 °C, 1 min) of each cycle were carried out in Applied Biosystems 7300 Real Time PCR System (AB Bioscience, Allston, USA).

Standard deviation for the screening system was determined after transforming the Dpo4 WT plasmid cloned in pALXtreme-5b and transformed into the *E. coli* BL21 Gold (DE3) *laqI*<sup>q1</sup> strain. Colonies from LB<sub>Amp</sub> agar plates were transferred to 96-well microtiter plates and screened as described above. To get the true standard deviation, standard deviations of the slopes calculated for the first 15 cycles were calculated on the relative fluorescence values obtained from the 96-well microtiter plate (90 Dpo4 WT clones, 3 empty vector controls, 3 wells filled with growth media only).

### 2.11. Construction and screening of epPCR libraries of Dpo4

For the mutagenic PCR (98 °C 2 min (1x); 98 °C 30 sec, 56 °C 30 s, 72 °C 1 min (25x); 72 °C 3 min (1x)), 2.5 U of *Taq* DNA polymerase, 0.20 mM of dNTP mix, 50 ng of plasmid template (pALXtreme-5b harboring sso), 0.1 – 0.5 mM MnCl<sub>2</sub>, and 10 pmol of each primer (CCG GAT CGG ACT ACT AGC AGC TGT AAT AC and CGT GAC ATA ACT AAT TAC ATG ATG CGG) were

used. PCR products were purified by using a NucleoSpin PCR Purification Kit (Macherey-Nagel, Düren, Germany), re-amplified with primers P29 and P30 (Table 3) and cloned using PLICing as previously described (Blanusa et al., 2010b).

epPCR libraries cloned in pAXLtreme-5b and transformed in *E. coli* BL21 Gold (DE3)  $\text{lacI}^q$  cells were plated on  $\text{LB}_{\text{Amp}}$  plates supplemented with 0.1 mM IPTG and incubated at 37 °C for 18 h until sizeable uniform colonies were present. All cells were harvested from the agar plates and lysed by sonication and the supernatant was clarified by centrifugation (4 °C,  $3,220 \times g$ , 20 min) using Eppendorf 5810R centrifuge (Eppendorf AG, Hamburg, Germany). The resulting cell-free extract was used in a forward primer extension assay (see section 2.7) whereby the amount of product after 20 thermal cycles was analyzed on a 25 % acrylamide gel.

## 2.12. Construction of saturation mutagenesis libraries of Dpo4

In total, 19 non-conserved amino acid residues positioned within 4 - 6 Å from the substrate were selected for site-saturation mutagenesis based on Dpo4 crystal structure with bound matching and mismatching DNA template/primer (PDB codes: 1S9F and 1S97, respectively (Trincao et al., 2004)), HotSpot Wizard and conservation analysis. A two-step approach was utilized for the generation of SSM whereby pALXtreme-5b-*dpo4* WT, an universal forward primer binding upstream of *dpo4* and deleting a XbaI restriction site, and a site-specific reverse primer were used for generation of „megaprimers“. Reactions contained 1x Phusion High-Fidelity buffer and 2 U *Taq* Phusion polymerase (New England Biolabs, Frankfurt, Germany), 0.4  $\mu\text{M}$  of each primer (Table 3; P1-P21), 0.2 mM of each dNTP and template (10 ng plasmid/50  $\mu\text{L}$  mastermix). The following PCR protocol was used: 98 °C 30 sec (1 cycle); 98 °C 15 sec, 60 °C 30 sec, 72 °C 30 sec/kb template (18 cycles); 72 °C 3 min (1 cycle). The gel extracted „megaprimers“ (Nucleospin Extract II kit, Macherey Nagel, Düren, Germany) were cloned by MEGAWHOP (Miyazaki, 2011). Following template DNA removal through digestion by DpnI and XbaI (New England Biolabs, Frankfurt am Main, Germany), the libraries were transformed into *E.coli* BL21-Gold (DE3)  $\text{lacI}^q$  and plated on  $\text{LB}_{\text{Amp}}$  agar plates. Next, master microtiter plates were prepared by transferring individual clones to 96-well, flat bottom microtiter plates (transparent polystyrene MTPs (Greiner Bio-One, Frickenhausen, Germany) containing 100  $\mu\text{L}$   $\text{LB}_{\text{Amp}}$ , grown until saturation in a microtiter plate shaker

(Multitron II, Infors, Einsbach, Germany; 37 °C, 900 rpm, 70 % relative humidity) then stored at -80 °C after addition of sterile glycerol solution (25 % (v/v) final concentration).

## 2.13. Purification of Dpo4 variants for biochemical characterization

### 2.13.1. Flask Expression

Flask expression was carried out in 500 mL Erlenmeyer flasks filled with 50 mL TYM-5052<sub>Kan</sub> auto-induction media (Studier, 2005) by inoculating with 1% (v/v) overnight culture of *E. coli* BL21 Gold (DE3) *lacIq1 pALXtreme-1a-dpo4* prepared, cultivating the expression culture at 37 °C, 250 rpm for 18 hours and harvesting by centrifugation (4 °C, 3 220 × g, 20 min, Eppendorf 5810R centrifuge, Eppendorf, Hamburg, Germany), washed with phosphate-buffered saline (PBS, pH 7.4) and used immediately or stored at -20 °C.

### 2.13.2. Purification

Cell pellets were thoroughly re-suspended in 1xLEW washing buffer (Macherey Nagel GmbH, Düren, Germany) such that 40% (w/v) cell suspension was prepared. Cells were lysed by sonication (5 x 1 min, 40% sonication amplitude) using Vibracell VCX130 sonicator (Sonics & Materials, Newton, CT, USA), followed by heat treatment (80 °C, 10 min) to remove most *E.coli* proteins and centrifugation (4 °C, 16 000 x g, 20 min, Eppendorf 5804R, Eppendorf, Hamburg, Germany). The crude cell extracts was further cleared by filtration through a low-protein-binding filter (0.45 µm; Minisart RC 25 disposable syringe filter; Sartorius, Hamburg, Germany) before applying them to His-tag purification, pre-packed Protino<sup>®</sup> Ni-IDA columns (Macherey Nagel, Düren, Germany). After extensive washing with 1x LEW buffer (5 x 5 mL), the His-tagged protein was eluted in 3 mL 1x Elution Buffer (Macherey Nagel GmbH, Düren, Germany) whereby 10 - 12 fractions of 300 µL each were collected, analyzed on 10% SDS acrylamide gel and only the purest fractions were taken for buffer exchange by dialysis against storage buffer (10 mM Tris-HCl, 100 mM KCl, 1 mM DTT, 10 mM EDTA, pH 7.9). On the following day, the purified protein was either used for activity assays and characterization or stored at -20 °C after addition of glycerol (50% (v/v), final concentration).

## 2.14. Detailed SeSaM-Tv protocol

### 2.14.1 Biological and chemical materials

- 1) Purified plasmid containing the gene of interest suspended in nuclease-free water or TE buffer: 10 mM Tris-Cl (pH 7.5), 1 mM EDTA.
- 2) Six PCR primer sets for the preliminary steps and steps 1 to 4 of SeSaM (also included in Table 3). As a rule, the primers should not exceed 50 nucleotides in length and should preferably be of HPLC purity grade. The melting temperature of the primers should be calculated only for the gene-specific portion excluding the attached synthetic sequences (where applicable).

- i. Primer set 1 amplifies fragment(s) that contains the region targeted for random mutagenesis not exceeding 1.5 kb in length (as this would negatively affect the efficiency of step 3), preferably including gene-flanking restriction sites. Each primer binding site should be located 20 – 80 bp up- or downstream of the region to be mutagenized. The gene specific primer region should have a length of 18 - 25 bp. Name these primers “fwd\_X” and “rev\_X”, where X stands for the abbreviated name of the gene of interest
- ii. Primer set 2 comprises the gene specific forward and reverse primers (i.e. identical to the primer set 1 sequences) but with the following artificial oligos attached at 5`-ends. Name these primers “SeSaM\_fwd\_X” and “SeSaM\_rev\_X”, respectively.

SeSaM\_fwd sequence: 5'- CACTACCGCACTCCGTCG -3'

SeSaM\_rev sequence: 5'- GTGTGATGGCGTGAGGCAGC -3'

- iii. Primer set 3 has gene specific sequences identical to primers set 1 with the following SeSaM\_up or SeSaM\_dn sequences attached at the 5' ends of the forward and reverse primer, respectively. Name the primers “SeSaM\_up\_X” and “SeSaM\_dn\_X”.

SeSaM\_up sequence: 5'- CGCCTGTCAC -3'

SeSaM\_dn sequence: 5'- GCGGACAGTG -3'

- iv. Primer set 4 comprises solely the following sequences:

SeSaM\_fwd sequence: 5'- CACTACCGCACTCCGTCG -3'

SeSaM\_rev sequence: 5'- GTGTGATGGCGTGAGGCAGC -3'

- v. Primer set 5 comprises the following primer sequences, each of which additionally harbors a biotin tag (Bio-TEG, a biotin attached to a 15-atom mixed polarity triethylene glycol spacer) at 5'-ends. Name these primers "Bio\_SeSaM\_fwd" and "Bio\_SeSaM\_rev":

Bio\_SeSaM\_fwd: 5'- [Biotin]CACACTACCGCACTCCGTCG -3'

Bio\_SeSaM\_rev: 5'- [Biotin]GTGTGATGGCGTGAGGCAGC -3'

- vi. Primer 6 is labeled with the fluorescent dye (FITC, fluorescein isothiocyanate) at the 5'-end. This primer is used for quality control of SeSaM step 2.

FITC: 5'- [FITC]GCAATACGCAAACAGTCTCTCT -3'

- 3) NucleoSpin Extract II PCR Purification Kit (Macherey Nagel, Düren, Germany) and the associated buffers NT, NT3 as well as NTC buffer. NTC buffer which is separately supplied by Macherey Nagel, Düren, Germany, is indispensable for efficient purification of ssDNA in steps 1 and 2 of SeSaM. This system is highly recommended since the SeSaM protocol has been validated with this particular kit.
- 4) Nuclease-free water (for PCR amplification and elution of DNA after purification) and deionized water (for buffer preparation).
- 5) Solution M: 100 mM NaOH in deionized water. Store at 4 °C.
- 6) Solution R: 0.1 % SDS (w/v) in deionized water. Store at 4 °C.
- 7) Binding and Washing buffer 1 (1x B&W): 10 mM Tris-HCl (pH 7.5), 1 mM EDTA (pH 8.0), 2 M NaCl.
- 8) Binding and Washing buffer 2 (2x B&W): 10 mM Tris-HCl (pH7.5), 1 mM EDTA (pH 8.0), 1 M NaCl.
- 9) dNTP buffer (pH 7.5): 1 mM Tris-HCl (pH 7.5), 0.1 mM DTT. Store at -20 °C.
- 10) Ethanol (99% purity).
- 11) Solution C: 20 mM Iodine in ethanol (99%). Store solution C for not more than 1 week at 4 °C.
- 12) Oligo buffer: 10 mM Tris-HCl (pH 7.5), 0.1 mM EDTA (pH 8.0). Store at -20 °C
- 13) 10x Tris borate EDTA buffer (10x TBE): 0.89 M Tris-HCl (pH 8.0), 0.89 M boric acid, 20 mM EDTA (pH 8.0). Store at room temperature.
- 14) Tris borate EDTA loading dye (6x TBE-loading dye): 60% 10 x Tris borate EDTA buffer (v/v), 30% glycerol (v/v), 1 mg mL<sup>-1</sup> (w/v) bromophenol blue. Store at 4 °C.
- 15) 40% acrylamide-bis-acrylamide solution. Store at 4 °C.

- 16) 10% (w/v) Ammonium persulfate solution (10% APS), prepared in deionized water. Store at -20 °C.
- 17) N,N,N',N'- Tetramethylethylenediamine (TEMED, 99% purity). Store at room temperature.
- 18) *Taq* DNA polymerase and corresponding buffer (e.g. New England Biolabs, Frankfurt, Germany). Store at -20 °C.
- 19) High fidelity (HF) DNA polymerase and corresponding buffer (e.g. Phusion® DNA polymerase and 5x Phusion® buffer; New England Biolabs, Frankfurt, Germany). Store at -20 °C.
- 20) Terminal deoxynucleotidyl transferase (TdT) (20 U  $\mu\text{L}^{-1}$ ) and 10x TdT buffer (New England Biolabs, Frankfurt, Germany). Store at -20 °C.
- 21) 3D1 polymerase (5 U  $\mu\text{L}^{-1}$ ) and 10x Super*Taq* reaction buffer (HT BioTechnology, Cambridge, U.K.). Store enzymes at -20 °C. 3D1 polymerase was expressed and purified to homogeneity according to published protocols (d' Abbadie et al., 2007a; Mundhada et al., 2011a). Vent (exo-) DNA polymerase and ThermoPol buffer from New England Biolabs (Frankfurt, Germany) can be used as alternative enzymes in step 3 of SeSaM. Using the Vent (exo-) instead of the 3D1-polymerase generates fewer consecutive nucleotide exchanges in SeSaM-Tv-classic libraries. Vent (exo-) should be the polymerase of choice in step 3 of SeSaM-R protocol. The reaction conditions with either polymerase remain the same.
- 22) Streptavidin coated magnetic beads (e.g. M-PVA SAV1, Chemagen, Baesweiler, Germany). Store at 4 °C.
- 23) 50x TAE-buffer: 2 M Tris, 1 M Acetic acid 50 mM EDTA, with pH adjusted to 8.5 using NaOH. Store at room temperature.
- 24) 0.8% (w/v) agarose solution in 1x TAE-buffer. Store at 65 °C for up to 2 weeks.
- 25) 0.001 % (v/v) ethidium bromide prepared in deionized water. Store at room temperature.
- 26) 6x DNA Loading dye: 67% (v/v) glycerol, 1 mg  $\text{mL}^{-1}$  bromophenol blue, 10% 10x TAE-buffer. Store at -20 °C.
- 27) Nucleotide analog dPTP $\alpha$ S as well as phosphorothioated nucleotides (dATP $\alpha$ S, dCTP $\alpha$ S, dTTP $\alpha$ S, dGTP $\alpha$ S) which can be purchased from BIOLOG Life Sciences Institute (Bremen,

Germany); standard dNTPs are available from NEB (New England Biolabs, Frankfurt, Germany). Prepare the following dNTP solutions in oligo buffer and store at -20 °C:

- i. individual solutions of dATP, dGTP, dCTP and dTTP; stock concentration of 100 mM;
- ii. dNTP mix: 10 mM each of dATP, dGTP, dTTP, dCTP;
- iii. individual solutions of phosphorothioate nucleotides (dATP $\alpha$ S, dGTP $\alpha$ S, dCTP $\alpha$ S and dTTP $\alpha$ S); stock concentration of 5 mM;
- iv. d(GTP/TTP/CTP)-mix: 10 mM each of dGTP, dTTP, dCTP;
- v. d(ATP/TTP/CTP)-mix: 10 mM each of dATP, dTTP, dCTP;
- vi. d(ATP/TTP/GTP)-mix: 10 mM each of dATP, dTTP, dGTP;
- vii. d(ATP/GTP/CTP)-mix: 10 mM each of dATP, dGTP, dCTP; dPTP $\alpha$ S (6-(2-deoxy- $\beta$ -d-ribofuranosyl)-3,4-dihydro-8H-pyrimido-[4,5-C][1,2]oxazin-7-one); stock concentration of 20  $\mu$ M; The rate of incorporation of dPTP $\alpha$ S by TdT is lower relative to that observed for non-phosphorothioate dPTP. This allows more precise control over the addition pattern obtained in SeSaM step 2.
- viii. d(ATP/ATP $\alpha$ S/GTP/TTP/CTP)-mixes: in total 8 mixes with 0%, 10%, 15%, 20%, 25%, 30%, 35% and 40% dATP $\alpha$ S should be prepared and labeled "A0" to "A40", respectively; on these labels "A" stands for adenine and the number (e.g. 10) signifies the percentage of dATP $\alpha$ S in relation to the total amount of adenine species (dATP $\alpha$ S + dATP); the final concentration for each nucleotide in the mix should be 1 mM (e.g. to prepare 50  $\mu$ L A10 mix combine 9  $\mu$ L dATP (5 mM stock), 1  $\mu$ L dATP $\alpha$ S (5 mM stock), 5  $\mu$ L d(GTP/TTP/CTP)-mix (10 mM stock), 35  $\mu$ L dNTP buffer);
- ix. d(GTP/GTP $\alpha$ S/ATP/TTP/CTP)-mixes: in total 8 mixes with 0%, 10%, 15%, 20%, 25%, 30%, 35% and 40% dGTP $\alpha$ S should be prepared and labeled "G0" to "G40", respectively; "G" stands for guanine and the number (e.g. 10) signifies the percentage of dGTP $\alpha$ S in relation to the total amount of guanine species (dGTP $\alpha$ S + dGTP); the final concentration for each nucleotide in the mix should be 1 mM; (e.g. to prepare 50  $\mu$ L G10 mix combine 9  $\mu$ L dGTP (5 mM stock), 1  $\mu$ L dGTP $\alpha$ S (5 mM stock), 5  $\mu$ L d(ATP/TTP/CTP)-mix (10 mM stock), 35  $\mu$ L dNTP buffer);
- x. d(CTP/CTP $\alpha$ S/ATP/TTP/GTP)-mixes: in total 8 mixes with 0%, 10%, 15%, 20%, 25%, 30%, 35% and 40% dCTP $\alpha$ S should be prepared and labeled "C0" to "C40", respectively; "C" stands for cytosine and the number (e.g. 10) signifies the percentage of dCTP $\alpha$ S in relation to the total amount of cytosine species (dCTP $\alpha$ S + dCTP); the



final concentration for each nucleotide in the mix should be 1 mM; (e.g. to prepare 50  $\mu$ L C10 mix combine 9  $\mu$ L dCTP (5 mM stock), 1  $\mu$ L dCTP $\alpha$ S (5 mM stock), 5  $\mu$ L d(ATP/TTP/GTP)-mix (10 mM stock), 35  $\mu$ L dNTP buffer);

- xi. d(TTP/TTP $\alpha$ S/ATP/GTP/CTP)-mixes: in total 8 mixes with 0%, 10%, 15%, 20%, 25%, 30%, 35% and 40% dTTP $\alpha$ S should be prepared and labeled "T0" to "T40", respectively; "T" stands for thymine and the number (e.g. 10) signifies the percentage of dTTP $\alpha$ S in relation to the total amount of thymine species (dTTP $\alpha$ S + dTTP); the final concentration for each nucleotide in the mix should be 1 mM; (e.g. to prepare 50  $\mu$ L T10 mix combine 9  $\mu$ L dTTP (5 mM stock), 1  $\mu$ L dTTP $\alpha$ S (5 mM stock), 5  $\mu$ L d(ATP/GTP/CTP)-mix (10 mM stock), 35  $\mu$ L dNTP buffer);

### 2.14.2. Equipment

- 1) Microcentrifuge (e.g. Eppendorf, Hamburg, Germany) with rotor for 1.5 mL and 2 mL tubes.
- 2) Thermocycler with gradient and temperature decrement option (e.g. Thermocycler proS, Eppendorf, Hamburg, Germany).
- 3) Small volume photometer for quantitation of DNA concentrations in sample volumes of 1 – 2  $\mu$ L (e.g. NanoDrop photometer, NanoDrop Technologies, Wilmington DE, USA).
- 4) Agarose gel electrophoresis equipment.
- 5) Phosphorimager with capability to measure Excitation 494 nm; Emission 518 nm (e.g. FLA-3000 from Fujifilm, Düsseldorf, Germany).
- 6) Acrylamide gel electrophoresis system (e.g. Mini-Protean Tetra Cell system from Bio-Rad, München, Germany).
- 7) Recommended: Automated gel electrophoresis system (e.g. Experion system and High Sens RNA chip from Bio-Rad, München, Germany).
- 8) PCR tubes (0.2 mL).
- 9) Microcentrifuge tubes (1.5 mL or 2 mL).
- 10) 4 magnets (diameter 20 - 30 mm).

### 2.14.3. SeSaM preliminary experiments

When SeSaM is performed for the first time for a given gene, the following two preliminary experiments should be performed.

#### 2.14.3.1. SeSaM preliminary experiment I: SeSaM step 1 and step 3 template generation

In the first preliminary experiment, linear DNA templates for step 1 and step 3 of SeSaM must be generated. These templates harbor artificial flanking sequences which play an important role in SeSaM step 1 and 4.

- 1) Assemble the following pre-mastermix (200  $\mu$ L) on ice in a 0.2 mL tube. Thaw all non-enzyme components at room temperature and enzymatic components on ice, mix by vortexing and collect by a brief centrifugation.

<u>Pre-mastermix (160 <math>\mu</math>L):</u>	<u>Volume:</u>
5x Phusion <sup>®</sup> polymerase buffer	40 $\mu$ L
dNTP-mix (10 mM)	4 $\mu$ L
Plasmid template	40 ng/kb of plasmid <sup>A</sup>
Phusion <sup>®</sup> polymerase (2 U/ $\mu$ L) *	2 $\mu$ L
Nuclease-free H <sub>2</sub> O	to a total volume of 160 $\mu$ L

\* High-fidelity polymerases such as Phusion<sup>®</sup> are recommended in the preliminary step I to ensure error-free synthesis of templates. High-fidelity polymerases should not be used in steps 1 to 4 of SeSaM.

<sup>A</sup> e.g. for a plasmid with a total length of 5 kb, use 2  $\mu$ L of a 100 ng/ $\mu$ L purified plasmid.

- 2) Mix the pre-mastermix by pipetting.
- 3) Divide the pre-mastermix into 4x 40  $\mu$ L aliquots, then add to each tube add a total of 10  $\mu$ L from the following primer pairs (*i.e.* 5  $\mu$ L apiece of forward and reverse primer).
  - i. SeSaM\_fwd\_X and Rev\_X; Recommended labeling of the PCR reaction and product - "step 1\_fwd\_X".
  - ii. SeSaM\_Rev\_X and fwd\_X; Recommended labeling of the PCR tube reaction and product - "step 1\_rev\_X".
  - iii. SeSaM\_fwd\_X and SeSaM\_dn\_X; Recommended labeling of the PCR tube reaction and product - "step 3\_rev\_X".

- iv. SeSaM\_rev\_X and SeSaM\_up\_X; Recommended labeling of the PCR tube reaction and product - "step 3\_fwd\_X".
- 4) If the optimal annealing temperature for these primer sets has been confirmed experimentally, skip points 5 - 7 and proceed to point 8.
- 5) Split each PCR reaction (50  $\mu$ L) into 4 x 12  $\mu$ L aliquots in fresh PCR tubes. This allows each aliquot to be tested at a different annealing temperature in a gradient PCR in order to identify the most suitable annealing temperature for each primer set.
- 6) Run PCR samples using a temperature gradient in the range 50  $^{\circ}$ C to 70  $^{\circ}$ C as follows: 98  $^{\circ}$ C 30 sec (1 cycle); 98  $^{\circ}$ C 15 sec, gradient Tm 30 sec, 72  $^{\circ}$ C 20 sec/kb of the SeSaM template (18 x); 72  $^{\circ}$ C 5 min (1 x); followed by a 4  $^{\circ}$ C hold to protect the samples.
- 7) Analyze PCR products on a 0.8% TAE agarose gel (7.5 V/cm; 40 min). Identify the primer annealing temperature at which the highest amount of specific product is formed and use this annealing temperature for the following PCR reactions.
- 8) Set up PCR mastermixes (as outlined in points 1 - 3) whereby a total volume of 200  $\mu$ L per primer pair is prepared and split into four aliquots of 50  $\mu$ L before thermocycling using the PCR program with the selected Tm outlined in point 6, above. Combine each 4x 50  $\mu$ L of identical PCR products after the PCR is completed.
- 9) Confirm that PCR products were successfully obtained by loading a 4  $\mu$ L aliquot from each reaction on a 0.8% TAE agarose gel (7.5 V/cm; 40 min). It is important that the SeSaM step 1 template is not contaminated with non-specific products. If nonspecific products are present, load and separate the PCR product on agarose gel and perform a gel extraction/ purification.
- 10) Individually column-purify the remainder of each reaction using the NucleoSpin Extract II PCR purification kit and determine the concentration of dsDNA photometrically (*e.g.* using NanoDrop spectrophotometer). Store the purified samples at -20  $^{\circ}$ C until needed.

### **2.14.3.2. SeSaM preliminary experiment II: determination of the optimal concentration of phosphorothioate nucleotides for step 1 of SeSaM**

When SeSaM is performed for the first time for a given gene, it is also crucial to determine the optimal concentration of phosphorothioate nucleotides to use in step 1. This ensures that fragments with uniform size distribution spanning the entire gene sequence are generated in the first step of SeSaM.

- 1) Prepare a “forward” (fwd) and a “reverse” (rev) mastermix (180 µL each) in 1.5 mL tubes on ice. Thaw all non-enzymatic components at room temperature, mix by vortexing and collect by brief centrifugation. Keep enzymes on ice.

<b>Forward (fwd) pre-mastermix (144 µL):</b>	<b>Volume:</b>
10x <i>Taq</i> DNA polymerase buffer	18 µL
SeSaM_fwd primer (5 µM)	18 µL
Rev_X primer (5 µM)	18 µL
SeSaM_fwd_step 1 template	96 ng/kb of SeSaM template *
<i>Taq</i> DNA polymerase (5 U/µL)	3.6 µL
Nuclease-free H <sub>2</sub> O	to a total volume of 144 µL

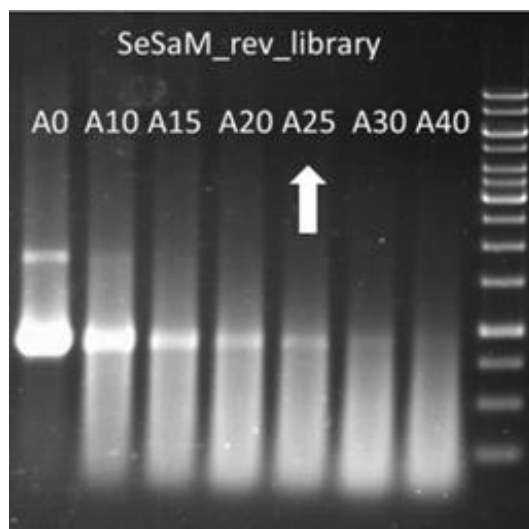
<b>Reverse (rev) pre-mastermix (144 µL):</b>	<b>Volume:</b>
10x <i>Taq</i> DNA polymerase buffer	18 µL
SeSaM_rev primer (5 µM)	18 µL
fwd_X primer (5 µM)	18 µL
SeSaM_rev_step 1 template	96 ng/ 1 kb SeSaM template *
<i>Taq</i> DNA polymerase (5 U/µL)	3.6 µL
Nuclease-free H <sub>2</sub> O	to a total volume of 144 µL

\* e.g. for a gene of 1 kb, a total of 4.8 µL of a 20 ng/µL purified SeSaM\_step 1 template is used.

- 2) Prepare the PCR reactions for the test library on ice using 16x 8 µL aliquots of either reverse or forward pre-mastermix in PCR tubes. Label tubes “AF0” - “AF40” for A-forward test libraries, “AR10” - “AR40” for A-reverse test libraries; “GF0”-“GF40” for G-forward libraries and “GR10”-“GR40” for G-reverse test libraries. To each tube add 2 µL of the corresponding phosphorothioate nucleotide premix (“A0” to “A40” or “G0” to “G40”, respectively), prepared as described in Materials, point 27. When performing preliminary experiment II for SeSaM-R protocol d(CTP/CTPαS/ATP/TTP/GTP)-mixes (C0-C40) or d(TTP/TTPαS/ATP/GTP/CTP)-mixes (T0-T40) should be used to target T- and C-sites. No other modifications in reaction conditions are required.
- 3) Run all samples in a PCR cycler for 15 cycles using the following PCR settings: 94 °C 120 sec (1 x); 94 °C 30 s, optimal T<sub>m</sub> of primers 30 s, 72 °C 30 s/kb of SeSaM template (14 x); 72 °C 180 sec (1 x); followed by a 4 °C hold to protect the samples.

- 4) After completion of the PCR reaction, add 1.1  $\mu\text{L}$  of freshly prepared solution C (the iodine solution used to cleave phosphorothioester bonds) to each PCR tube. Mix each tube immediately and vigorously by vortexing (10 sec).
- 5) Incubate the samples in a PCR cycler for 30 min at 70 °C. Make sure the lid of the thermal cycler is pre-heated to 80 °C to prevent sample evaporation. After 30 min the solutions should change color from brownish-orange to transparent.
- 6) Run the samples (4  $\mu\text{L}$ ) on a 0.8% TAE agarose gel (7.5 V/cm; 40 min). Evaluate the gel images to determine the condition under which the most homogeneous distribution of DNA fragments is achieved, i.e., select the concentration of dNTP $\alpha$ S at which the fragments are homogeneously distributed over the entire gene length and only traces of the uncleaved, full length DNA band are visible (Box 1).

**BOX 1**



An exemplary agarose gel used for evaluation of the optimal phosphorothioate nucleotide concentration. A0 to A40 indicates the fraction of phosphorothioate adenine in relation to adenine expressed in percentage. Uncleaved PCR product (A0) is loaded in the leftmost lane; in the samples prepared with dNTP-mix that contains a certain percentage of phosphorothioate nucleotides (right of A0), one can observe a decrease in intensity of the full length band as well as a smearing below that increases with dNTP $\alpha$ S concentration. This smearing signifies DNA cleavage at sites where phosphorothioate nucleotides have been incorporated. The lane labeled A40 represents a case where the concentration of dNTP $\alpha$ S nucleotides is too high as hardly any uncleaved fragment are visible and the distribution of fragments is uneven (i.e., small fragments are predominant). Lane A10 shows a case when the concentration of phosphorothioate nucleotides is too low, *i.e.* uneven smearing with prevailing large fragments and a very high fraction of uncleaved PCR product is observed. A final concentration of phosphorothioate nucleotides of 25% (A25) for the SeSaM\_rev\_library seems to be the optimal condition in the presented example.

#### 2.14.4. SeSaM step 1: Generation of ssDNA fragment pool with random length distribution

In step 1 of SeSaM, the incorporation of phosphorothioate nucleotides into PCR products is carried out. The optimal phosphorothioate nucleotide concentration is determined as explained in SeSaM preliminary step II. In SeSaM-Tv classic protocol, A- and G-sites are targeted with the pyrimidine analog dPTP, therefore dGTP $\alpha$ S or dATP $\alpha$ S are randomly incorporated in the gene sequence. In step 1 of SeSaM-R protocol, instead of preparing an A- and G- library, carry out a T- and C-library (*e.g.* for preparation of a T-library replace the d(ATP/TTP/CTP)-mix with d(ATP/GTP/CTP)-mix; similarly dGTP should be replaced by dTTP and dGTP $\alpha$ S by dTTP $\alpha$ S).

SeSaM step 1 comprises two parts – the incorporation of the phosphorothioate nucleotides and the iodine cleavage of the PCR products and resulting biotinylated ssDNA fragments are isolated using streptavidin-coated magnetic beads.

1) Assemble the following “forward” (fwd) and “reverse” (rev) pre-mastermixes (2 x 423  $\mu$ L) on ice. Thaw all non-enzyme components at room temperature, mix by a brief vortex and collect by a brief centrifugation. Keep all enzymatic components on ice.

<b>Fwd pre-mastermix (423 <math>\mu</math>L):</b>	<b>Volume:</b>
SeSaM template (SeSaM step 1_fwd_X template)	360 ng/kb SeSaM template *
10x <i>Taq</i> DNA polymerase buffer	45 $\mu$ L
<i>Taq</i> DNA polymerase (5 U $\mu$ L <sup>-1</sup> )	9 $\mu$ L
SeSaM_fwd_Bio primer (5 $\mu$ M)	45 $\mu$ L
rev_X primer (5 $\mu$ M)	45 $\mu$ L
Nuclease-free H <sub>2</sub> O	to a total volume of 423 $\mu$ L

<b>Rev pre-mastermix (423 <math>\mu</math>L):</b>	<b>Volume:</b>
SeSaM template (SeSaM step 1_rev_X template)	360 ng/kb SeSaM template *
10x <i>Taq</i> DNA polymerase buffer	45 $\mu$ L
<i>Taq</i> DNA polymerase (5 U $\mu$ L <sup>-1</sup> )	9 $\mu$ L
SeSaM_rev_Bio primer (5 $\mu$ M)	45 $\mu$ L
fwd_X primer (5 $\mu$ M)	45 $\mu$ L
Nuclease-free H <sub>2</sub> O	to a total volume of 423 $\mu$ L.

\* *e.g.* for a 1 kb gene use a total of 6  $\mu$ L of a 60 ng/ $\mu$ L purified SeSaM\_fwd\_step 1 template

2) Pipette 188  $\mu\text{L}$  from each mastermix into each of two new PCR tubes and keep the remaining volume of each pre-mastermix (47  $\mu\text{L}$ ) in the original PCR tube. Add the respective nucleotide mix according to the pipetting scheme as follows.

<b><i>A-fwd library mastermix:</i></b>	<b>Volume:</b>
<b>fwd</b> pre-mastermix:	188 $\mu\text{L}$
d(GTP/TTP/CTP)-mix (10 mM)	4 $\mu\text{L}$
dATP (5 mM)	X $\mu\text{L}$ *
dATP $\alpha$ S (5 mM)	Y $\mu\text{L}$ *
Total volume:	200 $\mu\text{L}$

<b><i>A-rev library mastermix:</i></b>	<b>Volume:</b>
<b>rev</b> pre-mastermix	188 $\mu\text{L}$
d(GTP/TTP/CTP)-mix (10 mM)	4 $\mu\text{L}$
dATP (5 mM)	X $\mu\text{L}$ *
dATP $\alpha$ S (5 mM)	Y $\mu\text{L}$ *
Total volume:	200 $\mu\text{L}$

<b><i>G-fwd library mastermix:</i></b>	<b>Volume:</b>
<b>fwd</b> pre-mastermix	188 $\mu\text{L}$
d(ATP/TTP/CTP)-mix (10 mM)	4 $\mu\text{L}$
dGTP (5 mM)	X $\mu\text{L}$ *
dGTP $\alpha$ S (5 mM)	Y $\mu\text{L}$ *
Total volume:	200 $\mu\text{L}$

<b><i>G-rev library mastermix:</i></b>	<b>Volume:</b>
<b>rev</b> pre-mastermix	188 $\mu\text{L}$
d(ATP/TTP/CTP)-mix (10 mM)	4 $\mu\text{L}$
dGTP (5 mM)	X $\mu\text{L}$ *
dGTP $\alpha$ S (5 mM)	Y $\mu\text{L}$ *
Total volume:	200 $\mu\text{L}$

\*The total volume of dATP and dATP $\alpha$ S (dGTP and dGTP $\alpha$ S, respectively) should add up to 8  $\mu\text{L}$ ; X  $\mu\text{L}$  is the volume of dATP (dGTP, respectively) and Y  $\mu\text{L}$  the volume of dATP $\alpha$ S (dGTP $\alpha$ S, respectively) used the respective reactions (e.g. to prepare dNTP mix (A25) containing 25 % dATP $\alpha$ S, one would add 2  $\mu\text{L}$  of dATP $\alpha$ S and 6  $\mu\text{L}$  of dATP to the reaction mixture. The percentage of dATP $\alpha$ S used for a given gene should be determined experimentally as outlined in SeSaM Preliminary Step II).



<b>Positive control mastermix:</b>	<b>Volume:</b>
Pre-mastermix	47 $\mu$ L
Nuclease-free H <sub>2</sub> O	2 $\mu$ L
dNTP-mix (10 mM)	1 $\mu$ L
Total volume:	50 $\mu$ L

- 3) Split each mastermix (200  $\mu$ L) into 4 x 50  $\mu$ L aliquots in 0.2 mL PCR-tubes and run all samples in a PCR cycler using the following parameters: 94 °C 120 sec (1 x); 94 °C 30 sec, 60 °C 30 sec, 72 °C 30 sec/kb of the SeSaM template-DNA (19 x); 72 °C for 3 min (1 x); followed by a 4 °C hold to protect the samples.
- 4) Pool the four identical PCR reactions together and run 4-5  $\mu$ l from each sample on an 0.8% TAE agarose gel (7.5 V/cm; 40 min). A specific band corresponding to the length of the gene of interest should be clearly visible on the gel.
- 5) Column-purify the dsDNA products using NucleoSpin Extract II PCR Purification Kit and NT buffer according to manufacturer's recommendation. Elute DNA in 92  $\mu$ L elution buffer (NE-buffer or nuclease-free water) and determine dsDNA concentration.
- 6) Combine 80  $\mu$ L of the purified products with 10  $\mu$ L of 10x *Taq* DNA polymerase buffer in a fresh 0.2 mL tube. Add 10  $\mu$ L of freshly prepared solution C to each tube, and thoroughly mix by pipetting. Incubate for 30 min at 70 °C in a thermal cycler. During this incubation period proceed with steps 7 - 10.
- 7) Resuspend the magnetic beads by a brief vortex, until all beads are in suspension.
- 8) Transfer 50  $\mu$ L of the magnetic bead solution into each of four 0.2 mL tubes. Magnetic beads tend to precipitate from solution so make sure they are fully suspended before aliquoting them. Place each tube horizontally on a magnet and wait until the beads have settled down. Discard the liquid while keeping the tube pressed against the magnet.
- 9) Add 100  $\mu$ L 2x B&W buffer to each tube and resuspend by pipetting. Place tube on the magnet and wait until the beads have settled down. Prolonged contact between the PCR tubes and the magnet may cause the magnetic beads to firmly stick to the wall of the tube and make resuspension difficult. Remove the supernatant by pipetting.
- 10) Add 100  $\mu$ L 2x B&W buffer to each tube and resuspend by pipetting.
- 11) Run 4  $\mu$ L of the cleaved DNA on an 0.8% TAE agarose gel in order to check the fragmentation pattern of samples after cleavage and before purification (use 7.5 V/cm; 40 min), as explained in SeSaM preliminary step II.

- 12) Transfer the remainder of each cleaved library to one of the four tubes containing the magnetic beads resuspended in 2x B&W buffer, and mix by pipetting. Incubate each tube for 30 min at room temperature to allow the biotin tag to interact to the streptavidin-coated beads. After the incubation period is over, place each tube on a magnet and wait until the beads have settled down. Remove the supernatant by pipetting.
- 13) To each tube, add 100  $\mu\text{L}$  solution M and resuspend by pipetting. Incubate each tube for 2 min at room temperature. Place each tube on the magnet and wait until the beads have settled down. Remove the supernatant by pipetting.
- 14) Add 100  $\mu\text{L}$  solution M and resuspend by pipetting. Place tube on a magnet and wait until the beads have settled down. Remove the supernatant by pipetting (it is important to make sure that all traces of solution M are removed).
- 15) Add 100  $\mu\text{L}$  1x B&W buffer and resuspend by pipetting. Place tube on a magnet and wait until the beads have settled down. Remove the supernatant by pipetting.
- 16) Add 75  $\mu\text{L}$  solution R and resuspend by pipetting. Incubate the samples in a PCR cycler for 10 min at 98 °C. Make sure the lid is pre-heated to 104 °C to avoid sample evaporation. Do not let the tube cool down!
- 17) Place the hot tube on a magnet, wait until the beads have settled down and immediately recover the supernatant by pipetting in a fresh 1.5 mL tube. Keep the supernatant. Individually column-purify the recovered ssDNA fragments from each tube with NucleoSpin Extract II PCR Purification Kit. Use NTC buffer (300  $\mu\text{L}$  per 75  $\mu\text{L}$  ssDNA) for binding DNA to the column instead of NT buffer. Elute DNA in 30  $\mu\text{L}$  NE-buffer or nuclease-free water. Measure the concentration spectrophotometrically using appropriate settings for ssDNA. The concentration of ssDNA fragments after SeSaM step 1 is usually in the range of 10 - 20  $\text{ng } \mu\text{l}^{-1}$ .
- 18) If available, we recommend using a microfluidic electrophoresis device (*e.g.* Experion Automated Electrophoresis System from BioRad) after DNA purification to confirm and accurately evaluate the size distribution of ssDNA fragments, especially those in the low molecular weight range. The fragments generated should be homogeneously distributed. If the majority of PCR product remains uncleaved for a particular fragment library, repeat SeSaM step 1 by increasing the concentration of dNTP $\alpha$ S by 5% for the corresponding library. On the other hand, if many short fragments are observed, repeat SeSaM step 1 by decreasing the concentration of dNTP $\alpha$ S by 5%.

### 2.14.5. SeSaM step 2: Enzymatic elongation of single stranded DNA with degenerate base

In SeSaM step 2, the ssDNA fragments generated in step 1 are elongated with degenerate base analog dPTP $\alpha$ S in a TdT-catalyzed reaction.

- 1) Prepare five reactions (A-fwd, A-rev, G-fwd, G-rev libraries and the positive control) in 0.2 mL tubes, on ice, as detailed below. Reaction components should be added in the listed order. Make sure TdT is added last. All non-enzyme components should be thawed at room temperature and all enzymatic components on ice.

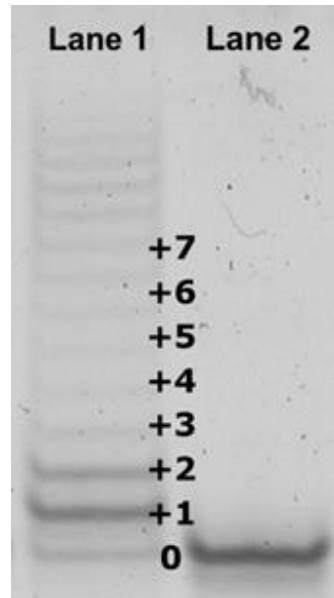
<b>Degenerate base addition reaction mix:</b>	<b>Volume:</b>
10x TdT buffer	2.5 $\mu$ L
10x CoCl <sub>2</sub> (25 mM)	2.5 $\mu$ L
SeSaM step 1 product	1 pmol of step 1 product
dPTP $\alpha$ S (20 $\mu$ M)	2 $\mu$ L
TdT (20 U $\mu$ L <sup>-1</sup> )	2 $\mu$ L
Nuclease-free H <sub>2</sub> O	to a total volume of 25 $\mu$ L

<b>Positive control (25 <math>\mu</math>L):</b>	<b>Volume:</b>
10x TdT buffer	2.5 $\mu$ L
10x CoCl <sub>2</sub> (25 mM)	2.5 $\mu$ L
FITC-labeled primer (2 $\mu$ M)	2 $\mu$ L
dPTP $\alpha$ S (5 $\mu$ M)	2 $\mu$ L
TdT (20 U $\mu$ L <sup>-1</sup> )	2 $\mu$ L
Nuclease-free H <sub>2</sub> O	14 $\mu$ L

- 2) Run all samples in a PCR cycler at 37 °C for 120 min, immediately followed by heat inactivation of TdT for 30 min at 75 °C. Make sure the lid of thermal cycler is pre-heated to 40 °C during first incubation step and to 80 °C during the inactivation step to prevent evaporation.
- 3) Keep the positive control aside (SeSaM step 2 positive control; store at -20 °C), and purify the other 4 samples using the NucleoSpin Extract II PCR Purification Kit and NTC buffer according to manufacturer's recommendation. Elute DNA in 25  $\mu$ L elution buffer or nuclease-free water.

- 4) Measure the concentration of ssDNA spectrophotometrically and store samples (SeSaM step 2 products) at -20 °C until the positive control has been analyzed.
- 5) Run the positive control of SeSaM step 2 on a 25% acrylamide gel in order to evaluate the resulting degenerate nucleotide addition pattern. To cast a 25% acrylamide gel, mix in a clean vessel 1.7 mL of deionized H<sub>2</sub>O, 2 mL of 5x TBE buffer, 6.25 mL of 40% acrylamide-bis-acrylamide solution, 100 µL of APS and 8 µL of TEMED. Allow the gel to solidify for at least 15 min. Place the gel in the running chamber filled with fresh 1x TBE buffer and run the gel for 5 min at 100 V prior to applying your samples. In the meantime, mix 5 µL of 6x TBE loading dye with the 25 µL positive control reaction. Prepare a negative control by mixing together 2 µL FITC primer (2 µM), 8 µL deionized water and 2 µL of 6x acrylamide loading dye. Load 6 µL of positive control sample and 4 µL of negative control samples per well. Do not use the outermost wells of the gel if possible. Run the gel at 170 V and 300 mA for 2.5 h. Visualize the gel without removing it from the glass casting chamber using phosphorimager/ scanner (ex. 494 nm/ em. 518 nm). Based on this gel, a conclusion about the number of nucleotide additions can be drawn. Ideally, all FITC primers should be “tailed” whereby fragments elongated with 2 or 3 degenerate base analog additions should be the predominant product (see Box 2).

**BOX 2**



Elongation of FITC labeled primer by TdT separated on acrylamide gel and visualized on phosphoroimager (ex. 494 nm/ em. 518 nm). Each stair of the ladder in the lane 1 (positive control) represents an addition of one universal base to the FITC labeled oligonucleotide. In lane 2 (negative control) FITC primer that has not been subjected to elongation reaction is loaded as a reference.

### 2.14.6. SeSaM step 3: Synthesis of the full length gene

In step 3 of SeSaM, the “tailed” DNA fragments (step 3 products from Section 3.3) are elongated to full length genes. In this step the forward and reverse libraries are elongated by combining the “SeSaM\_step3\_ templates” (generated in SeSaM preliminary experiment II) with the step 2 products (from step 4 of Section 3.3). If a recombined library with increased mutational load is desired, in SeSaM step 3 use equal amounts (80 ng each) of SeSaM step 2 forward and reverse products. No SeSaM\_step3\_ templates should be supplied as forward and reverse products serve to template each other’s elongation.

- 1) Prepare pre-mastermix for four reactions of 50  $\mu\text{L}$  (2 for fwd and 2 for rev libraries) on ice in 0.2 mL tubes.

<b>Pre-mastermix (120 <math>\mu\text{L}</math>):</b>	<b>Volume:</b>
10x SuperTaq buffer	20 $\mu\text{L}$
dNTP-mix (10 mM)	5 $\mu\text{L}$
3D1 polymerase (5 U $\mu\text{L}^{-1}$ )	4 $\mu\text{L}$
Nuclease-free H <sub>2</sub> O	to a total volume of 120 $\mu\text{L}$

- 2) Split the pre-mastermix into 4 x 30  $\mu\text{L}$  aliquots in fresh 0.2 mL tubes.
- 3) To each aliquot, one of the following combinations of SeSaM\_step 3\_templates and SeSaM step 2 products should be added; 80 ng from each in a total volume of 50  $\mu\text{L}$  is recommended
  - v. Mastermix + step 3\_fwd template + A-Fwd step 2 library
  - vi. Mastermix + step 3\_fwd template + G-Fwd step 2 library
  - vii. Mastermix + step 3\_rev template + A-Rev step 2 library
  - viii. Mastermix + step 3\_rev template + G-Rev step 2 library

*e.g.* for a 1 kb gene a total of 2  $\mu\text{L}$  of a 40 ng/ $\mu\text{L}$  purified SeSaM\_step 3\_template from preliminary experiment II, 8  $\mu\text{L}$  of a 10 ng/ $\mu\text{L}$  purified SeSaM step 2 product and 10  $\mu\text{L}$  Nuclease-free H<sub>2</sub>O should be used.
- 4) Run all samples in a PCR cycler for 20 cycles and preheated lid to 105 °C as follows: 94 °C 120 sec (1 x); 94 °C 30 sec, 52 °C 1 min, 72 °C 1 min/kb of the SeSaM template-DNA (19 x); followed by a 4 °C hold to protect the samples.

- 5) When Dpo4 polymerase was employed the mastermix was pre-heated to 98 °C for 2 min and cooled down slowly to room temperature, followed by enzyme addition and incubation at 55 °C for 120 minutes. The resulting products were purified (NucleoTrap Extract II kit) in preparation for the final step of SeSaM.
- 6) Column-purify the elongated ssDNA fragments individually using the Nucleospin Extract II PCR Purification Kit and NTC buffer according to the manufacturer's recommendation. Elute DNA in 25 µL elution buffer or nuclease-free water.
- 7) Measure the concentration of dsDNA spectrophotometrically and store samples at -20 °C.

#### 2.14.7. SeSaM step 4: Replacement of degenerate base

In the last step of SeSaM, *Taq* polymerase is used in a nested PCR to specifically amplify mutated DNA as well as to exchange the previously introduced degenerate bases with standard nucleotides. The PCR products can be directly used for cloning in an expression vector.

- 1) Prepare the following pre-mastermix (400 µL) on ice in a 1.5 mL tube:

<b>Pre-mastermix (360 µL):</b>	<b>Volume:</b>
10x <i>Taq</i> DNA polymerase buffer	40 µL
SeSaM_fwd primer (5 µM)	40 µL
SeSaM_rev primer (5 µM)	40 µL
dNTP mix (10 mM)	8 µL
<i>Taq</i> DNA polymerase (5 U µL <sup>-1</sup> )	8 µL
Nuclease-free H <sub>2</sub> O	to a total volume of 360 µL

- 2) Divide the pre-mastermix into 4 aliquots of 90 µL in 0.2 mL PCR-tubes.
- 3) Add 50 ng per kb of the respective SeSaM step 3 products and adjust the final volume to 100 µL with nuclease-free water, if necessary.
- 4) Split each mastermix into 2x 50 µL in 0.2 mL PCR-tubes.
- 5) Run the reactions in the PCR cycler using the following program: 94 °C for 120 sec (1 x); 94 °C 30 sec, 60 °C 30 sec, 72 °C 45 sec/kb of the SeSaM template-DNA (20 x); 72 °C 3 min (1 x); followed by a 4 °C hold to protect the samples.

- 6) Combine each set of duplicate aliquots in one tube and run 5  $\mu$ L of each sample on a 0.8% agarose gel (7.5 V/cm; 40 min) to evaluate the quality of PCR amplification. PCR clean-up the products using Nucleospin PCR clean-up kit and NT buffer and elute in 25  $\mu$ L in elution buffer or nuclease-free water. If unspecific bands are observed, the specific product bands can be isolated in an agarose gel extraction step.
- 7) Store the completed SeSaM library at -20 °C or proceed to cloning in an expression vector using your preferred methods.



### 3. Results and discussion

This part is divided into 3 sections, each concluded by a short summary and discussion of the most significant findings. The identification of putative Y-polymerase genes by data mining and their subsequent cloning, expression, and preliminary biochemical characterization with focus on elongation of subsequent transversion mismatches is described in section 3.1. The work relevant to development of a suitable high-throughput screening system is summarized in section 3.2. In section 3.3, the application of the developed screening system in protein engineering of the Y-polymerase Dpo4 from *Sulfolobus sulfataricus* is presented. From this engineering campaign, a Dpo4 mutant with an increased transversion mismatch elongation capability was identified. The application of this mutant in the preparation of a model SeSaM library and the incurred implications on the quality of the generated library are also discussed in section 3.3.

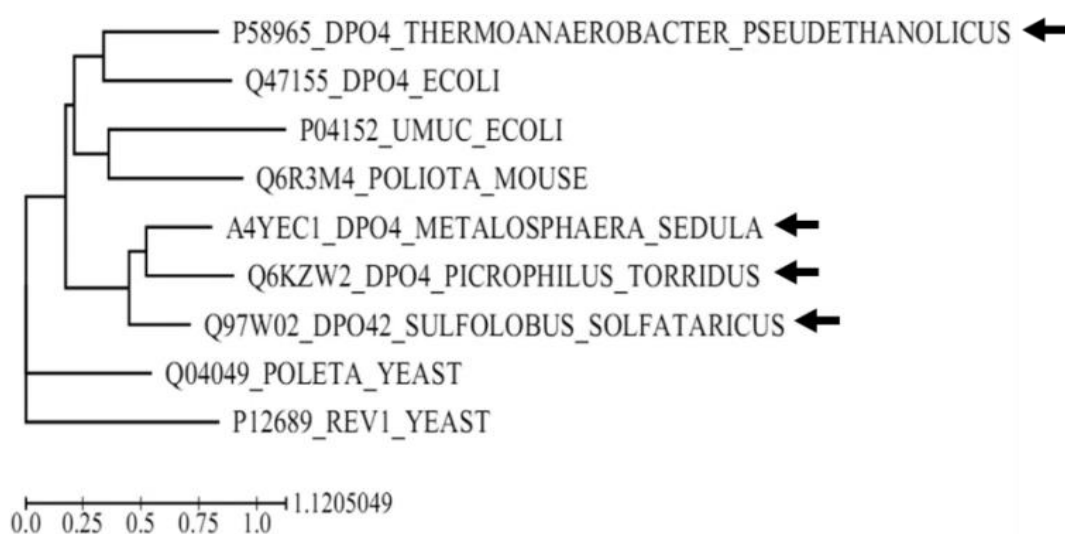
#### 3.1. Identification and preliminary characterization of Y-family polymerases with suitability for SeSaM

This section presents the identification of putative Y-polymerase genes by data mining, cloning, expression, and preliminary biochemical characterization with focus on elongation of subsequent transversion mismatches. At the end of the section, the reasons for selecting Dpo4 polymerase from *Sulfolobus sulfataricus* for the follow-up work are presented.

##### 3.1.1. Identification of Y-family polymerases by database mining

Dpo4 from *Sulfolobus sulfataricus* P2 has been extensively described in literature, both structurally and functionally (Boudsocq et al., 2001; Eoff et al., 2007; Wu et al., 2011; Zhang et al., 2009). This polymerase can be regarded as an archetype of the Y-family of DNA polymerases (Eoff et al., 2009). In search of proteins with similar properties, a BLASTp search (<http://blast.ncbi.nlm.nih.gov>) against amino acid sequence of Dpo4 (UniProt Accession number Q97W02) was carried out. The query identified more than 250 putative polymerase sequences exhibiting at least 30 % sequence homology to the query. Two of highest scoring sequences from the thermophilic archaeal species *Metallosphaera sedula* DSM 5348 (58 % identity) and *Picrophilus torridus* DSM 9790 (48 % identity) together with a low scoring sequence from *Thermoanaerobacter pseudethanolicus* ATCC 33223 (37 % identity) as well as

that Dpo4 from *Sulfolobus sulfataricus* were chosen for wet lab investigations. At the time of selection, none of these sequences was covered by patents which made them suitable for a copy protection with regard to an envisioned commercial application of a SeSaM mutagenesis kit. The cDNA for each amino acid sequence was designed to contain flanking NdeI and XhoI restriction sites for cloning. A codon optimization of the gene sequences for expression in *E. coli* was also performed. The synthesis of the resulting DNA sequences was outsourced to a specialized provider.



**Figure 10.** Phylogenetic tree based on amino acid sequence data, showing the estimated relationship between a selection of Y-family polymerases and the ones chosen for preliminary biochemical investigations (marked with an arrow). The tree was calculated using the BLASTp algorithm.

### 3.1.2. Cloning of synthetic genes in expression vectors

The putative polymerase gene sequences were obtained as synthetic genes from a specialized provider in pMA synthetic vectors. Each of 4 genes was re-cloned in 4 different vectors suitable for expression in *E. coli* via restriction cloning. The synthetic vectors were digested with NdeI and NotI restriction enzymes to cut out the encoding DNA fragment. The digested gene fragments were purified from agarose gel, then cloned in vectors pET28a(+), pBADN, pALXtreme-1a and pALXtreme-5b (derivatives of pET28a(+) and pET22b(+) vectors, respectively). pET28/22 and pALXtreme-1a/5b vectors harbor a T7 promoter-based expression system. The control of expression from pBADN is regulated by *araBAD* promoter.

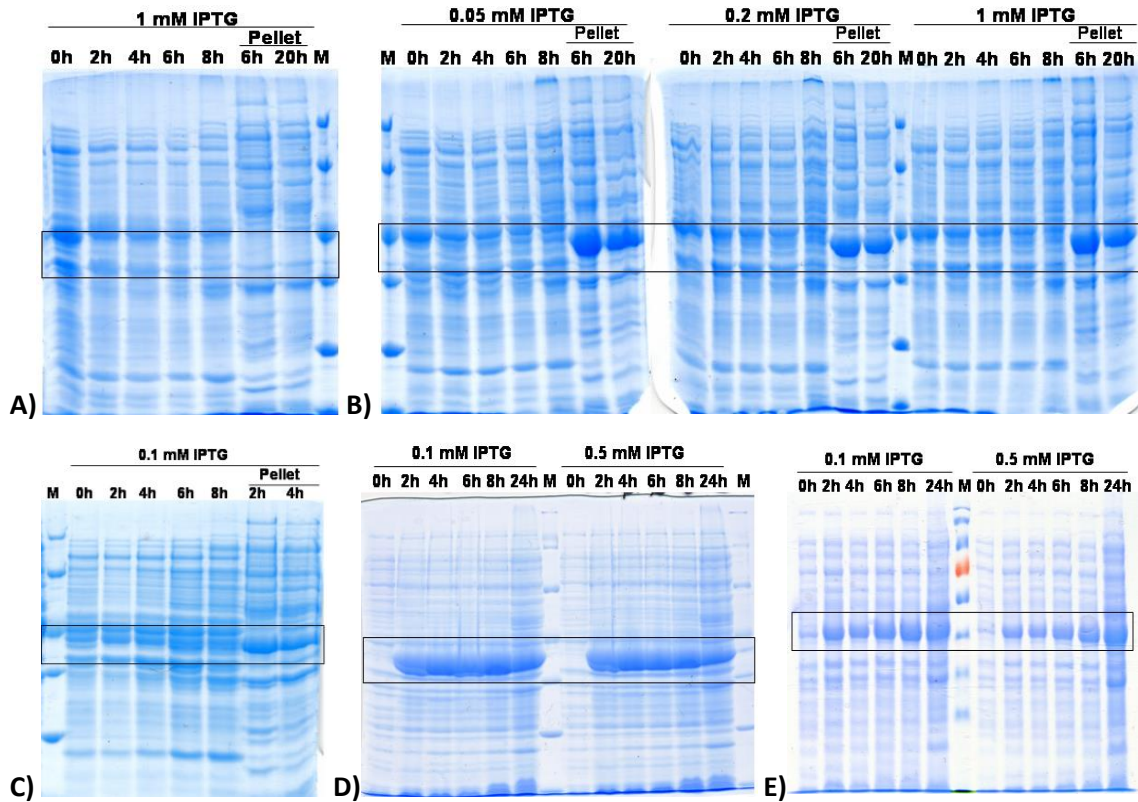
pET vectors were used in pilot expression and purification studies. In the directed evolution experiments of the most suitable Y-polymerase, pALXtreme vectors were employed due to their small size which facilitated DNA manipulations, e.g., for cloning of random mutant libraries and site-saturation mutagenesis PCR. Cloning of genes was verified by colony PCR using vector specific primers and after plasmid isolation, analytical digestion from least 2 positive colonies was performed as a second confirmation. An exemplary gel of the double digestion to verify cloning in pBAD vectors is presented in Fig. 11. In this figure, the correct insertion of the genes is marked with an arrow at ~1.1 kb corresponding to the average size of the investigated genes. Important to mention is that cloning in pET28a(+) and pALXtreme-1a was done in frame with a N-terminal histidine-tag which facilitated protein purification.



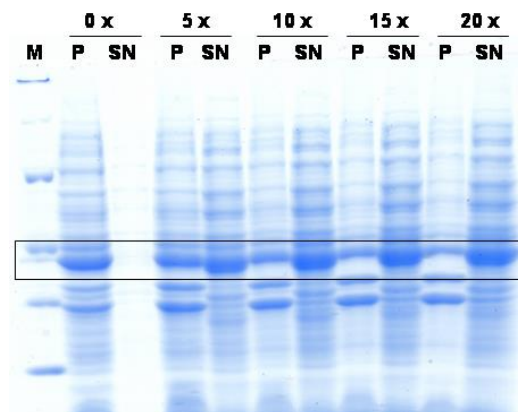
**Figure 11.** Cloning of the synthetic polymerase genes in pBADN expression vector. An exemplary gel showing the result of analytical digestion of pBADN vectors harbouring polymerase genes. For each pBADN construct two positive clones after colony PCR were grown in liquid media and the contained plasmids were isolated and digested with *NdeI* and *XhoI* restriction enzymes. As positive control (marked “+”) in this experiment, previously prepared and verified by sequencing pALXtreme-5b vectors with the correct inserts were also digested and loaded on the gel. The digestion of pBADN and pALXtreme-5b by *NdeI* and *XhoI* exhibited different restriction patterns due to difference in size of both vector backbones; however, the contained inserts were identical. The appearance of the highlighted specific bands at ~1.1 kb signified a successful cloning procedure. The analytical digestion was carried out for a short period of time and was not driven to conclusion, thus the presence of additional bands on the gel corresponding to undigested, circular vector.

### 3.1.3. Preliminary expression studies

Preliminary expression studies of the putative polymerases were carried out in *E. coli* BL21 Gold (DE3) - pET28a and *E. coli* BL21 Gold (DE3)  $\text{lacI}^{\text{q1}}$  - pALXtreme-1a expression systems using TB growth medium. Expression was carried out in 500 mL flasks containing 50 mL media inoculated with saturated pre-culture such that the starting  $\text{OD}_{600}$  of the main cultures was  $\sim 0.1$ . Cells were grown at 37 °C until  $\text{OD}_{600}$  reached  $\sim 0.6$  then induced with different concentrations of IPTG. Samples were collected at regular intervals over a period of 24 h. At harvesting of Dpo4, Tps and Mse cultures,  $\text{OD}_{600}$  of  $\sim 12$  and wet cell weight from 50 mL culture of  $\sim 2$  grams were reached. Pto expression yielded  $\text{OD}_{600}$  of 5 and wet cell weight from 50 mL culture of  $\sim 0.9$  gram. In the case of Dpo4 from *Sulfolobus sulfataricus* and Mse from *Metallospheara sedula*, pellets were incubated with B-PER lysis reagent, according to manufacturer's recommendation to separate cytosolic proteins from cell debris and inclusion bodies, then both fractions were loaded on denaturing polyacrylamide gels (Fig. 12B and 12C). Expression of Pto from *Picrophilus torridus* and Tps from *Thermoanaerobacter pseudethanolicus* was checked by preparing whole cell samples for SDS-PAGE analysis (Fig. 12D and 12E). Fig. 12 showed that overexpression was achieved for all four polymerases, however the experiment carried out with B-PER lysis reagent suggested that proteins were expressed in insoluble form as the expected protein band was found in pellet fractions. Subsequent experiments using sonication as a lysis method showed that the enzymes were solubly expressed and that the originally employed B-PER lysis was inefficient (Fig. 13). An additional benefit of using sonication was that lysis by ultrasound sheared genomic DNA (gDNA). The fractionation of gDNA ensured that in follow-up work, the interference of gDNA which can potentially serve to template unspecific DNA amplification in PCR or affect the interactions between protein and purification matrix, could be minimized.



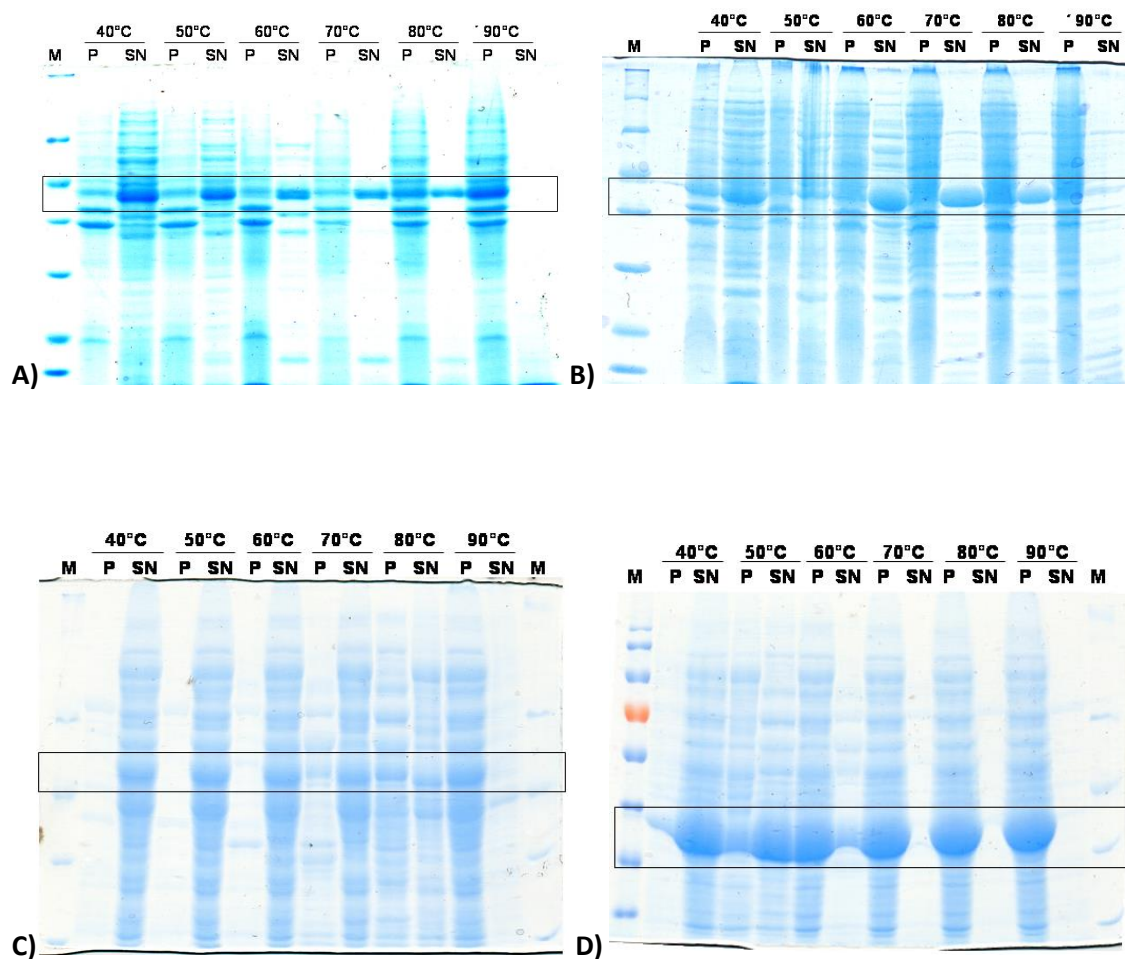
**Figure 12.** Expression profiles of Y-polymerases in *E. coli* BL21 Gold(DE3) pET28a expression system and TB media. **A)** empty vector; **B)** Dpo4 from *Sulfolobus sulfataricus*; **C)** Mse from *Metallospheara sedula*; **D)** Pto from *Pictophilus torridus*; **E)** Tps from *Thermoanaerobacter pseudethanolicus*. In the case of Dpo4 and Mse, cell lysis was done by treatment of whole cells with B-PER reagent, expression cultures of Pto and Tps were lysed by sonication. Soluble and insoluble (pellet) fractions were loaded on each gel. Expected protein sizes were in the range of 40 kDa and the corresponding region on each gel is highlighted.



**Figure 13.** Utilization of sonication as lysis method for lysis of *E. coli* expressing Dpo4 polymerase. The assay was performed by preparing 30 % cell suspension of expression culture and sonicating on ice for 20 min (1 min intervals) using 40 % amplitude. After every 5 sonication cycles samples were taken and insoluble fraction (marked P for pellet) were separated from the soluble fraction (marked SN for supernatant) and analyzed on a denaturing acrylamide gel. Expected protein size – ca. 40 kDa – is highlighted.

### 3.1.4. Thermostability studies with Y-family polymerases in crude cell extracts

One important criterion for using polymerases in *in vitro* DNA assays is stability at elevated temperatures ( $\geq 60$  °C). Elevated thermostability is also an important pre-requisite for application of Y-polymerases in SeSaM step 3 in which truncated gene fragments harboring multiple degenerate bases at their 3'-ends have to be accurately hybridized to a template and elongated to full length. This can only be achieved if the reaction mixture is iteratively brought to high temperature. In addition, the envisioned activity and screening assays for polymerases also required thermocycling conditions. The three putative Y-polymerases genes were chosen based on sequence similarities to the moderately thermostable Dpo4 from *Sulfolobus sulfataricus*. In addition, the selected genes originated from thermophilic microorganisms. These two criteria alone did not guaranteed that the resulting proteins would indeed be sufficiently thermostable for the envisioned application. Therefore, a simple experiment to evaluate the ability of the selected proteins to withstand elevated temperature was carried out. Clarified cell extracts were subjected to heat treatment at temperatures ranging from 40 °C to 90 °C for 10 minutes. Precipitated and soluble protein fractions were separated by centrifugation and separately loaded on SDS acrylamide gel. The resulting gel images are presented in Fig. 5. Dpo4 and Mse showed a similar degree of thermostability. These proteins remained soluble even after incubation at 80 °C for 10 minutes (Fig. 14A and 14B). Tps could withstand up to 60 °C but precipitated when exposed to higher temperatures (Fig. 14C). Pto exhibited much lower stability at elevated temperatures. The latter protein was found in the insoluble fraction after incubation at mild temperature (40 °C) and no soluble protein could be detected after heat treatment at 70 °C (Fig. 14D). Based on these observations it was concluded that Pto from *Picrophilus torridus* is not a suitable polymerase for use in SeSaM (unless it was first evolved for increased thermostability). Nevertheless, Pto was employed together with the other three enzymes in preliminary activity tests. Thermostability valuation also provided important hints on how to efficiently purify polymerases that exhibited resistance to treatment at elevated temperature. Simply by heating up crude cell extracts most *E. coli* proteins could be precipitated leaving only a fraction of *E. coli* native proteins to be removed by chromatographic methods.

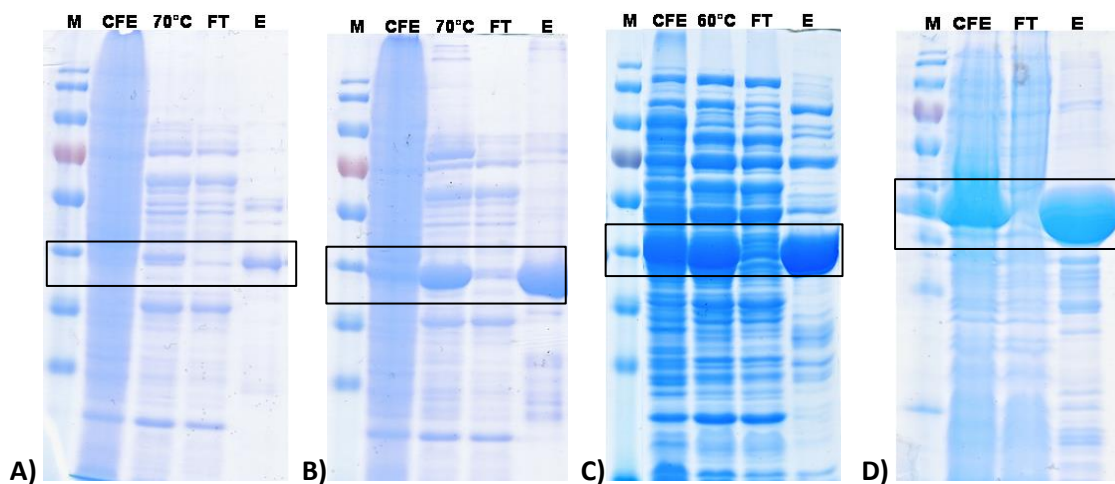


**Figure 14.** Thermostability studies for **A)** Dpo4; **B)** Mse; **C)** Tps; **D)** Pto. Expected Y-polymerase sizes of ca. 40 kDa are highlighted on each gel picture. Soluble fractions are marked with SN and insoluble fractions with P.

### 3.1.5. Purification of Y-family polymerases

The cloning of Y-family polymerases in pET28a(+) and pALXtreme-1a using a NdeI restriction site allowed protein expression in frame with a N-terminal His-tag and, consequently, purification by affinity chromatography. Expression of the four polymerases was carried out as outlined in the previous paragraph. Crude extracts containing Dpo4, Mse and Tps were heat treated at 80 °C (60 °C for Tps) for 10 min prior to applying the clarified cell extract on Protino Ni-IDA pre-packed columns. Cell extract of Pto were loaded on the Ni-NTA column without heat denaturation of *E. coli* proteins. Column washing and elution was conducted as described in “Materials and Methods” section. The resulting gels containing samples at different stages of purification are presented in Fig. 15. The fraction of target protein in the purified samples varied from ~60 % (for Dpo4) to ~80 % (for Pto) and protein concentration

reached up to 5.5  $\mu\text{g}/\text{mL}$  (for Pto). These purity levels were sufficient to conduct preliminary activity assays with purified enzymes. At a later stage, the purification protocol for Dpo4 was refined by supplying more stringent heat denaturation, washing and sample elution (see section 3.3).



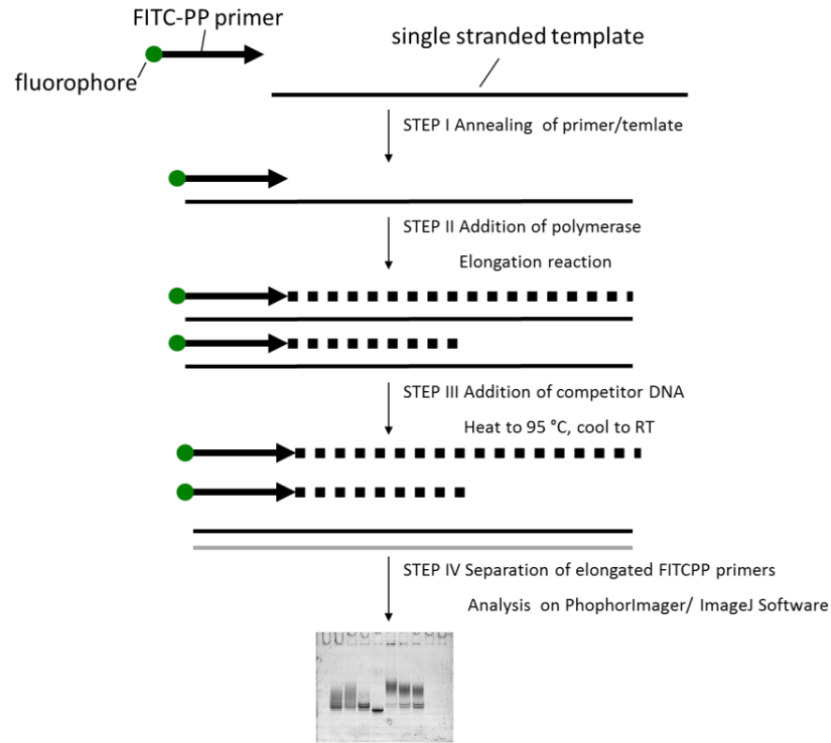
**Figure 15.** SDS-PAGE gel images after pilot purification of **A)** Dpo4; **B)** Mse; **C)** Tps; **D)** Pto. Whole cell extracts (CFE) were clarified and heat treated at 80 °C (or 70 °C) for 10 minutes. Heat denaturation step was omitted in the case of Pto. Soluble fractions were applied to pre-packed Ni-IDA column and washed extensively with LEW buffer supplied with the Protino Ni-IDA purification kit. Flow-through fractions (FT) were also loaded on the gel. Protein of interest was eluted from the column by addition of elution buffer supplied with the Protino Ni-IDA purification kit. The proteins of interest (highlighted in a rectangular box) could only be detected in the elution fractions (marked as E).

### 3.1.6. Forward primer elongation assay with purified Y-polymerases

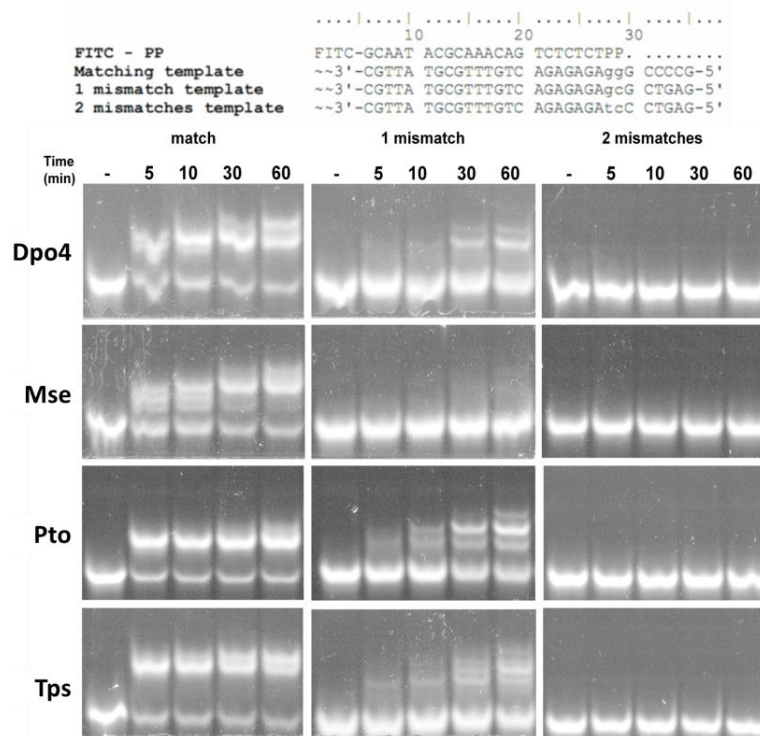
Forward primer elongation assay has been commonly used as an activity assay to probe the specific activity of polymerases with non-canonical primer-template pairs. This assay is especially useful for characterization of distributive (i.e., low-processive) polymerases (Avkin et al., 2004; McCulloch and Kunkel, 2008; Prakash et al., 2005). The assay procedure comprises DNA polymerase-catalysed elongation of a fluorescently labeled primer on a short single-stranded template, separation of DNA products on an acrylamide gel and visualization of the fluorescent probe on a fluorescence imager (Fig. 16A). In this study, forward primer elongation assay was used to evaluate the ability of the purified Y-polymerases to elongate primer-template pairs forming one or multiple transversion mismatches at 3'-end of primers. In the employed setup, the fluorescently-labeled primer oligo was designed to contain two degenerate pyrimidine base analogs (dPTP; pairs with A and G) attached at 3'-



end of the primer. This nucleotide was used because it is routinely employed in the preparation of transversion-enriched SeSaM mutant libraries. Templates forming a match, a single or double mismatch at the 3'-end of the primer provided the possibility to probe the capability of the investigated polymerases to "read" mismatches of increasing complexity (no mismatch vs a single transversion mismatch vs a double transversion mismatch). Relative mismatch elongation efficiency of Dpo4, Mse, Tps and Pto was estimated after incubating pre-hybridized primer-template pairs with equal amount enzyme at 37 °C for 60 min (Fig. 16B). This preliminary test showed that all four enzymes could efficiently elongate a dPTP-tailed primer when a matching template was supplied. Elongation of a single transversion mismatch could be carried out with low efficiency by Dpo4, Pto and Tps but hardly any elongation could be catalyzed by Mse under the aforementioned reaction conditions. Two consecutive transversion mismatches could not be elongated by any of the Y-polymerases.



A)



B)

**Figure 16. A)** A schematic outlining forward primer elongation assay procedures; **B)** Monitoring of the extension of FITC-labeled primer with two 3'-attached dPTP bases by four Y-family polymerases. FITC-labeled products separated on acrylamide gel and visualized on phosphoroimager (ex. 494 nm/em. 518 nm). Each band represents an addition of nucleotide(s) to the FITC labeled oligonucleotide. In the lanes labeled “-” (negative control), reaction mastermix subjected to incubation without addition of enzyme is loaded as a reference sample. See Box 2 for further explanations on how to interpret results.

### **3.1.7. Summary: Identification and preliminary characterization of Y-family polymerases with suitability for SeSaM**

To resolve the bottleneck in step 3 of the SeSaM method by achieving more efficient elongation of consecutive nucleotide mismatches at 3'-end of primers, 4 genes (3 putative + 1 described) encoding Y-family DNA polymerases were selected for preliminary studies. Members of this polymerase family are exclusively involved DNA synthesis past damaged DNA (Ohmori et al., 2001) and were, therefore, perceived as suitable starting point for this study. All four genes were successfully cloned, expressed and purified in preparation for preliminary characterization. Activity assay with purified enzyme confirmed that the genes identified from genetic databases were correctly assigned and indeed encoded DNA polymerases. The results from the biochemical and thermostability tests were useful for ranking of the polymerases in terms of suitability for application in SeSaM step 3. After the thermostability tests, Pto polymerases was excluded due to its low stability at elevated temperatures. After testing for the desired mismatch elongation activity, the least efficient polymerase (Mse) which under the applied conditions was not able to elongate efficiently even a single transversion mismatch was also eliminated from the selection. Moreover, the activity assay showed that none of the polymerases is capable of efficiently elongating a consecutive transversion mismatch, reinforcing that protein engineering of polymerases would be necessary in order to improve the efficiency of SeSaM step 3 and further advance the SeSaM method. Taking into account that valuable biochemical and structural data for Dpo4 was readily available in literature, Dpo4 polymerase from *Sulfolobus solfataricus* was to initiate a protein evolution campaign aiming at identifying a polymerase variant with improved mismatch elongation efficiency.

### **3.2. Development of a screening system for high-throughput screening of Y-family polymerases**

In order to engineer Dpo4, a high-throughput screening system that reflects well the envisioned application in SeSaM is indispensable. The development of such screening system is described in this section. The activity assay (i.e., forward primer elongation assay) used for the preliminary characterization of Y-family polymerases, although reliable, lacked the required throughput to screen large mutant libraries and identify a mutant with improved

performance. Finding an alternative screening approach with a high-throughput for testing consecutive transversion mismatch elongation efficiency was therefore crucial for the successful outcome of the entire project. Three approaches to the challenge were tested: 1) compartmentalized self-replication (CSR), 2) a screening approach based on dsDNA quantification using a double stranded DNA binding dye and 3) a duplex DNA probe used for real time quantitative PCR (termed Scorpion probe). The duplex Scorpion probe-based screening system in 96-well plate format for directed evolution of polymerases was eventually selected and optimized for screening applications.

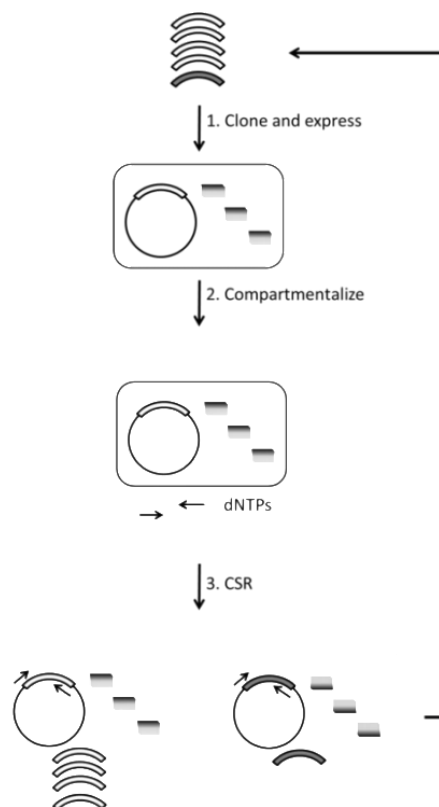
### **3.2.1. Emulsion-based screening system for directed evolution of the Y-polymerase**

At first, an emulsion-based selection system termed compartmentalized self-replication (CSR) was employed with the Y-family polymerase Dpo4. CSR has been successfully used to evolve replicative polymerases (Ghadessy and Holliger, 2007; Tubeleviciute and Skirgaila, 2010), (Ghadessy et al., 2001). CSR uses a simple feedback loop involving a polymerase that self-replicates only its own encoding gene (Fig. 17). Self-replication occurs in separate, non-interacting compartments formed by a heat-stable water-in-oil emulsion. Compartmentalization mediates the linkage of phenotype and genotype by ensuring that each polymerase replicates only its own encoding gene. Single cells and other components necessary for PCR reactions (DNA primers, dNTPs,  $Mg^{2+}$ , etc.) are compartmentalized individually in thermostable water-in-oil emulsions. Individual compartments contain one cell and consequently a single Y-polymerase variant. The emulsions are incubated under PCR conditions and only active Y-polymerase mutants can form a PCR product. The supplied primer pairs are designed such that Y-polymerase gene which encode active Y-polymerase mutants are exponentially amplified. In addition, the primers used form a transversion mismatch between two degenerate base analogs (dPTPs) at the 3'-ends of primers and two pyrimidine bases in the template strand in accordance with the requirements of the SeSaM technology. The attempts to adapt CSR to screening of mutant libraries of Dpo4 polymerase proved unsuccessful. Consequently, only the main results are discussed in short.

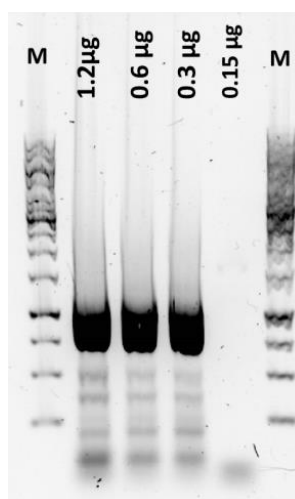
In the early experiments, CSR was attempted by emulsifying cells expressing Dpo4 polymerase, however polymerase expression level was insufficient for gene self-

amplification to take place. Even though Dpo4 is known to amplify long DNA fragments when sufficiently high enzyme concentration is provided in the mastermix (Boudsocq et al., 2001), it remains a distributive polymerase largely limited in its processivity. The achieved Dpo4 expression level from a single cell in emulsions (an estimated  $0.06 \text{ ng } \mu\text{L}^{-1}$ , under optimized expression conditions) corresponded to only 0.6 % of the enzyme amount ( $\sim 10 \text{ ng } \mu\text{L}^{-1}$  mastermix) needed for PCR amplification (Fig. 18). The addition of a small amount *Taq* polymerase to the reaction mix (in order facilitate DNA amplification after Dpo4-catalysed elongation of the transversion mismatch) also did not lead to substantial increase in product formation (data not shown). As a next step, an *in vitro* Dpo4 expression coupled to PCR amplification in emulsions was attempted. The *in vitro* expression of Dpo4 was successful as confirmed via forward primer elongation assay (Fig. 19A); however, the *in vitro* expression level of Dpo4 was not sufficient to generate full-length products by PCR in tubes or PCR in emulsions as illustrated in Fig. 19B.

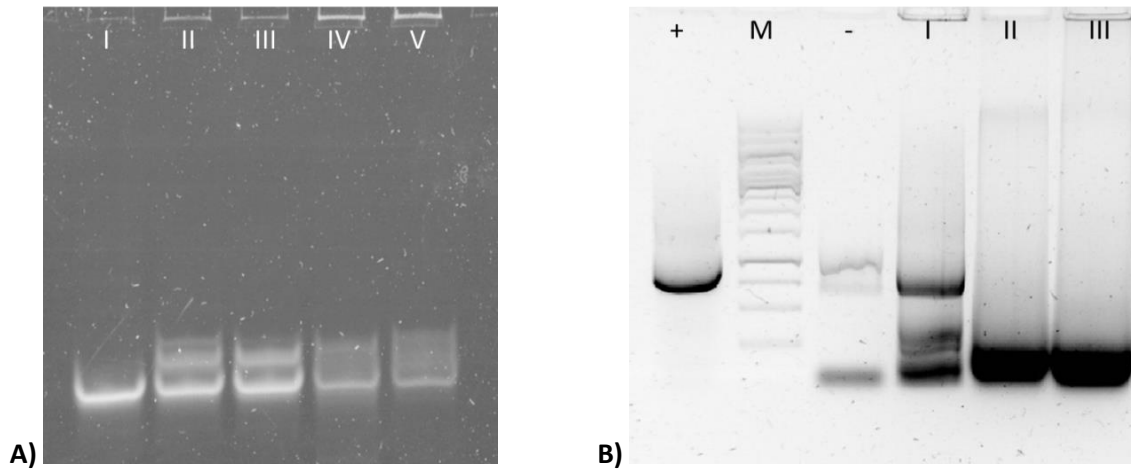
DNA-binding domain, Sso7d, can be fused to polymerases in order to increase their processivity of polymerases (Wang et al., 2004). These fusion polymerases are characterized by very high processivity and are among the most efficient, commercially available DNA polymerases (e.g., Phusion<sup>®</sup> polymerase from New England Biolabs). This strategy was applied in order to increase the processivity of Dpo4 and decrease the necessary enzyme load per emulsion compartment. Therefore, Dpo4 was genetically fused to Sso7d and the resulting construct was expressed and purified (Fig. 20A). A simple PCR-based experiment was carried out to check whether the fusion polymerase, Dpo4S, exhibited increased processivity. DNA sequences of different length (0.2 to 2.5 kb) were amplified with varied amount of purified enzyme (30 to 240 ng per 25  $\mu\text{L}$  reaction). Amplification by Dpo4S of the 0.2 kb target sequence could be observed only at an enzyme dose of  $\sim 10 \text{ ng } \mu\text{L}^{-1}$  (Fig. 20B). Therefore, Sso7d domain did not affect positively the processivity of Dpo4 and the approach was not pursued any further.



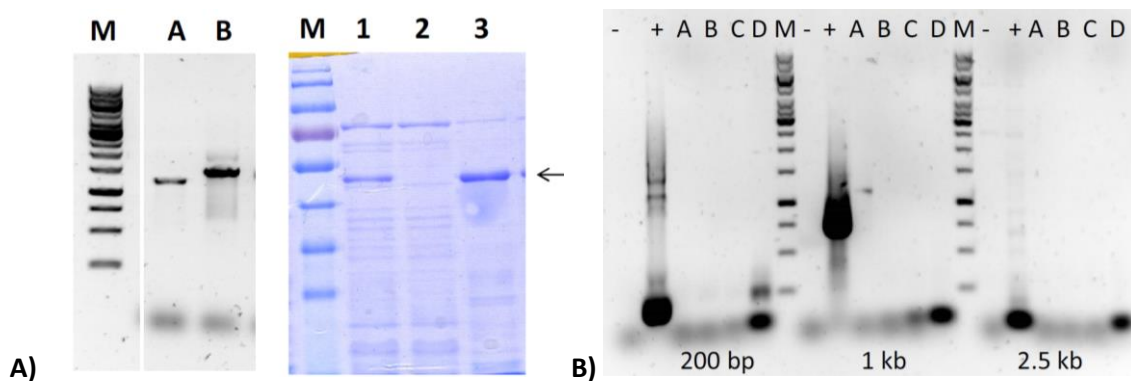
**Figure 17.** Schematic of the compartmentalized self-replication selection system. A library of polymerase genes is cloned and expressed in *E. coli*. Spheres represent active polymerase molecules. Bacterial cells containing the polymerase and encoding gene are suspended in reaction buffer containing flanking primers and dNTPs and segregated into aqueous compartments. The polymerase enzyme and encoding gene are released from the cell, allowing self-replication to proceed. Poorly active polymerases (white hexagons) fail to replicate their encoding gene. The offspring polymerase genes are released and re-cloned for another cycle of CSR (adapted from Ghadessy et al., 2001).



**Figure 18.** PCR amplification with varied concentration of purified Dpo4 polymerase in order to determine the minimal Y-polymerase concentration necessary for amplification of a 1 kb DNA fragment. Legend: M, marker; 1.2 µg Dpo4/25 µl mastermix; 0.6 µg Dpo4/25 µl mastermix; 0.3 µg Dpo4/25 µl mastermix; 0.15 µg Dpo4/25 µl mastermix.



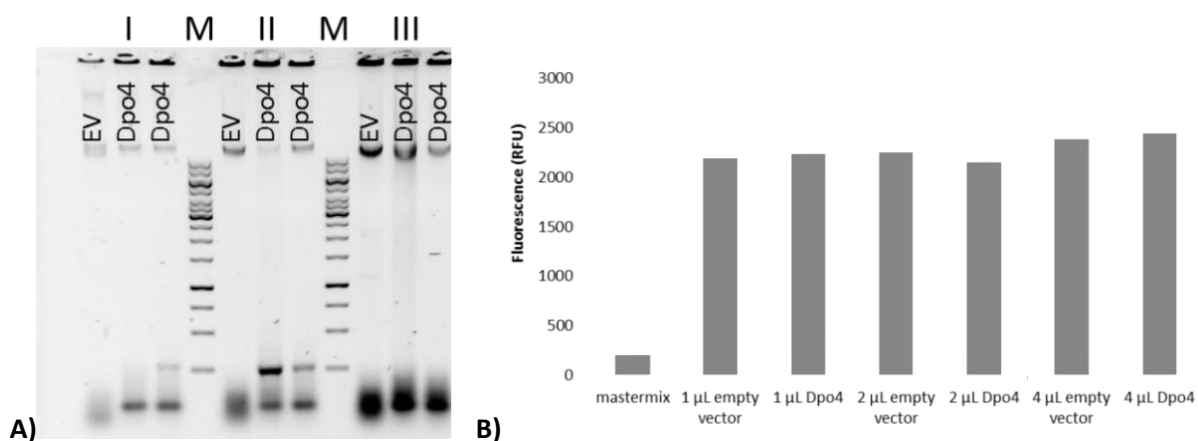
**Figure 19. A)** *In vitro* expression of Dpo4 coupled to forward primer elongation assay in emulsions. Legend: I - negative control; II – non-emulsified purified Dpo4; III – emulsified purified Dpo4; IV – non-emulsified *in vitro*/reaction mastermix; V – emulsified *in vitro*/reaction mastermix; **B)** Comparison of the activity (no mismatch supplied) of purified and *in vitro* expressed Dpo4 enzyme for PCR reactions in emulsions. The positive control *Taq* polymerase (+) shows the target band of an ordinary PCR which could also be obtained with purified Dpo4 (I). The negative control without enzyme showed no target band (-). Dpo4 enzyme from *in vitro* expressions is not visible by ordinary target amplification PCR (II) or in the emulsion PCR sample (III).



**Figure 20. A)** Cloning and purification of Dpo4S; Left) a DNA gel showing the linearized genetic construct is presented, M - DNA marker, A – Dpo4, B - Dpo4S; Right) an image of the polyacrylamide gel after expression of Dpo4S is shown, M - protein standard, 1 - Soluble cell-free extract after heat precipitation 80 ° C, 2 – flow through after HisTag purification, 3 - eluate with Dpo4S. The arrow indicates the purified fusion protein Dpo4S; **B)** Result of the amplification of sequences of varied length (0.2 kb, 1 kb and 2.5 kb) using 30 ng (A) 60 ng (B), 120 ng (C) and 240 ng (D) Dpo4S in 25  $\mu$ l reaction volume. Also shown are DNA marker (M) and a positive control with *Taq* polymerase (+) and negative control without added enzyme (-).

### 3.2.2. PicoGreen® double stranded DNA (dsDNA) dye-based screening system

One common way of quantifying DNA accumulation (e.g., in qPCR) is by the use of dsDNA intercalating dyes such as PicoGreen® (Dragan et al., 2010; Holden et al., 2009; Singer et al., 1997). These dyes interact specifically with dsDNA and upon intercalation become fluorescent. The intensity of fluorescent signal is proportionate to the amount of DNA formed. This property can be used to measure the degree to which DNA polymerases are able to generate dsDNA product in a defined time span. In the envisioned screening system setup, Dpo4 polymerase was expressed in microtiter plates (MTPs), then cells were lysed and the cell supernatant was used to amplify short DNA fragment (200 bp long). The amount of DNA formed after thermocycling was detected via PicoGreen® dsDNA dye as recommended by the manufacturer. Preliminary studies in which a short template was amplified with primers forming no mismatch showed that even though DNA product could be generated and visualized on agarose gel, the background fluorescence was too high for reliable dsDNA quantification. The high background was due to the fact that crude cell extracts were used, therefore *E. coli*-derived dsDNA was also present in the cell supernatant (Fig. 21). This idea was abandoned in favor of a more specific, fluorescence-based probe for detection of polymerase activity which is described next.

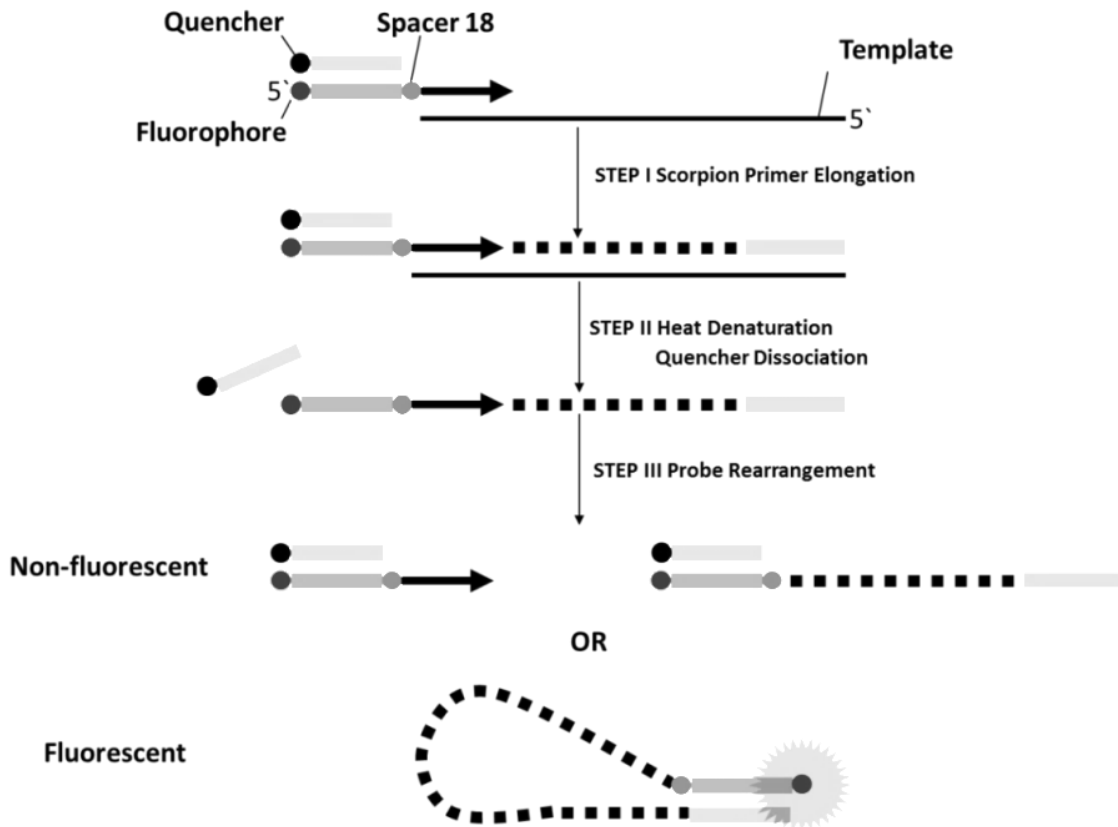


**Figure 21. A)** Agarose gel visualization of 200 bp product produced with 1 μL (I), 2 μL (II), or 4 μL (III) of concentrated cell lysate containing Dpo4 or expressing no polymerase (EV); **B)** PicoGreen® dye quantification of PCR products produced with 1 μL, 2 μL or 4 μL of cell lysate containing Dpo4. The values for lysate of cells harboring empty vector is an average of 3 measurements and the values for lysates of cells harboring Dpo4 encoding gene is an average of 6 measurements.



### 3.2.3. Establishment of a Scorpion probe-based screening system for Y-family polymerases

Earlier attempts to establish a screening system, even though unsuccessful, showed that an ideal screening system for directed evolution of Y-polymerases comprises of a highly specific, sensitive, closed-tube, robust assay that allows for continuous monitoring of activity that is capable of detecting enzymatic activity independent of polymerases' processivity, thermostability, displacement and exonuclease activity. Bearing these pre-requisites in mind, a fluorescent probe for polymerases originally employed in genotypic assays (Carters et al., 2008), single nucleotide polymorphisms or allelic discrimination (Solinas et al., 2001) was finally selected as a suitable reporter for polymerase activity. Scorpion probes are bi-functional molecules in which a primer oligo is linked in a flexible manner to a probing oligo. The probe element which is covalently attached a fluorophore can hybridize to a specific DNA sequence generated after elongation of the primer. If the primer is not elongated, the fluorescence is quenched by a separate quencher-bound oligo that pairs with the probe element. In the presence of the intramolecular target sequence which appears when the primer element is elongated, the fluorophore and the quencher separate and the elongated Scorpion rearranges which leads to an increase in the fluorescence signal as outlined in Fig. 22. The kinetics of such a rearrangement mechanism are very rapid, thus not a rate-limiting step. The specificity of binding of the probe element to its target (or quencher, respectively) ensures reliability of probing and high signal-to-noise ratio. Additionally, the stoichiometry of the reaction (one amplicon leads to one fluorescent signal) and concentration independency of probing provides reliable quantification and enhances overall sensitivity (Carters et al., 2008).

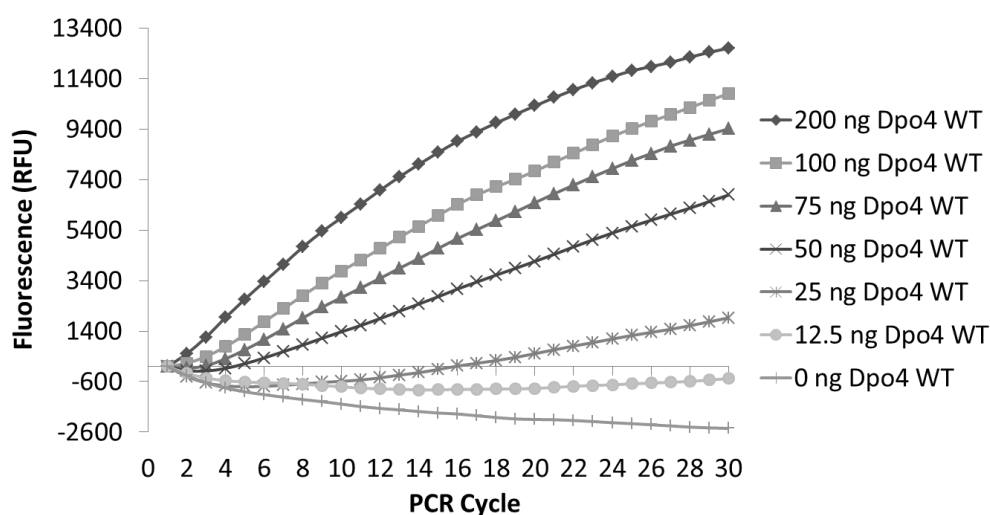


**Figure 22.** Principle of action of the Scorpion probe assay adapted for detection of DNA polymerase activity. A Scorpion probe comprises a primer covalently bound through a spacer molecule to a probe oligo with a fluorophore attached at its 5'-end. A separate quencher molecule consists of a quencher dye attached to an oligonucleotide complementary to the probe element. Prior to primer elongation on a synthetic, single stranded template designed to contain a specific sequence at its 5'-end which upon replication results in a region complementary to the probe sequence, the probing element is attached to the quencher, thus fluorescence levels are kept low. After extension of the primer element, the probe element is on the same strand as its complementary target and the probe can bind to its complement by a rapid intramolecular rearrangement. As a result, the fluorophore becomes unquenched which leads to an increase in specific fluorescence.

The probe design and reaction conditions were selected in accordance with the corresponding protocol published in the book „Molecular Beacons: Signalling Nucleic Acid Probes, Methods, and Protocols“ (Carters et al., 2008) with minor modifications. To adapt the probe for application in enzyme screening, a synthetic, single-stranded DNA template was employed such that a 73-mer oligo harbouring the target sequence at its 3'-end is elongated. Such an amplicon length ensured sufficient flexibility for molecular rearrangement to take place and short enough to minimize formation of non-specific but stable secondary DNA structures. Further flexibility is achieved through the incorporation of

a spacer molecule between the primer and probe element and ensures that the probe element is not incorporated in PCR product.

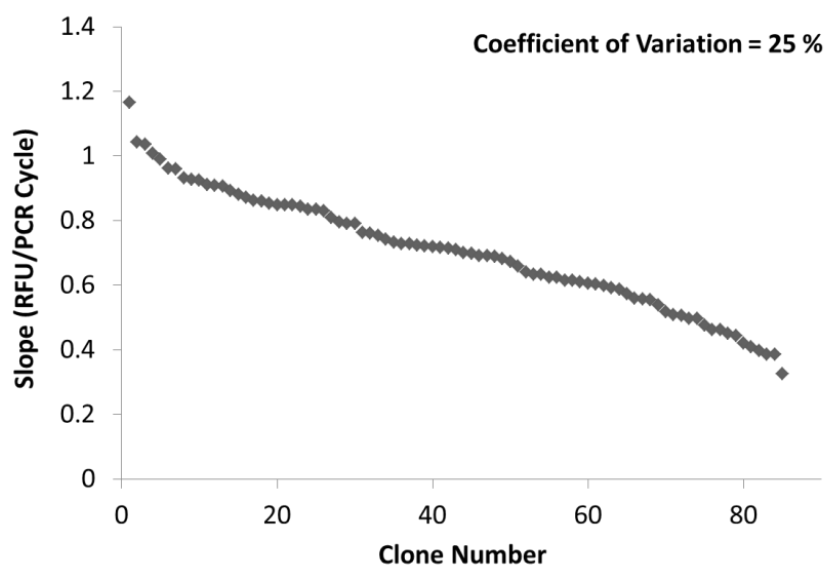
The Scorpion probe screening system was optimized for subsequent transversion mismatch elongation by first employing varying concentrations of purified Dpo4 wild type (WT) in order to determine the assay's working range and compensate for the inherent fluorescence loss at low protein concentrations (Fig. 23). From this experiment it was deduced that ~200 ng of Dpo4 is the lower detection limit for reliable determination of Dpo4 activity. In the latter case, a linear response was achieved from the first to at least the 15th thermal cycle.



**Figure 23.** Dpo4 polymerase concentration (per 25  $\mu$ L) in response of Scorpion probe fluorescence; 50 ng Dpo4 are sufficient for detection of polymerase activity; in the screening of mutant libraries ~200 ng Dpo4 were employed.

The Scorpion screening system was parallelized in 96-well microtiter plates (MTP) using crude cell extracts of *E.coli* expressing Dpo4 WT. Cell lysates were prepared such that the pre-determined amount of polymerase (accounting for roughly 20 % of total protein in the soluble fraction) could be supplied after cell lysis in 96-well MTP format and resulted in a standard deviation of 25 % (Fig. 24). Deviations of ~20 % have been reported in successful directed evolution campaigns (Lan Tee and Schwaneberg, 2007). Stringent selection conditions are important prerequisites to identify improved mutants; therefore only hits

with at least a 2- fold relative increase in activity were considered for re-screening and characterization studies.



**Figure 24.** Activity values in descending order of Dpo4 catalyzed transversion mis-match elongation in a 96-well microtiter plate using optimized Scorpion probe screening system. The measured coefficient of variation was 25 %.

### 3.2.4. Summary: Development of a screening system for high-throughput screening of Y-family polymerases

In summary, three different strategies to establishing a screening system for Y-family polymerases and Dpo4 from *Sulfolobus solfataricus* in particular have been pursued. The application of compartmentalized self-replication (CSR) approach were not suitable for Y-family polymerases due to inherent low processivity and, therefore, inability to amplify long DNA fragments at low enzyme load and self-amplify its encoding gene. Thus, despite the big advantages with regard to throughput, the idea of using CSR as a selection system for Dpo4 variants capable of elongating consecutive transversion mismatches was abandoned and other microtiter plate-based screening systems were investigated. The use of dsDNA binding dye which was attempted as a second approach was also unsuccessful. This was due to the high background fluorescence as a result of genomic and plasmid DNA added with crude cell extracts. The third strategy to adapt a fluorescent probe for real-time PCR to screening of

polymerase libraries was successful. From the plethora of RT-PCR probes available, Scorpion probe was identified to be suitable for the envisioned application, also with regard to the peculiarities of Y-family polymerases (i.e., low processivity, no exonuclease and displacement activity). The Scorpion probe could be adapted to screening of polymerase mutant libraries. After some optimization, good enzyme concentration dependent response and an acceptable coefficient of variation were achieved. The system was further validated by screening of mutant libraries of Dpo4 polymerase. The screening campaign and characterization of hits from mutant libraries is described in section 3.3.

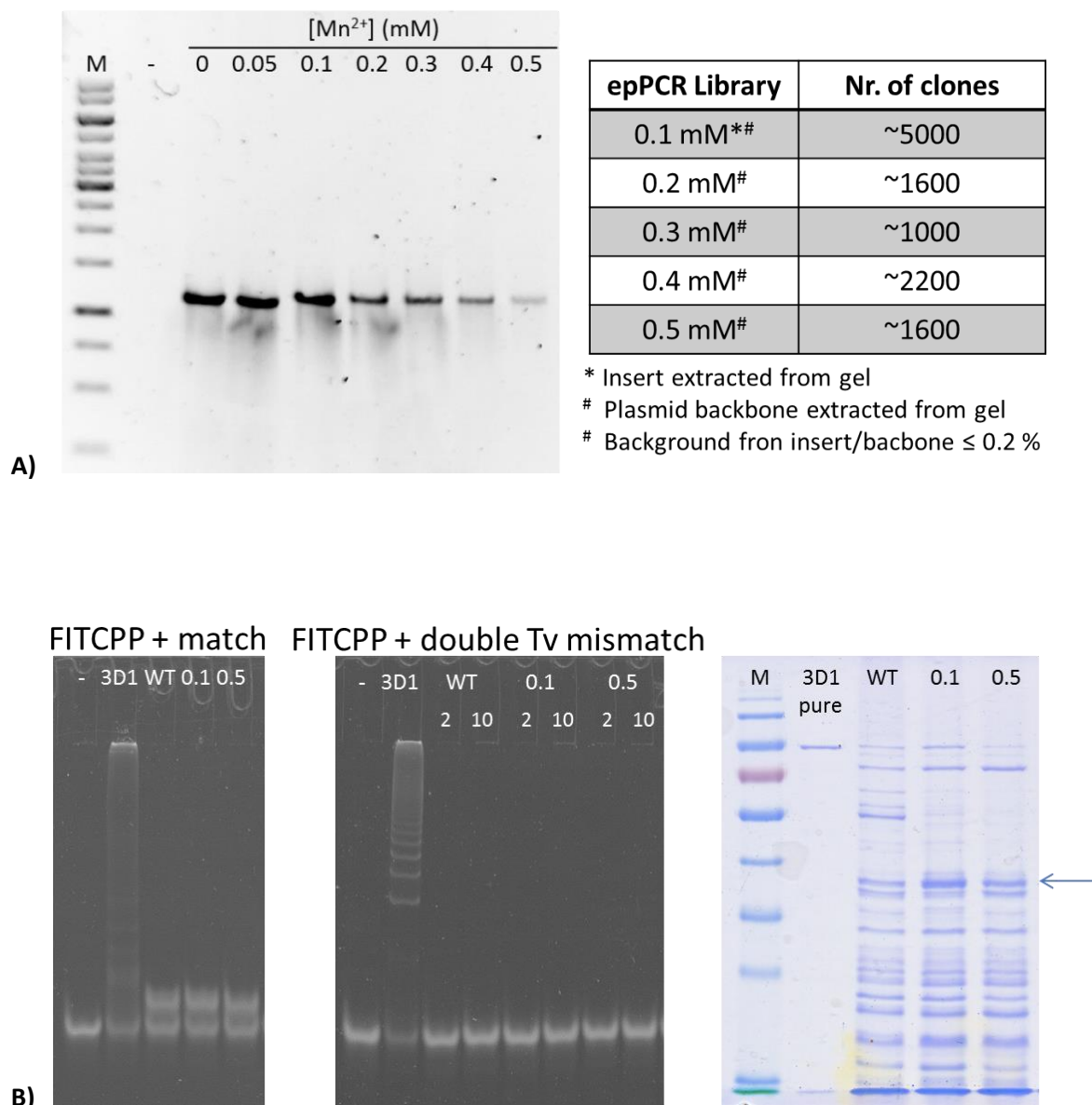
### **3.3. Engineering of the Y-family polymerase Dpo4 for application in directed evolution**

The engineering of Dpo4 polymerase from *Sulfolobus solfataricus* towards more efficient elongation of consecutive transversion mismatches is presented in this section. To achieve this objective, the Scorpion probe screening system was used to screen saturation mutagenesis libraries of Dpo4. The screening campaign resulted in the identification of a few hits which were subsequently purified and characterized with respect to the desired activity. Finally, one of the improved Dpo4 mutants was employed in SeSaM step 3. The resulting library contained a significantly increased fraction of the desired mutation type in comparison to a reference libraries prepared Dpo4 WT as well as 3D1 polymerase, which has been routinely employed in SeSaM. These favorable results also served to validate the developed screening system for distributive polymerases. Other aspects, such as the selection of position for mutagenesis and the generation of site specific mutagenesis libraries as well as of random mutagenesis libraries and their pre-screening are also discussed in this section.

#### **3.3.1. Generation and pre-screening of random mutagenesis libraries**

The generation of epPCR libraries was performed with *Taq* polymerases, balanced dNTPs and varied concentration of  $Mn^{2+}$ . The libraries were cloned into pALXetrme-5b using a ligase independent method, PLICing (Blanusa et al., 2010b). The cloned libraries were transformed first into *E. coli* DH5 $\alpha$  and re-transformed in *E. coli* BL21 Gold (DE3)  $\text{Lacl}^{q1}$  cells for screening. As expected, the amount of epPCR product and transformation efficiency

decreased with increasing  $Mn^{2+}$ . However, at least 1500 clones were obtained even for the PCR product obtained even with the highest manganese concentration employed for mutagenesis (Fig. 25A). Due to throughput limitations of the available screening methodology, it was opted for a “quick-and-dirty” test to check for desired activity in the epPCR libraries with the lowest and highest mutational load. The test was carried out by expressing mutant libraries on LB agar supplemented with IPTG. Next, individual clones were pooled and crude cell extracts of the mix of mutants was prepared. The cell-free lysate was used for elongation of a FITC-labeled primer on short linear templates with which the fluorescent primer formed either a match or a double transversion mismatch at the primer’s 3’-end. Visualization of the resulting products on acrylamide gels was performed as a final step. Protein expression was confirmed by visualizing the products on a protein gel. In addition, enzymatic activity could be detected when a matching template was employed, reaffirming that target proteins were successfully produced. However, no activity could be detected when a template forming a consecutive transversion mismatch was employed in the reaction even when at high enzyme load. Expression and screening results are summarized in Fig. 25B. The negative result and the presumably better prospects of finding an improved mutant by semi-rational design, made screening of individual clones from epPCR libraries in microtiter plates undesirable. In stead, more efforts were focused on screening of saturation mutagenesis libraries.

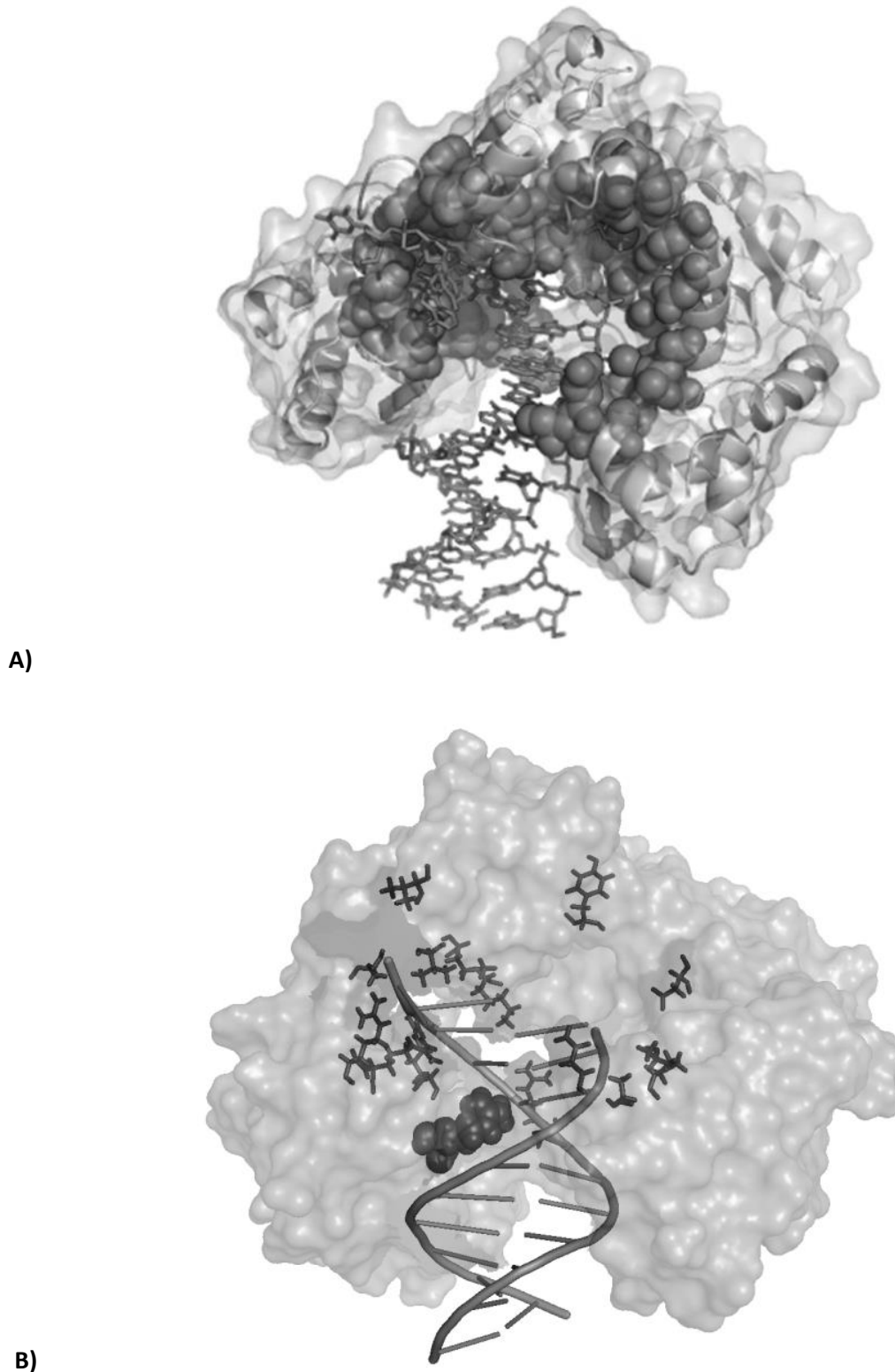


**Figure 25.** Generation and screening of random mutagenesis libraries by epPCR using *Taq*, balanced dNTPs and increasing concentration of  $Mn^{2+}$  and cloned using PLICing. **A)** On the left, epPCR products visualized on an agarose gel. On the right, a table showing the transformation efficiencies obtained for epPCR libraries; **B)** Screening of entire epPCR libraries 0.1 mM and 0.5 mM for elongation of a double transversion mismatch elongation using the forward primer elongation assay. Reactions with lysate of expression culture comprising a mix of epPCR mutants was carried out with FITC-PP primer and matching as well as double transversion mismatch templates; 2  $\mu$ L and 10  $\mu$ L of crude cell extract was used in the case of double transversion elongation.

### 3.3.2. Selection of amino acid positions for saturation mutagenesis studies.

A total of 21 amino acid residues for saturation mutagenesis were selected, namely Y10, V32, A34, A44, V62, M76, K78, D156, A181, P184, 188, A191, R242, R247, K275, V289, L293, R331, R332 and R336. The selection of residues for saturation mutagenesis was done in a semi-rational manner whereas computational tools and literature search were used to identify potential sites for mutagenesis. Firstly, a Hotspot Wizard (Pavelka et al., 2009) analysis of the amino acid sequence and crystal structure of Dpo4 was carried out. Hotspot Wizard is a program that combines bioinformatics databases and computational tools to perform simultaneous structural and evolutionary analyses for automated identification of sites for engineering of substrate specificity, activity or enantioselectivity of enzymes. Residues identified as “hotspots” for mutagenesis are rated according to their mutability and are accompanied with the structural and functional information. The statement of variability at the respective positions enable rational predictions about the mutability of the amino acids and serves as an indicator of the possible (functional) diversity. The top scoring positions from Hotspot Wizard analysis of the amino acid sequence of Dpo4 were visualized on the available crystal structure in order to identify those residues that participate in substrate binding (i.e., were found less than 6 Å from the DNA substrate). As a second selection criterion, the difference in B-factors for Dpo4 crystal structures with matching (PDB: 1S97) and mismatching (PDB: 1S9F) template-primer-DNA substrate and these values were used to refine the primary selection of residues for mutagenesis. With the help of the B-FITTER program (Reetz and Carballeira, 2007), one can analyze the electron density distributions in crystal structures and study the flexibility and importance of individual amino acids (reflected in their B-factor values) on properties such as thermostability and substrate specificity. The least flexible residues from the primary assortment were excluded from the selection. In a last selection step, an in-depth literature search was carried out in order to find experimental evidence for the importance of as many positions from the secondary assortment and to narrow down the final selection to ~20 positions. The residues that comprised the primary selection as well as those that were finally selected for mutagenesis are presented in Fig. 26.

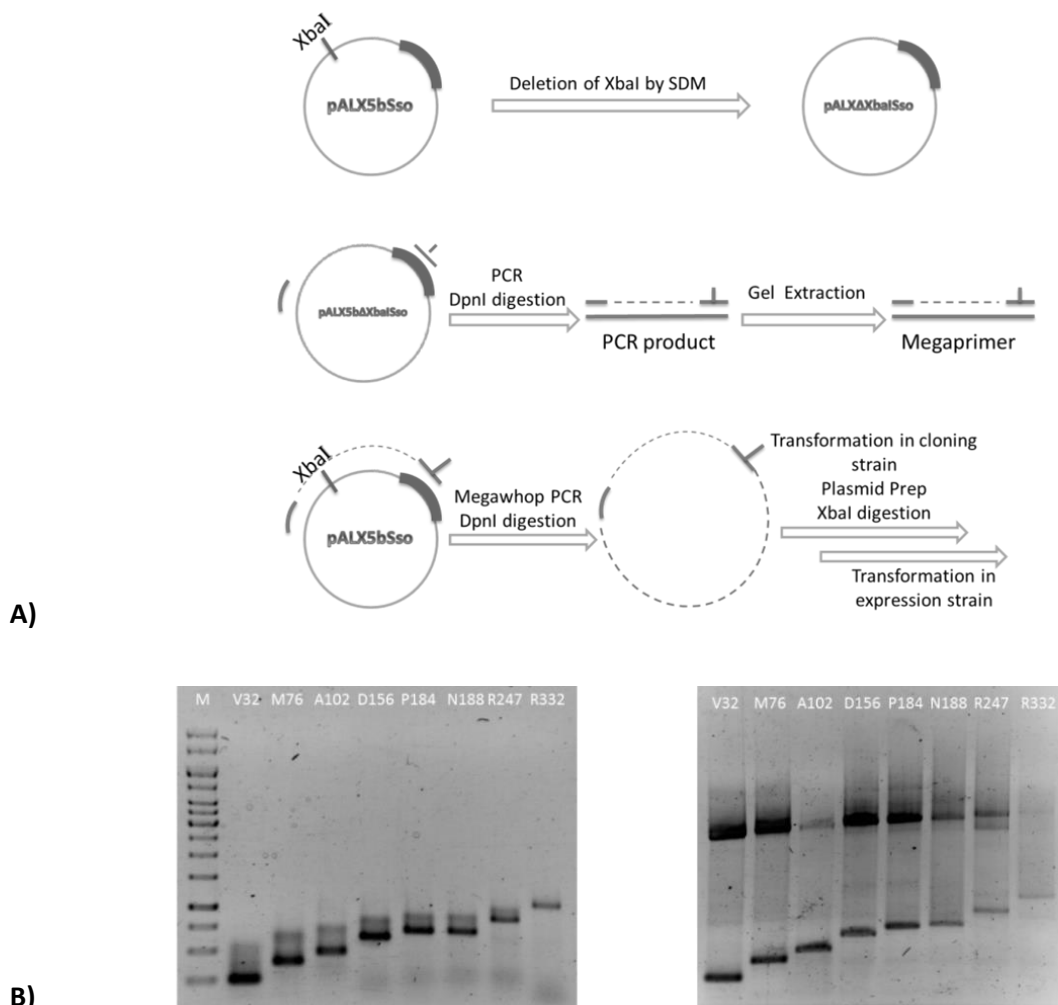




**Figure 26.** Crystal structure of DNA polymerase IV from *Sulfolobus solfataricus* bound to G:T mispaired DNA substrate in the presence of an incoming nucleotide (PDB ID 1S97) **A)** Amino acid residues within 6 Å from the DNA substrate; **B)** 21 residues were selected for saturation mutagenesis; 20 residues (Y10, V32, A34, A44, V62, M76, K78, D156, A181, P184, P188, A191, R242, R247, K275, V289, L293, R331, R332) are shown as sticks and residue R336, later identified in this study mutagenic hotspot, is represented in balls.

### 3.3.3. Generation of saturation mutagenesis libraries

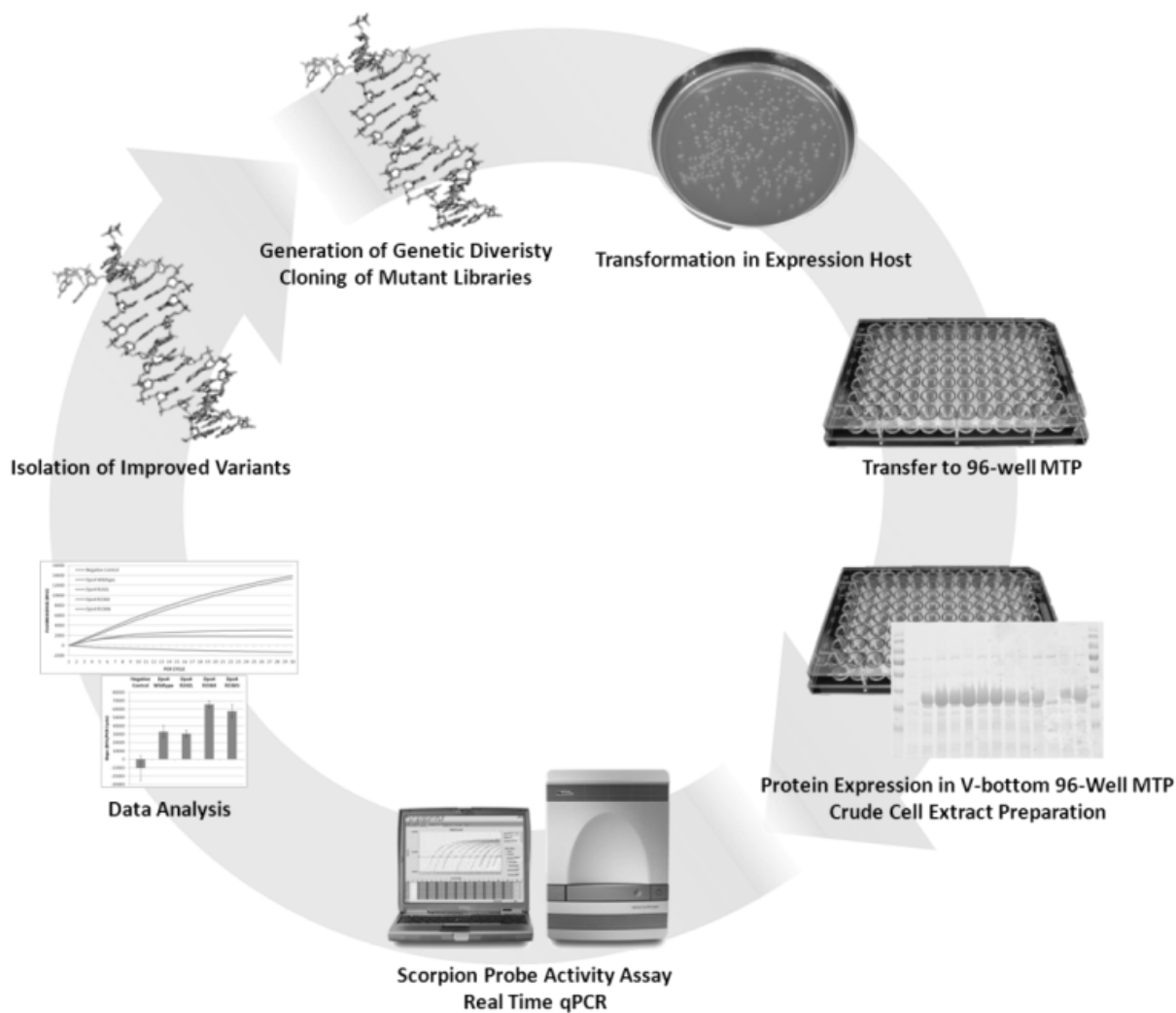
The generation of saturation mutagenesis libraries was carried out using a two-step PCR procedure involving amplification of “megaprimers” harboring point mutations and the subsequent cloning of these DNA fragments using MEGAWHOP. The “megaprimers” were generated with a degenerate forward primer and a site-specific reverse primer harboring the mutagenic codon. The overall strategy is presented in Fig. 27A. The PCR product from the “megaprimer” and MEGAWHOP PCRs are shown in an exemplary gel for 8 positions in Fig. 27B. The transformation efficiencies for the MEGAWHOP primers varied significantly, however they were sufficient for the preparation of 2 MTPs per position (i.e., at least 200 clones were obtained).



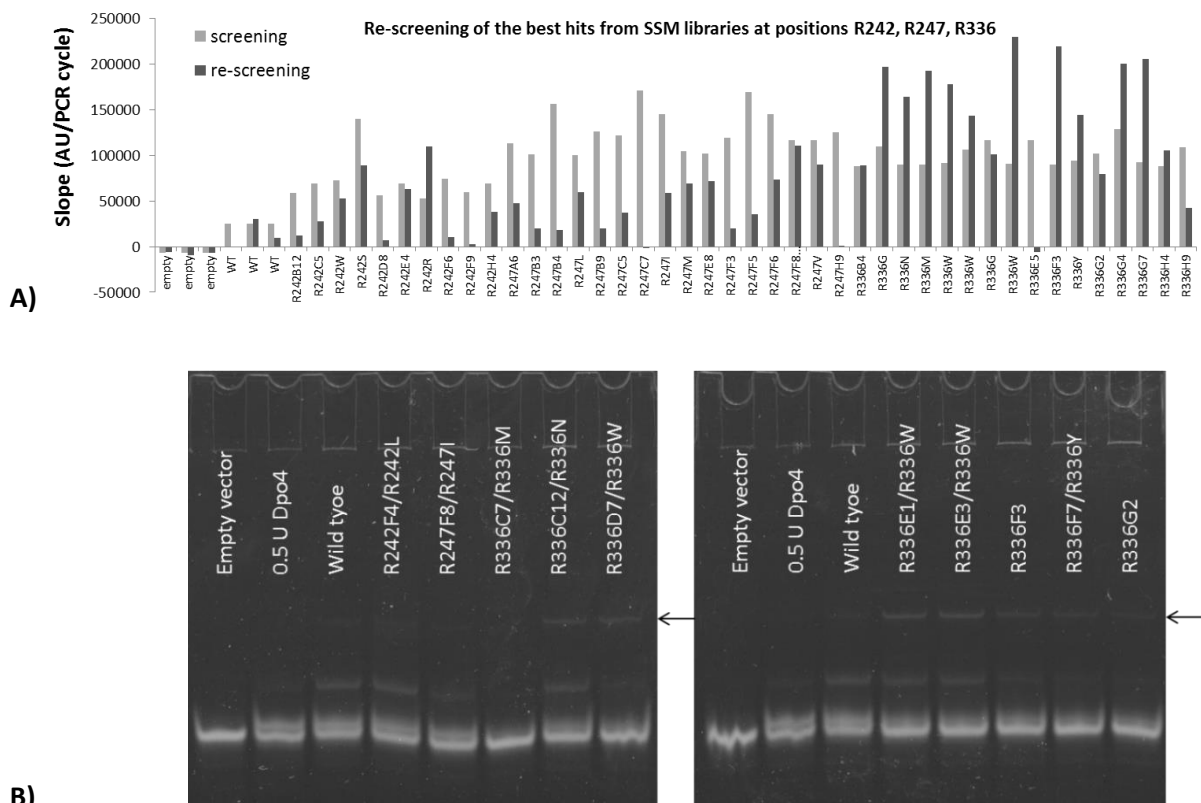
**Figure 27.** Generation of saturation mutagenesis libraries. **A)** Schematic outlining the 2 step PCR procedure involving a “megaprimer” PCR and a MEGAWHOP PCR for generation of saturation mutagenesis libraries as well as the preliminary template generation step for the “megaprimer” PCR; **B)** An exemplary agarose for 8 positions after “megaprimer” PCR (left) and MEGAWHOP PCR (right).

### 3.3.4. Screening of saturation mutagenesis libraries

A total of 21 saturation mutagenesis (SSM) libraries were generated at positions Y10, V32, A34, A44, V62, M76, K78, D156, A181, P184, N188, A191, R242, R247, I248, K275, V289, R298, L293, R331, R332 and R336. The respective mutant libraries were first screened for improved elongation of a defined double transversion mismatch (i.e., GC in the primer stand facing GC in the template strand) using Scorpion detection system in MTPs. The overall screening strategy is shown in Fig. 28. The data from the primary screening was analyzed and based on the result 3 new 96-well MTPs were prepared comprising only hits from the initial screening. A new test with the help of the Scorpion detection system was carried out in duplicate (Fig. 29A), and only the most reproducible results were considered for further characterization by forward primer elongation assay. Fifty positive hits were tested using forward primer extension assay for elongation of degenerate double transversion mismatch, i.e., PP in the primer strand and CT in the template strand, where „P“ is the degenerate pyrimidine analogue, dPTP, employed in SeSaM (Fig. 29B). The clones were also sequenced to identify the mutations contained by individual hits. The highest activities were detected in SSM libraries at positions R242 and especially R336. Consequently, one mutant from library R242 and two mutants from library at position R336 were selected for characterization in purified form.



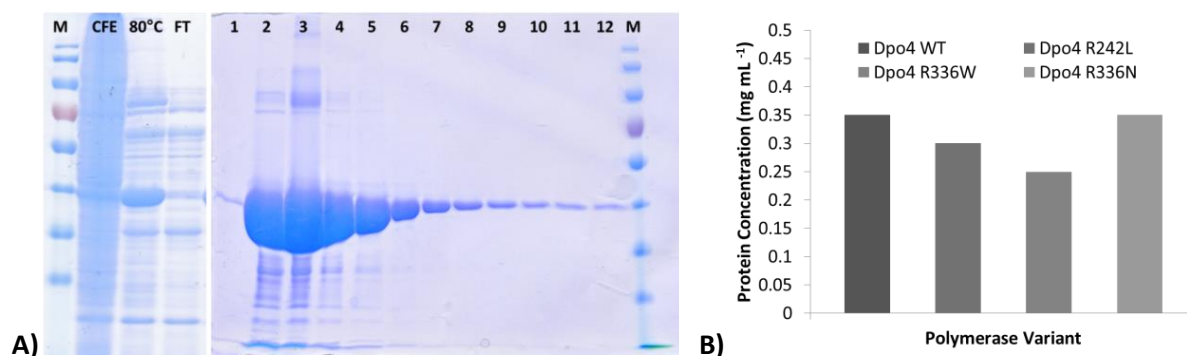
**Figure 28.** Scorpion probe screening system for relative quantification of polymerase activity in 96-well microtiter plate format. A Scorpion probe comprises a primer that is covalently bound via a spacer (Spacer18, hexaethylene glycol) to a probe oligo with a fluorophore (6-FAM, 6-carboxyfluorescein) attached at its 5'-end. A separate quenching primer consists of a quencher molecule (BHQ-1, black hole quencher 1) that is attached to an oligo with a complementary sequence to the 6-FAM oligo. If the probe is not elongated, the quenching primer hybridizes to the probe oligo and fluorescence is quenched. In case that the oligo probe is elongated an intramolecular hybridization sequence without a quencher is generated. As a result the probe rearranges and the fluorophore becomes unquenched generating a fluorescence signal (ex: 495 nm; em: 520 nm)



**Figure 29.** Exemplary results from re-screen of positive hits from site-saturation mutagenesis libraries at positions R242, R247 and R336 using **A)** Scorpion assay and **B)** forward primer elongation assay; full length product indicated by an arrow.

### 3.3.5. Purification of improved Dpo4 variants

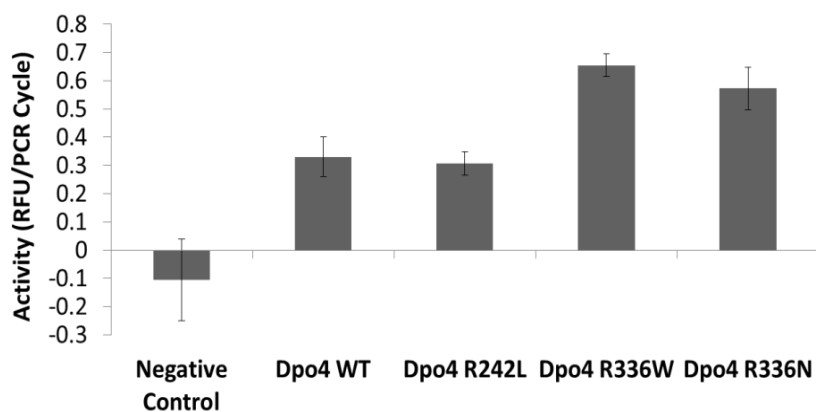
Dpo4 mutants R242L, R336N and R336W were selected for characterization and together with Dpo4 WT were expressed in flasks and purified via affinity chromatography (His-tag) as described in Materials and Methods. The elution fractions were subjected to SDS-PAGE analysis to determine the purity of individual samples (Fig. 30 A). Purity above 95 % could be achieved by a combination stringent washing and sampling procedures. After dialysis of the purest fractions, protein concentration was measured using bicinchoninic acid assay (BCA) in preparation for activity tests. Protein concentrations of above 200 mg/mL could be achieved for all purified polymerase variants. The purified proteins were used immediately after purification for activity assays.



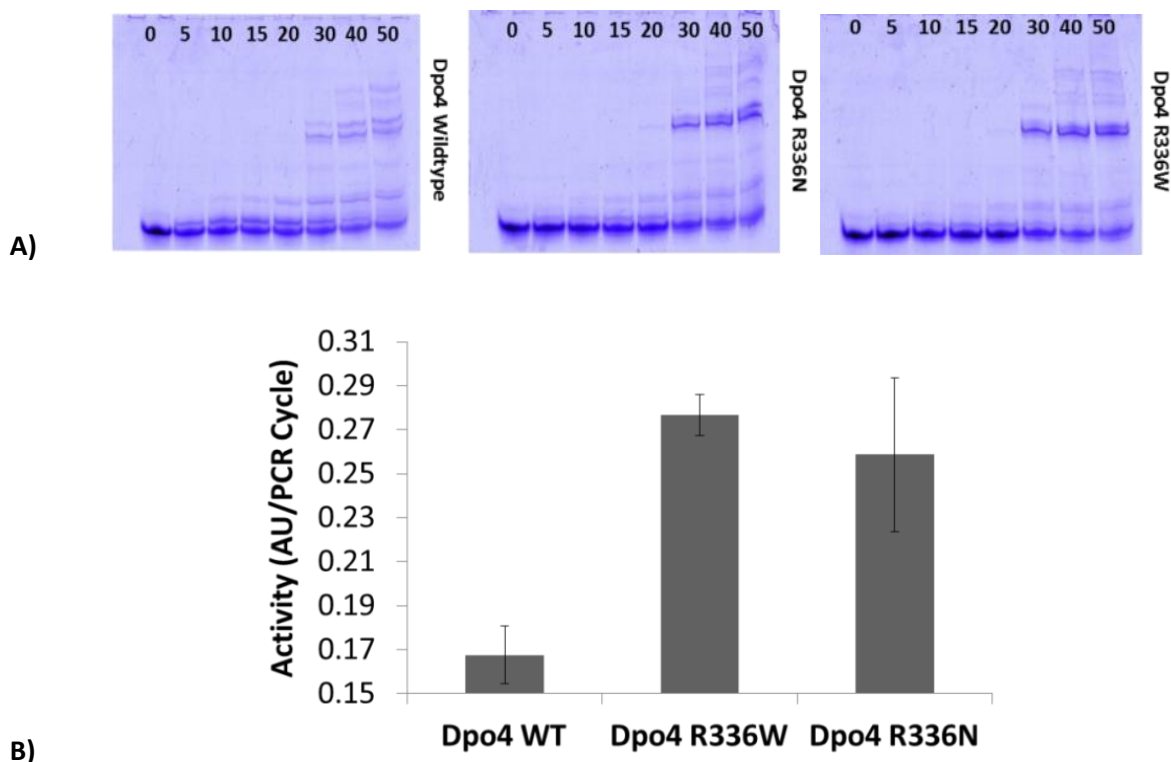
**Figure 30.** Purification of Dpo4 WT and 3 Dpo4 mutants. **A)** Exemplary gel after column purification. Dpo4 WT and mutants were from *E. coli* BL21 Gold DE3 lacI<sup>q1</sup>-pALXtreme-1a and purified by His-tag affinity chromatography. Crude cell extracts were heat treated to precipitate *E. coli* host proteins prior to the chromatography step. Aliquots were collected during purification and analysed by SDS-PAGE. Only the purest fractions were taken for dialysis against polymerase storage buffer. M – marker; CFE – cell free extract; 80 °C – heat precipitation of *E. coli* proteins at 80 °C for 10 minutes; FT – flow through fraction after washing; 1-12 – elution fractions; **B)** Concentration determination of Dpo4 polymerase variants via BCA assay.

### 3.3.6. Transversion mismatch elongation tests with purified Dpo4 variants

Dpo4 variants R242L, R336N and R336W were first characterized using the Scorpion probe activity assay for elongation efficiency of a defined consecutive transversion mismatch (GC:GC) after normalization to protein concentration (Fig. 31). This tests showed that Dpo4 R336N and R336W were approximately twice as efficient in elongating a GC:GC mismatch when compared to Dpo4 WT; the activity of Dpo4 R242L after purification was comparable to that of Dpo4 WT. Consequently, R336N and R336W were investigated for their ability to elongate a degenerated double Tv mismatch (PP:CT, where „P“ is a degenerate pyrimidine analogue, also employed in SeSaM using the well established forward primer elongation assay. Figure 32 shows that R336N and R336W were roughly 1.5-fold more efficient in elongating a consecutive transversion mismatch of the aforementioned type.



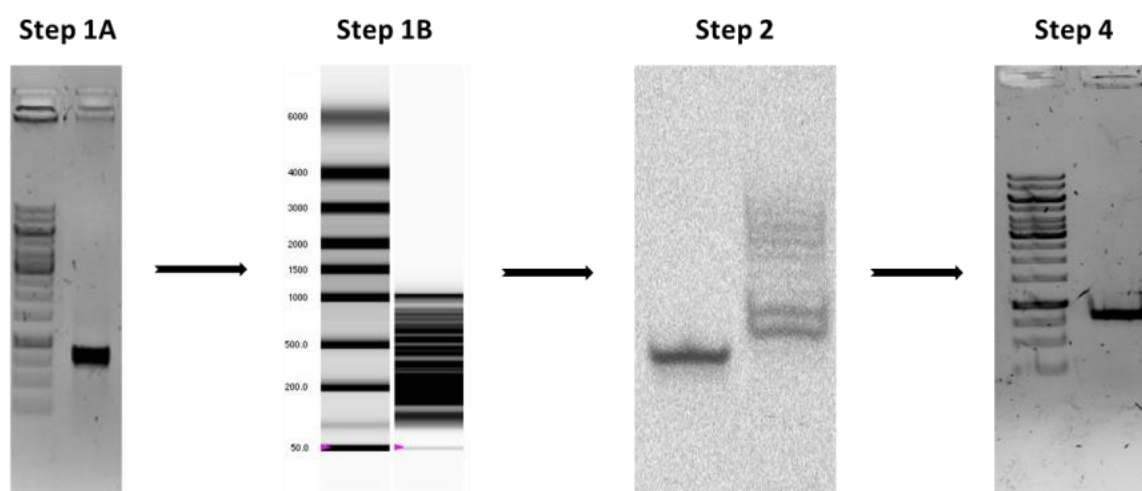
**Figure 31.** Scorpion probe assay based characterization of purified Dpo4 WT and mutants R242L, R336W, R336N. Mutants are evaluated for their ability to elongate a double Tv mismatch (GC:GC) by comparing the slope of first 10 PCR cycles of each reaction.



**Figure 32.** Forward primer elongation assay was performed with purified Dpo4 WT and the Dpo4 variants R336W and R336N. R336W and R336N were evaluated for their ability to elongate a double transversion mismatch with the degenerated P-base (PP:TC; “P”—a degenerated pyrimidine analogue dPTP). A) Exemplary acrylamide gels on which product profiles of reactions with Dpo4 WT and the Dpo4 variants R336W and R336N are visualized. Artificial color is added for more clear representation; B) Densitometric analysis of three independent sample series. The depletion of non-elongated primer is quantified by ImageJ software (Abramoff et al., 2004) and plotted over PCR cycle and the respective slopes are plotted as an activity measure and plotted in Microsoft Excel. Error bars represent the calculated standard deviation.

### 3.3.7. Validation of Dpo4 R336N in Step III of SeSaM

Three model SeSaM libraries of *bsla* (546 bp-long gene encoding lipase A from *Bacillus subtilis*) were prepared in which 3D1, Dpo4 WT, and Dpo4 R336N have been employed in Step III of SeSaM. The sequencing results of all three SeSaM libraries are summarized in Table 4; transitions were roughly twice as abundant as transversions in all SeSaM libraries. The percentage of non-mutated DNA ranged between 18.9 % for the 3D1 library, 12.2 % for the Dpo4 R336N library, and 7.9 % for Dpo4 WT. The percentage of consecutive mutations ranged from 20.2 % (3D1) to 9.8 % (Dpo4 R336N). Most notably, R336N generates >2.5 times more subsequent transversions than Dpo4 WT. Dpo4 also creates more consecutive transversions than 3D1, suggesting that even Dpo4 WT is already an attractive polymerase for diversity generation in random mutagenesis methods such as SeSaM. The mutational spectra of 3D1 differs from the previously reported one (Mundhada et al., 2011b) likely due to differences in the composition of the employed genes (*bsla* is shorter and AT-richer than *egfp*).



**Figure 33.** DNA modification steps from I to IV of the SeSaM random mutagenesis method.



**Table 4.** Mutational spectrum generated with the SeSaM-Tv-II protocol employing 3D1, Dpo4 WT or Dpo4 R336N in Step III. The fraction of consecutive Tv mutations is shown at the bottom of the table and highlighted in bold.

SeSaM G-fwd bsIa gene	3D1 Pol	Dpo4 WT	Dpo4 R336N
<b>Nr. of clones sequenced</b>	94	97	92
<b>AT--&gt;GC ( %)</b>	31.7	41.7	32.4
<b>GC--&gt;AT ( %)</b>	18.3	11.5	10.6
<b>AT--&gt;TA ( %)</b>	7.3	8.3	6.5
<b>AT--&gt;CG ( %)</b>	1.8	1.6	3.7
<b>GC--&gt;TA ( %)</b>	11.6	8.3	4.2
<b>GC--&gt;CG ( %)</b>	6.1	6.0	6.9
<b>Ts/Tv ratio</b>	1.9	2.2	2.0
<b>Deletions ( %)</b>	4.9	16.3	26.9
<b>WT ( %)</b>	18.9	7.9	12.2
<b>Ratio of Consecutives mutations/total clones (%)</b>	<b>20.2</b>	<b>13.4</b>	<b>9.8</b>
<b>Ratio of TvTv mutations/Total nr. of consecutives (%)</b>	<b>5.3</b>	<b>15.4</b>	<b>44</b>

#### 4. Summary and conclusions: Engineering of DNA polymerase for Application in directed protein evolution

In random mutagenesis methods, polymerases especially capable of introducing subsequent mutations and transversion mutations are essential in order to generate diverse mutant libraries. Consecutive nucleotide exchanges result in doubling of the generated chemical diversity on amino acid level (Wong et al., 2006c, 2006d, 2007d). Thus far, despite extensive directed evolution campaigns on polymerases, only few variants have been reported to achieve consecutive nucleotide exchanges. One example is the 3D1 polymerase (a StEP variant derived from three genes of the genus *Thermus*), which has been shown to extend consecutive mismatches (d' Abbadie et al., 2007b). When 3D1 polymerase was employed in the Sequence Saturation Mutagenesis (SeSaM) method (mismatch elongation in SeSaM step 3 is highly desired), most observed consecutive mutations were double mutations of the type transversion followed by transition (TvTs) (Mundhada et al., 2011b). These resulted in a mutant library with around 40 % increase of unobtainable or "hardly-obtainable" exchanges in comparison to other random mutagenesis methods. However, a complete, unbiased randomization spectrum was not achieved with consecutive TvTs mutations alone. SeSaM can reach its full potential in diversity generation if two or more consecutive transversion mismatches could be elongated in SeSaM step 3 (Wong et al., 2007b). A polymerase that meets these criteria is unlikely to be found in nature since transitions (which preserve chemical diversity on the protein level) are by far the predominant type of mutations that occur *in vivo* (Wong et al., 2007b). Thus, engineering a polymerase with an enriched transversion mutational spectrum is highly desirable for random mutagenesis methods such as SeSaM. The Y-family of DNA polymerases has been suggested as a promising source for enzymes with the right set of properties for the desired application.

Dpo4 from *Sulfolobus solfataricus* was selected as a starting point for this engineering campaign. As an archetype for the Y-family of polymerases (which naturally function in translesion DNA synthesis), Dpo4 holds promise with regard to elongation of consecutive mismatches since it has medium-low fidelity, moderate thermostability, and lacks displacement or exonuclease activity (Boudsocq et al., 2001); Dpo4's crystal structure has also been solved (Ling et al., 2001) enabling rational protein engineering. Site saturation at

21 amino acid residues in Dpo4 yielded two variants with around 2-fold increased elongation efficiency of a defined transversion mismatch (GC:GC) and about a 1.5-fold enhanced efficiency for a degenerate base mismatch (PP:TC, dPTP – pyrimidine analogue). Both Dpo4 mutants harbour a single amino acid exchange at position R336 which is located in the little finger (LF) domain of Dpo4. The LF domain is a unique feature of Y-polymerases known to play a pivotal role in determining Dpo4's performance in terms of fidelity and lesion bypass capabilities (Boudsocq et al., 2004). Furthermore, a previous computational study of Dpo4 has shown that R336 is a key residue within the LF domain that is responsible for DNA binding and DNA translocation after phosphodiester bond formation past damaged DNA (Wang et al., 2006b). The molecular dynamics studies revealed that after phosphodiester bond formation the side chain of R336 rotates by 120° which shortens its length by 1.5 Å. R336 interacts with the phosphate backbone via two hydrogen bonds (H-bonds) before the rotation; after rotation one H-bond is broken and the remaining H-bond is strengthened. In this experimental study, Dpo4 variants R336N and R336W which were selected for elongating subsequent (GC:GC) transversions, also yielded similar results with a degenerated transversion mismatch. Taken together, the findings from computational and experimental studies indicate that position 336 is of global importance for the elongation of non-canonical primer-template pairs by Dpo4. Additionally, previously unreported amino acid positions were also identified during screening as hotspots that could be mutagenized to improve the translesion DNA synthesis capability of Dpo4, notably V32, R242, R247. Mutagenizing these positions resulted in a moderate improvement in performance of Dpo4 in cell lysate in comparison to the wild-type but lower activity than that measured for the variants harboring mutation at position 336.

Dpo4 R336N performance was directly validated under application conditions in step 3 of SeSaM which employs the degenerate pyrimidine analogue dPTP. Comparison of the mutational spectra of 3D1 polymerase with Dpo4 WT showed that the selection of Y-polymerase for generating subsequent transversion mutations was indeed a suitable choice. This was backed up by the fact that the use of Dpo4 WT in SeSaM resulted in 10 % more subsequent transversions than 3D1 (Table 4). Furthermore, the mutant R336N generates twice as many transversion exchanges than Dpo4 WT. The latter also served to demonstrate that the Scorpion probe system was applicable to screening for polymerases with a higher

elongation efficiency of transversion mutations. Interestingly, Dpo4 WT and Dpo4 R336N outperformed 3D1 not only in the frequency of transversion mutations but also in overall mutation frequency (3D1:  $1.16 \times 10^{-3}$ ; Dpo4 WT:  $2.09 \times 10^{-3}$ ; Dpo4 R336N:  $1.44 \times 10^{-3}$ ). However, a higher mutation frequency comes at the expense of a higher deletion rate (3D1: 4.9 %; Dpo4 WT: 16.3 %; Dpo4 R336N: 26.9 %). The deletion frequency is not surprising since Dpo4 is known to generate single base deletions at high frequencies in repetitive sequences using a template slippage mechanism (Wu et al., 2011). Overall, despite the increased fraction of nonsense sequences (WT and deletions) compared to the reference libraries, Dpo4 R336N with its current performance is already a significant advancement in subsequent transversion generation and especially for the SeSaM random mutagenesis method to complement the mutational spectrum of epPCR based methods.

**CHAPTER II. Regioselective Synthesis of Mono- and Di-hydroxy Products from Benzenes using P450 BM3 Monooxygenase in Purified Form and as Whole Cell Catalyst**

## 6. Introduction

### 6.1. Industrial biotechnology

Industrial biotechnology explores sustainable routes to chemical synthesis. This applied approach employs a set of methodologies from e.g., molecular and computational biology, microbiology, biochemistry and engineering to develop bio-based industrial and environmental products and processes. To achieve this goal, enzymes either in isolated form or as whole-cell catalysts are utilized for the production of chemical compounds, feed and foods, textiles, in waste water treatment and biomining, etc. (Philp et al., 2013; Wenda et al., 2011; Woodley et al., 2013). Industrial biotechnology holds great promise for widespread application and has the potential to minimize or even completely eliminate most drawbacks associated with classical chemical methods such as the dependency on non-renewable resources and the associated high environmental costs (Philp et al., 2013). The use of biotechnology in industrial processes can change not only the way products are made but also to generate novel products, concepts and opportunities for business and academia alike (Erickson et al., 2012; Wenda et al., 2011). However, due to the fact that modern industrial biotechnology is still in its early stages of development and covers only a small segment of industrial processes, the benefits and potential of it are only partially exploited by industry, policymakers, and consumers (Woodley et al., 2013).

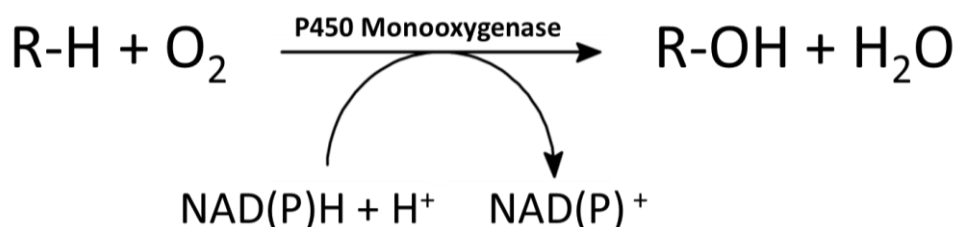
Rudimentary biotechnology has been an integral part of human development (Buchholz and Collins, 2013). The production of wine and beer, pickled food and cheese can be viewed as early biotechnological processes carried out the human kind. Since the 19<sup>th</sup> century, with intensification of scientific efforts in chemistry and biology, people started to become aware of the existence of microorganisms and their enormous practical potential. From this moment on, it didn't take long to start exploiting this potential. The true biotechnological revolution began with the end of World War II (Buchholz and Collins, 2013) and its importance to the human kind has been exponentially growing ever since. At present there are numerous examples of processes partially or entirely relying on enzymes, some the most prominent example being the production of amino acids (e.g., lysine), antibiotics (e.g., penicillin), vitamins (e.g., ascorbic acid), alcohols (e.g., butanol) (Soetaert and Vandamme, 2010).

Despite the outstanding versatility and reaction potential of enzymes as well as quickly advancing methodologies for their identification and characterization, a number of drawbacks associated with biocatalytic reactions exist (Wenda et al., 2011). Some major disadvantages are associated with the often insufficient activity of the employed biocatalysts, compromised stability outside their natural environment and substrate/product inhibition effects. Other stem from carrying out chemical reactions in aqueous environment e.g., low substrate solubility and complicated downstream processing steps (Buchholz et al., 2012). Even though these limitations can be generally overcome, the implementation of biosynthetic strategies requires significant monetary investments and still limits the broader implementation of enzymes in industrial processes (Soetaert and Vandamme, 2010). Nevertheless, emerging tendencies for more environmentally friendly processes that also provide the possibility to run more selective or even chemically unfeasible reactions coupled to the growing political pressure to refurbish economy towards a more sustainable one promises a bright future to industrial biotechnology (Erickson et al., 2012).

The workhorses of industrial biotechnology, enzymes, are generally classified according to the type of reactions they catalyze. Enzymes fall in one of six distinct classes – EC 1: oxidoreductases (catalyze redox reactions; e.g., laccase), EC 2: transferases (catalyze the transfer a chemical group; e.g., glucosyltransferase), EC 3: hydrolases (catalyze hydrolysis; e.g., cellulases), EC 4: lyases (catalyze the cleavage of C-C, C-O, C-N and other bonds by means other than by hydrolysis; e.g., pectate lyase), EC 5: isomerases (catalyze isomerization of a single molecule; e.g., glucose isomerase), EC 6: ligases (catalyze the fusion of two molecules with concomitant hydrolysis of the di-phosphate bond in ATP or a similar triphosphate; e.g., DNA ligase) (Schomburg et al., 2012). Members of all but the latter class (EC 6) are currently being used in industrial processes (Drauz, 2012). In line with the objectives of this study, P450 monooxygenases which fall in the class of oxidoreductase enzymes are introduced in the following section.

## 6.2. Cytochrome P450 Monooxygenases

Cytochrome P450 monooxygenases (P450s) are heme-iron containing oxido-reductases that catalyze the insertion of molecular oxygen into a non-activated carbon-carbon bond as shown in Fig. 34 (Bernhardt, 2006; Hollmann et al., 2011). P450s are ubiquitously found and form one of the largest enzyme superfamilies (Bernhardt, 2006). The name of this enzyme class is derived from the observation that the reduced form of the iron-containing heme prosthetic group with bound carbon monoxide causes a shift in absorbance maximum from 420 to 450 nm (Omura and Sato, 1964). P450s naturally function in catabolic and anabolic pathways alike. Their ability to catalyze various reactions such as reduction, desaturation, cleavage of esters and aldehydes, ring expansion and formation, dehydration, one-electron oxidation, coupling reactions by integrating molecular oxygen selectively into non-activated carbon atoms under mild conditions makes P450s attractive for applications in chemical synthesis (Guengerich, 2001).

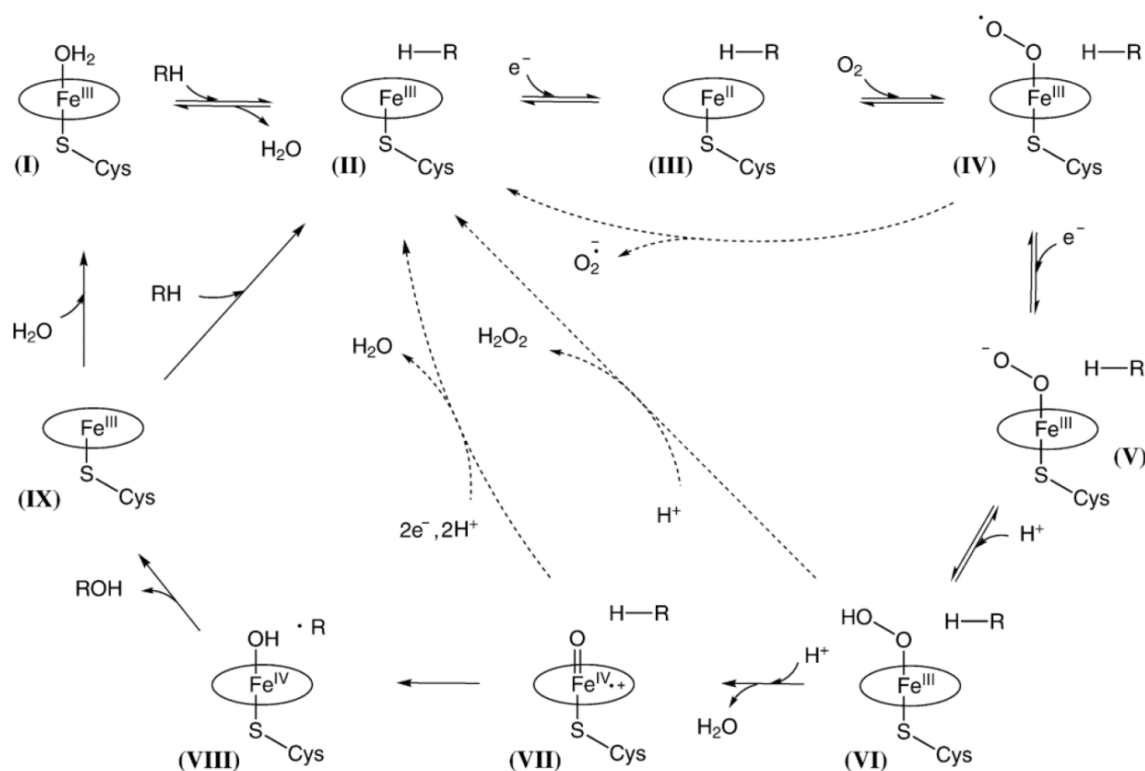


**Figure 34.** Schematic of a typical reaction catalyzed by cytochrome P450 monooxygenases.

Cytochrome P450s are cofactor (NADH or NADPH) dependent enzymes and as such they require electron transfer partners to perform their function (Montellano, 2005). Upon binding of a substrate, the change of oxidation potential of the heme-containing domain induces oxidation of the NAD(P)H by a reductase and initiates the transfer of the electrons to the enzyme's active site often via a mediator molecule (ferredoxin or flavin). A large diversity with regards to electron transfer systems for P450s exists; 10 classes of electron transfer systems or chains have been identified (Hannemann et al., 2007). These include multicomponent systems and fusions of the P450 and reductase components into one polypeptide chain that can be either membrane-bound or cytosolic.



The catalytic cycle of P450s is a complex, multistep process which has to be carried out meticulously to ensure efficient substrate hydroxylation. Fig. 25 gives an overview of the catalytic mechanism. In the resting state, a water molecule is found at the heme iron (I). The water leaves upon substrate binding (II). This leads to an increase of the oxidizing power of the heme. One electron from a NAD(P)H cofactor is then transferred via a reductase/redox mediator molecule (III) to the heme iron. The reduced heme iron (ferrous state) can bind a molecule of oxygen (IV; oxy-complex). A further reduction by another electron takes place and a peroxy-complex (V) is formed. After protonation of this complex (VI), compound 0 is formed. After a second protonation step, release of a water molecule follows and formation of compound I (VII) takes place. The latter represents the active species capable of abstracting H-atom from the substrate molecule (VIII). The catalytic cycle is reinitiated with the product release the active site. In addition to hydroxylation via rebounding, epoxidation of substrates without H-abstraction from the substrate molecule can take place (de Visser and Shaik, 2003). P450-catalyzed epoxidation involves a directly attack by compound I (VII) at unsaturated carbon atoms (de Visser and Shaik, 2003; Whitehouse et al., 2011). Last but not least, in some instances an “uncoupling” reaction can take place in which the electrons derived from NADPH are wrongly transferred during the catalytic cycle resulting in hydrogen peroxide. The frequency of occurrence of this shuttle pathway is mainly dependent on the employed (non-natural) substrate and it is seen as an unwanted artefact in biocatalytic reactions with P450s (Grinkova et al., 2013).

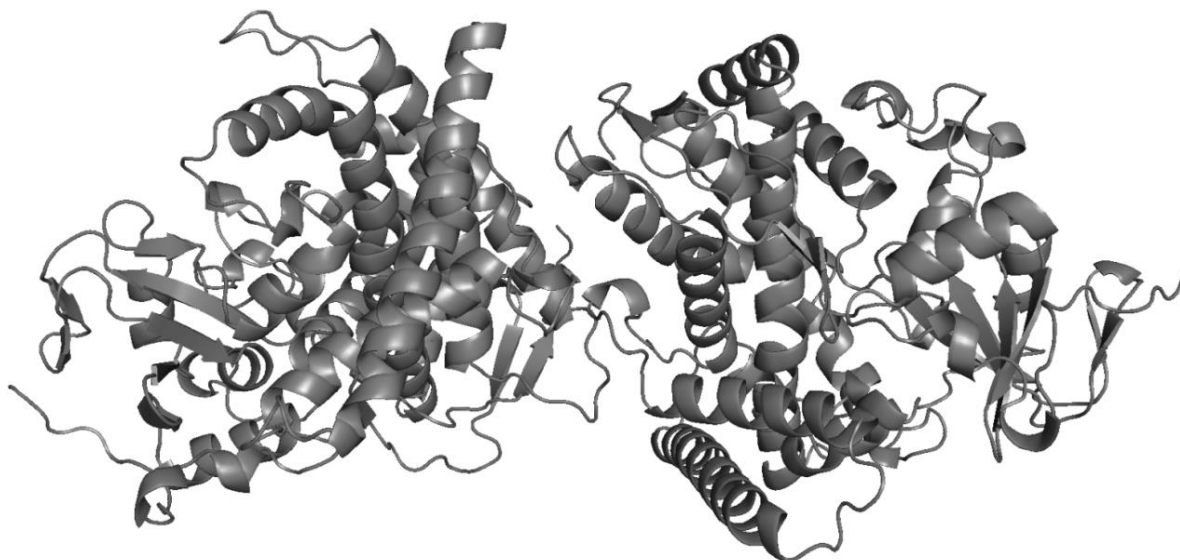


**Figure 35.** P450 catalytic cycle (taken from Whitehouse et al., 2012).

### 6.3. Cytochrome P450 (P450 BM3)

Cytochrome P450 from *Bacillus megaterium* (EC 1.14.14.1), for short P450 BM3, is one of the best studied members of the P450 enzyme family. P450 BM3, first described in the late 1980s (Narhi and Fulco, 1987), is a water-soluble, NADPH-dependent monooxygenase in which both heme and reductase domains are contained on a single polypeptide chain. Although the precise natural function of this monooxygenase remains unknown, there is strong evidence that the natural substrate of P450 BM3 are middle and long chain fatty acids for which P450 BM3 exhibits excellent NADPH coupling efficiency (Munro et al., 2002). This enzyme (and its mutants) was shown to hydroxylate a wide range of substrates, e.g., propane, cycloalkanes, naphthalene and testosterone (Whitehouse et al., 2012). The crystal structure of P450 BM3 has been solved and reveals that P450 is a dimeric protein formed by two identical peptide chains (Fig. 36). The reductase and monooxygenase domains in each chain are connected via a short amino acid linker (Govindaraj and Poulos, 1995, 1996) ensuring the efficient electron transfer between both monomers. In this thesis, P450 BM3 is

used in the context of regioselective biotransformations of aromatic substrates with isolated enzyme and as a whole cell catalyst.



**Figure 36.** The crystal structure of the complex between the heme- and FMN-binding domains of bacterial cytochrome P450 BM3 (PDB ID. 1BU7) (Sevrioukova et al., 1999).

#### 6.4. P450 BM3 catalyzed hydroxylation of aromatic hydrocarbons

Cytochrome P450 monooxygenase from *Bacillus megaterium* (P450 BM3) holds a significant potential for application in organic synthesis. The promiscuity of P450 BM3 which is naturally optimized for hydroxylation and epoxidation of long-chain unsaturated fatty acids, has been vastly enhanced by means of protein engineering, making it one of the most versatile and valuable monooxygenases available to chemists (Whitehouse et al., 2012). In a recently published study, a triple mutant of P450 BM3 (P450 BM3 M2 (R47S/Y51W/I401M)) was engineered into an efficient catalyst for the synthesis of monosubstituted benzenes (Dennig et al., 2013). Phenolic compounds are important precursors for chemical synthesis with a yearly turnover of 9 megatons (Rappoport, 2004). Preliminary studies have indicated that phenols could undergo a second P450 BM3-catalyzed hydroxylation to yield the respective 1,4-hydroquinones. Similar to phenols, 1,4-hydroquinones find applications in the chemical and pharmaceutical industries, e.g., as chemical precursors or antioxidant and bleaching agents for the production of rubber, polymers, dyes and pigments, fungicides and

herbicides, etc. (Fujimori and Nakamura, 1985; Hua et al., 2002; Hudnall, 2000; Krumenacker et al., 2000; Pérez et al., 2000; Ran et al., 2001; Rendon and Horwitz, 2012; Song et al., 2007).

Hydroquinones can be chemically synthesized from benzenes via the Hock process (rendering phenol), followed by hydroxylation of phenol using, e.g., Fremy's salt or hydrogen peroxide and zeolite as catalysts (Deya et al., 1987; Hudnall, 2000; Krumenacker et al., 2000; Rappoport, 2004). This process, accounting for approximately 60 % of the global hydroquinone production (Rappoport, 2004), is characterized by low isolated product yields and formation of side products (e.g., catechol). Electrochemical hydroxylation of phenol can provide alternative for carrying out the second oxidation step (Rautenbach, 2007), yet it also generates a mixture of dihydroxylated products. Direct synthesis of hydroquinone from benzene was also reported for n-alkane assimilating bacteria without specifying performance parameters (Yoshida et al., 1990). Benzene oxidation can be catalyzed by aromatic peroxygenase (APO) from *Agroclybe aegerita* and leads to a variety of side products, including 1,4-hydroquinone, in the course of benzene conversion (Karich et al., 2013). Hydroquinones can be produced from other starting substrates as well. For example, the three step chemical process from aniline uses manganese dioxide (20 % excess, in sulfuric acid) yielding benzoquinone which then has to be purified by steam stripping and reduced to hydroquinone using an aqueous iron suspension (Laha and Luthy, 1990). Biochemically, hydroquinone can be produced from quinic acid by a modified E. coli strain using glucose as starting material (Ran et al., 2001). Despite that most of the glucose is used for biomass and metabolite production (79 to 85 %), the final step is performed in the presence of stoichiometric amount of strong oxidizing reactants ( $K_2S_2O_8$  and  $Ag_3PO_4$ ;  $(NH_4)_2Ce(SO_4)_3$ ) with hydroquinone yields of 51 to 91 %.

With its hydroxylating potential, P450 BM3 can provide a greener and direct way for synthesis of substituted mono- and di-hydroxybenzenes from benzenes and/or phenols as substrates. This prospective has been investigated and results are described in section 7.1.

## 6.5. The “cofactor challenge” in catalysis with P450 BM3

Due to the fact that P450 BM3 catalysis is dependent on the use of cofactor (NAD(P)H), significant effort has been made in alleviating the need to use stoichiometric amounts of

cofactor to drive P450-catalysed reactions. While a number of alternatives have been pursued, e.g., non-natural electron transfer systems (Nazor et al., 2008), the *in situ* regeneration of the oxidized nicotinamide cofactor remains the most common approach to tackle the “cofactor challenge” (Zhao and van der Donk, 2003). When isolated enzymes are used, the pool of reduced cofactor can be restored through the use of enzymatic regeneration systems (Schrittweiser et al., 2011) or be replenished by chemical, electrochemical or photochemical setups (Hollmann et al., 2011). Cofactor regeneration with enzymes (e.g. glucose dehydrogenase) is the most efficient and commonly employed approach when cofactor replenishment is necessary. Nevertheless, cofactor regeneration systems add significant expenses to the overall process and are generally applied mostly in research labs for chemical synthesis of high value products, less commonly in bulk synthesis (Liu and Wang, 2007). Regeneration by chemical, electrochemical or photochemical setups is a concept creating much excitement but to date no industrially-applicable processes are known. Therefore, these approaches are largely at the stage of basic research and validation (Schrittweiser et al., 2011). A conceptually different way of carrying out *in vitro* enzymatic synthesis that requires cofactor regeneration is the use of intact whole cells (de Carvalho, 2011).

## 6.6. Whole cell catalysis

The use of whole cells in organic synthesis is an attractive prospect as it allows for a cost-effective cofactor regeneration and eliminates the need for labor-intensive enzyme purification. In addition, whole cells provide several advantages over cell-free setups such as higher stability and recyclability of biocatalysts. This in turn confers reduction in the costs for cofactor regeneration and downstream processing (de Carvalho, 2011). However, in whole-cell catalysis the limited permeability of the outer membrane of industrially relevant bacteria such as *E. coli* presents a major drawback preventing their broader application in industrial processes (Chen, 2007).

## 6.7. The outer membrane (OM) of gram-negative bacteria

The outer membrane (OM) of gram-negative bacteria such as *E. coli* provides an effective barrier for macromolecules, hydrophobic molecules and anionic detergents (Vaara, 1992;

Wiener and Horanyi, 2011). As such, it serves a defensive role against potentially harmful agents and ensures the cell's integrity. Selective passage of a limited number of essential compounds (e.g., nutrients) is possible either by diffusion through porin channels (e.g., OmpF and OmpA of *E. coli*) or active uptake through substrate-specific channels (e.g., TonB-dependent transporters of *E. coli*).

## 6.8. Permeability issues in whole cell catalysis

In whole-cell biocatalytic processes, the limited permeability of the cell envelope of industrially relevant microorganisms is often a major drawback causing sub-optimal product formation. The limited permeability of membranes results in inadequate intracellular substrate availability resulting in reduced reaction rates and overall productivities. Increasing the permeability of cell membranes could lead to a significant boost of productivity in whole cell catalysis (Julsing et al., 2012; Chen, 2007).

The issue of substrate permeability in whole-cell biocatalysts has been addressed in a number of ways. One strategy to alleviate the problem is the chemical pre-treatment of cells. Organic solvent (e.g., toluene, diethyl ether, chloroform for treatment of *P. rhodesiae*), detergents (e.g. Triton X-100 used for pre-treatment of *P. pseudoalcaligenes* and CTAB treatment of *Z. mobilis*), salt (e.g., NaCl used for permeabilization of *E. coli* and *R. leguminosarum*) or other chemical agents (e.g., pre-treatment of *E. coli* with EDTA, polyethylene imine treatment of the yeast *Achizosaccharomyces pombe* and Polymixin B treatment of *E.coli*) have all been successfully used as pre-treatment agents (Chen, 2007). In other studies, physical (freeze/thaw), osmotic stress or a combination of chemical and mechanical treatment have been reported to increase the overall rate of whole-cell biotransformations (Chen, 2007). These strategies, even though effective, require a multiple trials and errors and ultimately results in several extra process steps that complicate large scale production. In addition, chemical and physical pre-treatment of cells can cause excessive cell lysis or significant damage to membrane-associated systems for cofactor regeneration, and complicate product isolation.

The modulation of membrane permeability can also be achieved by varying the cell growth conditions. For example, this strategy has been applied to increase the membrane permeability of the gram-positive *R. erythropolis* (Sokolovská et al., 2003). This strategy,

however, is rather uncommon and reported efforts often resulted in only moderate increase of permeability after an extensive trial and error optimization process (McGarrity and Armstrong, 1981; Sokolovská et al., 2003).

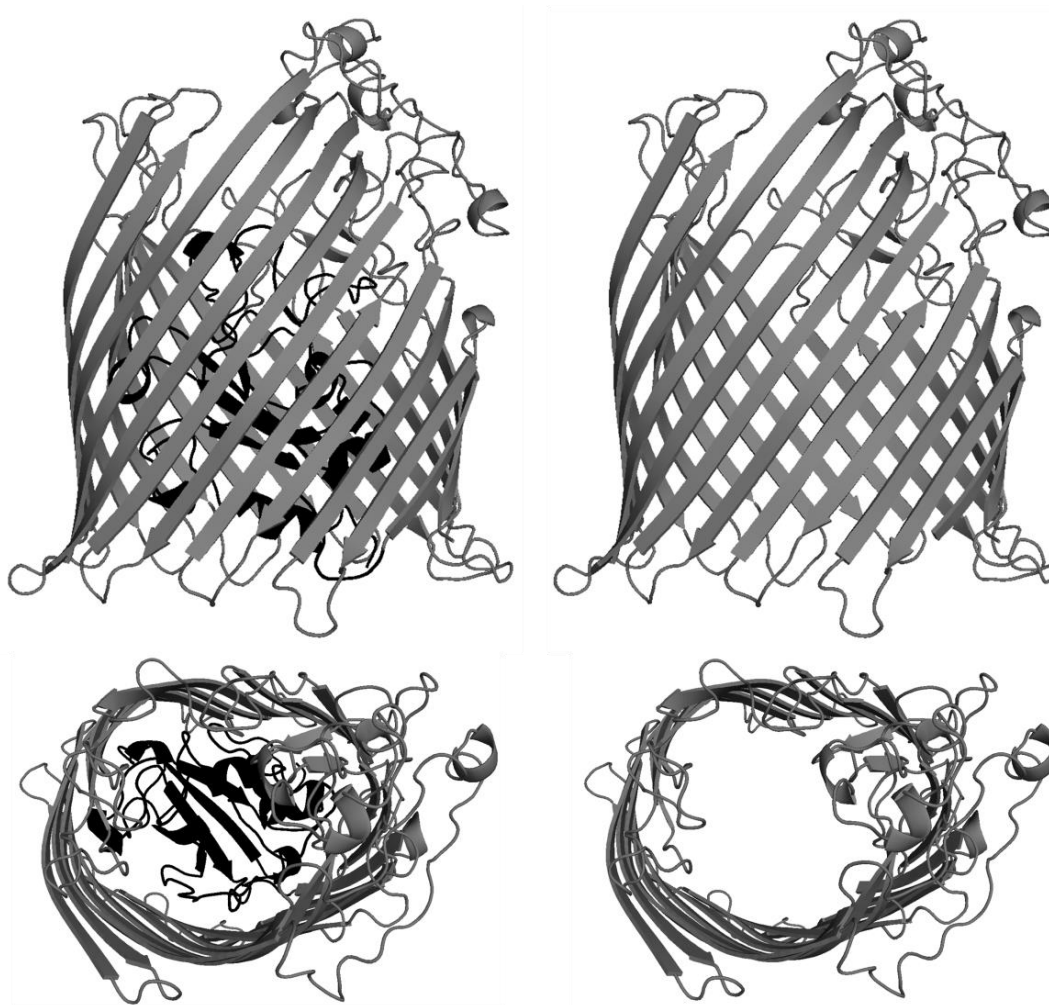
More recently, molecular engineering strategies have emerged as a more subtle and better alternative to classical approaches. For example, the permeability issue in whole cell catalysis can be circumvented by displaying enzymes on the surface of microorganisms such as *E. coli* and *P. syringae* (Chen, 2007; Sroga and Dordick, 2002). However, cell surface display can be used to present a limited number of enzymes to the cell's surface and impractical to do with cofactor-dependent enzymes. Increase in permeability of bacteria (e.g., *E. coli*) has also been achieved by genetically modifying the lipid composition of their membrane (Ni and Chen, 2004). Modifying the lipid composition may affect cells' viability and generally gives only moderate improvements; however, it also provides the possibility for combination with classical permeabilizing methods. Modifications that lead to increase in permeability can also be achieved by expressing membrane-active peptides (e.g. cinnamonin) and membrane proteins (e.g., AlkL) to the cell membrane in order to increase intracellular substrate availability (Cornelissen et al., 2013; Hofzumahaus and Schallmeyer, 2013; Julsing et al., 2012). For example, the co-expression of the outer membrane protein, AlkL, in *E. coli* enhanced the hydroxylation of octane (4-fold), nonane (40-fold), dodecanoic acid methyl ester (28-fold) (Julsing et al., 2012), and monoterpene hydroxylation rates by a factor of 2 (Cornelissen et al., 2013). AlkL protein was recently shown to be an alkane import enabling the uptake of C7-C16 n-alkanes (Grant et al., 2014). The latter concept is particularly interesting as it sets forward a general framework for applying membrane proteins in whole cell biocatalysis. In this thesis, this concept was applied to the *E.coli*'s native outer membrane protein FhuA, a mutant of which (i.e., FhuA  $\Delta$ 1-160) has been shown to form a large passive diffusion channel. In this study, the effect of co-expressing FhuA  $\Delta$ 1-160 diffusion channel in the outer membrane of *E. coli* on the whole cell biocatalysis of aromatic substrates was investigated. The P450 BM3-catalyzed conversion of the coumarin substrate, BCCE, of toluene and anisole was studied and results are presented in section 7.2. The mass transfer of substrates in whole cell biocatalysis employing P450 BM3 as catalyst is especially important to address due to the need for cofactor regeneration and the hydrophobic (and therefore non-permeating) nature of a typical P450 substrate.

## 6.9. Ferric-hydroxamate uptake protein component A (FhuA)

To satisfy their iron requirement, most bacteria have evolved diverse high-affinity acquisition systems for low molecular weight iron complexes (Clarke et al., 2001). Uptake is usually mediated by energy-dependent transporters that bind iron chelators with high affinity and mediate their uptake across the outer membrane of Gram-negative bacteria (Cornelissen, 2003). The energy required to translocate these compounds is derived from the proton motive force of the cytoplasmic membrane as transduced by the TonB-ExbB-ExbD complex (Ferguson, Coulton et al. 2000).

FhuA is a  $\beta$ -barrel transmembrane protein from *E. coli* that serves in the active transport of ferrichrome (Noinaj et al., 2010). The  $\beta$ -barrel is built up of 22 anti-parallel  $\beta$ -sheets and a globular “cork” domain (Ferguson, Hofmann et al. 1998) within the barrel structure formed by a four-stranded  $\beta$ -sheet and four short  $\alpha$ -helices (Fig. 37). The “cork” blocks most of the cross section of the barrel from the periplasmic side which ensures high specificity of the transporter for its target molecule. Ferrichrome binds via hydrogen and van der Waals contacts in a specific pocket built of 4 amino acid residues from the cork domain and 6 residues from the barrel domain slightly above the external outer membrane surface (Braun, Killmann et al. 1999). In addition to binding ferrichrome, FhuA also functions as the primary receptor for some structurally related antibiotics, for several bacteriophages (T1, T5, UC-1, and  $\Phi$ 80), for the antimicrobial peptide Microcin 25 and for the bacterial toxin Colicin M (Ferguson, Hofmann et al. 1998). Notably, FhuA could be converted into a large passive diffusion channel by deleting the globular “cork” domain (Braun et al., 2002). FhuA  $\Delta$ 1-160 can still fold into the characteristic  $\beta$ -barrel structure and insert into the outer membrane of *E. coli* to form a permanently open channel. Previous investigations of this FhuA variant have shown that its presence in the OM of *E. coli* conferred increased susceptibility to large antibiotics (Ferguson et al., 2001) and even allowed the translocation of single stranded DNA when FhuA  $\Delta$ 1-160 was incorporated into synthosomes (Nallani et al., 2006). These findings indirectly indicated that FhuA  $\Delta$ 1-160 could also positively affect mass transfer in whole cell processes.





**Figure 37.** Three dimensional structure of the outer membrane protein FhuA (PDB ID. 2FCP). The native structure including a “cork” domain (in black) and FhuA  $\Delta$ 1–160 without a “cork” are presented.

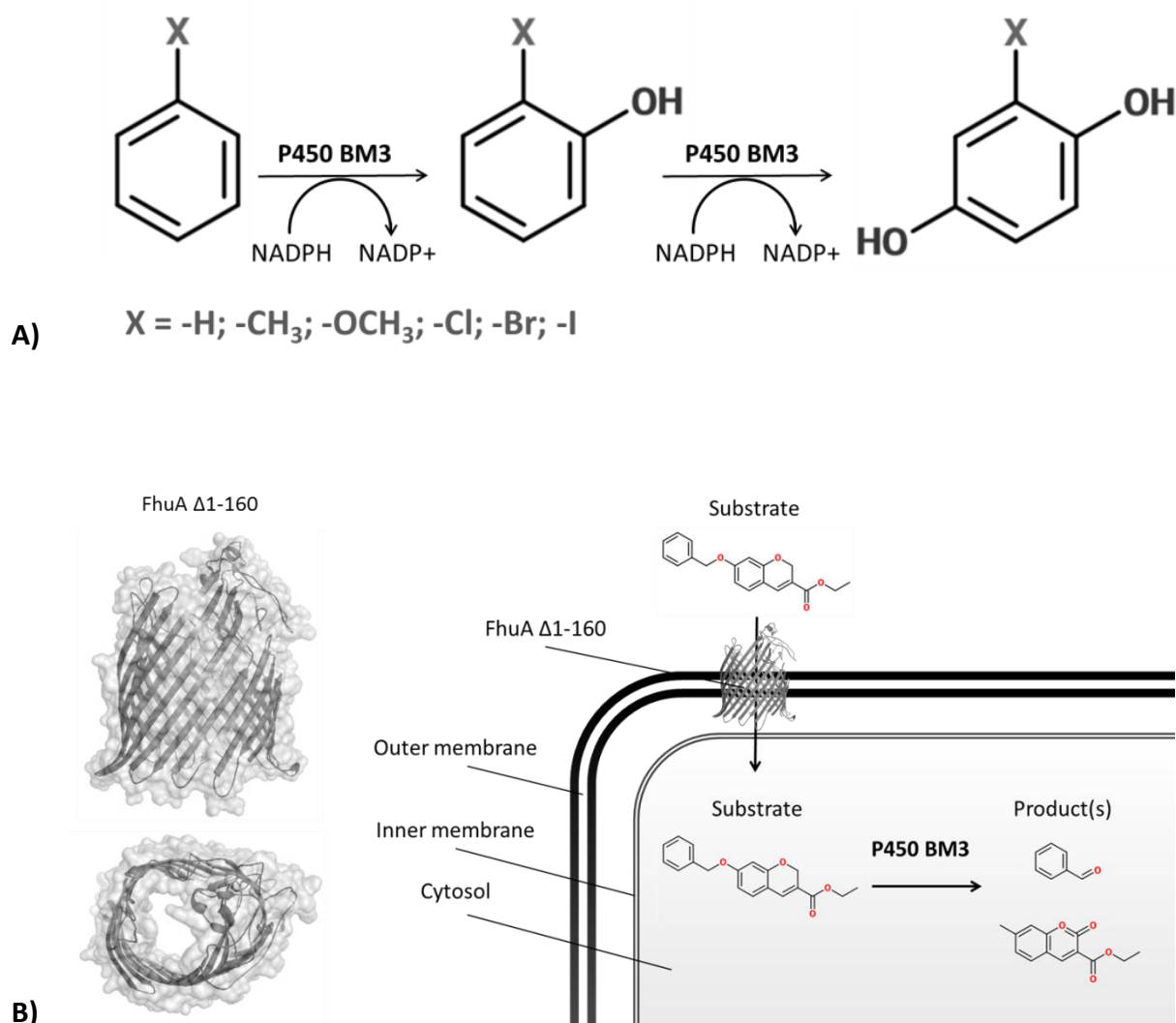
## 6.10. Aim and objectives

### 6.10.1. P450 project I: One step enzymatic synthesis of hydroquinones from monosubstituted benzenes

In P450 project I, the one step synthesis of 1,4-hydroquinones from monosubstituted benzenes by P450 BM3 was assessed. For this purpose, the P450 BM3-catalysed hydroxylation of phenols was studied first. A P450 BM3 variant M2 (R47S/Y51W/I401M) has been previously reported to carry out regiospecific, aromatic hydroxylation (Dennig et al., 2012). The initial product formation rate, coupling efficiency and total turnover number of the P450 BM3 variant M2 were determined and compared to those of P450 BM3 WT. The product formation and product profile in 24h reactions with varied substrate, the time dependent product accumulation and substrate preference of P450 BM3 M2-mediated synthesis of hydroquinones from monosubstituted benzenes were also reported. The objectives of P450 project I are schematically illustrated in Fig. 38 A.

### 6.10.2. P450 project II: Improving the rate of P450-mediated whole cell biotransformations by co-expressing a passive diffusion channel in the outer membrane of *E. coli*.

In P450 project II, one of the main drawbacks of whole cell catalysis with *E. coli*, i.e., the mass transfer across outer membrane of this microorganism was tackled. The hydroxylation of monosubstituted benzenes by P450 BM3 was used as a model reaction due to the low permeation of this class of substrates across the outer membrane of *E. coli* as well as the requirement for NADPH cofactor regeneration for the aforementioned biocatalytic reaction. The implications of the large passive diffusion channel, FhuA  $\Delta$ 1-160, on the mass transfer across the membrane barrier of *E. coli* was studied. This was achieved by comparing the performance in whole-cell biocatalysis of an *E. coli* strain co-expressing FhuA  $\Delta$ 1-160 and P450 BM3 M2 (R47S/Y51W/I401M) to the strain expressing solely the P450 BM3 catalyst. Two monosubstituted benzenes (anisole and toluene) as well as a more bulky fluorogenic P450 substrate (BCCE) were used as model substrates to test this concept. A graphical summary of this project is presented in Fig. 38 B.



**Figure 38.** Schematic summarizing the objectives and goals of the biocatalytic studies with P450 BM3 monooxygenase as a catalyst. **A)** One step synthesis of hydroquinones from monosubstituted benzenes; **B)** Improving the mass transfer across the outer membrane of *E. coli* in whole cell P450-catalyzed biotransformations.

## 7. Materials and methods

### 7.1. Chemicals

All chemicals were obtained at the highest available purity (i.e., analytical grade) from Sigma Aldrich (Steinheim, Germany), ABCR (Karlsruhe, Germany), Fluka (Neu-Ulm, Germany) and AppliChem (Darmstadt, Germany).

### 7.2. Cloning of FhuA WT and FhuA $\Delta$ 1-160 in pALXtreme-1a-P450 BM3 M2

Expression vector pALXtreme-1a harboring gene that encode P450 BM3 M2 (R47S/Y51W/I401M) and pALXtreme-1a empty vector were available from previous study (Dennig et al., 2012). Construction of pALXtreme-1a-P450 BM3 M2+FhuA  $\Delta$ 1-160 was done similarly to a previously described construct for co-expression of P450 BM3 and an alcohol dehydrogenase (Müller et al.). Both genes (encoding FhuA  $\Delta$ 1-160 and P450 BM3 M2) were placed under the control of a T7 promoter via ligase independent cloning, PLICing (Blanusa et al., 2010b) such that polycistronic mRNA containing two ribosome binding sites in front of each gene sequence was transcribed from the genetic construct. pALXtreme-1a-P450 BM3 M2 DNA fragment was amplified from pALXtreme-1a P450 BM3 M2 vector with PTO primers (**GCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTG** and **CCACTTATCCGGATGATTA CTAGTGGGTTATTACCCAGCCCACACGTC**; phosphorothioated part **in bold**) using 10 ng plasmid template, 0.4 mM of each primer, 0.2 mM dNTPs, 2.5 U *Phusion*<sup>®</sup> High-Fidelity DNA *Polymerase* and 1x *Phusion*<sup>®</sup> buffer (New England Biolabs, Frankfurt am Main, Germany) and the following thermocycling program: 25 cycles; volume: 50  $\mu$ L; 105° C heated lid) : 98° C 2 min (1x); 98° C 30 sec, 59° C 30 sec, 72 °C 2 min (24x); Hold 8° C. The DNA fragment containing the gene encoding FhuA  $\Delta$ 1-160 was amplified from pPR-IBA FhuA  $\Delta$ 1-160 plasmid (Tenne and Schwaneberg, 2012) with PTO primers (**CCGGATAAGTGG AATAAGGAGGTATACCATGGCGGTTCCAAACTGC** and **GGGCTTTGTTAGCTTATTAGAAACGG AAGGTTGCGGTTGC**; phosphorothioated nucleotides **in bold**) using 10 ng plasmid template, 0.4 mM of each primer, 0.2 mM dNTPs, 2.5 U *Phusion*<sup>®</sup> High-Fidelity DNA *Polymerase* and 1x *Phusion*<sup>®</sup> buffer (New England Biolabs, Frankfurt am Main, Germany) and the following thermocycling program: 25 cycles; volume: 50  $\mu$ L; 105° C heated lid) : 98° C 2 min (1x); 98° C 30 sec, 61° C 30 sec, 72 °C 1 min (24x); Hold 8° C. The PCR products were purified using PCR purification kit (Nucleospin Extract II kit, Macherey Nagel, Düren, Germany), ligated and

transformed as previously described (Blanusa et al., 2010b). *E. coli* DH5 $\alpha$  was the strain of choice during cloning and for propagation of the genetic constructs.

### 7.3. Shake flask expression of P450 BM3 wild type and M2

*E. coli* BL21 (DE3)  $\text{lacI}^{\text{q1}}$  transformed with pALXtreme-1a P450 BM3 WT and M2 variant were grown until culture saturation (250 rpm, 12-14 h, 37 °C) in 4 mL  $\text{LB}_{\text{Kan}}$  media. A main culture containing 50 mL TB media (10 g  $\text{L}^{-1}$  tryptone, 5 g  $\text{L}^{-1}$  yeast extract, 4 g  $\text{L}^{-1}$  glycerol, 0.1 M KPi pH 7.0) supplemented with 0.05 mg  $\text{L}^{-1}$  kanamycin and trace elements (final concentrations: 0.5 mg  $\text{L}^{-1}$   $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ , 0.18 mg  $\text{L}^{-1}$   $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ , 0.1 mg  $\text{L}^{-1}$   $\text{MnSO}_4 \cdot \text{H}_2\text{O}$ , 20.1 mg  $\text{L}^{-1}$   $\text{Na}_2\text{-EDTA}$ , 16.7 mg  $\text{L}^{-1}$   $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$ , 0.16 mg  $\text{L}^{-1}$   $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$  dissolved in ddH $_2\text{O}$ ) was prepared into a sterile Erlenmeyer flask (500 mL), inoculated with 500  $\mu\text{L}$  pre-culture and grown until  $\text{OD}_{600}$  reached  $\sim 0.6$ . P450 BM3 expression was initiated by addition of 0.1 mM IPTG, 0.5 mM aminolevulinic acid and 0.1 mg thiamine. After 24 h, cells were harvested by centrifugation (4000 x g, 15 min, 4 °C) using Eppendorf centrifuge 5810 R (Eppendorf, Hamburg, Germany). The cell pellets were stored at -20 °C.

### 7.4. Shake flask co-expression of P450 BM3 and outer membrane protein FhuA $\Delta 1$ -160

The co-expression of P450 BM3 M2 and FhuA variants was carried out as described in the previous section with minor changes. The expression strain *E. coli* BL21 (DE3)  $\text{lacI}^{\text{q1}}$  was transformed with pALXtreme-1a vectors harboring both P450 BM3 M2 gene and FhuA gene variants. Pre-cultures were grown for 8 hours and expression was terminated after 16-18 hours after induction.

### 7.5. Purification of P450 BM3

Purification of P450 was carried out as previously described (Schwaneberg et al., 1999) with minor modifications. In short, cell pellet containing expressed P450 BM3 were re-suspended in Tris-HCl buffer (0.1 M Tris-HCl, pH 7.8) and lysed by extensive sonication (40% amplitude, 30 sec on/off cycles, 15 min sonication time) using Vibracell VCX130 sonicator (Sonics & Materials, Newton, CT, USA). Cell-free extract was prepared by centrifugation (5000 x g, 20 min, 4 °C) and further clarified using a 0.45 micron filter to remove residual cell debris. For

the purification a glass chromatography column filled with anion exchange matrix (Toyopearl DEAE 650S, Tosoh Bioscience, Stuttgart, Germany) was used. During all purification steps a flow rate of 5 mL min<sup>-1</sup> was used and pressure, UV absorbance at 280 nm, flow rate, conductivity were constantly monitored. The column was equilibrated column with 10 column volumes (CV) of washing buffer (0.1 M Tris-HCl, pH 7.8) prior to sample application then the filtered cell extract was loaded on the equilibrated column using washing buffer inlet. The anion exchange matrix washed with 10 CV of washing buffer to remove proteins that did not bind to the matrix. Next, a second washing with 14 % elution buffer (1 M NaCl, 0.1 M Tris-HCl, pH 7.8) employing 10 CV; finally, P450 BM3 was eluted with 10 CV at ~24% elution buffer. Elution fractions of 3 mL were collected. Fractions containing P450 BM3 and purity >70% were pooled and desalted using a gel filtration column (PD-10 column, GE Healthcare, München, Germany) equilibrated with phosphate buffer (50 mM KPi, pH 7.5). Purified and desalted P450 BM3 protein was frozen in liquid N<sub>2</sub> and lyophilized under vacuum using freeze dryer Alpha 1-2 LD plus Christ (Osterode am Harz, Germany). Lyophilized P450 BM3 protein was stored at -20 °C until used.

## 7.6. BCCE activity assay

The assay was performed in 96-well flat bottom black microtiter plates (Greiner Bio-one, Frickenhausen, Germany) in a total volume of 200 µL. The reaction mix contained 20 µL normalized cell suspension (OD<sub>600</sub> of 10), 1 µM 7-benzyloxy-3-carboxy-coumarin ethyl ester (BCCE), 2% DMSO and 50 mM glucose in potassium phosphate buffer (KPi 50 mM, pH 7.5). The increase of fluorescence was continuously recorded over 20 sec intervals for 20 min (λ<sub>Ex</sub>: 400 nm, λ<sub>Em</sub>: 440 nm) using an Infinite M1000 microtiter plate reader (Tecan Group, Männedorf, Switzerland).

## 7.7. Flow cytometry analysis of cell populations expressing P450 BM3 and co-expressing P450 BM3 and FhuA channel protein variants

In order to examine the effect of additives on the overall morphology of cell populations, the cell suspensions were directly subjected to a FACS analysis or pre-treated with either 60 µM Polymixin B or with 1 µM BCCE and 2% DMSO for 20 min before flow cytometry analysis. The analysis was carried out by diluting the (pre-treated) cells suspensions in sterile phosphate

buffer saline (1:10 dilution) and sorting of 10 000 events using BD Influx™ cell sorter (BD Biosciences, San Jose, California). Whole cell suspensions (final OD<sub>600</sub> of 1) of cells expressing P450 BM3 M2 or co-expressing P450 BM3 BM3 and FhuA Δ1-160 were prepared in KPi buffer (50 mM, pH 7.4). The data for side-scatter (SSC) and forward-scatter (FSC) were recorded for each sample. Stringent gating conditions were selected based on the population appearance of untreated cells and the same parameters were applied to analyze the distribution of cells that were treated with permeabilizing agent (Polymixin B) or incubated with fluorogenic substrate (BCCE and DMSO).

### **7.8. Long term conversion of toluene and anisole by whole cells expressing P450 BM3 or co-expressing P450 BM3 and FhuA channel protein**

Normalized cell suspensions (OD<sub>600</sub> of 40) were prepared for each cell type in potassium phosphate buffer (50mM, pH 7.4) supplemented with 50 mM glucose. Whole cell biotransformations were carried out in 10 mL glass vials at room temperature under constant stirring with Eppendorf Mix Mate shaker (Eppendorf AG, Hamburg, Germany) using 8 mm x 4 mm stirring bar and stirring speed of 990 rpm. 5 mL normalized cell suspensions (final OD<sub>600</sub> of 40), 10 mM substrate (toluene or anisole) and co-solvent (2% DMSO). Samples (500 μL) were collected over 72 hours and extracted with equal volume of MTBE containing internal standard (20 mM guaiacol in case of reaction with toluene and 20 mM phenol in case of reaction with anisole) and finally subjected to analysis by gas chromatography. All reactions were performed in triplicate. The obtained values were corrected for small differences in optical density of cell suspensions and the P450 concentration measured for each suspension.

### **7.9. Long term conversion of benzenes or phenols to hydroquinone with purified P450 BM3**

To investigate product titers (product concentrations, TTN, time dependent product accumulation) and product profile of P450 variants long term (24 h up to 48 h) reactions with a NADPH cofactor regeneration system (glucose dehydrogenase (GDH) based) were carried out. The long term conversions were performed with purified monooxygenase in 5 mL glass vials sealed with a septum and lid. The reactions were carried out in a total volume

of 2 mL at room temperature and under constant stirring (500 rpm) and contained the following components: 1  $\mu\text{M}$  P450 BM3 WT or M2, 4 U GDH, 60 mM glucose, 2400 U catalase, 2 to 20 mM substrate, 2% DMSO, 500  $\mu\text{M}$  NADPH in phosphate buffer with ascorbic acid (50 mM  $\text{K}_2\text{HPO}_4$ , 10 mM ascorbic acid, pH 7.5). DMSO was added to the reactions with phenolic substrates which are only available in crystalline form and with low concentrations (2 or 4 mM) of benzylic substrates, in order to achieve higher reproducibility. Co-solvent (DMSO) was omitted from the reaction mixtures with high concentrations (20 mM) of benzylic substrates. Before initiating the monooxygenase reaction by addition of NADPH, the solution was always stirred for 5 min.

### 7.10. Carbon monoxide differential spectroscopy

The concentration of active P450 BM3 monooxygenase in solution was measured using CO-binding assay (Omura and Sato, 1964). A sprinkling of sodium dithionite was dissolved in P450 BM3 containing solution in 1 mL polystyrene cuvette (Carl Roth, Karlsruhe, Germany) and the baseline absorption spectrum between was recorded using a Varian Cary 50 UV spectrophotometer (Agilent Technologies, Darmstadt, Germany). Next, the same solution was treated with CO gas for around 45 seconds and the absorption spectrum (600 nm and 400 nm) was recorded again.

The concentration of active P450 BM3 monooxygenase could be calculated according to the following equation:

$$c(\text{P450\_BM3}_{\text{active}}) [\mu\text{M}] = \frac{(A_{450,\text{CO}^+} - A_{450,\text{CO}^-}) - (A_{500,\text{CO}^+} - A_{500,\text{CO}^-}) * \text{DF} * 1000 \frac{\mu\text{mol}}{\text{mmol}}}{d * \epsilon}$$

in which  $\epsilon = 91 \text{ mmol cm}^{-1} \text{ L}^{-1}$ , path length  $d=1\text{cm}$ , DF=dilution factor,  $A_{450,\text{CO}^+}$  is absorbance at 450 nm after saturation with carbon monoxide,  $A_{450,\text{CO}^-}$  is absorbance at 450 nm before saturation with carbon monoxide,  $A_{500,\text{CO}^+}$  is absorbance at 500 nm after saturation with carbon monoxide and  $A_{500,\text{CO}^-}$  is absorbance at 500 nm before saturation with carbon monoxide.



### 7.11. Determination of NADPH turnover frequency, coupling efficiency and initial turnover rate

NADPH turnover frequency for phenolic compounds by P450 BM3 variants was measured via the change of absorbance at 340 nm in the samples after addition of NADPH using Varian Cary 50 UV spectrophotometer (Agilent Technologies, Darmstadt, Germany). The reaction mix in 50 mM KPi buffer, pH7.4 as medium contained 10 mM ascorbic acid, 2mM substrate in DMSO (2% final), 1 to 2  $\mu$ M purified enzyme (depending on P450 variant and substrate used) and was prepared in UV cuvettes (Carl Roth, Karlsruhe, Germany). After 5 min incubation the solution was used as blank and absorbance measurement at 340 nm was started. When absorbance value was constant the hydroxylation reaction was initiated by addition of 0.2 mM NADPH. The final volume of reaction was 1 mL. Absorbance was measured over time until no further decrease could be observed. The enzyme concentration in each reaction was chosen so that NADPH was depleted in the time interval between 2 and 10 min after initiation. A series of reactions without substrate were prepared in the same way to determine the background absorbance change in the solution. Immediately after no further absorbance decrease was observed. Reactions were stopped by addition of 250  $\mu$ L acetonitrile which was added to 500  $\mu$ L reaction mix. The samples were centrifuged for 1 min at 13 000 x g to remove solid particles and the debris-free sample was transferred to a glass vial in preparation for product quantification by HPLC.

The NADPH turnover number (N) was calculated according to the following equation:

$$N \left[ \frac{\mu\text{M NADPH}}{\mu\text{M P450} * \text{min}} \right] = \frac{\frac{A_1 - A_0}{t_1 - t_0}}{\varepsilon_{\text{NADPH},340\text{nm}} * c(\text{P450}_{\text{BM3}}) * d}$$

in which  $\varepsilon_{\text{NADPH},340\text{nm}}=6.3*10^3 \mu\text{M}^{-1} \text{cm}^{-1}$ , path length  $d=1 \text{ cm}$ ,  $c(\text{P450}_{\text{BM3}})$  is the concentration of active P450 BM3 in  $\mu\text{M}$ ,  $A_1$  and  $A_0$  are the absorbance values at 340 nm at time points  $t_1$  and  $t_0$  in min. Time points  $t_1$  and  $t_0$  were chosen after addition of NADPH when absorbance decrease is linear and maximal.

NADPH coupling efficiency C was determined as follow:

$$C \left[ \frac{\% * \mu M \text{ product}}{\mu M \text{ NADPH}} \right] = \frac{c(\text{product})}{c(\text{depleted NADPH})} * 100 \%$$

in which c(product)= concentration of product in  $\mu\text{M}$ , and c(depleted) is the concentration of depleted NADPH in  $\mu\text{M}$ . The concentration of product in the initial solution was determined by HPLC quantification.

Product formation rates (PFR) are calculated as shown below:

$$PFR \left[ \frac{\mu M \text{ product}}{\mu M \text{ P450} * \text{min}} \right] = \frac{C * N}{100 \%}$$

## 7.12. HPLC measurement of phenols and hydroquinones

High performance liquid chromatography (HPLC) was used to analyze the formation of mono- and di-hydroxylated products from benzenes and phenols by P450 BM3 WT and M2. For this purpose samples of 500  $\mu\text{L}$  were taken from the reaction mixtures as described above, quenched with 250  $\mu\text{L}$  acetonitrile, thoroughly mixed, and centrifuged to remove insoluble particles (vortex 2 min; centrifugation 20 000 x g for 5 min). Prepared samples were separated on a NUCLEODUR® C18ec column (Macherey Nagel, Düren, Germany; flow rate of 1 mL min<sup>-1</sup>, ~68 bar pressure, 50°C) with a mobile phase consisting of 50% (v/v) ddH<sub>2</sub>O and 50% ACN (v/v). Baseline separations were obtained for all six investigated benzenes and absorbance spectra were recorded at 280 nm. Mono- and di-hydroxylated benzenes were identified with commercially available standards.

## 7.13. GC analysis

Hydroxylated products from P450 BM3 WT and M2 conversions of benzenes were quantified on the GC2010 system (Shimadzu GmbH, Duisburg, Germany) equipped with Supreme-5ms capillary column (CS-Chromatographie GmbH, Langerwehe, Germany). Products were extracted by partitioning (two-phase extraction) employing methyl tert-butyl ether (MTBE) supplemented with an internal standard (20 mM cyclododecanol). Reaction mixtures and extraction solvent were mixed at a ratio of 2:1 (vortex 2 min). Both phases were separated by centrifugation (20000 x g, 5 min, RT), the organic phase was removed and dried over

anhydrous  $\text{Na}_2\text{SO}_4$  to remove residual water. As last step, the dried organic phase was centrifuged to remove residual salts and supernatants were transferred into glass vials containing 200  $\mu\text{L}$  glass inlets. Baseline separation programs were established with commercial standards.

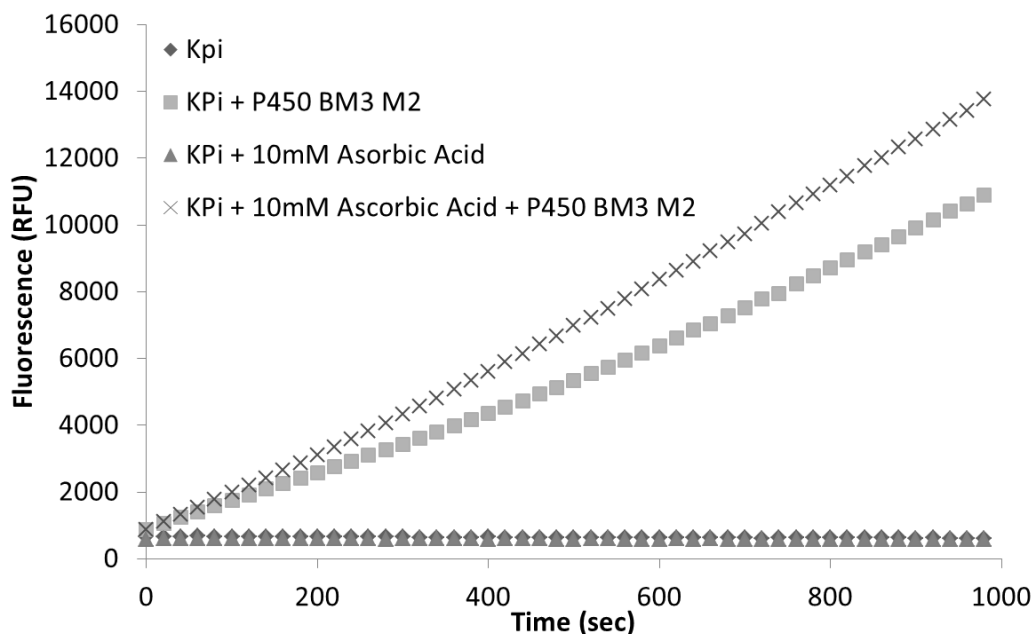
## 8. Results and discussion

### 8.1. P450 project I: One step enzymatic synthesis of hydroquinones from monosubstituted benzenes

Section 8.1 is divided in three parts that outline (I) the selection of phenolic substrates for di-hydroxylation by P450 BM3 WT and M2 and optimization of reaction conditions, (II) the production of hydroquinones from the corresponding phenols, and (III) the development of one step synthesis route to hydroquinones from the corresponding benzenes.

#### 8.1.1. Selection of phenolic substrates and reaction conditions

Six monosubstituted phenols (phenol, 2-methylphenol (o-cresol), 2-methoxyphenol (guaiacol), 2-chlorophenol, 2-bromophenol, and 2-iodophenol) were selected as suitable substrates for P450 BM3 WT and P450 BM3 M2 (R47S/Y51W/I401M). The substrate selection was made bearing in mind previous findings on selectivity and productivity P450 BM3 M2 with monosubstituted benzenes as substrates (Dennig et al., 2013). Only those o-phenols that were produced by variant M2 with selectivity exceeding 95% were chosen as substrates in the present study. In all experiments, P450 BM3 was used in purified form whereas protein purification was carried out as previously described (Dennig et al., 2012). Product quantification was performed by GC or by HPLC taking advantage of the characteristic absorbance of aromatic compounds in the middle-UV range and employing commercial standards (not available for 2-iodobenzoquinone) for product identification and quantification. The reaction buffer (50 mM KPi, pH 7.4) was supplemented with ascorbic acid (10 mM) as antioxidant to minimize the autocatalytic oxidation of the hydroquinone products by molecular oxygen. The buffer was prepared such that the optimal pH for P450 was maintained and the activity of P450 was not negatively affected. This was demonstrated in a reference experiment in which the initial conversion rate of a fluorogenic P450 BM3 substrate, 7-benzyloxy-3-carboxy-coumarin ethyl ester (BCCE), was largely unchanged by the modified composition of the buffering solution (Fig. 39).



**Figure 39.** Conversion of 7-benzyoxy-3-carboxy-coumarin ethyl ester (BCCE) by P450 BM3 M2 in 50 mM phosphate buffer, pH 7.5 and in 50 mM phosphate buffer with 10 mM ascorbic acid, pH 7.5.

### 8.1.2. P450-catalyzed production of hydroquinones from phenols as substrates

The initial product formations and cofactor coupling efficiencies of the two investigated P450 variants with phenols were determined by measuring the initial NADPH-consumption rate of reactions with different substrates and relating the latter to the amount of product formed. Total turnover numbers for phenolic compounds were calculated from 24 h reactions employing 1  $\mu$ M catalyst and 4 mM substrates.

Table 5 summarizes the catalytic performance of P450 BM3 WT and variant M2 towards hydroxylation of the phenols. In general, the product profiles obtained from conversion of phenols was similar for both investigated P450 variants. The only enzymatically-derived products detected in the experiments with phenolic substrates were the respective hydroquinones (1,4-benzenediols). In long term conversions with cofactor regeneration system some benzoquinones (oxidized hydroquinones, not exceeding 2 - 4% of the total product) were also observed. The presence of side products which were formed due to susceptibility of hydroquinones to oxidation by molecular oxygen appeared to be influenced by the ring substituents (-H, -CH<sub>3</sub>, -OCH<sub>3</sub>, -Cl, -Br, or -I) on the phenolic substrates.

The catalytic performance of WT with phenolic substrates was generally low. No NADPH depletion and product formation could be detected in the reactions with limited cofactor supply (200  $\mu$ M). The yield from 4 mM phenolic substrate in 24 h reactions with 1  $\mu$ M WT enzyme did not exceed 50% for any of the investigated substrates and was especially inefficient for the non-halogenated phenols. The WT enzyme did not hydroxylate guaiacol at all while only 1.7% and 8.8% substrate conversion in 24 h were obtained for phenol and *o*-cresol, respectively. For halogenated substrates, the highest conversion ( $\sim$ 50%) was achieved with 2-iodophenol as substrate and the lowest with 2-chlorophenol (18.8%). The conversion of 2-bromophenol (40.4%) was about twice as efficient in comparison to that of 2-chlorophenol and comparable to that of 2-iodophenol.

Similarly for P450 BM3 WT, variant M2 exhibited higher reactivity towards halogenated phenols than to non-halogenated substrates; however, in relative terms, the overall performance of variant M2 was significantly improved. NADPH depletion and products formation under limiting cofactor supply could be monitored; therefore, cofactor coupling efficiencies and initial product formation rates could be determined with the P450 BM3 mutant. The coupling efficiency of M2 lied between 39% (*o*-cresol, guaiacol) and  $\sim$ 70% (2-iodophenol). The coupling was on average 10 % higher for the halogenated phenols than for non-halogenated substrates. The product formation rates and total turnover numbers of variant M2 were significantly improved in comparison to WT. The conversion of 4 mM halogenated phenols by variant M2 was 2-fold (for 2-bromo and 2-iodophenol) up to 5-fold (for 2-chlorophenol) higher. The conversion of non-halogenated phenols by the mutant was especially improved by a factor of 6 (for phenol conversion) or higher. Notably, in contrast to WT which did not catalyze the hydroxylation of guaiacol under the employed reaction conditions, variant M2 could produce up to 1.6 mM of dihydroxylated product with the same substrate. The highest yield was achieved for 2-iodophenol ( $\sim$ 4 mM, 99% conversion). With 3.6 mM and 3.8 mM of chlorohydroquinone and bromohydroquinone, respectively formed in 24 hours, variant M2 proved to be equally efficient catalyst for hydroxylation of halogenated phenols.

**Table 5.** Catalytic performance of P450 BM3 WT and variant M2 towards selected phenols.

P450 BM3 Variant	Substrate	N <sup>1</sup>	C [%] <sup>2</sup>	PFR <sup>3</sup>	TTN <sup>4</sup>
WT	phenol	n.d. <sup>a</sup>	n.d. <sup>a</sup>	n.d. <sup>a</sup>	68 ± 16
M2		23.3 ± 2.6	n.d. <sup>a</sup>	n.d. <sup>a</sup>	353 ± 62
WT	o-cresol	n.d. <sup>a</sup>	n.d. <sup>a</sup>	n.d. <sup>a</sup>	121 ± 20
M2		23.3 ± 1.8	38.6 ± 1.3	9.0	1875 ± 138
WT	guaiacol	n.d. <sup>a</sup>	n.d. <sup>a</sup>	n.d. <sup>a</sup>	n.d. <sup>a</sup>
M2		20.5 ± 2.1	39.5 ± 4.4	8.1	1580 ± 411
WT	chlorophenol	n.d. <sup>a</sup>	n.d. <sup>a</sup>	n.d. <sup>a</sup>	753 ± 97
M2		60.8 ± 5.1	47.2 ± 3.0	28.7	3686 ± 219
WT	bromophenol	n.d. <sup>a</sup>	n.d. <sup>a</sup>	n.d. <sup>a</sup>	1615 ± 327
M2		107.9 ± 10.6	53.5 ± 4.2	57.7	3832 ± 366
WT	iodophenol	39.1 ± 6.5	n.d. <sup>a</sup>	n.d. <sup>a</sup>	ca. 2000 <sup>b</sup>
M2		370.8 ± 25.9	ca. 70 <sup>b</sup>	ca. 260 <sup>b</sup>	ca. 4000 <sup>b</sup>

<sup>1</sup> N = NADPH turnover frequency [ $\mu\text{mol NADPH } \mu\text{mol P450}^{-1} \text{ min}^{-1}$ ];

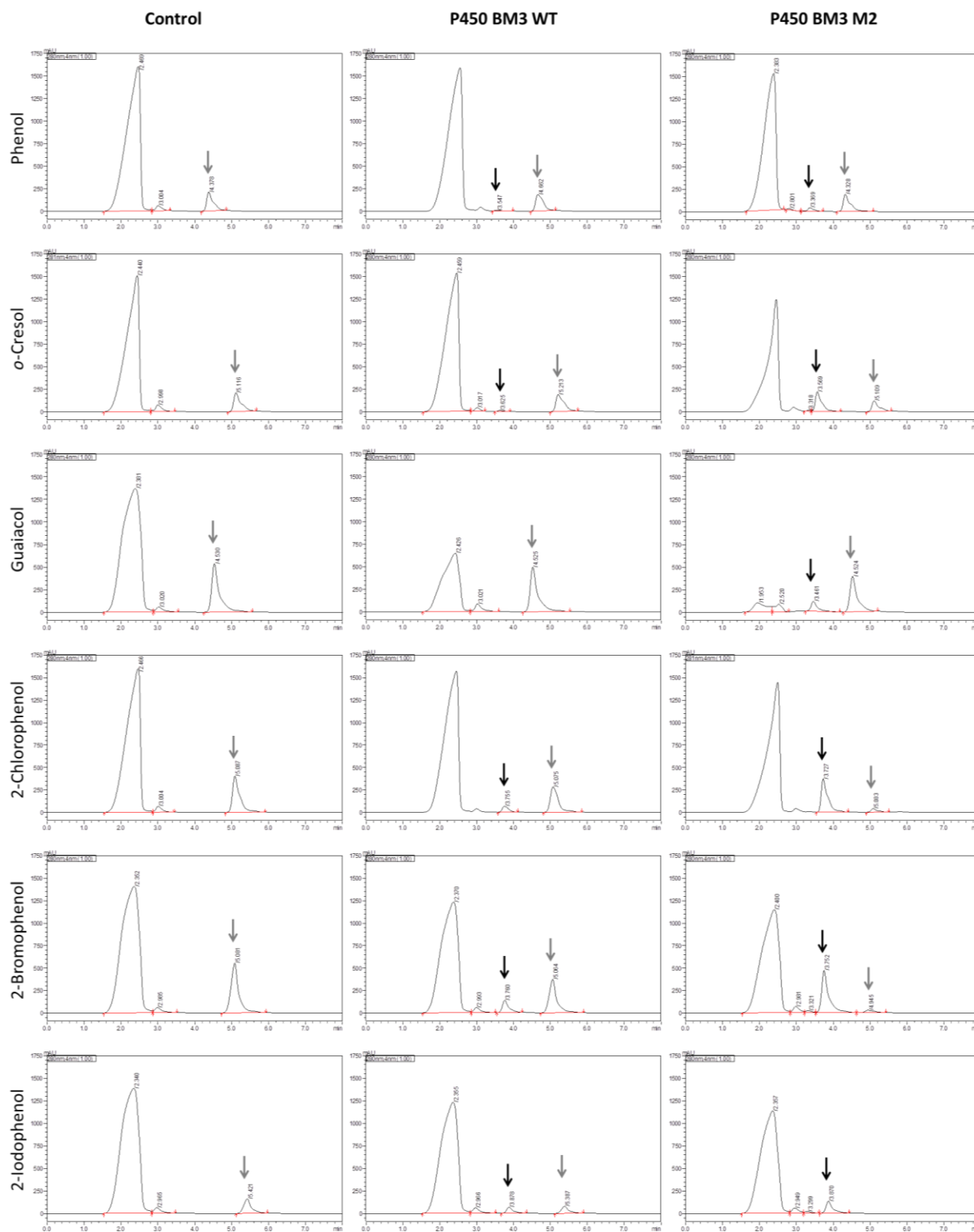
<sup>2</sup> C [%] = coupling efficiency [ $\mu\text{mol}_{\text{product}} \mu\text{mol}_{\text{NADPH}}^{-1}$ ]\*100;

<sup>3</sup> PFR = initial product formation rate [ $\mu\text{mol}_{\text{product}} \mu\text{mol P450}^{-1} \text{ min}^{-1}$ ];

<sup>4</sup> TTN = total turnover frequency [ $\mu\text{mol}_{\text{product}} \mu\text{mol}_{\text{P450}}^{-1} \text{ 24 h}^{-1}$ ]; 24 hour reactions using 1  $\mu\text{M}$  enzyme and 4  $\mu\text{M}$  substrate;

<sup>a</sup> not detected;

<sup>b</sup> estimated value calculated with a calibration curve for hydroquinone due to unavailability of commercial standard iodohydroquinone;

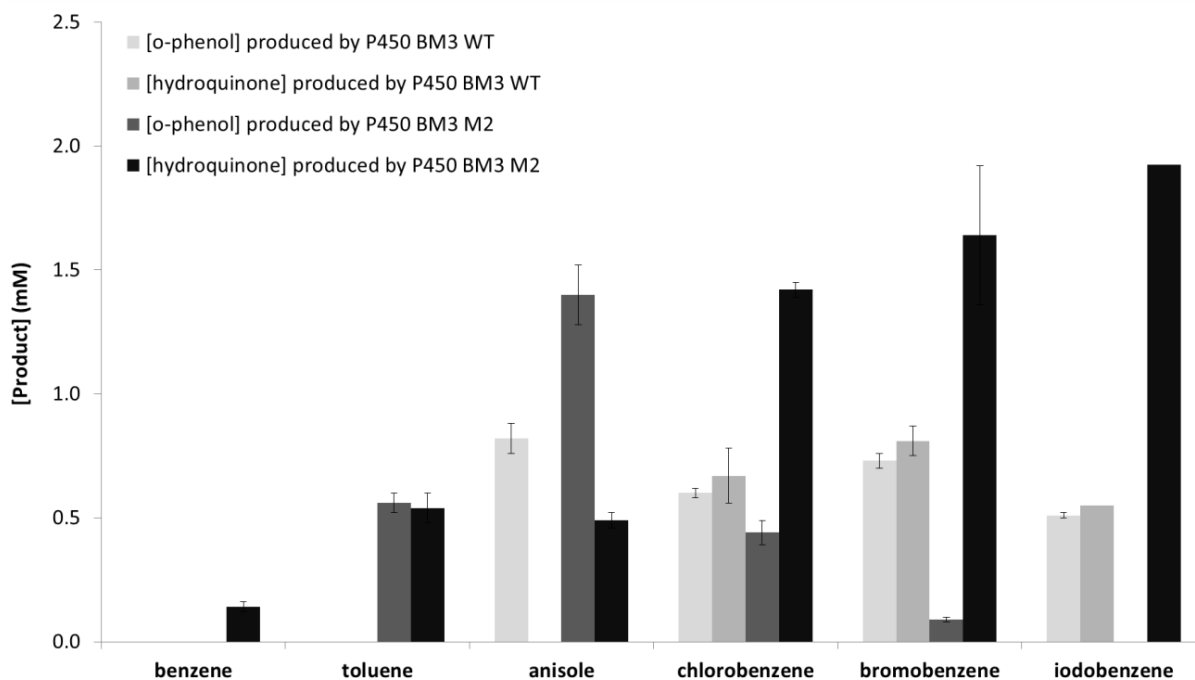


**Figure 40.** HPLC chromatograms after 24 h conversion of phenolic compounds to hydroquinones using P450 BM3 WT and variant M2. In the leftmost column chromatograms of negative control reactions (without enzymes) are shown; the middle column contains the chromatograms obtained with P450 BM3 WT; the rightmost column, the chromatograms with P450 BM3 M2. The large peak at ca. 2.3 and small peak at ca. 3 min are derived from reaction mix components (ascorbic acid and NADP<sup>+</sup>) which absorb light in the middle-UV range. Substrate (phenols) and product (hydroquinone) peaks are highlighted by light grey and black arrows, respectively.



### 8.1.3. One step synthesis of hydroquinones from monosubstituted benzenes

The production of hydroquinones from benzenes by P450 BM3 WT and variant M2 was evaluated in long term conversions with cofactor regeneration system. The reactions were at first carried out for 24 h with 4 mM substrates and 1  $\mu$ M enzyme. Fig. 41 summarizes the product concentrations measured only for the 2 main products produced by P450 BM3. The concentrations of *o*- and 1,4-dihydroxylated products are plotted while side products identified either as *p*-hydroxylated phenols ( $\leq 5\%$  of the total product, in accordance with previous findings (Dennig et al., 2013)) are not included in the graph. P450 BM3 WT did not convert benzene and toluene at a detectable level and produced *o*-cresol (0.82 mM) from anisole as a sole product. The halogenated benzenes were converted by WT to both hydroxylated products and achieved nearly equal concentrations of each product (from 0.55 to 0.8 mM). Variant M2 could hydroxylate all six substrates and generate both mono- and dihydroxylated products, the only exception being benzene for which only hydroquinone product (0.14 mM) was detected. The highest concentrations of hydroquinones (between 1.3 and 2 mM) were achieved with halogenated benzenes as substrates, especially with bromo- and iodobenzenes. The latter were converted by variant M2 almost exclusively to dihydroxylated product. Halophenol could be detected in the samples with the chlorinated and brominated benzenes (0.44 mM and 0.1 mM, respectively) but not with iodobenzene. In contrast, the P450 mutant converted anisole mainly to the corresponding cresol (1.4 mM) and produced 0.5 mM methylhydroquinone. In the reactions with toluene about 0.5 mM of mono- and dihydroxylated was measured.



**Figure 41.** Product titers of the catalytic cascade [benzenes → corresponding phenols → corresponding hydroquinone] generated by P450 BM3 WT (1  $\mu$ M) or P450 BM3 M2 (1  $\mu$ M) in 24-hour reactions (4 mM benzylic substrates, 2 vol% DMSO) and NADPH cofactor supplied via a cofactor regeneration system (60 mM glucose and 4 U glucose dehydrogenase). Light grey bars show the amount of phenols and hydroquinones generated from benzenes by P450 BM3 WT in one step reactions. Dark grey and black bar show the amount of phenols and hydroquinones generated from benzenes by P450 BM3 M2 in one step reactions. Hydroquinone oxidation products which accounted for less than 5% and *p*-hydroxylated phenols from benzenes which were produced in small amounts from iodobenzene ( $\leq$ 5%) are not included in the graph. Quantification was performed by HPLC in all reactions and product formation of iodohydroquinone is calculated using a calibration curve for benzohydroquinone.

In a second set of experiments the product titers and product profile generated by P450 BM3 M2 with 2 non-halogenated benzenes (toluene and anisole) and two halogenated benzenes (2-chlorobenzene and 2-bromobenzene) at a low (2 mM) and high (20 mM) substrate concentrations were assessed (Table 6). Halogenated benzenes were predominantly converted to the respective hydroquinone products irrespective of substrate load. The concentration of phenolic products was at least twice as low as that of hydroquinone. The increase of substrate concentration in the reaction maintained the product profile and did not lead to a significant change in titers with these substrates. The long term reactions with non-halogenated substrates rendered more phenolic product than dihydroxylated product. The increase of substrate concentration increased the product amounts of the reaction and changed the obtained product profiles with non-halogenated

substrates. In long term reactions with anisole as substrate, the predominantly formed product was the respective *o*-phenol (*o*-cresol) exceeding the concentration of hydroquinone (guaiacol) 2 to 4 times depending on the employed substrate concentration. The increased concentration of non-halogenated substrates increased the overall product titers ~8-fold. The substrate concentration in long term conversion of toluene affected the product concentrations (ca. 3-fold increase) and the product profile as well. At substrate concentration of 2 mM, *o*-cresol and methylhydroquinone were produced at nearly equimolar concentration; however, the increase of substrate to 20 mM shifted the ratio of products to 2:1 in favor of the phenolic compound.

**Table 6.** Catalytic performance of P450 BM3 M2 at 2 different concentrations of benzenes (2 and 20 mM) and NADPH cofactor provided by a cofactor regeneration system (60 mM glucose and 4 U glucose dehydrogenase). Co-solvent DMSO (2 vol%) was only used in the reactions with 2 mM but not in the reactions with substrate load of 20 mM. Products were quantified after 24 h conversion by GC (not HPLC) employing commercial standards for product identification. (data collected in a collaboration work; used with permission from (Dennig, 2014).

Substrate	Substrate concentration (mM)	Yield (%)	Ph:HQu (%)	Product concentration (g L <sup>-1</sup> )
Toluene	2	62	41:59	0.14
	20	21	69:31	0.47
Anisole	2	46	71:29	0.11
	20	33	81:19	0.83
2-chlorobenzene	2	65	18:82	0.15
	20	9	24:76	0.24
2-bromobenzene	2	54	24:76	0.26
	20	16	27:73	0.32

Ph = *o*-phenol; HQu = hydroquinone;

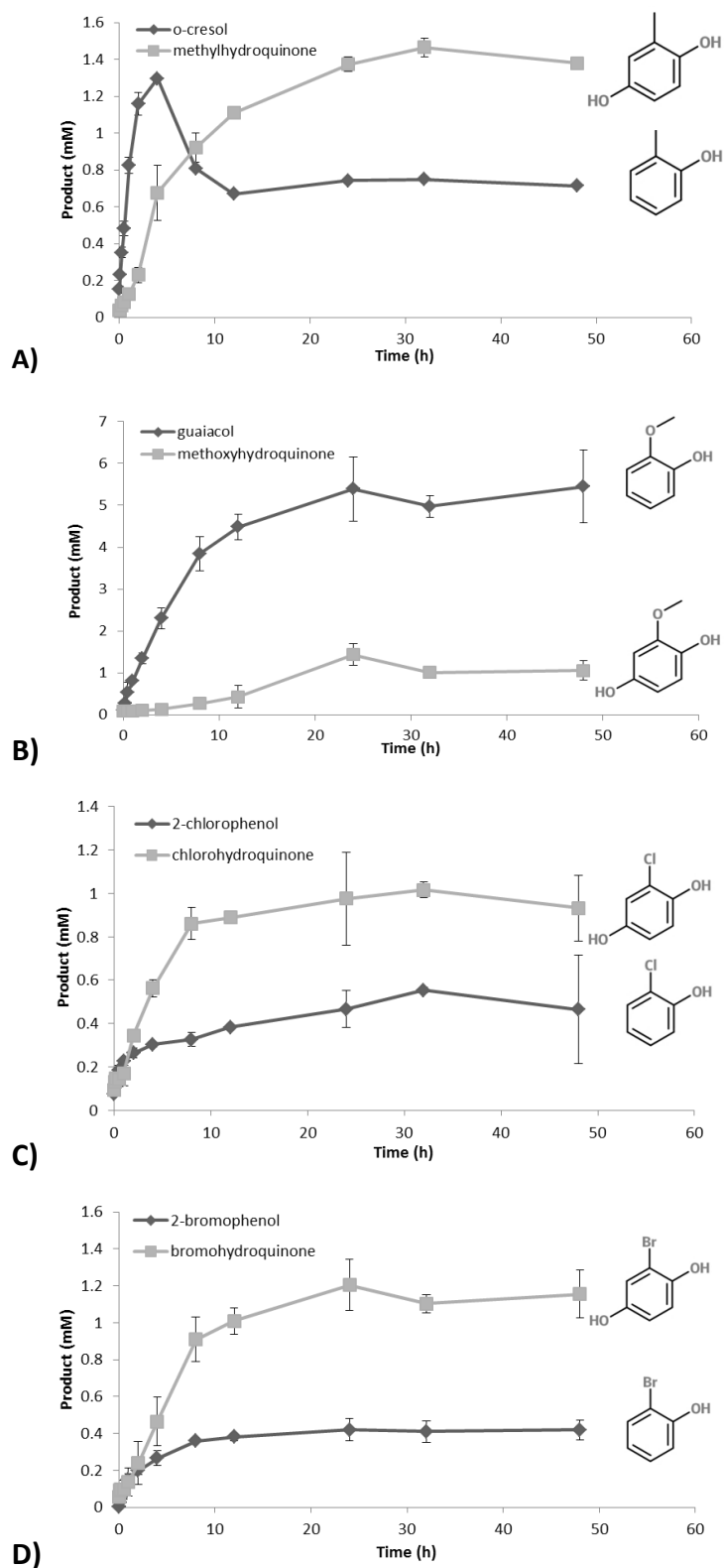
Time dependent conversion of toluene, anisole, chlorobenzene and bromobenzene was performed in order to monitor the switch in substrate preference from benzene to phenol for variant M2. In these reactions an increased substrate load of 20 mM was used .The

product profile is exemplified with toluene, anisole and chlorophenol as substrates (Fig.3) The conversion of toluene exhibited formation of two distinct products, *o*-cresol and methylhydroquinone (Fig 3 A). *o*-Cresol was predominantly formed within the first 4 hours reaching the maximum of 1.3 mM. After this time point the concentration of *o*-cresol decreased and remained steady at 0.65 mM until end of the reaction. The concentration of methylhydroquinone increased to around 1.3 mM in 24 h and remained constant until the end of the reaction. These product concentrations correspond to 0.07 g L<sup>-1</sup> *o*-cresol and 0.16 g L<sup>-1</sup> methylhydroquinone. The observed product profiles are different that those measured by GC in 24 hour reactions and could be due to evaporation of the highly volatile benzylic substrate (toluene).

The time dependent product formation employing anisole as substrate revealed that in the first 8 h guaiacol is produced reaching a concentration of ca. 4 mM product (Fig. 3 B). After 8 h reaction time enough guaiacol is produced so that variant M2 could start producing methoxyhydroquinone. The highest product concentration was reached after 24 h with 5.4 mM guaiacol and 1.4 mM methoxyhydroquinone that corresponds to 0.67 g L<sup>-1</sup> guaiacol and 0.2 g L<sup>-1</sup> methoxyhydroquinone.

In the conversion of chlorobenzene, chlorohydroquinone was the predominantly formed product. A product ratio of 2:1 (hydroquinone to phenol) was observed after 48 h (Fig. 3B). The final product concentration after 24 h reached 0.55 mM 2-chlorophenol and 1 mM chlorohydroquinone. The concentration of chlorohydroquinone exceeded the amount of 2-chlorophenol after only two hours reaction time. The formation of the respective chlorohydroquinone started after 1 h when the concentration of 2-chlorophenol was 0.23 mM. After 48 h a product concentration of 0.22 g L<sup>-1</sup> was obtained. Conversion of 2-bromobenzene displayed a similar product formation profile to chlorobenzene, i.e., product ratios and concentrations for bromohydroquinone and 2-bromophenol.

The total product concentrations obtained for variant M2 and 20 mM of the respective substrates showed the following order: anisole (6.8 mM) > toluene (2.1 mM) > bromobenzene (1.6 mM) > chlorobenzene (1.5 mM). The highest concentration for *o*-phenols was obtained using anisole (5.4 mM) as substrate. In case of other substrates, the *o*-phenolic product did not exceed concentrations of 0.7 mM.



**Figure 42** Conversion of **A)** toluene, **B)** anisole, **C)** chlorobenzene and **D)** bromobenzene by P450 BM3 M2 employing a cofactor regeneration system (60 mM glucose and 4 U glucose dehydrogenase). Time-dependent formation of phenols and hydroxyquinones are quantified by HPLC analysis over a period of 48 h. 1  $\mu$ M P450 BM3 M2 with a substrate load of 20 mM without co-solvent (DMSO) was employed in all conversions.(data generated in collaboration work; used with permission from (Dennig, 2014).

#### **8.1.4. Summary and conclusions: One step enzymatic synthesis of hydroquinones from benzenes**

To generate a deeper insight into the operational performance and synthetic potential of P450 BM3 WT and variant M2, parameters such as regioselectivity, total turnover, and product concentrations for a selection of phenolic substrates were investigated in this study. In addition, to access information on time dependent product formation and the effect on substrate selectivity, the catalytic performance of variant M2 with benzylic substrates was analyzed by monitoring the formation of phenol and hydroquinone for 48 h. The engineered P450 BM3 variant demonstrated higher product formations than WT for all tested substrates. This could be attributed to the increased coupling efficiency and initial activity compared to P450 BM3 WT for both phenolic compounds and benzylic substrates. Variant M2 was very efficient in hydroxylating halogenated phenols (chloro-, bromo- and iodophenol) to hydroquinones with selectivities exceeding 99% and conversions higher than 90%. The conversion of non-halogenated phenols did not exceed 50% but was still performed with excellent regioselectivities. The good conversion of halogenated phenols could be partially attributed to their better solubility in aqueous buffer. The long term (24 h) conversion of benzenes by the P450 BM3 M2 increased product concentrations up to 0.83 g L<sup>-1</sup> and rendered different product profiles depending on the employed substrate (halogenated vs non-halogenated) and, in case of non-halogenated benzenes, depending on substrate concentration. Long term conversion of halogenated substrates with the P450 mutant produced predominantly hydroquinones irrespective of the employed substrate load. However, the increase of substrate concentration did not lead to a significant increase of product formation and decreased yields. In addition, from the time-dependent product formation experiments, it was observed that most product formation took place within the first 12 h of reaction indicating inhibition effect of halogenated benzenes on the P450 catalyst at high substrate load. In contrast, in case of non-halogenated substrates, the substrate load affected productivity and in case of toluene, the obtained product profile as well. In case non-halogenated substrates, the product concentrations significantly increased at high substrate concentrations but yields were negatively affected. The product profile in the case of toluene was shifted with the increase of substrate concentration. The predominant product formed at low concentration was benzohydroquinone; however

phenol formation was favored at 10-times higher substrate concentration. This, however, might be an artefact from the experimental design; samples were withdrawn from the same vessel multiple times and each opening likely lead to loss of toluene through evaporation. *o*-Cresol was predominantly produced from anisole at low and high substrate load; the product ratios were not significantly affected. Some reasons for the observed substrate selectivity of variant M2 are the varying substrate concentrations, but also the pKa values and difference in electronegativity of the phenolic substituents.

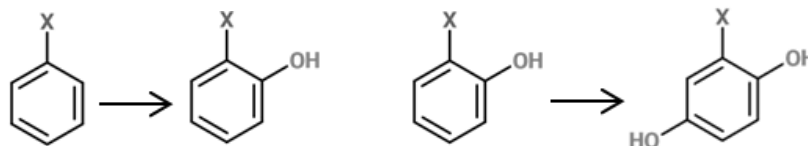
Comparison of chemical and biocatalytic production of mono- and di-hydroxylated aromatic compounds from benzenes shows that the process described in this study has several advantages over chemical synthesis (Table 7). Firstly, P450 BM3 M2-mediated synthesis minimizes the need for toxic chemicals. Chemical hydroxylation of benzenes requires large amounts of corrosive reactants (e.g., caustic soda, sulfuric acid, hydrochloric acid), metal catalysts (e.g., copper/iron catalyst) or strong oxidizing agents (e.g. H<sub>2</sub>O<sub>2</sub>, K<sub>2</sub>S<sub>2</sub>O<sub>8</sub> and Ag<sub>3</sub>PO<sub>4</sub>). In P450 BM3-mediated synthesis organic co-solvent (e.g., 2% DMSO) can be added to the reaction mixtures in order to increase the substrate solubility. However, co-solvent addition was necessary only for crystalline substrates (i.e., phenolic substrates). Conversions of liquid substrates (i.e., benzylic compounds) at 20 mM substrates load (~2 g L<sup>-1</sup>) was performed in a solvent-free aqueous solution and rendered good overall product titers, e.g., with the poorly water-soluble anisole, 0.83 g L<sup>-1</sup> total product was formed. Secondly, P450 BM3-catalyzed phenolic hydroxylation is advantageous with regard to reaction conditions and energy efficiency. Individual steps of the chemical synthesis routes may require temperature control (cooling) or high temperature (up to 400 °C) and high pressure (up to 70 bar). Monooxygenase hydroxylation of benzenes is performed under mild reaction conditions (room temperature, atmospheric pressure) which translate into significant energy cost reduction of the overall process. Thirdly, P450 BM3-mediated production of hydroquinone is carried out in a rigorous manner with excellent regioselectivity for mono- (≥95%) and di-hydroxylation (≥99%). Chemical synthesis of hydroquinones proceed with side product formation (1,2- and 1,3-dihydroxylated isomers). P450 BM3 produces exclusively 1,4-benzenediols which subsequently simplifies downstream processing and reduces expenses and losses in separation of ortho- and para-isomers(Hudnall, 2000)(Krumenacker et al., 2000). Fourthly, since P450 BM3-catalyzed di-hydroxylation is a one step process, the

need to isolate reaction intermediates is eliminated and the whole process is intensified in comparison to classical syntheses routes. In the fifth place, P450 BM3 is a promiscuous catalyst which as demonstrated in the present study generates mono- and di-hydroxylated products from a variety of monosubstituted benzenes. Chemical synthesis routes are applicable to a narrow set of substrates, e.g., hydroperoxidation of diisopropylbenzene can be applied solely to benzohydroquinone synthesis; halogenated hydroquinones are not synthesized from halogenated benzenes rather via halogenation of benzoquinone.

In summary, the possibility to produce hydroquinones from benzene educts using P450 BM3 allows a new and attractive route to building blocks for vitamins (tocopherols), antioxidants, pharmaceuticals and polymerization inhibitors (Netscher, 2007; Rappoport, 2004). P450 BM3 M2 is an excellent catalyst for the one step double oxidation of mono-substituted aromatic hydrocarbons and offers opportunities to explore new synthetic routes to the production of 1,4-dihydroxylated products, especially of methyl-, chloro-, bromo- and iodohydroquinone. P450 BM3 M2 performs the double hydroxylation as single catalyst with an excellent regioselectivity. Thereby, the oxidation of benzenes rather than phenols becomes the rate limiting step in the one step synthesis of hydroquinones. Nevertheless, in order to bring this process to closer an industrial scale, productivity has to further be improved and whole cell cofactor regeneration systems have to be developed.



**Table 7.** Comparison of chemical and enzymatic processes for synthesis of phenols and hydroquinones



Name (commercial: Yes/No)	Raw materials/ Catalyst	Products (selectivity %)	% Yield	Raw materials/ Catalyst	Products (selectivity %)	% Yield	Comments/ Reference
<b>Hock Process (Yes)</b>	Benzene, propylene, radical initiator, H <sub>2</sub> SO <sub>4</sub> , H <sub>3</sub> PO <sub>4</sub> , NaOH, temperature control, heat, pressure	Phenol (14 %), acetone (8 %)	5 % (on benzene)	-	-	-	Multi-step process; 90% of world's phenol production Ref. (Zakoshansky, 2007)
<b>Direct conversion of benzene to phenol using VCl<sub>3</sub> in the two-phase system (No)</b>	Benzene, benzonitrile, water, air, VCl <sub>3</sub>	Phenol (98.1 %)	3.6 % (on benzene)	-	-	-	Not suited for practical applications Ref. (Battistel et al., 2003)
<b>Direct oxidation of benzene using Pt/Al<sub>2</sub>O<sub>3</sub>/V(acac)<sub>3</sub> catalysts (No)</b>	Benzene, air, Pd/Al <sub>2</sub> O <sub>3</sub> /V(acac) <sub>3</sub> catalyst, acetic acid, pressure	Phenol (99 %)	26 % (on benzene)	-	-	-	Ref. (Jiang et al., 2013)
<b>Direct hydroxylation of benzene using zeolites or acidic resins (No)</b>	Benzene, air, ethanol, water, Nafion/silica composites	Phenol (56 %)	4.6 % (on benzene)	-	-	-	Ref. (Laufer and Hoelderich, 2002)
<b>Direct hydroxylation of benzene by the use of palladium membrane (No)</b>	Benzene, air, Pd surface, heat	Phenol (80-90 %)	2-16 % (on benzene)	-	-	-	Ref. (Niwa et al., 2002)
<b>Direct hydroxylation of benzene with nitrous oxide and ZSM-5 zeolites (No)</b>	Benzene, air, N <sub>2</sub> O, ZSM-5 zeolites, heat	Phenol (98 %)	26 % (on benzene)	-	-	-	Highly pure N <sub>2</sub> O needed; rapid catalyst deactivation Ref. (Ebner et al., 1999)
<b>Direct oxidation of benzene to phenol by hydrogen peroxide and redox-active metals (No)</b>	Benzene, air, H <sub>2</sub> O <sub>2</sub> , iron catalyst, Fe(II) complex, trifluoroacetic acid	Phenol (97 %)	8.6 % (on benzene)	-	-	-	Ref. (Bianchi et al., 2000)
<b>Direct hydroxylation of benzene by the use of titanium-containing molecular sieves (No)</b>	Benzene, air, H <sub>2</sub> O <sub>2</sub> , titaniumsilicate catalyst (TS-1)	Phenol (99 %)	13.3 % (on benzene)	-	-	-	Ref. (Remias et al., 2003)

CHAPTER II. Synthesis of Mono- and Di-hydroxy Benzenes using P450 BM3 Monooxygenase  
P450 Project I – Summary and Conclusions

<b>Direct oxidation of benzene to phenol by hydrogen peroxide and oxovanadium catalyst (No)</b>	Benzene, air, H <sub>2</sub> O <sub>2</sub> , oxovanadium Schiff base moieties/silica, acetonitrile	Phenol (99 %)	30.8 % (on benzene)	-	-	-	Ref. (Reddy et al., 1992)
<b>Hydroperoxidation of diisopropylbenzene (Yes)</b>	-	-	-	Diisopropyl benzene, radical initiator, H <sub>2</sub> SO <sub>4</sub> , H <sub>3</sub> PO <sub>4</sub> , NaOH, heat, temperature control	Hydroquinone (94 %)	62 % (on diisopropyl benzene)	Multi-step process; 60 % of world's hydroquinone production Ref. (Hudnall, 2000)
<b>Electrochemical hydroxylation of aromatic substrates (No)</b>	-	-	-	Phenol, acid, organic solvent, electric current	Hydroquinone (70 %)	Up to 34 % (on phenol)	Ref. (Rautenbach, 2007)
<b>Hydroxylation of phenol with hydrogen peroxide using titanium-containing molecular sieves (Yes)</b>	-	-	-	Phenol, H <sub>2</sub> O <sub>2</sub> , TS-1 catalyst	Hydroquinone (up to 50 %)	92 % (on phenol)	Ref. (Perego et al., 2001)
<b>Gas phase oxidation of phenol in gas phase with nitrous oxide and ZSM-5 zeolites (No)</b>	-	-	-	Benzene, N <sub>2</sub> O, ZSM-5 zeolites, heat	Hydroquinone (up to 90 %)	76 % (on N <sub>2</sub> O)	Ref. (Ivanov et al., 2002)
<b>Aromatic peroxygenase (APO) mediated synthesis (No)</b>	Benzene, H <sub>2</sub> O <sub>2</sub> , AaeAPO	Phenol (n.r.)	n.r.	Phenol, H <sub>2</sub> O <sub>2</sub> , AaeAPO	Hydroquinone (n.r.), catechol (n.r.), resorcinol (n.r.), trihydroxy benzene (n.r.),	n.r.	Ref. (Karich et al., 2013)
<b>Mixed microbial and chemical synthesis of hydroquinone (No)</b>	-	-	-	Glucose, <i>E. coli</i> QP1.1, K <sub>2</sub> S <sub>2</sub> O <sub>8</sub> , Ag <sub>3</sub> PO <sub>4</sub> , (NH <sub>4</sub> ) <sub>2</sub> Ce(SO <sub>4</sub> ) <sub>3</sub>	Hydroquinone (99 %)	51 % (on glucose)	Multi-step process Ref. (Ran et al., 2001)
<b>P450 BM3 M2 catalyzed hydroxylation of non-halogenated benzenes (No)</b>	Benzene, NADPH, P450 BM3 M2	Phenol (99 %)	-	Phenol, NADPH, P450 BM3 M2	Hydroquinone (99 %)	Up to 33 % (on benzene) Up to 46 % (on phenol)	This study
<b>P450 BM3 M2 catalyzed hydroxylation of halogenated benzenes (No)</b>	Halogenated benzene, NADPH, P450 BM3 M2	Phenol (95 – 99 %)	-	Halogenated phenol, NADPH, P450 BM3 M2	Hydroquinone (99 %)	Up to 50 % (on benzene) Up to 99 % (on phenol)	This study

n.r. = not reported

## 8.2. P450 project II: Improving the rate of P450-mediated whole cell biotransformations by co-expressing a passive diffusion channel in the outer membrane of *E. coli*

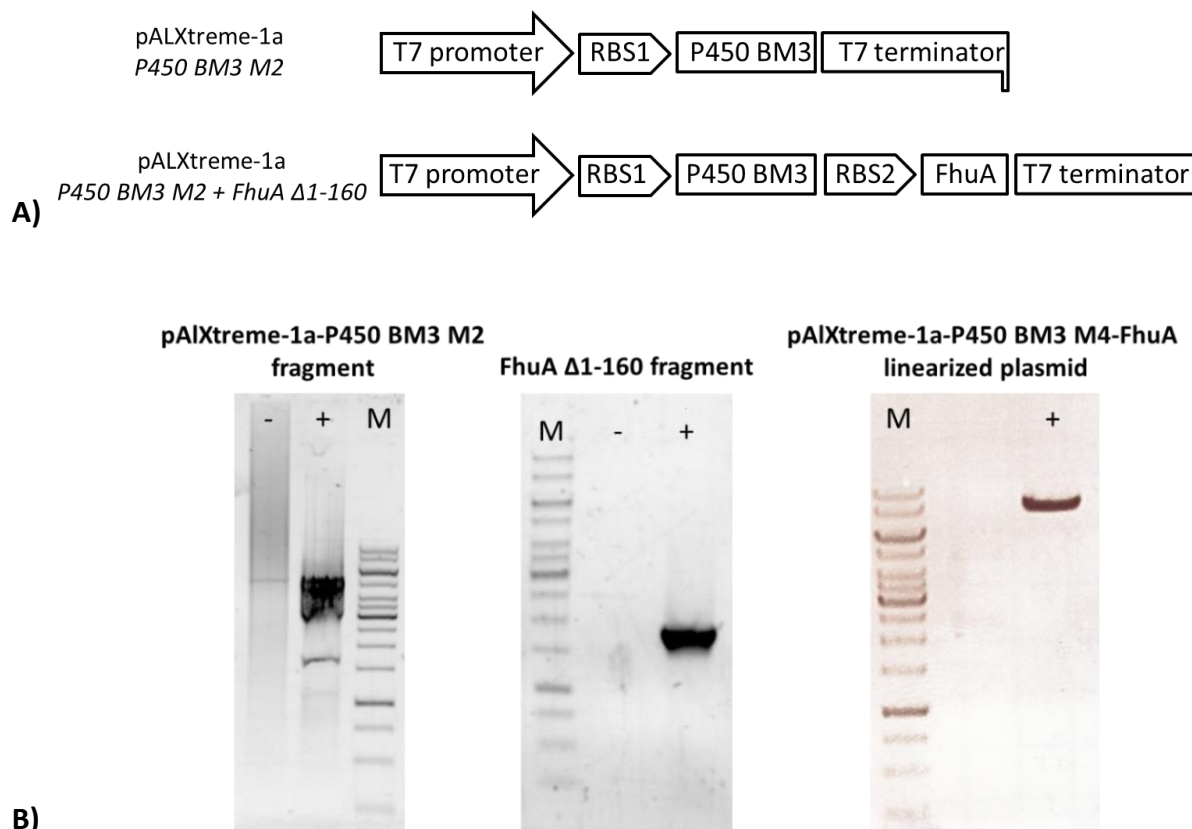
This section outlines I) the generation of genetic construct for co-expression of P450 BM3 M2 and FhuA  $\Delta$ 1-160, II) the conversion of the model substrate BCCE and III) the whole cell conversion of toluene and anisole with cells expressing either P450 BM3 or cells co-expressing a channel protein in their outer membrane.

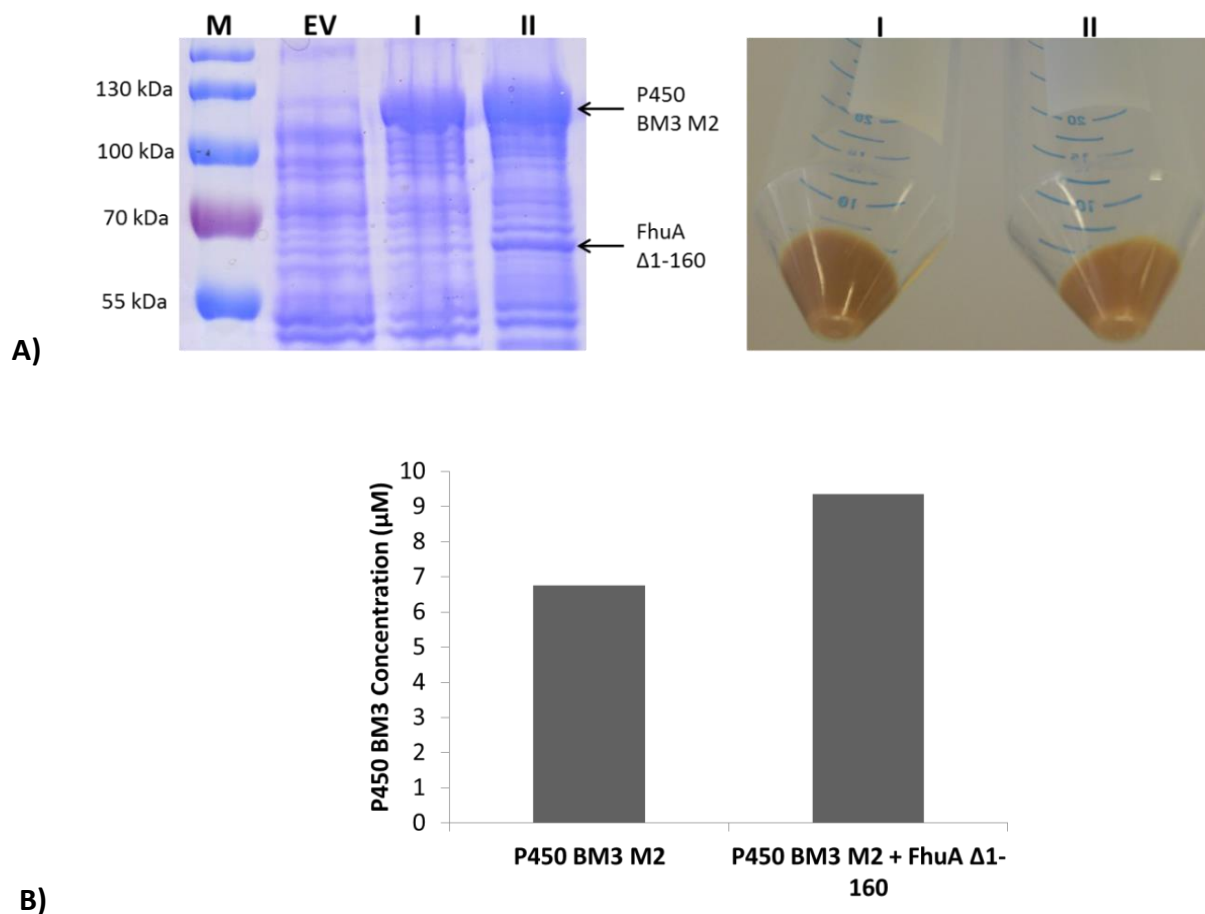
### 8.2.1. Generation of genetic constructs and pilot expressions

A simple synthetic operon was constructed for the simultaneous transcription of genes encoding P450 BM3 M2 and FhuA  $\Delta$ 1-160 whereas the gene that encodes FhuA  $\Delta$ 1-160 protein also contained an outer membrane targeting sequence. The genes were placed under the control of a T7 promoter so that a single polycistronic mRNA molecule containing two native *E. coli* ribosome binding sites in front of each gene was transcribed. This construct was assembled into an expression vector (pALXtreme-1a) using a phosphorothioate ligase independent cloning, PLICing, and introduced into *E. coli* BL21 Gold (DE3)  $\text{lacI}^{\text{q1}}$  cells in preparation for preliminary expression studies. The design of the genetic constructs is shown in Fig. 43. *E. coli* transformed with pALXtreme-1a harboring solely the P450 gene or no insert were available from a previous study (Dennig et al., 2012) and used as controls in the follow-up experiments.

Pilot expression studies were carried out using the optimized expression conditions for P450 BM3 resulting in similar wet cell weight at the end of fermentation ( $\sim 0.02 \text{ g mL}^{-1}$  growth after 18 h cultivation at 30 °C) and comparable expression level of P450 BM3. Images of the protein gel and cell pellets after expression are presented in Fig. 44 A. In the total protein extract of the co-expressing *E. coli* strain an additional band corresponding to the molecular weight of FhuA  $\Delta$ 1-160 could be observed (not present in empty vector control sample). The measurement by CO-binding assay of specific P450 concentration in normalized (to OD) cell suspensions of both recombinant strains confirmed that P450 could be expressed in soluble form. The concentration of catalyst in the co-expressing strain was not severely compromised due to the production of a second recombinant protein. Specific

concentrations of 6.75 and 9.35  $\mu\text{M}$  were measured in normalized cell suspension ( $\text{OD}_{600}$  of 40), for the strain co-expressing FhuA  $\Delta 1-160$  and the one producing solely P450, respectively (Fig. 44 B). The expression quality and quantity of both proteins were viewed as adequate and, therefore used throughout the rest of this study without further optimization of expression conditions.



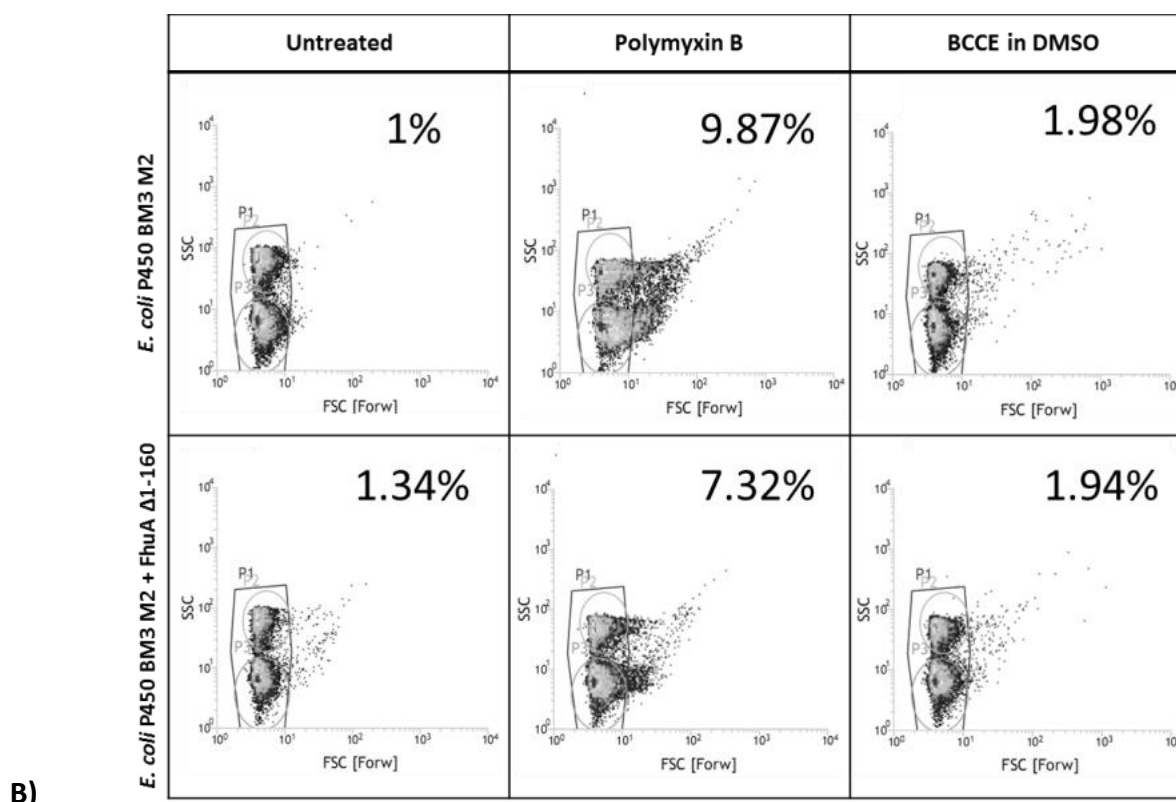
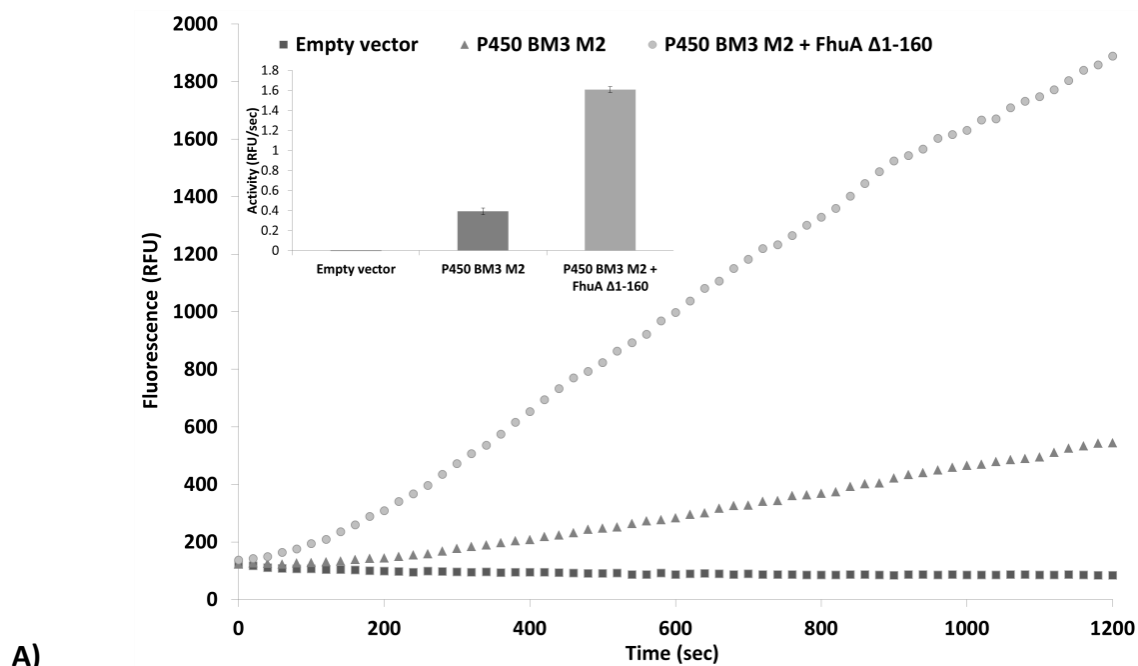


**Figure 44.** Verification of protein expression. **A)** SDS-PAGE of whole cells with expressed P450 BM3 or co-expressed P450 BM3 and FhuA  $\Delta$ 1-160; recombinantly expressed proteins are marked with an arrow (lane EV – cells harboring pALXtreme-1a empty vector; lane I – cells harboring pALXtreme-1a P450 BM3 M2; lane II – cells harboring pALXtreme-1a P450 BM3 M2 + FhuA  $\Delta$ 1-160; expected protein sizes –  $\sim$ 120 kDa for P450 BM3 M2 and  $\sim$ 55 kDa for FhuA  $\Delta$ 1-160). The picture on the right shows the wet cell pellets from 50 mL culture obtained after harvesting of *E. coli* transformed with pALXtreme-1a P450 BM3 M2 (I) and pALXtreme-1a P450 BM3 M2 + FhuA  $\Delta$ 1-160 (II); **B)** Graphical representation of the specific P450 concentration of normalized cell suspensions ( $\text{OD}_{600}$  of 40) of *E. coli* transformed with pALXtreme-1a P450 BM3 M2 and pALXtreme-1a P450 BM3 M2 + FhuA  $\Delta$ 1-160 measured by CO - binding assay (Omura and Sato, 1964).

### 8.2.2. Whole cell biotransformation of 7-benzyoxy-3-carboxy-coumarin ethyl ester (BCCE) – a model fluorogenic substrate for P450 BM3

In a proof-of-principle experiment, the fluorogenic 7-benzyoxy-3-carboxy-coumarin ethyl ester (BCCE, Ex. 400 Em. 440), was used as substrate with whole cell suspensions of 3 *E. coli* strains – one harboring an empty vector (negative control), another expressing P450 BM3 and a third co-expressing P450 and FhuA  $\Delta$ 1-160. The initial rate of BCCE conversion with cell suspensions normalized to small differences in the optical density as well as specific P450 BM3 concentration was 4.1-fold higher for the strain expressing the channel protein compared to the *E. coli* strain containing only the P450 catalyst (Fig. 44A).

In order to evaluate the effect of BCCE and the co-solvent used (DMSO) in this assay on cells' membrane integrity, a flow-cytometry analysis of 10 000 events from the two cell types was performed (Fig. 45B). The light scattered upon passage of individual cells through the laser beam of a flow cytometer gives information about cell's dimensions (forward scatter, FSC) and complexity/ smoothness of the cell's membrane (side scatter, SSC) and the plots of cells distributions across FSC vs SSC plot can be used to monitor changes in physical appearance of the cell population. Cells incubated with permeabilizer (Polymixin B) and untreated cells served as positive and negative control for the determination of the region (gate) in the FSC vs SSC histogram corresponding to intact cells and cells with compromised integrity. This analysis showed that under the reaction conditions employed in the BCCE assay (1  $\mu$ M substrate, 2% DMSO, 20 min incubation), the cell populations remained largely unaffected; a population shift less than 1% outside the gated region was observed for either of the tested cell types (Fig. 45 B). Taken together, these findings suggested that the increased rate of BCCE hydroxylation is due to an increased mass transfer of substrate across the cell wall due to the presence of the large passive diffusion channel in the outer membrane of *E. coli*.

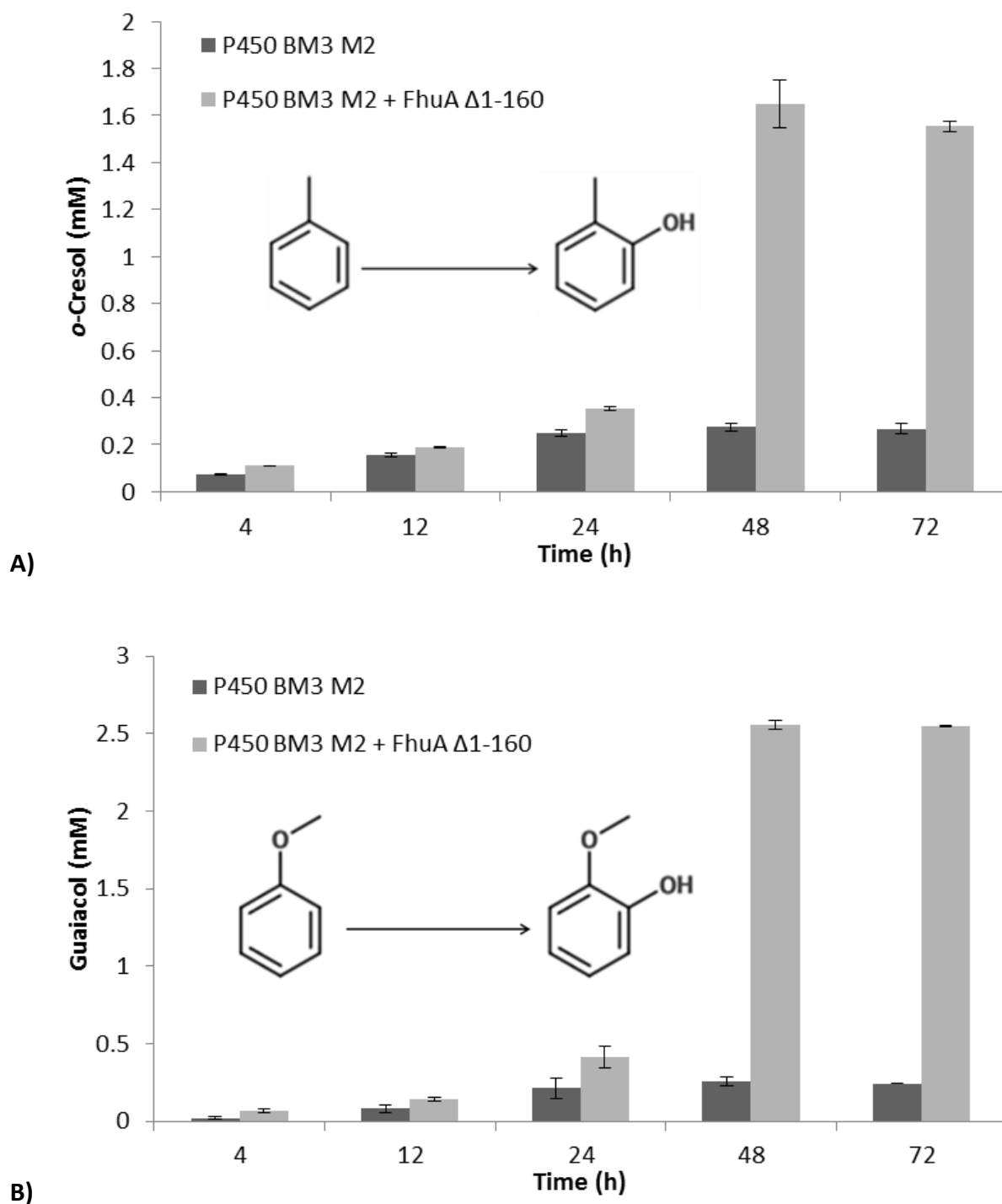


**Figure 45. A)** Initial conversion rates of 7-benzyoxy-3-carboxy-coumarin ethyl ester (BCCE) with whole cells with empty vector and recombinantly expressed P450 BM3 M2 or co-expressed P450 BM3 M2 and FhuA Δ1-160; **B)** Analysis of populations of *E. coli* cells (10 000 events) harboring P450 BM3 and co-expressed P450 BM3 M2 and FhuA Δ1-160. Before analysis OD-normalized cell suspensions were either left untreated or were incubated with permeabilizer (60 μM Polymyxin B) or BCCE substrate in DMSO (1 μM BCCE, 2% DMSO) for 20 min prior to population analysis. Gate P1 was selected according to the appearance the population of untreated *E. coli* cells harboring solely P450. The fraction (in %) of the outlying population is shown for each strain type and treatment method.

### 8.2.3. Long term biotransformation of toluene and cresol with whole cells expressing P450 BM3 M2 or co-expressing P450 BM3 M2 and FhuA $\Delta$ 1-160

Whole cell biocatalysis of two monosubstituted benzenes (toluene and anisole) was attempted as a second validation step of the proposed concept. The reactions were performed with resting cells over a 72-hour period. Glucose was added to the reaction mixes to ensure co-factor regeneration (NAD(P)H), necessary for P450-mediated phenolic hydroxylation of the selected substrates. Substrate concentration of  $\sim 1 \text{ g L}^{-1}$  (0,1% (v/v)) and 2% DMSO were used so that no major disturbances in the membrane integrity are caused due to the presence of organic compounds. When toluene was employed as substrate, a total of 1.6 mM of *o*-cresol (2-methyl phenol) was formed by the *E. coli* strain co-expressing FhuA  $\Delta$ 1-160 in 48 hours (Fig. 46 A). Interesting, most of the product was formed in the time interval between 24 and 48 hours. Cell expressing solely the P450 catalyst produced 6 times less *o*-cresol, most of it formed in the first 24 hour of biotransformation. A similar trend was observed in case of whole cell biotransformation of the more polar monosubstituted benzene, anisole. *E. coli* co-expressing FhuA  $\Delta$ 1-160 produced 2.6 mM guaiacol (2-methoxy phenol) over 48 hours with  $\sim 75\%$  of the product being generated in the period between 24 and 48 hours (Fig 46 B). This accounted for 10 times higher product formed compared to the reaction carried out with *E. coli* without the outer membrane protein. Side product formation (*p*-hydroxylated products) with either substrate accounted for less than 5% of the total products formed (not shown) and did not affect the reported trend. Production of hydroquinones was surprisingly not detected in these experiments although this could be assumed from conversion with purified enzyme. This is possibly attributable to the analytical method and thermal program settings used in the study (GC vs HPLC in the study described in section 3.1). Nevertheless, monitoring of only first product formed in the catalytic reaction still provides strong indications about the influence of the channel protein on the whole cell bioconversion of benzenes.





**Figure 46.** Formation of **A)** *o*-cresol from toluene and **B)** guaiacol from anisole in whole cell catalysis with *E. coli* expressing P450 BM3 and co-expressing P450 BM3 and FhuA  $\Delta$ 1-160. Samples were collected over 72 hour period and specific product formation was quantified by gas chromatography. The sharp increase in product concentrations between 24 and 48 h was a reproducible result; however, the reasons for this were not investigated in detail and remain unclear.

#### **8.2.4. Summary and conclusions: Improving the rate of P450-mediated whole cell biotransformations by co-expressing a passive diffusion channel in the outer membrane of *E. coli***

The barrier formed by the outer membrane of gram-negative bacteria is a primary reason for the suboptimal productivity of whole-cell processes. In this study, an improvement of the mass transfer across the outer membrane of *E. coli* was realized by recombinant co-expression of an outer membrane mutant protein, FhuA  $\Delta$ 1-160, which forms a passive diffusion channel in the outer membrane of *E. coli*. This approach was applied to P450-catalyzed conversion a fluorogenic coumarin substrate (BCCE) and 2 monosubstituted benzenes (toluene and anisole). Our data indicated that the heterologous co-expression of FhuA  $\Delta$ 1-160 has a beneficial effect on whole cell conversion of the coumarin substrate BCCE and the monosubstituted benzenes - anisole and toluene. The whole cell conversion of BCCE was improved 4.1-fold for *E. coli* co-expressing FhuA  $\Delta$ 1-160 and P450 BM3 in comparison to cells containing solely the P450 catalyst. In addition, ~10-fold relative increase of phenolic product formation was observed in a 48-hour long biotransformation with cells co-expressing FhuA  $\Delta$ 1-160 and P450 BM3 M2 (R47S/Y51W/I401M), a variant engineered for regiospecific, aromatic hydroxylation (Dennig et al., 2012, 2013).

The membrane modification was achieved via a genetic construct carried on a plasmid vector. The plasmid-borne nature of the artificial operon allows for easy genetic manipulation (e.g., gene exchange) and straightforward transfer to any *E. coli* strain of interest by a simple transformation. This is important as the presented approach might be relevant to whole cell biocatalytic processes in general as the permeability change effected by the presence of a big diffusion channel is possibly of global nature.

The observed improvements were not at the expense of severely compromising the growth and expression level of the strain co-expressing the channel protein. Nevertheless, the marked relative increase in product formation with the engineered whole-cell catalyst suggests that the outer membrane is likely significantly altered by the presence of the protein forming a diffusion channel. Yet, cells co-expressing FhuA  $\Delta$ 1-160 yielded nearly as much wet cell weight and as high P450 concentration as the strain without membrane modifications, despite that these cells harbored a genetic construct of larger size from which

two individual genes were produced. Still, the cells co-expressing a membrane component were able to nearly completely compensate for the membrane defects that can potentially interfere with the proton motive force across the semi-permeable outer membrane and compromise the cell's integrity. In addition, the small difference in growth and yield of catalyst caused by the expression of FhuA  $\Delta$ 1-160 seemed to pose little obstacle in the biocatalytic process, especially taking into account that the gain in product amounts in 48 h was generally substantial for the investigated substrates – toluene, anisole and BCCE. The means by which cells manage to cope with this extra burden remain unclear. The elucidation of this rescuing mechanism can help to better explain the different performance of the strains in the biotransformation process as well.

The observed relative increase in productivity with whole-cells suggests that the modification of cellular membrane composition and mass transfer barrier properties through co-expression of a channel protein could be after further optimization a general approach for whole cell biocatalysis. The relative gain in performance of the whole cell catalysis due to the presence of an outer membrane channel appears to be a general trend, however we cannot rule out the dependency of mass transfer on the molecular features of the substrate molecules. To investigate further the relation between permeability and substrate structure, the presented approach should be tested with substrates of systematically varied size, hydrophobicity, and ionic charges. Finding a correlation between mass transfer across the modified outer membrane and the molecular features of substrates will provide further information for rationalizing the experimental data. It can be expected that the properties of the employed substrate as well as substrate concentrations to certainly cause a difference in the mass transport rates across the modified outer membrane of *E. coli*. As a final remark, the relative gains in performance of the whole cell catalysis due to the presence of an outer membrane channel are possibly of global nature, and may provide a general platform for alleviating the issue of limited mass transfer in whole cell biocatalysis.

## 9. References

- D' Abbadie, M., Hofreiter, M., Vaisman, A., Loakes, D., Gasparutto, D., Cadet, J., Woodgate, R., Pääbo, S., and Holliger, P. (2007a). Molecular breeding of polymerases for amplification of ancient DNA. *Nat. Biotechnol.* *25*, 939–943.
- D' Abbadie, M., Hofreiter, M., Vaisman, A., Loakes, D., Gasparutto, D., Cadet, J., Woodgate, R., Paabo, S., and Holliger, P. (2007b). Molecular breeding of polymerases for amplification of ancient DNA. *Nat Biotech* *25*, 939–943.
- Agresti, J.J., Antipov, E., Abate, A.R., Ahn, K., Rowat, A.C., Baret, J.-C., Marquez, M., Klibanov, A.M., Griffiths, A.D., and Weitz, D.A. (2010). Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc Natl Acad Sci USA* *107*, 4004–4009.
- Arango Gutierrez, E., Mundhada, H., Meier, T., Duefel, H., Bocola, M., and Schwaneberg, U. (2013). Reengineered glucose oxidase for amperometric glucose determination in diabetes analytics. *Biosens. Bioelectron.* *50*, 84–90.
- Avkin, S., Goldsmith, M., Velasco-Miguel, S., Geacintov, N., Friedberg, E.C., and Livneh, Z. (2004). Quantitative Analysis of Translesion DNA Synthesis across a Benzo[a]pyrene-Guanine Adduct in Mammalian Cells. *J. Biol. Chem.* *279*, 53298–53305.
- Battistel, E., Tassinari, R., Fornaroli, M., and Bonoldi, L. (2003). Oxidation of benzene by molecular oxygen catalysed by vanadium. *J. Mol. Catal. Chem.* *202*, 107–115.
- Bernhardt, R. (2006). Cytochromes P450 as versatile biocatalysts. *J. Biotechnol.* *124*, 128–145.
- Bianchi, D., Bortolo, R., Tassinari, R., Ricci, M., and Vignola, R. (2000). A Novel Iron-Based Catalyst for the Biphasic Oxidation of Benzene to Phenol with Hydrogen Peroxide. *Angew. Chem.* *112*, 4491–4493.
- Blanusa, M., Schenk, A., Sadeghi, H., Marienhagen, J., and Schwaneberg, U. (2010a). Phosphorothioate-based ligase-independent gene cloning (PLICing): An enzyme-free and sequence-independent cloning method. *Anal. Biochem.* *406*, 141–146.
- Blanusa, M., Schenk, A., Sadeghi, H., Marienhagen, J., and Schwaneberg, U. (2010b). Phosphorothioate-based ligase-independent gene cloning (PLICing): An enzyme-free and sequence-independent cloning method. *Anal. Biochem.* *406*, 141–146.
- Bommarius, A.S., Blum, J.K., and Abrahamson, M.J. (2011). Status of protein engineering for biocatalysts: how to design an industrially useful biocatalyst. *Curr. Opin. Chem. Biol.* *15*, 194–200.
- Bornscheuer, U.T., and Pohl, M. (2001). Improved biocatalysts by directed evolution and rational protein design. *Curr. Opin. Chem. Biol.* *5*, 137–143.
- Bornscheuer, U.T., Huisman, G.W., Kazlauskas, R.J., Lutz, S., Moore, J.C., and Robins, K. (2012). Engineering the third wave of biocatalysis. *Nature* *485*, 185–194.
- Boudsocq, F., Iwai, S., Hanaoka, F., and Woodgate, R. (2001). *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic pol{eta}. *Nucleic Acids Res.* *29*, 4607–4616.

- Boudsocq, F., Kokoska, R.J., Plosky, B.S., Vaisman, A., Ling, H., Kunkel, T.A., Yang, W., and Woodgate, R. (2004). Investigating the role of the little finger domain of Y-family DNA polymerases in low fidelity synthesis and translesion replication. *J. Biol. Chem.* *279*, 32932–32940.
- Bowler, P.J. (1996). *Charles Darwin: The Man and His Influence* (Cambridge University Press).
- Brakmann, S., and Grzeszik, S. (2001). An Error-Prone T7 RNA Polymerase Mutant Generated by Directed Evolution. *ChemBioChem* *2*, 212–219.
- Bratkovič, T. (2010). Progress in phage display: evolution of the technique and its applications. *Cell. Mol. Life Sci.* *67*, 749–767.
- Braun, M., Killmann, H., Maier, E., Benz, R., and Braun, V. (2002). Diffusion through channel derivatives of the Escherichia coli FhuA transport protein. *Eur. J. Biochem. FEBS* *269*, 4948–4959.
- Buchholz, K., and Collins, J. (2013). The roots—a short history of industrial microbiology and biotechnology. *Appl. Microbiol. Biotechnol.* *97*, 3747–3762.
- Buchholz, K., Kasche, V., and Bornscheuer, U.T. (2012). *Biocatalysts and Enzyme Technology* (John Wiley & Sons).
- Camps, M., and Loeb, L.A. Use of Pol I-Deficient E. coli for Functional Complementation of DNA Polymerase. In *Directed Enzyme Evolution*, (New Jersey: Humana Press), pp. 11–18.
- Carters, R., Ferguson, J., Gaut, R., Ravetto, P., Thelwell, N., and Whitcombe, D. (2008). Design and Use of Scorpions Fluorescent Signaling Molecules. In *Molecular Beacons: Signalling Nucleic Acid Probes, Methods, and Protocols*, A. Marx, and O. Seitz, eds. (Humana Press), pp. 99–115.
- De Carvalho, C.C.C.R. (2011). Enzymatic and whole cell catalysis: Finding new strategies for old processes. *Biotechnol. Adv.* *29*, 75–83.
- Chandani, S., Jacobs, C., and Loechler, E.L. (2010). Architecture of Y-Family DNA Polymerases Relevant to Translesion DNA Synthesis as Revealed in Structural and Molecular Modeling Studies. *J. Nucleic Acids* *2010*, 1–20.
- Chen, R. (2007). Permeability issues in whole-cell bioprocesses and cellular membrane engineering. *Appl. Microbiol. Biotechnol.* *74*, 730–738.
- Clarke, T.E., Tari, L.W., and Vogel, H.J. (2001). Structural Biology of Bacterial Iron Uptake Systems. *Curr. Top. Med. Chem.* *1*, 7–30.
- Cornelissen, C.N. (2003). Transferrin-iron uptake by gram-negative bacteria. *Front. Biosci.* *8*, d836.
- Cornelissen, S., Julsing, M.K., Volmer, J., Riechert, O., Schmid, A., and Bühler, B. (2013). Whole-cell-based CYP153A6-catalyzed (S)-limonene hydroxylation efficiency depends on host background and profits from monoterpene uptake via AlkL. *Biotechnol. Bioeng.* *110*, 1282–1292.
- Credo, G.M., Su, X., Wu, K., Elibol, O.H., Liu, D.J., Reddy, B., Tsai, T.-W., Dorvel, B.R., Daniels, J.S., Bashir, R., et al. (2012). Label-free electrical detection of pyrophosphate generated from DNA polymerase reactions on field-effect devices. *The Analyst* *137*, 1351.
- DeLucia, A.M., Grindley, N.D.F., and Joyce, C.M. (2003). An error-prone family Y DNA polymerase (DinB homolog from *Sulfolobus solfataricus*) uses a “steric gate” residue for discrimination against ribonucleotides. *Nucleic Acids Res.* *31*, 4129–4137.

- Dennig, A. (2014). Engineering of cytochrome P450 monooxygenases for application in phenol synthesis.
- Dennig, A., Shivange, A.V., Marienhagen, J., and Schwaneberg, U. (2011). OmniChange: The Sequence Independent Method for Simultaneous Site-Saturation of Five Codons. *PLoS ONE* 6, e26222.
- Dennig, A., Marienhagen, J., Ruff, A.J., Guddat, L., and Schwaneberg, U. (2012). Directed Evolution of P 450 BM 3 into a p-Xylene Hydroxylase. *ChemCatChem* 4, 771–773.
- Dennig, A., Lülldorf, N., Liu, H., and Schwaneberg, U. (2013). Regioselective o-Hydroxylation of Monosubstituted Benzenes by P450 BM3. *Angew. Chem. Int. Ed.* 52, 8459–8462.
- Deya, P.M., Dopico, M., Jeronimo Morey, A.G.R., and Saa, J.M. (1987). On the regioselectivity of the fremy's salt oxidation of phenols. *Tetrahedron* 43, 3523–3532.
- Doublé, S., Sawaya, M.R., and Ellenberger, T. (1999). An open and closed case for all polymerases. *Structure* 7, R31–R35.
- Dragan, A.I., Casas-Finet, J.R., Bishop, E.S., Strouse, R.J., Schenerman, M.A., and Geddes, C.D. (2010). Characterization of PicoGreen Interaction with dsDNA and the Origin of Its Fluorescence Enhancement upon Binding. *Biophys. J.* 99, 3010–3019.
- Drauz, K. (2012). *Enzyme Catalysis in Organic Synthesis: A Comprehensive Handbook* (John Wiley & Sons).
- Ebner, J.R., Felthouse, T.R., and Fentress, D.C. (1999). Catalytic hydroxylation of an aromatic compound with nitrous oxide in the presence of a zeolite catalyst containing a group 8 metal; reduced carbon monoxide production during regeneration.
- Eoff, R.L., Angel, K.C., Egli, M., and Guengerich, F.P. (2007). Molecular basis of selectivity of nucleoside triphosphate incorporation opposite O6-benzylguanine by *Sulfolobus solfataricus* DNA polymerase Dpo4: steady-state and pre-steady-state kinetics and x-ray crystallography of correct and incorrect pairing. *J. Biol. Chem.* 282, 13573–13584.
- Eoff, R.L., Stafford, J.B., Szekely, J., Rizzo, C.J., Egli, M., Guengerich, F.P., and Marnett, L.J. (2009). Structural and functional analysis of *Sulfolobus solfataricus* Y-family DNA polymerase Dpo4-catalyzed bypass of the malondialdehyde-deoxyguanosine adduct. *Biochemistry (Mosc.)* 48, 7079–7088.
- Erickson, B., Nelson, and Winters, P. (2012). Perspective on opportunities in industrial biotechnology in renewable chemicals. *Biotechnol. J.* 7, 176–185.
- Erlich, H.A. (2013). Development and Evolution of PCR. *Genet. Eng. Biotechnol. News* 33, 32–33, 45.
- Fabbrizzi, L., Marcotte, N., Stomeo, F., and Taglietti, A. (2002). Pyrophosphate Detection in Water by Fluorescence Competition Assays: Inducing Selectivity through the Choice of the Indicator. *Angew. Chem. Int. Ed.* 41, 3811–3814.
- Ferguson, A.D., Ködding, J., Walker, G., Bös, C., Coulton, J.W., Diederichs, K., Braun, V., and Welte, W. (2001). Active Transport of an Antibiotic Rifamycin Derivative by the Outer-Membrane Protein FhuA. *Structure* 9, 707–716.
- Filée, J., Forterre, P., Sen-Lin, T., and Laurent, J. (2002). Evolution of DNA Polymerase Families: Evidences for Multiple Gene Exchange Between Cellular and Viral Proteins. *J. Mol. Evol.* 54, 763–773.

- Friedberg, E.C. (2003). DNA damage and repair. *Nature* 421, 436–440.
- Friedberg, E.C., Walker, G.C., and Siede, W. (1995). DNA repair and mutagenesis (ASM Press).
- Fujimori, O., and Nakamura, M. (1985). Protein A gold-silver staining method for light microscopic immunohistochemistry. *Arch. Histol. Jpn. Nihon Soshikigaku Kiroku* 48, 449–452.
- Ghadessy, F.J., and Holliger, P. (2007). Compartmentalized Self-Replication: a novel method for the directed evolution of polymerases and other enzymes. In *Protein Engineering Protocols*, pp. 237–248.
- Ghadessy, F.J., Ong, J.L., and Holliger, P. (2001). Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci. U. S. A.* 98, 4552–4557.
- Glieder, A., Farinas, E.T., and Arnold, F.H. (2002). Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase. *Nat. Biotechnol.* 20, 1135–1139.
- Gomes, S., Leonor, I.B., Mano, J.F., Reis, R.L., and Kaplan, D.L. (2012). Natural and genetically engineered proteins for tissue engineering. *Prog. Polym. Sci.* 37, 1–17.
- Goodman, M.F. (2002). Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu. Rev. Biochem.* 71, 17–50.
- Govindaraj, S., and Poulos, T.L. (1995). Role of the Linker Region Connecting the Reductase and Heme Domains in Cytochrome P450BM-3. *Biochemistry (Mosc.)* 34, 11221–11226.
- Govindaraj, S., and Poulos, T.L. (1996). Probing the structure of the linker connecting the reductase and heme domains of cytochrome P450BM-3 using site-directed mutagenesis. *Protein Sci. Publ. Protein Soc.* 5, 1389–1393.
- Grant, C., Deszcz, D., Wei, Y.-C., Martínez-Torres, R.J., Morris, P., Folliard, T., Sreenivasan, R., Ward, J., Dalby, P., Woodley, J.M., et al. (2014). Identification and use of an alkane transporter plug-in for applications in biocatalysis and whole-cell biosensing of alkanes. *Sci. Rep.* 4.
- Grinkova, Y.V., Denisov, I.G., McLean, M.A., and Sligar, S.G. (2013). Oxidase uncoupling in heme monooxygenases: human cytochrome P450 CYP3A4 in Nanodiscs. *Biochem. Biophys. Res. Commun.* 430, 1223–1227.
- Guengerich, F.P. (2001). Common and Uncommon Cytochrome P450 Reactions Related to Metabolism and Chemical Toxicity. *Chem. Res. Toxicol.* 14, 611–650.
- Guzman, L.M., Belin, D., Carson, M.J., and Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* 177, 4121–4130.
- Hannemann, F., Bichet, A., Ewen, K.M., and Bernhardt, R. (2007). Cytochrome P450 systems--biological variations of electron transport chains. *Biochim. Biophys. Acta* 1770, 330–344.
- Heid, C.A., Stevens, J., and Livak, K.J. (1996). Real time quantitative PCR. *Genome Res.* 6, 986–994.
- Hilvert, D. (2013). Design of Protein Catalysts. *Annu. Rev. Biochem.* 82, 447–470.
- Hofzumahaus, S., and Schallmey, A. (2013). Escherichia coli-based expression system for the heterologous expression and purification of the elicitor  $\beta$ -cinnamomin from *Phytophthora cinnamomi*. *Protein Expr. Purif.* 90, 117–123.

- Holden, M.J., Haynes, R.J., Rabb, S.A., Satija, N., Yang, K., and Blasic, J.R. (2009). Factors Affecting Quantification of Total DNA by UV Spectroscopy and PicoGreen Fluorescence. *J. Agric. Food Chem.* *57*, 7221–7226.
- Hollmann, F., Arends, I.W.C.E., Buehler, K., Schallmeyer, A., and Bühler, B. (2011). Enzyme-mediated oxidations for the chemist. *Green Chem.* *13*, 226–265.
- Hua, D.H., Tamura, M., Huang, X., Stephany, H.A., Helfrich, B.A., Perchellet, E.M., Sperflage, B.J., Perchellet, J.-P., Jiang, S., Kyle, D.E., et al. (2002). Syntheses and bioactivities of substituted 9,10-dihydro-9,10-[1,2]benzoanthracene-1,4,5,8-tetrones. Unusual reactivities with amines. *J. Org. Chem.* *67*, 2907–2912.
- Huang, M.M., Arnheim, N., and Goodman, M.F. (1992). Extension of base mispairs by Taq DNA polymerase: implications for single nucleotide discrimination in PCR. *Nucleic Acids Res.* *20*, 4567–4573.
- Hudnall, P.M. (2000). Hydroquinone. In *Ullmann's Encyclopedia of Industrial Chemistry*, (Wiley-VCH Verlag GmbH & Co. KGaA),.
- Huggett, J., Dheda, K., Bustin, S., and Zumla, A. (2005). Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun.* *6*, 279–284.
- Hull, D.L. (1973). *Darwin and his critics: the reception of Darwin's theory of evolution by the scientific community* (Harvard University Press).
- Ivanov, D.P., Sobolev, V.I., Pirutko, L.V., and Panov, G.I. (2002). New Way of Hydroquinone and Catechol Synthesis using Nitrous Oxide as Oxidant. *Adv. Synth. Catal.* *344*, 986–995.
- Jaeger, K.E., Eggert, T., Eipper, A., and Reetz, M.T. (2001). Directed evolution and the creation of enantioselective biocatalysts. *Appl. Microbiol. Biotechnol.* *55*, 519–530.
- Jiang, T., Wang, W., and Han, B. (2013). Catalytic hydroxylation of benzene to phenol with hydrogen peroxide using catalysts based on molecular sieves. *New J. Chem.* *37*, 1654–1664.
- Johnson, K.A. (2010). The kinetic and chemical mechanism of high-fidelity DNA polymerases. *Biochim. Biophys. Acta BBA - Proteins Proteomics* *1804*, 1041–1048.
- Julsing, M.K., Schrewe, M., Cornelissen, S., Hermann, I., Schmid, A., and Bühler, B. (2012). Outer Membrane Protein AlkL Boosts Biocatalytic Oxyfunctionalization of Hydrophobic Substrates in *Escherichia coli*. *Appl. Environ. Microbiol.* *78*, 5724–5733.
- Karich, A., Kluge, M., Ullrich, R., and Hofrichter, M. (2013). Benzene oxygenation and oxidation by the peroxygenase of *Agrocybe aegerita*. *AMB Express* *3*, 5.
- Kelman, Z., and White, M.F. (2005). Archaeal DNA replication and repair. *Curr. Opin. Microbiol.* *8*, 669–676.
- Krumenacker, L., Costantini, M., Pontal, P., and Sentenac, J. (2000). Hydroquinone, Resorcinol, and Catechol. In *Kirk-Othmer Encyclopedia of Chemical Technology*, (John Wiley & Sons, Inc.),.
- Laha, S., and Luthy, R.G. (1990). Oxidation of aniline and other primary aromatic amines by manganese dioxide. *Environ. Sci. Technol.* *24*, 363–373.



- Laufer, W., and Hoelderich, W.F. (2002). New direct hydroxylation of benzene with oxygen in the presence of hydrogen over bifunctional ion-exchange resins. *Chem. Commun.* 1684–1685.
- Leemhuis, H., Kelly, R.M., and Dijkhuizen, L. (2009). Directed evolution of enzymes: Library screening strategies. *IUBMB Life* 61, 222–228.
- Li, Z., Roccatano, D., Lorenz, M., and Schwaneberg, U. (2012). Directed evolution of subtilisin E into a highly active and guanidinium chloride- and sodium dodecylsulfate-tolerant protease. *Chembiochem Eur. J. Chem. Biol.* 13, 691–699.
- Ling, H., Boudsocq, F., Woodgate, R., and Yang, W. (2001). Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell* 107, 91–102.
- Liu, W., and Wang, P. (2007). Cofactor regeneration for sustainable enzymatic biosynthesis. *Biotechnol. Adv.* 25, 369–384.
- Lutz, S. (2010). Beyond directed evolution—semi-rational protein engineering and design. *Curr. Opin. Biotechnol.* 21, 734–743.
- Martinez, R., Jakob, F., Tu, R., Siegert, P., Maurer, K.-H., and Schwaneberg, U. (2012). Increasing activity and thermal resistance of *Bacillus gibsonii* alkaline protease (BgAP) by directed evolution. *Biotechnol. Bioeng.*
- McCulloch, S.D., and Kunkel, T.A. (2008). The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res* 18, 148–161.
- McDonald, J.P., Hall, A., Gasparutto, D., Cadet, J., Ballantyne, J., and Woodgate, R. (2006). Novel thermostable Y-family polymerases: applications for the PCR amplification of damaged or ancient DNAs. *Nucleic Acids Res.* 34, 1102–1111.
- McGarrity, J.T., and Armstrong, J.B. (1981). The effect of temperature and other growth conditions on the fatty acid composition of *Escherichia coli*. *Can. J. Microbiol.* 27, 835–840.
- Mills, D.R., Peterson, R.L., and Spiegelman, S. (1967). An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc. Natl. Acad. Sci. U. S. A.* 58, 217–224.
- Miyazaki, K. (2011). MEGAWHOP cloning: a method of creating random mutagenesis libraries via megaprimer PCR of whole plasmids. *Methods Enzymol.* 498, 399–406.
- Montellano, P.R.O. de (2005). *Cytochrome P450: Structure, Mechanism, and Biochemistry* (Springer Science & Business Media).
- Müller, C.A., Dennig, A., Welters, T., Winkler, T., Ruff, A.J., Hummel, W., Gröger, H., and Schwaneberg, U. Whole-cell double oxidation of n-heptane. *J. Biotechnol.*
- Mundhada, H., Marienhagen, J., Scacioc, A., Schenk, A., Roccatano, D., and Schwaneberg, U. (2011a). SeSaM-Tv-II generates a protein sequence space that is unobtainable by epPCR. *Chembiochem Eur. J. Chem. Biol.* 12, 1595–1601.
- Mundhada, H., Marienhagen, J., Scacioc, A., Schenk, A., Roccatano, D., and Schwaneberg, U. (2011b). SeSaM-Tv-II Generates a Protein Sequence Space that is Unobtainable by epPCR. *ChemBioChem* 12, 1595–1601.

- Munro, A.W., Leys, D.G., McLean, K.J., Marshall, K.R., Ost, T.W.B., Daff, S., Miles, C.S., Chapman, S.K., Lysek, D.A., Moser, C.C., et al. (2002). P450 BM3: the very model of a modern flavocytochrome. *Trends Biochem. Sci.* *27*, 250–257.
- Nair, D.T., Johnson, R.E., Prakash, S., Prakash, L., and Aggarwal, A.K. (2004). Replication by human DNA polymerase-[iota] occurs by Hoogsteen base-pairing. *Nature* *430*, 377–380.
- Nallani, M., Benito, S., Onaca, O., Graff, A., Lindemann, M., Winterhalter, M., Meier, W., and Schwaneberg, U. (2006). A nanocompartment system (Synthosome) designed for biotechnological applications. *J. Biotechnol.* *123*, 50–59.
- Narhi, L.O., and Fulco, A.J. (1987). Identification and characterization of two functional domains in cytochrome P-450BM-3, a catalytically self-sufficient monooxygenase induced by barbiturates in *Bacillus megaterium*. *J. Biol. Chem.* *262*, 6683–6690.
- Nazor, J., Dannenmann, S., Adjei, R.O., Fordjour, Y.B., Ghampson, I.T., Blanusa, M., Roccatano, D., and Schwaneberg, U. (2008). Laboratory evolution of P450 BM3 for mediated electron transfer yielding an activity-improved and reductase-independent variant. *Protein Eng. Des. Sel. PEDS* *21*, 29–35.
- Nelson, J.R., Lawrence, C.W., and Hinkle, D.C. (1996). Deoxycytidyl transferase activity of yeast REV1 protein. *Nature* *382*, 729–731.
- Netscher, T. (2007). Synthesis of Vitamin E. In *Vitamins & Hormones*, Gerald Litwack, ed. (Academic Press), pp. 155–202.
- Ni, Y., and Chen, R.R. (2004). Accelerating whole-cell biocatalysis by reducing outer membrane permeability barrier. *Biotechnol. Bioeng.* *87*, 804–811.
- Niwa, S., Eswaramoorthy, M., Nair, J., Raj, A., Itoh, N., Shoji, H., Namba, T., and Mizukami, F. (2002). A One-Step Conversion of Benzene to Phenol with a Palladium Membrane. *Science* *295*, 105–107.
- Noinaj, N., Guillier, M., Barnard, T.J., and Buchanan, S.K. (2010). TonB-dependent transporters: regulation, structure, and function. *Annu. Rev. Microbiol.* *64*, 43–60.
- Nyrén, P. (1987). Enzymatic method for continuous monitoring of DNA polymerase activity. *Anal. Biochem.* *167*, 235–238.
- Obeid, S., Schnur, A., Gloeckner, C., Blatter, N., Welte, W., Diederichs, K., and Marx, A. (2011). Learning from Directed Evolution: *Thermus aquaticus* DNA Polymerase Mutants with Translesion Synthesis Activity. *ChemBioChem* *12*, 1574–1580.
- Ohmori, H., Friedberg, E.C., Fuchs, R.P., Goodman, M.F., Hanaoka, F., Hinkle, D., Kunkel, T.A., Lawrence, C.W., Livneh, Z., Nohmi, T., et al. (2001). The Y-family of DNA polymerases. *Mol. Cell* *8*, 7–8.
- Omura, T., and Sato, R. (1964). The Carbon Monoxide-binding Pigment of Liver Microsomes. *J. Biol. Chem.* *239*, 2370–2378.
- Orlando, C., Pinzani, P., and Pazzagli, M. (1998). Developments in Quantitative PCR. *Clin. Chem. Lab. Med.* *36*, 255–269.
- Pata, J.D. (2010). Structural diversity of the Y-family DNA polymerases. *Biochim. Biophys. Acta BBA - Proteins Proteomics* *1804*, 1124–1135.

- Pavelka, A., Chovancova, E., and Damborsky, J. (2009). HotSpot Wizard: a web server for identification of hot spots in protein engineering. *Nucleic Acids Res.* *37*, W376–W383.
- Perego, C., Carati, A., Ingallina, P., Mantegazza, M.A., and Bellussi, G. (2001). Production of titanium containing molecular sieves and their application in catalysis. *Appl. Catal. Gen.* *221*, 63–72.
- Pérez, J.M., López-Alvarado, P., Avendaño, C., and Menéndez, J.C. (2000). Hetero Diels–Alder Reactions of 1-Acetylamino- and 1-Dimethylamino-1-azadienes with Benzoquinones. *Tetrahedron* *56*, 1561–1567.
- Philp, J.C., Ritchie, R.J., and Allan, J.E.M. (2013). Biobased chemicals: the convergence of green chemistry with industrial biotechnology. *Trends Biotechnol.* *31*, 219–222.
- Prakash, S., Johnson, R.E., and Prakash, L. (2005). Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. *Annu. Rev. Biochem.* *74*, 317–353.
- Pritchard, L., Corne, D., Kell, D., Rowland, J., and Winson, M. (2005). A general model of error-prone PCR. *J. Theor. Biol.* *234*, 497–509.
- Ran, N., Knop, D.R., Draths, K.M., and Frost, J.W. (2001). Benzene-free synthesis of hydroquinone. *J. Am. Chem. Soc.* *123*, 10927–10934.
- Rappoport, Z. (2004). *The Chemistry of Phenols*, 2 Volume Set (John Wiley & Sons).
- Rasila, T.S., Pajunen, M.I., and Savilahti, H. (2009). Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, *Escherichia coli* mutator strain, and hydroxylamine treatment. *Anal. Biochem.* *388*, 71–80.
- Rautenbach, D. (2007). *The electrochemical hydroxylation of aromatic substrates*. Thesis.
- Reddy, J.S., Sivasanker, S., and Ratnasamy, P. (1992). Hydroxylation of phenol over ts-2, a titanium silicate molecular sieve. *J. Mol. Catal.* *71*, 373–381.
- Reetz, M.T., and Carballeira, J.D. (2007). Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* *2*, 891–903.
- Remias, J.E., Pavlosky, T.A., and Sen, A. (2003). Catalytic hydroxylation of benzene and cyclohexane using in situ generated hydrogen peroxide: new mechanistic insights and comparison with hydrogen peroxide added directly. *J. Mol. Catal. Chem.* *203*, 179–192.
- Rendon, M., and Horwitz, S. (2012). Topical treatment of hyperpigmentation disorders. *Ann. Dermatol. Vénérologie* *139 Suppl 4*, S153–S158.
- Rogozin, I.B., and Pavlov, Y.I. (2003a). Theoretical analysis of mutation hotspots and their DNA sequence context specificity. *Mutat. Res.* *544*, 65–85.
- Rogozin, I.B., and Pavlov, Y.I. (2003b). Theoretical analysis of mutation hotspots and their DNA sequence context specificity. *Mutat. Res.* *544*, 65–85.
- Ruff, A.J., Dennig, A., Wirtz, G., Blanusa, M., and Schwaneberg, U. (2012a). Flow Cytometer-Based High-Throughput Screening System for Accelerated Directed Evolution of P450 Monooxygenases. *ACS Catal.* *2*, 2724–2728.

- Ruff, A.J., Marienhagen, J., Verma, R., Roccatano, D., Genieser, H.-G., Niemann, P., Shivange, A.V., and Schwaneberg, U. (2012b). dRTP and dPTP a complementary nucleotide couple for the Sequence Saturation Mutagenesis (SeSaM) method. *J. Mol. Catal. B Enzym.* *84*, 40–47.
- Ruff, A.J., Dennig, A., and Schwaneberg, U. (2013a). To Get What We Aim For: Progress in Diversity Generation Methods. *FEBS J.* n/a – n/a.
- Ruff, A.J., Dennig, A., and Schwaneberg, U. (2013b). To Get What We Aim For: Progress in Diversity Generation Methods. *FEBS J.* n/a – n/a.
- Ruff, A.J., Kardashliev, T., Dennig, A., and Schwaneberg, U. (2014). The Sequence Saturation Mutagenesis (SeSaM) Method. In *Directed Evolution Library Creation*, E.M.J. Gillam, J.N. Copp, and D. Ackerley, eds. (Springer New York), pp. 45–68.
- Sale, J.E., Lehmann, A.R., and Woodgate, R. (2012). Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Rev. Mol. Cell Biol.* *13*, 141–152.
- Schomburg, I., Chang, A., Placzek, S., Söhngen, C., Rother, M., Lang, M., Munaretto, C., Ulas, S., Stelzer, M., Grote, A., et al. (2012). BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res.* gks1049.
- Schrittwieser, J.H., Sattler, J., Resch, V., Mutti, F.G., and Kroutil, W. (2011). Recent biocatalytic oxidation-reduction cascades. *Curr. Opin. Chem. Biol.* *15*, 249–256.
- Schwaneberg, U., Sprauer, A., Schmidt-Danner, C., and Schmid, R.D. (1999). P450 monooxygenase in biotechnology. I. Single-step, large-scale purification method for cytochrome P450 BM-3 by anion-exchange chromatography. *J. Chromatogr. A* *848*, 149–159.
- Seo, K.Y., Yin, J., Donthamsetti, P., Chandani, S., Lee, C.H., and Loechler, E.L. (2009). Amino acid architecture that influences dNTP insertion efficiency in Y-family DNA polymerase V of *E. coli*. *J. Mol. Biol.* *392*, 270–282.
- Sevrioukova, I.F., Li, H., Zhang, H., Peterson, J.A., and Poulos, T.L. (1999). Structure of a cytochrome P450–redox partner electron-transfer complex. *Proc. Natl. Acad. Sci.* *96*, 1863–1868.
- Shendure, J., and Ji, H. (2008). Next-generation DNA sequencing. *Nat Biotech* *26*, 1135–1145.
- Shivange, A.V., Marienhagen, J., Mundhada, H., Schenk, A., and Schwaneberg, U. (2009). Advances in generating functional diversity for directed protein evolution. *Curr. Opin. Chem. Biol.* *13*, 19–25.
- Shivange, A.V., Serwe, A., Dennig, A., Roccatano, D., Haefner, S., and Schwaneberg, U. (2012). Directed evolution of a highly active *Yersinia mollaretii* phytase. *Appl. Microbiol. Biotechnol.* *95*, 405–418.
- Silvian, L.F., Toth, E.A., Pham, P., Goodman, M.F., and Ellenberger, T. (2001). Crystal structure of a DinB family error-prone DNA polymerase from *Sulfolobus solfataricus*. *Nat Struct Mol Biol* *8*, 984–989.
- Singer, V.L., Jones, L.J., Yue, S.T., and Haugland, R.P. (1997). Characterization of PicoGreen Reagent and Development of a Fluorescence-Based Solution Assay for Double-Stranded DNA Quantitation. *Anal. Biochem.* *249*, 228–238.

- Soetaert, W., and Vandamme, E.J. (2010). *Industrial Biotechnology: Sustainable Growth and Economic Success* (John Wiley & Sons).
- Sokolovská, I., Rozenberg, R., Riez, C., Rouxhet, P.G., Agathos, S.N., and Wattiau, P. (2003). Carbon source-induced modifications in the mycolic acid content and cell wall permeability of *Rhodococcus erythropolis* E1. *Appl. Environ. Microbiol.* *69*, 7019–7027.
- Solinas, A., Brown, L.J., McKeen, C., Mellor, J.M., Nicol, J., Thelwell, N., and Brown, T. (2001). Duplex Scorpion primers in SNP analysis and FRET applications. *Nucleic Acids Res.* *29*, E96.
- Song, S., Lee, H., Jin, Y., Ha, Y.M., Bae, S., Chung, H.Y., and Suh, H. (2007). Syntheses of hydroxy substituted 2-phenyl-naphthalenes as inhibitors of tyrosinase. *Bioorg. Med. Chem. Lett.* *17*, 461–464.
- Sroga, G.E., and Dordick, J.S. (2002). A strategy for in vivo screening of subtilisin E reaction specificity in *E. coli* periplasm. *Biotechnol. Bioeng.* *78*, 761–769.
- Stemmer, W.P. (1994). DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci. U. S. A.* *91*, 10747–10751.
- Studier, W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* *41*, 207–234.
- Stynen, B., Tournu, H., Tavernier, J., and Dijck, P.V. (2012). Diversity in Genetic In Vivo Methods for Protein-Protein Interaction Studies: from the Yeast Two-Hybrid System to the Mammalian Split-Luciferase System. *Microbiol. Mol. Biol. Rev.* *76*, 331–382.
- Swan, M.K., Johnson, R.E., Prakash, L., Prakash, S., and Aggarwal, A.K. (2009). Structure of the human Rev1-DNA-dNTP ternary complex. *J. Mol. Biol.* *390*, 699–709.
- Takata, M., Sasaki, M.S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi-Iwai, Y., Shinohara, A., and Takeda, S. (1998). Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *EMBO J* *17*, 5497–5508.
- Tee, K.L., and Wong, T.S. (2013). Polishing the craft of genetic diversity creation in directed evolution. *Biotechnol. Adv.* *31*, 1707–1721.
- Lan Tee, K., and Schwaneberg, U. (2007). Directed Evolution of Oxygenases: Screening Systems, Success Stories and Challenges. *Comb. Chem. High Throughput Screen.* *10*, 197–217.
- Tenne, S.-J., and Schwaneberg, U. (2012). First Insights on Organic Cosolvent Effects on FhuA Wildtype and FhuA  $\Delta$ 1-159. *Int. J. Mol. Sci.* *13*, 2459–2471.
- Thelwell, N., Millington, S., Solinas, A., Booth, J., and Brown, T. (2000). Mode of action and application of Scorpion primers to mutation detection. *Nucleic Acids Res.* *28*, 3752–3761.
- Tolia, N.H., and Joshua-Tor, L. (2006). Strategies for protein coexpression in *Escherichia coli*. *Nat. Methods* *3*, 55–64.
- Trincao, J., Johnson, R.E., Escalante, C.R., Prakash, S., Prakash, L., and Aggarwal, A.K. (2001). Structure of the Catalytic Core of *S. cerevisiae* DNA Polymerase [eta]: Implications for Translesion DNA Synthesis. *Mol. Cell* *8*, 417–426.

- Trincao, J., Johnson, R.E., Wolfle, W.T., Escalante, C.R., Prakash, S., Prakash, L., and Aggarwal, A.K. (2004). Dpo4 is hindered in extending a G.T mismatch by a reverse wobble. *Nat. Struct. Mol. Biol.* *11*, 457–462.
- Tubeleviciute, A., and Skirgaila, R. (2010). Compartmentalized self-replication (CSR) selection of *Thermococcus litoralis* Sh1B DNA polymerase for diminished uracil binding. *Protein Eng. Des. Sel.* *23*, 589–597.
- Uljon, S.N., Johnson, R.E., Edwards, T.A., Prakash, S., Prakash, L., and Aggarwal, A.K. (2004). Crystal Structure of the Catalytic Core of Human DNA Polymerase Kappa. *Structure* *12*, 1395–1404.
- Urban, A., Neukirchen, S., and Jaeger, K.-E. (1997). A rapid and efficient method for site-directed mutagenesis using one-step overlap extension PCR. *Nucleic Acids Res.* *25*, 2227–2228.
- Vaara, M. (1992). Agents that increase the permeability of the outer membrane. *Microbiol. Rev.* *56*, 395–411.
- Verma, R., Schwaneberg, U., and Roccatano, D. (2012). MAP2.03D: A Sequence/Structure Based Server for Protein Engineering. *ACS Synth. Biol.* *1*, 139–150.
- De Visser, S.P., and Shaik, S. (2003). A proton-shuttle mechanism mediated by the porphyrin in benzene hydroxylation by cytochrome p450 enzymes. *J. Am. Chem. Soc.* *125*, 7413–7424.
- Wang, W., and Malcolm, B.A. (1999). Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange site-directed mutagenesis. *BioTechniques* *26*, 680–682.
- Wang, T.-W., Zhu, H., Ma, X.-Y., Zhang, T., Ma, Y.-S., and Wei, D.-Z. (2006a). Mutant library construction in directed molecular evolution. *Mol. Biotechnol.* *34*, 55–68.
- Wang, Y., Prosen, D.E., Mei, L., Sullivan, J.C., Finney, M., and Horn, P.B.V. (2004). A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. *Nucleic Acids Res.* *32*, 1197–1207.
- Wang, Y., Arora, K., and Schlick, T. (2006b). Subtle but variable conformational rearrangements in the replication cycle of *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4) may accommodate lesion bypass. *Protein Sci. Publ. Protein Soc.* *15*, 135–151.
- Wenda, S., Illner, S., Mell, A., and Kragl, U. (2011). Industrial biotechnology—the future of green chemistry? *Green Chem.* *13*, 3007.
- Whitcombe, D., Theaker, J., Guy, S.P., Brown, T., and Little, S. (1999). Detection of PCR products using self-probing amplicons and fluorescence. *Nat. Biotechnol.* *17*, 804–807.
- Whitehouse, C.J.C., Rees, N.H., Bell, S.G., and Wong, L.-L. (2011). Dearomatisation of o-xylene by P450BM3 (CYP102A1). *Chem. Weinh. Bergstr. Ger.* *17*, 6862–6868.
- Whitehouse, C.J.C., Bell, S.G., and Wong, L.-L. (2012). P450(BM3) (CYP102A1): connecting the dots. *Chem. Soc. Rev.* *41*, 1218–1260.
- Wiener, M.C., and Horanyi, P.S. (2011). How hydrophobic molecules traverse the outer membranes of Gram-negative bacteria. *Proc. Natl. Acad. Sci.* *108*, 10929–10930.

- Wong, J.H., Fiala, K.A., Suo, Z., and Ling, H. (2008a). Snapshots of a Y-Family DNA Polymerase in Replication: Substrate-induced Conformational Transitions and Implications for Fidelity of Dpo4. *J. Mol. Biol.* *379*, 317–330.
- Wong, T.S., Tee, K.L., Hauer, B., and Schwaneberg, U. (2004). Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. *Nucleic Acids Res.* *32*, e26.
- Wong, T.S., Zhurina, D., and Schwaneberg, U. (2006a). The diversity challenge in directed protein evolution. *Comb. Chem. High Throughput Screen.* *9*, 271–288.
- Wong, T.S., Roccatano, D., Zacharias, M., and Schwaneberg, U. (2006b). A statistical analysis of random mutagenesis methods used for directed protein evolution. *J. Mol. Biol.* *355*, 858–871.
- Wong, T.S., Zhurina, D., and Schwaneberg, U. (2006c). The diversity challenge in directed protein evolution. *Comb. Chem. High Throughput Screen.* *9*, 271–288.
- Wong, T.S., Roccatano, D., Zacharias, M., and Schwaneberg, U. (2006d). A Statistical Analysis of Random Mutagenesis Methods Used for Directed Protein Evolution. *J. Mol. Biol.* *355*, 858–871.
- Wong, T.S., Roccatano, D., and Schwaneberg, U. (2007a). Challenges of the genetic code for exploring sequence space in directed protein evolution. *Biocatal. Biotransformation* *25*, 229–241.
- Wong, T.S., Roccatano, D., and Schwaneberg, U. (2007b). Are transversion mutations better? A Mutagenesis Assistant Program analysis on P450 BM-3 heme domain. *Biotechnol. J.* *2*, 133–142.
- Wong, T.S., Wong, T.S., Roccatano\*, D., and Schwaneberg, U. (2007c). Challenges of the genetic code for exploring sequence space in directed protein evolution. *Biocatal. Biotransformation* *25*, 229–241.
- Wong, T.S., Roccatano, D., and Schwaneberg, U. (2007d). Steering directed protein evolution: strategies to manage combinatorial complexity of mutant libraries. *Environ. Microbiol.* *9*, 2645–2659.
- Wong, T.S., Roccatano, D., Loakes, D., Tee, K.L., Schenk, A., Hauer, B., and Schwaneberg, U. (2008b). Transversion-enriched sequence saturation mutagenesis (SeSaM-Tv+): a random mutagenesis method with consecutive nucleotide exchanges that complements the bias of error-prone PCR. *Biotechnol. J.* *3*, 74–82.
- Woodley, J.M., Breuer, M., and Mink, D. (2013). A future perspective on the role of industrial biotechnology for chemicals production. *Chem. Eng. Res. Des.* *91*, 2029–2036.
- Wu, Y., Wilson, R.C., and Pata, J.D. (2011). The Y-Family DNA Polymerase Dpo4 Uses a Template Slippage Mechanism To Create Single-Base Deletions. *J. Bacteriol.* *193*, 2630–2636.
- Yang, W. (2005). Portraits of a Y-family DNA polymerase. *FEBS Lett.* *579*, 868–872.
- Yang, W., and Woodgate, R. (2007). What a difference a decade makes: Insights into translesion DNA synthesis. *Proc. Natl. Acad. Sci.* *104*, 15591–15598.
- Yoshida, S., Yoshikawa, A., and Terao, I. (1990). Microbial production of hydroquinone. *J. Biotechnol.* *14*, 195–202.
- Zakoshansky, V.M. (2007). The cumene process for phenol-acetone production. *Pet. Chem.* *47*, 273–284.

- Zhang, H., Eoff, R.L., Kozekov, I.D., Rizzo, C.J., Egli, M., and Guengerich, F.P. (2009). Versatility of Y-family *Sulfolobus solfataricus* DNA polymerase Dpo4 in translesion synthesis past bulky N2-alkylguanine adducts. *J. Biol. Chem.* *284*, 3563–3576.
- Zhao, H., and van der Donk, W.A. (2003). Regeneration of cofactors for use in biocatalysis. *Curr. Opin. Biotechnol.* *14*, 583–589.
- Zhao, H., Giver, L., Shao, Z., Affholter, J.A., and Arnold, F.H. (1998). Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat Biotech* *16*, 258–261.
- Zhao, J., Kardashliev, T., Ruff, A.J., Bocola, M., and Schwaneberg, U. (2014). Lessons from diversity of directed evolution experiments by an analysis of 3000 mutations. *Biotechnol. Bioeng.* n/a – n/a.
- Zhou, B.-L., Pata, J.D., and Steitz, T.A. (2001a). Crystal Structure of a DinB Lesion Bypass DNA Polymerase Catalytic Fragment Reveals a Classic Polymerase Catalytic Domain. *Mol. Cell* *8*, 427–437.
- Zhou, B.-L., Pata, J.D., and Steitz, T.A. (2001b). Crystal structure of a DinB lesion bypass DNA polymerase catalytic fragment reveals a classic polymerase catalytic domain. *Mol. Cell* *8*, 427–437.
- Zhu, L., Verma, R., Roccatano, D., Ni, Y., Sun, Z.-H., and Schwaneberg, U. (2010). A potential antitumor drug (arginine deiminase) reengineered for efficient operation under physiological conditions. *Chembiochem Eur. J. Chem. Biol.* *11*, 2294–2301.



## FURTHER SCIENTIFIC CONTRIBUTIONS

### 1) “Lessons from diversity of directed evolution experiments by an analysis of 3,000 mutations”

*Jing Zhao, Tsvetan Kardashliev, Anna Joëlle Ruff, Marco Bocola and Ulrich Schwaneberg  
Accepted in Biotechnology and Bioengineering*

Diversity generation by random mutagenesis is often the first key step in directed evolution experiments and screening of 1,000–2,000 clones is in most directed evolution campaigns sufficient to identify improved variants. For experimentalists important questions such as how many positions are mutated in the targeted gene and what amino acid substitutions can be expected after screening of 1,000–2,000 clones are surprisingly not answered by a statistical analysis of mutant libraries. Therefore three random mutagenesis experiments (epPCR with a low- and a high-mutation frequency and a transversion-enriched sequence saturation mutagenesis method named SeSaM-Tv P/P) were performed on the lipase BSLA and in total 3,000 mutations were analyzed to determine the diversity in random mutagenesis libraries employed in directed evolution experiments. The active fraction of the population ranged from 15% (epPCR-high), to 52% (SeSaM-Tv P/P), and 55% (epPCR-low) which correlates well with the average number of amino acid substitutions per protein (4.1, 1.6 and 1.1). In the epPCR libraries transitions were the predominant mutations (>72%), and >82% of all mutations occurred at A- or T-nts. Consecutive nucleotide (nt) mutations were obtained only with a low fraction (2.8%) under highly error-prone conditions. SeSaM-Tv P/P was enriched in transversions (43%; >1.7-fold more than epPCR libraries), and consecutive nt mutations (30.5%; 11-fold more than epPCR-high). A high fraction of wild-type BSLA protein (33%) was found in the epPCR-low mutant library compared to 2% in epPCR-high and 13% in SeSaM-Tv P/P. An average of 1.8–1.9 amino acid substitutions per residue was obtained with epPCR-low and -high compared to 2.1 via SeSaM-Tv P/P. The chemical composition of the amino acid substitutions differed, however, significantly from the two epPCR methods to SeSaM-Tv P/P.

### 2) “A competitive flow cytometry screening system for arginine-metabolizing enzymes under physiological conditions with suitability in cancer treatment”

*Feng Cheng, Tsvetan Kardashliev, Christian Pitzler, Aamir Shehzad, Hongqi Lue, Jürgen Bernhagen, Leilei Zhu, Ulrich Schwaneberg  
Accepted in ACS Synthetic Biology*

A ligand-mediated *gfp*-expression system (LiMEx) was developed as a novel flow cytometry screening platform that relies on a competitive conversion/binding of arginine between arginine deiminase and ArgR. In contrast to standard product-driven detection systems, the competitive screening platform allows to evolve enzymes to efficiently operate at low substrate concentrations under physiological conditions. The principle of this competitive screening system was validated using the example of the anti-tumor therapeutic arginine deiminase.

## FINAL SUMMARY AND CONCLUSIVE REMARKS

This doctoral thesis addressed three fundamental questions from the related fields of protein engineering and biocatalysis: 1) advancing methods for diversity generation in directed protein evolution, 2) establishing biocatalytic processes for oxy-functionalization of generally unreactive aromatic C-H bonds and 3) postulating a strategy for improving the mass transfer in whole cell biocatalysis.

In CHAPTER I, the challenge of diversity generation in protein engineering has been tackled by addressing the remaining bottleneck of a highly advanced random mutagenesis method, the Sequence Saturation Mutagenesis (SeSaM) method. Improvement of SeSaM's mutational spectrum was accomplished by increasing the frequency of consecutive transversion nucleotide substitutions in SeSaM libraries. To achieve this, a DNA polymerase was engineered for more efficient transversion mismatch elongation; polymerases with such capabilities are essential in SeSaM step 3. For the purposes of the engineering campaign, a novel screening system for directed evolution of DNA polymerases employing a fluorescent Scorpion probe as a reporter was developed. The screening system was validated in a directed evolution experiment of a distributive polymerase from the Y-polymerase family (Dpo4 from *Sulfolobus solfataricus*) which was improved in elongation efficiency of consecutive mismatches. The engineering campaign yielded improved Dpo4 polymerase variants and identified previously unreported mutational hotspots that can modulate the translesion DNA synthesis capabilities of Dpo4, notably V32, R242, R247, R336. One of the identified mutants (Dpo4 R336N) was successfully benchmarked in a Sequence Saturation Mutagenesis experiment especially with regard to the desirable consecutive transversion mutations (2.5-fold increase in frequency relative to a reference library prepared with Dpo4 WT). The Scorpion probe screening system enabled to re-engineer polymerases with low processivity and fidelity, and no secondary activities (i.e. exonuclease activity or strand displacement activity) to match demands in diversity generation for directed protein evolution. In addition, the identified Dpo4 mutant with its current performance was already a significant advancement in subsequent transversion generation and especially for the SeSaM random mutagenesis method to complement the mutational spectrum of epPCR based methods.

In CHAPTER II of this thesis, P450 BM3 monooxygenase was studied in the context of regioselective biotransformations of benzenes with isolated enzyme (P450 Project I) and as a whole cell catalyst (P450 Project II). In P450 Project I, the one step synthesis of hydroquinones from benzene derivatives was performed with the monooxygenase P450 BM3 (CYP102A1) as a sole catalyst. The catalytic conversion of six aromatic hydrocarbons (toluene, anisole, chlorobenzene, bromobenzene, iodobenzene) to the corresponding *o*-phenols (*o*-cresol, guaiacol, 2-chlorophenol, 2-bromophenol, 2-iodophenol, respectively) by the P450 BM3 variant M2 (R47S/Y51W/I401M) was reported to proceed with high regioselectivity (*o*-hydroxylation  $\geq 95\%$ ). Furthermore, P450 BM3 M2 could catalyze the hydroxylation of mono-hydroxylated benzenes to di-hydroxylated products with high regioselectivity ( $>99\%$ ) and halogenated phenols were converted with high catalytic activities ( $28.7$  up to  $\sim 260 \text{ nmol}_{\text{product}} \text{ nmol}_{\text{p450}}^{-1} \text{ min}^{-1}$ ) and yields  $>90\%$ . Hydroxylation of non-halogenated phenols was performed with catalytic activity of up to  $9 \text{ nmol}_{\text{product}} \text{ nmol}_{\text{p450}}^{-1} \text{ min}^{-1}$  and yields  $\leq 50\%$ . One step double hydroxylation of the aforementioned benzenes to hydroquinones was performed at prolonged conversion times (24 h; 4 mM substrate;  $1 \mu\text{M}$  P450 BM3 M2). Halohydroquinones were the predominant products formed ( $\sim 1.5$  mM up to  $\sim 2$  mM) from halogenated benzenes and especially the conversion of iodobenzene yielded only 2-iodo-1,4-benzenediol. A mixture of *o*-phenols and hydroquinones was generated from toluene and anisole as substrates; in case of toluene nearly equimolar concentrations of mono- and di-hydroxylated products were obtained ( $\sim 0.5$  mM each). For anisole, the concentration of *o*-phenol ( $\sim 1.5$  mM) was approximately 3 times higher than that of the respective hydroquinone. Finally, under non-optimized reaction conditions, total product concentrations between  $0.24$  up to  $0.83 \text{ g L}^{-1}$  were obtained in one step syntheses. These results indicate that P450 BM3 M2-catalyzed hydroxylations offer alternative routes for semi-preparative scale synthesis of phenols and hydroquinones from the corresponding benzenes. This biosynthetic route provides significant advantages over chemical synthesis methods in terms of reduction of toxic chemical use, mild reaction conditions and energy efficiency, process intensification (one step process), regioselectivity, promiscuity and reaction yields (especially for halogenated substrates).

In P450 project II, an improvement of the mass transfer across the outer membrane of *E. coli* was achieved by recombinant co-expression of an outer membrane mutant protein, FhuA  $\Delta 1-160$ , which forms a passive diffusion channel in the outer membrane of *E. coli*. This

approach was applied to P450 BM3-catalyzed conversions due to the cofactor dependency of this enzyme. A fluorogenic coumarin substrate (BCCE) and 2 monosubstituted benzenes (toluene and anisole) were tested with the aforementioned setup. The recorded data indicated that the heterologous co-expression of FhuA  $\Delta$ 1-160 was beneficial for increasing product titers in whole cell catalysis of the coumarin substrate BCCE and both investigated monosubstituted benzenes. The whole cell conversion of BCCE was 4.1-fold higher for *E. coli* co-expressing FhuA  $\Delta$ 1-160 and P450 BM3 in relation to cells containing solely the P450 BM3 enzyme. In addition, ~10-fold relative increase of phenolic product formation was observed in a 48-hour long biotransformation with cells co-expressing FhuA  $\Delta$ 160 and P450 BM3 M2 (R47S/Y51W/I401M), a variant engineered for regiospecific, aromatic hydroxylation (Dennig et al., 2012, 2013). The relative gains in performance of the whole cell catalysis due to the presence of an outer membrane channel appeared to be of global nature, and possibly provides a general platform for alleviating the issue of limited mass transfer in whole cell biocatalysis.

As a final point, the scientific contributions made during my PhD studies include:

- Identification, expression and preliminary characterization of 3 putative polymerase gene;
- Development of Scorpion probe-based screening system for directed evolution of DNA polymerases and engineering of DNA polymerase towards improved efficiency of mismatch elongation;
- Advancement of the Sequence Saturation Mutagenesis (SeSaM) protocol;
- Development of an one step biosynthetic route to regioselective mono- and di-hydroxylation of monosubstituted benzenes;
- Development of a concept for improving the mass transfer in whole cell biocatalysis;

# APPENDIX

## Table of Figures

Figure 1. Schematic illustrating the main steps of a directed evolution campaign. ....	10
Figure 2. Mutagenesis methods embraced in 100 randomly selected papers.. ....	13
Figure 3. Graphic representation of selected methods for diversity generation. ....	16
Figure 4. Classification of random mutagenesis methods. ....	17
Figure 5. The importance of subsequent nucleotide substitutions. ....	21
Figure 6. Illustration of the two preliminary SeSaM experiments. ....	24
Figure 7. Schematic of the SeSaM method comprising four steps. ....	25
Figure 8. The obtainable mutational spectrum with degenerate base dPTP.....	27
Figure 9. Structural features of the Y-family polymerase Dpo4 from <i>Sulfolobus solfataricus</i> .....	35
Figure 10. Phylogenetic tree based on amino acid sequence data, showing the estimated relationship between a selection of Y-family polymerases.....	72
Figure 11. Cloning of the synthetic polymerase genes in pBADN expression vector. ....	73
Figure 12. Expression profiles of Y-polymerases.....	75
Figure 13. Utilization of sonication as lysis method for lysis of E. coli expressing Dpo4 polymerase..	75
Figure 14. Thermostability studies for A) Dpo4; B) Mse; C) Tps; D) Pto.. ....	77
Figure 15. SDS-PAGE gel images after pilot purification of A) Dpo4; B) Mse; C) Tps; D) Pto. ....	78
Figure 16. A) A schematic outlining forward primer elongation assay procedures; B) Monitoring of the extension of FITC-labeled primer with two 3'-attached dPTP bases by four Y-family polymerases. ...	80
Figure 17. Schematic of the compartmentalized self-replication selection system.. ....	84
Figure 18. PCR amplification with varied concentration of purified Dpo4 polymerase.....	84
Figure 19. A) <i>In vitro</i> expression of Dpo4 coupled to forward primer elongation assay in emulsions..	85
Figure 20. A) Cloning and purification of Dpo4S; B) Result of the amplification of sequences of varied length.....	85
Figure 21. A) Agarose gel visualization of 200 bp product; B) PicoGreen® dye quantification of PCR products produced with 1 µL, 2 µL or 4 µL of cell lysate containing Dpo4.....	86
Figure 22. Principle of action of the Scorpion probe assay.....	88
Figure 23. Dpo4 polymerase concentration in response of Scorpion probe fluorescence.....	89
Figure 24. Activity values in descending order of transversion mis-match elongation. ....	90
Figure 25. Generation and screening of random mutagenesis libraries by epPCR.....	93
Figure 26. Crystal structure of DNA polymerase IV from <i>Sulfolobus solfataricus</i> .....	95
Figure 27. Generation of saturation mutagenesis libraries.. ....	96
Figure 28. Scorpion probe screening system for relative quantification of polymerase activity in 96-well microtiter plate format.....	98
Figure 29. Exemplary results from re-screen of positive hits from site-saturation mutagenesis libraries at positions R242, R247 and R336.....	99
Figure 30. Purification of Dpo4 WT and 3 Dpo4 mutants. ....	100
Figure 31. Scorpion probe assay based characterization of purified Dpo4 WT and mutants R242L, R336W, R336N. ....	101
Figure 32. Forward primer elongation assay was performed with purified Dpo4 WT and the Dpo4 variants R336W and R336N. R336W and R336N. ....	101
Figure 33. DNA modification steps from I to IV of the SeSaM random mutagenesis method. ....	102
Figure 34. Schematic of a typical reaction catalyzed by cytochrome P450 monooxygenases. ....	110

Figure 35. P450 catalytic cycle (taken from Whitehouse et al., 2012).....	112
Figure 36. The crystal structure of the complex between the heme- and FMN-binding domains of bacterial cytochrome P450 BM3 (PDB ID. 1BU7) (Sevrioukova et al., 1999). .....	113
Figure 37. Three dimensional structure of the outer membrane protein FhuA. ....	119
Figure 38. Schematic summarizing the objectives and goals of the biocatalytic studies with P450 BM3 monooxygenase as a catalyst.....	121
Figure 39. Conversion of 7-benzyoxy-3-carboxy-coumarin ethyl ester (BCCE) by P450 BM3 M2.....	131
Figure 40. HPLC chromatograms after 24 h conversion of phenolic compounds to hydroquinones using P450 BM3 WT and variant M2.....	134
Figure 41. Product titers of the catalytic cascade generated by P450 BM3 WT (1 $\mu$ M) or P450 BM3 M2 (1 $\mu$ M) in 24-hour reactions (4 mM benzylic substrates, 2 vol% DMSO). .....	136
Figure 42 Conversion of A) toluene, B) anisole, C) chlorobenzene and D) bromobenzene by P450 BM3 M2 employing a cofactor regeneration system. ....	139
Figure 43. A) Schematic of the genetic constructs employed in this study; B) Exemplary agarose gels of the PCR products for PLICing and the final genetic construct verification after isolation from an <i>E. coli</i> host. ....	146
Figure 44. Verification of protein expression.....	147
Figure 45. A) Initial conversion rates of 7-benzyoxy-3-carboxy-coumarin ethyl ester (BCCE) with whole cells with empty vector and recombinantly expressed P450 BM3 M2 or co-expressed P450 BM3 M2 and FhuA $\Delta$ 1-160; B) Analysis of populations of <i>E. coli</i> cells (10 000 events) harboring P450 BM3 and co-expressed P450 BM3 M2 and FhuA $\Delta$ 1-160.....	149
Figure 46. Formation of A) <i>o</i> -cresol from toluene and B) guaiacol from anisole in whole cell catalysis with <i>E. coli</i> expressing P450 BM3 and co-expressing P450 BM3 and FhuA $\Delta$ 1-160. ....	151
Figure 47. Vector maps of Y-family polymerase genes from <i>Metallosphaera sedula</i> DSM 5348 ( <i>mse</i> ) and <i>Pictophilus torridus</i> DSM 9790 ( <i>pto</i> ), <i>Thermoanaerobacter pseudethanolicus</i> ATCC 33223 ( <i>tps</i> ) and Dpo4 from <i>Sulfolobus sulfataricus</i> ( <i>ssu</i> ) cloned in pET28a vector .....	175
Figure 48. Vector maps of Y-family polymerase genes from <i>Metallosphaera sedula</i> DSM 5348 ( <i>mse</i> ) and <i>Pictophilus torridus</i> DSM 9790 ( <i>pto</i> ), <i>Thermoanaerobacter pseudethanolicus</i> ATCC 33223 ( <i>tps</i> ) and Dpo4 from <i>Sulfolobus sulfataricus</i> ( <i>ssu</i> ) cloned in pALXtreme-1a vector.....	176
Figure 49. Vector maps of Y-family polymerase genes from <i>Metallosphaera sedula</i> DSM 5348 ( <i>mse</i> ) and <i>Pictophilus torridus</i> DSM 9790 ( <i>pto</i> ), <i>Thermoanaerobacter pseudethanolicus</i> ATCC 33223 ( <i>tps</i> ) and Dpo4 from <i>Sulfolobus sulfataricus</i> ( <i>ssu</i> ) cloned in pBADN vector. ....	177
Figure 47. Vector maps of the genetic constructs used in P450 projects I and II. ....	178

## List of Tables

Table 1. List of bacterial strains used in this study.....	38
Table 2. List of plasmids used in this study. ....	38
Table 3. List of primers, DNA probes and other oligos used in the study .....	39
Table 4. Mutational spectrum generated with the SeSaM-Tv-II protocol employing 3D1, Dpo4 WT or Dpo4 R336N in Step III.....	103
Table 5. Catalytic performance of P450 BM3 WT and variant M2 towards selected phenols. ....	133
Table 6. Catalytic performance of P450 BM3 M2 at 2 different concentrations of benzenes and NADPH cofactor provided by a cofactor regeneration system. ....	137
Table 7. Comparison of chemical and enzymatic processes for synthesis of phenols and hydroquinones .....	143

## Abbreviations

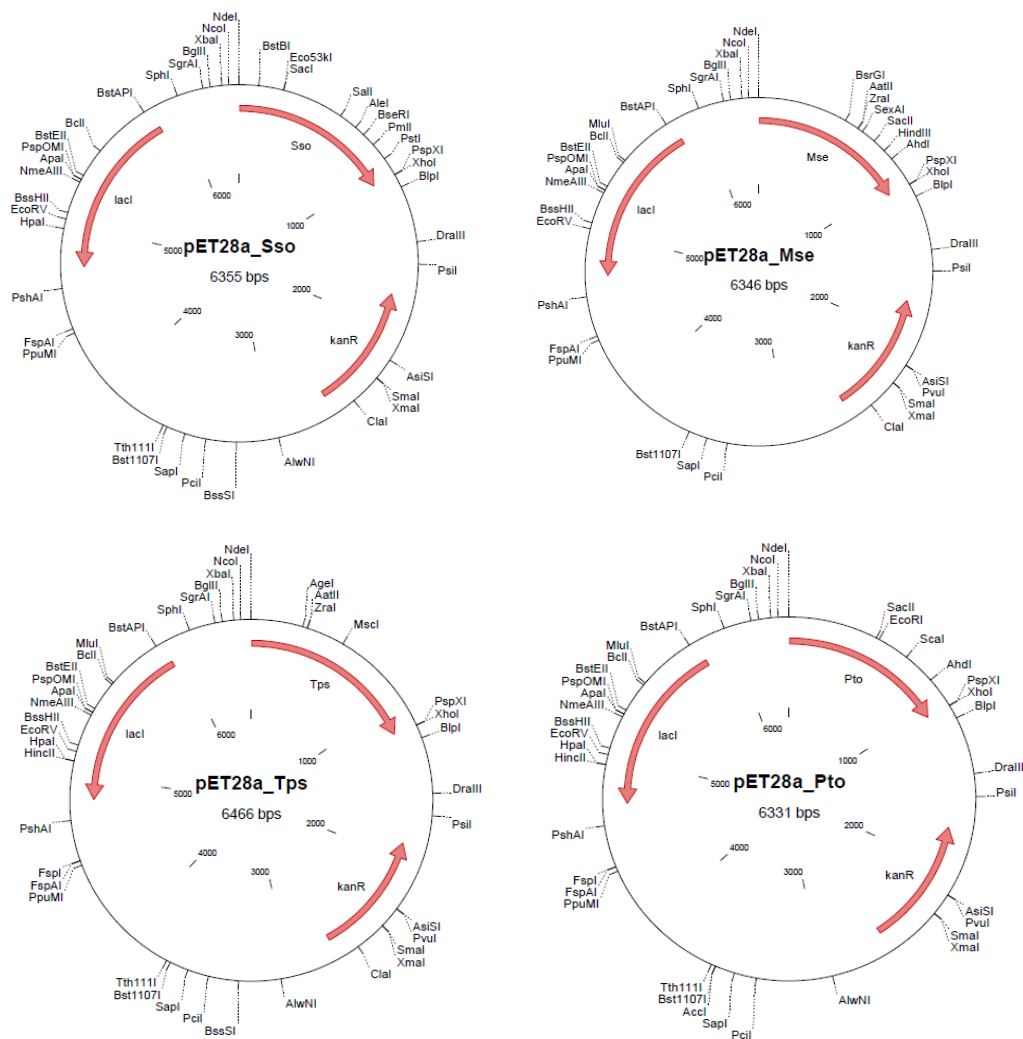
$\alpha$ S	$\alpha$ -Phosphorothioate
$^{\circ}$ C	Degree Celsius
AA	Amino acids
Amp	Ampicillin
ACN	Acetonitrile
ALA	Aminolevulinic acid
APS	Ammonium persulfate
atm	Atmospheric pressure
bar	Pressure
bp	Base pairs
BM	<i>Bacillus megaterium</i>
BSA	Bovine Serum Albumin
BSLA	<i>Bacillus subtilis</i> Lipase A
CYP	Cytochrome P450 monooxygenase
CV	Column volume
ddH <sub>2</sub> O	Double-distilled water
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
DNA	Deoxyribonucleic acid
DMSO	Dimethyl sulfoxide
dNTP	Deoxynucleotide triphosphate
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytosine triphosphate
dGTP	Deoxyguanosine triphosphate
dTTP	Deoxythymidine triphosphate
dsDNA	Double stranded Deoxyribonucleic acid
<i>E. coli</i>	<i>Escherichia coli</i>
e.g.	<i>exempli gratia</i> (for example)
epPCR	Error-prone polymerase chain reaction
et al.	<i>et alli</i>
EtOH	Ethanol
EV	Empty vector

FAD	Flavin adenine dinucleotide
FDH	Formate dehydrogenase
FID	Flame ionization detector
FITC	Fluorescein isothiocyanate
FMN	Flavin mononucleotide
Fw	Forward
g	Gravitational force
GC	Gas chromatography
GDH	Glucose dehydrogenase
h	Hour(s)
HTS	High Throughput Screening
HPLC	High performance liquid chromatography
HQu	Hydroquinone
IPTG	Isopropyl $\beta$ -D-thiogalactopyranoside
Kan	Kanamycin
kb	Kilobase(s)
kcat	Catalytic activity
KM	Michaelis Menten constant/substrate affinity constant
KPi	Potassium phosphate buffer with inorganic phosphate
LB	Lysogeny broth
Min	Minute(s)
mRNA	Messenger ribonucleic acid
mM	Milli molar
MTBE	Methyl <i>tert</i> -butyl ether
MTP	Microtiter plate
NADP <sup>+</sup>	Nicotine amide adenine dinucleotide phosphate (oxidized form)
NADPH	Nicotine amide adenine dinucleotide phosphate (reduced form)
nt	Nucleotide
OD600	Optical density at a wavelength of 600 nm
P450	Cytochrome P450 monooxygenase
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PFR	Product formation rate
pH	Decimal logarithm of the reciprocal of the hydrogen ion activity
Ph	Phenol
pKa	Logarithmic measure of the acid dissociation constant
PLICing	Phosphorothioate-based ligase-independent gene cloning
pmol	Picomoles
qPCR	Quantitative Polymerase Chain Reaction
Rv	Reverse
rpm	Revolutions per minute
RT	Room temperature
SDM	Site Directed Mutagenesis
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
sec	Seconds
SeSaM	Sequence Saturation Mutagenesis
ssDNA	Single stranded Deoxyribonucleic acid
SOC	Super optimal broth with catabolite repression

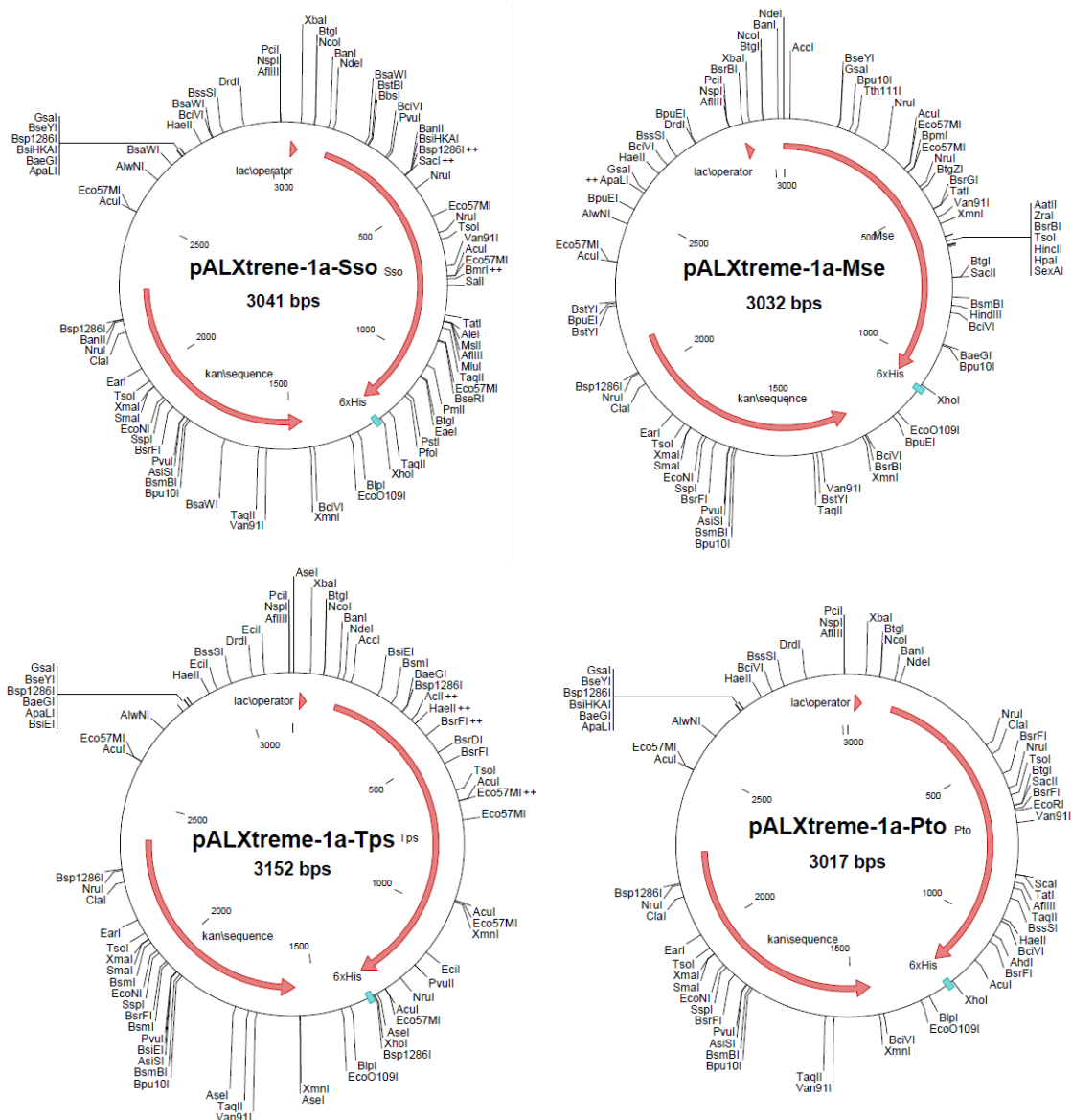


SSM	Site Saturation Mutagenesis
TB	Terrific broth
TEMED	N,N,N',N'-tetramethylethylenediamine
TTN	Total turnover number
SOC	Super optimal broth for catabolite repression
TB	terrific broth
TEMED	N,N,N',N'-tetramethylethylene diamine
TdT	Terminal deoxynucleotidyl transferase
Ts	Transitions
Tv	Transversions
UV	Ultraviolet
vs.	versus

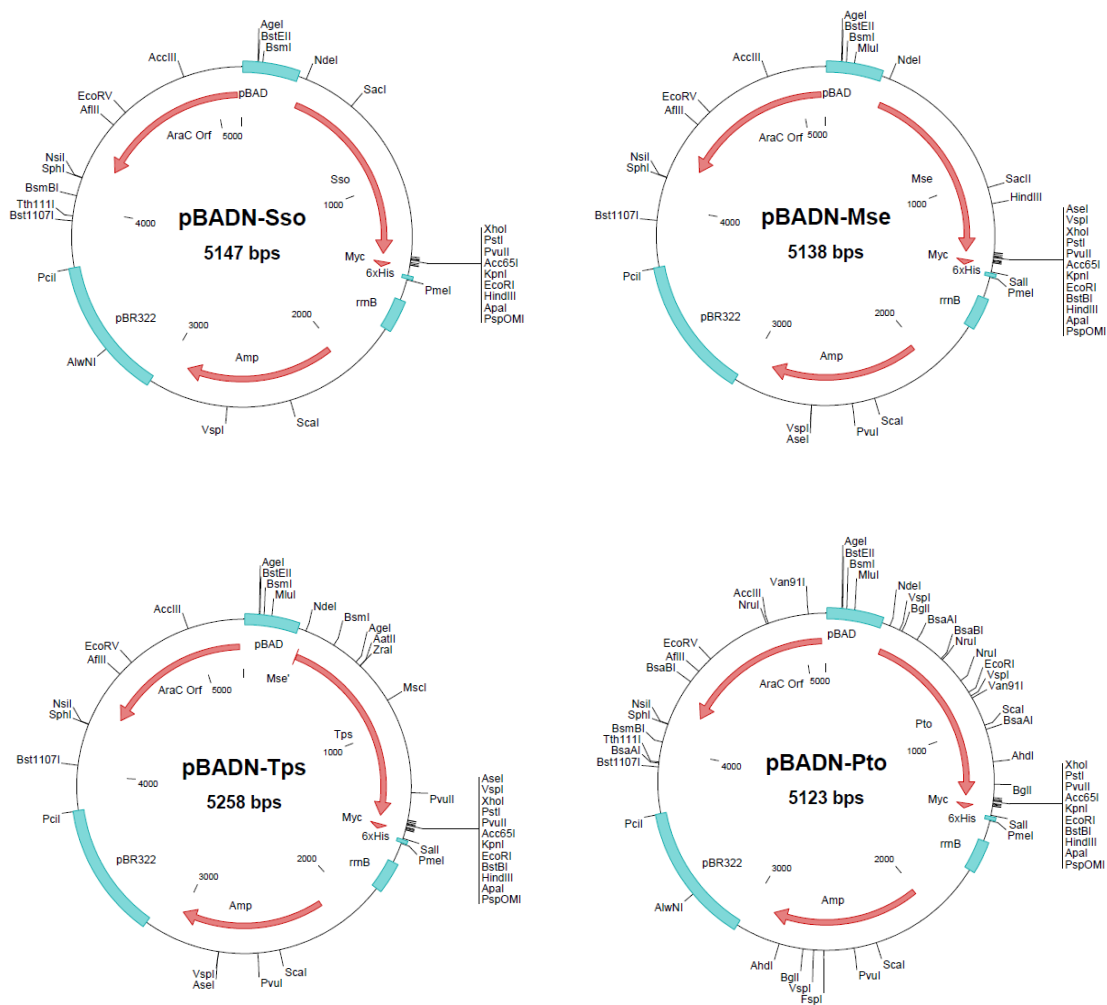
### Vector maps used in this thesis



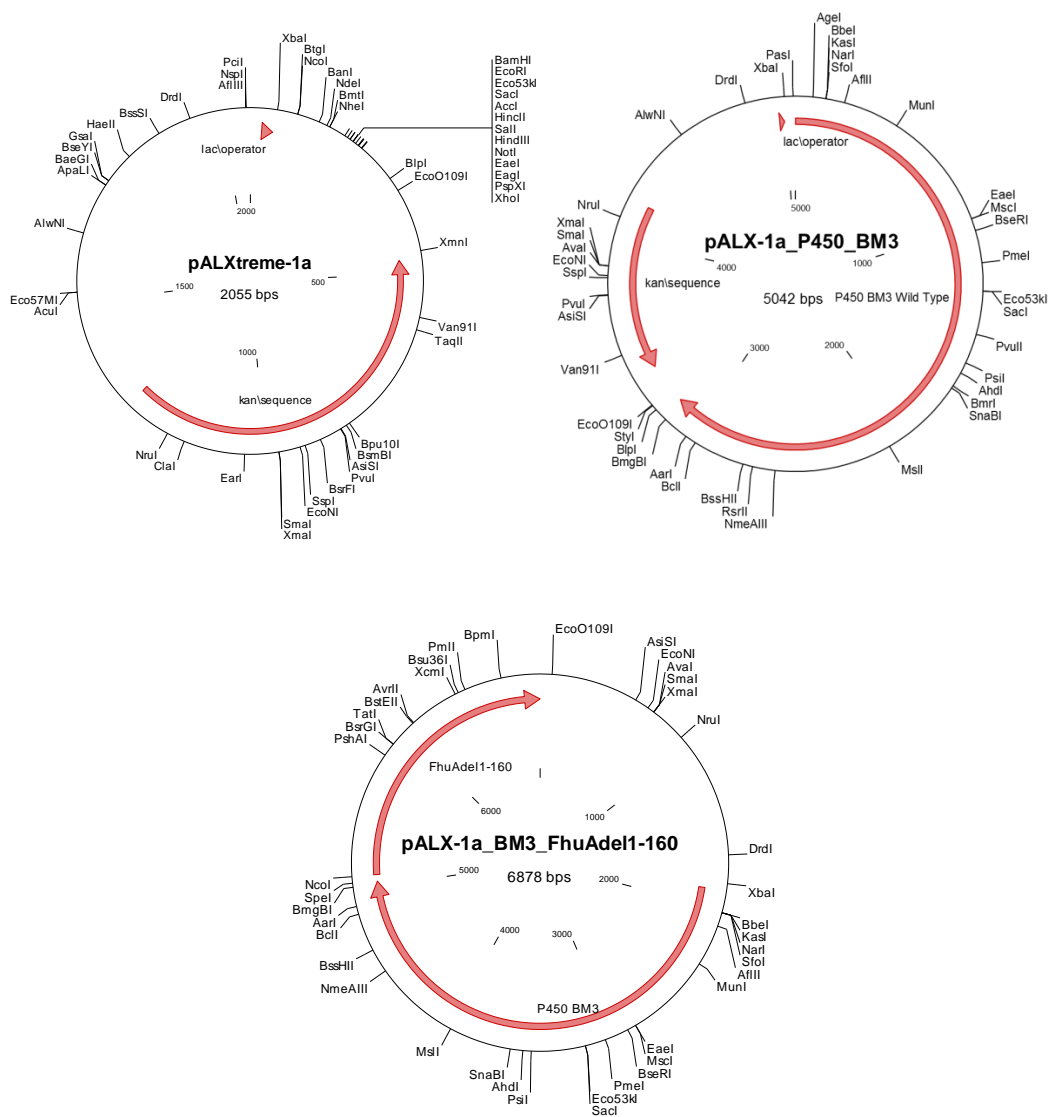
**Figure 47.** Vector maps of Y-family polymerase genes from *Metallosphaera sedula* DSM 5348 (*mse*) and *Picrophilus torridus* DSM 9790 (*pto*), *Thermoanaerobacter pseudethanolicus* ATCC 33223 (*tps*) and Dpo4 from *Sulfolobus sulfataricus* (*sso*) cloned in pET28a vector



**Figure 48.** Vector maps of Y-family polymerase genes from *Metallosphaera sedula* DSM 5348 (*mse*) and *Pictophilus torridus* DSM 9790 (*pto*), *Thermoanaerobacter pseudethanolicus* ATCC 33223 (*tps*) and Dpo4 from *Sulfolobus sulfataricus* (*sso*) cloned in pALXtreme-1a vector.



**Figure 49.** Vector maps of Y-family polymerase genes from *Metallosphaera sedula* DSM 5348 (*mse*) and *Pictophilus torridus* DSM 9790 (*pto*), *Thermoanaerobacter pseudethanolicus* ATCC 33223 (*tps*) and Dpo4 from *Sulfolobus sulfataricus* (*sso*) cloned in pBADN vector.



**Figure 50.** Vector maps of the genetic constructs used in P450 projects I and II. The gene encoding P450 BM3 was cloned in pALXtreme-1a vector as a standalone gene and used in both P450 projects. A pALXtreme-1a harboring a simple synthetic operon comprising of P450 BM3 and FhuA $\Delta$ 1-160 was generated and used in P450 Project II.

## **ACKNOWLEDGEMENTS**

Many people contributed to my personal and professional development over the last 4.5 years.

First, I would like to thank Prof. Dr. Ulrich Schwaneberg for his guidance and support throughout my PhD studies.

Thanks also to Prof. Dr. Lothar Elling and Prof. Dr. Lars Blank for being members of my dissertation committee.

I acknowledge the financial support provided by BMBF and DWI-Leibniz Institut für Interaktive Materialien.

I wish to thank all my scientific collaborators, the cluster partners from BIODIVERSITY2021, my supervisors and students guided by me over the last 4 years. I would like to thank all members of AG Schwaneberg for the technical support, scientific discussions and for sharing good times outside the lab.

Finally, I must acknowledge my parents, my brother, my friends and my girlfriend without whom I would have never made it that far. Благодаря ви от сърце!

## **STATEMENT**

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, dass alle Stellen der Arbeit, die wörtlich oder sinngemäß aus anderen Quellen übernommen wurden, als solche kenntlich gemacht sind und dass die Arbeit in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegt wurde.

Tsvetan Kardashliev

# LEBENS LAUF

## Persönliches

Name: Tsvetan Dinkov Kardashliev  
Geburtsdatum: 25.07.1986  
Geburtsort: Stara Zagora, Bulgarien  
Staatsangehörigkeit: Bulgarisch

## Ausbildung

2010 –2014

Promotion an der RWTH Aachen University, Lehrstuhl für Biotechnologie

2008 –2010

Jacobs University, Bremen, Deutschland

MSc Molecular Life Science

Master's Thesis: "Rational protein engineering of cytochrome P450cin monooxygenase for non-natural electron transfer"

2005 –2008

Jacobs University, Bremen, Deutschland

BSc Biochemistry and Cell Biology

Bachelor's Thesis: "Development of a FACS-based screening system for cytochrome P450 monooxygenases" and "Identification of outer membrane components in the antibiotic resistance of Providencia"

1999 –2005

Romain Rolland Foreign Language School, Stara Zagora, Bulgarien

Abitur