Development of a combined *in vitro* transcription and translation system for rapid cell-free screening of metagenomic DNA

Dissertation

to obtain the academic degree Doctor rerum naturalium (Dr. rer. nat.)

at the Department of Biology

Subdivision of the Faculty of Mathematics, Informatics and Natural Sciences

of the University of Hamburg

by Mareike Janus

born on 06.02.1989

in Hamburg

Genehmigt vom Fachbereich Biologie der Universität Hamburg

auf Antrag von Prof. Dr. J. Kehr

Weiterer Gutachter der Dissertation: Prof. Dr. S. J. Wittke

Tag der Disputation: 19.06.2020

"When life gives you lemons, make lemonade."

Table of Contents

List of figures

List of tables

Abbreviations

aa(s)	Amino acid(s)
AmpR	Ampicillin resistance
approx.	Approximately
ATP	Adenosine triphosphate
bidest	Bidistilled water
bp	Base pair(s)
BCIP	5-Bromo-4-chloro-3-indolyl phosphate
BSA	Bovine serum albumin
CmR	Chloramphenicol resistance
conc.	Concentrated
C. sambhunathii	*Chelatococcus sambhunathii*
DEPC	Diethylpyrocarbonate
CFPS	Cell-free protein synthesis
DNA	Deoxyribonucleic acid
DNTP(s)	Deoxyribonucleotide triphosphate
DMF	Dimethylene formamide
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
E. coli	*Escherichia coli*
E-cup	Eppendorf cups (reaction tubes)
et al.	*et alii*
EtOH	Ethanol
EDTA	Ethylene diaminetetraacetic acid
FPLC	Fast protein liquid chromatography
G	Gram
gDNA	Genomic DNA
GmR	Gentamicin resistance
HTS	High-throughput screening
I. hospitalis	*Ignicoccus hospitalis*
IgG	Immunoglobulin G
IMAC	Immobilized metal chelate affinity chromatography

IPTG	Isopropyl β-D-1-thiogalactopyranoside
ivTT	*In vitro* transcription and translation
g	Radial centrifugal force
G. thermoleoverans	*Geobacillus thermoleoverans*
KanR	Kanamycin resistance
kb	Kilobase(s)
LB	Luria Bertani or lysogeny broth
Mb	Mega base pair(s)
MCS	Multiple cloning site
4-MU	4-methylumbelliferyl
MW	Molecular weight
MWCO	Molecular weight cut off
N/A	Not available
NBT	Nitro blue tetrazolium chloride
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
NTA	Nitrilotriacetic acid
OD	Optical density
ORF	Open reading frame
P. antarctica	*Pseudomonas antarctica*
PAGE	Polyacrylamide gel electrophoresis
PB	Phosphate buffer
PBD	Polybutadien
PCR	Polymerase chain reaction
PDMS	Polydimethylsiloxane
PEO	Poly(ethylene oxide)
PET	Polyethylene terephthalate
Pfu	*Pyrococcus furiosus* (polymerase)
PMSF	Phenylmethylsulfonyl fluoride
pNP(s)	Para-nitrophenyl ester(s)[=4-nitrophenyl ester(s)]
PPase	Pyrophosphatase (inorganic)
psi	Pounds per square inch
rRNA	Ribosomal ribonucleic acid

RNA	Ribonucleic acid
RNAP(s)	RNA polymerase(s)
$RNAP_E$	RNA polymerase from elephant metagenome
RNase	Ribonuclease A
RNTP(s)	Ribonucleoside tri-phosphate
rpm	Rotation per minute
RT	Room temperature
S	Svedberg unit
SDS	Sodium dodecyl sulfate
(sf)GFP	(Superfolder) green fluorescent protein
SIGEX	substrate-induced gene expression screening
sp.	Species
T.	*Thermus*
TAE	Tris-acetate EDTA
Taq	*Thermus aquaticus* (polymerase)
TBS	Tris-buffered saline
Tc^R	Tetracycline resistance
TEMED	N, N, N´, N´ tetramethyl-ethane-1,2-diamine
T_{Ann}	Annealing temperature
T_m	Melting temperature
t_R	Retention time
U	Unit (enzyme activity)
UV	Ultra violet
Vis	Visible
v/v	Volume per volume
w/v	Weight per volume
X-Gal	5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Abbreviations of nucleic acids

A	Adenine	C	Cytosine
G	Guanine	T	Thymine

Abstract

Metagenomics is a powerful tool in microbial research, which allows the discovery of novel robust and highly versatile biocatalysts with exciting new functions relevant for industrial applications (Ferrer *et al.*, 2005; Ferrer *et al.*, 2007). Conventional function-based screening methods are extremely time consuming, expensive and laborious. Current approaches for function-based screening of metagenomic libraries have several limitations that still do not allow this technology to access the large enzyme variety of a metagenome (Beloqui *et al.*, 2008; Ferrer *et al.*, 2009; Nevondo, 2016). Thereby, the heterologous protein expression in a suitable host is the most limiting factor in function-based metagenomics. For library construction, recombinant protein expression and the screening for a specific desired function, *E. coli*-based systems are still the first choice despite the well-known problems of protein expression in bacterial host systems. Several studies are indicating that the expression of metagenomic-derived genes in *E. coli* is limited to about 40% (Gabor *et al.*, 2004; McMahon *et al.*, 2012; Felczykowska *et al.*, 2015). The genetic machinery of the host often fails in recognizing promoter sequences and transcriptional signals in the metagenome. Differences in the codon bias, the inadequate recognition of translational sinals, as well as the inability of carrying out posttranslational modifications, limit the translation rate of active proteins decisively (Uchiyama & Miyazaki, 2009). Protein misfolding and/or intracellular accumulation, the lack of an appropriate secretion system of the host cell or the degradation of the recombinant protein represents another major limitation in functional metagenomics. In addition, it is not uncommon, that heterologously expressed proteins are toxic to the host cell and significantly inhibit cell growth.

This work lays the foundation for a cell-free screening approach for function-based metagenomics. A cell-free protein synthesis (CFPS) system should overcome many limitations associated with heterologous protein expression, representing a promising alternative to conventional function-based screening methods, being host-independent, time-saving and less labor-intensive. The focus of this cell-free expression system, whose initial experiments are carried out in this work, is the *in vitro* transcription and translation of “thermozymes”, for which there is a high demand for industrial bioprocesses, that require elevated temperatures to take place (DeCastro *et al.*, 2016; Mirete *et al.*, 2016).

To improve the transcription of metagenomic genes, a bacterial RNA polymerase (RNAP) from the extreme-thermophile bacterium *Thermus thermophilus* (*T. thermophilus*) and a new, viral metagenome-derived RNAP ($RNAP_E$) were cloned, heterologously expressed and purified. *In vitro* transcription assays were carried out using both RNAPs, whereas only with the $RNAP_E$ mRNA could be synthesized. This new RNAP was discovered using sequence-based metagenomic analysis of a

microbial community from elephant feces and could be expressed heterologously with His_6-tag in *E. coli* BL21. The purification using immobilized metal chelate affinity chromatography (IMAC) resulted in up to 200 mg protein/l cell culture. The $RNAP_E$ is characterized by having only an identy of 29% in amino acid sequence compared to the T7 RNAP, but shows comparable preferences regarding assay conditions with an equeal level of transcriptional activity on a broad variety of DNA templates. Within this work, *in vitro* transcriptions with (i) uncloned, genomic DNA from different bacteria as found in a metagenome, as well as (ii) fosmids from a metagenomic library and (iii) metagenomic-derived genes coding for already characterized enzymes, were successfully carried out.

To capture the broad spectrum of potential new biocatalysts in a metagenome, *in vitro* expression experiments were performed based on cell extracts from various bacteria, both gram-negative and gram-positive strains. In addition to the commonly used extract of *E. coli* (using the strains (1.) MRE600 and (2.) CodonPlus RIL), robust cell extracts from the psychrophilic bacterium (3.) *Pseudomonas antarctica*, a mesophile (4.) *Bacillus subtilis* strain, two thermophile *Geobacillus* species, the inhouse designed strain (5.) *Geobacillus* sp. GHH01 and (6.) *Geobacillus thermoleoverans*, as well as from the thermophile strain (7.) *Chelatococcus sambhunathii* and the hyper-thermophile (8.) *T. thermophilus* were prepared. These should improve the expression of proteins, that need special conditions like extreme temperatures, to be active. CFPS with extracts from the *Geobacillus* species, as well as from *Pseudomonas antartica* could successfully been demonstrated for the first time in this work. Unfortunately, *in vitro* translations with cell extracts from *B. subtilis*, *C. sambhunathii* and *T. thermophilus* were without success.

To investigate the applicability of *in vitro* expression technologies for function-based metagenomics, model experiments were carried out with already characterized enzymes of metagenomic origin as target proteins. Heat-tolerant hydrolases, whose demand for industrial applications has been increasing rapidly in recent years, were successfully expressed *in vitro*, including metagenomic-derived, thermostable lipases from the hyperthermophilic archaeon *Ignicoccus hospitalis* (Kobus *et al.*, 2019) and recently published polyethylene terephthalate (PET)-degrading hydrolases (Danso *et al.*, 2018). In addition, a simple and convenient method has been developed for immobilizing *in vitro* expressed His_6-tagged enzymes to Ni^{2+}-NTA agarose beads and subsequent performance of function-based screening.

Finally, the -in this work- designed CFPS system was successfully tested for compatibility with advanced *in vitro* compartmentalization. By coupling CFPS packed into polymersomes and fluorescence-activated cell sorting (FACS) techniques, a new *in vitro*-based technique can be designed to overcome the low throughput rate of classical function-based metagenomic screening.

Zusammenfassung

Die Metagenomik ist ein leistungsvolles Werkzeug in der mikrobiellen Forschung, welches die Entdeckung robuster und vielseitiger Biokatalysatoren mit aufregenden, neuen Funktionen für die industrielle Nutzung, ermöglicht (Ferrer *et al.*, 2005; Ferrer *et al.*, 2007). Konventionelle funktionsbasierte Screeningmethoden sind extrem zeitaufwändig, teuer und mühsam. Die derzeitigen Ansätze für das funktionsbasierte Screening von Metagenombibliotheken unterliegen diversen Limitierungen, die es dieser Technologie bislang nicht ermöglichen, auf die große Enzymvielfalt eines Metagenoms zuzugreifen (Beloqui *et al.*, 2008; Ferrer *et al.*, 2009; Nevondo, 2016). Dabei ist die heterologe Proteinexpression in einem geeigneten Wirtsorganismus der limitierendste Faktor. Für die Konstruktion einer Metagenombibliothek, die rekombinante Proteinexpression und das Screening hinsichtlich einer spezifischen Funktion, sind *E. coli*-basierte Systeme -trotz der bekannten Probleme der Proteinexpression in bakteriellen Wirtssystemen- noch immer die erste Wahl. Diverse Studien weisen darauf hin, dass die Expressioneffizienz metagenomisch abgeleiteter Gene in *E. coli* auf etwa 40% begrenzt ist (Gabor *et al.*, 2004; McMahon *et al.*, 2012; Felczykowska *et al.*, 2015). Die genetische Maschinerie des Wirts hat häufig Schwierigkeiten bei der Erkennung von Promotorsequenzen und Transkriptionssignalen im Metagenom. Unterschiede im Codon-Bias, die unzureichende Erkennung von Translationssignalen, sowie die Unfähigkeit, posttranslationale Modifikationen vorzunehmen, begrenzen die Translationsrate aktiver Proteine entscheidend (Uchiyama und Miyazaki 2009). Proteinfehlfaltung und/oder intrazelluläre Akkumulation, das Fehlen eines geeigneten Sekretionssystems der Wirtszelle oder der Abbau des rekombinanten Proteins stellen weitere wesentliche Limitierungen in der funktionellen Metagenomik dar. Darüber hinaus ist es nicht ungewöhnlich, dass heterolog exprimierte Proteine für die Wirtszelle toxisch sind und das Zellwachstum signifikant hemmen.

Diese Arbeit legt den Grundstein für einen zellfreien Screening-Ansatz für funktionsbasierte Metagenomik. Ein zellfreies Proteinsynthesesystem (CFPS) soll die Schwierigkeiten, die mit der heterologen Expression von Proteinen metagenomischen Ursprungs verbunden sind, überwinden und eine vielversprechende Alternative zu herkömmlichen funktionsbasierten Screeningmethoden darstellen. Das System ist unabhängig vom Wirtorganismus, zeitsparend und weniger arbeitsintensiv. Der Fokus dieses zellfreien Expressionssystems, dessen erste Experimente in dieser Arbeit durchgeführt wurden, liegt auf der *in vitro* Transkription und Translation von „Thermozymen“. Für diese thermoresistenten Enzyme besteht ein hoher Bedarf für die Anwendung in industriellen Bioprozessen, die erhöhte Reaktionstemperaturen erfordern (DeCastro *et al.*, 2016; Mirete *et al.*, 2016). Um die Transkription metagenomischer Gene zu verbessern, wurde eine bakterielle RNA-Polymerase (RNAP) aus dem extrem thermophilen Bakterium *Thermus thermophilus* (*T. thermophilus*)

und eine neue, virale RNAP metagenomischen Ursprungs ($RNAP_E$) kloniert, heterolog exprimiert und gereinigt. *In vitro* Transkriptionsversuche wurden unter Verwendung beider RNAPs durchgeführt, während lediglich mit der $RNAP_E$ mRNA synthetisiert werden konnte. Diese $RNAP_E$ wurde mittels sequenzbasierter Analyse eines Metagenoms der mikrobiellen Gemeinschaft in Elefantenkot entdeckt und konnte heterolog mit His_6-Tag in *E. coli* BL21 exprimiert werden. Die Reinigung mittels immobilisierter Metallchelat-Affinitätschromatographie (IMAC) resultierte in bis zu 200 mg Protein/l Zellkultur. Die $RNAP_E$ zeichnet sich durch eine Aminosäuresequenzähnlichkeit von 29% im Vergleich zur T7 RNAP aus und zeigt vergleichbare Präferenzen hinsichtlich der Reaktionsbedingungen und eine vergleichbar effiziente Transkriptionsaktivität mit einer breiten Vielfalt von DNA-Templates.
In dieser Arbeit wurden erfolgreich *in vitro* Transkriptionen mit (i) ungeklonter, genomischer DNA verschiedener Bakterienstämme, wie sie typischerweise in einem Metagenom vorliegt, sowie (ii) Fosmiden aus einer Metagenombibliothek und (iii) von Genen metagenomischen Ursprungs, die für bereits charakterisierte Enzyme kodieren, durchgeführt.

Um das breite Spektrum potenzieller neuer Biokatalysatoren in einem Metagenom einzufangen, wurden *in vitro* Translationsexperimente durchgeführt, die auf Zellextrakten verschiedener Bakterien, sowohl gram-negativer als auch gram-positiver Stämme, basieren. Zusätzlich zu dem üblicherweise verwendeten Extrakt von *E. coli* (die Stämme (1.) MRE600 und (2.) CodonPlus RIL) wurden robuste Zellextrakte aus dem psychrophilen Bakterium (3.) *Pseudomonas antarctica*, einem mesophilen (4.) *Bacillus subtilis*-Stamm, zweier thermophiler *Geobacillus*-Stämme, dem hauseigens entwickelten Stamm (5.) *Geobacillus* sp. GHH01 und (6.) *Geobacillus thermoleoverans*, sowie aus dem thermophilen Stamm (7.) *Chelatococcus sambhunathii* und dem hyper-thermophilen (8.) *T. thermophilus* präpariert. Diese sollten die Expression von Proteinen implementieren, die spezielle Bedingungen wie extreme Temperaturen benötigen, um aktiv zu sein. So konnte in dieser Arbeit erstmals erfolgreich die CFPS mit Extrakten aus *Geobacillus* sowie aus *Pseudomonas antartica* gezeigt werden. Leider blieben *in vitro* Translationen mit Zellextrakten aus *B. subtilis*, *C. sambhunathii* und *T. thermophilus* erfolglos.

Um die Anwendbarkeit von *in vitro* Expressionstechnologien für funktionsbasierte Metagenomik zu überprüfen, wurden Modellexperimente mit bereits charakterisierten Enzymen metagenomischen Ursprungs als Zielproteine durchgeführt. Hitzestabile Hydrolasen, deren Nachfrage für die industrielle Anwendung in den letzten Jahren rapide zugenommen hat, wurden erfolgreich *in vitro* exprimiert. Dazu gehören thermostabile Lipasen aus dem hyperthermophilen Archaeon *Ignicoccus hospitalis* (Kobus *et al.*, 2019) und kürzlich publizierte Polyethylenterephthalat (PET) abbauende Hydrolasen (Danso *et al.*, 2018).
Zusätzlich wurde eine einfache und anwenderfreundliche Methode entwickelt, um *in vitro* exprimierte

His_6-getaggte Enzyme an Ni^{2+}-NTA Agarose-Beads zu immobilisieren und anschließend ein funktions-basiertes Screening durchzuführen.

Schließlich wurde das in dieser Arbeit entwickelte CFPS-System erfolgreich auf Kompatibilität mit fortschrittlicher *in vitro*-Kompartimentierung getestet. Durch die Kopplung von in Polymersomen verpacktem CFPS und fluoreszenzaktivierte Zellsortierungstechniken (FACS) kann eine neue *in vitro*-basierte Technologie entwickelt werden, um die Problematik geringer Durchsatzraten im klassischen, funktions-basierten Metagenomscreening zu lösen.

1 Introduction

1.1 Metagenomics

Microorganisms and their communities occupy every biological niche on this planet and represent the biggest part of the global biodiversity. Our current knowledge about microorganisms is mainly based on laboratory experiments with microbiological pure cultures. In fact, more than 99% of all microorganisms in the environment cannot be cultivated under laboratory conditions (Whitman *et al.*, 1998). Thus, only a tiny part of the world of microorganisms can be understood by classical experiments (Torsvik *et al.*, 1990; Amann *et al.*, 1995; Rappé & Giovannoni, 2003). The field of metagenomics was created to exploit the great potential of uncultivated microbiological organisms and to provide access to a multitude of new genes and interesting proteins, especially enzymes (Handelsman *et al.*, 1998). The "metagenome" is the entirety of the DNA of all organisms in a sample. Metagenomics uses the latest genomic technologies and bioinformatic tools to access the genetic profile of all organisms in a certain environment. It is based on the culture-independent investigation of different habitats by direct isolation of the DNA from environmental samples and subsequent sequence- or function-based analyses of the genes. In particular, genes from sources with extreme conditions, such as oceans (Mou *et al.*, 2008) and lakes (Messina *et al.*, 2016), hydrothermal vents at the bottom of the deep sea (Xie *et al.*, 2011; Placido *et al.*, 2015), arctic permafrost (Krivushin *et al.*, 2015) or deserts (Patel *et al.*, 2015) are of great interest because they represent a promising source of new, robust enzymes for biotechnology.

There are two main approaches of metagenomic studies, (1) the sequence-based and (2) the function-based approach (Figure 1). Both begin with sampling and subsequent isolation of the DNA. (1) involves direct analysis and sequencing of the DNA from an environmental sample and consists of three main steps before primer design and amplification of the gene is possible. This involves the sequencing of the metagenome, the identification of reads containing protein coding sequences (gene prediction) and the annotation of sequences by comparing with already known genes with presumed function (Venter *et al.*, 2004; Tringe & Rubin, 2005). The use of sequence-based metagenomics has become particularly popular in recent years through the application of high-throughput next generation sequencing (NGS), as it enables the decoding of several gigabases from metagnomic DNA without the necessity for time-consuming cloning (Margulies *et al.*, 2005). There are different strategies for the sequence-based approach. Techniques such as dot blot hybridisation or PCR with degenerated primers for conserved enzyme functions can lead to the detection of new functional genes (Shanks *et al.*, 2006). The use of microarrays to detect specific genes is also possible (Gentry *et al.*, 2006). Metagenomic libraries can also be screened for phylogenetic markers and the identification of new bioactive

molecules. For phylogenetic investigations, the ribosomal RNA of an environmental sample is isolated, sequenced and classified in a phylogenetic tree.

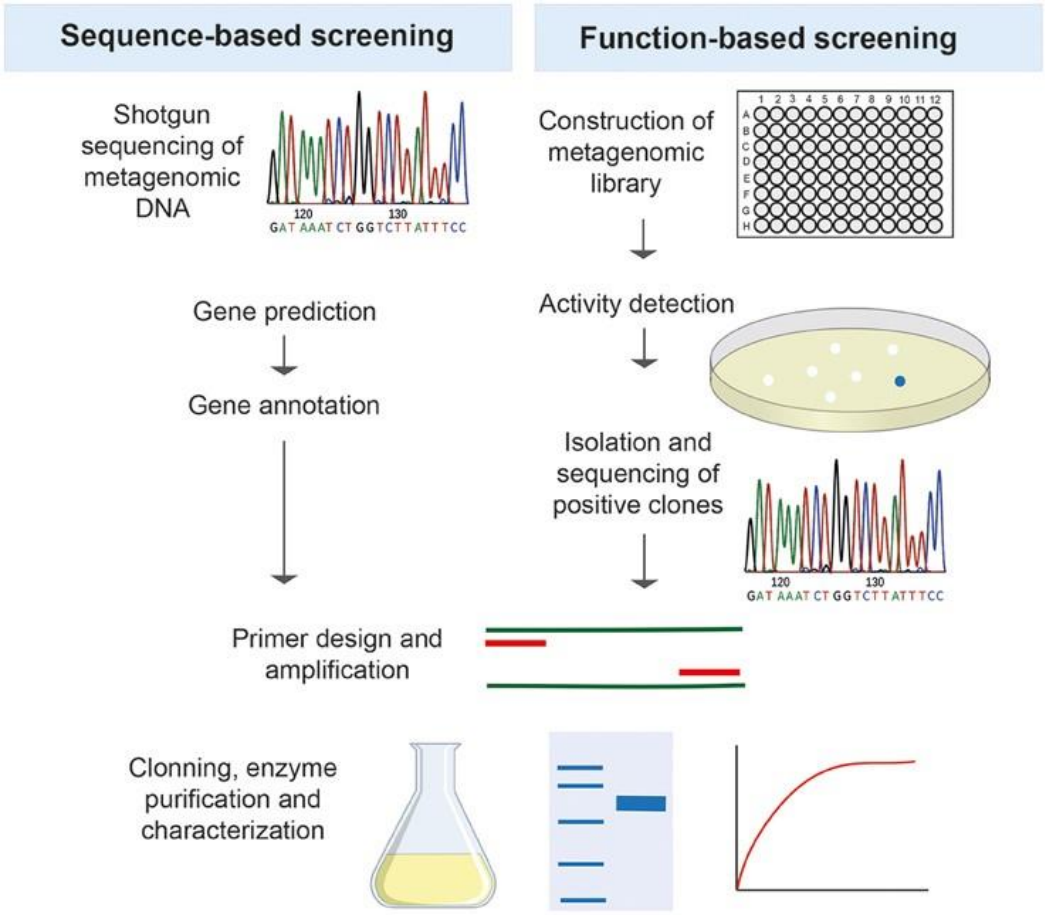

Figure 1. The two major strategies for screening for new enzymes with metagenomics.

The entire DNA or RNA is isolated from an environmental sample (not shown). To screen for new biocatalysts, there are two basic approaches: The sequence-based screening (left) is based on homologies of already known gene sequences. The function-based screening (right) is a direct approach to detect genes that show the desired function. Both result in promising genes for possible new proteins, which can then be characterized. Figure reprinted from (DeCastro *et al.*, 2016).

Sequence-based screening methods already made the identification of many new biocatalysts and pharmaceutical agents from diverse environments possible. These include esterases/lipases (Ferrer *et al.*, 2005; Rhee *et al.*, 2005), proteases from Antarctic sea water (Acevedo *et al.*, 2008), chitinases from freshwater and saline lakes, estuarine water and the central Arctic Ocean (LeCleir *et al.*, 2004), alkane hydroxylases from the Pacific deep-sea sediment (Xu *et al.*, 2008) and antibiotics (Gillespie *et al.*, 2002).

1.1.1 Functional screening for new enzymes in metagenomics

Since sequence-based screening is based on homology, no real novel enzymes can be identified and, despite sequence similarity, functionality is not guaranteed (Streit & Schmitz, 2004). The effectiveness of the sequence-based approach is largely dependent on the accuracy of genome annotation and the completeness of the sequence present in databases (Ferrer *et al.*, 2009; Ngara & Zhang, 2018). This approach, for example, is limited when the novel gene has a weak similarity to genes whose products have already been characterized biochemically. To exploit the full potential of metagenomics, the function-based approach is being used more and more frequently.

The function-based screening is a straight forward approach to detect genes that show the desired function. It begins with the construction of a metagnomic library (Figure 1), whereas cosmid- and fosmid-based systems are often preferred due to their high cloning efficiency and the large and cosistent size of their insert (Lam *et al.*, 2015). The constructs are then transformed into a suitable host, usually *E. coli*, which expresses the metagenomic-derived genes by induction. After library construction, the clones can be tested for their activity on a specific substrate in a high-throughput procedure, or entire substrate profiles can be created (Reyes-Duarte *et al.*, 2012; Reyes-Duarte *et al.*, 2018). Function-based screening is often microtiter plate-based and is performed fluorometrically or colorimetrically in small volumes by conversion of the substrate. Alternatively, plate tests are performed on substrate containing culture media to which the clones are transferred. Positive hits are detected, for example, by the growth of the clones themselves, a color change or the formation of halos by substrate degradation on the plate. This is followed by a more intensive examination of the genetic material of the positive cosmid/fosmid clones by sequencing. As with the sequence-based approach, the next step is the amplification and cloning of the gene of interest and the subsequent overexpression and purification of the enzyme. If the enzyme can be successfully produced in sufficient quantities by the expression host, it is characterized in detail to explore its biotechnological potential. Function-based metagenomics, thus, offers the possibility to discover novel proteins with either already known or completely new functions (Daniel, 2005).

An important benefit of function-based metagenomics is the elucidation of the rich biochemical diversity of enzymes adapted to extreme conditions. These include extremely low and high temperatures, extreme pH values, salt concentrations and pressures (Popovic *et al.*, 2015). In this way, novel molecular mechanisms for adaptation to extreme conditions can be investigated by biochemical characterization of metagenomic-derived enzymes (Ferrer *et al.*, 2007; Alcaide *et al.*, 2015; Tchigvintsev *et al.*, 2015). The majority of all metagenomic-derived enzymes originates from function-based screenings. These are in particular hydrolytic enzymes such as lipases/esterases (Simon & Daniel,

2009; Reyes-Duarte *et al.*, 2018), cellulases from deep sea archaea (Leis *et al.*, 2015b; Escuder-Rodríguez *et al.*, 2018), proteases (Popovic *et al.*, 2015) and chitinases (Berini *et al.*, 2017).

1.1.2 Challenges and limitations

However, there are several factors, that limitate the full potential offered by functional metagenomics. The challenges of the function-based approach begin with the extraction of sufficient amounts of DNA of necessary quality from the enviromental sample. Unfortunately, the extracted DNA often shows insufficient purity and integrity (Kinfu *et al.*, 2017). In comparison to the sequence-driven approach, function-based metagenomics is much more time-consuming and labor-intensive, as well as more expensive. In addition to the cloning steps, the heterologous expression in the appropriate host organism and the subsequent detection of the activity, in particular, is a time-consuming process. The main problem of classical function-based screening, however, is based on problems of gene expression and incorrect processing of the proteins by the host (Streit *et al.*, 2004). Many reports are indicating that the expression of metagenomic-derived genes in the most popular host organism - *E. coli* - is limited to about 40% (Gabor *et al.*, 2004; McMahon *et al.*, 2012; Felczykowska *et al.*, 2015). The reason for that are the significant differences in the expression machineries between the different taxonomic groups of the organisms found in a metagenomic sample.

Problems appear on 3 different levels: (1) Transcription; (2) Translation; (3) Protein folding, processing, toxicity, etc.: (1) The genetic machinery of the host often fails in recognizing promoter sequences and transcriptional signals in the metagenome. Furthermore, the metagenome might contain sequences which are innocuous in the natural host cell, but acts as termination signals in *E. coli*, which leads to a premature termination of the gene expression. A problem on the level of translation (2) is the codon bias. Each organism has its individual bias towards preferred codons which corresponds to the concentration of different tRNAs within the cell and the efficiency to recognize the different codons. If a metagenome contains a high amount of unfavoured codons, the translation efficiency of the host cell is reduced, which leads to a low amount of synthesized protein (Gomes *et al.*, 2016). Another challenge is the recognition of translational signals in the metagenome. For example, the initiation codon for translation in *E. coli* is biased toward AUG, whereas some organisms prefer GUG and UUG as well (Uchiyama & Miyazaki, 2009). Last but not least, the step of protein folding and processing (3) represents a decisive limitation in the recombinant expression of metagenomic proteins. *E. coli* is limited in the ability to form extensive disulfide bonds and is not able to carry out posttranslational modification like N and O linked glycosylation, phosphorylation and fatty acid acylation. This often leads to misfolding of the secondary, tertiary and quaternary structures of the protein of interest which looses its characteristic function or becomes unsoluble and forms unclusion bodies. In addition to

protein misfolding and/or intracellular accumulation, the lack of an appropriate secretion system of the host cell or the degradation of the recombinant protein represents a major limitation in functional metagenomics. In addition, it is not uncommon, that heterologously expressed proteins are toxic to the host cell and significantly inhibit cell growth.

There are different starting points to overcome the host-dependent limitations. One possible approach is the engineering of the transcription and translation machinery of the host organism (Bernstein *et al.*, 2004). To overcome problems caused by a different codon usage, which leads to no translation of the protein at all, the host organism can be supplemented with additional tRNA genes. Strains with extended codon use due to additional plasmid-coded eukaryotic tRNA genes are already available (e.g. *E. coli* expression strains Rosetta, BL21(DE3) CodonPlus RIL, NOVAGEN) (Christian Leggewie, 2005). In addition, the supplementation of the expression host with different chaperones from various organisms could be beneficial for obtaining the correct enzyme folding/conformation.
The heterologous host could also be equipped with an additional secretion machinery, to enable the secretion of the metagenomic-derived protein, which can be essential for the detection of enzyme activity.
Another approach is the use of alternative expression hosts. Alternative bacterial systems for function-based metagenomics are already in use, for example *Streptomyces lividans*, *Pseudomonas putida*, *Sinorhizobium meliloti, Rhizobium leguminosarum*, *Bacillus subtilis* and *Thermus thermophilus* (*T. thermophilus*) (Wang *et al.*, 2000; Courtois *et al.*, 2003; Martinez *et al.*, 2004; Li *et al.*, 2005; Wexler *et al.*, 2005; Angelov *et al.*, 2009; Cheng *et al.*, 2014).

Also, eukaryotic expression hosts are available. Protein expression systems based on filamentous fungi are used for large-scale industrial fermentation (Nevalainen *et al.*, 2005). Other eukaryotic hosts used for heterologous protein expression are single-cell yeasts like *Saccharomyces cerevisiae* and *Pichia pastoris*. They are characterized by special features, like posttranslational modifications, disulfide bond formation, glycosylation, protein folding and processing (Weidner *et al.*, 2010). Especially the expression of eukaryotic genes and metatranscriptome samples was successfully demonstrated with the use of *S. cerevisiae* (Damon *et al.*, 2011; Kellner *et al.*, 2011).

In addition to host-dependent limitations, the activity screening step is another challenge. Activity screening should ensure a high throughput of samples, as most of the metagenomic library clones do not show any activity due to host-dependant limitations, was well as the use of specific reaction conditions that address only a few enzymes. Furthermore, there are still too few substrates available for screening (Fernandez-Arrojo *et al.*, 2010).

1.2 Cell-free protein synthesis

Cell-free protein synthesis (CFPS) is a transcription and translation system that is independent of metabolic activities associated with cell maintenance and growth (Shrestha *et al.*, 2012). This *in vitro* approach allows the direct expression of proteins by incubation of a complex mixture. Aminoacyl-tRNA synthetases, ribosomes, translation initiation and elongation factors, chaperons and others provided by cell extracts (Kim *et al.*, 1996), and additional components like amino acids, an energy-regeneration system, cofactors, salts and nucleotides, which are essential for the transcription and translation. The *in vitro* transcription itself is usually initiated by addition of an RNA polymerase (RNAP). The CFPS thus takes place in small volumes in the µl range and is performed by incubation at a desired temperature for about 2 to 6 hours or longer (Carlson *et al.*, 2012).

Basically there are 2 different approaches of the CFPS; (1) the coupled and (2) the linked approach. These differ mainly in the kind of template and the reaction space. In (1), DNA is used as a template, with transcription and translation running spatially and temporally in parallel. This system is fundamentally comparable to the protein expression process in prokaryotes. While the RNAP transcribes, translation of the generated mRNA can be initiated at the ribosomes in parallel. The coupled system thus runs in one step. Alternatively, (2) can be performed in which the *in vitro* transcription takes place followed by RNA purification. In a second step, the purified RNA is used for *in vitro* translation, which is performed in another reaction tube. The transcription and translation reactions thus take place separated from each other, like in the eukaryotic expression system (Richter, 2014).

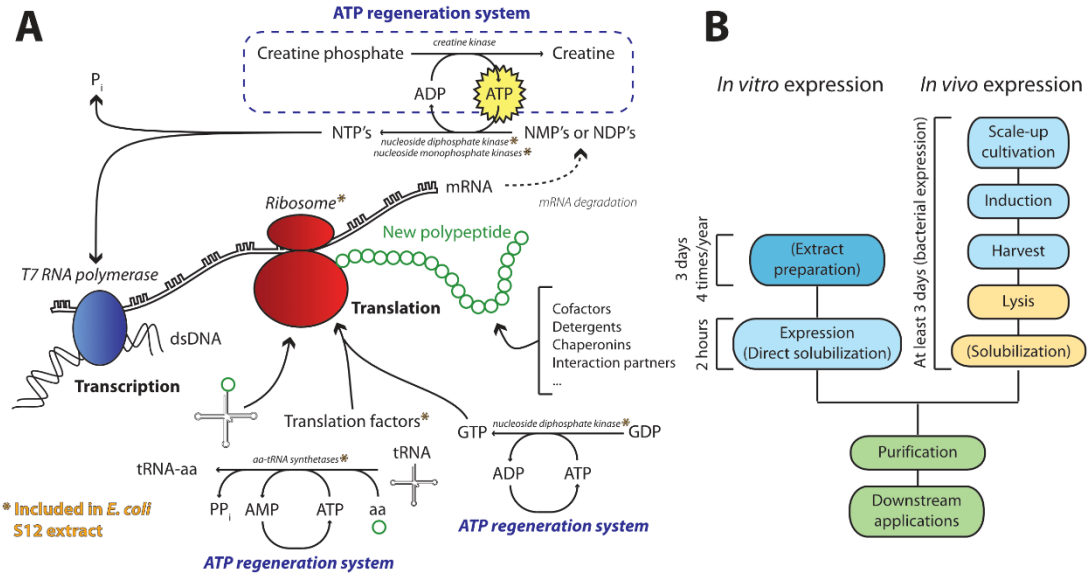

Figure 2 (A) General scheme of the process of protein biosynthesis with the labelling (*) of the individual components provided by the cell extract in the CFPS.

These include in particular the ribosomes and other essential components for translation, as well as enzymes for energy regeneration. **(B) Comparison of *in vitro* and *in vivo* protein expression in terms of time consumption.** Figure reprinted from https://nmr.gu.se/english/ research/cell-free-expression

CFPS is used to perform mutagenesis studies and to verify the identity of cloned genes, to study protein-protein, protein-nucleic acid, and protein-drug interactions (Goyal, 2018). In the past years, CFPS has rapidly become a highly preferred approach for functional and structural studies of proteins in high throughput scale. It has become a versatile tool for synthetic biology, *in vitro* protein evolution, synthesis of protein libraries for functional genomics, the expression of virus-like particles, the production of personalized medicines and others (Smutzer, 2001; Carlson *et al.*, 2012). However, *in vitro* transcription could only be made possible by the identification of (bacterio-)phage DNA-dependent RNAPs and their promoters. In contrast to the complex bacterial RNAPs, viral RNAPs (like those from bacteriophages T3 or T7) are single subunit enzymes, which, have up to five times faster transcription rates compared with bacterial RNAPs (Butler & Chamberlin, 1982; Moll *et al.*, 2004).

1.2.1 DNA-dependant RNA polymerases for *in vitro* transcription

There are four basic components needed to accomplish CFPS. The genetic template (mRNA or DNA) encoding the desired protein, an RNAP for mRNA sythesis, a reaction solution supplied with the above-mentioned components, and a cell extract. The step of transcription is usually made possible by a viral RNAP, whereas template DNA can be linear, a circular plasmid or a PCR amplificate. The key factor is the promoter sequence upstream of the gene to be transcribed. Commercial viral systems are often used for *in vitro* transcription. Three different phage RNAPs (T3, T7 and SP6) are available (*Cell-Free Protein Expression | Thermo Fisher Scientific - DE*).

Table 1. Comparison of the promoter sequences recognized by the viral RNAP T3, T7 and SP6 and the promoter of the well-known bacterial *lac* operon.

RNAP	Promotor sequence (5´- 3´)
T3	AATTAACCCTCACTAAAGGG
T7	TAATACGACTCACTATAGGG
SP6	AATTTAGGTGACACTATAGAA
lac	GGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGA

Commercially available cloning vectors containing the specific promoter sequences for bacteriophage RNAPs, like the pET-system, can be used for *in vitro* transcription. Alternatively, template DNA can be generated via PCR using gene-specific primers containing the promotor-sequence at the 5´end of the forward (or upstream) primer. Single-subunit RNAPs, encoded by bacteriophages, such as T3 or T7, differ fundamentally from multi-subunit RNAPs (Figure 3) by recognizing specific promoter sequences without the need of accessory and regulatory factors (Werner & Grohmann, 2011).

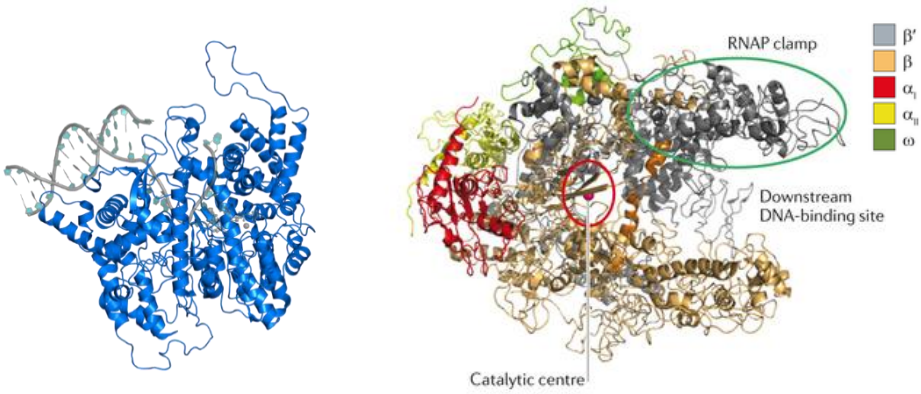

Figure 3. Comparison of the basic structures of viral and bacterial RNAPs.

The structures of the viral RNAP as a T7-promoter complex (left; based on the *Enterobacteria* phage T7 Protein Data Bank entry 2pi4) and the bacterial RNAP core enzyme (right; based on the *Thermus aquaticus* Protein Data Bank entry 1I6V, source: Werner and Grohmann 2011). The T7 RNAP is a single-polypeptide-chain RNAP, whereas the bacterial RNAP core-enzyme consists of 5 subunits designated α, α, β', β, and ω. The holo enzyme contains an additional σ-factor, which is important for initiation of the transcription at specific promoter sites. Individual subunits are color coded.

1.2.1.1 The T7 RNAP

The viral T7 RNAP is a member of the single-subunit RNAPs which include phage RNAPs such as the T3, K11, SP6 and others, as well as mitochondrial RNAPs (McAllister & Raskin, 1993; Sousa & Mukherjee, 2003). Therefore, the T7 RNAP is also closer related to mitochondrial RNAPs than, for example, the DNA-dependent RNAP from *E. coli*. The most important difference between viral and mitochondrial RNAPs is that they exhibit no sequence similarity in the N-terminal domain, which participates in the interaction of the RNAP at promoter sequences and changes its conformation to enlarge the active site and create an exit tunnel for the transcript. This fact leads to the conclusion that the N-terminal domain is an additional, evolutive feature and hence the starting point of the divergent development of yeast and phage enzymes (Masters *et al.*, 1987).

The T7 RNAP consists of a single amino acid chain (878 amino acids) with a total molecular weight of approximately 99 kDa. It is characterized by its high promoter specificity and the transcription direction is downstream of the T7 promoter. It features a very high activity, elongating mRNA strands five times faster than the RNAP from *E. coli* and terminates transcription less frequently (Golomb & Chamberlin, 1974).

The T7 RNAP is an often-used tool in molecular biology applications, as it excels in a high transcription rate and a very low error rate. For the synthesis of RNA it requires double strand DNA as template and Mg^{2+} ions as cofactor.

1.2.1.2 Bacterial RNAPs

Bacterial RNAPs (Figure 3) belong to the family of multi-subunit RNAPs, whose core is composed of at least 5 subunits conserved in all three domains of life (Werner & Grohmann, 2011). The core enzyme consisting of the subunits α (two copies), β, β', and ω with a total molecular mass of approximately 400 kDa, is responsible for DNA-binding (Cramer, 2002). The core of the RNAP is evolutionarily conserved in sequence, structure and function from archaea, bacteria to all eukaryotes (Ebright, 2000). Together with the σ-factor, the holo enzyme forms, which allows promotor-specific transcription (Gross *et al.*, 1998). The σ-factor, composed of alpha helices connected by turns and loop structures, disconnects to the core RNAP after synthesis of 9-12 nucleotide RNA and the elongation process starts. ß and ß´, the two largest subunits form a "crab claw pincer" structure (Figure 3), in which the DNA can bind (Vassylyev *et al.*, 2002). The α subunit homodimer is essential for RNAP assembly and also is involved in the regulation of the transcription because it interacts with many transcription factors and is important for binding of the β and β' subunit (David Marcey & Nathan Silva, 2006; Murakami, 2015). The role of the ω subunit is not yet completely understood, but it possibly assists the final step of the assembly of the RNAP core by binding to the ß´subunit (Gunnelius *et al.*, 2014).

1.2.2 *In vitro* transcription and translation

1.2.2.1 Additional components for transcription

As already mentioned, for the transcription process, especially for the initation and termination step, further proteins other than the RNAP are needed. Those are, for example, the cAMP binding protein and a variety of sigma factors, which play an important role for initation in prokaryotic systems. Sigma factors are responsible for recognition and binding of specific promoters at the -35 sequence (that is TTGACA) and initiate the transcription.

The Nus-A protein for example can affect the elongation step. For the dissociation from the DNA template and the release of the transcript, prokaryotic RNAPs need termination signals, either protein-based (rho-dependent) or RNA-based (rho-independent). First is controlled by the rho factor (rho protein), which binds to the mRNA and tracks along behind the RNAP during transcription of the mRNA. The rho-independent termination is controlled by a C-G rich region, which is transcribed in the end and forms a hairpin structure, which causes the RNAP to pause. The hairpin structure is followed by a region rich in A-T nucleotides, which forms an unstable interaction with the template DNA and leads to the release of the RNAP and the transcript (Clark *et al.*, 2019).

To generate transcripts of mature mRNA for translation experiments, additional features of the template are needed. For example, prokaryotic systems require the Shine Dalgano sequence, which is

recognized by the ribosomes as part of the ribosomal binding site (RBS) and thus marks the start of translation. The Shine Dalgano sequence is a purine-rich region, which is complementary to the 3' end of the 16s rRNA in the 30S ribosomal subunit and contains the initiation codons AUG, GUG and UUG. This allows the pairing between the mRNA and the ribosome (Cooper, 2000).

1.2.2.2 Cell extract for *in vitro* translations

There are a variety of different CFPS systems. The main components of a cell-free reaction are the DNA with the desired gene, a RNAP, cofactors such as magnesium, components for energy regeneration, a supply of amino acids and cell extract containing ribosomes (Figure 2). In addition, tRNAs, RNase inhibitors, as well as transcription and translation factors can be supplied.

Cell extracts contribute many important components of the transcriptional and in particular the translational machinery for *in vitro* protein expression. These include, in addition to the amino acids and tRNAs, intact ribosomes, enzymatic cofactors and cellular components for translation and subsequent correct folding of the proteins like aminoacyl-tRNA synthetases, nucleases, translation initiation and elongation factors, etc. Cell extracts provide essential components for energy regeneration. In order to translate larger quantities of protein, extremely high amounts of key molecules for adenosine triphosphate (ATP) regeneneration are also added, such as the glycolytic intermediates phosphoenolpyruvate and 3-phosphoglycerate or acetyl phosphate and creatine phosphate (Calhoun & Swartz, 2007). For example, Anderson *et al.* have developed a cell-free yeast-based system that uses glucose for energy regeneration (Anderson *et al.*, 2015).

To prepare bacterial cell extract for *in vitro* experiments, the cells of the desired strain are harvested in the mid-exponential phase and lysed. Cell membranes, genomic DNA and other debris are usually removed by centrifugation. The remaining cytosolic and organelle components of the cells are added to the *in vitro* protein expression assay in different proportions. The commonly used term "S30 extract" is due to the fact that the extract is usually obtained by high-speed centrifugation of disrupted cells at 30,000 xg.

The first kinds of lysates developed for *in vitro* protein expression were derived from prokaryotes. Early protocols for *E. coli*-based cell-free extract preparation were published in the 1960s (LAMBORG & ZAMECNIK, 1960; Nirenberg & Matthaei, 1961; Moore & Shaner, 1968; Chong, 2014). Since then many improvements have been made with the focus on higher efficiency and reducing costs. By using high density fermantors the biomass yield can be drastically increased (Liu *et al.*, 2005) and for cell lysis French press and sonication proved to be a simple, cost-effective method (Shrestha *et al.*, 2012).

1.2.3 Existing systems

As already mentioned, first cell-free protein expression experiments were carried out in the early 1960s (Nirenberg & Matthaei, 1961), whereby protein yields in the nmol range were only achieved from the end of the 1980s on (Spirin *et al.*, 1988). A variety of different proteins that are difficult to express *in vivo*, could be synthesized by cell-free expression. These include for example membrane proteins, virus particles and toxic proteins (Carlson *et al.*, 2012).

The choice of the appropriate CFPS system depends largely on the end product application and the protein origin. In addition, factors such as protein yield, purity, cost and required time play an important role. In principle, cell extracts used for CFPS can be obtained from all culturable cell types. The most widely used system for expression of proteins without posttranslational modifications is based on *E. coli* extract due to its simplicity and well-known molecular machinery (Carlson *et al.*, 2012). Alternative systems based on cell extract of prokaryotic origin, such as *Vibrio natriegens* or *Rhodococcus erythropolis*, however, are rare to find (Nevondo, 2016; Des Soye *et al.*, 2018). To date, mainly extracts of *E. coli*, yeast, wheat germ, rabbit reticulocytes and insect cells are used (Erickson & Blobel, 1983; Jackson & Hunt, 1983; Spirin *et al.*, 1988; Richter, 2014). Furthermore, the successful preparation of cell extracts from human HeLa cells, archaea, *Xenopus* eggs and *Drosophila* embryos could be shown (Mikami *et al.*, 2006; Richter, 2014).

There are different CFPS systems commercially available. Most of these systems, such as the "PURExpress® *In Vitro* Protein Synthesis Kit" (New England Biolabs, Frankfurt, Germany), consist of heterologously expressed and purified proteins rather than cell extracts. Usually, commercially available CFPS systems provide DNA expression vectors for cloning the gene of interest followed by a coupled transcription and translation reaction (Chong, 2014).

In addition to the choice of cell extract, the CFPS systems also differ in their spatial and temporal structure. The reaction can take place in a batch process, ie as a coupled reaction in a single compartment, with the reaction time mostly lasting from 30 minutes to 4 hours. This method is particularly suitable for high-throughput screening (HTS) with a variety of conditions. Also, CFPS can be continuously fed with new substrates and thus the protein yield can be massively increased over long reaction times (Stech *et al.*, 2014). Alternatively, CFPS reactions can be performed in a variety of formats, including continuous flow or exchange through a dialysis/ semipermeable membrane, or microfluidic formats (1.4). Continuous reactions use a two-chamber system for supplementation of rectants and removement of products or byproducts (Gregorio *et al.*, 2019). In the continuous exchange system, the CFPS reaction and a reactant-rich feed solution are separated via a semi-permeable membrane, whereas the protein product stays in the reaction compartment (Hong *et al.*, 2015). In the continuous flow system, the feed solution is pumped into the reaction chamber, while

products and byproducts are pushed out continuously through an utrafiltration membrane (Volyanik *et al.*, 1993).

1.2.4 Advantages and disadvantages

CFPS is a simplified and effective technique for transcribing and translating a particular protein without the necessity to pay attention to the vitality of a living cell/organism. In contrast to the classical *in vivo* approach, this method is time-saving, simple and can be carried out in high throughput for a variety of proteins at the same time. In particular, toxicity and solubility issues confronted with when expressing *in vivo* can be overcome. Challenging proteins such as functional membrane proteins can be produced by *in vitro* technologies in sufficient amounts, so that even crystallographic and biochemical analyses can be carried out (Abdine *et al.*, 2012; Stech *et al.*, 2014). Proteins that have previously been toxic to most host organisms by suppressing cell metabolism were expressed for the first time. mRNA produced during transcription of foreign genes is highly susceptible to digestion from exonucleases when they have no significance to the matebolic activities from the host organism (Baneyx, 1999). This often leads to the fact, that genes of interest which have fewer copies in a metagenome are entirely degraded at mRNA level by the host cell. And even translated proteins are susceptible to degradation by host proteases (Dedhia *et al.*, 2010). By adding RNase and protease inhibitors in CFPS, these processes can be completely avoided. Alternatively, protease or RNase I knockout mutants can be used as strains for cell extracts preparation like for example the *E. coli* strain MRE600 (Kurylo *et al.*, 2016).

The disadvantages of CFPS are mainly the often very low protein yield and the cost factor. In contrast to the usual culturing of the expression strain, for example *E. coli*, the costs of the partially commercially purchased individual components of the *in vitro* reaction are much higher. Especially with larger reaction volumes, the cost of e.g. the substrates for ATP regeneration or additional RNase inhibitors can add up to an inappropriate level.

1.3 Cell-free protein expression for function-based metagenomics

The major “bottleneck” in function-based metagnomics is the heterologous gene expression (Uchiyama & Miyazaki, 2009). Gabor et al. estimated by *in silico* analysis that only 40% of the enzymatic potential can be identified by cloning of metagenomic DNA in *E. coli* (Gabor *et al.*, 2004). Other studies have also shown that the heterologous expression host is the most challenging limitation in the discovery of new, promising biocatalysts from environmental samples (Ferrer *et al.*, 2016). While cell culture-independent technologies are finding increasing application in the field of biochemistry, their potential to discover completely new enzymes remains unexploited. Therefore, an innovative, host independent system for function-based metagenomics is being developed (Kinfu *et al.*, 2017). Which means, that a metagenomic library is generated on DNA-level and proteins are expressed *in vitro* with the use of an CFPS system established in this work. Elaborate heterologous protein expression in appropriate host strains, as well as cell harvest, disruption and protein purification omits in this system.

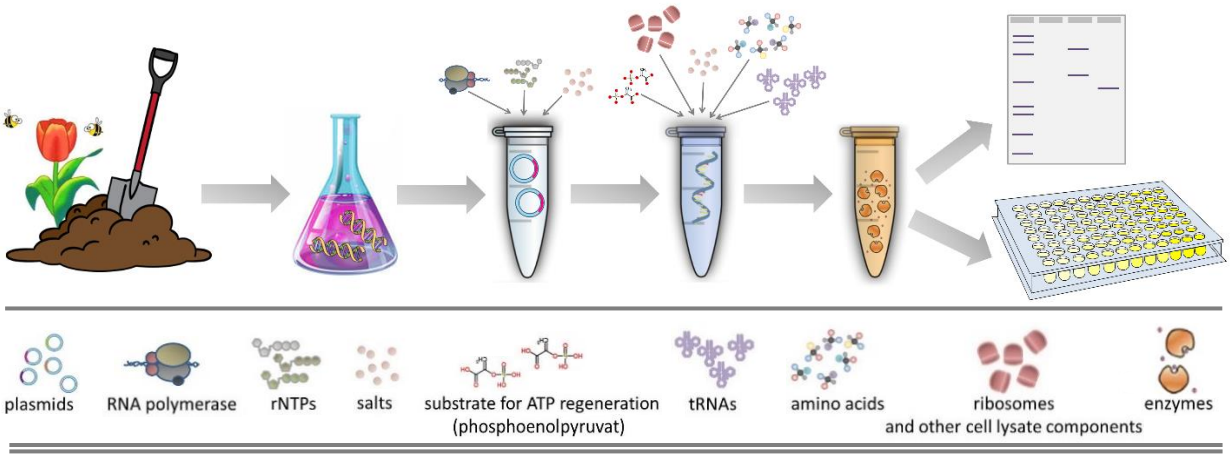

Figure 4. Scheme of the combined *in vitro* transcription and translation strategy for rapid cell-free screening of metagenomic DNA.

Metagenomic DNA is isolated from an environmental sample and used with as little cloning as possible for cell-free, function-based screening. The *in vitro* protein expression is either spatially and temporally coupled, as common in prokaryotic living systems. Or the protein synthesis proceeds in 2 steps, the *in vitro* transcription, an optional RNA purification and the *in vitro* translation. The individual components of the “expression cocktail” are shown exemplarily in the scheme. For the final step in the screening process, microtiter plate-based activity tests are performed and positive hits are analyzed more closely. In parallel, tagged target proteins can be detected, for example, by Western blot analysis.

The first step involves sampling and subsequent isolation of the total DNA of an environmental sample. Then there are 2 options. Either the DNA is directly used for cell-free protein expression without extensive cloning steps or a metagenomic library in a suitable vector system is constructed and individual clones are screened. Instead of transformation/transfection of a suitable host and

performing classical heterologous expression in, for example, *E. coli*, the cell-free system is used. A cocktail with all essential components for transcription and translation is mixed with metagenomic DNA constructs in a µl scale and incubated for 4 to 6 hours at moderate temperatures. Subsequently, it is possible to screen directly with suitable substrates for the desired biochemical activity and to detect positive hits by e.g. Western blot analysis. Due to the low reaction volume and the simple experimental set-up, the cell-free technology would allow the screening of an entire metagenomic library within a few hours.

The goal is the cost-effective and rapid, functional screening of a metagenome using a variable CFPS system. This consists of different RNAPs and cell extracts to recognize a higher number of promoters and synthesize more proteins from its environmental sample than the classic approach (1.1.1). Since transcription in the most commonly used CFPSs is based on commercially available phage-derived RNAPs, their applicability for the transcription of varied/mixed template DNA, as found in a metagenome, is severely limited because recognition of promoter sequences, distinctly different from their own, is hardly possible. To expand the recognition and transcription spectrum and thus capture a wide variety and diversity of different genes from a metagenome, endogenous RNAP in cell extracts have already been proposed (Shin & Noireaux, 2010; Garamella *et al.*, 2016). One approach is, that a selection of different, partly new RNAPs from different phylogenetic origin could solve the problems of recognition of promoters given in the gene library.
The advantage of *in vitro*-based metagenome screening is the overcoming of host-specific limitations (1.1.2). Low protein expression rates based on codon bias differences between the source and the expression organism should be increased by providing an oversupply of all tRNA variants, because codon bias is directly correlated with a bias of tRNA abundance within each organism. By adding all essential amino acids in high concentrations and optional chaperones, the *in vitro* translation is performed without accumulation of the protein or difficulties of the host organism with its secretion (Felczykowska *et al.*, 2015). Instead of overcoming host-based problems by replacing the host organism and using alternative strains like *Bacillus*, *Pseudomonas* or *Streptomyces* (Lorenz & Eck, 2005; Aakvik *et al.*, 2009), cell extracts from organisms of various phyla can be prepared and used for *in vitro* translations. By the use of particularly thermophilic strains, robust systems can be established, that make the heterologous expression of thermostable biocatalysts at elevated temperatures possible. This is described in detail in chapter 2.

Going "*in vitro*" also allows the direct detection of the target activity since the protein is in soluble form without any secretion or cell disruption steps. Thus, cell-free expression techniques can be combined with microtiter plate-based HTS with specific fluorophores or chromophores.

1.4 *In vitro* compartmentalization

In addition to the plate-based screening of metagenomic libraries, functional metagenomics can also be coupled with fluorescence-activated cell sorting (FACS) techniques. For example, Uchiyama *et al.* used FACS to increase the throughput of a metagnomic library screening (Uchiyama *et al.*, 2005). The so-called substrate-induced gene expression screening (SIGEX) was designed to overcome the low throughput rate of classical functional metagenomic screening. The SIGEX system consists of different steps including the construction of a metagenomic library using the operon-trap p18GFP vector and transformation of a host strain like *E. coli*, removal of false-positive clones and the selection of clones expressing GFP in the presence of a specific substrate through FACS (Figure 5). The last step is the cultivation of sorted cells on agar plates, subjected to FACS again to verify their activity and then used for other thermophilic *Bacillus* strains sis and protein expression (Uchiyama & Watanabe, 2007).

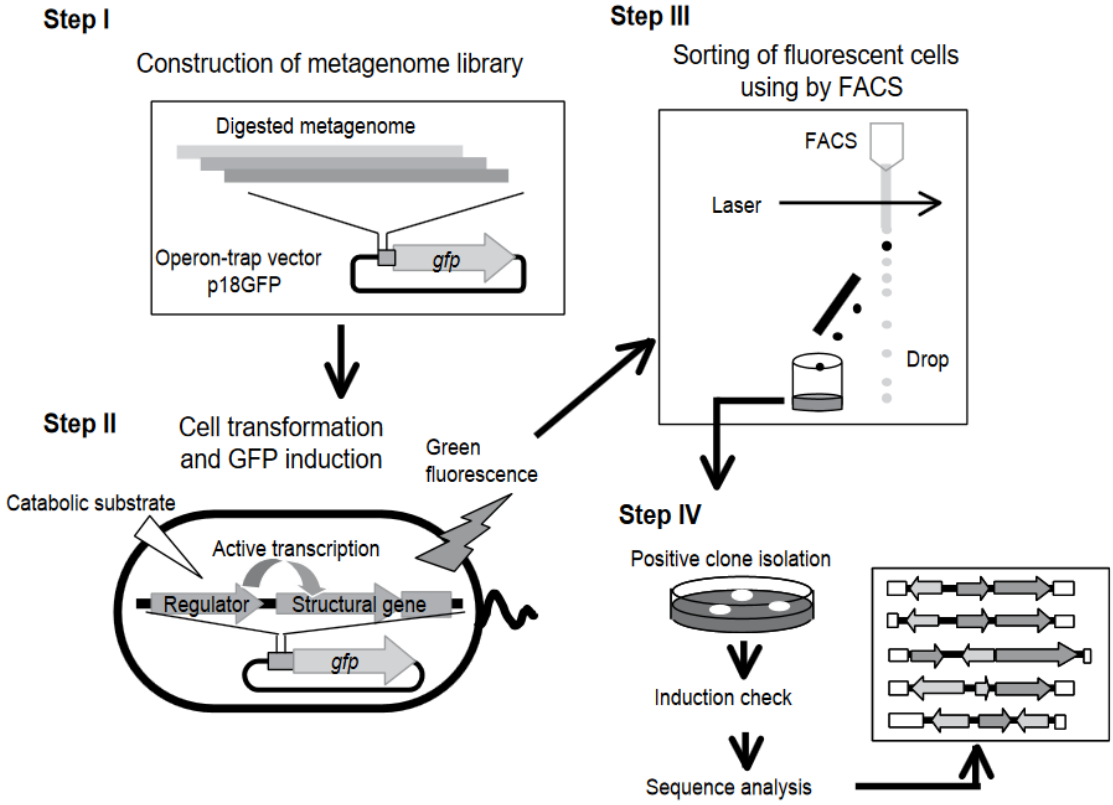

Figure 5. Scheme of the SIGEX system.

The four main steps are described in the figure. The main goal of the SIGEX is the screening of a metagenomic library for obtaining genes with catabolytic potential, whose expression is induced by a chemical compound like a response to an environmental stimulus. Figure reprinted from (Uchiyama & Watanabe, 2007).

By coupling FACS techniques to the construction of a metagenomic library, the screening efficiency can be increased, but time-consuming and labor-intensive steps such as library construction, including all

cloning steps, cannot be unified. Also, it is only possible to screen for the reporter protein like GFP after exposure to a particular stimulus.

A simpler way to couple function-based metagenome screening with FACS techniques is cell-free protein expression in micro-compartments. CFPS reactions can be compartimentalized, allowing the synthesis of proteins from uncloned metagenomic DNA and a subsequent FACS screening in a high throughput scale. The process of encapsulating CFPS reactions in cell-like oil droplets is described by Bernath *et al.* and called "*in vitro* compartimentalization" (IVC) (Bernath *et al.*, 2004). Usually oil droplets are used, whereas the CFPS reaction is forming the aqueous phase. CFPS is carried out by incubation of the *in vitro* compartments in a reaction tube and then filtering out the positive hits using FACS. For that, model membranes such as liposomes and polymersomes can be used. Liposomes are a simplified version of a cell membrane, consisting of a phospholipid bilayer without any integral components. Liposomes are highly biocompatible, but also unstable and sensitive. A promising alternative to liposomes are compartments composed of a bilayer shell of diblock amphiphilic copolymers and an aqueous core as used by the "Schwaneberg group" (Pitzler, 2015). Those polymersomes share some general physical properties with lipids, because they are both composed of a polar end, which is covalently bound to a hydrophobic tail. As a result, both liposomes and polymersomes are capable of self-assembly (Rideau *et al.*, 2018). The key advantage of polymersomes over liposomes is the high versatility and significantly higher stability. For example, polymersomes are stable for many weeks at a wide range of temperatures (Bartenstein *et al.*, 2016).

2 Aim of this work

Microorganisms living at temperatures above 55°C represent an exciting source of a large variety of biocatalysts, needed for industrial bioprocesses, that require elevated temperatures to take place (DeCastro *et al.*, 2016; Mirete *et al.*, 2016). The performance of biotechnological processes at high temperatures is characterized by numerous benefits such as a reduced risk of microbial contamination even in long term reactions, and a higher solubility of reagents (Mirete *et al.*, 2016). Robust enzymes of thermophilic-origin (so-called "thermozymes" (DeCastro *et al.*, 2016)) distinguish themselves over their mesophilic counterparts as they perform efficiently unter extreme conditions like high temperatures, high substrate concentrations, high pressure and extreme pH values (Sarmiento *et al.*, 2015). Some thermozymes are highly resistant to organic solvents and denaturing agents (Fan *et al.*, 2011; Roh & Schmid, 2013; DeCastro *et al.*, 2016). In addition, separation from heat-labil proteins is easily possible through heat incubation (Pessela *et al.*, 2004).

Metagenomics is a promising source for these biocatalysts with exciting new functions relevant for industrial applications (Ferrer *et al.*, 2005; Ferrer *et al.*, 2007). However, it is well known that the current approaches for function-based screening of metagenomic libraries have several limitations that still do not allow this technology to access the full metabolic potential of a metagenome (Beloqui *et al.*, 2008; Ferrer *et al.*, 2009; Nevondo, 2016).
As already described (1.1.2), one main obstacle in function-based screening is the heterologous protein expression in a suitable host, which is strongly dependent on the experimental set-up (Felczykowska *et al.*, 2015). For library construction, recombinant protein expression and the screening for a specific desired function, *E. coli*-based systems are still the first choice despite the typical problems of protein expression in prokaryotic host systems described in 1.1.2. Alternative host systems such as yeast or insect cells are the most commonly used solutions that can only partially overcome these hurdles (1.2.3). The CFPS, however, represents a promising alternative, which is host-independent, time-saving and less labor-intensive.

The aim of this study is the development of a cell-free screening approach for function-based metagenomics as described in 1.3. A combined *in vitro* transcription and translation system for rapid cell-free screening of metagenomic DNA should overcome many limitations associated with heterologous protein expression and conventional function-based screening methods. The focus of this cell-free system, whose initial experiments are carried out in this work, is in particular the *in vitro* transcription and translation of thermozymes for functional-based metagenomics.

***In vitro* transcription with different RNAPs**

Even though gene expression is a multi-step process involving transcription, RNAP processing, translation, post-translational modification, etc., transcription is the most limiting step in the expression of metagenomic genes in a surrogate host. Based on that, the first part of this work is to improve the transcription of metagenomic genes in particular of thermophilic origin, through the application of *in vitro* reactions with heterologously expressed RNAPs.

For effective *in vitro* transcriptions, alternative RNAPs, both bacterial (RNAP from the extreme-thermophile *T. thermophilus*) and viral (metagenome-derived RNAP) should be cloned, synthesized and purified. These should cover a broad recognition spectrum of different promoters in metagenomic DNA and allow the performance of *in vitro* transcription reactions at elevated temperatures especially for thermozymes coding genes. For this purpose, a variety of different templates are selected to investigate the application spectrum of the RNAPs for *in vitro* transcription experiments. Transcription reactions with (i) uncloned, genomic DNA from different bacteria as found in a metagenome, as well as (ii) fosmids from a metagenome library and (iii) metagenomic-derived genes coding for already characterized enzymes, should be carried out.

***In vitro* translation with different cell extracts**

As a second part of this work, the general applicability of cell-free protein expression for metagenomic-derived genes coding for thermostable enzymes should be tested. Already characterized enzymes of metagenomic origin are used as target proteins, including archaeal and bacterial lipases and PET hydrolases. These have already been identified as active, heat-stable biocatalysts by either sequence-based or function-based metagenomic screening and serve as ideal protein candidates for performing model systems for cell-free metagenomics.

For *in vitro* translations at elevated temperatures, in addition to the commonly used extract of *E. coli*, different, robust cell extracts (1.2.2.2) from various bacteria, both gram-negative and gram-positive strains should be prepared, covering a wide temperature range. These should implement the expression of proteins, that need special conditions like extreme temperatures, to be active.

Advanced function-based screening

Furthermore, a direct microtiter plate-based method for activity screening of *in vitro* expressed enzymes should be developed, which can be applied without elaborate purification steps in high-throughput scale. Finally, the system should be tested for compatibility with advanced *in vitro* compartmentalization (1.4). In cooperation with the “Schwaneberg Group” of the RWTH Aachen, cell-free reactions in polymersomes are planned, whereas the expression of the desired enzymes should be detected fluoromerically.

3 Material and Methods

3.1 Bacterial strains

Table 2. Bacterial Strains.

Strain	Characteristics[1)]	Reference/source
***E. coli* DH5α**	*supE*44 Δ*lacU*169 (Φ80 lacZΔM15) *hsdR*17 *recA*1 *endA*1 *gyrA*96 *thi*-1 *rel*A1	Invitrogen (Karlsruhe, Germany; Hanahan 1983)
***E. coli* BL21 (DE3)**	F^-, *ompT*, *hsdS* B ($r_B^-m_B^-$) *gal*, *dcm*, λDE3	Novagen/Merck (Darmstadt, Germany)
***BL21*-CodonPlus (DE3)-RIL**	F^-, *ompT*, *hsdS* B ($r_B^-m_B^-$) dcm^+ Tet^r *E. coli* galλ (DE3) *end*A Hte [*argU ileY leuW* Cam^r]	Minna Eklung (2006)
***Rosetta-gami*™ 2 (DE3)**	Δ(*ara-leu*)*7697* Δ*lacX74* Δ*phoA Pvu*II *phoR araD139 ahpC galE galK rpsL* (DE3) F′[lac^+ $lacI^q$ *pro*] *gor522*::Tn*10 trxB* pRARE2 (Cam^R, Str^R, Tet^R)	Novagen/Merck (Darmstadt, Germany)
***E. coli* MRE600**	F^-, *rna*	unknown
***Thermus thermophilus HB8* (DSM- no. 579)**	Wild-type strain	DSMZ (German Collection of Microorganisms and CellCultures, Braunschweig, Germany; Oshima and Imahori 1974)
Geobacillus* sp. GHH01	Wild-type strain	Isolated from an enrichment culture (Botanischer Garten, University of Hamburg, Germany, Dr. U. Rabausch)
***Geobacillus thermoleoverans* LEH-1, R-35636 (DSM-no. 5366)**	Wild-type strain	DSMZ (German Collection of Microorganisms and CellCultures, Braunschweig, Germany; Zarilla and Perry 1988)
***Pseudomonas antarctica* (DSM- no. 15318)**	Wild-type strain	DSMZ (German Collection of Microorganisms and CellCultures, Braunschweig, Germany; Reddy *et al.* 2004)
***Chelatococcus sambhunathii* HT4 (DSM-no. 18167)**	Wild-type strain	DSMZ (German Collection of Microorganisms and CellCultures, Braunschweig, Germany; Panday and Das 2010)
***Bacillus subtilis* TEB 1030**	*His, nprE, aprE, bpf, ispI, ΔlipA, ΔlipB* and protease knockout	(Eggert *et al.*, 2003)

1) Geno- and phenotypes according to (Bachmann, 1983). II *Sequence analysis has shown that other *Bacillus* strains cannot be excluded as contaminants. Further details are described in 5.3.1.2.

3.2 Vectors and constructs

Table 3. Vectors.

Vector	Characteristics[1)]	Size (kb)	Reference/source
pDrive	TA-cloning vector, *oriEc*, P_{lac}*lacZ*, Amp^R, Kan^R, T7-promotor	3.85	QIAGEN (Hilden, Germany)
pBluescript II SK(+)	Cloning vector, P_{lac}*lacZ*, Amp^R, T3- and T7- promotor	2.96	Stratagene /Agilent Technologies (Santa Clara, CA, USA)
pET-21a(+)	Expression vector, *lacI*, Amp^R, T7-promotor, C-terminal His_6-tag coding sequence	5.44	Novagen/Merck (Darmstadt, Germany)
pET-28a(+)	Expression vector, *lacI*, Kan^R, T7-promotor, N-terminal and C-terminal His_6-tag coding sequence	5.37	Novagen/Merck (Darmstadt, Germany)

1) Geno- and phenotypes according to (Bachmann 1983)

Table 4. Constructs.

Construct	Characteristics[1)]	Reference/source
pET-21a(+)::*rpoA*	α-subunit of *T. thermophilus* RNAP in pET-21a(+)	This work
pET-21a(+)::*rpoB*	β-subunit of *T. thermophilus* RNAP in pET-21a(+)	This work
pET-21a(+)::*rpoC*	β'-subunit of *T. thermophilus* RNAP in pET-21a(+)	This work
pET-21a(+)::*rpoD*	σ^{70}-factor of *T. thermophilus* RNAP in pET-21a(+)	This work
pET-21a(+)::*rpoZ*	ω-subunit of *T. thermophilus* RNAP in pET-21a(+)	This work
pET-28a(+)::*rpoA*	N-His_6-tagged α-subunit of *T. thermophilus* RNAP in pET-28a(+)	This work
pET-28a(+)::*rpoB*	N-His_6-tagged β-subunit of *T. thermophilus* RNAP in pET-28a(+)	This work
pET-28a(+)::*rpoC*	N-His_6-tagged β'-subunit of *T. thermophilus* RNAP in pET-28a(+)	This work
pET-28a(+)::*rpoD*	N-His_6-tagged σ^{70}-factor of *T. thermophilus* RNAP in pET-28a(+)	This work
pET-28a(+)::*rpoZ*	N-His_6-tagged ω-subunit of *T. thermophilus* RNAP in pET-28a(+)	This work
XXX::*GFP*	Not described „control vector“ (probably pIVEX) from the kit, containing GFP	Biotechrabbit RTS 100 *E. coli* HY Kit

pDrive-*vioABCDE*	pDrive encoding the *vioA-E* genes under the native promoter	(Hornung, 2013)
pET-21a(+)::*PET2*	C-His_6-tagged PET hydrolase 2 in pET-21a(+)	(Danso *et al.*, 2018)
pET-21a(+)::*PET5*	C-His_6-tagged PET hydrolase 5 in pET-21a(+)	(Danso *et al.*, 2018)
pET-21a(+)::*PET6*	C-His_6-tagged PET hydrolase 6 in pET-21a(+)	(Danso *et al.*, 2018)
pET-21a(+)::*PET12*	C-His_6-tagged PET hydrolase 12 in pET-21a(+)	(Danso *et al.*, 2018)
pET-28a(+)::*AmpR*	C-His_6-tagged beta-lactamase in pET-28a(+)	This work
pET-21a(+)::*sfGFP*	C-His_6-tagged superfolder GFP in pET-21a(+)	Kinfu
pET28a(+)::*CelA2 H288F*	C-His_6-tagged improved variant of the cellulase CelA2	(Körfer *et al.*, 2016)
pET-21a(+)::*igni13*	C-His_6-tagged lipase 13 from *Ignicoccus hospitalis* in pET-21a(+)	König; Holzscheck
pET-21a(+)::*igni14*	C-His_6-tagged lipase 14 from *Ignicoccus hospitalis* in pET-21a(+)	König; Holzscheck
pET-21a(+)::*igni15*	C-His_6-tagged lipase 15 from *Ignicoccus hospitalis* in pET-21a(+)	König; Holzscheck
pET-21a(+)::*igni16*	C-His_6-tagged lipase 16 from *Ignicoccus hospitalis* in pET-21a(+)	König; Holzscheck
pET-21a(+)::*igni18*	C-His_6-tagged lipase 18 from *Ignicoccus hospitalis* in pET-21a(+)	König; Holzscheck

3.3 Primers

rpoAHB-NdeI_F	CATATGTTGGATTCCAAGCTCAAGGCCC
rpoAHB-EcoRI_R	GAATTCTCACTCCTTCAGGGTGAAGCC
rpoBHB-NdeI_F	CATATGGAGATCAAGCGGTTCGGTCG
rpoBHB-EcoRI_R	GAATTCTCACCGCTTGGAGGCCAACCC
rpoCHB-NdeI_F	CATATGAAAAAAGAGGTTCGTAAGGTTCGC
rpoCHB-EcoRI_R	GAATTCTTAAGCCTGCTTGCCGGGCTGCTCC
rpoCHB-NdeI_F2	CATATGGCCTCCCCGGAGAAGATCC
rpoZHB-NdeI_F	CATATGGCGGAACCGGGCATTGACAAGC
rpoZHB-EcoRI_R	GAATTCCTACTCCTCCCGCTCCACCGGGTAGA
rpoDHB-NdeI_F	CATATGAAGAAGAGCAAGCGCAAGAACGCC
rpoDHB-EcoRI_R	GAATTCTTAGTCCAGGAAGTCCCTGAGC
B_lactamase_for-N	CTCGCATATGAGTATTCAACATTTCCG
B_llactamase_rev-N	CGAGTCGACCCAATGCTTAATCAGTGAG
Vio_HindIII_rev-New	AGCAAGCTTCGGCCATTTAGGTGACACTATAGAATAC

3.4 Culture media and supplements

All media were autoclaved at 121°C and 2 bar for 20 minutes. Antibiotics were sterile filtrated and added to the media after cooling down.

LB medium (Sambrook 2001)

Tryptone	10 g
Sodium chloride	5 g
Yeast extract	5 g
(Agar	12g)
H_2O_{bidest}	*ad* 1000 ml

Standard I medium (DSMZ)

Peptone from meat	7.8 g
Peptone from casein	7.8 g
Sodium chloride	5.6 g
Yeast extract	2.8 g
D(+)-Glucose	1.0 g
H_2O_{bidest}	*ad* 1000

pH 7.5 ± 0.2 at 25°C

Medium I (DSMZ)

Peptone	5.0 g
Meat exract	3.0 g
H_2O_{bidest}	*ad* 1000 ml
	pH 7.0

***Thermus* medium**

Solution A:

Nitrilotriacetic acid	1 g	$CaSO \cdot 2\ H_2O$	0.6 g
$MgSO_4 \cdot 7\ H_2O$	1 g	NaCl	80 mg
KNO_3	1 g	$NaNO_3$	6.9 g
Na_2HPO_4	1.1 g	H_2O_{bidest}	*ad* 1000 ml

Solution B:

$FeCl_3$	28 mg	H_2O_{bidest}	*ad* 1000 ml

Solution C:

H_2SO_4 (conc.)	0.5 ml	$MnSO_4 \cdot H_2O$	220 mg
$ZnSO_4 \cdot 7\ H_2O$	50 mg	H_3BO_3	50 mg
$CuSO_4$	1.6 mg	$Na_2MoO_2 \cdot H_2O$	2.5 mg
H_2O_{bidest}	*ad* 1000 ml		

Medium D:

Tryptone	1 g	Yeast extract	1 g

Solution A	100 ml	Solution B	10 ml
Solution C	10 ml	H_2O_{bidest}	*ad* 1000 ml
			pH 8.2

Table 5. Antibiotics and other supplements.

Antibiotic/ supplement	Stock solution	Working concentration
Ampicillin	100 mg/ml	100 µg/ml
Chloramphenicol	50 mg/ml	50 µg/ml
Kanamycin	25 mg/ml	25 µg/ml
Tetracyclin	7 mg/ml	7 µg/ml
IPTG	100 mg/ml	100 µg/ml
X-Gal	50 mg/ml	50 µg/ml

3.5 Culture conditions

3.5.1 Cultivation of bacteria

Liquid cultures were inoculated either with material from a single colony from an agar plate, from a glycerol stock or with max. 2% of a liquid pre-culture. Bacterial strains were grown on agar plates, test tubes (5 ml) and Erlenmeyer flasks (25 ml to 1 l) at 130 to 200 rpm on a shaker (Infors HT minitron, Infors AG, Switzerland). Supplements and antibiotics are summarized in Table 5. They were sterile filtered with a pore size of 0.22 µm.

***E. coli* culture conditions**

E. coli strains were grown on LB medium appropriately supplemented with antibiotics at 37°C overnight.

***T. thermophilus* culture conditions**

T. thermophilus was cultured in bottles stirring on a heating block at 75°C in *Thermus* medium overnight.

***Geobacillus/Bacillus* culture conditions**

Geobacillus strains were grown on Standard I medium at 50-55°C overnight.

***Pseudomonas antarctica* (*P. antarctica*) culture conditions**

P. antartica was grown on LB medium at 22°C overnight.

3.5.2 Strain maintenance

Bacterial colonies can be stored on agar plates up to 4 weeks. For long-term storage, glycerol stocks (33% (v/v) final glycerol concentration) were applied and stored at -70°C.

3.5.3 Measurement of optical density (OD)

The optical density of a bacterial culture was measured in a 1 cm thick cuvette (Sarstedt, Nümbrecht, Germany) at a wave length of 600 nm (OD_{600}) using a photometer (SmartSpec™ Plus Spectrophotometer, BIO RAD, Hercules, CA, USA). Sterile medium was used for Blank and for dilution of the culture if the OD_{600} exceeded 0.8 to avoid measurement errors.

3.5.4 Cell harvesting

Liquid cell cultures were harvested by centrifugation. For extraction of DNA, small volumes were transferred to E-cups and sedimented in a tabletop centrifuge (5415D, Eppendorf, Hamburg, Germany) at room temperature at 13,000 rpm for 2 min. Larger volumes were harvested either using a falcon centrifuge (5804R, rotor A-4-44, Eppendorf, Hamburg, Germany / Varifuge 3.0R, Heraeus, Hanau, Germany) or a Sorvall RC6+ centrifuge (150-350 ml, rotor F14S-6x250y or F10S-6x500y; Thermo scientific, Braunschweig, Germany) at 4,000-6,000 rpm and 4°C for 20 to 30 min.

3.6 Preparation of cell extracts for *in vitro* translation

The protocol for extract preparation is a modified version from (Kwon & Jewett, 2015).

All buffers for working with RNA and *in vitro* protein expression were prepared with DEPC-treated water. It is prepared by incubating with 0.1% diethylpyrocarbonate (DEPC) overnight and then autoclaved to remove the DEPC.

To produce cell extracts, a growth curve was created for each bacterial strain to determine the exponential phase. All bacterial strains were grown to mid exponential phase and harvested.

Table 6. Harvesting points for cell extract preparation.

Bacterial strain	OD_{600} for harvesting
E. coli BL21 CodonPlus	~ 2.0
E. coli MRE600	~ 2.8
Geobacillus sp. GHH01	~ 0.9
Geobacillus thermoleoverans	~ 0.9
Thermus thermophilus HB27	~ 1.0
Pseudomonas antarctica	~ 1.0
Bacillus sp.	~ 1.0

E. coli strains, *T. thermophilus* and *P. antartica* were centrifugated for 15 min, *Geobacillus* strains were centrifugated for 30 min (10,000 rpm, 4°C). The pellets were resuspended in ice-cold Buffer A supplied with 2 mM DTT just before use and transferred to a preweighted falcon tube.

Buffer A:

Tris/Acetat, pH 8.2	10 mM
MgOAc	14 mM
KGlu	60 mM
DTT	2 mM

A further centrifugation step and 2 washing steps with ice-cold Buffer A followed. The pellet was then weighed, flash frozen in liquid nitrogen and stored at -70°C overnight. The pellet was thawed on ice the next day and resuspended in 1 ml buffer A per 1 g pellet. Cell disruption was performed by French press (3 cycles, 1000 psi). The lysate was centrifuged twice at 16,100 rpm, 4°C for 30 min to remove cell debris and other insoluble components. The total amount of protein in cell extract was quantified by Bradford assay (3.8.5) after diluting the sample in the same buffer. The cell extract was aliquoted (50-200 µl) in RNase-free PCR tubes, immediately flash frozen in liquid nitrogen stored at -70°C until use.

3.7 Standard techniques for working with DNA

3.7.1 Isolation of DNA

3.7.1.1 Isolation of genomic DNA

In vitro transcription assays with RNAP from *T. thermophilus* were carried out with genomic DNA (gNDA) as a template. gDNA from *T. thermophilus* was isolated with the peqGOLD Bacterial DNA Kit (PEQLAB Biotechnologie GmbH, Erlangen, Germany). Up to 100 ml of an overnight culture were used for isolation and purification of the gDNA. According to the manufacturer´s protocol, the cell wall was degraded with lysozyme (100 µg) for 10 min at 30°C. After digestion with proteinase K (400 µg) and RNase A (300 µg) treatment for 30 min at 70°C, the lysate was loaded onto a DNA column containing a silica membrane. After two washing steps, the gDNA was eluted with up to 50 µl H_2O_{bidest}. The quality and quantity of the gDNA were analyzed by agarose gel electrophoresis subsequently.

3.7.1.2 Plasmid isolation with a plasmid mini kit

Plasmid DNA was isolated using “High-Speed Plasmid mini kit” (Avegene life science, Taipei, Taiwan, China) according to manufacturer´s instruction using 5-8 ml of an overnight culture. The plasmid DNA was eluted with 30 to 50 µl of preheated H_2O_{bidest}. If used as template for *in vitro* protein expression, DNA was eluated with DEPC-treated water.

3.7.2 Purification and concentration of DNA

The “Gel/PCR DNA Fragments Extraction kit” (Avegene life science, Taipei, Taiwan, China) was used to purify DNA after performing a PCR reaction or cutting out of an agarose gel after running an electrophoresis. The purification was carried out according to each manual. 20 to 40 µl of H_2O_{bidest} were used for elution. Bigger volumes of low concentrated DNA in aqueous solutions were concentrated using a vacuum concentrator (Concentrator 5301, Eppendorf, Hamburg, Germany) at 45°C for up to 30 min.

3.7.2.1 Analysis of DNA concentration and purity

The total DNA yield was quantificated using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, Germany). DNA absorbance was measured at 260 nm and protein absorbance at 280 nm against H_2O_{bidest}.

3.7.3 Agarose gel electrophoresis for DNA

Genomic DNA, plasmid DNA and PCR products were analyzed by agarose gel electrophoresis. A 0.8% (w/v) agarose gel was used for all DNA samples. Before loading on the gel, the samples (1-10 µl) were mixed with 1-2 µl loading dye. The electrophoresis was performed at 120 V for at least 20 min with a power supply PowerPacTM Basic (BioRad, Munich, Germany) in an electrophoresis gel chamber (HE-33 mini horizontal submarine unit, Hoefer™, Holliston, MA, USA) filled with 1x TAE buffer. The DNA was stained in an ethidium bromide solution (10 µg/ml) for 5 to 15 min and detected under UV light of 254 nm in a Molecular Imager® (GelDocTM XR+ Imaging System, BioRad, Munich, Germany). The documentary of the gel was done with a Quantity One 1-D analysis software (version 4.6.9, BioRad, Munich, Germany). The GeneRuler™ 1 kb DNA Ladder (Fermentas, St. Leon-Rot, Germany) was used to estimate the fragment sizes.

TAE buffer (50x):

Tris-Ac	2 M
EDTA	100 mM
Dissolved in H_2O_{bidest}	
The pH was adjusted to 8.0 with acetic acid.	

Loading dye:

Glycerol (30%)	60 ml
EDTA	50 mM
Bromphenol blue (0.25%)	0.5 g
Xylencyanol (0.25%)	0.5 g
H_2O_{bidest} *ad* 200 ml	

3.7.3.1 Gel extraction of DNA

After PCR or digestion, electrophoresis was performed and the DNA fragments were extracted from the agarose gel. The “Gel/PCR DNA Fragments Extraction kit” (Avegene life science, Taipei, Taiwan, China) was used according to the manufacturer´s instructions. The DNA was eluted in 30 µl H_2O.

3.7.4 PCR

PCR was performed to amplify the *rpo* genes as well as the ß-lactamase gene. Primers are given in 3.3. The standard PCR mixture consists of:

DNA template	1-2 µl
Buffer (10x)	5 µl
dNTPs (2 mM)	2 µl
Forward primer (10 µM)	1 µl
Reverse Primer (10 µM)	1 µl

DMSO	2 µl
DNA polymerase (2.5 U/µl)	0.25-0.5 µl
H_2O_{bidest}	*ad* 50 µl

The PCR reaction was carried out in a thermocycler (Mastercycler® personal, Eppendorf, Hamburg, Germany) using conditions as indicated in Table 7. An elongation time of 1 kb per min was estimated for the Taq DNA polymerase and 3 kb per min for the Phusion DNA polymerase.

Table 7. Standard PCR reaction:

PCR Step		Temperature	Time
Initial denaturation		95°C	2 min
Denaturation		95°C	30 sec
Annealing	35x	$T_{ann} = T_m$- 5°C	30 sec
Elongation		72°C	15 sec- 1min/kb
Final elongation		72°C	5 min

3.7.4.1 Colony PCR

Colony PCR was used to check clones for the ligation and the correct insert size. A small amount of cell material was used directly as template and therefore was resuspended in the reaction mix.

Taq polymerase buffer (10x)	2.5 µl
dNTPs (2 mM)	1 µl
Forward primer (10 µM)	1 µl
Reverse primer (10 µM)	1 µl
Taq DNA polymerase	0.25 µl
$H2O_{bidest}$	ad 25 µl

The PCR reaction conditions were as indicated in Table 7.

3.7.5 Cloning of the *T. thermophilus* RNAP subunits

The *rpo* genes from *T. thermophilus* were amplified using the DreamTaq DNA Polymerase (Thermo scientific, Braunschweig, Germany) with genomic DNA from the strain HB27 as template. The standard PCR protocol (3.7.4) was performed.

The amplified *rpo* genes from *T. thermophilus* were cloned via 3´ A overhangs added by the DreamTaq DNA Polymerase (Thermo scientific, Braunschweig, Germany) into the pDrive vector using the QIAGEN PCR Cloning Kit (Hilden, Germany). For the reaction, the 2x Ligation Master Mix was mixed with the PCR fragment and the linealized pDrive vector which carries U overhangs at each 3' end.

2x Ligation Master Mix	2.5 µl
PCR fragment (insert)	2 µl (200 ng/µl)
pDrive cloning vector	0.5 µl (50 ng/µl)

The ligation was incubated for 4h at 16°C and subsequently used for transformation of competent *E. coli* DH5α cells. Recombinants were selected by blue-white screening (3.7.9) on LB agar plates and analyzed using colony PCR (3.7.4.1) using both gene-specific and vector-specific primers or DNA restriction (analytical scale).

DNA restriction was also used to prepare the pET-21a and pET-28a vector as well as the *rpo* genes for cloning (preparative scale). For that, restriction sites for EcoRI and NdeI were inserted into the *rpo* gene primers used for PCR amplification. The *rpo* genes could then be cut out of the pDrive vector.

Vector DNA	10-20 µl (1-2 µg)
EcoRI (10 U/µl)	2 µl
NdeI (10 U/µl)	2 µl
Buffer (10x)	5 µl
H_2O_{bidest}	*ad* 50 µl

The reaction was incubated for 2 h to overnight at 37°C (optimal temperature for the enzymes used) and stopped afterwards by heat inactivation of the restriction enzymes at 65°C to 80°C for 20 min. The analytical restriction was then separated by gel electrophoresis (3.7.3) and the band pattern was visualized.

The vector was cut in its multiple cloning site and then dephosphorylated on its 5´-end to ensure that the vector does not re-circularize during the ligation process.

Preparative digestion of the vector	17 µl
Antarctic phosphatase buffer (10x)	2 µl
Antarctic phosphatase (5 U/µl)	1 µl

The reaction was catalyzed by the Antarctic phosphatase (New England Biolabs, Frankfurt am Main, Germany) for 20 min at 37°C and then stopped by heat inactivation (60°C, 5 min) of the enzyme. All

restriction endonucleases and their appropriate buffers were purchased from Fermentas (St. Leon-Rot, Germany).

After DNA restriction, genes were ligated into the pET-vectors with the ATP-dependent T4 DNA Ligase (Fermentas, St. Leon-Rot, Germany). Usually, three to five times the amount of DNA fragment is used for a simple part of vector DNA. To obtain the ideal ratio of DNA fragment and vector, samples were analyzed by gel electrophoresis (3.7.3) to be able to estimate the quantities to be used on the basis of size and band intensity.

The standard ligation mixture had a total volume of 10 µl.

		Final concentration
Insert DNA (digested)	4-6 µl	70-140 ng
pET-21a/pET-28a (digested)	1-2 µl	20-40 ng
T4 DNA ligase (400 U/µl)	1 µl	400 U
T4 ligase buffer (10x)	1 µl	1x
ATP (10 mM)	1 µl	1 mM

The reaction was incubated for 2 h at room temperature or overnight at 16°C. The ligation was checked by agarose gel electrophoresis with non-ligated control and then used directly for transformation of a host.

3.7.6 Cloning of the $RNAP_E$

The DNA sequences of the new RNAP candidates (as the $RNAP_E$) were transmitted to Eurofins Genomics (Eurofins MWG Operon, Ebersberg, Germany) for gene synthesis. The sequences were codon optimized for expression in *E. coli*. Additional restriction sides at 5' (NheI) and 3' (SalI) were added for cloning. The genes were inserted into a vector with ampicillin resistance (pEX-A258 or pEX-A128) by Eurofins.

3.7.7 Cloning of templates for *in vitro* protein expression

3.7.7.1 ß-Lactamase

The ß-lactamase was amplified via PCR using the pET-21a vector as template. The “Gel/PCR DNA Fragments Extraction kit” (Avegene life science, Taipei, Taiwan, China) was used according to the manufacturer´s instructions to purify the PCR product. After digestion with NdeI and SalI, whose restriction sites were inserted by the primers, the reaction was stopped by heat inactivation (60°C, 5

min) and directly used for ligation into the pET-28a vector, which was digested and dephosphorylated. Transformants of BL21 were selected directly via LB agar with IPTG, ampicillin and kanamycin.

3.7.7.2 Violacein operon

The violacein operon was cloned by Dr. Hornung into pDrive (see (Hornung, 2013) for details). For *in vitro* transcription experiments, the construct was cut with the restriction enzyme HindIII behind the vioE gene and subsequently isolated (3.7.2). In parallel, primers were created to enable cloning of the operon into other vectors.

3.7.8 Heat shock transformation of *E. coli*

The transformation was carried out by heat shock method. For this, one aliquot (100 µl) of cells were thawed on ice for 5 min. Afterwards 5 µl of the ligation micture was added to the cells and stirred gently. The cells were incubated for 20 min on ice before heat shock was carried out at 42°C for 60-90 sec. After cooling down on ice for 5 min, 500 µl of liquid LB-medium were added before incubating at 37°C for 30 to 45 min. Finally, the cells were plated out on LB-agar plates containing selective antibiotics and then incubated overnight at 37°C.

3.7.9 Blue-white screening

To screen for bacterial clones with recombinant pDrive with *rpo*-genes, the Blue-white screening was used. After transformation, the cells were plated out on agar plates containing appropriate antibiotics as well as the inductor IPTG and the substrate X-gal. White colonies indicate the presence of an insert in the MCS of the plasmid, whereas blue colored colonies indicate the presence of an empty vector. White colonies were picked, checked for the right insert size (3.7.4.1) and cultivated overnight for plasmid preparation (3.7.1.2) and further cloning steps.

3.7.10 Sequencing

Sequencing was performed by Eurofins (Elsberg, Germany) based on the Sanger method. Samples were purified (3.7.1.2), adjusted to a DNA concentration of 100 ng/µL (plasmids) in H_2O_{bidest} and mixed with 1 µl of primer (4.8 µM).

3.8 Standard techniques for working with proteins

3.8.1 Heterologous expression in *E. coli*

3.8.1.1 Induction and expression

Protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at an approximate OD_{600} of 0.6. Detailed expression conditions are listed in Table 8.

Table 8. Expression conditions for the *T. thermophilus* subunits.

Subunit	Size (kDa)	*E. coli* host strain	Temperature	Time
α	**40**	BL21 (DE3)	37°C	3h
β	**140**	BL21 (DE3)/ Rosetta-gami™ 2 (DE3)	37°C	5-6h
β´	**170**	BL21 (DE3)/ Rosetta-gami™ 2 (DE3)	37°C 22°C	3h overnight
ω	**10**	BL21 (DE3)	37°C	3h
σ-70	**90**	BL21 (DE3)	37°C	3h

Test expressions in a volume of 20-100 ml were carried out to find optimal conditions for the production of each RNAP subunit. Samples were taken at different time points of the expression. To determine the expression level of the recombinant protein, samples were centrifugated (13,200 rpm, 2 min, RT), boiled for 10 min at 95°C in 1x SDS loading dye and analyzed by SDS-PAGE (3.8.6).

Cells were disrupted by 1 min bursts of sonication (duty cycle: 50%, output: 0.5) 4 to 5 times. During sonication, the falcon tube with the cell suspension was kept on ice. Between the sonications steps, the suspension was cooled down for 1 min. The lysate was next centrifuged at 13,200 rpm for 30 min at 4°C and the supernatant was stored at 4°C until protein purification.

3.8.2 Purification

3.8.2.1 Heat denaturation

Soluble subunits from the RNAP from *T. thermophilus* were crudely purified via heat denaturation of the majority of the *E. coli* proteins. For that the cell extracts were incubated at 60-80°C for up to 1h depending on the subunit. After incubation on ice for 20 min, precipitated proteins were removed by centrifugation (13,200 rpm, 20 min, 4°C).

3.8.2.2 Inclusion bodies

The ω-subunit of the *T. thermophilus* RNAP was found as inclusion bodies. Therefore the pellet was thawed, washed in 5 ml of lysis wash buffer, then centrifuged at 5,000 xg and resuspended in 5 ml of

inclusion body buffer. The suspension was incubated for 30 min at room temperature and sonified meanwhile (duty cycle: 50%, output: 0.5). After centrifugation (20 min, 18,000 xg, 4°C) the washing step with inclusion body buffer was repeated two times, whereas the buffer without sodium deoxycholate was used in the last step. The pellet was then stored at −20°C until further use.

To solubilize the protein, the pellet was resuspended on ice in 500 µl of denaturation buffer containing 6M Urea for at least 30 min and vortexing regularly. After centrifugation (1h, 18,000 xg, 4°C) the soluble fraction was used for *in vitro* reconstitution (3.8.4).

Inclusion body buffer:

Tris-HCl, pH 8.0	20 mM
NaCl	0.2 M
Sodium deoxycholate	0.5%
EGTA	2 mM

Denaturation buffer:

Tris-HCl, pH 7.9	50 mM
$MgCl_2$	10 mM
$ZnCl_2$	10 µM
Glycerol	10% (v/v)
EDTA	1 mM
DTT	10 mM
Urea	6 M

3.8.2.3 FPLC purification of His_6-tagged proteins

For the purification of the $RNAP_E$, the cell pellet from 1 l of *E. coli* BL21 expression culture was first resuspended in 5 ml lysis buffer containing 0.2 mM phenylmethylsulfonyl fluoride (PMSF) and 1 mM DTT. The cells were lysed using French press (3 cycles, 1,300 psi) and then incubated with 200 units DNase I for 30 min at 37°C supplemented with $MgCl_2$ to a final concentration of 2 mM. Afterwards the cell suspension was centrifuged at 4,500 xg for 10 min at 4°C. The clear cell lysate was used for fast protein liquid chromatography (FPLC) based purification.

For purification of His_6-labeled proteins, a special form of affinity chromatography, the immobilized metal chelate affinity chromatography (IMAC) was used. Nitrioltriacetic acid (NTA) bound to sepharose beads were used as the column material (HisTrap FF, 5 ml, Amersham Biosciences, GE Healthcare, Freiburg, Germany). IMAC is based on the specific covalent bond of amino acids, especially histidine, to metals, in this case Ni^{2+}.

Lysis buffer:		**Wash buffer:**		**Elution buffer:**	
NaH_2PO_4, pH 8.0	50 mM	NaH_2PO_4, pH 8.0	50 mM	NaH_2PO_4, pH 8.0	50 mM
NaCl	300 mM	NaCl	500 mM	NaCl	300 mM
Imidazol	20 mM	Imidazol	40 mM	Imidazol	250 mM
		Glycerin	1% (v/v)		

All buffers for FPLC based purification were filtered, degassed and cooled before use.

Ni^{2+}-chelating NTA-column was first preequilibrated with at least 5 column volumes of lysis buffer. Subsequently, the cell lysate was injected onto the column and the flow through was collected. The column was then washed using at least 5 column volumes of washing buffer and a sample of the washing fraction was taken. Bound protein was finally eluted using elution buffer and fractions of 1 ml were collected. The protein purification was checked using SDS-PAGE (3.8.6). By comparison with the chromatogram those fractions were pooled which showed the elution peak of the $RNAP_E$.

3.8.3 Buffer exchange

Finally, the $RNAP_E$ was re-buffered in 2x T7 storage buffer (VIVASPIN® 6 concentrator, MWCO 30,000) and glycerol was added (final 50% glycerol and 1x T7 storage buffer).

2x T7 storage buffer:

Tris-HCl, pH 7.9	100 mM
NaCl	300 mM
Triton X-100	0.2%
2-ME	40 mM

3.8.4 *In vitro* reconstitution of the RNAP of *T. thermophilus*

The reconstitution of the RNAP core enzyme ($RNAP_C$) from *T. thermophilus* was accomplished by mixing all subunits ($\alpha_2\beta\beta'\omega$) under denaturing conditions and slowly removing the denaturant through gradient dialysis. For this, the protocol from (Kuznedelov & Severinov, 2009) was slightly modified. All five subunits were mixed in denaturation buffer in a defined molar ratio of 2:8:4:2 (α:β:β':ω) for soluble and briefly purified subunits. The total protein concentration of the reconstitution mixture was adjusted to 0.5 mg/ml by adding denaturation buffer and was then dialyzed (Servapor(R) dialysis tube, MWCO 12-14,000) stepwise at different temperatures against up to 250 volumes of reconstitution buffer. The urea concentration in the buffer is reduced from 6 M to 3 M to 0 M with every buffer exchange. Because of the advantage of the heat-stability of *Thermus* proteins, the reconstitution was

carried out not only at 4°C but also at room temperature and 60°C. The RNAP holoenzyme ($RNAP_H$) was assembled by addition of purified σ^{70}-factor in an equimolar concentration and incubating at 37°C for 1h. The reconstitution efficiency was visualized using a seminative PAGE (3.8.7). The RNAP was directly tested for transcription activity (3.9.1) without long-term storage.

Reconstitution buffer:

Tris-HCl, pH 7.9	50 mM
KCl	200 mM
$MgCl_2$	10 mM
$ZnCl_2$	10 µM
Glycerol	10% (v/v)
EDTA	1 mM
DTT	1 mM

3.8.5 Protein quantification (Bradford, 1976)

Soluble proteins or protein mixtures were quantified according to the Bradford assay (Bradford, 1976). For that, the Roti®-Quant solution (Carl Roth GmbH, Karlsruhe, Germany) was applied according to the manufacturer´s protocol. After incubation of 10 min in the dark, protein concentration was measured at 595 nm against a blank with buffer. Bovine serum albumin (BSA) was used as a reference protein for standardization.

3.8.6 SDS - polyacrylamide gel electrophoresis (SDS-PAGE; (Laemmli, 1970))

Sodiumdodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out on discontinuous polyacrylamide gels according to standard techniques. Mostly, a 7% stacking gel (pH 6.8) and a 12% separating gel (pH 8.8) were prepared. The gels were poured and the electrophoresis was performed in the MiniProtean equipment (BioRad, Munich, Germany).

Stacking gel:

Acrylamide (40%)	X% (v/v)
Tris/HCl, pH 6.8	120 mM
APS	0.5% (w/v)
TEMED	0.1%

Separation gel:

Acrylamide (40%)	X%	(v/v)
Tris/HCl, pH 8.8	375	mM
APS	0.5%	(w/v)
TEMED	0.1%	

4x SDS loading buffer:

Glycerol	50% (w/v)
SDS	4% (w/v)
Tris-HCl, pH 6.8	150 mM
NaCl	30 mM
Dithiothreitol (DTT)	100 mM
Bromphenol blue	0.02% (w/v)
EDTA	1 mM
H_2O_{bidest}	*ad* 10 ml

10x SDS running buffer:

Tris	250 mM
SDS	1% (w/v)
Glycine	1.92 M

The SDS loading dye was stored as aliquots of 1 ml at -20°C.

Before application to the gel, the samples were mixed with SDS loading buffer (final concentration 1x) and denatured at 95°C for at least 5 minutes. To estimate the protein sizes of the sample, the PageRuler™ Unstained Protein Ladder (19-200 kDa, Thermo Scientific) was used. For subsequent Western Blot analysis (3.8.10) the PageRuler™ Prestained Protein Ladder (10-180 kDa, Thermo Scientific) was used. The protein samples run in 1x running buffer at 20 mA and 30 mA (PowerPac™ Basic, BioRad, Munich, Germany) until the dye line reached the end of the gel.

3.8.7 Semi-native gels

Stacking gel (4%):

Acrylamide (40%)	200 µl
Tris/HCl, 1 M, pH 6.8	250 µl
APS	24 µl
TEMED	4 µl
H_2O_{bidest}	*ad* 2 ml

Separation gel (10%):

Acrylamide (40%)	1.25 ml
Tris/HCl, 1.5 M, pH 8.8	1,25 ml
APS	25 µl
TEMED	5 µl
H_2O_{bidest}	*ad* 5 ml

4x native loading buffer:

Tris-HCl, pH 6.8	150 mM
Glycerol	50% (w/v)
Bromphenol blue	0.02% (w/v)

10x native running buffer:

Tris-HCl	250 mM
Glycine	1.92 M

To check the reconstitution efficiency of the *T. thermophilus* RNAP, a native, as well as a seminative PAGE was performed. For that a non-reducing and non-denaturing gel system was combined with native loading buffer as well as SDS loading buffer. To check the running behavior through the gel and to estimate the protein sizes of the sample, bovine serum albumin (BSA) was used.

3.8.8 Coomassie staining of proteins and estimation of molecular weight

For staining, SDS gels were soaked in Coomassive Coomassie Brilliant Blue R250 solution (Gerbu Biotechnik GmbH, Gaiberg, Germany) for at least 1h.

Staining solution (1 l):		**Destaining solution (1 l):**	
Coomassie Brilliant Blue R-250 (10% solution)	4 ml	Ethanol	0.4 l
Ethanol	0.4 l	Acetic acid	0.1 l
Acetic acid	0.1 l	H_2O_{bidest}	0.5 l
H_2O_{bidest}	0.5 l		

Afterwards, the gels were destained using destaining solution until protein bands became visible and clear.

3.8.9 Transfer of proteins on nitrocellulose membranes (Western Blot)

Tank blots as well as semi-dry transfer electrophoresis were performed using a nitrocellulose membrane. Therefore, the SDS-gel was pre-incubated in transfer buffer for 5 min.

3.8.9.1 Tank blot

The tank blot is a more complex and time-consuming process compared to the semi-dry blot (3.8.9.2), but promises a quantitatively better transfer (Mahmood & Yang, 2012). The blot sandwiches were prepared according to standard protocols (e.g. www.bio-rad.com/ProteinBlottingGuide) using the Biorad Tank Blotting System. The tank was completely filled with transfer buffer and mixed with cooling elements to lower the temperature during the transfer. The proteins as well as the PageRuler™ Prestained Protein Ladder (10-180 kDa, Thermo Scientific) were transferred to the membrane with a power supply (PowerPacTM Basic, BioRad, Munich, Germany) with 350 mA for 90 min.

3.8.9.2 Semi-dry blot

Before the transfer, all layers were equilibrated in transfer buffer for at least 10 min. Semi-dry blots were performed in a blotting cassette (Trans-Blot SD cell, BioRad, Munich, Germany) with a constant voltage of 10 V and a max. of 5.5 mA/cm^2 for 30 min.

The quality of the transfer could be estimated from the intensity of the marker bands on the membrane.

Transfer buffer:

Tris/HCl, pH 8.6	125 mM
Glycine	192 mM
Methanol (100%)	20% (v/v)
H_2O_{bidest}	*ad* 1 l

3.8.10 Immunodetection of six-fold histidines

After the transfer, the nitrocellulose membrane was washed twice for 5 min with TBS buffer. Then the blot was incubated in 5% milk powder solution made with TBST buffer for at least 1 h to block unspecific binding-sites. Subsequently, the primary His_6-antibody (polyclonal from rabbit, MoBiTec, Göttingen, Germany), dissolved 1:5000 in TBST buffer with 5% milk powder was added to the membrane for 1 h to overnight. After rinsing 3 times with TBST buffer the blot was incubated with the secondary anti-rabbit IgG antibody conjugated with an alkaline phosphatase (dissolved 1:10000 in 5% milk powder TBST solution; derived from goat, Sigma, Munich, Germany) for 1-2h at RT. After a final washing step, the membrane was equilibrated with detection buffer for 3 min. Detection was carried out in the dark with 10 ml of BCIP/NBT-staining solution until signals became clearly visible. *In vivo* expressed His_6-tagged and purified protein with a defined concentration was used as blot control.

TBS(T) buffer (pH 7.5):

Tris	100 mM
NaCl	0.9% (w/v)
(Tween-20	0.1% (v/v))

BCIP/NBT-staining solution:

NBT-solution (75 mg/ml NBT in 70% DMF)	66 µ
BCIP-solution (50 mg/ml BCIP in 100% DMF)	33 µl
Detection buffer	*ad* 10 ml

Detection buffer:

Tris/HCl, pH 9.5	100 mM
NaCl	100 mM

3.9 Techniques for working with RNA

3.9.1 Transcriptional activity of RNAPs

The transcriptional activity of the RNAP from *T. thermophilus* was tested via *in vitro* transcription with genomic DNA and analysis in an RNA gel without subsequent translation. For *in vitro* transcription with bacterial RNAP the following buffer system was used.

10x Transcription (TB) buffer:

Tris-HCl, pH 8.0	400 mM
$MgCl_2$	200 mM
Spermidine	20 mM

5x Transcription (TB2) buffer:

Tris-HCl pH 7.8	200 mM
Spermidine	10 mM
DTT	50 mM
$MgCl_2$	30 mM
NaCl	50 mM

The *in vitro* transcription mixtures had a total volume of 10-20 µl and were incubated at 37°C to 60°C up to 2h.

***In vitro* transcription with $RNAP_{T. thermophilus}$**

10x TB buffer	2 µl
DTT (1 M)	0.1 µl
rNTPs (each 25 mM)	3.2 µl
PPase (2 U/µl)	1 µl
Ribolock (40 U/µl)	1 µl
DNA	400-1000 ng
RNAP	*ad* 20 µl

***In vitro* transcription with $RNAP_E$**

5x TB2 buffer	2 µl
rNTPs (each 25 mM)	2 µl
PPase (2 U/µl)	0.5 µl
Ribolock (40 U/µl)	0.5 µl
DNA	200-250 ng
RNAP	0.5 µl
H_2O_{DEPC}	*ad* 10 µl

The transcription efficiency was visualized by a formaldehyde agarose gel (3.9.5).

3.9.2 Cleaning & concentration of RNA

For quantification of RNA, the samples were treated with RNase-free DNase I (NEB) and subsequently cleaned & concentrated with the „RNA Clean & Concentrator™“ Kit from Zymo Research according to the manufacturer´s instructions.

3.9.3 DNase I treatment of RNA samples

RNA samples were RNase free DNase I (NEB) treated in order to remove contaminating genomic DNA or the template DNA from *in vitro* transcription assays. 200 µl samples were incubated with 27 U of DNase I for 15min at 37°C.

3.9.4 Measurement of RNA concentration

The RNA concentration was measured using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, Germany). Nucleic acid absorbance was measured at 260 nm and protein absorbance at 280 nm against DEPC-treated H_2O_{bidest}. The extinction coefficient of RNA and DNA are slightly different which affects the ratio of OD_{260} to OD_{280}. In theory, for absolutely pure RNA, the ratio is 2.0 and for absolutely pure DNA, it is 1.8.

3.9.5 Agarose gel electrophoresis for RNA

The size and quantity of RNA were analyzed by electrophoresis in denaturing formaldehyde agarose gels. In the presence of formaldehyde all secondary structures of the RNA are removed via heating. All materials were incubated overnight with 0.5% SDS in H_2O_{DEPC} and then washed with H_2O_{DEPC} before use. Electrophoresis was performed in a 1.2% (w/v) RNase-free agarose gel.

10x FA buffer:

MOPS-NaOH, pH 7.0	200 mM
NaAc	50 mM
EDTA	10 mM

Formaldehyde agarose gel:

Agarose	0.48 g
10x FA buffer	4 ml
37% Formaldehyde	720 µl
H_2O_{DEPC}	*ad* 40 ml

Running Buffer:

10x FA buffer	40 ml
37% Formaldehyde	8 ml
H_2O_{DEPC}	*ad* 400 ml

Samples as well as the marker (RiboRuler HR, Thermo Scientific) were mixed 1:2 or 1:3 with Loading dye (Thermo Scientific) which contains bromophenol blue, xylene cyanol FF as well as ethidium bromide. The RNA was denatured at 70°C for 5 min. The samples were then cooled on ice for 3 minutes and briefly centrifuged. 5-20 µl were applied to the gel. The electrophoresis was performed at 100 V for 60 min with a power supply PowerPacTM Basic (BioRad, Munich, Germany) in an electrophoresis

gel chamber (HE-33 mini horizontal submarine unit, Hoefer™, Holliston, MA, USA) filled with Running Buffer. The RNA was visualized under UV light of 254 nm in a Molecular Imager® (GelDocTM XR+ Imaging System, BioRad, Munich, Germany). The documentary of the gel was done with a Quantity One 1-D analysis software (version 4.6.9, BioRad, Munich, Germany).

3.10 *In vitro* protein expression

3.10.1 Coupled ivTT

Coupled *in vitro* transcription and translation (ivTT) was performed in a volume of 50–150 µl at 30 or 37°C in a thermocycler (Mastercycler® personal, Eppendorf, Hamburg, Germany). Mostly, a total volume of 50 µl reaction mixture was prepared in a 200 µl RNase-free PCR-tube.

Table 9. Coupled ivTT mixture.

	volume	**Final**
10x Puffer ivTT buffer	5 µl	1x
NAD/ATP (3,3 mM/12 mM)	5 µl	0.33 mM/1.2 mM
Phosphoenolpyruvate (330 mM)	5 µl	33 mM
10x rNTPs (s.o.)	5 µl	1x
E. coli tRNA mix (7.2 mg/ml)	0.85 µl	120 µg/ml
Mix of amino acids (20 mM each)	5 µl	2 mM each
Ribolock (40 U/µl)	1 µl	1,2 U/µl
Cell extract	13.5 µl	27% (v/v)
RNAP*	X µl	X U/µl
DNA		665 ng
H_2O_{DEPC} *ad* 50 µl		

*T7-RNAP (20 U/µl, *Thermo Scientific*): 2.5 µl → 1 U/µl

10x IvTT buffer:

Folinic acid	340 mg/ml
Potassium glutamate	1.3 M
Ammonium acetate	100 mM
Magnesium glutamate	120 mM
Coenzyme A	2.7 mM
Spermidine	15 mM
Putrescine	10 mM
Sodium oxalate	40 mM

3.10.2 Linked ivTT

Linked *in vitro* transcription and translation was performed to check the amount and quality of RNA and to increase the concentration for *in vitro* translation. The *in vitro* transcription mixtures, which were subsequently used for *in vitro* translation, had a total volume of 200 µl and were incubated for 2h at 37°C.

Table 10. *In vitro* transcription mixture.

	volume	final
10x T7 Transcription buffer (Kit)	20 µl	1x
10x rNTPs (s.o.)	20 µl	1x
Ribolock (40 U/µl)	4 µl	1,2 U/µl
RNAP*	10 µl	X U/µl
DNA		2.66 µg
H_2O_{DEPC} *ad* 200 µl		

*T7-RNAP (20 U/µl, *Thermo Scientific*): 2.5 µl → 1 U/µl
*$RNAP_{elephant}$ (unknown U/µl)

Afterwards, the *in vitro* transcription was treated with RNase free DNase I (NEB), cleaned and concentrated (3.9.2), the amount of transcript was measured (3.9.4) and the quality was monitored using RNA gel electrophoresis (3.9.5).

For *in vitro* translation, cleaned RNA was used in as high amounts as possible.

Table 11. *In vitro* translation mixture.

	volume	final
10x Puffer ivTT buffer	5 µl	1x
NAD/ATP (3,3 mM/12 mM)	5 µl	0.33 mM/1.2 mM
Phosphoenolpyruvate (330 mM)	5 µl	33 mM
E. coli tRNA mix (7.2 mg/ml)	0.85 µl	120 µg/ml
Mix of amino acids (20 mM each)	5 µl	2 mM each
Ribolock (40 U/µl)	1 µl	1,2 U/µl
Cell extract	13.5 µl	27% (v/v)
RNA (cleaned & concentrated)		> 1 µg
H_2O_{DEPC} *ad* 50 µl		

The reaction was incubated for at least 4h at 37°C.

3.11 Screening methods for activity of *in vitro* expressed enzymes

3.11.1 *Para*-nitrophenol (pNP) ester assay

For the screening of *in vitro* expressed lipases from *Ignicoccus hospitalis (I. hospitalis)*, a *para*-nitrophenol (*p*NP) ester assay was performed. To reduce backround signals coming from the cell extract, the *in vitro* mix was incubated at 70°C for 2h after *in vitro* expression and centrifugated for 10 min (10,000 rpm, 4°C). The activity test was performed by incubating the *in vitro* mix supernatant with 1 mM *para*-nitrophenyl palmitate (*p*NP-C16) (Sigma Aldrich, Munich, Germany) in 0.1 M potassium phosphate buffer (PB, pH 8.0) at 90°C. The success of *the vitro* expression was analyzed by the colorimetic quantification of the released yellow *p*NP. *P*NP is the product of the enzymatic hydrolysis of *p*NP-C16, which was measured at 405 nm in a microtiter plate spectrophotometer (Benchmark, Bio-Rad, Hercules, CA, USA) against a buffer control.

3.11.2 Methylumbelliferyl (MU) ester assay with immobilized enzymes

In addition to the use of *p*NP esters, fluorescent substrates were used for the detection of *in vitro* expressed enzymes with esterolytic activity.

3.11.2.1 Immobilisation of His_6-tagged proteins for MU assays

MU assays with *in vitro* expressed His_6-tagged PET hydrolases were performed after immobilisation on Protino® Ni-NTA beads (Machery Nagel) using immobilized metal ion affinity chromatography (IMAC). The beads consist of 6 % cross-linked agarose on which the chelating ligand nitrilotriacetic acid (NTA) is immobilized (Figure 27). Because the resin is precharged with Ni^{2+} ions it can be used directly for IMAC. 30 µl of the Ni-NTA bead suspension were equilibrated twice for 5 min with 300 µl of Soerensen buffer (centrifugation 2000xg, 2 min, RT). Subsequently the *in vitro* translation mix was diluted 1:5 with equilibration buffer and pipetted onto the beads. For binding, the beads were incubated for 2 h at 4°C while shaking gently. After incubation, the beads were sedimented (3 min, 1000xg, RT) and the supernatant was discarded. Afterwards the beads were washed twice with 1 ml Soerensen buffer and the activity assay with 4-methylumbelliferyl (4-MU) octanoate was performed.

Soerensen buffer (pH 8.0):

Potassiumdihydrogenphosphate solution (50 mM)	3.1 ml
Dinatriumdihydrogenphosphate solution (50 mM)	96.9 ml

PET hydrolase activity buffer:

Soerensen buffer (pH 8.0)

0.1% (w/v) gum arabic

5 mM sodium deoxycholate

3.11.2.2 Methylumbelliferyl (MU) ester assay

4-methylumbelliferone esters do not fluoresce unless they are cleaved by specific enzymes which leads to the release of the fluorophore. Enzymatic activity can be detected by measuring the fluorescent 4-MU which emits light at 460 nm when excited by 365 nm light.

For the detection of *in vitro* expressed PET hydrolases 4-MU octanoate was used as substrate. 230 µl of PET hydrolase activity buffer was mixed with 34.5 µl of DMSO and 34.5 µl of 4-MU octanoate (10mM stock solution in isopropanol). The activity assay was performed directly with immobilized enzymes for 30 min at 60°C. As a buffer control washed Ni-NTA beads were incubated with the reaction mixture. Afterwards the beads were sedimented and and the reaction was cooled down (3 min, 1000xg, 4°C). The supernatant was transferred to a 96 well microtiter plate and measured in a microtiter plate spectrophotometer (Benchmark, Bio-Rad, Hercules, CA, USA).

3.11.3 ß-Lactamase

For screening for ß-lactamase activity, nitrocefin was used as substrate. Nitrocefin is a chromogenic cephalosporin which is routinely used as a substrate to detect the presence of beta-lactamases produced by various bacteria (Ghavami *et al.*, 2015). Nitrocefin contains a beta-lactam ring which hydrolysation leads to a color change from yellow (max. at pH 7.0 = 390 nm) to red (max. at pH 7.0 = 482 nm) (Figure 6). The reaction can be followed by measuring absorbance at 480-500 nm. The nitrocefin was dissolved in DMSO and diluted with phosphate buffer (100 mM, pH 7.2) to a final concentration of 1 mg/ml. The solution was stored at -20°C for a few weeks.

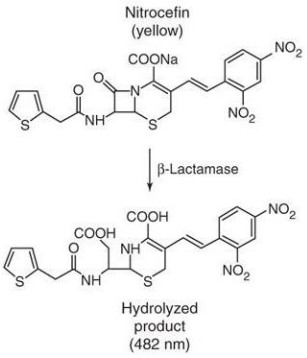

Figure 6. Hydrolysis of nitrocefin. The amide bond in the beta-lactam ring is hydrolysed by the ß-lactamase, which leads to the color change. Figure reprinted from (Remy *et al.*, 2007)

For the activity test, equal amounts of phosphate buffer, nitrocefin stock solution and sample were mixed and incubated at 22°C until a visible color change was observed. The whole spectrum measurements were carried out with the NanoPhotometer® (Implen) on a 2 µl scale. The hydrolyzed nitrocefin was measured at a single wavelength in a microtiter plate spectrophotometer (Benchmark, Bio-Rad, Hercules, CA, USA).

3.12 *In vitro* compartmentalization

The microfluidic chips were prepared in the Leibniz Institute for Interactive Materials (DWI) Aachen by Richard Meurer. The functionalization consisting of hydrophobic sol-gel coating and hydrophilic acrylic acid coating was UV-polymerized.

Outer phase: 2 wt.% Poly(vinyl alcohol) (MW: 31-50 kDa; 87-89% hydrolyzed) in DI-H_2O with 175 µl/h

Middle phase: 7.1 mg / mL PBD-b-PEO in toluene: Chloroform (2: 1 (v/v)) with 125 µl/h

Internal phase: *In vitro* expression mixture with 30 µl/h

Coupled *in vitro* expression mixtures were encapsulated in polymersomes consisting of PBD-b-PEO as shown in Figure 7.

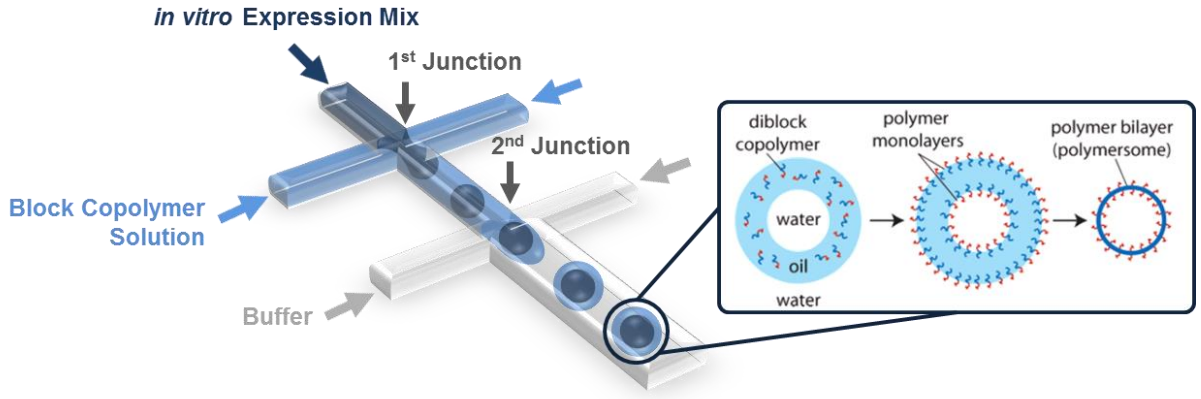

Figure 7. Scheme of sol-gel-coated and acrylic acid-coated microfluidic device for forming polymersomes.
The device design with two cross junctions enables the separate injection of the inner, water-based phase (the in vitro expression mix), the diblock copolymer solution and the buffer. In the first junction drops are formed, which are encapsulated by passing the second junction. The double-emulsion templated polymersomes consist of a bilayer shell of diblock copolymers and an aqueous core (Thiele *et al.*, 2010). Figure unpublished and reprinted from Richard Meurer (RWTH Aachen).

After formation of the polymersomes, they were collected in 200 µl Eppendorf tubes containing phosphate buffer and incubated for 4h at 30°C or 37°C. *In vitro* expression of GFP as well as CelA2 (pIX3.0-RMT7-CelA2) which was substituted with resorufin-ß-D-cellobioside, the fluorescent cellulose substrate, was performed. The expression success was monitored by fluorescence microscopy.

3.13 Computational analysis

The following programs and databases were used to analyze the nucleotide and amino acid sequences used in this work:

3.13.1 Programs

BioEdit (http://www.mbio.ncsu.edu/bioedit/bioedit.html)

Clone Manager Suite 9 (SciCentral Software, licenced)

Mega 6 (http://www.megasoftware.net/)

Quantity One (Bio-Rad Laboratories, Munich, Germany)

3.13.2 Databases

IMG (integrated microbial genomes; https://img.jgi.doe.gov/cgi-bin/er/main.cgi)

NCBI Database (http://www.ncbi.nlm.nih.gov/)

- BLAST Alignment tools (http://blast.ncbi.nlm.nih.gov/Blast.cgi)
- GenBank® Sequence database (http://www.ncbi.nlm.nih.gov/genbank/)

Pfam (http://pfam.sanger.ac.uk/)

UniProt (Swiss-Prot and TrEMBL, http://www.uniprot.org/)

4 Results

4.1 Recombinant RNAPs

Transcription is the most limiting step in the expression of metagenomic genes in a surrogate host (1.1.2). To improve the transcription of metagenomic genes, especially of thermophilic origin, coding for potentially new thermozymes, the first part of this work was the heterologous expression of exciting RNAPs.

4.1.1 Recombinant RNAP from *T. thermophilus*

For effective *in vitro* transcriptions at elevated temperatures, first, the bacterial RNAP from the extreme-thermophile *T. thermophilus* wascloned, expressed and purified. *T. thermophilus* has already been proposed as a promising host candidate for function-based screening of esterase-active thermozymes (Angelov *et al.*, 2009; Leis *et al.*, 2015a).

4.1.1.1 Cloning and heterologous expression

The genes encoding the subunits of the RNAP from *T. thermophilus* as well as the sigma^{70} factor were cloned into the expression vectors pET-21a and pET-28a. To verify the correct insertion of the genes, direct colony PCR was carried out (3.7.4.1) and the positive clones were sequenced (3.13). Subsequently, the constructs were transformed into *E. coli* BL21 (DE3) or Rosetta gami 2 and the optimal expression conditions for each gene were found (Table 12). The subunits were expressed individually and then analyzed for solubility via SDS-PAGE (3.8.6) using both the extract and the pellet fraction.

Table 12. Overview of the main characteristics of the individual RNAP subunits.

Gene	Length (bp)	Subunit	Size (kDa)	Soluble?	Stability	Time (min)
rpoA	957	α	40	yes	90°C	< 120
rpoB	3369	β	140	yes	60°C	10
rpoC	4584	β´	170	yes	60°C	10
rpoZ	309	ω	10	no	-	-
rpoD	1281	σ-70	90	yes	90°C	< 60

4.1.1.2 Purification

Soluble subunits were roughly purified using heat denaturation, whereas the insoluble ω -subunit was expressed as inclusion bodies and needed to be recovered using urea (3.8.2.2).

In general, larger amounts of soluble protein from the non-tagged subunits could be expressed heterologously.

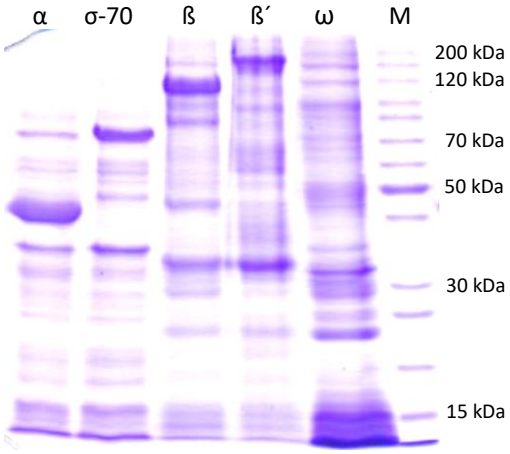

Figure 8. Rough purification of single subunits.

Soluble subunits (α, ß, ß´and σ-70) were crudely purified via heat treatment. The ω subunit (with His_6-tag) was solubilized from inclusion bodies by several washing steps with 0.2% Na-deoxycholate containing lysis buffer and resuspension in 6 M urea containing denaturation buffer. All subunits were separated in a 12% SDS-polyacrylamide gel.

4.1.1.3 Reconstitution

The reconstitution of recombinant *T. thermophilus* RNAP was performed according to the procedure of (Kuznedelov & Severinov, 2009) but modified in temperature due to the thermophilic environment of the parent organism. Reconstitution by dialysis overnight was tested at 3 different temperatures and then checked for efficiency by semi native-gel electrophoresis (3.8.7).

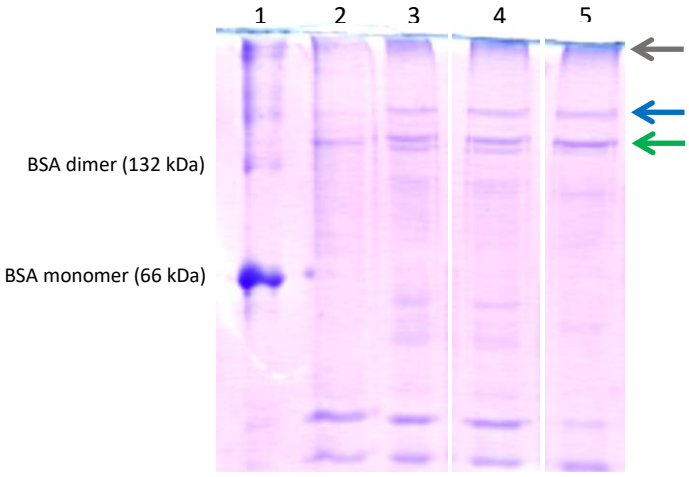

Figure 9. *In vitro* reconstitution of the RNAP from *T. thermophilus*.

1: BSA (5 µg)**,** 2: Reconstitution mix before incubation, 3: after dialysis at 4°C, 4: after dialysis at RT, 5: after dialysis at 60°C. 16 µl of each sample were loaded onto the gel followed by semi-native gel electrophoresis using a 4% stacking gel and a 10% separation gel. Afterwards the gel was stained in Coomassie. The grey arrow indicates the presence of the reconstituted RNAP (~ 400 kDa), the blue arrow indicates the ß´subunit (170 kDa) and the green arrow indicates the ß-subunit (140 kDa).

The resulting band at the upper edge of the gel suggests the successful reconstitution of the core enzyme (indicated by the grey arrow), as it becomes clearly visible only in sample 3, 4 and 5, ie after dialysis (Figure 9). Based on the band intensity, the amount of reconstituted RNAP does not differ after dialysis at different temperatures. This initial result would have to be further tested, for example, by purification of the RNAP via size exclusion chromatography (SEC) and subsequent analysis of individual peaks in an agarose gel. Semi-native gel electrophoresis over a gradient, for example 4-20%, would allow better separation and thus allow a more accurate statement about the actual size of the proteins.

The blue arrow marks the band corresponding to the ß'subunit (170 kDa) and the green arrow marks the band corresponding the ß-subunit (140 kDa), indicating that either the ratio of the individual subunits was not optimally chosen, or that the reconstitution did not work completely. It is noticeable that sample 2 shows no clear bands that indicate the individual subunits.

4.1.1.4 *In vitro* transcription

The activity of reconstituted, recombinant RNAP from *T. thermophilus* was analyzed by *in vitro* transcription (3.9.1) and subsequent analysis in the RNA agarose gel (3.9.5). Genomic DNA from *T. thermophilus* HB27 and *Geobacillus* sp. GHH01 were used as template. The amount of template, the temperature, the buffer components, and the duration of the *in vitro* transcription were varied. Unfortunately, no *in vitro* transcription could be detected with the RNAP from *T. thermophilus* at any conditions (no picture shown).

4.1.2 Recombinant viral RNAP from the elephant metagenome

Since it was shown that *in vitro* transcription with bacterial RNAPs (*Geobacillus* sp. GHH01, (Kinfu *et al.*, 2017); *T. thermophilus*, this work) resulted in insufficient RNA yields with high error rates, the transcription was switched to a viral system.

Viral RNAP consist of only one subunit (1.2.1.1), which significantly reduces the cloning and expression effort. This also eliminates the critical step of reconstituting the core/holo enzyme from individual subunits. Viral/phage RNAP like the commonly used T7 RNAP are characterized by their very high activity, elongating mRNA strands about five times faster than the *E. coli* RNAP (Golomb & Chamberlin, 1974). Another advantage of a viral RNAP is its ability to generate very long mRNAs as needed for transcription of metagenomic DNA. The fact, that viral transcription systems are poorly terminated by unrelated transcription terminators found in metagenomes, encourages the use of a phage RNAP for cell-free, function-based metagenomics (McAllister *et al.*, 1981).

4.1.2.1 Metagenome search and gene synthesis

For this purpose, different metagenomes were screened sequence-based for viral genes with the lowest possible sequence homology to common systems such as T3, T7 or SP6 (Figure 29). This was done by Dr. Simon Güllert and is not described in detail in this work.

In addition, the goal was to find viral RNAPs that are more thermostable than the commercial RNAPs to perform *in vitro* transcriptional assays of genes of thermophilic origin at elevated temperatures.

4.1.2.2 Cloning and heterologous expression

The gene of the $RNAP_E$ (RNAP“E” for "elephant", since it´s the RNAP from the microbial community from elephant feces) was synthesized and inserted into a pEX-vector with ampicillin resistance (pEX-A258) by Eurofins Genomics (Eurofins MWG Operon, Ebersberg, Germany). Additional restriction sides at 5' (NheI) and 3' (SalI) were used for cloning into the expression vector pET21a (3.7.6), kindly taken over by Dr. Maike Jahnke. The $RNAP_E$ could be successfully produced as a soluble protein in *E. coli* BL21 with an N-terminal His_6-Tag (Figure 11; 2).

4.1.2.3 Purification and storage

The $RNAP_E$ could be expressed heterologously in *E. coli* BL21 and then purified via IMAC as described in 3.8.2.3. The results of the purification and corresponding chromatogram are shown in Figure 10. The $RNAP_E$ was eluted in one peak, whereby 5 fractions of 0.8 ml were collected. The flow through, the wash fraction, and the two prominent peaks were examined in an SDS-PAGE for the presence of the $RNAP_E$.

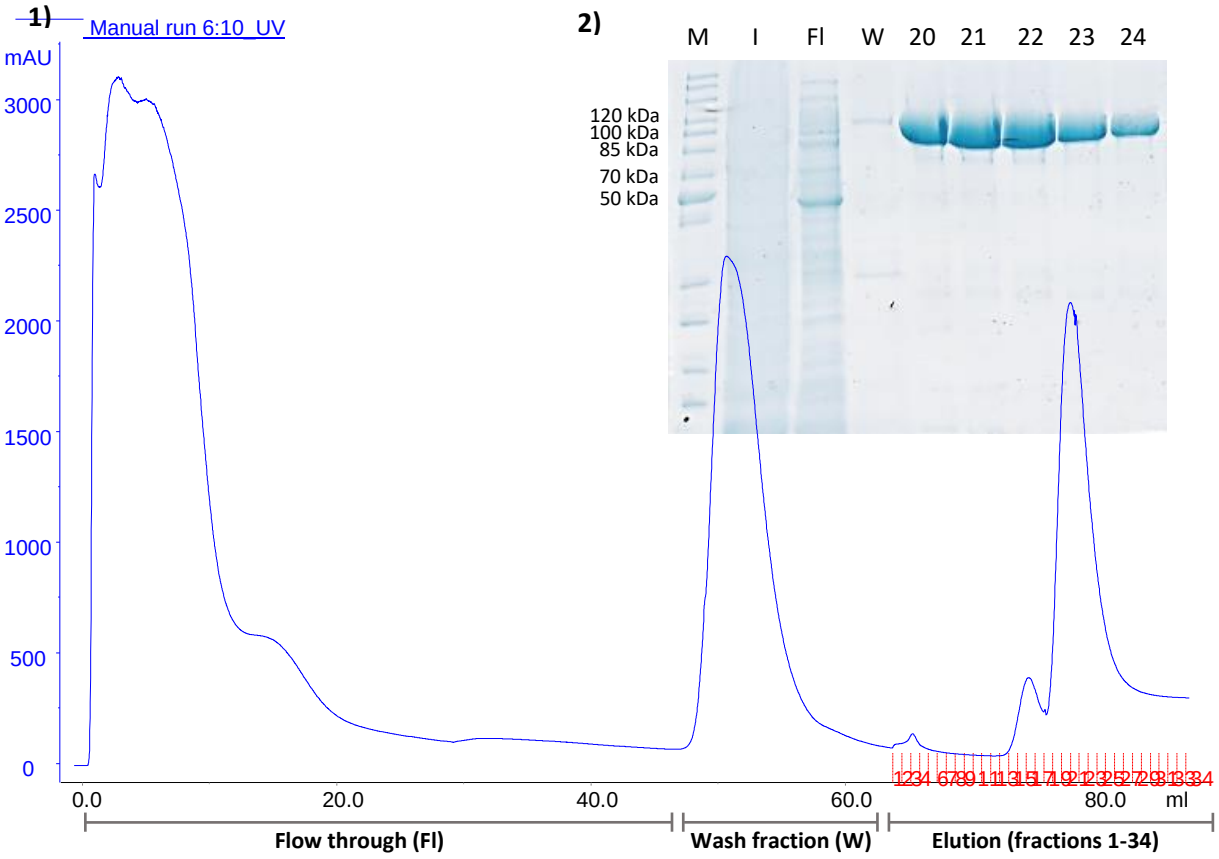

Figure 10. The $RNAP_E$ was purified on the FPLC system with a HisTrap FF, 5 ml column.

1) Chromatogram of the purification of the RNAP from 200 ml cell culture. The column (5 ml) was washed with 25 ml of wash buffer, whereas the middle peak (about 50 min) is showing slightly matrix-bound proteins. The $RNAP_E$ was eluted by a high imidazole concentration (250 mM) with a prominent peak at 80 ml (fraction 20-24)

2) Samples taken from the purification were analyzed in 12% SDS gel, which was stained with Coomassie Brilliant Blue. The washing fraction was collected when the first high peak appeared. The strongest signal in the elution fractions in the gel (21 and 22) coincides with the appearance of the peak in the chromatogram. M: Protein marker (6 µl), I: Input (8 µl), Fl: Flow trough (12 µl), W: Wash fraction (16 µl), 20-24: Elution fractions (4 µl).

From 1 liter of culture induced with 1 mM IPTG at OD_{600} of 0.6, up to 200 mg of purified enzyme could be produced. After purification, the $RNAP_E$ was rebuffered (3.8.3) and stored at -20°C for several months without loss of activity.

4.1.2.4 *In vitro* transcription

The $RNAP_E$ was tested for activity. For this purpose, *in vitro* transcriptions (3.9.1) were performed, whereby various buffer systems, temperatures, promoters and genes were tested. The transcription efficiency was quantified by measuring the RNA (3.9.4) and visualized by gel electrophoresis (3.9.5).

4.1.2.4.1 Buffer conditions and promotor search

To perform an initial characterization of the $RNAP_E$, a promoter search was carried out by Dr. Jahnke in addition to the temperature profile, and various buffers were tested. To date, the $RNAP_E$ shows the highest promoter specifity by using the T7 promoter. In addition, the $RNAP_E$ shows the highest transcriptional activity when using the transcription buffer from the commercial T7 polymerase kit (New England Biolabs), compared to the transcription rate with TB2 and ivTT buffer.

4.1.2.4.2 Temperature profile

To determine the temperature profile of $RNAP_E$, *in vitro* transcriptions were performed in a gradient cycler. The metagenomic-derived cellulase CelA2 (in pIX3.0-RMT7) was used as a template. The relative amount of RNA was analyzed directly after transcription by RNA electrophoresis (3.9.5). It shows that the optimum reaction temperature of the $RNAP_E$ is around 35°C (Figure 11). This is comparable to the temperature optimum of commercial RNAPs such as the T7. By increasing the temperature by approximately 2°C, the reaction power decreases. Above 45°C, no activity was detectable.

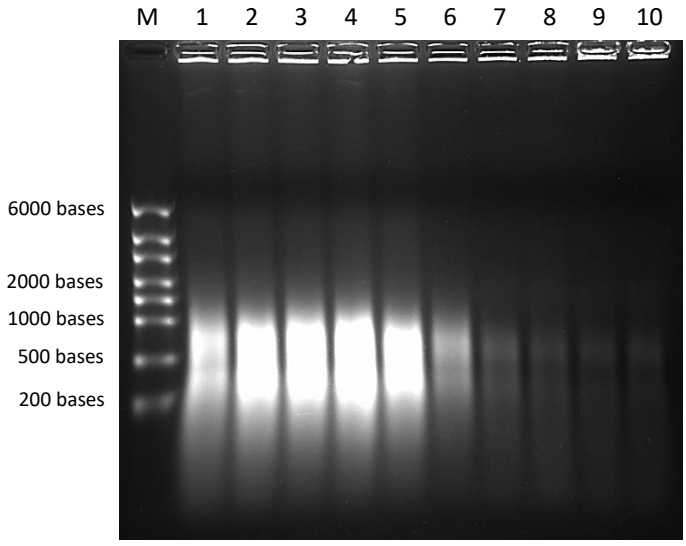

Figure 11. *In vitro* transcription activity of $RNAP_E$ at different temperatures.

RNA was run in a 1.2% agarose gel containing 0.7% (v/v) formaldehyde and stained with ethidium bromide, which is included in the loading dye (ThermoFisher Scientific GmbH, Schwerte, Germany). M: RNA ladder (2 µl), 1: 30.0°C, 2: 32.1°C, 3: 33.7°C, 4: 35.5°C, 5: 37.4°C, 6: 39.3°C, 7: 41.1°C, 8: 42.6°C, 9: 43.8°C, 10: 44.5°C (14 µl each).

4.1.2.4.3 General activity

To investigate the applicability of the $RNAP_E$ for function-based metagenomics, *in vitro* transcriptions with a diversity of templates were performed. The focus was on the transcription of metagenomic derived genes coding for heat-tolerant hydrolases, whose demand for industrial applications has been increasing rapidly in recent years. In addition to GFP, which is a common control for cell-free protein expression, model experiments were carried out mostly with genes coding for already characterized enzymes of metagenomic origin, as well as a β-lactamase from *E. coli*. The former includes a bacterial lipase (LipS), a bacterial cellulase (CelA2), thermostable lipases (Igni13, Igni15, Igni18) from the hyperthermophilic archaeon *I. hospitalis* (Kobus *et al.*, 2019) and recently published polyethylene terephthalate (PET)-degrading hydrolases (PET2, PET5, PET6, PET12) (Danso *et al.*, 2018). In addition, *in vitro* transcription of the violacein operon from *Janthinobacterium* sp., which consists of 5 genes, was performed. For this, the T7 promoter was used and all genes were amplified with a His_6-tag coding sequence at the 3 'end.

It was shown that the yield of transcript is about 2 to 5 ug mRNA from 100 µl *in vitro* transcription using a pET vector construct as template. With the gene for the cellulase CelA2, which is cloned into the pIX vector, by far the highest transcription rate was achieved (23.1 µg mRNA from 100 µl *in vitro* transcription).

Table 13. Transcription potential of $RNAP_E$ on different DNA templates.

In vitro transcription reactions with volumes of 200 µl were each carried out using 2.66 µg of plasmid DNA with different inserted genes.

gene	µg mRNA from 100 µl *in vitro* transcription
sfGFP	4.8
celA2	23.1
lipS	1.9
ß-lactamase	3.9
vioABCDE	6.3
igni13	1.8
igni14	2.2
igni15	2.3
igni16	2.0
igni18	1.9
pet2	2.4
pet5	2.0
pet6	2.3
pet12	2.0

Comparing the transcript levels of *in vitro* transcriptions with the commercial T7 RNAP and the $RNAP_E$, the $RNAP_E$ shows a slightly lower activity.

In addition to the amount of transcript, the quality of the mRNA is critical for the subsequent in *vitro* translation. To get an initial assessment of the actual amount as well as the correct size and quality of the mRNA, the transcripts were separated and visualized in a formaldehyde-containing agarose gel (3.9.5). For this, the samples were subjected to DNAse digestion after transcription to eliminate the template DNA. After photometrical quantification (3.9.4), about 1 to 2 µg mRNA were loaded on to the gel. It turned out that it is reasonable to purify and concentrate the RNA in a further step. Without this step, the amount of RNA, as well as the purity, is insufficient for analysis in an agarose gel. While the transcripts appear as "clouds" in the range of less than 200 bases before the purification, cleared and concentrated samples mostly show distinct bands corresponding to the predicted size of the gene (data not shown).

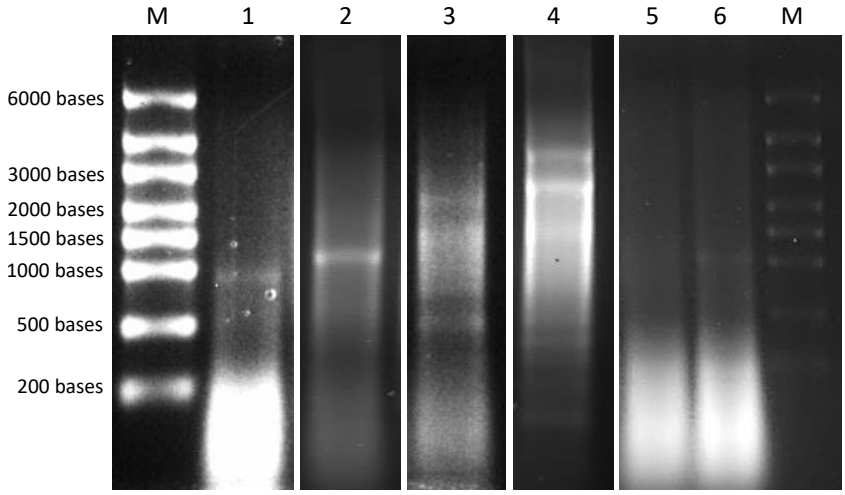

Figure 12. *In vitro* transcription of varios templates with $RNAP_E$.

After purification, 1-2 µg RNA of each sample was run in a 1.2% agarose gel containing 0.7% (v/v) formaldehyde and stained with ethidium bromide. M: RNA ladder (2 µl), 1: sfGFP, 2: ß-lactamase, 3: vioABCDE, 4: Cel_{A2}, 5: PET2, 6: PET6.

Before performing RNA gel electrophoresis, all RNA samples were boiled at 70°C to dissolve secondary structures. As shown in Figure 12, all transcription reactions with circular plasmid constructs containing different genes as templates show completely different mRNA profiles on the gel. In addition to low molecular weight RNA “clouds”, each sample shows at least one stringent band, some of very low intensity. All transcripts showed the band of the calculated molecular size of the gene including the additive sequences of the vector flanking the target gene. Sample 1 shows the transcript of sfGFP (~ 720 bases) from pET-21a (additional ~ 230 bases) with a unique band of nearly 1000 bases. In addition to this band, the gel shows high signals at a size of less than 200 bases, which is due either to degration of the RNA during the purification process or to a too short denaturation time and thus remaining

secondary structures. The transcript of the beta-lactamase (860 bases) from the pET-21a-vector shows a clear band between 1000 and 1500 bases, which confirms the calculated transcript size of about 1100 bases. The transcript profile of the violacein operon (sample 3) is more complex. The template used was the cloning vector pDrive, which has no terminator sequence for the transcription. Therefore, it was previously restricted after the stop codon of the vioE gene. All five vio genes (vioA, vioB, vioC, vioD and vioE) are controlled by one promoter sequence. The band pattern shows 3 distinct bands and a concentration of several bands between 1000 and 1500 bases. The lowest band at about 500 bases can be associated with the transcript of vioE (574 bases). The very weak band signal at 3000 bases suggests the transcript of vioB (2998 bases). The transcripts of vioA (1264 bases), vioC (1288 bases) and vioD (1120 bases) can be expected in the range between 1000 and 1500 bases. A band corresponding to the transcript size of the whole violacein operon (7345 bases) is not evident. Sample 4 shows the transcript of the cellulase CelA2 from the pIX3.0-RMT7 vector, which is a modified (Schwaneberg Group (RWTH Aachen)) version of the vector pIX3.0 from Qiagen. The RNA shows a complex banding pattern, whereas the signal between 1500 and 2000 bases is assigned to the desired transcript size due to the gene length of CelA2 of 1815 bases. The other bands are due to possible termination of the transcription by the polymerase, the degradation of RNA, RNA-RNA interactions and secondary structures of RNA. This can happen, for example, by too short boiling or the refolding of the RNA by subsequent cooling. Signals above the calculated transcript length are due to RNA in nicked form, whereas signals below the calculated transcript length suggest the supercoiled form. Sample 5 and 6 show the result of the transcription of PET hydrolases PET2 and PET6 from pET21a. Each sample shows at least one stringent band in addition RNA "clouds" at low molecular weight. The band signal slightly above 1000 bases is very weak, but the size can be confirmed by the calculated transcript size of about 1160 bases (length of the PET hydrolase gene is 927 bases, additional 230 bases from the pET-21a vector).

In conclusion, all transcriptions resulted in a sufficient amount of mRNA. On the basis of RNA gel electrophoresis, it was also possible to get an initial assessment of the quality of the transcripts with regard to the correct length. However, no statement can be made about the actual quality of the transcripts regarding to the error rate and translatability.

4.1.2.5 Transcription of genomic DNA from different bacteria

To test the applicability of the $RNAP_E$ for cell-free metagenomics, *in vitro* transcriptions with genomic DNA of different bacteria were performed (3.9.1). Genomic DNA from 5 different gram-negative and positive bacteria was isolated (3.7.1.1) and used as template for production of RNA.

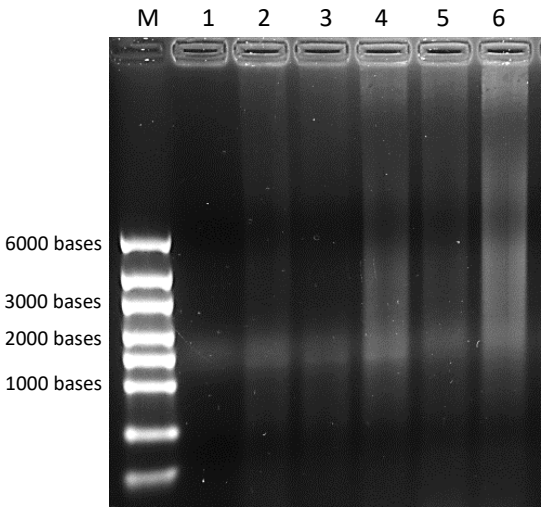

Figure 13. Transcription potential of $RNAP_E$ on genomic DNA templates from different bacteria.

Transcription reactions were carried out using genomic DNA templates provided by Dr. Maike Jahnke. 12 µl of the *in vitro* transcription were loaded onto the gel without purification. The samples were DNase treated. RNA of each sample was run in a 1.2% agarose gel containing 0.7% (v/v) formaldehyde and stained loading dye. M: Riboruler high range RNA ladder (ThermoFisher Scientific GmbH, Schwerte, Germany), 1: Negative control (without DNA template), 2-6: *In vitro* transcription using genomic DNA from 2: *Geobacillus* sp. GHH01, 3: *Bacillus subtilis* TEp1031, 4: *C. sambhunathii*, 5: *E. coli* BL21 Codon Plus RIL, 6: *P. antartica.*

It was possible to successfully transcribe genomic DNA of different bacteria *in vitro*. Figure 13 shows that the $RNAP_E$ has an individual transcription profile. Thus, the reactions with genomic DNA of *Chelatococcus sambhunathii* (*C. sambhunathii*) (sample 4) and *P. antartica* (sample 6) resulted in the highest production of mRNA. However, the information about which genes are transcribed and which promoter sequences are recognized can only be obtained by sequencing analyses. With RNA sequencing, it would also be possible to evaluate how many regions in the bacterial genome are of viral origin.

4.1.2.6 Transcription of fosmid clones from metagenomic libraries

An important step in the classical metagenomic approach is the screening of metagenomic libraries. Metagenomic DNA is usually cloned into fosmids and then transformed into a suitable host for function-based screening. For this reason, the $RNAP_E$ was tested for its ability to transcribe fosmid clones. Metagenomic-derived fosmid clones, which showed hydrolytic activity in classical function-based screening, were kindly provided by Dr. Antonio García-Moyano (University of Bergen, Norway). The fosmids were used as described in the protocol for *in vitro* transcription with plasmid DNA with the $RNAP_E$ (3.9.1). The fosmid clones contain a T7 promoter. As this is unpublished data from another working group, no further information is given regarding the clones and their activity. The result of the

in vitro transcription experiments with fosmid clones was verified by photometric mRNA quantification (3.9.4) after DNase digestion (0) and purification (3.9.2), as well as by electrophoretic analysis (3.9.5).

Table 14. Transcription of six fosmid clones from a metagenomic library.

Fosmid	**µg mRNA from 100 µl *in vitro* transcription**
TB2	1.3
JP	4.4
MSeaJ2	1.4
MBOO6	11.0
MSea M5	2.8
MSedi12	8.4

All six clones of the metagenomic fosmid-library could be transcribed *in vitro* with $RNAP_E$ (Figure 14), whereby the yield of mRNA was comparable to plasmid-derived genes (Table 13). Between 1.3 and 11.0 µg mRNA from 100 µl *in vitro* transcription could be quantified.

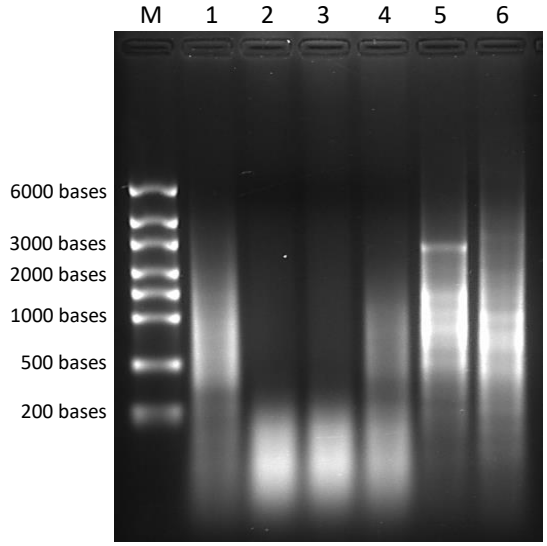

Figure 14. mRNA produced by the $RNAP_E$ using metagenomic-derived fosmid clones as template.

12 µl of the *in vitro* transcription were loaded onto the gel after purification and DNase treatment. Each sample was run in a 1.2% agarose gel containing 0.7% (v/v) formaldehyde and stained loading dye. M: Riboruler high range RNA ladder (ThermoFisher Scientific GmbH, Schwerte, Germany), 1 to 6: Transcripts of six different fosmid clones.

In the agarose gel (Figure 14) all six clones show an individual band pattern, whereby no clear statements can be made about the correctness of the band sizes. Also, no statement about the quality of the transcript is possible.

4.2 Cell-free protein synthesis

4.2.1 Coupled vs linked cell-free expression

There are two different approaches of CFPS, the coupled and the linked approach. These differ in two essential aspects, the template and the reaction space. In coupled protein expression, DNA is used as a template, with transcription and translation running spatially and temporally in parallel (1.2). In linked protein expression, *in vitro* transcription takes place followed by RNA purification. In a second step, the purified RNA is used for *in vitro* translation. The transcription and translation thus take place temporally and spatially separated from each other.

To test extracts for their translation efficiency, *in vitro* protein expression was first carried out as coupled reaction. With this system, no statement can be made about the success of the stranscription step. In addition, the reaction temperature in the coupled protein expression is dependent on the optimal temperature of the RNAP, whereby the temperature profile for the translation cell extracts cannot be determined. Therefore, the system was decoupled later (linked CFPS). First, the protein expression rate between coupled expression and linked expression was compared (unpublished data from Dr. Maike Jahnke). It was also tested whether the amount of protein and the amount of RNA are proportional.

GFP with His_6-tag under control of the T7 promoter (pIX3.0 vector) was used as the test protein. Reactions were carried out using the T7 RNAP and extract of *E. coli* BL21 CodonPlus.

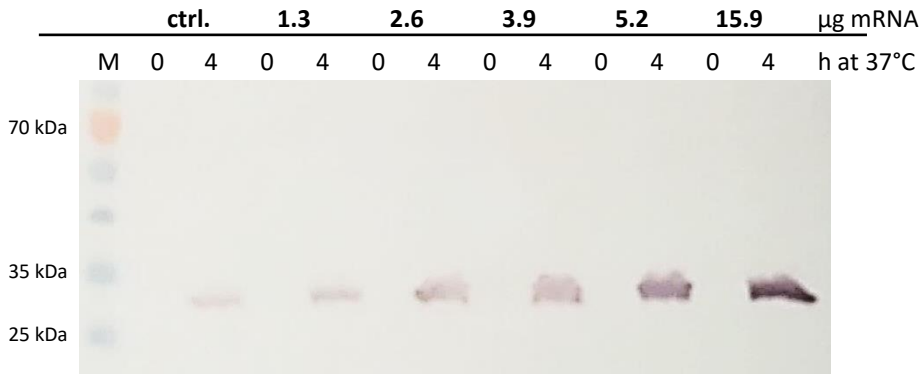

Figure 15. Western Blot immunoassay for detection of His_6-tagged *in vitro* expressed GFP.

His_6-specific antibody was used as primary antibody, the secondary antibody was conjugated with an alkaline phosphatase for detection with NBT and BCIP. M: Marker (Fermentas prestained protein molecular weightmarker #SM0671). *In vitro* expression reactions were loaded onto the SDS gel without purification. The blot picture was kindly provided by Dr. Maike Jahnke.

The Western Blot shows the result of the *in vitro* expression of GFP. As a control (left) a coupled *in vitro* expression according to the standard protocol (3.10.1) with 665 ng plasmid DNA/ 50µl reaction was

performed at 37°C for 4h. In addition to the coupled, linked *in vitro* expressions were performed with an *in vitro* transcription step for 2h at 37°C. The mRNA was then DNAse treated, purified, concentrated and quantified. To investigate the dependency between the amount of mRNA and the resulting amount of protein, *in vitro* translations with different amounts of purified transcript were carried out for 4h at 37°C. Each reaction was sampled before incubation (0h) and after 4h of incubation. Subsequently, the same amount of each *in vitro* expression was loaded onto an SDS gel, the proteins were separated by electrophoresis and the His_6-tagged protein of interest (GFP) was detected by Western blot. The band size was verified using the prestained protein marker and the band intensities were compared.

All *in vitro* expressions show one band between 25 kDa and 35 kDa, which confirms the molecular mass of GFP with 28.9 kDa (238 amino acid residues). The signal of the control (coupled ivTT) is the weakest compared to the linked ivTT. As the mRNA concentration increases, the band thickness and intensity of the signal on the blot increases continuously. By using 1.3 µg of mRNA in a linked translation, the protein yield can be increased in comparison to the coupled system. The amount of RNA template is proportional to the amount of translated protein (tested with a maximum of 15.9 µg mRNA in 60 µl *in vitro* translation).

4.2.2 Comparison of different bacterial translation extracts

To capture the broad spectrum of potential new biocatalysts in a metagenome, *in vitro* expression experiments were performed based on cell extracts from various bacteria, both gram-negative and gram-positive strains. For this purpose, extracts from a variety of different bacterial strains were prepared. These were checked for their applicability to *in vitro* translations.

Figure 16. Cell extracts of the different bacterial strains for *in vitro* translation.

differ in both color and protein concentration, which affects color intensity. Left: *E. coli*; middle: *Bacillus subtilis*, right: *Geobacillus* sp.

In addition to the commonly used extract of *E. coli* (using the strains (1.) MRE600 and (2.) CodonPlus RIL), robust cell extracts from the psychrophilic bacterium (3.) *P. antarctica*, a mesophile (4.) *Bacillus subtilis* strain, two thermophile *Geobacillus* species, the inhouse designed strain (5.) *Geobacillus* sp. GHH01 and (6.) *Geobacillus thermoleoverans (G. thermoleoverans)*, as well as from the thermophile strain (7.) *C. sambhunathii* and the hyper-thermophile (8.) *T. thermophilus* were prepared. These should implement the expression of proteins, that need special conditions like extreme temperatures, to be active.

The different strains were cultivated and growth curves were created to determine the range of exponential growth. Subsequently, a timepoint, or an optical density in the early to mid exponential phase was determined at which the cells were harvested (see Table 6). Cell extracts were prepared (3.6) and the protein concentration determined by Bradford assay (Bradford, 1976). To achieve the best possible cell lysis, parallel experiments were carried out using different lysis methods (data not shown). It turned out that cell disruption by sonification produced the most inactive cell extract. The most concentrated and vital cell extract was prepared by mechanical disruption with the French Press. In addition, (*Geo*-)*bacillus* cell extracts were generated from enzymatically lysed cells (by incubation with lysozyme), but this did not positively affect the vitality of the extract compared to mechanical disruption. The extracts were stored at -70°C and gently thawed on ice before use.
To test the translational ability of the extracts, first, coupled *in vitro* expression with T7 RNAP and His_6-tagged GFP as template were performed. Simultaneously, negative controls without DNA template were run. The expressed GFP was analyzed via Western Blot using a His_6-specific antibody.

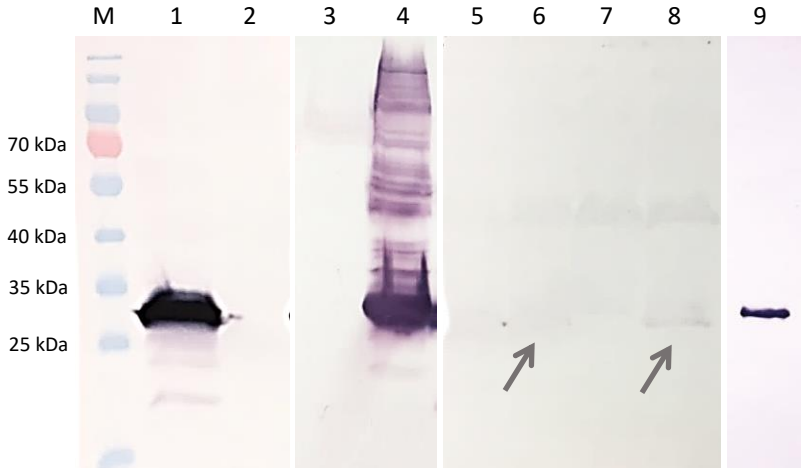

Figure 17. Coupled *in vitro* expression of His6-tagged GFP with cell extracts from different bacteria.
16 µl of each expression mixture were loaded onto a 12% SDS-gel followed by Western Blot analysis. Polyclonal His_6-specific antibody was used as primary antibody, the secondary antibody was conjugated with an AP for detection with NBT and BCIP. M: Marker (Fermentas prestained protein molecular weightmarker #SM0671), 1-8: *In vitro* expression of GFP with different bacterial extracts at different temperatures, 1: *E. coli* BL21 CodonPlus RIL (30°C), 2: Corresponding negativ control, 4: *Geobacillus* sp. GHH01 (37°C), 3: Corr. negativ control, 6: *P.antarctica* (22°C), 5: Corr. negativ control, 8: *P. antarctica* (30°C), 7: Corr. negativ control, 9: Heterologously expressed and purified His_6-tagged GFP.

After coupled *in vitro* expression for 4 hours at different temperatures, the samples were stored at 4°C for several hours or at -20°C for longer storage until analysis. 16 µl of each expression mixture were loaded onto an SDS-gel. Western Blot analysis of GFP *in vitro* expressed with different cell-extracts show a signal of different intensities between 25 and 35 kDa (Figure 17), which corresponds to the molecular mass of GFP (28.9 kDa). Sample 9 shows 100 ng of *in vivo* expressed GFP with His_6-Tag, which was purified, as a control for the correct size of the protein and the efficiency of the Western Blot analysis. Sample 1, which corresponds to the expression of GFP with *E. coli BL21* CodonPlus extract, shows a thick band of the right size, with the negative control showing no signal. The negative control is the same *in vitro* expression reaction run for the same time under the same conditions, but without template DNA. The result of the *in vitro expression* of GFP with extract of *Geobacillus* sp. GHH01 (sample 3: negative control, sample 4: GFP expression) shows, in addition to a strong signal of the expressed GFP, many weaker bands with mainly higher molecular mass. One explanation for the many additional bands might be that the amount of protein, loaded onto the gel was too high. Those bands do not result from proteins from the *Geobacillus* extract, since the negative control has only a weak band above 70 kDa. The additional bands of smaller size in sample 1 and 4 may be due to degradation products of the GFP by proteases from the cell extract. The coupled *in vitro* expression of GFP with extract of *P. antarctica* was performed at 2 different temperatures. Sample 5 and 6 show the result of the expression at 22 °C, sample 7 and 8 at 30 °C. Both expressions show a barely detectable band of the correct size (samples 6 and 8), which does not appear in the respective negative control (samples 5 and 7). However, the intensity of the band in sample 7 is clearly stronger. Obviously, the expression with *P. antartica* extract is more effective at 30°C than at 22°C, with the temperature optimum of the bacterium at 22°C. One reason for that, is the transcription activity of RNAP, which is most efficient at 30 to 37°C. Thus, the low transcription rate ultimately limits the protein yield in coupled expression at reduced temperatures. In addition, the reaction rate increases in enzyme-catalyzed reactions by increasing the temperature, which may be a reason for a higher translation rate in this temperature range. In addition to the band which corresponds to the expressed GFP, the samples 5 to 8 show more bands, in particular at the size of about 50 kDa. Presumably, these proteins are due to proteins from the *P. antarctica* extract, which have a histidine rich part in their amino acid sequence, which leads to the binding of His_6-specific antibodies.

Table 15 lists the various cell extracts tested for their applicability to *in vitro* translations. The total of 8 different bacterial strains consisting of representatives of both Gram-positive and Gram-negative bacteria, cover a wide range of different temperature optima, from 22°C to 70°C.

Table 15. Cell extracts of different bacteria were prepared and subsequently used for *in vitro* protein expression. Both coupled and linked approaches were carried out at different temperatures.

Bacterial strain	Gram +/-	Growth temperature	Temperature optimum	mg/ml Protein in the extract	*In vitro* translation successful
E. coli MRE600	-	30-40 °C	37 °C	32.1	yes
E. coli CodonPlus RIL	-	30-40 °C	37 °C	53.0	yes
Geobacillus sp. GHH01	+	45-55 °C	55 °C	19.5	yes
Geobacillus thermoleoverans	+	60-70 °C	65 °C	16.6	yes
Pseudomonas antarctica	-	4-30 °C	22 °C	38.0	yes
Bacillus subtilis TEB 1030	+	30-40 °C	37 °C	21.4	no
Chelatococcus sambhunathii	-	37-55 °C	50 °C	N/A	no
Thermus thermophilus HB27	-	60-75 °C	70 °C	N/A	no

In addition to the different color of the cell extracts (Figure 16), the extracts differed particularly in their protein concentration which is, in particular, attributable to the effectiveness of the cell lysis. The extracts from *E. coli* had the highest protein concentration of up to 53 mg/ml. The lowest protein concentration was achieved with *G. thermoleoverans* as a Gram-positive organism, which is due to the less effective cell disruption. It was found that the protein concentration of the cell extract is not proportional to its translation efficiency. Figure 17 clearly shows that the protein yield from *in vitro* expressions based on *E. coli* extract is comparable to the protein yield synthesized with *Geobacillus extract*, although both extracts differ extremely in their protein concentration (Table 15). In addition, the protein concentration measured with the Bradford method gives only information about the total protein amount. Nothing can be said about the amount of ribosomes and enzymes involved in the translation process. With an OD_{600} shift of 0.1 at the point of cell harvesting, the efficiency of the cell extract significantly decreased (data not shown).

The standard system for CFPS is based on cell extracts of mesophilic bacteria, mainly *E. coli*. In direct comparison, a higher expression rate could be achieved with extract of *E. coli* CodonPlus RIL than with extract of *E. coli* MRE600. In addition, cell extract of another mesophilic bacterium, *Bacillus* sp., was tested for *in vitro* translation. However, no expressed protein could be detected. Since thermoresistant proteins from metagenomic libraries are of great interest, in particular thermophilic strains were searched for the extract preparation. These should allow an improved translational rate of thermoresistant enzymes at high temperatures. Extracts of moderate mesophilic representatives, such as *C. sambhunathii*, *Geobacillus* sp. GHH01 and *G. thermoleoverans* were used too for *in vitro* translations. The *in vitro* protein expression with *Chelatococcus* extract, as well as the extract preparation itself, was kindly taken over by Dr. Maike Jahnke. Unfortunately, no *in vitro* expression

could be achieved with this system (data not shown). With extract of *Geobacillus* sp. GHH01 and *G. thermoleoverans* metagenomic-derived genes could be expressed, although this was only possible at 37°C. Because of the aim to establish a cell-free protein expression system based on *T. thermophilus*, cell extract for *in vitro* translation was prepared. The extract of *T. thermophilus*, should enable the translation of thermoresistant enzymes (as described in 2) not only at elevated but extremely high temperatures. Unfortunately, no *in vitro* translation could be achieved with *T. thermophilus* extract, neither at 37°C, 45°C, 50°C nor 60°C (data not shown). When incubating the *in vitro* mix at high temperatures, even the proteins originating from extremely thermophilic organisms denature (data not shown). A coupled *in vitro* expression with viral RNAP at high temperatures was also not possible yet, since the $RNAP_E$ show no transcriptional activity at temperature above 45°C (Figure 11).

In order to cover the large temperature range for *in vitro* protein expression and thus enable the expression of cold-adapted proteins, cell extract of the psychrophilic bacterium *P. antartica* was prepared. As already described in Figure 17 coupled *in vitro* protein expression was detectable at both 22°C and 30°C.

4.2.3 *In vitro* expression of metagenomic-derived genes

4.2.3.1 Archael lipases

To investigate the applicability of *in vitro* expression technologies for function-based metagenomics, model experiments were carried out with already characterized enzymes of metagenomic origin as target proteins, especially those, that are difficult or impossible to express *in vivo* in bacterial systems.

The main goal was the time- and cost-saving expression of heat-stable biocatalysts that are of great interest to biotechnology (2). These include in particular hydrolases, which are not limited to a bacterial origin. In the last years, the demand for thermostable lipases for industrial applications has increased. It is noticeable that the origins of most heat-stable lipases are hyperthermophilic archaea. Besides many bacterial genes, the recombinant over-expression of especially archaeal genes comes with many limitations. In particular, when using common bacterial expression hosts such as *E. coli*, only low yields can be produced or the target protein is insoluble or misfolded (Smith & Robinson, 2002). To overcome those limitations, there are various alternative systems for the expression of archaeal proteins. One possibility is to change the expression host, from the prokaryotic to the eukaryotic system. For example, the methylotrophic yeast *Pichia pastoris* has been used successfully in recent years as a system for heterologous expression of archael genes. However, the *P. pastoris* expression system is laborious and time consuming. For these reasons, *in vitro* expression of archaeal lipases of metagenomic origin was carried out.

Igni13, as well as Igni15 and Igni18, hyperthermophilic lipases from *I. hospitalis* KIN4/Iren, which

belongs to the *Crenarchaeota*, were cloned and kindly provided as pET-21a constructs from Pablo Pérez-García. These are derived from a metagenomic library through function-based screening, whereas Igni18 was extensively characterized (Kobus *et al.*, 2019). Igni18 has a temperature optimum of 90°C, which corresponds to the optimal growth temperature of *I. hospitali*s.

Both Igni13, Igni15 and Igni18 could be expressed *in vitro*, with all three, showing activity with pNP substrates (Figure 19). In addition to the activity, the successful expression of Igni18 was verified by Western Blot analysis (3.8.9) (Figure 18). The *in vitro* transcription was performed with the T7 RNAP and the translation with different cell extracts (3.10). With both the coupled and the linked system, the Igni lipases could be expressed. With the linked system, more reliable results could be achieved at 37°C. In contrast to Igni13 and Igni15, which were insufficiently detectable in the Western blot, Igni18 could be clearly detected.

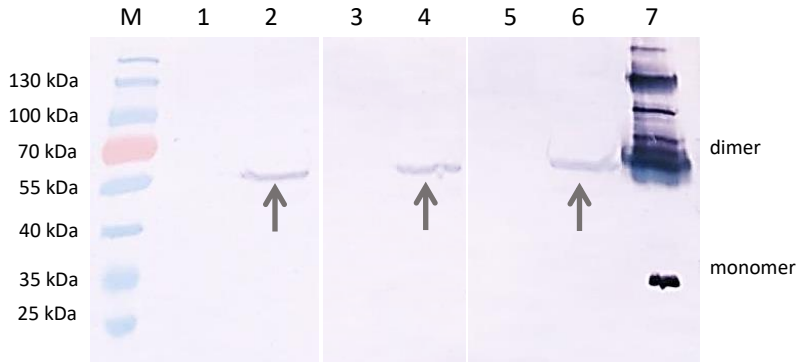

Figure 18. *In vitro* expression of His6-tagged Igni18 with cell extracts from different bacteria.

16 µl of each expression mixture were loaded onto a 12% SDS-gel followed by Western Blot analysis. His_6-specific antibody was used as primary antibody, the secondary antibody was conjugated with an alkaline phosphatase for detection with NBT and BCIP. M: Marker (Fermentas prestained protein molecular weightmarker #SM0671), 1-7: I*n vitro* expression of Igni18 with different bacterial extracts at 37°C. 2: *E. coli BL21* CodonPlus RIL, 1: Corresponding negativ control, 4: *Geobacillus* sp. GHH01, 3: Corr. negativ control, 6: *G. thermoleoverans*, 5: Corr. negativ control, 7: Igni18, heterologously expressed in *Pichia pastoris* and His_6-tag purified (3 µg).

Figure 18 shows the Western blot analysis of the linked *in vitro* expression of Igni18. *In vitro* translations showed the highest efficiency with all extracts at 37°C despite their partially thermophilic origin (data not shown). It can be seen, that the *in vitro* translations with *Geobacillus* sp. GHH01 extract (sample 4), as well as with *G. thermoleoverans* extract (sample 6) show a comparably intense band between 55 and 70 kDa as with *E. coli* BL21 CodonPlus extract (sample 2). The corresponding negative controls (sample 1, 3 and 5) show no signal. In addition to the prominent band between 55 and 70 kDa, the blot shows weak bands at about 100 and 130 kDa, and below 55 kDa. The control (sample 7) with *in vivo* expressed and purified Igni18 shows a complex band pattern. It was recently found out that Igni18 formes stable homodimers and homotrimers and has a half-life of 46h at 90°C (unpublished

data). The signal below 35 kDa shows the monomeric form of Igni18, the signal between 55 and 70 kDa results from the homodimer and one of the signals between 70 and 100 kDa shows the homotrimer of Igni18. Incubation at 95°C for only 10 minutes does not denature Igni18 due to its extreme thermostability, and tertiary and quaternary structures were only partially resolved. The amount of purified protein loaded on the gel was rather high, which, in combination with the different structural variants of the protein, results in the band pattern in sample 7. The prominent bands in samples 2, 4 and 6 thus show *in vitro* expressed Igni18 in its dimer form, with the weaker bands above being corresponding to the homotrimer. The Western blot not only shows that Igni18 can be expressed with different cell extracts *in vitro*, but also that it forms homodimers and -trimers.

4.2.3.1.1 Verification of the activity of the *in vitro* expressed archaeal lipases

Since the CFPS system is established for the use in function-based metagenomics, the *in vitro* expressed model enzymes were tested for their specific activity. To demonstrate the activity of Igni lipases, activity assays were carried out with 1 mM para-nitrophenyl palmitate (pNP-C16) in PB (pH 8.0) at 90°C (3.11.1). Hydrolysis of pNP-C16 by Igni lipases released the yellow pNP, which was measured photometrically. Since it was discovered that enzymes in the cell extract itself can hydrolyze the substrate, the *in vitro* mix was incubated at 70°C for 2h after translation. This resulted in the denaturation of most of the proteins from the cell extract, while the *in vitro* expressed Igni lipases remained intact. Since the turnover rate of Igni lipases is very low, the incubation time was set at 5h.

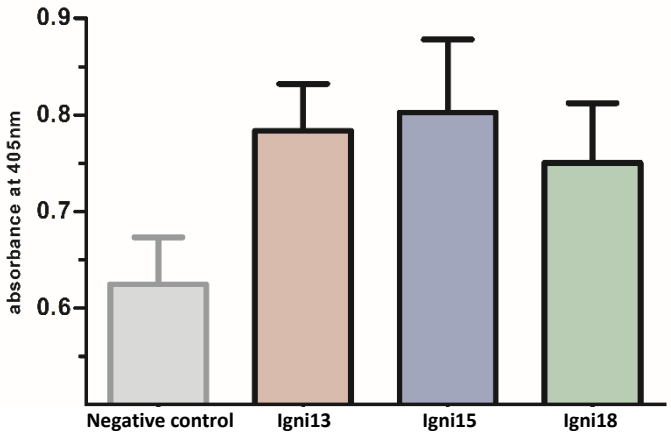

Figure 19. Results of the activity assay for *in vitro* expressed lipases using pNP-substrate.

The released pNP was measured at 405 nm in a microtiter plate spectrophotometer. The data are corrected by the buffer control. The graph shows the average measurement data from both biological and technical replicates. Error bars show the standard deviation from exemplary biological duplicates of two trial days. Gray bar: Negative control, which corresponds to an *in vitro* translation without mRNA as template. Colored bars: Results of the activity tests of *in vitro* expressed Igni lipases.

Figure 19 shows the result of the activity test with the cell-free expressed archaeal lipases Igni13, Igni15 and Igni18 (using T7 RNAP and *E. coli* BL21-CodonPlus (DE3)-RIL extract) with the substrate *p*NP-C16. Lipase activity was determined by the quantification of produced *p*NP, measured at 405 nm in a microtiter plate. Preliminary tests have shown that the absorbance is linearly related to the lipase concentration (data not shown) and thus directly attributable to the enzymatic hydrolysis of the pNP substrate. Both the substrate and the enzyme concentration were chosen so that autohydrolysis effects appear as low as possible, but the enzymatic substrate degradation is clearly detectable. All data were corrected by a buffer control composed of substrate and PB. It is striking that the negative control (grey bar), which corresponds to the *in vitro* translation without mRNA as a template, shows lipolytic activity. This is probably due to enzymes in the cell extract, which are temperature stable and show hydrolytic activity with long-chain *p*NP substrates. The lipolytic activity of the *in vitro* expressed Igni lipases was determined as the activity above the translation control. Numerous replications of the experiment showed that higher measured values of Igni13 (red bar), Igni15 (blue bar) and Igni18 (green bar) are clearly attributed to the *in vitro* expression of the respective Igni lipase compared to the control.

4.2.3.2 PET hydrolases

To test the applicability of CFPS for function-based screening of metagenomic libraries, additional, heat-stable, metagenomic-derived hydrolases were selected. Recently, a paper on “The Function and Global Distribution of Polyethylene Terephthalate (PET) Degrading Bacteria and Enzymes in Marine and Terrestrial Metagenomes” (Danso *et al.*, 2018) was published. Metagenomic-derived hydrolases, capable of degrading PET, have been expressed in *E. coli*, purified and screened for function. To simplify and shorten this procedure, cell-free expression experiments of the so-called PET hyrolases were carried out. Vector constructs (pET-21a, T7 promoter, N-terminal His_6-tag) of the two already characterized PET hydrolases PET2 and PET6 were kindly provided by Dr. Dominik Danso. PET2 was found by sequence-based screening of a marine metagenomics data set and PET6 is derived from the *Vibrio gazogenes* strain DSM-21264.

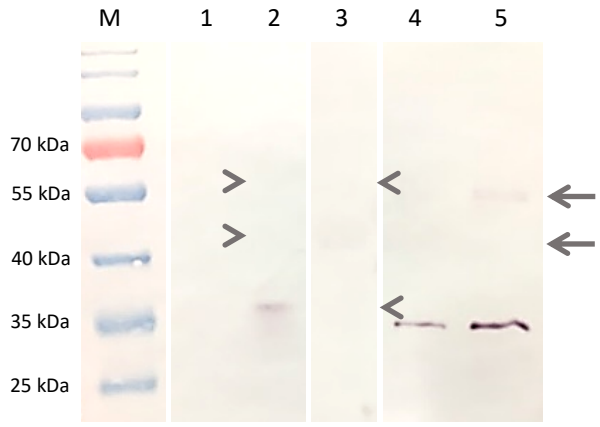

Figure 20. *In vitro* expression of His_6-tagged PET hydrolases PET2 and PET6.

16 µl of each expression mixture were loaded onto a 12% SDS-gel followed by Western Blot analysis. His_6-specific antibody was used as primary antibody, the secondary antibody was conjugated with an alkaline phosphatase for detection with NBT and BCIP. M: Marker (Fermentas prestained protein molecular weightmarker #SM0671), 1: Negative control (without template mRNA). 2: *In vitro* expressed PET2; 3: *In vitro* expressed PET6; 4: PET6 heterologously expressed in *E. coli* BL21 and His_6-tag purified (10 ng); 5: PET6, heterologously expressed in *E. coli* BL21 and His_6-tag purified (100 ng). The arrows mark very weak signals, which appear in sample 2 and 3, as well as the control sample 5. They may correspond to different quaternary structures of the PET hydrolases.

The two PET hydrolases PET2 and PET6 were successfully expressed *in vitro* (Figure 20). Also here, the linked method proved to be more promising in terms of expression efficiency. Both the T7 RNAP and the $RNAP_E$ were able to produce transcripts of the two PET hydrolases (Figure 12, data with T7 RNAP not shown). The *in vitro* translation could be demonstrated with extract of *E. coli* BL21 Codon Plus RIL as well as *Geobacillus* sp. GHH01. Figure 20 shows the Western blot analysis of PET2 and PET6, where heterologously expressed and purified PET6 in different concentrations was used as control. It could be shown that the detection limit of the analysis is below 10 ng protein (sample 4) and very weak signals of the *in vitro* expressed PET hydrolases (sample 2 and sample 3) show that the expression yield is extremely low. PET2 and PET6 have a molecular weight of about 30 kDa and PET6 (the control, sample 4 and 5) shows 3 visible bands (the weakest is marked with an arrow) in the gel. The sample with *in vitro* expressed PET2 shows a clear band at about 35 kDa and a very weak band marked with an arrow between 55 and 70 kDa. This may be the monomeric and dimeric form of the heat-stable protein. Sample 3 (PET2) shows 3 very weak band signals, which are also marked by arrows. These are similar in size to those of the control and are also due to stable tertiary structures that were maintained despite heat incubation.

4.2.3.2.1 Activity assay with *in vitro* expressed and immobilized PET hydrolases

Activity tests with the *in vitro* expressed lipases from *I. hospitalis* already showed difficulties in the use of *p*NP-substrates. The high background activity from cell extracts components when using *p*NP-substrates made it difficult to detect very small amounts (in the range of ng and µg) of *in vitro* expressed enzymes. This was also demonstrated by the use of *p*NP-octanoate as a substrate for the detection of PET hydrolases. Therefore, an alternative enzyme assay was performed with 4-methylumbelliferone (4-MU) octanoate as a substrate (0). Esters of 4-MU do not fluoresce unless cleaved to release the fluorophores and represent an extremely sensitive detection method for enzyme activities.

Both PET hydrolases showed a particularly high activity with short-chain *p*NP substrates but could also implement long-chain substrates (>C_{10}). PET2 and PET6 showed high thermostability and a temperature optimum of 55°C to 70°C (Danso *et al.*, 2018). Therefore, the activity tests of the *in vitro* expressed PET hydrolases were carried out at 60°C.

In order to minimize the background signals from the cell extract in the best possible way, an activity assay with the *in vitro* expressed enzymes coupled to Ni-NTA beads was developed (0). The *in vitro* expression micture was incubated with Protino® Ni-NTA beads (Machery Nagel). The additional His_6-tag enables the immobilization of the *in vitro* expressed PET hydrolases to the beads via metal affinity. Subsequent washing steps removed components of the expression mix and re-buffered the reaction mixture simultaneously (3.11.2.1). Instead of eluting the enzymes, the activity test was carried out directly with the immobilized enzymes by adding the substrate and incubation at 60°C. The supernatant was measured fluorometrically in a microtiter plate using 4-MU octanoate as substrate (3.11.2.2).

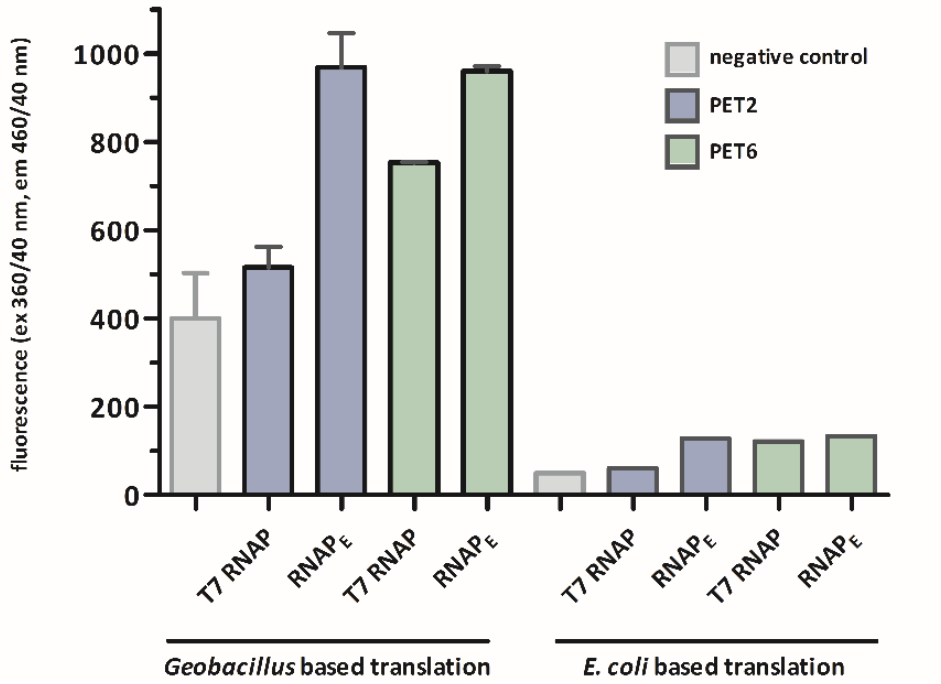

Figure 21. The success of the *in vitro* expression of PET2 and PET6 was analyzed by performing an activity assay with 4-MU octanoate as substrate.

The fluorescence was measured with an excitation at 360/40 nm and an emission at 460/40 nm in a microtiter plate spectrophotometer. The data are corrected with the buffer control. The graph shows the average measurement data from two biological replicates from different trial days (*Geobacillus* based translation experiments). Unfortunately, the *E. coli*-based translation is only represented by one value (average measurement data from three technical replicates). The error bars show the standard deviation. The grey bars show the data of the negative controls, which corresponds to *in vitro* translations without mRNA as template. The blue bars show the result of the activity tests of *in vitro* expressed PET2 and the green bars of *in vitro* expressed PET6. The left bars show the results of *Geobacillus* sp. GHH01 based translation and the right bars show those of translation with extract of *E. coli* BL21 CodonPlus. The samples also differ in the used RNAP for the *in vitro* transcription step (shown in the diagram).

To compare the new cell-free expression system consisting of the $RNAP_E$ for the transcription step and the cell extract of *Geobacillus* for the *in vitro* translation with the classical T7 RNAP/ *E. coli* system, linked approaches were performed (3.10.2). In the first step, the transcription of the two PET hydrolases PET2 and PET6 was performed in parallel with the T7 RNAP and the $RNAP_E$. The purified mRNA was then used in equal amounts for translations (37°C) with both *E. coli* BL21 CodonPlus extract and *Geobacillus* sp. GHH01 extract. After coupling the His_6-tagged PET hydrolases to Ni-NTA beads, the activity test was performed as described and the results are compared directly (Figure 21). Only one representative data set of the *E. coli*-based translations is shown, since biological replicates of this experiment were difficult to compare, as the background signals coming from parts of the cell extracts, bound to the beads, differ greatly. Nevertheless, the results of the individual trial days always showed the tendency as shown in Figure 21. Especially noticeable is the comparison between the *Geobacillus*-

based and the *E. coli*-based translation. The *in vitro* translation with *Geobacillus* extract results in a higher activity, which most likely correlates to a higher protein yield. Interestingly, translations with mRNA produced by $RNAP_E$ tend to provide better results. The error bars show that the negative control varies more than the expression of the PET hydrolases. The results clearly show that this alternative system for CFPS for some proteins demonstrates comparable or even better results than the classical approach. Despite the unequivocally detected activity using this method, of most in vitro expressed enzymes, only very few samples showed a Western blot signal. This indicates that the benefit of this detection method lies in the detection especially of low enzyme concentrations.

4.2.4 Photometric detection of *in vitro* expressed enzymes

One of the goals of cell-free expression of metagenomic-derived genes is to develop a rapid and unambiguous method of detecting the expressed enzymes. The idea is that this microtiter plate-based assay works without further purification steps and allows the photometric detection of positive hits. In this case, the *in vitro* expressed enzyme converts a specific substrate, which leads to a color change, which is visible optically and measurable on basis of the absorbance spectrum. For this purpose, the linked *in vitro* expression (3.10.2) of a beta-lactamase was used as test protein. Even the smallest amounts of enzyme can be detected via an activity test with the substrate nitrocefin.

After performing the *in vitro* translation, the mixture was used directly for the activity test. The $RNAP_E$ was used for the transcription and the translation was carried out both with extract of *E. coli* BL21 CodonPlus and *Geobacillus* sp. GHH01. The enzymatic digestion of nitrocefin was monitored over time by measuring the spectrum from 300 nm to 700 nm (3.11.3). The detection of the ß-lactamase activity with nitrocefin is possible because of a shift of the absorbance maximum from UV to visible light (~390 nm to ~500 nm) when it is hydrolized. This can also be seen as a color shift from yellow to red.

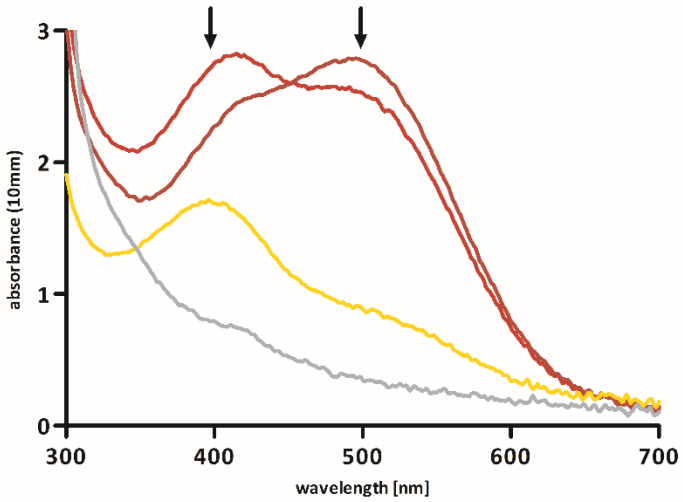

Figure 22. Spectrum of the activity test mixture consisting of nitrocefin, PB and *in vitro* translation mixture.
The spectrum was measured after incubation for 10 minutes at 30°C. Different samples were measured: A control consisting of PB and *in vitro* translation mixture (grey), a control consisting of PB, nitrocefin and *in vitro* translation mixture without mRNA as template (yellow) and 2 biological replicates (red and red-brown) consisting of PB, nitrocefin and *in vitro* translation mixture with ß-lactamase mRNA as template. The black arrows mark the absorbance maximum of the intact nitrocefin (390 nm) and the hydrolyzed product (~500 nm). The measurements were carried out with the NanoPhotometer (Implen) on a 2 µl scale.

The activity assay with the *in vitro* expressed lipases using pNP substrates indicated that enzymatic background activities from the cell extract are a major problem after incubation. In addition, absorbance spectrum of the cell extracts interfere with the absorbance of the desired end product and can thus lead to false-positive readings. Therefore, to investigate the specific of the *in vitro* expressed β-lactamase, first, not only one wavelength was measured, but a section of the entire spectrum of the sample. In this way, the exact maximum absorbance of the hydrolyzed nitrocefin could be determined.

Figure 22 shows the spectrum from 300 nm to 700 nm of the controls (gray and yellow) and two biological replicates (red and red-brown). The grey spectrum shows that the *in vitro* translation mixture itself has not always a noticeable absorbance, which could falsify the measurements at 390 nm (intact nitrocefin) and 500 nm (hydrolyzed product). The control of *in vitro* expression itself (without mRNA as template) shows a prominent peak at 390 nm and a low peak at about 500 nm in the spectrum. This indicates a minimal degradation of the substrate. The spectra of the activity tests with replicates of *in vitro* expressed β-lactamase clearly show 2 peaks. One peak is attributable to the not yet hydrolyzed substrate (390 nm) and one peak to the hydrolyzed substrate (500 nm). The additional peak is thus clearly based on the enzymatic activity of the β-lactamase.

Furthermore, the activity test was carried out microtiter plate-based and only the absorbance at 390 nm and 500 nm was measured.

Figure 23. The success of the *in vitro* expression of the ß-lactamase was analyzed by performing an activity assay with nitrocefin.

The activity of ß-lactamase can be visualized as a shift from yellow to red. The yellow sample (left) corresponds to the buffer control. The red samples are the activity tests with the *in vitro* expression mixtures. In the middle is the test of the negative control (*in vitro* translation without mRNA as template), on the right is the test with *in vitro* expressed ß-lactamase.

Figure 23 shows the result of the activity test of the *in vitro* expressed ß-lactamase with nitrocefin, which hydrolyzation can be detected as a color shift from yellow to red. A difference between the negative control (*in vitro* translation without mRNA) and the expression of ß-lactamase could be detected optically and could be verified by measuring the absorbance at 500 nm (data not shown), whereas again the problem of the high background signals becomes clear. Transcription of the ß-lactamase gene was performed with $RNAP_E$ and *in vitro* translation was successfully demonstrated with extract from *Geobacillus* sp. GHH01. A difficulty is the high background activity of the cell extract combined with the low expression rate per se.

4.3 *In vitro* protein expression in compartments

Finally, the -in this work- designed CFPS system was successfully tested for compatibility with advanced *in vitro* compartmentalization. By coupling CFPS packed into polymersomes and fluorescence-activated cell sorting (FACS) techniques, a new *in vitro*-based technique may be designed to overcome the low throughput rate of classical function-based metagenomic screening.

In cooperation with the research group of Prof. Dr. Schwaneberg at the RWTH Aachen, a system for CFPS in polymersomes was developed. For this purpose, the CFPS system established in this work was used both with extract of *E. coli* and of *Geobacillus* sp. GHH01. GFP and the metagnomic-derived cellulase CelA2 were used as model proteins.
Coupled *in vitro* expression mixtures (3.10.1) were encapsulated in polymersomes consisting of PBD-b-PEO (3.12). Polymersomes are assembled from amphiphilic diblock copolymers and are similar in structure to liposomes (Figure 7). In comparison, they show enhanced mechanical stability and can therefore easily be collected after production and incubated for many hours at different temperatures. Because of the lower permeability compared to liposomes, all components from the *in vitro* expression mixture stay in the inner phase of the polymersomes.

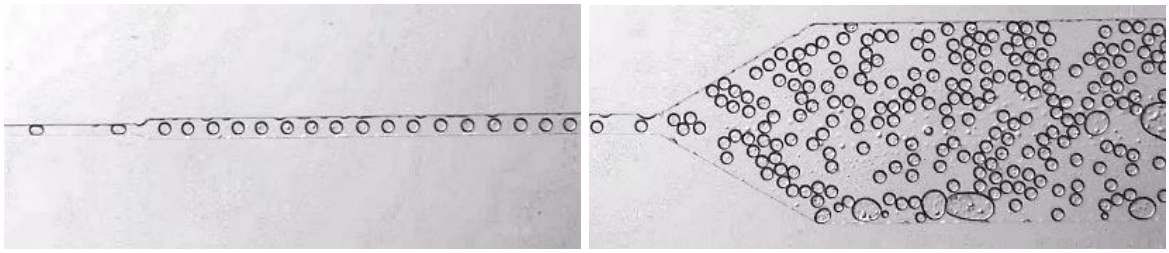

Figure 24. Microscopic picture (light field) of the final part in the polymersome formation.
The polymersomes are composed of a bilayer shell of diblock amphiphilic copolymers and an aqueous core, in this case the *in vitro* protein expression mixture. The microscopic picture shows the arrival of stable polymersomes after forming through the PDMS device coated with the diblock copolymer. The exact procedure is shown schematically in Figure 7. The polymers give the polymersomes their typical dark ring and are therefore easy to detect.

In cooperation with Volkan Besirlioglu and Richard Meurer, *in vitro* expression mixtures were successfully encapsulated and polymersomes were produced. Thus, the compatibility of the two systems could be proven. In addition, CFPS in polymersomes was performed with 2 different extracts and the results were analyzed by fluorescence microscopy. It turned out that the production of the polymersomes posed the greatest challenge, since the system is very sensitive to high flow rates and the narrow channels are easily blocked by the diblock polymer coating. If the flow rates of the individual phases are too low, no polymers are formed, while too high flow rates cause the double emulsions to burst before the polymers can form correctly. With correct fine tuning, large quantities

of robust polymersomes can be produced, which can be easily converted into Eppendorf-tubes and incubated for days. Even after microscopic analysis, the polymersomes remained intact.

4.3.1 GFP

GFP was used as a control for successful protein expression in the polymersomes. For this purpose, an expression mixture for a coupled expression was encapsulated with plasmid DNA as template, T7 RNAP and *E. coli* BL21 CodonPlus extract and afterwards the polymersomes were incubated for 4 h at 30°C. After overnight storage at 4°C for maturation of the GFP, the fluorescence was analyzed microscopically.

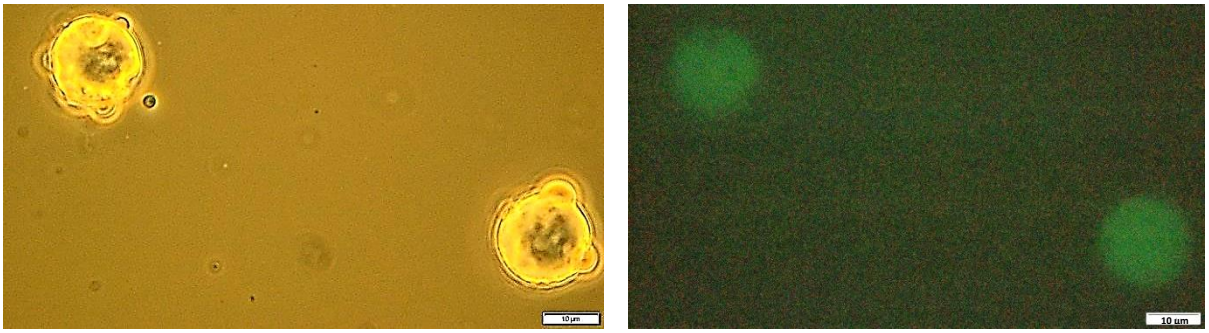

Figure 25. Microscopic analysis of the polymersomes after expression of GFP inside.

Left: Light field image of two polymersomes. The polymer can be seen as a dark ring.

Right: Microscopic image of the fluorescence of 2 polymersomes in which the *in vitro* expression of GFP has taken place. Ex.: 457-487 nm; em.: 502-538 nm. The scale bar is 10 µm.

The microscopic images show polymersomes with a diameter of about 15-20 µm, which is comparatively small. In the literature, polymersomes are described with a diameter of up to 100 µm (Bartenstein *et al.*, 2016). Nonetheless, GFP could be successfully expressed in the polymersomes, which is clearly indicated by the fluorescence images. The green fluorescence is uniformly distributed in the inner phase of the polymersomes. Three quarters of all polymersomes fluoresced.

Unfortunately, no negative controls could be performed without plasmid DNA because many device channels were blocked by the polymer during the process of polymersome formation. As a result, some experiments could not be performed. Nevertheless, the fluorescence of the polymersomes can only be archieved by a successful *in vitro* expression of the GFP, since some polymersomes did not show any fluorescence at all. The bilayer with the polymer does not fluoresce itself, which can also be seen from the overlay of the fluorescence channel and the highfield (not shown). The fluorescence is clearly localized in the internal phase of the polymersome where the *in vitro* protein expression takes place.

4.3.2 CelA2

After successful performance of CFPS in polymersomes, the metagenomic-derived cellulase CelA2 (H288F variant) was tested for expression. CelA2 is derived from a metagenomic library (Pottkämper *et al.*, 2009), which was constructed from a biogas plant and changed through direct evolution in order to increase its specific activity and the resistance to choline chloride (Ilmberger *et al.*, 2012; Lehmann *et al.*, 2012; Martinez & Schwaneberg, 2013; Körfer *et al.*, 2016). The direct evolution resulted in an 8-fold increase in activity of the variant compared to the wildtype CelA2 from the metagenomic library.

For the expression of the CelA2 in polymersomes, extract of *Geobacillus* sp. GHH01 whas used. For subsequent detection of the expressed enzyme, the substrate was added directly to the expression mixture prior to incapsulation. Instead of using 4-MU substrates, resorufin-ß-D-cellobioside was used. The presence of endogenous fluorescence from cellular components and proteins in the cell extract make it difficult to detect the degradation of 4-MU substrates, as they have similar absorption and fluorescence wavelength. The substrate resorufin-ß-D-cellobioside has a long-wavelenght fluorecence and can be easily detected by a red signal. The cellulase activity of the encapsulated CelA2 could thus be directly detected microscopically by the released red fluorescent fluorophore resorufin (Ex: 571 nm, Em: 585 nm) (Coleman *et al.*, 2007).

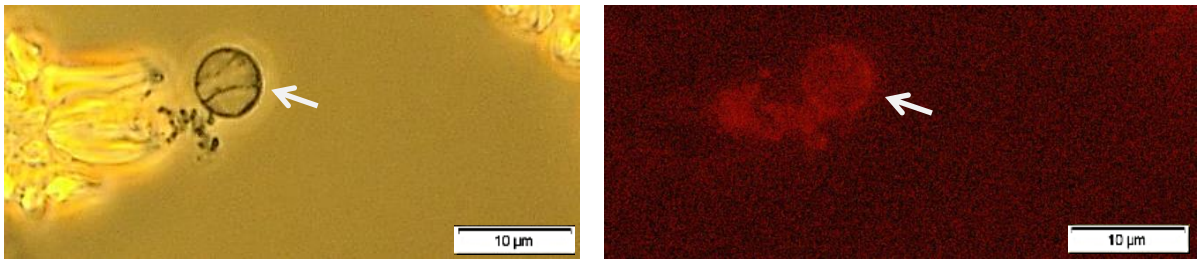

Figure 26. Microscopic analysis of the polymersomes after expression of CelA2 and incubation with resorufin-ß-D-cellobioside.

Left: Light field image of a polymersome with expression of $CelA_2$ in the inner phase.

Right: Microscopic image of the fluorescence of 2 polymeromes in which the *in vitro* expression of CelA2 has degraded the substrate resorufin-ß-D-cellobioside and released the red fluorescent resorufin. The scale bar is 10 µm.

The microscopic images show one of many polymersome with a diameter of about 7-8 µm, which is much smaller than the ones expressing GFP. Nonetheless, the polymersome shows the red fluorescence from the released resorufin, which is due to the presence of the expressed CelA2. The red fluorescence is mainly located at the outer phase of the polymersomes which leads to the assumption that it binds to the polymer, still, not all polymersomes fluoresced. Unfortunately, no negative controls could be performed without plasmid DNA because of the above-mentioned reason. But the fluorescence of the polymersomes indicates a successful *in vitro* expression of the active CelA2,

since some polymersomes did not show any red fluorescence. In addition to the intact polymersomes, microscopic images show fragments of burst polymersomes and aggregated polymer (dark and crystalline structure). This also partly shows a red fluorescence. This might be due to the binding of resorufin, since the expression mixture was released into the buffer by the burst polymersomes and the expression of CelA2 also could have taken place there. It is also possible that during storage or transfer to the microscope slides, polymersomes containing the released resorufin may have burst and then the resorufin attached to the polymer artifacts in the buffer.

5 Discussion

5.1 Recombinant RNAP from *T. thermophilus*

As a first step to overcome problems associated with insufficient transcription rates of metagenomic DNA, the development of a robust *in vitro* transcription system was focused. For this purpose, RNAPs of thermophilic origin were recombinantly synthesized and subsequently assays for transcription of metagenomic DNA were carried out.

5.1.1 Cloning, heterologous expression and *in vitro* reconstitution

In this work, the RNAP from the extremely thermophilic eubacterium *T. thermophilus* was cloned and heterologously expressed. The individual subunits were all successfully cloned into pET-vectors, expressing the protein either with or without His_6-tag. The best expression conditions for each subunit were found, with variants without additional His_6-tag showing greater expression success. Soluble subunits were roughly purified using heat denaturation, whereas inclusion bodies forming ω-subunit was recovered using urea-treatment. Classical reconstitution of the core enzyme ($RNAP_C$) composed of α2ββ'ω subunits, was performed following a slightly modified protocol described by (Kuznedelov & Severinov, 2009) by mixing the soluble subunits under denaturing conditions and a smooth removement of the denaturant through gradient dialysis.

Semi-native gel electrophoresis suggested the successful reconstitution of the $RNAP_c$ at 4°C as well as room temperature and at 60°C (4.1.1.3). The correct arrangement of the subunits, folding and formation of the active center, however, can only be concluded by crystal structure analysis and activity tests (Vassylyev *et al.*, 2002; Vassylyeva *et al.*, 2002).

5.1.2 Activity

In vitro transcriptions were performed with both the core enzyme and the holo enzyme (generated by incubation of the $RNAP_C$ with the σ^{70}-factor) of the *T. thermophilus* RNAP. For this, genomic DNA of various bacteria, i.a. *T. thermophilus* HB27 itself were used as templates. However, no transcriptional activity could be detected in agarose gel analyses. Transcriptional assays were performed at 30°C to 70°C, template amount and incubation times varied, and the assay was controlled by performing reactions with the commercial *E. coli* RNAP and reconstituted RNAP from *Geobacillus* sp. GHH01 (Kinfu *et al.*, 2017). This suggests that the reconstitution was flawed, the holoenzyme was probably not formed correctly, and thus, it might be that promoter recognition was impaired. Urea-treatment of the ω-subunit might have damaged the protein and inhibited its function, which is the promotion of

the RNAP assembly, comparable to a chaperone (David Marcey & Nathan Silva, 2006; Gunnelius *et al.*, 2014).
In addition, the contamination with remaining *E. coli* proteins, may hamper the reconstitution efficiency.

Furthermore, it is possible that the magnesium concentration in the reconstitution buffer (10 mM) was insufficient, although this concentration has been described for the successful reconstitution of bacterial RNAP in the literature (Markov *et al.*, 1999). The crystal structure of the *Thermus* RNAP showed that numerous Mg^{2+}-ions bind to the holo-enzyme and form a coat on the surface of the protein. Vassylyev *et al.* assume that metal ions play an important role in the binding of the DNA, as known for example in DNA polymerases (Vassylyev *et al.*, 2002). Crystal structure analysis of the RNAP from *T. aquaticus* revealed zinc binding motifs on the β'-subunit which shows that Zn^{2+} may play a role in enzyme assembly (Robert Schenck; Markov *et al.*, 1999). However, a concentration of 10 µM in the reconstitution buffer, as used in this work, should be completely sufficient (Kuznedelov & Severinov, 2009).

Transcription experiments were carried out without additive transcription factors. It is known that the bacterial transcription machinery consists of a multitude of regulatory proteins as described in 1.2.2.1. In the absence of additional regulatory components, the transcription level depends mostly on the strength of the promoter and the RNAP concentration. For example, transcriptional activity could be enhanced by the addition of transcription factors such as GreA and GreB which can suppress RNAP pausing and arrest and stimulate RNAP activity (Hogan *et al.*, 2002). *In vitro* transcriptions with the RNAP of *Geobacillus* sp. GHH01 showed that the addition of GreA significantly improves transcriptional activity (Kinfu, 2018). Also, the supplementation with the multifunctional transcription elongation factor NusA, as well as the transcription-repair coupling factor Mfd could have a positive effect in *in vitro* transcription reactions (Borukhov *et al.*, 2005). To make a clear statement about the *in vitro* transcription ability of the *T. thermophilus* RNAP, the implementation of reactions with different, additional transcription factors would be crucial.

While performing these experiments, unpublished RNA sequencing data showed that another bacterial RNAP had a tremendous error rate and an extreme promoter unspecificity. It was found that, in addition to the genes to be transcribed, *E. coli* genes which were linked to the heterologously expressed RNAP were transcribed as well. This unspecificity of the bacterial RNAP allows recognition of completely new promoters in metagenomic samples, on the contrary it is not possible to use incorrect transcripts for the *in vitro* expression for functional metagenomics, since a translation is thus impossible. This leads to the conclusion that bacterial RNAPs may be less well suited for the transcription of metagenomic genes, and viral enzymes might be the better choice.

5.2 Recombinant viral RNAP from the elephant metagenome

Since the bacterial-based *in vitro* transcription system showed no mRNA yield, it was searched for a viral alternative.

Viral RNAPs are characterized by the fact that they consist of only one subunit, which significantly reduces the cloning and expression effort. This also eliminates the critical step of reconstituting the core/holo enzyme from individual subunits. Unlike bacterial and eukaryotic RNAPs, viral RNAPs do not require additional transcription factors for efficient mRNA production (1.2.1.1). In addition, they are characterized by their extremely high promoter specificity, consequencing, that they recognize less metagenomic promoters, but only transcribe the desired gene without any mRNA byproducts. These features make viral RNAPs excellent tools for *in vitro* transcriptions.

5.2.1 Metagenome search, gene synthesis and expression

For this purpose, a large number of metagenomic datasets of the working group were screened for potential new RNAPs of viral origin. These include metagenomes from freshwater and seawater samples, biogas plants, fish tanks, marine hydrothermal vents, cow rumen and feces from elephants from the zoo "Hagenbeck's Tierpark" in Hamburg. Bioinformatic analyses of datasets were performed by Dr. Simon Güllert. Results are not described in detail in this work. From a pool of interesting RNAP candidates, which have low sequence similarity to commercially available and routinely used RNAPs such as SP3, SP6 and T7, 4 genes were synthesized by Eurofins Genomics, subsequently cloned into pET-vectors and heterologously expressed. Using *in vitro* transcription experiments, a RNAP from the microbial community from the elephant feced ($RNAP_E$) clearly showed potential for the synthesis of metagenomic RNA. The $RNAP_E$ shows only 29% identity in amino acid sequence compared to T7 RNAP, which is relatively low. The $RNAP_E$ could easily be heterologously expressed with His_6-tag in *E. coli* BL21 and purified using immobilized metal chelate affinity chromatography (IMAC) with yields up to 200 mg/l cell culture (4.1.2.3).

5.2.2 Activity

5.2.2.1 Temperature profile

In terms of activity, the $RNAP_E$ shows similarities to the commercial T7 RNAP (data not shown). It prefers a temperature range of about 30°C to 40°C (4.1.2.4.2). This can be explained by the fact that viruses optimize their proteins to the respective host temperature, in the case of $RNAP_E$, to the microbial community of the intestinal tract of an elephant (4.1.2.2). With a temperature optimum of 35°C the RNAP meets the demands of classical *in vitro* transcription experiments. For high temperature

applicability, it would be conceivable to genetically engineer the $RNAP_E$ improve its thermostability. For example, single amino acid substitutions are sufficient to create a thermostable variant like the thermo T7 RNAP from "Toyobo" with a given optimum reaction temperature of 50°C (http://www.toyobo-global.com). This polymerase was additionally used for *in vitro* transcription experiments and showed activity in own experiments up to 50°C with an optimum at 40 to 42°C (Figure 30).

A promising alternative, to improve *in vitro* transcriptions at elevated temperatures, would be the use of an RNAP from a naturally thermophile bacteriophage. Potential sources of thermophilic viral enzymes have been described, for example, by Liu *et al.*, who isolated two new bacteriophages of thermophilic bacteria from deep-sea hydrothermal fields (Liu *et al.*, 2006). In addition, Marks and Hamilton characterized a thermophilic bacteriophage of *Geobacillus kaustophilus* (Marks & Hamilton, 2014).

5.2.2.2 General activity

A broad spectrum of transcripts can be generated using the $RNAP_E$ (4.1.2.4.3). The *in vitro* transcription of sfGFP and metagenomic-derived genes coding for already chacaterized hydrolases, such as the bacterial cellulase CelA2, the bacterial lipase LipS, bacterial PET hydrolases, archaeal lipases, as well as the complete violacein operon could be demonstrated successfully. With up to 23.1 µg mRNA from a 100 µl *in vitro* transcription, the $RNAP_E$ represents a serious alternative to the typical T7 system, which can also be produced with low costs in high quantities and good quality.
Interestingly, the $RNAP_E$ also has the ability to transcribe uncloned genomic DNA, which plays an important role in the applicability to metagenomic screening. Also, the transcription of diverse fosmid clones of a metagenomic library was demonstrated using the new $RNAP_E$.

Furthermore, it was shown that the transcripts generated by the $RNAP_E$ can also be translated into functional proteins (4.2.3.2.1). This was not possible, for example, with bacterial transcription systems (*Geobacillus* sp. GHH01 and *T. thermophilus*) because of the inadequate amounts of RNA and the probably high error rate of the RNAP. Surprisingly, activity tests with *in vitro* expressed PET hydrolases suggest that transcripts synthesized by the $RNAP_E$ tend to be more translatable than transcripts generated by the T7 RNAP. At least, this manifests itself in different levels of substrate turnover, which of course is not directly attributable to a higher protein concentration.
One reason could be the quality of the mRNA, possibly due to a more correct transcription by the $RNAP_E$. Two major limitations of the T7 RNAP in *in vitro* experiments are common knowledge, namely poor transcription rate of G-rich initial sequences (Dunn & Studier, 1983) and the formation of multiple undesired mRNA byproducts of different length. Least is caused by premature terminations at the 3´-

end or the addition of extra nucleotides at the 5´- and 3´-end of the transcript due to self-priming (Pleiss *et al.*, 1998 ; Helm *et al.*, 1999; Gholamalipour *et al.*, 2018). All these inaccuracies of the T7 RNAP result in a decreased number of correctly transcribed, full-length mRNA products (Wichłacz *et al.*, 2004). However, the length profile of synthesized mRNA provided by electrophoresis is too imprecise to demonstrate this phenomenon. The $RNAP_E$ may synthesize more correct transcripts than the T7 polymerase, which results in higher translation yields despite the use of the same amout of mRNA. Precise sequences, however, can only be determined by RNA-sequencing.

It should be noted that the experiments were performed with plasmids containing the T7 promoter. Extensive promoter studies might be used to identify the appropriate promoter sequence, which may significantly increase the transcription rate of the $RNAP_E$. Besides the fact, that the $RNAP_E$ is capable of recognizing transcription initiation signals in uncloned, genomic DNA, or clones from a fosmid library, a T7-like promoter is still needed for high transcription levels. Lussier *et al.* proposed the use of the bifunctional cosmid vector pFX583 allowing T7 RNAP-directed transcription for construction and functional screening of a meganomic library (Lussier *et al.*, 2011). pFX583 features a T7 promoter and a *Lambda* phage *cos* sequence that allows its use for cloning of large DNA fragments as given in a metagenome. Combined with *E. coli* and *Streptomyces lividans* strains that inducibly produce the T7 RNAP, Lussier *et al.* constructed a metagenomic cosmid library with pFX583 using extracted DNA from the biomass of an enriched fed-batch reactor. Function-based screening identified 17 positive hits with lipolytic activity from approximately 2,000 clones, which corresponds to a very high hit rate. The application of the T7-based transcription using the pFX583 cosmid has the potential to increase the transcription efficiency of metagenomic genes, which makes the use of phage-based RNAPs for function-based metagenomics very attractive. The availability of phage promoter-containing fosmids/cosmids (CopyControl™ Fosmid Library Production Kit with pCC1FOS™ Vector) eliminates the need for additional elaborate cloning steps than those common for metagenomic libraries and makes *in vitro* transcriptions with the $RNAP_E$ for function-based metagenomics especially interesting.

Another promising approach to enhance the expression rate of metagenomic genes i.a. by preventing premature transcription termination, through the combination of viral components with modified heterologous expression hosts, was successfully established by Terrón-González *et al*. (Terrón-González *et al.*, 2013). The use of a genetically modified pCC1FOS fosmid vector combined with an *E. coli* EPI300-T1 derivative strain expressing the T7 RNAP in low concentrations resulted in an enhanced expression rate of metagenomic genes. Through the incorporation of the T7 RNAP, additionally to the bacterial RNAP of the host strain, premature transcription termination could be prevented, because the phage RNAP is insensitive to most metagenomic termination signals. With this viral/bacterial expression system, 54,000 clones (2 GB in total) representing the metagenome of a coastal soil contaminated with oil spill, was screened for carbenicillin resistance. Compared to the commonly used

metagenomic approach based on the *E. coli* expression system, this enhanced system resulted in a 6-fold increase in the number of carbenicillin resistant clones.
In addition, the use of the modified pCC1FOS fosmid vector would allow the subsequent sorting of the fosmid clones through FACS by SIGEX technology (Figure 5), since successful transcription results in the co-expression of the reporter protein GFP.

These results reinforce the idea of using viral transcription components such as the $RNAP_E$ discovered in this work to enhance the gene detection frequency in function-based metagenome screening.

The use of cell extracts (4.2.2) from strains, which are additionally equipped with the plasmid coding for the $RNAP_E$, in coupled *in vitro* expression reactions with metagenomic fosmid clones, could increase the functional analysis potential in metagenomics. This would be a concept for future experiments.

5.3 Cell-free protein synthesis

CFPS has not only emerged as a powerful research tool for understanding transcription and translation reactions, as well as functional and structural of proteins. This technology also makes it possible to produce proteins that were previously difficult or impossible to produce in a short time with common expression hosts.

Due to the increasing variety of available CFPS systems (1.2.3) and the advantages shown in 1.2.4 over conventional *in vivo* protein expression, the applicability of CFPS for functional metagenomicswas examined. In addition to the successful *in vitro* transcription of metagenomic-derived genes, translation experiments were carried out. It was possible to show that the decoupling of transcription and translation tends to lead to higher protein yields (Figure 15). This can be explained by the fact that the quantity and quality of the mRNA is considerably increased by the purification step. Another reason is the unequal reaction rate between the T7 RNAP and the bacterial translation machinery in coupled reactions. In bacterial cells, processes of transcription and translation are tightly timed. While the RNAP synthesizes mRNA, the ribosomes initiate the translation of the nascent mRNA chain, in *E. coli*, for example, with a rate of ~20 amino acids per second (Bremer & Dennis, 2008). The tight coupling of transcription and translation prevents the formation of secondary structures, which can complicate or even block translation. In addition, the mRNA is protected from endonuclease degradation by the ribosomes. Because the T7 RNAP synthesizes mRNA about 4 to 5 times faster (~230 nucleotides/s) than the wild-type *E. coli* RNAP at 37°C (~50-60 nucleotides/s), the bacterial translation apparatus fails in its reaction speed, further complicated by the formation of secondary mRNA structures (Golomb & Chamberlin, 1974; Iskakova, 2005).

Furthermore, by decoupling both steps, optimal reaction conditions for the $RNAP_E$ for *in vitro* transcription experiments can be determined, and *in vitro* translation systems based on different cell extracts can be better compared.

To investigate the applicability of *in vitro* expression technologies for function-based metagenomics, model experiments were successfully carried out with already characterized enzymes of metagenomic origin as target proteins. Two exciting, heat-tolerant hydrolases were successfully expressed *in vitro*, including metagenomic-derived, archaeal lipases (Kobus *et al.*, 2019) and PET-degrading hydrolases (Danso *et al.*, 2018).

Alternative bacterial strains have been found for the preparation of cell extracts for *in vitro* translations (4.2.2). Finally, techniques have been developed that allow the functional screening of *in vitro* expressed enzymes microtiter-plate scale without the need for time-consuming purification steps (4.2.3.2.1).

5.3.1 Preparation and efficiency of different cell extracts

A crucial component for the CFPS system was the preparation of cell extract, also called "S30 extract". In principle, cell extracts can be obtained from all culturable cell types. The most widely used system for expression of proteins without posttranslational modifications is based on *E. coli* extract due to its simplicity and well-known molecular machinery (Carlson *et al.*, 2012). Alternative systems based on cell extract of prokaryotic origin, such as *Vibrio natriegens* or *Rhodococcus erythropolis*, are rare to find (Nevondo, 2016; Des Soye *et al.*, 2018). To date, mainly extracts of *E. coli*, yeast, wheat germ, rabbit reticulocytes and insect cells are used (Erickson & Blobel, 1983; Jackson & Hunt, 1983; Spirin *et al.*, 1988; Richter, 2014).

To enable functional screening especially for thermostable enzymes, alternative cell extracts from thermophilic and hyper-thermophilic bacteria were prepared. These were expected to be more resilient than commonly prepared extracts and should enable the *in vitro* translation at elevated temperatures, thus resulting in higher amounts of thermoresistent proteins. Extracts from 8 different bacterial strains were prepared and used for *in vitro* translation experiments, including *E. coli* MRE600 and CodonPlus RIL, *P. antarctica*, *B. subtilis*, *Geobacillus* sp. GHH01, *G. thermoleoverans*, *C. sambhunathii* and *T. thermophilus*. For example, the psychrophilic bacterium *P. antartica* has been used for the preparation of cell extracts to express, for example, cold-adapted enzymes at low reaction temperatures.

5.3.1.1 Preparation

The preparation of the cell extracts revealed that the time of cell harvest is a decisive factor for the subsequent *in vitro* expression rate. Unpublished data show that even with an OD_{600} shift of 0.1, the efficiency of the cell extract is significantly altered. In contrast to the preparation of the extract of *E. coli*, the preparation of the *Bacillus* extracts was more difficult and time consuming, since its generation time is much higher, lower cell densities were reached in the mid-exponential phase, the cell pellet was much more unstable and the cell disruption of Gram-positive bacteria is more difficult due to a protective cell wall. It turned out that disruption with a French Press is the most effective method. It is noticeable that the extracts differ significantly in their protein concentration, which is expressed in their colour (Figure 16). The most active *E. coli* extract was prepared from the strain CodonPlus RIL and contained more than 50 mg/ml protein.

The quality of the cell disruption is critical for the activity of the cell extracts, but the protein concentration itself is not a decisive factor for the *in vitro* translation efficiency. Finally, it could be demonstrated that the extracts of the *Geobacilli*, with a protein concentration of 16 to 20 mg/ml, produce comparable and even higher yields of *in vitro* translated protein than extracts of *E. coli* (Figure 17).

5.3.1.2 *In vitro* translation

In direct comparison of both *E. coli* strains, a higher amount of protein could be translated with extract of *E. coli* CodonPlus RIL than with extract of *E. coli* MRE600. The strain MRE600 lacks RNase I activity, because of a single mutation in the RNase I-encoding gene *rna*, which leads to a premature stop codon in its open reading frame (Kurylo *et al.*, 2016). *E. coli* CodonPlus RIL contains extra copies of the tRNA genes *argU*, *ileY* and *leuW*. These genes encode tRNAs that are responsible for the recognition of the codons AGA and AGG for arginine, the codon AUA for isoleucin and the codon CUA for leucin. The CodonPlus RIL strain is designed for translation of proteins with AT-rich mRNAs (Agilent). A similar genetic modification of the alternative strains could thus further influence *in vitro* translations positively.

Within the scope of this work, CFPS with extracts of *Geobacillus* species and *P. antartica* could be successfully shown for the first time (Figure 17). As already described in Figure 17 coupled *in vitro* protein expression with extract from *P. antarctica* was detectable at both 22°C and 30°C. Thus, *in vitro* translation of proteins from psychrophilic organisms may be more successful at lower temperatures than at standard temperatures of 30-37°C. This should be further investigated in the future. *In vitro*

expressions with cell extracts from *Bacillus subtilis*, *C. sambhunathii* and *T. thermophilus* HB27 were unfortunately without success (Table 15).

In vitro translations performed at high temperatures led to no positive results, when using *Geobacillus* or *T. thermophilus* extract. This may be due to the fact that the mRNA was denatured before it could be translated. Many proteins from the cell extract precipitated on incubation above 45°C, despite their thermophilic origin. This clearly shows that CFPS is just a very simplified form of the transcriptional and translational processes in the cell and that enzymes from a thermophilic organism are not necessarily thermostable or more active at elevated temperatures *in vitro*. For example, the protein ratios and salt concentration are different, as well as the distribution and organization of cell components.

A coupled *in vitro* expression with viral RNAP at high temperatures was also not possible yet, since the $RNAP_E$ show no transcriptional activity at temperature above 45°C (Figure 11) as discussed in 4.1.2.4.2.

Even though completely new *in vitro* translation systems were established and especially metagenomic-derived enzymes could be functionally synthesized in a very short time, there were crucial difficulties with the reproducibility. Despite compliance with published manufacturing protocols for protein extract preparation (Kwon & Jewett, 2015), partially low-grade or non-functional extract was generated. And although working aliquots of extracts were prepared and thawed only once for *in vitro* translations, translation efficiency varied from experiment to experiment using the same stock of extract. One reason for this could be the thawing process, in which important components of the extract can be destroyed. Experience has shown that cell extracts are sensitive to rough mixing such as vortexing, which, in some experiments, may explain the problems with the reproducibility of *in vitro* translation results. However, it remained unclear which parameters significantly influenced the quality of the cell extracts. Nevertheless, it was astonishing that the extracts were still active after storage at -70°C for more than one year.

Subsequent sequencing analyzes of the *Geobacillus sp.* GHH01 strain in the working group showed that contamination with other thermophilic *Bacillus* strains appeared in some cultures. Unfortunately it could not be proven where this contamination came from and when it was introduced, but this could be another factor that might affect the reproducibility of ivTT results. However, it could be shown that a variety of factors are essential for the translation efficiency of extracts, e.g. the cell density, and that cell extracts from only a few organisms are suitable for *in vitro* translations at all. When cultivating extremophilic organisms –in absence of antibiotics- in the future, it is advisable to carry out regular control experiments, as it is standard in industry as part of a quality control. However, it is not clear if and how a contamination could affect the efficiency of the cell extracts.

To significantly increase the efficiency of some extracts in their use for CFPS and to avoid degradation of the product on mRNA or protein level by RNases or proteases within the extract, the genetic modification of the bacterial strains could be a possible approach (1.2.4). Thus, for example, RNase I knockout mutants of the *Geobacillus* strains would be interesting variants, but also protease knockouts or strains equipped with additional chaperones would be conceivable.

To analyse the quality of the cell extracts and possibly increase the general yield of *in vitro* translated proteins, the quantification of intact ribosomes in the cell extract would be interesting (Failmezger, 2018). The ribosome concentration in cell extracts can be directly quantified by separation from other cellular components by sucrose gradient ultracentrifugation (Dong *et al.*, 1995; Graham & Rickwood, 1997). By creating a "polysome profile" including single ribosomal subunits, ribosomes and polysomes, the translational activity of a cell can be determined (Qin & Fredrick, 2013; Failmezger, 2018). However, this method is very elaborate and time consuming. Modern quantification of ribosomes relies on the assessment of the intracellular concentration of ribosomal RNA (rRNA) (Dennis *et al.*, 2004). This is done by capillary gel electrophoresis (CGE) and UV detection (Hjertén, 1983) or the detection by laser-induced fluorescence (LIF) (Failmezger, 2018).

5.3.2 Enzymes of archaeal origin

After the *in vitro* expression of the model protein GFP was shown successfully with the commercial T7 polymerase and self-prepared extract of *E. coli*, the system was transferred to difficult-to-express enzymes. For this purpose, three hyperthermophilic, putative metallohydrolases from *I. hospitalis* KIN4/I, found via function-based screening, were chosen. The heterologous expression of Igni13, Igni15 and Igni18 has not been possible in *E. coli* until now, only the eukaryotic expression system was successful. The expression of archaeal genes comes with many limitations, especially when using common bacterial hosts such as *E. coli*. Even the heterologous expression in their natural host has proven challenging because of the difficult growth conditions and low biomass yields of archaea. When using bacterial hosts, the expression of archaeal genes often fails due to low yield, insolubility or misfolding of the protein (inclusion bodies) (Smith & Robinson, 2002). Therefore, alternative expression systems, such as the methylotrophic yeast *Pichia pastoris*, are frequently used for the production of archaeal enzymes, which, however, are unbelievably time-consuming.

In this work, the cell-free expression of archaeal lipases was first demonstrated under standard expression conditions (3.10). The detailed characterized Igni18 was translated within 4 hours using the *E. coli* extract-based system as well as extracts of *Geobacillus* sp. GHH01 and *G. thermoleoverans*,

leading to comparable activities (4.2.3.1). Unfortunately, detection via Western Blot analysis with His_6-Tag antibodies could only be reproduced sufficiently with Igni18.

The fact that only Igni18 could be clearly detected in the Western Blot, but activity of all 3 Igni lipases was measured, suggests that the concentrations of the *in vitro* expressed lipases are in the range of the detection minimum of the Western blot analysis. In addition, it is well known, that in Western Blot analysis each protein shows different characteristics with regards to optimal transfer and detection conditions. It is highly likely that e.g. chosen conditions are well suitable for the detection of Igni18 but do not necessarily apply to all Igni lipases. This could result in a reduced signal for latter proteins even though they are expressed in higher amounts (Figure 19) than suggested by Western Blot results. Also, detection with Anti-His_6-antibodies might not be suitable for the detection of Igni13 and Igni15, as tertiary structures could make the His_6-tag inaccessible. Despite measured activity (Figure 19), no exact statement can be made about the specific yield of active enzyme expressed *in vitro*, since the background activity of the extract makes it difficult to quantify the absolute activity of the particular Igni lipases.

In addition, it can be assumed that the protein yield was at the detection limit of the Western blot analysis with secondary anti-rabbit IgG antibody conjugated with alkaline phosphatase, which is 100 pg. The Western blot analysis with fluorescent or chemiluminescent detection would be conceivable alternatives to detect the lowest protein levels (chemiluminescent HRP with Luminol: detection limit of fg to 1 pg), whereby the high total protein concentration of the samples will also increase the background signal. It should be considered that the problem is not the detection limit for the *in vitro* expressed proteins, but the low protein yield per se. Thus, approximately 1 µg of the archaeal enzyme could be generated from a 100 µl *in vitro* translation reaction (based on band thickness and intensity, compared to standards). Although this is not a satisfactory yield, considering that most archaeal enzymes can not be expressed in bacterial expression hosts at all and expression in *Pichia pastoris* requires weeks of cultivation, the cell-free expression of the Igni lipases within half a day is a great success.

Furthermore, it could be shown that all 3 lipases from *I. hospitalis* were active. This was shown in activity tests with pNP-C16 at 90°C (4.2.3.1.1). The fact that the enzymes apparently have been folded correctly, even with the translational machinery from *E. coli* and *Geobacillus* at moderate temperatures, is convincable. This fact is also supported by the band pattern of Igni18 in the Western blot, which suggests the formation of stable homodimers and homotrimers of the enzyme. Finally, previous studies showed that Igni18 forms stable multimers with a half-life of 46h at 90°C. The fact that the expression and folding of the protein proceeds slower *in vitro* than in a host cell, might explain the correct folding and formation of different structures of the lipases.

To sum up, CFPS is a suitable system for the expression of archaeal genes and thus represents a promising technology for the discovery of hyperthermophilic biocatalysts not only of bacterial origin. By using CFPS for function-based screening, enzymes could be detected for the first time, that are not functionally expressed in a metagenomic library based on a common host such as *E. coli*. This would significantly improve the hit rate in metagenomics.

5.4 Screening for specific activities of *in vitro* expressed enzymes

Classical function-based metagenomics typically consists of 3 processes: (1.) Isolation and cloning of metagenomic DNA and construction of gene libraries, (2.) heterologous expression in a suitable host organism, and (3.) the conception of efficient screening methods for the identification of clones of interest with specific activity. By using CFPS systems, the processes (1.) and (2.) can be revolutionized. However, the screening method, which still represents a major limitation in metagenomics, is decisive for the yield of positive "hits". In recent years, the design of functional screening methods exponentially increased, especially for the identification of hydrolases and oxidoreductases (Reyes-Duarte *et al.*, 2012).

The screening methods differ both in their set-up and in the choice of the type of substrate. For example, agar plate screening approaches are often used for function-based metagenomics, as they are simple to perform and active clones are visualized optically. For example, lipases/esterases can be identified agar plate-based by the use of an indoxyl ester (product: indigo) or a-naphthyl ester (product: diazo dye complex, brown precipitate) (Miller & Karn, 1980; Pazmiño & Fraaije, 2007). The reason why the agar plate-based screening was not used in this work is the insufficient sensitivity and visualization of differences in the catalytic rates of the *in vitro* expressed enzymes. This method makes it difficult to quantify the turnover rate of enzymes or the enzyme concentration itself. Therefore, microtiter plate-based screening methods were used for the screening for specific activities of *in vitro* expressed enzymes.

5.4.1 Detection of lipase activity

For the activity test for *in vitro* expressed Igni lipases, the substrate pNP-C16 was used. Although CFPS was particularly demonstrated by detection of activity, background activity is a major problem. While heating of the expression mixture allowed the denaturation of many of the cell extract components, the background was only lowered but not eliminated. An important aspect is the autohydrolysis of the substrate, which occurs in particular at neutral or basic pH values. Due to the slow turnover rate of the Igni lipases, the activity test was carried out for a relatively long time (at least 5 h), which naturally promotes autohydrolysis and unspecific hydrolysis by cell extract components. However, changes in the buffer composition for *in vitro* expressions resulted in poorer expression rates. Therefore, the lipolytic activity of the *in vitro* expressed Igni lipases can only be determined as the activity above a threshold corresponding, for example, to the highest value of the negative control. Unfortunately, no activity assays could be performed successfully with Igni lipases immobilized on Ni^{2+}-NTA agarose beads. To eliminate background signals, the aptamer-based affinity purification for His_6-tagged

proteins would be an alternative method (Lim *et al.*, 2013). Aptamers are characterized by their stability against high temperature and extreme pH, a long half-life and a reversible denaturation. Their robustness makes them a viable alternative for purifying *in vitro* expressed thermostable lipases.

5.4.2 MU assay with immobilized PET hydrolases

Since the focus of this work was on the CFPS of exciting enzymes of metagenomic origin, so-called PET hydrolases were chosen as target proteins as well. PET is one of the most widely used synthetic polymers in plastic products worldwide. Only a small part of the plastic is recycled, the majority accumulates in the environment and is only very slowly decomposed into microplastic particles. The environmental impact of plastic is a global problem, that´s why scientists are looking for ways to break down microplastics. Through the metagenomic screening of environments exposed to PET, in 2016, scientists were able to isolate a new bacterium that uses PET as the major energy and carbon source. It is capable of converting PET into terephthalic acid and ethylene glycol using 2 PET hydrolases (Yoshida *et al.*, 2016). For some years now, there has been a search for other enzymes that are capable of degrading PET to possibly find new solutions to solve the global "plastic problem".

Dr. Dominik Danso kindly provided me the cloned genes of four newly discovered PET hyrolases. For identification of PET hydrolases, a hidden Markov model (HMM) was developed to screen genome and metagenome databases for potential PET hydrolases. Over 1500 candidates were identified from which 12 new potential PET hydrolases were chosen for further analysis (Danso *et al.*, 2018). Four of these enzymes were heterologously expressed and extensively characterized. Based on this, *in vitro* transcription and translation experiments were carried out with two of these PET hydrolases (PET2 and PET6). Both showed the highest activity in plate assays with PET nanoparticles and PCL, and a temperature optimum of 55°C to 60°C, which makes them particularly interesting for the biotechnical application.

The *in vitro* transcription of PET hydrolases was demonstrated with both T7 RNAP and $RNAP_E$. However, the detection of the *in vitro* translated PET hydrolases proved to be more difficult as the signals in the Western blot were very weak (Figure 20). Unfortunately, activity tests with pNP-C8 did not produce any results, as the background signal seemed to be an even greater challenge than in the lipase assay. Therefore, 4-MU esters were selected as substrates which release the fluorescent 4-MU by enzymatic degradation. The advantage of using fluorescence detection is the greater specificity and the up to 1000-fold higher sensitivity compared to UV/Vis absorption methods. As a result, the measurements are in general more precise and detection of extremely low levels of enzymatic activity is possible. This method is particularly well suited for the detection of *in vitro* translated enzymes, since the

fluorescence intensity of 4-MU at basic pH (which corresponds to the pH of the *in vitro* mixure) is up to 100-fold higher than at neutral or acidic pH (Roberts, 1985; Gee *et al.*, 1999).

Since the problem of background signals was not solved by lowering the detection limit, a method has been developed for immobilizing *in vitro* translated PET hydrolases and subsequently performing the activity tests (3.11.2) (Figure 27). The *in vitro* expressed His_6-tagged PET hydrolases were immobilized on Ni^{2+}-NTA Agarose beads, a Ni^{2+}-charged affinity resin that is usually used to purify recombinant proteins containing a His_6-tag.

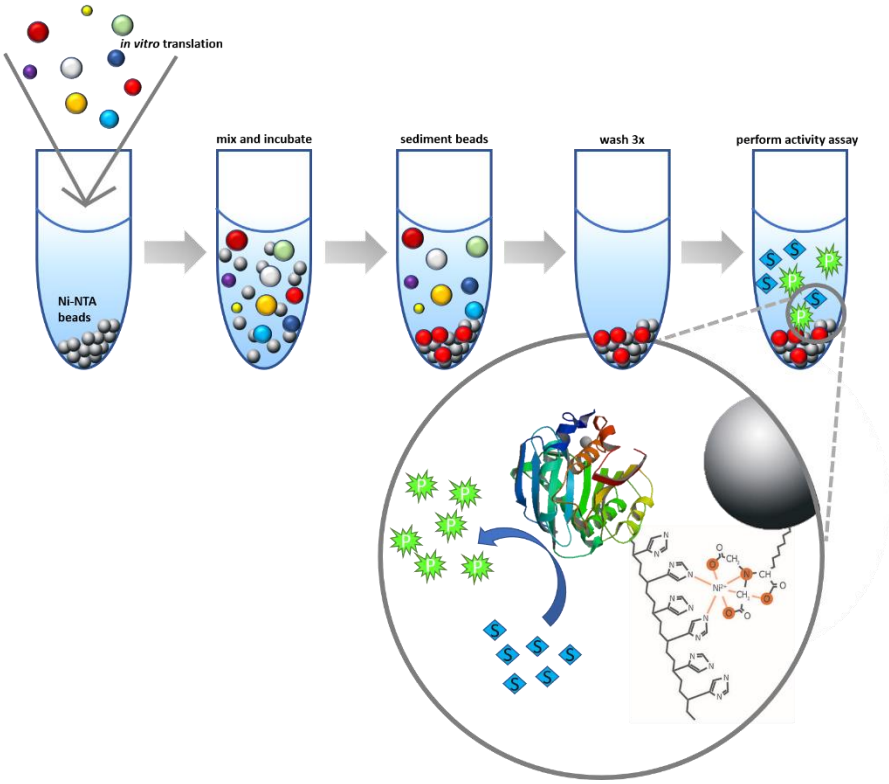

Figure 27. Scheme of immobilization of *in vitro* expressed enzymes and subsequent activity assay.
After *in vitro* translation, the sample is mixed with Ni^{2+}-NTA Agarose beads and incubated until the tagged proteins have bound. Subsequently, the beads are sedimented and washed 3 times to remove interfering components from the *in vitro* mixture. Thereafter, the enzyme reaction can be carried out directly with the enzymes coupled to the beads and the supernatant can be measured fluorometrically on a microtiter plate scale.
The bond between His_6-tag and Ni^{2+}-NTA resin: Nickel ions immobilize on NTA Agarose through coordination sites with three oxygen atoms and one nitrogen atom. Two of the six histidine residues of the tagged recombinant protein can then attach to the other two coordination sites (Bolanos-Garcia & Davies, 2006).

In this crude purification, the tagged enzyme was not eluted, the activity test was performed with coupled enzymes. By binding the tagged PET hydrolases to the Ni^{2+}-NTA Agarose beads, interfering cell extract components could be removed by several washing steps. At the same time, a buffer change

was made to provide the optimal conditions for the activity test. By this method, it was possible to minimize background signals and clearly demonstrate the presence of the *in vitro* expressed PET hydrolase.

Surprisingly, *Geobacillus*-based translations showed significantly higher fluorescence values, giving the impression that, for some proteins, the *Geobacillus* extract might be better suited for *in vitro* translation experiments (Figure 21). The standard deviation bars show that the whole process of *in vitro* transcription, translation, coupling to Ni^{2+}-NTA agarose beads, activity assay, and fluorescence measurement provides reliable results. By decoupling *in vitro* transcription and translation, it is also possible to compare both RNAPs in terms of their activity. In addition, it can be examined how translatable the respective transcripts are, since these are used in equal amounts for translation experiments. It is striking that translations with mRNA produced by the $RNAP_E$ achieve higher activities, ie higher protein yields, than translations with mRNA produced by the commercial T7 RNAP. This is the case with both *Geobacillus* and *E. coli*-based translation experiments. Since the same amount of mRNA was used in all experiments, this indicates a better quality of $RNAP_E$ transcripts.

It is also interesting that despite clear activity in the assay, signals were not always detected in Western blot analysis. This also proves that the assay developed in this work is a much more reliable and sensitive detection method especially for very low enzyme concentrations, such as those found in metagenomics.

For the screening of a metagenomic library, however, the presence of an affinity tag for the purification of the protein would be necessary. Typically used (fosmid/cosmid) vectors do not contain an additional tag. Inserting tags within a genomic fosmid clone at specific target sites through recombineering was proposed by Tursun *et al.* (Tursun *et al.*, 2009), but is not as easy to realize with metagenomic libraries as the sequences are unknown. Engineering of a fosmid vector like pCC1FOS™ containing at least two tags flanking the metagenomic insert, would be an approach, however, also results in the expression of non-tagged proteins. For function-based screening of thermoresistant enzymes, the subsequent incubation of the CFPS reaction at high temperatures would be possible to minimize background effects. The use of a tag is therefore not necessary.

5.4.3 Direct photometric detection of *in vitro* expressed enzymes

Although the *in vitro* expression of the ß-lactamase could be shown colorimetrically using the substrate nitrocefin for activity tests, the background signal was the biggest problem. Both the *in vitro* translation buffer and the cell extract could be found as cause. Desprite this, the presence of the enzyme

expressed *in vitro* could directly be visualized by a color change for the first time in this study. To obtain clearer results, the affinity coupling to beads would be a promising solution.

5.5 *In vitro* protein expression in polymersomes

Although microtiter plate-based screening allows the simultaneous screening of hundreds of genes or clones, it is cost- and time-consuming, despite the ever-decreasing volumes and the use of laboratory robots. As a result, more and more microfluidic technologies are emerging, permitting screening rates of thousands to millions of clones per day. These HTS strategies are made possible by FACS, for example, by the substrate-induced expression of a fluorescent marker or the direct enzymatic conversion of a substrate to a fluorescent product (Uchiyama & Watanabe, 2008).

The idea to revolutionize function-based metagenomics by FACS-based HTS methods is not new, but most approaches like the SIGEX-system are based on the construction of host-based metagenomic libraries incapsulated in microdroplets (Figure 5). However, this approach does not overcome the host-based expression hurdles. More effective would be a higher expression efficiency coupled with an increased screening rate.

Therefore, in cooperation with the "Schwaneberg Group" of the RWTH Aachen, *in vitro* compartmentalization using polymersomes was carried out. The main difficulty was the synthesis of the polymersomes themselves, since changes in the flow rates of the individual phases either did not lead to the formation of double-emulsions or led to the burst at the end of the device. Frequently, attachment of the coblock polymer to the thin channels could be observed, resulting in the blockage of these. An automated method for production and coating of the devices could solve these problems.

Despite the technical difficulties, the incapsulation of *in vitro* mixtures with both *E. coli* and *Geobacillus* extract was demonstrated. The polymersomes showed exceptional stability regarding the transfer from glass vessels to Eppendorf tubes and even to microscope slides. The incubation for more than 18 h was possible as well.

In addition, cell-free expression of both GFP and the metagenomic-derived cellulase CelA2 in polymersomes was demonstrated for the first time in this work. Incapsulation of the reaction mixture, subsequent incubation of the polymersomes in Eppendorf tubes and microscopic analysis allowed both non-fluorescent and fluorescent polymersomes to be detected. The reason for the few non-fluorescing polymersomes could be a heterogeneity in the component distribution of the *in vitro* mixture. Control reactions, for example with empty vectors or without the addition of the substrate, need to be done in the future, to prove that the fluorescence signals are indeed based on the successful CFPS of the desired enzyme within the polymersome.

Besides the expression of the reporter protein GFP, which is typically used as an expression signal in FACS analysis, cell-free expression of the cellulase CelA2 could be demonstrated by the direct conversion of a specific substrate to a fluorescent signal in the polymersome. For this purpose, a CFPS based on *Geobacillus* extract was performed. The reaction was prior supplemented with the substrate for the subsequent fluorometric detection. Thus, it was possible to show that both the *Geobacillus*-based translation works in polymersomes and that fluorescence-based detection can be coupled directly to protein expression by the addition of a suitable substrate.

The method of *in vitro* compartmentalization has become increasingly popular for several years, because these synthetic artificial cells allow the mimicking of reactions of a natural cell. This powerful tool is used for example for directed evolution of new enzymes (Griffiths & Tawfik, 2003). However, mostly simpler water-in-oil emulsions are used for that. The use of polymersomes for *in vitro* compartmentalization experiments is also not new, for example for the expression and aggregation analysis of cytoskeletal proteins (Martino *et al.*, 2012) or the expression of membrane proteins (Nallani *et al.*, 2011). But the outstanding stable polymersomes have not yet been coupled with CFPS reactions for function-based screenings.

The next step in the use of *in vitro* compartmentalization for functional metagenomics would be the application of FACS technologies for screening of positive clones.

6 Conclusion and outlook

6.1 Conclusion

Metagenomics is a powerful tool not only to describe the microbial biodiversity of the environment, but also to be a promising source for entirely new sequence classes and biocatalysts with exciting new functions relevant for industrial applications (Ferrer *et al.*, 2005; Ferrer *et al.*, 2007). The field of microbiology and biotechnology has undoubtedly been revolutionized by metagenomics. However, it is well known that the current approaches to screen metagenomic libraries for active enzymes have several limitations that still do not allow this technology to access the full metabolic potential and broad genetic spectrum of a metagenome (Beloqui *et al.*, 2008; Ferrer *et al.*, 2009; Nevondo, 2016). The main obstacles are the construction of metagenomic libraries and the protein expression in a suitable host for function-based screenings (Felczykowska *et al.*, 2015). Last is strongly dependent on the experimental set-up. For library construction, recombinant protein expression and the screening for a specific desired function, *E. coli*-based systems are still the first choice.
Typical problems of protein expression in prokaryotic host systems include translational termination due to a different codon usage, defective post-translational processing, the aggregation of misfolded proteins, and the lack of machinery for the insertion of membrane proteins of eukaryotic origin (Jansohn & Rothhämel, 2012). By using, for example, alternative host systems such as yeast or insect cells, these hurdles can only be partially overcome. CFPS, however, represents a real alternative, which is time-saving and less labor-intensive.

The major aim of this study was to develop a host-independent approach for screening enzymatic activities using metagenomic-derived DNA. This new screening system should be designed to eliminate time-consuming transformation and heterologous expression strategies and help to overcome the main limitations in function-based metagenomics. For that, three techniques, which are finding more and more applications in recent years, should be combined: Cell-free protein synthesis (CFPS), *in vitro* compartmentalisation (IVC) and Fluorescence Activated Cell Sorting (FACS).

The first *E. coli*-based CFPS system was introduced by Mattheai and Nierenberg in 1961 with the focus on studying the translation process. Since then, much has been done in the field of CFPS, so today there are a multitude of advanced systems based on extracts of organisms from all domains of life, including mammals, insects, plants, funghi, bacteria and archaea (Zemella *et al.*, 2015). Going “cell-free” allows to bypass many limitations existing in *in vivo* expression systems as it is highly flexible in the reaction conditions, the absence of a membrane allows the use of a high variety of substrates for function-based screenings. Furthermore, larger gene libraries can be produced, since the system is not limited by the transformation efficiency of the host cells (Charles *et al.*, 2017). In addition, *in vitro*

expression enables the incorporation of unnatural aminoacids and cofactors, as well as the synthesis of proteins that would otherwise be toxic to the host organism and severely limit cell viability.

To accomplish *in vitro* transcriptions of metagenomic-derived genes, both bacterial (RNAP from *T. thermophilus*) and viral (metagenome-derived $RNAP_E$) RNAP were cloned and heterologously expressed. With the $RNAP_E$, it succeeded to establish a promising alternative to the classic T7 system. In terms of temperature profile and efficiency, the $RNAP_E$ is in no way inferior to the commercial T7. In addition, it was shown that it has a broad recognition spectrum of different native promoters in uncloned, genomic DNA and is suitable for the transcription of fosmids of a metagenomic library. Since the transcripts synthesized by the $RNAP_E$ were successfully translated *in vitro*, it can be assumed that the $RNAP_E$ has a very low error rate. To obtain concrete data, RNA-sequencing of the transcripts would be an important next step.

In recent years, the search for alternative microbial model systems, which are more suitable for function-based screenings than *E. coli*, is in focus. These include also psychrophilic and thermophilic systems that could facilitate enzyme discovery especially from extreme sources (Charles *et al.*, 2017). With this idea in mind, besides the discovery of a promising new RNAP for *in vitro* transcription assays, alternative cell extracts from psychrophilic, as well as mesophilic and thermophilic bacteria were prepared and tested for their applicability for *in vitro* translations at different temperatures. A total of 8 different cell extracts were prepared and tested. From those 8, besides the *E. coli* systems, 3 completely new cell extracts showed activity *(Geobacillus* sp., GHH01, *G. thermoleoverans*, *P. antarctica*). Thus, a novel, in-house *Geobacillus*-based CFPS system could be established as a promising alternative to the typical *E. coli*-based cell-free system. This system convinces in the yield of expressed, active protein, but shows weaknesses in its reproducibility.

Nevondo has assumed that combining a variety of cell extracts into a multi-species cell-free cocktail might increase the number of transcribed and translated genes in an environmental sample (Nevondo, 2016). Thus, it would be possible to improve the functional screening hit rate in metagenomics by combining the transcription-translation machinery of different organisms, such as *E. coli* and *Geobacillus* sp. Therefore, the screening of more, various organisms for the preparation of cell extracts for CFPS systems would be an interesting part of future research.

Model experiments for CFPS of metagenomic-derived enzymes were successfully carried out, including archaeal lipases (Kobus *et al.*, 2019) and PET-degrading hydrolases (Danso *et al.*, 2018).

To achieve higher protein yields in the cell-free system, the supplementation of the reactions with additional cofactors would be useful (Kinfu, 2018). Also, the stabilization of the mRNA by inclusion of terminal stem-loop structures have been proven to be beneficial, as well as the creation of (RNase E)

variants from strains used for extract preparation (Ahn *et al.*, 2005). Jun *et al.* prepared extracts from *E. coli* BL21 derivated strains which were equipped with additional genes of disulfide isomerase or molecular chaperone, as well as RNase E-deficient variants to improve mRNA stability and enhance protein folding and solubility (Jun *et al.*, 2008).

The biggest challenge that appeared during this work is the background signal in activity assays from the cell extract and translation components as well as buffer components. Protein purification using heat denaturation could minimize this factor or the coupling to beads (as used for the detection of PET hydrolases). In the future, the coupling of heterologous proteins to magnetic beads using affinity ligand techniques (Safarik & Safarikova, 2004) in microtiter plates would be a simple solution for screening of metagenomic libraries. The magnetically separated isolation and purification of tagged enzymes in 96-well-microplates is a common technique that is supplemented by various companies. With the help of so-called microplate loaders, hundreds of activity tests could be carried out in parallel in a very short time. Applying robotic technologies would further simplify and automate the screening of metagenomic libraries.

6.2 Outlook

In order not only to carry out hundreds of reactions simultaneously, but also to automatically screen hundreds of thousands of clones in a very short time, the methods IVC and FACS would be promising. The best-known example in the literature for screening metagenomic libraries coupled with FACS technology is the SIGEX system as described in 1.4. However, the SIGEX system is based on the detection of functional genes by screening the reporter protein GFP and thus does not allow the direct selection for a specific, enzymatic reaction. In addition, this method is based on the transcription and translation machinery of *E. coli*, as only transformed cells are selected after a particular stimulus (Uchiyama & Watanabe, 2008). Although it is possible to screen hundreds of thousands of clones in high throughput scale, the limitations by the host organism *E. coli*, as described, still exist.

By coupling CFPS packed into microdroplets (IVC) and FACS techniques, not only the expression number of metagenomic-derived proteins can be increased, but also the screening and selection of extremely large libraries in a very short time becomes possible.

Already in 2009, Ferrer *et al.* proposed "*in vitro* metagenomics" as an approach to help overcoming those limitations associated with classical heterologous protein expression (Ferrer *et al.*, 2009). While the SIGEX system already uses FACS for screening in function-based metagenomics, the vision of "*in vitro* metagenomics" remains a not yet realized vision. A small, first step in the realization of cell-free functional-based metagenomics in high throughput scale using IVC in polymersomes and FACS

techniques could be demonstrated in this work: By the successful encapsulation and performance of the *Geobacillus*-based CFPS system for the expression and subsequent activity-based detection of the metagenomic-derived cellulase CelA2. These results are promising in terms of rapid screening of a metagenomic library, with FACS being the next important step.

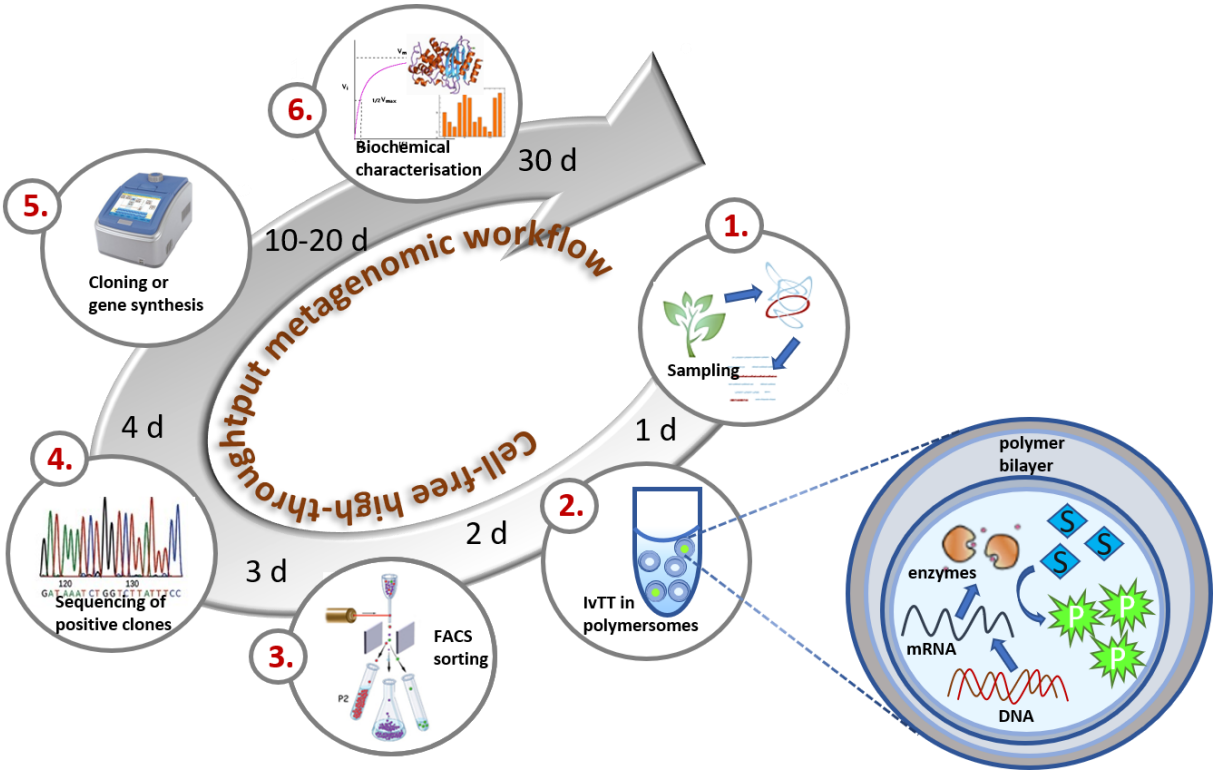

Figure 28. Scheme of the workflow for cell-free high-troughput metagenomics.

The process mainly consists of 6 steps, with the focus on step 2 and 3: Metagenomic DNA fragments are transcribed and translated in polymersomes. Positive „hits“ are detected through the enzymatic coversion of a non-fluorescent substrate into a fluorescent product and subsequent FACS. Steps 4 to 6 are typical steps in classical metagenomics. The idea of the scheme is based on a figure for the "*in vitro* flow cytometry-based screening platform for cellulase engineering" (Körfer *et al.*, 2016).

In the future, functional metagenomics might look like this (Figure 28):

Collecting an environmental sample, isolation of the DNA, restriction of the DNA and avoiding elaborate cloning steps (1.). Subsequent *in vitro* transcription and translation of the metagenomic DNA fragments in polymersomes (2.). A previously added, specific, non-fluorescent substrate converting into a fluorescent product through the enzymatic activity of a translated protein. Fluorescent polymersomes are separated from non-fluorescent polymersomes using FACS (3.). Subsequently, the polymersomes can be broken up by drying and osmotic pressure (Shum *et al.*, 2008) and the positive "polymersome clones" are sequenced (4.). This is followed by the typical cloning or gene synthesis (5.) and the expression and extensive biochemical characterization of the enzymes (6.). The time required

is difficult to estimate, but the classical screening takes about one to two years according to own experiences. This new process could possibly be realized in one to two months.

IVC-analysis of metagenomic libraries could be a promising tool for screening of billions of bp for exciting enzymatic activities within a few days, which to date is impossible with classical approaches (Ferrer *et al.*, 2009). This work lays the foundation to revolutionize function-based metagenomics, that is, to dramatically reduce costs, man-power, and time, increase hit-rates, and maximize the access to the metabolic potential of an environmental sample.

6.3 Functional metagenomics: modern biotechnology and future trends.

Functional-based metagenomics is a great source for new enzymes for industrial application. Why is the need for ever newer, more robust biocatalysts so great? Ferrer *et al.* published a paper in 2016 that sheds light on this issue (Ferrer *et al.*, 2016). The identification and use of new biocatalysts represents a “greener” alternative to chemical synthesis (Fernandez-Arrojo *et al.*, 2010; Bornscheuer *et al.*, 2012; Turner & Truppo, 2013; Ferrer *et al.*, 2016).

The benefits of enzymes are, that they are mostly non-toxic catalysts with low waste production, low energy consumption as a result of comparable moderate reaction conditions, low consumption of organic solvents and metals. Enzymes can therefore be seen as renewable feedstocks (Wenda *et al.*, 2011). While surpassing the maximum rate of oil extraction ('peak oil'), besides the search for non-fossil fuel sources, biotechnology is becoming an increasingly important component in the industrial sector, to make chemistry sustainable. Enzymatic catalysis made up 7% in 2016 (Brahmachari *et al.*, 2016), is expected to increase up to 40% for bulk chemical synthesis processes in 2030, which now require environmentally damaging organic solvents and high energy inputs (Sayawa 2010; (Ferrer *et al.*, 2016). In 2018, the global enzymes market size was USD 7.10 billion and is expected to reach USD 11.03 billion by 2026 (https://www.globenewswire.com, 2019); numbers vary depending on the source) because of its increasing application in pharmaceuticals, detergents and food (https://www.grandviewresearch.com).

Modern methods for the identification of new biocatalysts are a highly explosive field of research and a future trend.

7 References

Aakvik T., Degnes K.F., Dahlsrud R., Schmidt F., Dam R., Yu L., Volker U., Ellingsen T.E. & Valla S. (2009) A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species. *FEMS microbiology letters,* **296,** 149–158. DOI: 10.1111/j.1574-6968.2009.01639.x.

Abdine A., Park K.-H. & Warschawski D.E. (2012) Cell-free membrane protein expression for solid-state NMR. *Methods in molecular biology (Clifton, N.J.),* **831,** 85–109. DOI: 10.1007/978-1-61779-480-3_6.

Acevedo J.P., Reyes F., Parra L.P., Salazar O., Andrews B.A. & Asenjo J.A. (2008) Cloning of complete genes for novel hydrolytic enzymes from Antarctic sea water bacteria by use of an improved genome walking technique. *Journal of biotechnology,* **133,** 277–286. DOI: 10.1016/j.jbiotec.2007.10.004.

Ahn J.-H., Chu H.-S., Kim T.-W., Oh I.-S., Choi C.-Y., Hahn G.-H., Park C.-G. & Kim D.-M. (2005) Cell-free synthesis of recombinant proteins from PCR-amplified genes at a comparable productivity to that of plasmid-based reactions. *Biochemical and biophysical research communications,* **338,** 1346–1352. DOI: 10.1016/j.bbrc.2005.10.094.

Alcaide M., Stogios P.J., Lafraya A., Tchigvintsev A., Flick R., Bargiela R., Chernikova T.N., Reva O.N., Hai T., Leggewie C.C., Katzke N., La Cono V., Matesanz R., Jebbar M., Jaeger K.-E., Yakimov M.M., Yakunin A.F., Golyshin P.N., Golyshina O.V., Savchenko A. & Ferrer M. (2015) Pressure adaptation is linked to thermal adaptation in salt-saturated marine habitats. *Environmental microbiology,* **17,** 332–345. DOI: 10.1111/1462-2920.12660.

Amann R.I., Ludwig W. & Schleifer K.H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological reviews,* **59,** 143–169.

Anderson M.J., Stark J.C., Hodgman C.E. & Jewett M.C. (2015) Energizing eukaryotic cell-free protein synthesis with glucose metabolism. *FEBS Letters,* **589,** 1723–1727. DOI: 10.1016/j.febslet.2015.05.045.

Angelov A., Mientus M., Liebl S. & Liebl W. (2009) A two-host fosmid system for functional screening of (meta)genomic libraries from extreme thermophiles. *Systematic and applied microbiology,* **32,** 177–185. DOI: 10.1016/j.syapm.2008.01.003.

Bachmann B.J. (1983) Derivations and Genotypes of Some Mutant Derivatives ofEscherichia coliK-12.

Baneyx F. (1999) Recombinant protein expression in Escherichia coli. *Current opinion in biotechnology,* **10,** 411–421.

Bartenstein J.E., Robertson J., Battaglia G. & Briscoe W.H. (2016) Stability of polymersomes prepared by size exclusion chromatography and extrusion. *Colloids and Surfaces A: Physicochemical and Engineering Aspects,* **506,** 739–746. DOI: 10.1016/j.colsurfa.2016.07.032.

Beloqui A., Maria P.D. de, Golyshin P.N. & Ferrer M. (2008) Recent trends in industrial microbiology. *Current opinion in microbiology,* **11,** 240–248. DOI: 10.1016/j.mib.2008.04.005.

Berini F., Presti I., Beltrametti F., Pedroli M., Vårum K.M., Pollegioni L., Sjöling S. & Marinelli F. (2017) Production and characterization of a novel antifungal chitinase identified by functional screening of a suppressive-soil metagenome. *Microbial cell factories,* **16. DOI:** 10.1186/s12934-017-0634-8.

Bernath K., Hai M., Mastrobattista E., Griffiths A.D., Magdassi S. & Tawfik D.S. (2004) In vitro compartmentalization by double emulsions: sorting and gene enrichment by fluorescence activated cell sorting. *Analytical biochemistry,* **325,** 151–157.

Bernstein J.A., Lin P.-H., Cohen S.N. & Lin-Chao S. (2004) Global analysis of Escherichia coli RNA degradosome function using DNA microarrays. *Proceedings of the National Academy of Sciences of the United States of America,* **101,** 2758–2763. DOI: 10.1073/pnas.0308747101.

Bolanos-Garcia V.M. & Davies O.R. (2006) Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. *Biochimica et biophysica acta,* **1760,** 1304–1313. DOI: 10.1016/j.bbagen.2006.03.027.

Bornscheuer U.T., Huisman G.W., Kazlauskas R.J., Lutz S., Moore J.C. & Robins K. (2012) Engineering the third wave of biocatalysis. *Nature,* **485,** 185–194. DOI: 10.1038/nature11117.

Borukhov S., Lee J. & Laptenko O. (2005) Bacterial transcription elongation factors: new insights into molecular mechanism of action. *Molecular microbiology,* **55,** 1315–1324. DOI: 10.1111/j.1365-2958.2004.04481.x.

Bradford M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry,* **72,** 248–254.

Brahmachari G., Demain A.L. & Adrio J.L. (2016) *Biotechnology of Microbial Enzymes: Production, Biocatalysis and Industrial Applications.* Elsevier Science.

Bremer H. & Dennis P.P. (2008) Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus,* **3. DOI:** 10.1128/ecosal.5.2.3.

Butler E.T. & Chamberlin M.J. (1982) Bacteriophage SP6-specific RNA polymerase. I. Isolation and characterization of the enzyme. *The Journal of biological chemistry,* **257,** 5772–5778.

Calhoun K.A. & Swartz J.R. (2007) Energy systems for ATP regeneration in cell-free protein synthesis reactions. *Methods in molecular biology (Clifton, N.J.),* **375,** 3–17. DOI: 10.1007/978-1-59745-388-2_1.

Carlson E.D., Gan R., Hodgman C.E. & Jewett M.C. (2012) Cell-Free Protein Synthesis: Applications Come of Age. *Biotechnology advances,* **30,** 1185–1194. DOI: 10.1016/j.biotechadv.2011.09.016.

Cell-Free Protein Expression | Thermo Fisher Scientific - DE. URL https://www.thermofisher.com/de/de/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/cell-free-protein-expression.html [accessed on 28 November 2018].

Charles T.C., Liles M.R. & Sessitsch A. (2017) *Functional Metagenomics: Tools and Applications.* Springer International Publishing.

Cheng J., Pinnell L., Engel K., Neufeld J.D. & Charles T.C. (2014) Versatile broad-host-range cosmids for construction of high quality metagenomic libraries. *Journal of microbiological methods,* **99,** 27–34. DOI: 10.1016/j.mimet.2014.01.015.

Chong S. (2014) Overview of Cell-Free Protein Synthesis: Historic Landmarks, Commercial Systems, and Expanding Applications. *Current protocols in molecular biology / edited by Frederick M. Ausubel … [et al.],* **108,** 16.30.1-16.30.11. DOI: 10.1002/0471142727.mb1630s108.

Christian Leggewie (2005) *Neue Biokatalysatoren aus dem Metagenom: Expression, Identifizierung und biochemische Eigens chaften.* Dissertation, Heinrich-Heine-Universität, Düsseldorf.

Clark M.A., Choi J. & Douglas a.M. (2019) *Prokaryotic Transcription – Biology 2e.* URL https://opentextbc.ca/biology2eopenstax/chapter/prokaryotic-transcription/ [accessed on 22 August 2019].

Coleman D.J., Studler M.J. & Naleway J.J. (2007) A long-wavelength fluorescent substrate for continuous fluorometric determination of cellulase activity: resorufin-beta-D-cellobioside. *Analytical biochemistry,* **371,** 146–153. DOI: 10.1016/j.ab.2007.08.027.

Cooper G.M. (2000) *The Cell: A Molecular Approach. 2nd edition.: Translation of mRNA.* Sinauer Associates.

Courtois S., Cappellano C.M., Ball M., Francou F.-X., Normand P., Helynck G., Martinez A., Kolvek S.J., Hopke J., Osburne M.S., August P.R., Nalin R., Guerineau M., Jeannin P., Simonet P. & Pernodet J.-L. (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Applied and environmental microbiology,* **69,** 49–55.

Cramer P. (2002) Multisubunit RNA polymerases. *Current opinion in structural biology,* **12,** 89–97.

Damon C., Vallon L., Zimmermann S., Haider M.Z., Galeote V., Dequin S., Luis P., Fraissinet-Tachet L. & Marmeisse R. (2011) A novel fungal family of oligopeptide transporters identified by functional metatranscriptomics of soil eukaryotes. *The ISME journal,* **5,** 1871–1880. DOI: 10.1038/ismej.2011.67.

Daniel R. (2005) The metagenomics of soil. *Nature Reviews Microbiology,* **3,** 470. DOI: 10.1038/nrmicro1160.

Danso D., Schmeisser C., Chow J., Zimmermann W., Wei R., Leggewie C., Li X., Hazen T. & Streit W.R. (2018) New Insights into the Function and Global Distribution of Polyethylene Terephthalate (PET)-Degrading Bacteria and Enzymes in Marine and Terrestrial Metagenomes. *Applied and environmental microbiology,* **84. DOI:** 10.1128/AEM.02773-17.

David Marcey & Nathan Silva (2006) *RNA Polymerase of Thermus thermophilus.* URL http://earth.callutheran.edu/Academic_Programs/Departments/BioDev/omm/jsmol/rnapol/rnapol.html.

DeCastro M.-E., Rodríguez-Belmonte E. & González-Siso M.-I. (2016) Metagenomics of Thermophiles with a Focus on Discovery of Novel Thermozymes. *Frontiers in microbiology,* **7,** 1521. DOI: 10.3389/fmicb.2016.01521.

Dedhia P.H., Keeshan K., Uljon S., Xu L., Vega M.E., Shestova O., Zaks-Zilberman M., Romany C., Blacklow S.C. & Pear W.S. (2010) Differential ability of Tribbles family members to promote degradation of C/EBPalpha and induce acute myelogenous leukemia. *Blood,* **116,** 1321–1328. DOI: 10.1182/blood-2009-07-229450.

Dennis P.P., Ehrenberg M. & Bremer H. (2004) Control of rRNA synthesis in Escherichia coli: a systems biology approach. *Microbiology and molecular biology reviews : MMBR,* **68,** 639–668. DOI: 10.1128/MMBR.68.4.639-668.2004.

Des Soye B.J., Davidson S.R., Weinstock M.T., Gibson D.G. & Jewett M.C. (2018) Establishing a High-Yielding Cell-Free Protein Synthesis Platform Derived from Vibrio natriegens. *ACS synthetic biology,* **7,** 2245–2255. DOI: 10.1021/acssynbio.8b00252.

Dong H., Nilsson L. & Kurland C.G. (1995) Gratuitous overexpression of genes in Escherichia coli leads to growth inhibition and ribosome destruction. *Journal of bacteriology,* **177,** 1497–1504. DOI: 10.1128/jb.177.6.1497-1504.1995.

Dunn J.J. & Studier F.W. (1983) Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *Journal of molecular biology,* **166,** 477–535. DOI: 10.1016/s0022-2836(83)80282-4.

Ebright R.H. (2000) RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *Journal of molecular biology,* **304,** 687–698. DOI: 10.1006/jmbi.2000.4309.

Eggert T., Brockmeier U., Dröge M.J., Quax W.J. & Jaeger K.-E. (2003) Extracellular lipases from Bacillus subtilis: regulation of gene expression and enzyme activity by amino acid supply and external pH. *FEMS Microbiology Letters,* **225,** 319–324. DOI: 10.1016/S0378-1097(03)00536-6.

Erickson A.H. & Blobel G. (1983) Cell-free translation of messenger RNA in a wheat germ system. *Methods in enzymology,* **96,** 38–50.

Escuder-Rodríguez J.-J., DeCastro M.-E., Cerdán M.-E., Rodríguez-Belmonte E., Becerra M. & González-Siso M.-I. (2018) Cellulases from Thermophiles Found by Metagenomics. *Microorganisms,* **6. DOI:** 10.3390/microorganisms6030066.

Failmezger J. (2018) *Understanding limitations to increased potential of cell-free protein synthesis.* Dissertation, Universität Stuttgart.

Fan X., Liu X., Wang K., Wang S., Huang R. & Liu Y. (2011) Highly soluble expression and molecular characterization of an organic solvent-stable and thermotolerant lipase originating from the metagenome. *Journal of Molecular Catalysis B: Enzymatic,* **72,** 319–326. DOI: 10.1016/j.molcatb.2011.07.009.

Felczykowska A., Krajewska A., Zielińska S., Łoś J.M., Bloch S.K. & Nejman-Faleńczyk B. (2015) The most widespread problems in the function-based microbial metagenomics. *Acta Biochimica Polonica,* **62,** 161–166. DOI: 10.18388/abp.2014_917.

Fernandez-Arrojo L., Guazzaroni M.-E., Lopez-Cortes N., Beloqui A. & Ferrer M. (2010) Metagenomic era for biocatalyst identification. *Current opinion in biotechnology,* **21,** 725–733. DOI: 10.1016/j.copbio.2010.09.006.

Ferrer M., Beloqui A., Vieites J.M., Guazzaroni M.E., Berger I. & Aharoni A. (2009) Interplay of metagenomics and in vitro compartmentalization. *Microbial biotechnology,* **2,** 31–39. DOI: 10.1111/j.1751-7915.2008.00057.x.

Ferrer M., Golyshina O., Beloqui A. & Golyshin P.N. (2007) Mining enzymes from extreme environments. *Current opinion in microbiology,* **10,** 207–214. DOI: 10.1016/j.mib.2007.05.004.

Ferrer M., Martínez-Abarca F. & Golyshin P.N. (2005) Mining genomes and 'metagenomes' for novel catalysts. *Current opinion in biotechnology,* **16,** 588–593. DOI: 10.1016/j.copbio.2005.09.001.

Ferrer M., Martínez-Martínez M., Bargiela R., Streit W.R., Golyshina O.V. & Golyshin P.N. (2016) Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microbial biotechnology,* **9,** 22–34. DOI: 10.1111/1751-7915.12309.

Gabor E.M., Alkema W.B.L. & Janssen D.B. (2004) Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environmental microbiology,* **6,** 879–886. DOI: 10.1111/j.1462-2920.2004.00640.x.

Garamella J., Marshall R., Rustad M. & Noireaux V. (2016) The All E. coli TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology. *ACS synthetic biology,* **5,** 344–355. DOI: 10.1021/acssynbio.5b00296.

Gee K.R., Sun W.C., Bhalgat M.K., Upson R.H., Klaubert D.H., Latham K.A. & Haugland R.P. (1999) Fluorogenic substrates based on fluorinated umbelliferones for continuous assays of phosphatases and beta-galactosidases. *Analytical biochemistry,* **273,** 41–48. DOI: 10.1006/abio.1999.4202.

Gentry T.J., Wickham G.S., Schadt C.W., He Z. & Zhou J. (2006) Microarray applications in microbial ecology research. *Microbial ecology,* **52,** 159–175. DOI: 10.1007/s00248-006-9072-6.

Ghavami A., Labbe G., Brem J., Goodfellow V.J., Marrone L., Tanner C.A., King D.T., Lam M., Strynadka N.C.J., Pillai D.R., Siemann S., Spencer J., Schofield C.J. & Dmitrienko G.I. (2015) Assay for drug discovery: Synthesis and testing of nitrocefin analogues for use as beta-lactamase substrates. *Analytical biochemistry,* **486,** 75–77. DOI: 10.1016/j.ab.2015.06.032.

Gholamalipour Y., Karunanayake Mudiyanselage A. & Martin C.T. (2018) 3′ end additions by T7 RNA polymerase are RNA self-templated, distributive and diverse in character—RNA-Seq analyses. *Nucleic Acids Research,* **46,** 9253–9263. DOI: 10.1093/nar/gky796.

Gillespie D.E., Brady S.F., Bettermann A.D., Cianciotto N.P., Liles M.R., Rondon M.R., Clardy J., Goodman R.M. & Handelsman J. (2002) Isolation of antibiotics turbomycin a and B from a metagenomic library of soil microbial DNA. *Applied and environmental microbiology,* **68,** 4301–4306.

Golomb M. & Chamberlin M. (1974) Characterization of T7-specific ribonucleic acid polymerase. IV. Resolution of the major in vitro transcripts by gel electrophoresis. *The Journal of biological chemistry,* **249,** 2858–2863.

Gomes A.R., Byregowda S.M., Veeregowda B.M. & Balamurugan V. (2016) An Overview of Heterologous Expression Host Systems for the Production of Recombinant Proteins. *Advances in Animal and Veterinary Sciences,* **4,** 346–356. DOI: 10.14737/journal.aavs/2016/4.7.346.356.

Goyal M.R. (2018) *Scientific and Technical Terms in Bioengineering and Biological Engineering.* Apple Academic Press.

Graham J.M. & Rickwood D. (1997) *Subcellular Fractionation: A Practical Approach.* IRL Press at Oxford University Press.

Gregorio N.E., Levine M.Z. & Oza J.P. (2019) A User's Guide to Cell-Free Protein Synthesis. *Methods and Protocols,* **2,** 24. DOI: 10.3390/mps2010024.

Griffiths A.D. & Tawfik D.S. (2003) Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. *The EMBO journal,* **22,** 24–35. DOI: 10.1093/emboj/cdg014.

Gross C.A., Chan C., Dombroski A., Gruber T., Sharp M., Tupy J. & Young B. (1998) The functional and regulatory roles of sigma factors in transcription. *Cold Spring Harbor symposia on quantitative biology,* **63,** 141–155.

Gunnelius L., Hakkila K., Kurkela J., Wada H., Tyystjärvi E. & Tyystjärvi T. (2014) The omega subunit of the RNA polymerase core directs transcription efficiency in cyanobacteria. *Nucleic Acids Research,* **42,** 4606–4614. DOI: 10.1093/nar/gku084.

Handelsman J., Rondon M.R., Brady S.F., Clardy J. & Goodman R.M. (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology,* **5,** R245-R249. DOI: 10.1016/S1074-5521(98)90108-9.

Helm M., Brulé H., Giegé R. & Florentz C. (1999) More mistakes by T7 RNA polymerase at the 5' ends of in vitro-transcribed RNAs. *RNA,* **5,** 618–621.

Hjertén S. (1983) High-performance electrophoresis: the electrophoretic counterpart of high-performance liquid chromatography. *Journal of Chromatography A,* **270,** 1–6. DOI: 10.1016/S0021-9673(01)96347-2.

Hogan B.P., Hartsch T. & Erie D.A. (2002) Transcript cleavage by Thermus thermophilus RNA polymerase. Effects of GreA and anti-GreA factors. *The Journal of biological chemistry,* **277,** 967–975. DOI: 10.1074/jbc.M108737200.

Hong S.H., Kwon Y.-C., Martin R.W., Des Soye B.J., Paz A.M.d., Swonger K.N., Ntai I., Kelleher N.L. & Jewett M.C. (2015) Improving Cell-Free Protein Synthesis through Genome Engineering of Escherichia coli Lacking Release Factor 1. *ChemBioChem,* **16,** 844–853. DOI: 10.1002/cbic.201402708.

Hornung C.K. (2013) *Violacein-producing Janthinobacterium sp. HH01 : Genome analysis revealed a novel α-hydroxyketone-sensing reporter-strain,* Universität Hamburg, Hamburg.

https://www.globenewswire.com (2019) *Enzymes Market To Reach USD 11.03 Billion By 2026 | Reports And Data.* URL https://www.globenewswire.com/news-release/2019/04/10/1802216/0/en/Enzymes-Market-To-Reach-USD-11-03-Billion-By-2026-Reports-And-Data.html [accessed on 24 May 2019].

https://www.grandviewresearch.com *Enzymes Market Size, Share & Trends | Global Industry Report, 2024.* URL https://www.grandviewresearch.com/industry-analysis/enzymes-industry [accessed on 24 May 2019].

Ilmberger N., Meske D., Juergensen J., Schulte M., Barthen P., Rabausch U., Angelov A., Mientus M., Liebl W., Schmitz R.A. & Streit W.R. (2012) Metagenomic cellulases highly tolerant towards the presence of ionic liquids--linking thermostability and halotolerance. *Applied microbiology and biotechnology,* **95,** 135–146. DOI: 10.1007/s00253-011-3732-2.

Iskakova M. (2005) *Optimisation of a transcription-translation coupled in vitro system.* Dissertation, Freie Universität Berlin, Berlin.

Jackson R.J. & Hunt T. (1983) Preparation and use of nuclease-treated rabbit reticulocyte lysates for the translation of eukaryotic messenger RNA. *Methods in enzymology,* **96,** 50–74.

Jansohn M. & Rothhämel S. (2012) *Gentechnische Methoden: Eine Sammlung von Arbeitsanleitungen für das molekularbiologische Labor.* Spektrum Akademischer Verlag.

Jun S.Y., Kang S.H. & Lee K.-H. (2008) Continuous-exchange cell-free protein synthesis using PCR-generated DNA and an RNase E-deficient extract. *BioTechniques,* **44,** 387–391. DOI: 10.2144/000112690.

Kellner H., Luis P., Portetelle D. & Vandenbol M. (2011) Screening of a soil metatranscriptomic library by functional complementation of Saccharomyces cerevisiae mutants. *Microbiological research,* **166,** 360–368. DOI: 10.1016/j.micres.2010.07.006.

Kim D.M., Kigawa T., Choi C.Y. & Yokoyama S. (1996) A highly efficient cell-free protein synthesis system from Escherichia coli. *European journal of biochemistry,* **239,** 881–886.

Kinfu B.M. (2018) *Function-Based Searches for Selected Phosphotransferases and Establishing in VitroTranscriptionPlatformfor Cell-Free Metagenomics.* Dissertation, University of Hamburg, Hamburg.

Kinfu B.M., Jahnke M., Janus M., Besirlioglu V., Roggenbuck M., Meurer R., Vojcic L., Borchert M., Schwaneberg U., Chow J. & Streit W.R. (2017) Recombinant RNA Polymerase from Geobacillus sp. GHH01 as tool for rapid generation of metagenomic RNAs using in vitro technologies. *Biotechnology and bioengineering,* **114,** 2739–2752. DOI: 10.1002/bit.26436.

Kobus S., Perez-Garcia P., Hoeppner A., Holzscheck N., Kovacic F., Streit W.R., Jaeger K.-E., Chow J. & Smits S.H.J. (2019) Igni18, a novel metallo-hydrolase from the hyperthermophilic archaeon Ignicoccus hospitalis KIN4/I: cloning, expression, purification and X-ray analysis. *Acta Crystallographica Section F,* **75,** 307–311. DOI: 10.1107/S2053230X19002851.

Körfer G., Pitzler C., Vojcic L., Martinez R. & Schwaneberg U. (2016) *In vitro* flow cytometry-based screening platform for cellulase engineering. *Scientific reports,* 26128. DOI: 10.1038/srep26128.

Krivushin K., Kondrashov F., Shmakova L., Tutukina M., Petrovskaya L. & Rivkina E. (2015) Two Metagenomes from Late Pleistocene Northeast Siberian Permafrost. *Genome Announcements,* **3. DOI:** 10.1128/genomeA.01380-14.

Kurylo C.M., Alexander N., Dass R.A., Parks M.M., Altman R.A., Vincent C.T., Mason C.E. & Blanchard S.C. (2016) Genome Sequence and Analysis of Escherichia coli MRE600, a Colicinogenic, Nonmotile Strain that Lacks RNase I and the Type I Methyltransferase, EcoKI. *Genome biology and evolution,* **8,** 742–752. DOI: 10.1093/gbe/evw008.

Kuznedelov K. & Severinov K. (2009) Recombinant bacterial RNA polymerase: preparation and applications. *Methods (San Diego, Calif.),* **47,** 44–52. DOI: 10.1016/j.ymeth.2008.10.007.

Kwon Y.-C. & Jewett M.C. (2015) High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Scientific reports,* **5,** 8663. DOI: 10.1038/srep08663.

Laemmli U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* **227,** 680–685.

Lam K.N., Cheng J., Engel K., Neufeld J.D. & Charles T.C. (2015) Current and future resources for functional metagenomics. *Frontiers in microbiology,* **6. DOI:** 10.3389/fmicb.2015.01196.

LAMBORG M.R. & ZAMECNIK P.C. (1960) Amino acid incorporation into protein by extracts of E. coli. *Biochimica et biophysica acta,* **42,** 206–211.

LeCleir G.R., Buchan A. & Hollibaugh J.T. (2004) Chitinase Gene Sequences Retrieved from Diverse Aquatic Habitats Reveal Environment-Specific Distributions. *Appl. Environ. Microbiol.,* **70,** 6977–6983. DOI: 10.1128/AEM.70.12.6977-6983.2004.

Lehmann C., Sibilla F., Maugeri Z., Streit W.R., María P.D.d., Martinez R. & Schwaneberg U. (2012) Reengineering CelA2 cellulase for hydrolysis in aqueous solutions of deep eutectic solvents and concentrated seawater. *Green Chemistry,* **14,** 2719–2726. DOI: 10.1039/C2GC35790A.

Leis B., Angelov A., Mientus M., Li H., Pham V.T.T., Lauinger B., Bongen P., Pietruszka J., Gonçalves L.G., Santos H. & Liebl W. (2015a) Identification of novel esterase-active enzymes from hot environments by use of the host bacterium Thermus thermophilus. *Frontiers in microbiology,* **6,** 275. DOI: 10.3389/fmicb.2015.00275.

Leis B., Heinze S., Angelov A., Pham V.T.T., Thürmer A., Jebbar M., Golyshin P.N., Streit W.R., Daniel R. & Liebl W. (2015b) Functional Screening of Hydrolytic Activities Reveals an Extremely Thermostable Cellulase from a Deep-Sea Archaeon. *Frontiers in Bioengineering and Biotechnology,* **3. DOI:** 10.3389/fbioe.2015.00095.

Li Y., Wexler M., Richardson D.J., Bond P.L. & Johnston A.W.B. (2005) Screening a wide host-range, waste-water metagenomic library in tryptophan auxotrophs of Rhizobium leguminosarum and of Escherichia coli reveals different classes of cloned trp genes. *Environmental microbiology,* **7,** 1927–1936. DOI: 10.1111/j.1462-2920.2005.00853.x.

Lim H.K., Kim I.-H., Nam H.Y., Shin S. & Hah S.S. (2013) Aptamer-Based Alternatives to the Conventional Immobilized Metal Affinity Chromatography for Purification of His-Tagged Proteins. *Analytical Letters,* **46,** 407–415. DOI: 10.1080/00032719.2012.721105.

Liu B., Wu S., Song Q., Zhang X. & Xie L. (2006) Two novel bacteriophages of thermophilic bacteria isolated from deep-sea hydrothermal fields. *Current microbiology,* **53,** 163–166. DOI: 10.1007/s00284-005-0509-9.

Liu D.V., Zawada J.F. & Swartz J.R. (2005) Streamlining Escherichia coli S30 extract preparation for economical cell-free protein synthesis. *Biotechnology progress,* **21,** 460–465. DOI: 10.1021/bp049789y.

Lorenz P. & Eck J. (2005) Metagenomics and industrial applications. *Nature reviews. Microbiology,* **3,** 510–516. DOI: 10.1038/nrmicro1161.

Lussier F.-X., Chambenoit O., Cote A., Hupe J.-F., Denis F., Juteau P., Beaudet R. & Shareck F. (2011) Construction and functional screening of a metagenomic library using a T7 RNA polymerase-based expression cosmid vector. *Journal of industrial microbiology & biotechnology,* **38,** 1321–1328. DOI: 10.1007/s10295-010-0915-2.

Mahmood T. & Yang P.-C. (2012) Western blot: technique, theory, and trouble shooting. *North American journal of medical sciences,* **4,** 429–434. DOI: 10.4103/1947-2714.100998.

Margulies M., Egholm M., Altman W.E., Attiya S., Bader J.S., Bemben L.A., Berka J., Braverman M.S., Chen Y.-J., Chen Z., Dewell S.B., Du L., Fierro J.M., Gomes X.V., Godwin B.C., He W., Helgesen S., Ho C.H., Irzyk G.P., Jando S.C., Alenquer M.L.I., Jarvie T.P., Jirage K.B., Kim J.-B., Knight J.R., Lanza J.R., Leamon J.H., Lefkowitz S.M., Lei M., Li J., Lohman K.L., Lu H., Makhijani V.B., McDade K.E., McKenna

M.P., Myers E.W., Nickerson E., Nobile J.R., Plant R., Puc B.P., Ronan M.T., Roth G.T., Sarkis G.J., Simons J.F., Simpson J.W., Srinivasan M., Tartaro K.R., Tomasz A., Vogt K.A., Volkmer G.A., Wang S.H., Wang Y., Weiner M.P., Yu P., Begley R.F. & Rothberg J.M. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature,* **437,** 376–380. DOI: 10.1038/nature03959.

Markov D., Naryshkina T., Mustaev A. & Severinov K. (1999) A zinc-binding site in the largest subunit of DNA-dependent RNA polymerase is involved in enzyme assembly. *Genes & Development,* **13,** 2439–2448.

Marks T.J. & Hamilton P.T. (2014) Characterization of a thermophilic bacteriophage of Geobacillus kaustophilus. *Archives of virology,* **159,** 2771–2775. DOI: 10.1007/s00705-014-2101-8.

Martinez A., Kolvek S.J., Yip C.L.T., Hopke J., Brown K.A., MacNeil I.A. & Osburne M.S. (2004) Genetically Modified Bacterial Strains and Novel Bacterial Artificial Chromosome Shuttle Vectors for Constructing Environmental Libraries and Detecting Heterologous Natural Products in Multiple Expression Hosts. *Applied and environmental microbiology,* **70,** 2452–2463. DOI: 10.1128/AEM.70.4.2452-2463.2004.

Martinez R. & Schwaneberg U. (2013) A roadmap to directed enzyme evolution and screening systems for biotechnological applications. *Biological research,* **46,** 395–405. DOI: 10.4067/S0716-97602013000400011.

Martino C., Kim S.-H., Horsfall L., Abbaspourrad A., Rosser S.J., Cooper J. & Weitz D.A. (2012) Protein expression, aggregation, and triggered release from polymersomes as artificial cell-like structures. *Angewandte Chemie (International ed. in English),* **51,** 6416–6420. DOI: 10.1002/anie.201201443.

Masters B.S., Stohl L.L. & Clayton D.A. (1987) Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7. *Cell,* **51,** 89–99.

McAllister W.T., Morris C., Rosenberg A.H. & Studier F.W. (1981) Utilization of bacteriophage T7 late promoters in recombinant plasmids during infection. *Journal of molecular biology,* **153,** 527–544. DOI: 10.1016/0022-2836(81)90406-x.

McAllister W.T. & Raskin C.A. (1993) The phage RNA polymerases are related to DNA polymerases and reverse transcriptases. *Molecular microbiology,* **10,** 1–6.

McMahon M.D., Guan C., Handelsman J. & Thomas M.G. (2012) Metagenomic analysis of Streptomyces lividans reveals host-dependent functional expression. *Applied and environmental microbiology,* **78,** 3622–3629. DOI: 10.1128/AEM.00044-12.

Messina E., Sorokin D.Y., Kublanov I.V., Toshchakov S., Lopatina A., Arcadi E., Smedile F., La Spada G., La Cono V. & Yakimov M.M. (2016) Complete genome sequence of 'Halanaeroarchaeum sulfurireducens' M27-SA2, a sulfur-reducing and acetate-oxidizing haloarchaeon from the deep-sea hypersaline anoxic lake Medee. *Standards in genomic sciences,* **11,** 35. DOI: 10.1186/s40793-016-0155-9.

Mikami S., Masutani M., Sonenberg N., Yokoyama S. & Imataka H. (2006) An efficient mammalian cell-free translation system supplemented with translation factors. *Protein expression and purification,* **46,** 348–357. DOI: 10.1016/j.pep.2005.09.021.

Miller R.B. & Karn R.C. (1980) A rapid spectrophotometric method for the determination of esterase activity. *Journal of Biochemical and Biophysical Methods,* **3,** 345–354. DOI: 10.1016/0165-022X(80)90043-3.

Mirete S., Morgante V. & Gonzalez-Pastor J.E. (2016) Functional metagenomics of extreme environments. *Current opinion in biotechnology,* **38,** 143–149. DOI: 10.1016/j.copbio.2016.01.017.

Moll P.R., Duschl J. & Richter K. (2004) Optimized RNA amplification using T7-RNA-polymerase based in vitro transcription. *Analytical biochemistry,* **334,** 164–174. DOI: 10.1016/j.ab.2004.07.013.

Moore T.C. & Shaner C.A. (1968) Synthesis of indoleacetic acid from tryptophan via indolepyruvic acid in cell-free extracts of pea seedlings. *Archives of biochemistry and biophysics,* **127,** 61321.

Mou X., Sun S., Edwards R.A., Hodson R.E. & Moran M.A. (2008) Bacterial carbon processing by generalist species in the coastal ocean. *Nature,* **451,** 708–711. DOI: 10.1038/nature06513.

Murakami K.S. (2015) Structural Biology of Bacterial RNA Polymerase. *Biomolecules,* **5,** 848–864. DOI: 10.3390/biom5020848.

Nallani M., Andreasson-Ochsner M., Tan C.-W.D., Sinner E.-K., Wisantoso Y., Geifman-Shochat S. & Hunziker W. (2011) Proteopolymersomes: in vitro production of a membrane protein in polymersome membranes. *Biointerphases,* **6,** 153–157. DOI: 10.1116/1.3644384.

Nevalainen K.M.H., Te'o V.S.J. & Bergquist P.L. (2005) Heterologous protein expression in filamentous fungi. *Trends in biotechnology,* **23,** 468–474. DOI: 10.1016/j.tibtech.2005.06.002.

Nevondo W. (2016) *Development of a high throughput cell-free metagenomic screening platform.* Dissertation, University of the Western Cape.

Ngara T.R. & Zhang H. (2018) Recent Advances in Function-based Metagenomic Screening. *Genomics, Proteomics & Bioinformatics,* **16,** 405–415. DOI: 10.1016/j.gpb.2018.01.002.

Nirenberg M.W. & Matthaei J.H. (1961) The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *The Journal of biological chemistry,* **47,** 1588–1602. DOI: 10.1073/pnas.47.10.1588.

Patel R., Mevada V., Prajapati D., Dudhagara P., Koringa P. & Joshi C.G. (2015) Metagenomic sequence of saline desert microbiota from wild ass sanctuary, Little Rann of Kutch, Gujarat, India. *Genomics data,* **3,** 137–139. DOI: 10.1016/j.gdata.2015.01.003.

Pazmiño D.E.T. & Fraaije M.W. (2007) Chapter 6 - Discovery, redesign and applications of Baeyer-Villiger monooxygenases. In: *Future directions in biocatalysis* (ed. by T. Matsuda), pp. 107–127. Elsevier, Amsterdam.

Pessela B.C.C., Torres R., Fuentes M., Mateo C., Filho M., Carrascosa A.V., Vian A., Garcia J.L., Guisan J.M. & Fernandez-Lafuente R. (2004) A simple strategy for the purification of large thermophilic proteins overexpressed in mesophilic microorganisms: application to multimeric enzymes from Thermus sp. strain T2 expressed in Escherichia coli. *Biotechnology progress,* **20,** 1507–1511. DOI: 10.1021/bp049785t.

Pitzler C. (2015) *Novel flow cytometer-based platforms for directed evolution.* Dissertation, RWTH Aachen.

Placido A., Hai T., Ferrer M., Chernikova T.N., Distaso M., Armstrong D., Yakunin A.F., Toshchakov S.V., Yakimov M.M., Kublanov I.V., Golyshina O.V., Pesole G., Ceci L.R. & Golyshin P.N. (2015) Diversity of hydrolases from hydrothermal vent sediments of the Levante Bay, Vulcano Island (Aeolian archipelago) identified by activity-based metagenomics and biochemical characterization of new esterases and an arabinopyranosidase. *Applied microbiology and biotechnology,* **99,** 10031–10046. DOI: 10.1007/s00253-015-6873-x.

Pleiss J.A., Derrick M.L. & Uhlenbeck O.C. (1998) T7 RNA polymerase produces 5' end heterogeneity during in vitro transcription from certain templates. *RNA (New York, N.Y.),* **4,** 1313–1317. DOI: 10.1017/s135583829800106x.

Popovic A., Tchigvintsev A., Tran H., Chernikova T.N., Golyshina O.V., Yakimov M.M., Golyshin P.N. & Yakunin A.F. (2015) Metagenomics as a Tool for Enzyme Discovery: Hydrolytic Enzymes from Marine-Related Metagenomes. *Advances in experimental medicine and biology,* **883,** 1–20. DOI: 10.1007/978-3-319-23603-2_1.

Pottkämper J., Barthen P., Ilmberger N., Schwaneberg U., Schenk A., Schulte M., Ignatiev N. & Streit W.R. (2009) Applying metagenomics for the identification of bacterial cellulases that are stable in ionic liquids. *Green Chemistry,* **11,** 957–965. DOI: 10.1039/B820157A.

Qin D. & Fredrick K. (2013) Analysis of polysomes from bacteria. *Methods in enzymology,* **530,** 159–172. DOI: 10.1016/B978-0-12-420037-1.00008-7.

Rappé M.S. & Giovannoni S.J. (2003) The uncultured microbial majority. *Annual review of microbiology,* **57,** 369–394. DOI: 10.1146/annurev.micro.57.030502.090759.

Reyes-Duarte D., Coscolín C., Martínez-Martínez M., Ferrer M. & García-Arellano H. (2018) Functional-Based Screening Methods for Detecting Esterase and Lipase Activity Against Multiple Substrates. *Methods in molecular biology (Clifton, N.J.),* **1835,** 109–117. DOI: 10.1007/978-1-4939-8672-9_4.

Reyes-Duarte D., Ferrer M. & Garcia-Arellano H. (2012) Functional-based screening methods for lipases, esterases, and phospholipases in metagenomic libraries. *Methods in molecular biology (Clifton, N.J.),* **861,** 101–113. DOI: 10.1007/978-1-61779-600-5_6.

Rhee J.-K., Ahn D.-G., Kim Y.-G. & Oh J.-W. (2005) New thermophilic and thermostable esterase with sequence similarity to the hormone-sensitive lipase family, cloned from a metagenomic library. *Applied and environmental microbiology,* **71,** 817–825. DOI: 10.1128/AEM.71.2.817-825.2005.

Richter C. (2014) *Optimization of protein production development using a combination of cell-free expression and high-throughput protein analysis.* Dissertation, Karlsruhe.

Rideau E., Dimova R., Schwille P., Wurm F.R. & Landfester K. (2018) Liposomes and polymersomes: a comparative review towards cell mimicking. *Chemical Society Reviews,* **47,** 8572–8610. DOI: 10.1039/C8CS00162F.

Robert Schenck *Thermus aquaticus RNA Polymerase.* URL https://collab.its.virginia.edu/access/content/group/f85bed6c-45d2-4b18-b868-6a2353586804/2/Ch29_Schenck_R_RNA_Polymerase_(Thermus_aquaticus)-_-/index.html.

Roberts I.M. (1985) Hydrolysis of 4-methylumbelliferyl butyrate: A convenient and sensitive fluorescent assay for lipase activity. *Lipids,* **20,** 243–247. DOI: 10.1007/BF02534195.

Roh C. & Schmid R.D. (2013) Isolation of an organic solvent-tolerant lipolytic enzyme from uncultivated microorganism. *Applied biochemistry and biotechnology,* **171,** 1750–1758. DOI: 10.1007/s12010-013-0464-z.

Safarik I. & Safarikova M. (2004) Magnetic techniques for the isolation and purification of proteins and peptides. *Biomagnetic Research and Technology,* **2,** 7. DOI: 10.1186/1477-044X-2-7.

Sarmiento F., Peralta R. & Blamey J.M. (2015) Cold and Hot Extremozymes: Industrial Relevance and Current Trends. *Frontiers in Bioengineering and Biotechnology,* **3. DOI:** 10.3389/fbioe.2015.00148.

Shanks O.C., Santo Domingo J.W., Lamendella R., Kelty C.A. & Graham J.E. (2006) Competitive metagenomic DNA hybridization identifies host-specific microbial genetic markers in cow fecal samples. *Applied and environmental microbiology,* **72,** 4054–4060. DOI: 10.1128/AEM.00023-06.

Shin J. & Noireaux V. (2010) Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70. *Journal of biological engineering,* **4,** 8. DOI: 10.1186/1754-1611-4-8.

Shrestha P., Holland T.M. & Bundy B.C. (2012) Streamlined extract preparation for Escherichia coli-based cell-free protein synthesis by sonication or bead vortex mixing. *BioTechniques,* **53,** 163–174. DOI: 10.2144/0000113924.

Shum H.C., Kim J.-W. & Weitz D.A. (2008) Microfluidic fabrication of monodisperse biocompatible and biodegradable polymersomes with controlled permeability. *Journal of the American Chemical Society,* **130,** 9543–9549. DOI: 10.1021/ja802157y.

Simon C. & Daniel R. (2009) Achievements and new knowledge unraveled by metagenomic approaches. *Applied microbiology and biotechnology,* **85,** 265–276. DOI: 10.1007/s00253-009-2233-z.

Smith J.D. & Robinson A.S. (2002) Overexpression of an archaeal protein in yeast: secretion bottleneck at the ER. *Biotechnology and bioengineering,* **79,** 713–723. DOI: 10.1002/bit.10367.

Smutzer G. (2001) Cell-free Transcription and Translation. *The Scientist*.

Sousa R. & Mukherjee S. (2003) T7 RNA Polymerase. In: *Progress in Nucleic Acid Research and Molecular Biology,* pp. 1–41. Academic Press.

Spirin A.S., Baranov V.I., Ryabova L.A., Ovodov S.Y. & Alakhov Y.B. (1988) A continuous cell-free translation system capable of producing polypeptides in high yield. *Science (New York, N.Y.),* **242,** 1162–1164.

Stech M., Quast R.B., Sachse R., Schulze C., Wüstenhagen D.A. & Kubick S. (2014) A continuous-exchange cell-free protein synthesis system based on extracts from cultured insect cells. *PloS one,* **9,** e96635. DOI: 10.1371/journal.pone.0096635.

Streit W.R., Daniel R. & Jaeger K.-E. (2004) Prospecting for biocatalysts and drugs in the genomes of non-cultured microorganisms. *Current opinion in biotechnology,* **15,** 285–290. DOI: 10.1016/j.copbio.2004.05.006.

Streit W.R. & Schmitz R.A. (2004) Metagenomics--the key to the uncultured microbes. *Current opinion in microbiology,* **7,** 492–498. DOI: 10.1016/j.mib.2004.08.002.

Tchigvintsev A., Tran H., Popovic A., Kovacic F., Brown G., Flick R., Hajighasemi M., Egorova O., Somody J.C., Tchigvintsev D., Khusnutdinova A., Chernikova T.N., Golyshina O.V., Yakimov M.M., Savchenko A., Golyshin P.N., Jaeger K.-E. & Yakunin A.F. (2015) The environment shapes microbial enzymes: five cold-active and salt-resistant carboxylesterases from marine metagenomes. *Applied microbiology and biotechnology,* **99,** 2165–2178. DOI: 10.1007/s00253-014-6038-3.

Terrón-González L., Medina C., Limón-Mortés M.C. & Santero E. (2013) Heterologous viral expression systems in fosmid vectors increase the functional analysis potential of metagenomic libraries. *Scientific Reports,* **3,** 1107. DOI: 10.1038/srep01107.

Torsvik V., Goksoyr J. & Daae F.L. (1990) High diversity in DNA of soil bacteria. *Applied and environmental microbiology,* **56,** 782–787.

Tringe S.G. & Rubin E.M. (2005) Metagenomics: DNA sequencing of environmental samples. *Nature reviews. Genetics,* **6,** 805–814. DOI: 10.1038/nrg1709.

Turner N.J. & Truppo M.D. (2013) Biocatalysis enters a new era. *Current opinion in chemical biology,* **17,** 212–214. DOI: 10.1016/j.cbpa.2013.02.026.

Tursun B., Cochella L., Carrera I. & Hobert O. (2009) A toolkit and robust pipeline for the generation of fosmid-based reporter genes in C. elegans. *PloS one,* **4,** e4625. DOI: 10.1371/journal.pone.0004625.

Uchiyama T., Abe T., Ikemura T. & Watanabe K. (2005) Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nature biotechnology,* **23,** 88–93. DOI: 10.1038/nbt1048.

Uchiyama T. & Miyazaki K. (2009) Functional metagenomics for enzyme discovery: challenges to efficient screening. *Current opinion in biotechnology,* **20,** 616–622. DOI: 10.1016/j.copbio.2009.09.010.

Uchiyama T. & Watanabe K. (2007) The SIGEX scheme: high throughput screening of environmental metagenomes for the isolation of novel catabolic genes. *Biotechnology & genetic engineering reviews,* **24,** 107–116.

Uchiyama T. & Watanabe K. (2008) Substrate-induced gene expression (SIGEX) screening of metagenome libraries. *Nature protocols,* **3,** 1202–1212. DOI: 10.1038/nprot.2008.96.

Vassylyev D.G., Sekine S.-i., Laptenko O., Lee J., Vassylyeva M.N., Borukhov S. & Yokoyama S. (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature,* **417,** 712. DOI: 10.1038/nature752.

Vassylyeva M.N., Lee J., Sekine S.I., Laptenko O., Kuramitsu S., Shibata T., Inoue Y., Borukhov S., Vassylyev D.G. & Yokoyama S. (2002) Purification, crystallization and initial crystallographic analysis of RNA polymerase holoenzyme from Thermus thermophilus. *Acta crystallographica. Section D, Biological crystallography,* **58,** 1497–1500. DOI: 10.1107/S0907444902011770.

Venter J.C., Remington K., Heidelberg J.F., Halpern A.L., Rusch D., Eisen J.A., Wu D., Paulsen I., Nelson K.E., Nelson W., Fouts D.E., Levy S., Knap A.H., Lomas M.W., Nealson K., White O., Peterson J., Hoffman J., Parsons R., Baden-Tillson H., Pfannkoch C., Rogers Y.-H. & Smith H.O. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science (New York, N.Y.),* **304,** 66–74. DOI: 10.1126/science.1093857.

Volyanik E.V., Dalley A., Mckay I.A., Leigh I., Williams N.S. & Bustin S.A. (1993) Synthesis of Preparative Amounts of Biologically Active Interleukin-6 Using a Continuous-Flow Cell-Free Translation System. *Analytical biochemistry,* **214,** 289–294. DOI: 10.1006/abio.1993.1490.

Wang G.Y., Graziani E., Waters B., Pan W., Li X., McDermott J., Meurer G., Saxena G., Andersen R.J. & Davies J. (2000) Novel natural products from soil DNA libraries in a streptomycete host. *Organic letters,* **2,** 2401–2404.

Weidner M., Taupp M. & Hallam S.J. (2010) Expression of recombinant proteins in the methylotrophic yeast Pichia pastoris. *Journal of visualized experiments : JoVE. DOI:* 10.3791/1862.

Wenda S., Illner S., Mell A. & Kragl U. (2011) Industrial biotechnology—the future of green chemistry? *Green Chemistry,* **13,** 3007. DOI: 10.1039/c1gc15579b.

Werner F. & Grohmann D. (2011) Evolution of multisubunit RNA polymerases in the three domains of life. *Nature Reviews Microbiology,* **9,** 85. DOI: 10.1038/nrmicro2507.

Wexler M., Bond P.L., Richardson D.J. & Johnston A.W.B. (2005) A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase. *Environmental microbiology,* **7,** 1917–1926. DOI: 10.1111/j.1462-2920.2005.00854.x.

Whitman W.B., Coleman D.C. & Wiebe W.J. (1998) Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences of the United States of America,* **95,** 6578–6583.

Wichłacz A., Łęgiewicz M. & Ciesiołka J. (2004) Generating in vitro transcripts with homogenous 3' ends using trans-acting antigenomic delta ribozyme. *Nucleic Acids Research,* **32,** e39. DOI: 10.1093/nar/gnh037.

Xie W., Wang F., Guo L., Chen Z., Sievert S.M., Meng J., Huang G., Li Y., Yan Q., Wu S., Wang X., Chen S., He G., Xiao X. & Xu A. (2011) Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. *The ISME journal,* **5,** 414–426. DOI: 10.1038/ismej.2010.144.

Xu M., Xiao X. & Wang F. (2008) Isolation and characterization of alkane hydroxylases from a metagenomic library of Pacific deep-sea sediment. *Extremophiles : life under extreme conditions,* **12,** 255–262. DOI: 10.1007/s00792-007-0122-x.

Yoshida S., Hiraga K., Takehana T., Taniguchi I., Yamaji H., Maeda Y., Toyohara K., Miyamoto K., Kimura Y. & Oda K. (2016) A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science (New York, N.Y.),* **351,** 1196–1199. DOI: 10.1126/science.aad6359.

Zemella A., Thoring L., Hoffmeister C. & Kubick S. (2015) Cell-Free Protein Synthesis: Pros and Cons of Prokaryotic and Eukaryotic Systems. *ChemBioChem,* **16,** 2420–2431. DOI: 10.1002/cbic.201500340.

8 Appendix

Nr	Metagenome Name	qeuencing Cent	IMG Genome ID	Assembled size (Bp)	Gene count	IMG submission ID	Viral RNA Poly Hits (score>300)	Hits / Gb
16	Freshwater to marine saline gradient viral communities from Chesapeake Bay - CB_1508_1M Viral MetaG	DOE Joint Geno	3300007539	584,438,297	1,325,157	78177	32	54.75
15	Marine viral communities from the Subarctic Pacific Ocean - 10_ETSP_OMZ_AT15264 metaG	DOE Joint Geno	3300006754	624,766,230	1,370,303	75053	23	36.81
19	Marine viral communities from the Global Malaspina Expedition - Malaspina viral metaG DeepMed_b05	DOE Joint Geno	3300008219	467,982,578	942,640	83499	10	21.37
2	western Arctic Ocean co-assembly	DOE Joint Geno	3300010883	5,416,576,525	6,760,737	92899	98	18.09
13	Freshwater to marine salinity gradient microbial communities from Chesapeake Bay, USA - CPBay_Spr_31_0.8_	DOE Joint Geno	3300013010	1,523,809,060	3,122,886	111799	25	16.41
12	Marine microbial communities from the West Antarctic Peninsula - Coastal water metaG006-DNA	DOE Joint Geno	3300006165	1,399,272,735	4,615,274	71583	15	10.72
6	Freshwater to marine salinity gradient microbial communities from Chesapeake Bay, USA - CPBay_Sum_0.6_0_	DOE Joint Geno	3300010354	2,908,194,785	6,293,646	94474	20	6.88
14	Aqueous microbial communities from the Delaware River and Bay under freshwater to marine salinity gradien	DOE Joint Geno	3300006641	1,534,215,455	4,217,285	74787	8	5.21
4	Marine eukaryotic phytoplankton communities from the Atlantic Ocean - Atlantic ANT 2 Metagenome	DOE Joint Geno	3300012953	3,480,224,628	7,021,553	109643	17	4.88
11	Pelagic marine microbial communities from North Sea - COGITO_mtgs_120405	DOE Joint Geno	3300009071	1,375,835,412	2,733,107	87173	4	2.91
1	Pelagic marine sediment microbial communities from the LTER site Helgoland, North Sea, for post-phytoplank	DOE Joint Geno	3300004097	5,431,569,150	13,267,245	58716	15	2.76
25	Biogas_NH3_control	MPI Plön		1,129,848,321	2,198,445		3	2.66
9	Pelagic marine microbial communities from North Sea - COGITO_mtgs_120412	DOE Joint Geno	3300009508	1,570,047,347	2,969,905	88679	4	2.55
21	GS000c - Sargasso Station 3 Co Assebmly	J. Craig Venter I	3300002040	9,006,699,085	16,106,649	21831	22	2.44
7	Marine microbial communities from the Costa Rica Dome - CRUD Field 142mm St17 metaG	DOE Joint Geno	3300012928	2,630,446,705	5,914,726	108703	6	2.28
3	Marine microbial communities from the Costa Rica Dome - CRUD Field 142mm St18 metaG	DOE Joint Geno	3300012954	4,274,994,468	9,545,890	109645	9	2.11
5	Marine eukaryotic phytoplankton communities from Arctic Ocean - Arctic Ocean- Svalbard ARC20M Metagenor	DOE Joint Geno	3300009544	3,081,152,087	5,466,259	88943	6	1.95
22	Adult Elephant (Nele)	Uni Hamburg		929,519,943	1,005,402		1	1.08
26	Biogas_NH3_shock	MPI Plön		997,483,007	1,985,098		1	1.00
24	BioPara Biogas May 2013	Uni Hamburg		1,255,011,327	2,011,807		1	0.80
29	Fischtank Sample 2	MPI Plön		1,479,983,842	2,378,766		1	0.68
20	Marine microbial and viral communities from oxygen minimum zone, Eastern Pacific Ocean - ETNP201302SV91	DOE Joint Geno	3300006166	2,041,951,888	5,508,781	71568	1	0.49
8	Pelagic marine sediment microbial communities from the LTER site Helgoland, North Sea, for post-phytoplank	DOE Joint Geno	3300004113	2,131,370,674	6,051,073	59215	0	0.00
10	Marine hydrothermal vent microbial communities from Guaymas Basin, Gulf of California to study Microbial Di	DOE Joint Geno	3300010332	1,080,322,849	2,490,314	94478	0	0.00
17	Marine algal microbial communities from Sidmouth, United Kingdom - Sidmouth_Asex2 metaG	DOE Joint Geno	3300009421	552,079,437	92,860	87303	0	0.00
18	Oil polluted marine microbial communities from Coal Oil Point, Santa Barbara, California, USA - Sample 2 (Crud	DOE Joint Geno	3300001685	495,862,225	1,143,552	18818	0	0.00
23	Adult Elephant (Philipp)	Uni Hamburg		2,640,727,002	4,744,007		0	0.00
27	Cow Rumen (Hess et al.)			1,500,000,000	2,083,556		0	0.00
28	Fischtank Sample 1	MPI Plön		2,541,572,205	4,440,847		0	0.00
30	Fischtank Sample 3	MPI Plön		1,874,257,931	3,107,831		0	0.00
				65,960,215,198	130,915,601	Total hits	322	
				~66Gb DNA		Full length hits	107	

Figure 29: Metagenomic search for viral RNAP candidates.

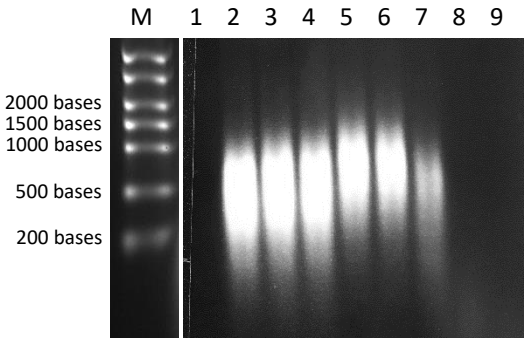

Figure 30. *In vitro* transcriptions with the thermo T7 RNA-polymerase from "Toyobo".

RNA was run in a 1.2% agarose gel containing 0.7% (v/v) formaldehyde and stained with ethidium bromide, which is included in the loading dye (ThermoFisher Scientific GmbH, Schwerte, Germany). M: RNA ladder (2 µl), 1: Negative control without DNA (45°C), 2: 37°C, 3: 40.1°C, 4: 42.3°C, 5: 45°C, 6: 47.7°C, 7: 50.5°C, 8: 53°C, 9: 55.2°C (14 µl of each *in vitro* transcription, GFP used as template).

Note:

During the preparation of this dissertation, a phd thesis with a very similar research goal emerged. Walter Nevondo succeeded in establishing the so-called metagenomic *in vitro* compartmentalization (mIVC-FACS), which is based on CFPS with cell extract of *Rhodococcus erythropolis* in double-emulsions.

9 Declaration on oath

I hereby declare, on oath, that I have written the present dissertation by my own and have not used any other than the acknowledged resources and aids.

Eidesstattliche Versicherung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Hamburg, 03.2020

_____

Mareike Janus

10 English language assessment

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

**Fakultät für
Mathematik, Informatik
und Naturwissenschaften**

UHH
Studienbüro Biologie
Petra Ziemer, Raum CvL/E.522
Ohnhorststr. 18

22609 Hamburg

Prof. Dr. Andrew Torda
Zentrum für Bioinformatik
Bundesstr. 43
20146 Hamburg

Tel. +49 40 42838-7331
Fax +49 40 42838-7332
torda@zbh.uni-hamburg.de
www.zbh.uni-hamburg.de/torda

19.02.20

English Language Assessment:

JANUS, Mareike

"Development of a ... screening of metagenomic DNA"

I have read books by Patrick White, Samuel Beckett, William Golding, Doris Lessing and seen films written by Harold Pinter. They all received Nobel prizes for literature, but few were as easy to read as Frau Janus's thesis.

The language in the thesis is a delight.

Andrew Torda

11 **Danksagung**

Zuallererst möchte ich mich bei Prof. Julia Kehr für die Übernahme der Betreuung und die großartige Unterstützung bedanken.

Ich bedanke mich bei Prof. Stefan Wittke für die freundliche Übernahme der Rolle als zweiter Gutachter.
Dankeschön an Prof. Andrew Torda für das tolle “English language assessment”.

Ein besonderer Dank gilt meinen ehemaligen Kollegen der Mikrobiologie & Biotechnologie, die mich in der intensiven Zeit immer unterstützt und motiviert haben. Eva, Katha, Pablo, Domm, es war eine spannende Zeit mit den ein oder anderen (Lach-)Tränen.

Ich danke meinen Mädels - Aline, Scarlett, Ila- für die Unterstützung, die vielen aufmunternden Worte und ihre Freundschaft.

Ich bin Anton sehr dankbar für die viele Unterstützung, die Abende mit Formatierungs-Problemen und richtig gutem Essen.

Und zu guter Letzt möchte ich mich bei meinen Maids bedanken: Maike, Andrea, ohne die das Projekt “Doktortitel” gescheitert wäre. Ihr habt mich motiviert, aufgebaut, wir waren frustriert, genervt, haben uns totgelacht, haben gesungen und getanzt.

DANKE.

“I was served lemons, but I made lemonade.” Hattie White