Function-Based Searches for Selected Phosphotransferases and Establishing *in Vitro* Transcription Platform for Cell-Free Metagenomics

Dissertation

with the aim of achieving the degree of Doctor rerum naturalium (Dr. rer. nat.) at the

Faculty of Mathematics, Informatics and Natural Sciences

Department of Biology

of the Universität Hamburg

Submitted by

Birhanu Mekuaninte Kinfu

2017 in Hamburg

The following evaluators recommend the admission of this dissertation:

Professor Dr. Wolfgang Streit

Professor Dr. Andreas Liese

Date of oral defense: 24.05.2018

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.

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.

Hamburg, 28.12.2017

_____

Birhanu Mekuaninte Kinfu

Dedicated to my parents for their endless love and sacrifices

Contributions to the quoted articles

Kinfu BM, Jahnke M, Janus M, Besirlioglu V, Roggenbuck M, Meurer R, Vojcic L, Borchert M, ... Wolfgang RS (2017) **Recombinant RNA Polymerase from *Geobacillus* sp. GHH01 as tool for rapid generation of metagenomic RNAs using in vitro technologies**. *Biotechnology and Bioengineering* **114**(12):2739-2752

- Planned and co-designed experiments, executed experimental work of His_6-tagged expression and purification of RNA polymerase subunits and transcription elongation factors, reconstitution efforts and biochemical characterizations of the RNAP, and its transcription activities on meta/genomic template DNAs. Co-written (majority of) the manuscript

Hardt N, **Kinfu BM**, Chow J, Schoenenberger B, Streit WR, Obkircher M & Wohlgemuth R (2017) **Biocatalytic Asymmetric Phosphorylation Catalyzed by Recombinant Glycerate-2-Kinase**. *Chembiochem: a European journal of chemical biology* **18**(15):1518-1522

- Planned and performed experimental work on cloning, heterologous expression, purification and biochemical analysis of glycerate-2-kinase
- Discussion of results and written part of the manuscript

Molla GS, **Kinfu BM**, Chow J, Streit W, Wohlgemuth R & Liese A (2017) **Bioreaction Engineering Leading to Efficient Synthesis of L-Glyceraldehyde-3-Phosphate**. *Biotechnology Journal* **12**(3):1600625

- Discussion of results, Read and manuscript checkup
- Planned and executed experimental work on searching and multiple expression attempts of acid-stable Archaeal glycerol kinases for use in glyceraldehyde phosphorylation

Kinfu BM, Chow J, Molla GS, Liese A, Wohlgemuth R & Streit WR. **Establishing selective screening method and functional search for ß-glycerophosphate-involving activities**. Manuscript under preparation

- Planned multiple enzyme search approaches and performed the experimental works. Prepared the draft manuscript

Book chapter (contributed)

- Gauss D, Schönenberger B, Molla GS, **Kinfu BM**, Chow J, Liese A, Streit WR & Wohlgemuth R (2016) Biocatalytic phosphorylation of metabolites, in Applied biocatalysis: from fundamental science to industrial applications (eds Hilterhaus L, Liese A, Kettling U. & Antranikian G. *et. al.*), Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.

Table of contents

Abstract

The shift in industrial production of specialty chemicals, pharmaceuticals, and metabolites towards biocatalysis has been the center of interest for several decades. The progress, however, is largely challenged due to the limited enzyme diversity to fulfil synthetic requirements. This has channeled a significant effort towards screening enzymes with unique features especially from extreme environments. While this has a huge potential to expand the enzyme toolkits, the existing functional screening methods are severely limited in tapping nature's unlimited potential. The aim of this study was to search for selected phosphotranferases, including rare and difficult-to-express enzymes, for use in biocatalytic phosphorylation of target metabolites. It thereby aimed at establishing *in vitro* transcription system to lay the foundation for cell-free functional metagenomic screening platform.

Functional screening of enzymes for phosphorylation of glycerol and some of its derivatives is presented here. While a previous report on glycerol-2-phosphotransferse activity was challenged, functional search of ß-glycerophosphate-involving activities were screened from environmental samples after enrichment. The use of a highly enantio-selective kinase integrated with energy regeneration system was also established for sustainable biocatalytic production of optically pure D-glycerate-2-phosphate and product analysis thereafter. Additionally, multiple expression attempts of acid-stable glycerol kinases from the most extreme acidophilic microbes ever reported was performed for biocatalytic synthesis of the very labile glyceraldehyde-3-phosphate. Alongside, engineering process parameters led to the use of an existing kinase for efficient synthesis of L-glyceraldehyde-3-

phosphate which suggests to strategically consider adapting available enzymes for desired applications.

In addition to the successes achieved here, the search for phosphotransferases faced a serious challenge of protein expression. Not surprisingly, any function-based screening approach including metagenomics suffers a similar impediment where difference in codon usage often leads to biased transcription ultimately resulting poor protein expression. To this end, an *in vitro* transcription system based on a recombinant RNA polymerase (RNAP) from *Geobacillus* sp. GHH01 was established. The RNAP can easily be reconstituted using a new 'thermo-reconstitution' method which requires only a brief thermal treatment saving time and resources compared to the classical method of reconstitution.

One of the striking features of this RNAP is its ability to successfully initiate transcription from diverse recognition factors as evidenced by the amount of transcript generated when employing meta/genomic DNA templates from different environmental samples and organisms of varying phylogenetic origins. While phage-derived RNAPs are limited due to their absolute stringency towards own recognition factors, the *Geobacillus* RNAP has a significant comparative advantage over the tested commercial *Escherichia coli* RNAP in terms of both the transcript generated and other important features. The *Geobacillus* RNAP has remarkable working and storage stabilities maintaining its maximum activity up to 55°C and can simply be stored at non-freezing temperatures. In this work, cell-free metagenomic screening and expression platform is suggested to overcome the multitude of challenges associated with host systems used for functional metagenome searches.

1 Introduction

1.1 Enzyme search and biocatalytic metabolite phosphorylation

Phosphorylation is a pivotal process that life has evolved to be impossible without. It is important in the biosynthesis of building blocks of macromolecules such as nucleic acids, phosphoproteins, and phospholipids as well as small molecules such as cofactors and several metabolites (Gauss*, et al.*, 2016, Westheimer, 1987). Not surprisingly, the ubiquitous presence of phosphorylated biomolecules is suggested by the vast cellular abundance of phosphotransferases/kinases catalyzing their biosynthesis (Gauss*, et al.*, 2016, Knowles, 1980, Westheimer, 1987, Wohlgemuth*, et al.*, 2017). Phosphorylation reactions are also immensely important in industry for the synthetic production of specialty chemicals, pharmaceuticals, and metabolites used in ranges of applications such as the study and characterization of metabolic enzyme functions and biochemical pathways, and as potential pharmaceuticals in biomedical research (Gauss*, et al.*, 2016, Wohlgemuth, 2009).

Although the number of commercially available phosphorylated metabolites has increased over time, many are synthesized solely via chemical methods and some are not accessible in their pure form yet. Chemical phosphorylation production often involves lengthy multi-step processes of protection and deprotection steps with the reputation of utilizing toxic reagents (Abderhalden and Eichwald, 1918, Ballou and Fischer, 1954, Ballou and Fischer, 1955, Fischer and Baer, 1941, Gray and Barker, 1971, Hartman and Wold, 1967, Kulkarni, 2013, Serianni*, et al.*, 1979). In addition, it usually lacks the required chemo-, regio- and stereo-selectivity, generating rather a mix of closely related by-products. This further complicates

downstream processing and is often unable to deliver products with the required optical purity.

Direct and straightforward enzymatic syntheses approaches have been the center of interest for production of phosphorylated metabolites and other biomolecules in general. On one hand, the simplicity and inherent specificity (both regio- and stereo-selectivity) make biocatalysts the best alternatives to chemical synthesis. On the other hand, the interesting promiscuity of enzymes has led to their use in catalyzing conversion of other substrate analogs yet maintaining the required level of specificity. For example, glycerol-3-kinase produces only L-glycerol-3-phosphate but not the D-optical isomer, while catalyzing selective phosphorylation of glyceraldehyde, dihydroxyacetone, and ranges of other glycerol analogs as well (Crans and Whitesides, 1985a, Crans and Whitesides, 1985b, Gauss*, et al.*, 2014).

Industrial applications of biophosphorylation have been limited for reasons of stability, cofactor requirements and lack of suitable functional enzymatic screening and analytical methods. The most decisive factors here are product stability and purity. Glyceraldehyde-3-phosphate, an intermediary in several metabolic pathways, is the best example where its stability only below pH 4 has been the biggest challenge for its industrial biosynthesis (Gauss*, et al.*, 2014, Gauss*, et al.*, 2016). As its D- and L- isomers have different biological properties and roles, the synthesis of enantiopure products of these compounds is very important. In fact, this is where the powerful chiral-selectivity nature of enzymes becomes super-advantageous. Accordingly, industrial biosynthesis of such metabolites has two general options to follow. One is to use “beat-the-clock” strategy of production with existing enzymes by optimizing reaction conditions and engineering process parameters for maximum

speed of conversion reaction and product recovery in compromise to a tolerable level of product lose.

The second approach is looking for unique enzymes from extremophiles which potentially suit the stability requirements of bioconversion processes. Since the reports of life in extreme environments, there has been a growing interest in biotechnological research and applications. The foremost important portions of these extremophiles in relation to phosphorylation are the extreme acidophiles especially Archaea. The two sister species *Picrophilus torridus* and *P. oshimae* are known not only to grow even in negative pH values but also to be able to maintain their intracellular pH the lowest known so far (Fütterer*, et al.*, 2004, Slonczewski*, et al.*, 2009). This is particularly important to mine acid-stable enzymes suitable for the synthesis of pH-sensitive phosphorylated metabolites primarily glyceraldehyde phosphate. Although it is highly promising with already proven success, applying extremozymes requires overcoming the often encountering challenge of protein expression. It is therefore important to consider improving expression host or more conveniently customize cell-free protein expression systems where host-related limitations can possibly be bypassed.

1.2 Phosphoryl donors for industrial biocatalytic conversions

High energy phosphate compounds are inevitable players in life processes and industrial biocatalysis driving many unfavorable biochemical reactions to meet the required energy balance. ATP is the preferred phosphoryl donor in most of the natural and biosynthetic reactions. However, there are other high energy reservoirs in nature which can potentially be used as alternative energy sources (Gauss*, et al.*, 2016). Best examples include phosphoenolpyruvate (Crans and Whitesides, 1985b,

Schoenenberger, *et al.*, 2017), inorganic polyphosphate (Kameda, *et al.*, 2001, Shiba, *et al.*, 2000), tripolyphosphate (Krawiec, *et al.*, 2003), other nucleotide triphosphates (NTP) (Hiles and Henderson, 1972, Ling and Lardy, 1954), acetylphosphate (Li, *et al.*, 2001), pyrophosphate and carbamylphosphate (Asano, *et al.*, 1999), methoxycarbonyl phosphate (Faber, 2011), and phosphoramidate (Fujimoto and Smith, 1960). There are also few interesting reports where existing enzymes have been engineered in order to shift their energy source preference from ATP to other cheaper and more stable phosphoryl donors (Nakamichi, *et al.*, 2013, Sanchez-Moreno, *et al.*, 2015).

The common practice in industries is using energy regeneration in coupled reaction systems by which ATP is sustainably formed using other cheaper and/or more stable high energy phosphoryl donors (Andexer and Richter, 2015, Crans, *et al.*, 1987, Faber, 2011, Wang, *et al.*, 2017). ATP regeneration from phosphorylation of ADP using phosphoenolpyruvate as a phosphoryl donor and catalyzed by pyruvate kinase is one of the commonly applied systems (Crans and Whitesides, 1985b, Gauss, *et al.*, 2014, Schoenenberger, *et al.*, 2017). The system has the advantage over the other handful ATP regeneration systems due to the high stability property of phosphoenolpyruvate in aqueous solution reactions, especially for slower bioconversions.

Recently, energy regeneration is also sought in cell-free protein expression systems where traces of endogenous metabolic machineries from cell extracts are exploited to sustainably produce ATP from cheaper substrates such as glucose (Kim and Swartz, 2001). This has transformed cell-free protein production scheme to be

more stable and feasible especially for continuous protein production and recovery (Kim and Kim, 2009, Whittaker, 2013).

1.3 Functional metagenomics and mining for biocatalysts

Metagenomics is a powerful tool for direct analysis of the identity and function of DNAs from complex environments without the need to culture source organisms. It has opened the opportunity to investigate a large portion of nature's biotechnological potential which has been out of touch with the classical culture-dependent approaches (Cowan, 2000, Streit and Schmitz, 2004). This analysis of total or enriched environmental samples is done based on either sequence homology or function-dependent approaches. Unlike sequence based metagenomics which depends on deducted consensus features from conserved sequences, functional search offers the opportunity to look for truly novel enzymes based on their activity/phenotype for known or entirely new functions. Functional screening has also the advantage to identify full-length coding genes or gene clusters responsible for enzymes or entire metabolic pathway.

Since its introduction in the early 1990s, functional metagenomics have been used for mining of a number of enzymes and other biomolecules from diverse environmental samples with a remarkable success (e.g. Boehmwald*, et al.*, 2016, Culligan*, et al.*, 2014, DeCastro*, et al.*, 2016, Ferrer*, et al.*, 2016, Handelsman, 2004, Iqbal*, et al.*, 2012, Santana-Pereira and Liles, 2017, Steele*, et al.*, 2009, Streit*, et al.*, 2004, Uchiyama and Miyazaki, 2009, Wilson and Piel, 2013). Metagenomics can also be effectively combined with enrichment of environmental samples using, for example, substrate selection which favors the growth of organisms with targeted genes/traits to enhance the chance of positive hits (Graham*, et al.*, 2011, Jacquiod*,*

et al., 2013, Knietsch*, et al.*, 2003). Functional metagenomics has seen recent developments in screening platforms such as droplet-based microfluidics technologies creating isolated compartments for independent sorting and analysis (Colin*, et al.*, 2015), as well as the use of substrate- and product-induced gene expression (SIGEX and PIGEX, respectively) approaches in high throughput functional screening (HTS) setups (Uchiyama*, et al.*, 2005, Uchiyama and Miyazaki, 2010).

1.4 Challenges and prospects of function-based metagenomics

Despite its applauded success and use of improved mining techniques, the much-anticipated promise of functional metagenomics is heavily hampered primarily by expression problems. Inefficient transcription and/or translation of target genes due to differences in codon usage with the surrogate host organism, incorrect folding and lack of appropriate secretion systems for gene products contribute to the average low number of positive hits from large-sized libraries (Ekkers*, et al.*, 2012, Kudla*, et al.*, 2009, Lam*, et al.*, 2015, Steele*, et al.*, 2009, Uchiyama and Miyazaki, 2009). More precisely, Gabor and colleagues estimated that the commonly used *E. coli* host expresses only about 40% in average of metagenome-derived genes from a total of 32 prokaryote genomes (Gabor*, et al.*, 2004). Worth noting is that the transcription profile was highly biased against genes originated from distantly related organisms.

The current approach to tackle expression problems in metagenomic screening focusses primarily on the strategy of diversifying alternative customized surrogate systems along with developing broad-host-range vectors for wider compatibility. Accordingly, alternative heterologous microbial host systems able to

uptake foreign DNA material and process to yield target activities have been developed based on different microbial genera (Lewin*, et al.*, 2017, Liebl*, et al.*, 2014). These include alternative host strains from *Pseudomonas putida*, *Streptomyces lividans*, *Sinorhizobium meliloti*, *Rhizobium leguminosarum, Desulfovibiro* sp., *Bacillus* and *Saccharomyces, Salmonella enterica*, *Serratia marcescens*, *Vibrio vulnificus* and *Enterobacter nimipressuralis*, *Ralstonia metallidurans Thermus thermophilus, Agrobacterium tumefaciens* (Craig*, et al.*, 2010, Kakirde*, et al.*, 2011, Katzke*, et al.*, 2017, Leis*, et al.*, 2015, Li*, et al.*, 2005, Maroti*, et al.*, 2009, Martinez*, et al.*, 2004, Rousset*, et al.*, 1998, Schallmey*, et al.*, 2011, Wang*, et al.*, 2006, Wexler*, et al.*, 2005). As the expression profile of each organism towards foreign DNAs is generally different from one another, none of these can qualify to be the best functional metagenomic hosts for screening target biomolecules from complex samples.

Further dissecting the expression bottleneck, transcription of metagenome-derived genes is indeed a decisive limiting step in functional searches which was revealed based on comparative transcriptome analyses of *E. coli* EPI300 carrying different metagenome clones (Jürgensen, 2014, Liebl*, et al.*, 2014). The *E. coli* host strain has a strong bias towards its own gene compared to the different fosmid inserts as depicted in Fig. 1 below. On a functional level, engineered *E. coli* strain carrying additional *rpoD* gene from *Clostridium cellulolyticum* showed increased detection frequency for hydrolytic enzymes while the parent strain resulted in the detection of rather differing fosmid clones suggesting a positive role of the added sigma factor (W. R. Streit, unpublished data).

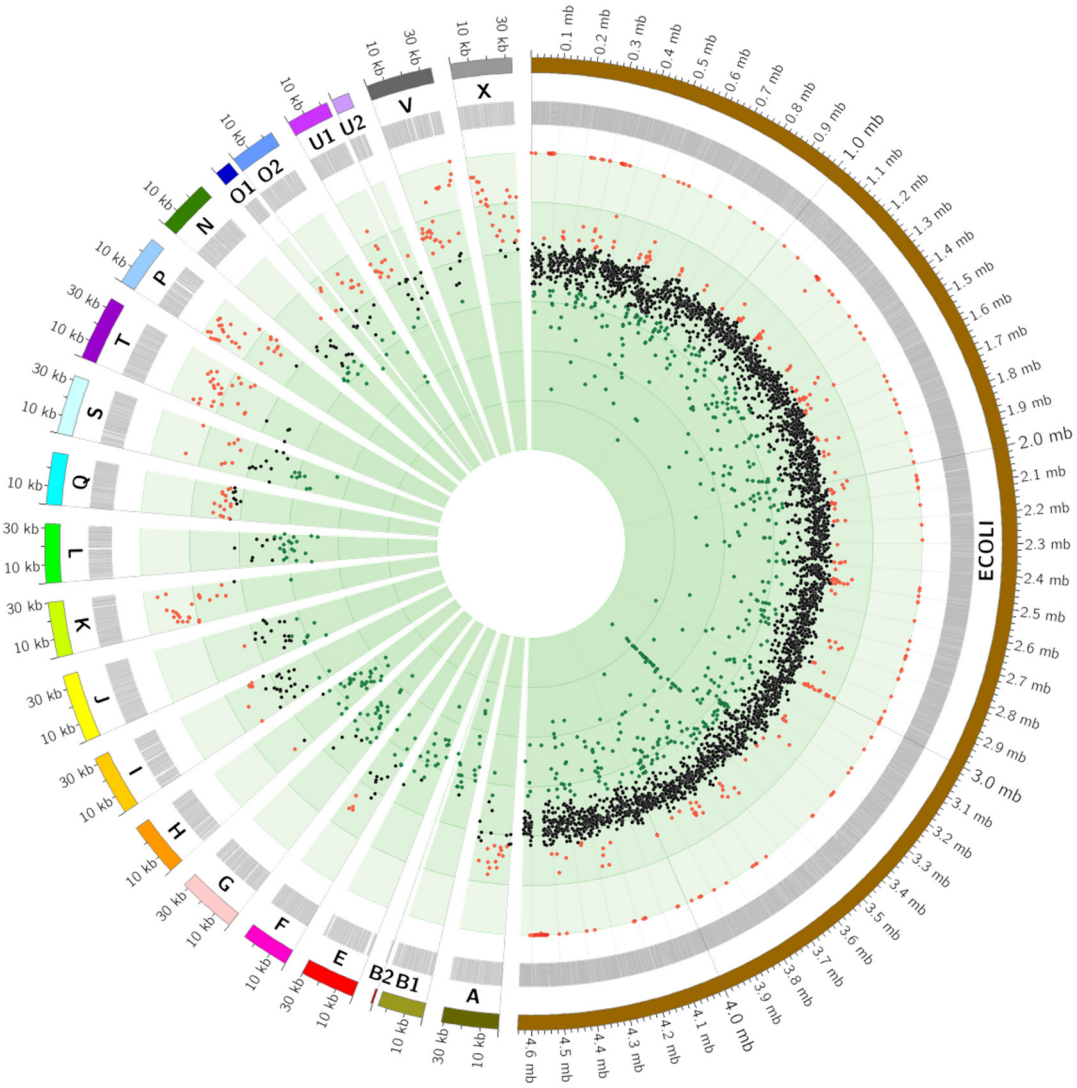


Figure 1: Transcription profile of different fosmid inserts in *E. coli* EPI300 host system and its own genes. Labels A to X represent different fosmid inserts originated from libraries of environmental samples. Green dots: range of the highest transcribed 10 % of *E. coli* genes, red dots: range of the lowest 10 %. Black dots display the average 80 % of the FPKM values (number of fragments per kilo base, per million mapped reads) of the *E. coli* genes (Jürgensen, 2014).

Not long after, Gaida and colleagues have investigated seven different sigma factors separately cloning into *E. coli* among which the strain carrying chromosomally integrated *rpoD* from *Lactobacillus plantarum* was able to initiate transcription of

genetic determinants of the latter from a genomic library construct (Gaida*, et al.*, 2015). As a result, the strong ethanol tolerance property could clearly be observed in the engineered *E. coli* strain but not in the wild-type.

Even though there is progress in establishing various alternative host systems, the inherent challenge of transcription bias disfavoring metagenome-derived genes originated from distantly related organisms still remains to be addressed. In addition, the construction of metagenomic libraries, activity-based screening and the final recovery of positive clones can be laborious and lengthy process especially when multiple alternative host systems have to be used to enhance expression coverage.

As much as the choice of an efficient and inclusive expression host and strategy is a pivot, the success of functional enzyme search or any other biomolecule depends very much on the use of effective screening platform. Any screening method in functional metagenomics should be able to deal with the requirements of feasibility, sensitivity, specificity, and adaptability. In functional searches, usually high number of clones has to be screened to get few positive hits and these methods should be sensitive enough to pick the usually low level of activity directly from metagenome clones. These screening methods vary ranging from the use of enrichments, growth selection with or without deletion mutant-based strains (such as the use of activity complementation) to the direct detection of target activity, co-/substrate, by/product, or any other agent that can provide information about the target phenotype. Screening techniques can be manual as in plates assays (Chow*, et al.*, 2012), thin-layer chromatography (TLC) (Rabausch*, et al.*, 2013) and high-performance liquid chromatography (HPLC) (Bohnke and Perner, 2015) or more automated high throughput screening (HTS) platforms (Acker and Auld, 2014).

1.5 Cell-free protein expression and its potential for functional screening

Cell-free expression has been popular for some time now with significant achievements in *in vitro* protein production and synthesis of other biomolecules used for ranges of applications (Carlson*, et al.*, 2012, Chong, 2014, Hiroyuki*, et al.*, 2010, Hodgman and Jewett, 2012, Karig, 2017, Körfer*, et al.*, 2016, Pardee*, et al.*, 2016, Shimizu*, et al.*, 2001, Tuckey*, et al.*, 2014, Zemella*, et al.*, 2015). It is transformed from its earlier use in the fundamental understanding of transcription and translation processes to its current use in large-scale production of complex proteins, unnatural amino acids incorporations and virus-like particles, cell-free metabolic engineering, and artificial cell engineering (Lu, 2017, Rosenblum and Cooperman, 2014). Cell-free protein expression systems have also been used for the production of difficult-to-express proteins such as toxic, disulfide-bonded, membrane and other Archaeal proteins (Bundy and Swartz, 2011, Hirao*, et al.*, 2009, Isaksson*, et al.*, 2012, Lu, 2017, Thoring*, et al.*, 2016). Moreover, the advanced combination of cell-free protein synthesis with high throughput screening (HTS) and modern sorting systems in droplet-based artificial cell compartments is emerging as a powerful tool for enzyme discovery and engineering (Catherine*, et al.*, 2013, Körfer*, et al.*, 2016).

Unlike living cells, cell-free expression systems offer unprecedented autonomy to the user for easier optimization and customization according to virtually unlimited biosynthesis formats and requirements. Apart from being typically cheaper and more feasible, the system is also energy efficient allowing resources to be directed for expression of target proteins which would otherwise be distributed for several other metabolic processes necessary for host cell growth and maintenance as well.

Many lines of cell-free protein synthesis systems have been developed from different organisms of both prokaryotic and eukaryotic origin (Villarreal and Tan, 2017, Zemella*, et al.*, 2015). The basic principle of establishing the systems is generally the same. Endogenous nucleic acid depleted cell extracts are used for expression upon addition of required supplements, although practically more complicated than it sounds. The presence of nucleases, proteases, several other cellular contaminants, and side activities from cell extract has limited the direct use of cell-free expression system, especially for enzyme screening.

Applying partly or completely purified expression components or reconstituted systems have been demonstrated to avoid or reduce unwanted background activities arising from crude extracts, reduce degradation of nucleic acid and protein, and/or prevent rapid depletion of energy charges. As a result, researchers established the use of reconstituted translation systems from all-purified recombinant components in combination with phage polymerase (Shimizu*, et al.*, 2001, Srivastava*, et al.*, 2016, Zhou*, et al.*, 2012). The system was named as Protein synthesis Using Recombinant Elements (PURE) system and later more optimized and commercialized by different companies (Whittaker, 2013). Although such systems have been applied for expression of some proteins, it requires recombinantly expressing and purifying a large number of translation components (Kuruma and Ueda, 2015). The use of reconstituted translation systems runs on minimal translation state and is not always suitable especially when complex post-translational modifications are required.

An alternative to this and widely applied approach for ranges of applications is to combine purified high-level recombinant transcription tools with well-processed cell-extract-based translation lines. This can be achieved via direct *in vitro* coupled

reaction mode or feeding pre-synthesized mRNA to the cell extract for subsequent translation and protein synthesis which is referred as uncoupled or linked method. In such a combination, the use of recombinant RNAP takes advantage of the rapid synthesis of transcripts, while cell extract-based mix has the benefit to provide all translation components for the more complex translation and post-translation modifications. The key strategy here is therefore the choice of RNA polymerase to produce a high quality and representative mRNA for a successful protein expression. This is an extremely important step especially when functional screening from of template DNAs from diverse origins in a homogeneous or heterogeneous state is sought.

Cell-free expression is routinely directed by the T7 monomeric highly processive phage polymerase. Transcription is based on specially prepared circular or PCR amplified linear templates, unlike *in vivo* systems. Avoiding the need for cloning and cell cultivation for propagation, expression from linear DNA template is more preferred despite being prone to degradation by endonucleases from cell extracts (Schinn*, et al.*, 2016). PCR amplified templates are particularly suitable for screening and engineering applications where a large number of template variants have to be screened for desired function (Körfer*, et al.*, 2016). In this specific example, selected promoter-fused cellulose gene is used to generate mutant library using PCR which is a different scenario to using multiple different template DNAs from the beginning.

The lack in modularity of the T7 or other phage-derived RNA polymerases to recognize promoter motifs other than their own limits its use for applications which require transcription from mixed/varied templates, such as functional screening of

enzymes from genomes or metagenomes. To improve the coverage of transcription when genomic or DNAs with different recognition factors are used as templates, endogenous RNA polymerases with in the cell extracts have been suggested (Fujiwara*, et al.*, 2017, Garamella*, et al.*, 2016, Shin and Noireaux, 2010). Therefore, investigating more robust and stable recombinant RNA polymerases is of at most interest especially for direct *in vitro* transcription of environmental DNAs using innate recognition sequences.

In this work, a heat-stable and robust recombinant RNA polymerase-based *in vitro* transcription system was developed from the recently isolated *Geobacillus* sp. GHH01 (Wiegand*, et al.*, 2013) for transcription of meta/genome-derived DNAs without the need for a special template preparation, and outlined how this can be combined with cell-extract translation systems for establishing cell-free functional metagenomics screening platform.

1.6 Aims of the study

The goal of this study was to search for selected, including rare and difficult-to-express, enzymes using host-dependent and host-independent functional meta/genomics and expression strategies for biocatalytic phosphorylation of target metabolites. It thereby aimed at establishing *in vitro* meta/transcription system by employing recombinant RNA polymerase from *Geobacillus* sp. GHH01 and lay a foundation for cell-free functional metagenomic screening platform.

2 Establishing selective screening method and functional search for ß-glycerophosphate-involving activities

Birhanu M. Kinfu[1], Jennifer Chow[1], Getachew S. Molla[2], Andreas Liese[2], Roland Wohlgemuth[3] and Wolfgang R. Streit*[1]

[1]Department of Microbiology and Biotechnology, Biocenter Klein Flottbek, University of Hamburg, Ohnhorststr. 18, D-22609 Hamburg, Germany; [2]Institute of Technical Biocatalysis, Hamburg University of Technology, Hamburg, Germany; [3]Sigma-Aldrich Chemie GmbH, Member of Merck Group, Buchs, Switzerland

Manuscript under preparation

***Correspondence addressed:**
Wolfgang R. Streit,
Department of Microbiology and Biotechnology,
Biocenter Klein Flottbek, University of Hamburg, Ohnhorststr. 18,
D-22609 Hamburg, Germany.
Tel.: +49 40 428 16 463;
Email: wolfgang.streit@uni-hamburg.de

Introduction

Glycerol-2-phosphate, a phosphorylated glycerol compound at its second carbon atom and commonly known as ß-glycerophosphate, is reported only in handful organisms such as in *Streptococcus pneumoniae*, *Arabidopsis thaliana* and fission yeast (Chaleckis*, et al.*, 2014, Morona*, et al.*, 1999, Wang*, et al.*, 2010). Other than few reports of the cellular presence of ß-glycerophosphate, little is known how it is biosynthesized in nature. Nonetheless, it is used in a wide range of applications including in cell culture media for *in vitro* calcification and cell differentiation studies (osteogenic, adipogenic and odontoplan differentiations) for stem cell research (Couble*, et al.*, 2000), to prepare calcium-phosphate coatings for bone implants (Park*, et al.*, 2007, Surmenev*, et al.*, 2014), to make thermosensitive hydrogels for drug delivery systems (Ding*, et al.*, 2013, Szymanska*, et al.*, 2016), for water treatment (Ding*, et al.*, 2015), in reaction buffers as a potent serine-threonine phosphatase inhibitor (Novak-Hofer and Thomas, 1985), in M17 medium for culturing probiotic bacteria and their phages, and for Lactococcus culture in recombinant protein expression (Samazan*, et al.*, 2015, Terzaghi and Sandine, 1975).

While glycerol is normally phosphorylated by glycerol-3-kinase (EC 2.7.1.30) to glycerol-3-phosphate and enters to glycolysis, it is little known about phosphorylation of glycerol at its 2nd carbon atom. It is also not clear weather alternative processes such as breakdown of bigger phosphorylated metabolites or isomeric rearrangements potentially yields ß-glycerophosphate within cells. Morona *et al.* hypothesized the synthesis of glycerol-2-phosphate from glyceraldehyde-2-phosphate by the *cps23fX* encoded putative glyceraldehyde-2-phosphate dehydrogenase enzyme in the biosynthesis of CDP-2-glycerol, a precursor for

capsular polysaccharide (Morona*, et al.*, 1999). Experimental work by Wang *et al.*, on the other hand, reported that there are indeed three enzymes involved in the biosynthesis of CDP-2-glycerol, however, with difference in the enzyme types from suggested above (Wang*, et al.*, 2010). Although the activity was not fully characterized, a second enzyme coded by a gene in the *cps* locus named as *gtp3* in the later study is claimed to have glycerol-2-phosphotransferase activity. It is therefore interesting to further examine whether this enzyme can practically be used for biocatalytic synthesis of ß-glycerophosphate. Biosynthesis of this metabolite has never been studied beyond.

In an enzyme target known so little about its presence in nature and process of its biosynthesis, functional search from environmental samples is a reasonable approach as it offers the advantage of screening new enzymes or enzyme functions without the requirement of background information. This can be achieved by establishing and employing a reliable detection method for the targeted activity. As there are different candidate substrates from which ß-glycerophosphate can possibly be biocatalytically derived, this study followed a top-down approach starting from the utilization of this unusual metabolite in bacteria with the aim to trace back its biosynthesis. In addition, it is wise to consider activity test and optimization of other known enzymes which phosphorylate glycerol derivatives with the required regio-specificity.

Materials & Methods

Bacterial strains, vectors, and constructs

Bacterial strains used in this study are listed on table 1, and plasmids and constructs on table 2.

Table 1: Bacterial strains used in this study.

Strains & isolates	Features	Reference/Source
Streptococcus pneumonia 23F	Wild-type strain	(Ramirez and Tomasz, 1998); (Wang*, et al.*, 2010)
E.coli EPI300	F−, *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC), Φ80d*lac*ZΔM15, Δ*lac*X74, *rec*A1, *end*A1, *ara*D139, Δ(*ara, leu*)7697, *gal*U, *gal*K, λ−, *rps*L, *nup*G, *trf*A, *ton*A, *dhfr*	Epicentre (Madion, WI, USA)
E. coli DH5α	F^- φ80 *lacZ*ΔM15 Δ(*lacZYA-argF*)U169 *recA1 endA1 hsdR*17(rk^-, mk^+) *phoA supE44* λ^- *thi*-1 *gyrA96 relA1*	Life Technologies (Frankfurt, Germany)
E. coli BL21(DE3)	F^- *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3)	Novagen/Merck (Darmstadt, Germany)
Four different isolates	Isolated from enrichment cultures and able to utilize ß-glycerophosphate as C-source	This work

Table 2: Plasmids and constructs.

Plasmids and constructs	Traits	Reference/Source
pDrive	3.85 kb TA-cloning vector, *oriEc*, Plac*lacZ*, AmpR, KanR, T7-promotor	QIAGEN (Hilden, Germany)
pET-28a(+)	5.37 kb expression vector, *lacI*, KanR, T7-promotor, N- and C-terminal His_6-tag coding sequence	Novagen/Merck (Darmstadt, Germany)
pCC1FOS™	8.139 kb, ChlR, *red*F, oriV, ori2, *rep*E, *par*A, *par*B, *par*C, cos, *lox*P, *lac*Z, T7-promoter	Epicentre (Madison, WI, USA)
pDrive::*Gtp3*	pDrive carrying *gtp3* gene	This work
pET-28a::*Gtp3*	pET-28a (+) carrying *gtp3* gene	This work
Fosmid clones	About 8000 fosmid clones constructed from enrichment isolates in PCC1FOS	This work

Culture media and supplements

M9 minimal salt medium was prepared according to Sambrook and Russel (Sambrook, 2001) after modification by including additional amino acids, vitamins, and other supplements. Unless mentioned otherwise, 30 mM ß-glycerophosphate was used as the only carbon source for enrichment and screening.

One liter of M9 minimal medium contained 200 ml of M9 salt solution (5X), 10 ml of trace elements solution (100X), 1 ml of vitamin stock (1000X), 0.9 ml of thiamine B1 (10 mg/ml), 0.4 ml of casamino acid (10 %), 0.94 ml of leucine (16 mg/ml), 1 ml of yeast extract (20 mg/ml), 2 ml of $MgSO_4$ (1 M), and 100 µl of $CaCl_2$ (1 M). $CaCl_2$ was first completely dissolved in 723 ml sterile deionized water before adding M9 salts followed by rest of the components. The M9 medium was optimized

for *E. coli* with varying concentrations of vitamin mix, thiamin (B_{12}), Leucine, yeast extract, and casamino acid.

M9 salt solution (5x) contains 85 g of $Na_2HPO_4.2H_2O$, 30 g of KH_2PO_4, 2.5 g of NaCl, 5 g of NH_4Cl dissolved in 1 L deionized water and autoclaved (Sambrook, 2001). Trace element stock solution (100X) contains 1 g of EDTA, 166 mg of $FeCl_3 \cdot 6H_2O$, 16.8 mg of $ZnCl_2$, 2.6 mg of $CuCl_2 \cdot 2H_2O$, 2 mg $CoCl_2 \cdot 2H_2O$, 2 mg H_3BO_3, 0.32 mg $MnCl_2 \cdot 4H_2O$ (Pfennig and Lippert, 1966). Once the 1 g EDTA is dissolved in 160 ml water after the pH is adjusted to 7.5 with NaOH, the rest of the components were added, filled to 200 ml with deionized water and filter sterilized. $MgSO_4$ and $CaCl_2$ were separately prepared in 1 M stock each and autoclaved. Vitamin stock solution (1000x) has components of 10 mg Ca-pantothenate, cyanocobalamine (B_{12}), nicotinic acid, pyridoxal-HCl (B_6), and riboflavin each and 1 mg biotin, folic acid, and *p*-amino benzoic acid each all dissolved in 100 ml.

LB medium has 1 % tryptone/peptone, 0.5 % yeast extract, 0.5 % NaCl autoclaved at 121°C for 20 min and supplemented with appropriate antibiotics and supplements listed in table 3. Agar media were prepared by adding 1.2 % of agar to same respective media compositions with supplements as necessary.

Table 3: Antibiotics and supplements

Antibiotics/ supplements	**Solvent**	**Stock solution**	**Working concentration**
Chloramphenicol	EtOH	50 mg/ml	50 µg/ml
Kanamycin	H_2O	25 mg/ml	25 µg/ml
IPTG	H_2O	1 M	0.1 mM

Enrichment of environmental samples and bacterial culture

Soil and mud samples were collected from Botanical garden, Klein flottbek (Hamburg, Germany) using sterile wares and processed immediately. The samples were solubilized with phosphate buffered saline and sieved. Serially diluted samples were then added to M9 defined medium supplemented with ß-glycerophosphate as a sole carbon source and incubated at 30°C. The cultures were washed and re-inoculated in fresh liquid medium for at least three cycles and then spread onto agar medium of same composition used for enrichment. The depletion of ß-glycerophosphate in liquid cultures was also checked using the TLC-based detection method.

Bacterial cultivation for ß-glycerophosphate utilization experiments and fosmid library screening was performed in M9 medium supplemented with ß-glycerophosphate in microtiter plates. Cell growth was monitored by measuring the optical density at a wavelength of 600 nm using a Synergy HT Microplate Reader with Gen5 software (BioTek Instruments, Winooski, VT). Cell growth density in flask cultures was measured using a photometer (SmartSpecTM Plus Spectrophotometer, BIO-RAD, Hercules, CA, USA) at a wavelength of 600 nm against the respective sterile medium.

Extraction of plasmid, fosmid and genomic DNAs

Genomic DNA (gDNA) extraction from bacterial isolates was done using peqGOLD Bacterial DNA Kit (*PEQLAB Biotechnologie GmbH*, Erlangen, Germany) according to the manufacturer's instruction and a standard CTAB/NaCl method (Wilson, 2001). Plasmid DNA isolation was done using either QIAGEN® plasmid mini

kit (Qiagen, Hilden, Germany) for sequencing or alkaline cell lysis method (Bimboim and Doly, 1979). Fosmid constructs were isolated using the high-speed plasmid mini kit (Geneaid Biotech, Taiwan). Nanodrop ND-2000 instrument (PEQLAB Biotechnologie GmbH, Erlangen, Germany) was used to analyze concentration and purity of DNA.

Partially sheared gDNAs by repeated pipetting for fosmid library construction, PCR product of glycerol phosphotransferase and restriction modified DNA fragments for cloning were extracted from agarose gel using Gel/PCR DNA fragment extraction kit (Geneaid Biotech, Taiwan).

Amplification and cloning of g*tp3* gene

The glycerol-2-phosphotransferase gene was amplified from genomic DNA of *S. pneumoniae* using *Gtp3_NdeI_*for (CATATGAAATTGACAAATAGAGTTGAT) and *Gtp3_HindIII_*rev (AAGCTTGACAATTCCTTTCCACATTTC) primer pairs. The gene was subcloned into pET28a vector from pDrive::*Gtp3* clone after digesting both with *NdeI* and *HindIII* DNA restriction enzymes (Thermo Fischer Scientific). Constructs were verified with automated sequencing using T7 primers and edited with the Pregap and Gap programs of the Staden sequence analysis package (Staden*, et al.*, 2003). Plasmid DNA construct with the right insert sequence was then heat-shock transformed into *E. coli* BL21 (DE3) for over-expression.

Construction of fosmid libraries and functional screening

The CopyControl™ Fosmid Library Production Kit with pCC1FOS vector™ (Epicentre, Madison, USA) was used for the construction of the large insert fosmid library, according to the manufacturer's instructions. For this, genomic DNAs from

four ß-glycerophosphate utilizing enrichment isolates were employed. Prior to library construction, the extracted gDNAs with two methods were pooled and subjected to optimized cycles of physical shearing. It was then gel purified as stated above and processed for packaging. Four fosmid libraries each with an average size of 2000 clones for about 3 times coverage of genome size were constructed. DNA of selected positive fosmid clones were digested with *BamHI* and presence and size polymorphism of insert was verified. The libraries were subjected to subsequent cycles of screening by transferring each clone into ß-glycerophosphate containing M9 medium supplemented with chloramphenicol in 96 well plates and by monitoring cell growth.

Heterologous protein expression and purification

Overnight culture of *E. coli* BL21 (DE3) host carrying the correct insert was used to inoculate 250 ml LB medium supplemented with 25 µg/ml kanamycin. The culture was grown at 37°C to an optical density (OD_{600}) of approximately 0.7 and induced protein expression by adding 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and further incubating 4 hours at 37°C. Cells were then harvested via centrifugation for 20 min at 5,000 rpm and 4°C using a Sorvall RC6+ centrifuge (rotor F10S-6x500y, Thermo Scientific, Braunschweig, Germany).

Cell pellet was resuspended in lysis buffer (50 mM $NaH_2PO_4 \cdot 2H_2O$ pH 8.0, 300 mM NaCl, 20 mM imidazole, 0.1 mM phenylmethylsulfonyl fluoride (PMSF), 1 mM dithiothreitol (DTT)) per 250 ml *E. coli* expression culture. French pressure cell press at 1300 psi was used to lyse cells and the lysate was incubated with 20 u DNase for 30 min at 37°C supplemented with 2mM $MgCl_2$. The mixture was centrifuged at 16,000g and 4°C for one hour. The clear cell lysate was then loaded

onto a Ni-chelating NTA-column (HisTrap columns, *GE Healthcare*, Freiburg, Germany) pre-equilibrated with lysis buffer and purified using a fast protein liquid chromatography (FPLC) system. Bound protein was first washed with 5 column volumes (CV) of wash buffer (50 mM $NaH_2PO_4 \cdot 2H_2O$ pH 8.0, 500 mM NaCl, 40 mM imidazole, 1% glycerol, 0.1 mM PMSF, 1 mM DTT) before finally eluting it using elution buffer (50 mM $NaH_2PO_4 \cdot 2H_2O$ pH 8.0, 300 mM NaCl, 250 mM to 500 mM imidazole, 0.1 mM PMSF, 1 mM DTT).

All buffers used for the FPLC system were filtered, degassed and cooled prior to use. Afterwards, the buffer was exchanged to lysis or tris-HCl buffer by at least three rounds of 5-fold volume dilution and concentration in a VIVASPIN® 6 concentrator with 10 kDa cut-off (MWCO). Protein quantification was done using Bradford assay (Bradford, 1976) taking BSA as control. Protein purity was monitored performing SDS-PAGE analysis (Laemmli, 1970) and further verified using western blot assay (Towbin*, et al.*, 1979).

Establishing TLC-based activity detection method and enzyme assay

Among the multiple approaches followed for developing an ideal screening method for ß-glycerophosphate separation and detection, TLC-based method was established using silica plate and a solvent mixture of methanol, 25% ammonia and water (6:3:1) for separation. After drying the TLC plate, phosphorylated metabolites were detected after staining by molybdate reagent with L-ascorbic acid and anthranilic acid. For selective detection of ß-glycerophosphate especially from its close α-isomer, samples were treated with periodic acid attacking vicinal hydroxyl functional groups. This subsequently oxidizes only α-glycerophosphate to formaldehyde and phosphoglycolaldehyde (Leva and Rapoport, 1943).

Kinase activity assays were performed in 50 mM reaction buffer (HEPES, PIPES, tris-HCl, phosphate buffers were tested) containing 30mM glycerol, 15mM ATP, 10mM $MgCl_2$. The reaction mixtures were incubated at 37°C up to 1 hour monitoring possible formation of ß-glycerophosphate using TLC-based detection method.

Results and discussion

Heterologous expression, purification and activity assay

With the aim of selective phosphorylation of glycerol at its second carbon atom, a glycerol-2-phosphotransferase coding gene from *Streptococcus pneumoniae* was heterologously expressed in *E. coli* Rosetta-gami 2(DE3) host cells as C-terminal His_6-tagged soluble protein. The protein was purified to high purity using ion-exchange chromatography. The protein was analyzed using SDS-PAGE and further verified by western blot immunoassay using penta-His-specific antibodies (Fig. 1, A and B).

While the other two enzymes in the biosynthesis pathway, gtp1 and gtp2, were investigated in detail in the work of Wang and coauthors, *gtp3* which was claimed to have glycerol-2-phosphotransferase activity was not fully characterized. In the same study, it was reported that the enzyme was active only when coupled with gtp2 which converts ß-glycerophosphate to CDP-2-glycerol (Wang*, et al.*, 2010). As continuous product recovery process is not the ideal way for biocatalytic production of this metabolite, the enzyme was independently investigated to retain its activity under different parameters in this study, with the potential to produce ß-glycerophosphate from glycerol as substrate. Despite all the attempts, there was no detectable activity

employing the newly established TLC-based assay method (Fig. 1C). In addition to *gtp3*, possible promiscuity of glycerate-2-kinase from *Thermotoga maritima* was tested for the phosphorylation of glycerol to ß-glycerophosphate under different parameters although no activity could be detected.

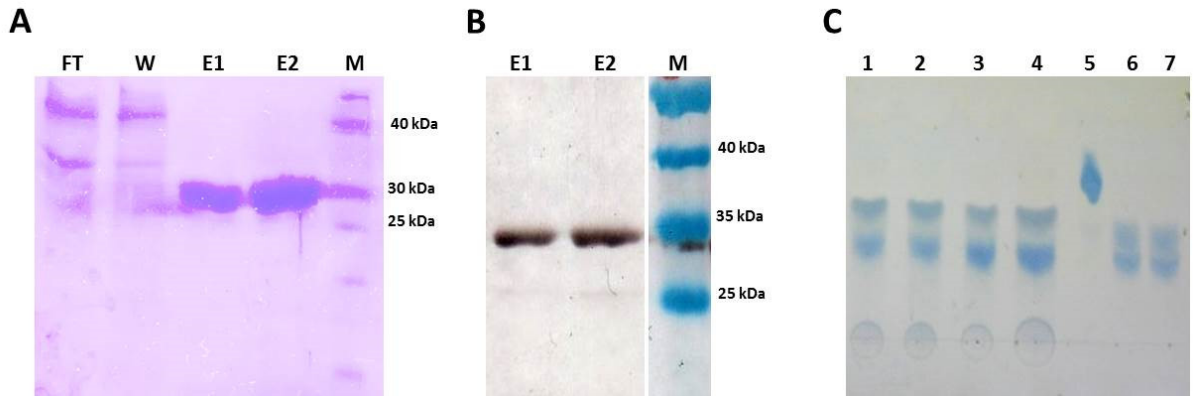


Figure 1: Ion-exchange purification of protein and activity assay. (A) SDS-PAGE analysis, (B) Western blotting protein analysis using penta-His-specific antibodies, (C) TLC-based activity assay of protein activity. Lane FT is for flow through, W for wash, E1 and E2 for elution fractions, and M for marker. Lanes 1 to 4 are reactions in different buffer systems, 5 is ß-glycerophosphate standard reference compound, 6 and 7 are for negative controls without enzymes.

Developing activity screening method for further searches

Establishing a reliable activity-based screening method is often the main bottleneck for functional search of enzymes with novel activities. This becomes even more challenging when specific separation of isomers like ß- and α-glycerophosphates with close properties is required. In this study, TLC-based method for screening of ß-glycerophosphate-involving activities was established. The α- and ß-glycerophosphate isomers were separated from the rest of possible reaction components by employing either silica or cellulose plates and a mixture of methanol,

25% ammonia and water (6:3:1 ratio) as mobile phase (Fig. 2A). Detection on plate after separation was possible using staining phosphate groups with molybdate reagent.

More interestingly, the system was further combined with selective oxidation of adjacent hydroxyl groups (Leva and Rapoport, 1943) removing only α-glycerophosphate while leaving ß-glycerophosphate untouched (Fig. 2B). The high alkaline nature of the solvent system has also additional advantage of dissociating other possible phosphorylated contaminant metabolites as they are often not stable at such high pH. Furthermore, it was possible to improve the sensitivity level of this method as low as 4.5 µg of sample concentration (Fig. 2C).

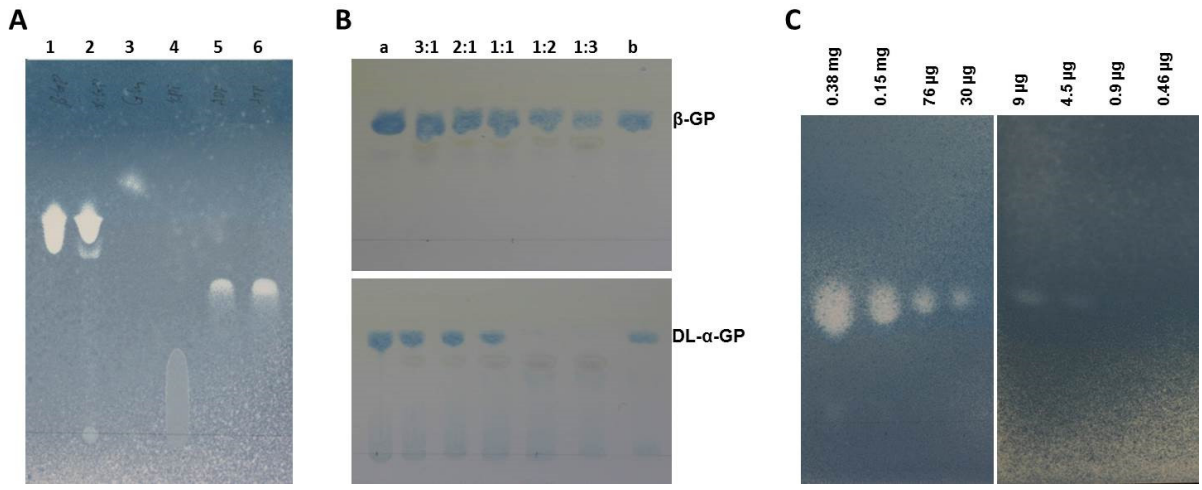


Figure 2: TLC-based screening of ß-glycerophosphate-involving activities. Both silica (left and right most) and cellulose (middle) plates were suitable. (A) The solvent-stationery phase combination is able to separate glycerophosphate isomers from the rest of possible reaction components. Lanes 1 to 2 are for ß-glycerophosphate (β-GP), α-glycerophosphate (DL-α-GP), 3 to 6 are for glycerol, inorganic phosphate, ADP and ATP, respectively. (B) Shows selective oxidation of α-glycerophosphate (bottom part) by periodic acid leaving β-glycerophosphate (top part). Different ratios of sample to periodic acid concentration are

shown from 2nd to 5th lanes. Sample compounds without any treatment are on lane a, and lane b depicts sample to water ration of 1:3 to show dilution effect. (C) Detection limit of the TLC-based screening method. From 1 to 8 shows detection of decreasing concentration of ß-glycerophosphate applied onto each spot.

Enrichment of environmental samples and construction of fosmid libraries

The presence of ß-glycerophosphate in cells has been thought to be unnatural for a long time until few studies report it in some organisms (Chaleckis*, et al.*, 2014, Morona*, et al.*, 1999, Wang*, et al.*, 2010). In studies conducted to investigate alternative phosphorous sources in nature and their uptake, it is described that the highly regulated phosphate acting enzymes mainly phosphatases and phosphate transport systems are activated only during phosphorus starvation (Matsuhisa*, et al.*, 1995, Torriani, 1960, Yang*, et al.*, 2009). It is also evidenced in this study that there was no cell growth recorded for *E. coli* EPI300 cells in minimal medium provided with other phosphate sources when ß-glycerophosphate was supplied as a sole carbon source. On the other hand, there was normal cell growth when the α-glycerophosphate isomer was supplemented instead (Fig. 3). The growth mixture of the two isomers did not show any shift or metabolic induction towards utilization of ß-glycerophosphate either. The cellular uptake of ß-glycerophosphate is described to be via *ugp*-encoded transporter the same way to the α-isomer (Yang*, et al.*, 2009).

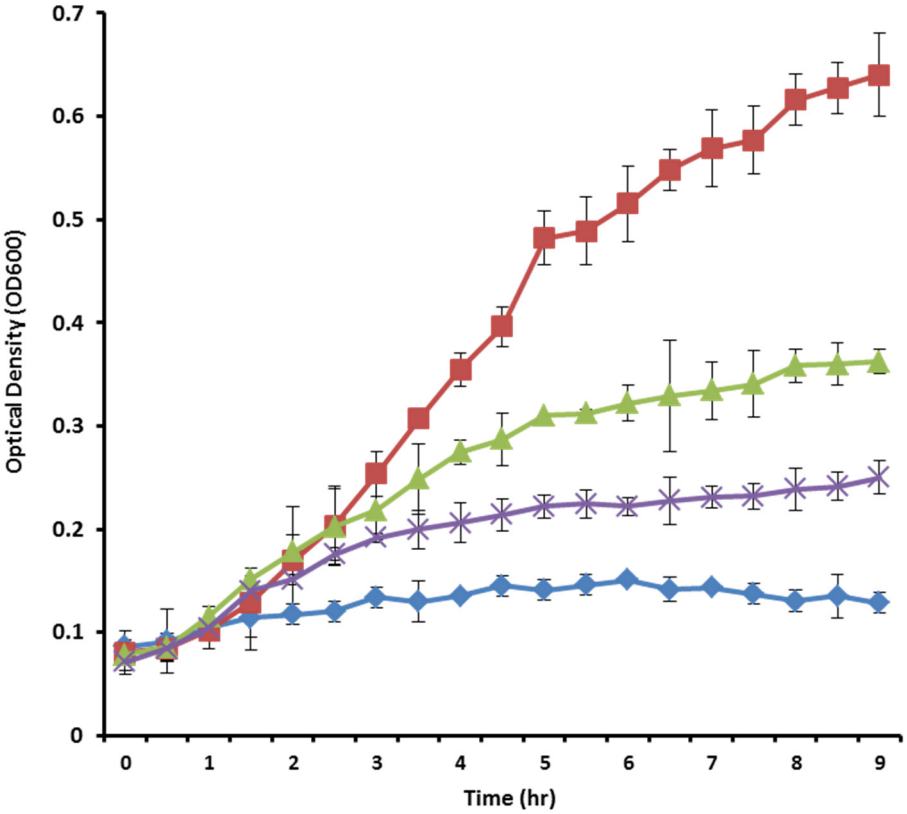


Figure 3: Utilization of glycerophosphates by *E.coli* EPI300 as a sole carbon sources in defined medium. Graphs represent growth of *E. coli* in M9 medium supplied with 10 mM of ß-glycerophosphate only (◊), 10 mM of α-glycerophosphate (□), α- and ß-glycerophosphate mix each 5 mM (△), and α- and ß-glycerophosphate mix each 2.5 mM (x).

Considering the rare report of its cellular synthesis in cells and the inability of one of the main hosts for library construction which is *E. coli* EPI300, enrichment of samples from the environment in minimal medium supplemented with ß-glycerophosphate as the only carbon source was done for functional searches. Interestingly, enrichment cultures resulted isolates which can utilize ß-glycerophosphate as carbon source (Fig. 4A and B). TLC-base assay also verified that cell growth was concomitant to depletion of ß-glycerophosphate (Fig. 4C). After 16S rRNA sequence analysis of selected isolates based on their vigorous growth, about 8000 pCCFOS fosmid clone libraries were generated from four isolates.

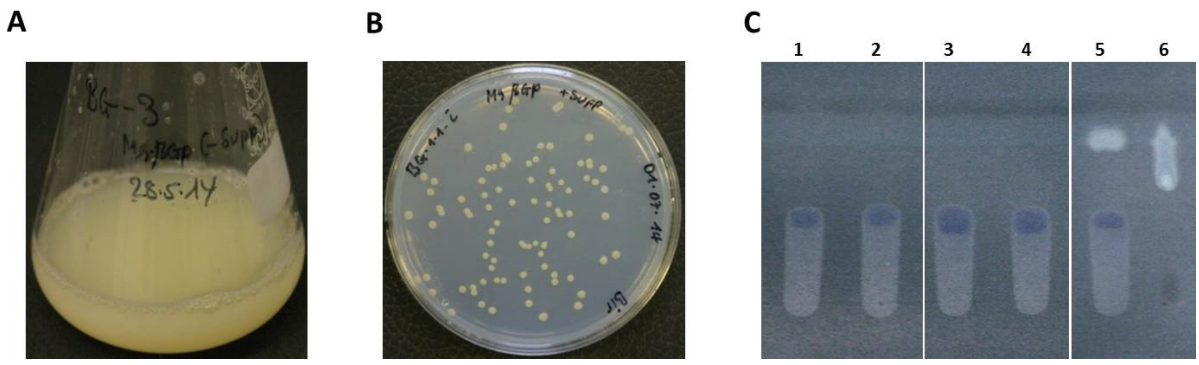


Figure 4: Enrichment of environmental samples using defined medium for ß-glycerophosphate utilization. (A) Liquid medium re-enriched for multiple cycles in newly prepared medium. **(B)** Pure isolate from the enrichment cultures on M9 minimal medium with ß-glycerophosphate. **(C)** TLC analysis of ß-glycerophosphate depletion during cell growth. Lanes 1 to 4 are ß-glycerophosphate utilizing isolates. Lane 5 is for the same medium incubated with *E. coli* cells and lane 6 is the standard ß-glycerophosphate sodium salt (85% purity) as reference.

Screening for fosmid libraries and activity complementation

Fosmid libraries of about 8000 clones constructed from genomic DNAs of four ß-glycerophosphate utilizing environmental isolates were screened based on growth using minimal growth medium with ß-glycerophosphate as a sole carbon and energy source using 96-well plates. Monitoring cell growth by measuring the optical density at a wavelength of 600 nm using a Synergy HT Microplate Reader with Gen5 software (Bio-Tek Instruments, Winooski, VT), 8 positive hits were found. Fosmid DNAs were extracted from these clones and further re-complementation experiments on the same medium composition yielded insufficient and inconsistent growth activity which most likely is due to problems associated with expression bias, as it is also evidenced by multiple studies involving functional meta/genomics systems (Uchiyama and Miyazaki, 2009).

Outlook and recommendation

As clearly evidenced in this work and presented by multiple other studies as well, expression problems are serious challenges facing functional enzyme search efforts especially with metagenomics mainly due to difference in codon usage leading to biased transcription, poor protein expression, protein misfolding or lack of appropriate secretion systems (Ekkers*, et al.*, 2012, Kudla*, et al.*, 2009, Lam*, et al.*, 2015, Steele*, et al.*, 2009, Uchiyama and Miyazaki, 2009). In this dissertation, a recombinant RNAP-based transcription system is established with the objective of setting cell-free metagenomics platform that can bypass host-related expression limitations during enzyme search in general and phosphotransferases in particular. Therefore, applying the use of such a system once it is fully optimized is highly recommended.

Alternatively, random or directed mutagenesis of existing enzymes such as glycerate-2-kinase which phosphorylate glycerol derivative with excellent regio-specificity can be a feasible alternative to modify its active site configuration that could fit the hydroxymethyl group of glycerol instead of the carboxyl group in glyceric acid.

3 Biocatalytic asymmetric phosphorylation catalyzed by recombinant glycerate-2-kinase

Norman Hardt[a], Birhanu M. Kinfu[b], Jennifer Chow[b], Bernhard Schoenenberger[a], Wolfgang R. Streit[b], Markus Obkircher[a] and Roland Wohlgemuth[a]

[a]Sigma-Aldrich, Member of Merck Group, Industriestrasse 25, CH-9470 Buchs, Switzerland; [b]Universität Hamburg, Biozentrum Klein Flottbek, Abteilung für Mikrobiologie und Biotechnologie, Ohnhorststrasse 18, 22609 Hamburg, Germany

Published in:

ChemBioChem, 18(15): 1518-1522. doi:10.1002/cbic.201700201

ChemPubSoc Europe

DOI: 10.1002/cbic.201700201

CHEMBIOCHEM Full Papers

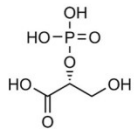
Biocatalytic Asymmetric Phosphorylation Catalyzed by Recombinant Glycerate-2-Kinase

Norman Hardt,[a] Birhanu M. Kinfu,[b] Jennifer Chow,[b] Bernhard Schoenenberger,[a] Wolfgang R. Streit,*[b] Markus Obkircher,[a] and Roland Wohlgemuth*[a]

The efficient synthesis of pure D-glycerate-2-phosphate is of great interest due to its importance as an enzyme substrate and metabolite. Therefore, we investigated a straightforward one-step biocatalytic phosphorylation of glyceric acid. Glycerate-2-kinase from *Thermotoga maritima* was expressed in *Escherichia coli*, allowing easy purification. The selective glycerate-2-kinase-catalyzed phosphorylation was followed by ^{31}P NMR and showed excellent enantioselectivity towards phosphorylation of the D-enantiomer of glyceric acid. This straightforward phosphorylation reaction and subsequent product isolation enabled the preparation of enantiomerically pure D-glycerate 2-phosphate. This phosphorylation reaction, using recombinant glycerate-2-kinase, yielded D-glycerate 2-phosphate in fewer reaction steps and with higher purity than chemical routes.

Introduction

D-Glyceric acid 2-phosphate (Scheme 1) is a key metabolite in central carbon metabolism, glycolysis/gluconeogenesis, pentose phosphate pathway, glycine, serine and threonine metabolism, methane metabolism, biosynthesis of plant secondary metabolites, phenylpropanoids, terpenoids, steroids, alkaloids derived from the shikimate pathway, antibiotics, and amino acids.



Scheme 1. D-Glyceric-acid-2-phosphate.

The glycolytic intermediate between D-glycerate-3-phosphate and phosphoenolpyruvate was first isolated from yeast and characterized as a crystalline D-glycerate-2-phosphate barium salt by Meyerhof and Kiessling.[1] An efficient, robust, and scalable route for the preparation of a soluble form of pure D-glycerate-2-phosphate has been of great interest since its discovery. It continues to be required both as a metabolite reference standard and as an enzyme substrate for measuring activities of enolases and their inhibitors; for example, enolase activities in a number of human diseases and disease-causing human pathogens such as *Streptococcus mutans* and the inhibitory activity of fluoride or phosphate in caries prevention, *Staphylococcus aureus*, *Streptococcus pyogenes* or the blood parasites *Schistosoma mansoni* causing schistosomiasis, and *Trypanosoma brucei* causing sleeping sickness.[2] The first definitive synthesis of D-glycerate-2-phosphate was described by Ballou and Fischer.[3] The identity and purity of the final product, which was obtained as the crystalline trisodium salt pentahydrate, was proven, but nine reaction steps were required to get from the starting compound (D-galactose) to the final product. Although a shorter synthesis starting from D-gluconolactone was later described,[4] the bottlenecks in chemical synthesis and the number of reaction steps using protection/deprotection and purification steps have created interest in an attractive direct enzymatic phosphorylation of glyceric acid.[5] The enzymatic phosphorylation of glyceric acid is also of clinical relevance for patients with D-glyceric aciduria, a rare inborn error of serine and fructose metabolism, as mutations in the *GLYCTK* gene have been revealed as the cause of D-glycerate kinase deficiency and D-glyceric aciduria.[6]

Site- and stereoselective kinases are very powerful and versatile catalysts for biocatalytic phosphorylations of small molecules, for example, in straightforward syntheses of phosphorylated metabolites.[7] As class II glycerate kinases catalyze the formation of D-glyceric-acid-2-phosphate in nature, a stable and highly selective glycerate kinase from this class (EC: 2.7.1.165) was selected for the biocatalytic phosphorylation of glycerate in the 2-position (Scheme 2).

Although a number of glycerate kinases from rat liver,[8] *Escherichia coli*,[9] *Pyrococcus horikoshii*,[10] *Picrophilus torridus*,[11] *Thermoplasma acidophilum*,[12] *Sulfolobus tokodaii*[13] and *Hyphomicrobium methylovorum*,[14] which are able to catalyze the formation of glycerate-2-phosphate, were investigated, detailed enzymatic information on the key enzyme, glycerate-2-kinase, was reported for the glycerate kinase of the hyperthermophilic archaeon *Thermoproteus tenax*.[15] From the comparison of the kinetic and biochemical properties of the characterized class II glycerate kinases of *T. tenax*, *P. torridus*, *T. acidophilum*, and

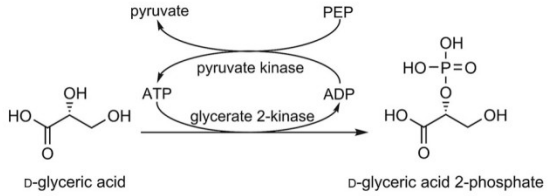
[a] Dr. N. Hardt, Dr. B. Schoenenberger, Dr. M. Obkircher, Dr. R. Wohlgemuth
Sigma–Aldrich, Member of Merck Group
Industriestrasse 25, 9470 Buchs (Switzerland)
E-mail: roland.wohlgemuth@sial.com

[b] B. M. Kinfu, Dr. J. Chow, Prof. W. R. Streit
Universität Hamburg, Abteilung für Mikrobiologie und Biotechnologie
Ohnhorststrasse 18, 22609 Hamburg (Germany)
E-mail: wolfgang.streit@uni-hamburg.de

Supporting information for this article can be found under: https://doi.org/10.1002/cbic.201700201.

ChemPubSoc Europe

CHEMBIOCHEM Full Papers



Scheme 2. Biocatalytic phosphorylation of D-glyceric acid by glycerate-2-kinase using the phosphoenolpyruvate(PEP)/pyruvate kinase system for regeneration of ATP.

H. methylovorum, the glycerate-2-kinase from *Thermotoga maritima* was selected due to its favorable biocatalytic properties and structure–function characterization.[16–17]

The 1.25 kb glycerate-2-kinase gene from *T. maritima* was expressed heterologously in *E. coli* as a fusion protein with the maltose-binding protein (MBP) and was purified by affinity chromatography. The active form of the protein was obtained by cleaving off the MBP. We investigated biocatalytic phosphorylations catalyzed by recombinant glycerate-2-kinase with quantitative ^{31}P NMR spectroscopy using a phosphoenolpyruvate (PEP)/pyruvate kinase system for ATP regeneration, starting with racemic and the enantiopure D- and L-glycerate as substrate. Nearly 100% conversion of D-glycerate to D-glycerate-2-phosphate was observed. With pure L-glycerate as substrate, no L-glycerate-2-phosphate was formed, whereas DL-glycerate gave 50% conversion with excellent enantioselectivity.

Results and Discussion

The glycerate-2-kinase gene from *T. maritima* was cloned into the pMAL-C2X vector and expressed in *E. coli* as a soluble fusion protein with maltose-binding protein (MBP, 87 kDa). The highest protein yield was obtained when expressed at a temperature of 17 °C. After harvesting and purification by affinity chromatography with amylose resins, the active glycerate-2-kinase (52 kDa) was obtained by cleaving off the MBP with factor Xa protease, maintaining its maximum activity without any tag (Figure 1).

Enantioselectivity of glycerate-2-kinase-catalyzed phosphorylation of glycerate

Enantioselectivity was determined by analyzing the kinetics of the glycerate 2-kinase-catalyzed conversion of equal amounts of DL-glyceric acid, D-glyceric acid, and L-glyceric acid. Every 15 min, a ^{1}H NMR spectrum (left) and a ^{31}P NMR spectrum (right) was measured. $T=0$ (red, bottom) was measured approximately 2 min after addition of the enzymes. The conversion can be tracked by the decrease in PEP signals (^{1}H: 5.0–5.4 ppm; ^{31}P: −0.9 ppm) and the increase in product signals of D-glyceric-acid-2-phosphate and pyruvate. (^{1}H: 4.4 ppm and 2.3 ppm; ^{31}P: 2.4 ppm), as shown in Figures 2 and 3. These NMR experiments clearly demonstrate the conversion of only

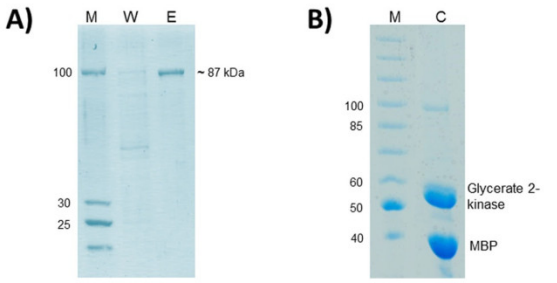


Figure 1. SDS-PAGE analysis of expression and purification of glycerate-2-kinase. A) Wash and elution fraction of purified fusion protein with MBP and B) cleaved protein after using factor Xa protease. Approximately 1.5 µg (A) and 35 µg (B) total proteins were loaded, respectively. M: marker, W: wash, E: eluate, C: cleaved protein.

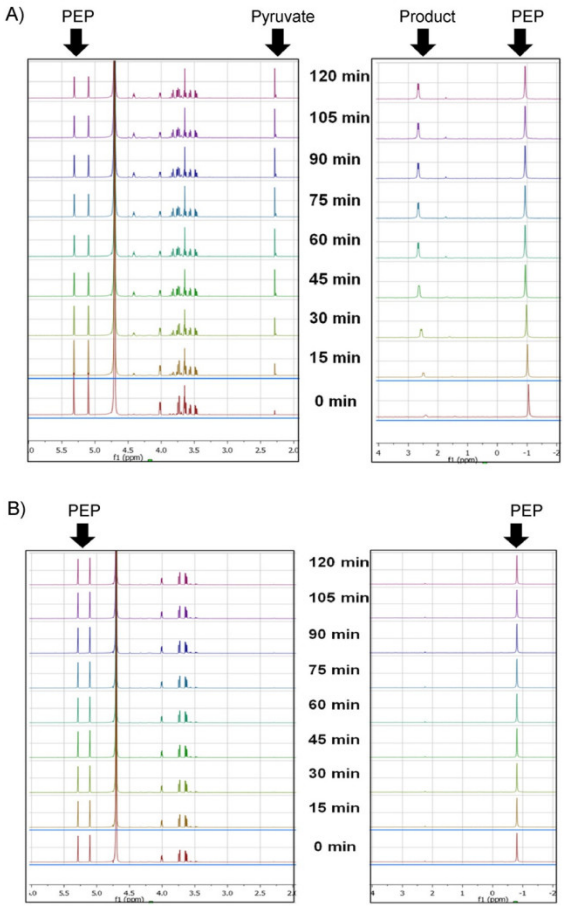


Figure 2. A) NMR kinetics of the enzymatic conversion of DL-glyceric acid. DL-glyceric acid (10 mg; 94 µmol, 1.0 equiv), 2.6 mg ATP (5 µmol, 0.05 equiv) and 17.5 mg PEP (85 µmol, 0.9 equiv) were dissolved in D_2O (1.7 mL) with $MgCl_2$ (10 mM). The pH was adjusted to 7. Subsequently, pyruvate kinase (50 units in 50 µL) and the prepared recombinant glycerate-2-kinase (*T. maritima*; 250 µL) were added. The enzymatic reaction was observed by NMR. B) NMR kinetics of enzymatic conversion of L-glyceric acid, whereby the same procedure as in (A) was followed except that instead of DL-glyceric acid as substrate, L-glyceric acid (10 mg, 94 µmol, 1.0 equiv) was used.

ChemPubSoc Europe
CHEMBIOCHEM Full Papers

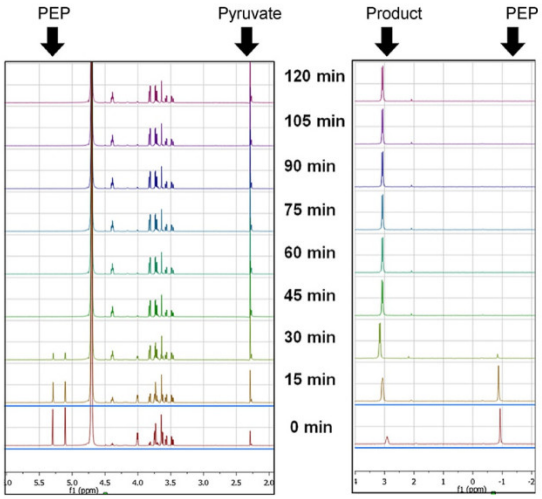


Figure 3. NMR kinetics of enzymatic conversion of D-glyceric acid. The same procedure as in Figure 2 was applied, but g D-glyceric acid (10 nm, 94 µmol, 1.0 equiv) was used as starting material.

the D-glyceric acid and not the L-glyceric acid to the corresponding product phosphorylated in the 2-position.

Kinetic measurements of glycerate-2-kinase from *Thermotoga maritima*

A K_m value of (0.13±0.03) mM to D-glyceric acid, a catalytic turnover number (k_{cat}) of $(5.4\pm0.7)\times10^{-2}\,s^{-1}$ and a k_{cat}/K_m value of $(4.0\pm0.05)\times10^{4}\,M^{-1}s^{-1}$ was determined at 37 °C for this glycerate-2-kinase (see the Experimental Section). The K_m value for our glycerate-2-kinase agrees well with the wild-type TM1585. Although the k_{cat} value for the D-glycerate was not determined in the wild-type TM1585, the temperature dependence of the specific activity showed a large increase from 37 to 80 °C, plateauing at 90 °C.[17]

As shown in Figure 4, glycerate-2-kinase from *T. maritima* was most active at neutral and alkaline pH, and its maximum activity was recorded at pH 9.2. The enzyme also maintained its full activity after storage at 4 °C for several months in 0.02% NaN_3 (*w/v*). Its excellent stability (both reaction and storage) and activity profile, thus allowing a direct coupling to the ATP regeneration system for glycerate-2-phosphate synthesis, is an advantage of this glycerate-2-kinase from *T. maritima*.

An optimized reaction system for the biocatalytic phosphorylation of D-glyceric acid is, however, not only dependent on the optimal properties of the glycerate-2-kinase but also on the optimum parameters for ATP regeneration and the stability profile of D-glyceric-acid-2-phosphate. The stability of D-glyceric-acid-2-phosphate was examined in the pH range of 6.0 to 11.0. A 20 mM solution of the lithium salt of D-glyceric-acid-2-phosphate in citrate buffer at pH 6.0, in Tris buffer at pH 7.0 and 8.0, and in carbonate buffer at pH 9.0, 10.0, and 11.0 was observed by ^{1}H and ^{31}P NMR. No decomposition was detected during the observation period, which was five days at pH 10.0 and 11.0, 11 days at pH 7.0, 8.0, and 9.0, and 18 days at pH 6.0. As the optimum pH for the ATP regeneration system is pH 7.0–7.5, the biocatalytic phosphorylation of D-glyceric acid at a pH near 7.0 is favored over a reaction pH at the optimum pH of the glycerate-2-kinase, although the stability profile of the D-glyceric-acid-2-phosphate is compatible.

To determine the enantiomeric purity of the product, capillary zone electrophoresis was performed with vancomycin as a chiral selector and confirmed a D/L enantiomeric ratio of 98.3:1.7 for the product D-glycerate 2-phosphate (see direct data on capillary zone electrophoresis in the Supporting Information). The development of suitable reaction conditions for complete conversion in phosphorylations was facilitated by ^{31}P NMR reaction monitoring,[18] as both product formation and PEP donor consumption could be measured simultaneously. From the time course of biocatalytic O-phosphorylation of D-glyceric acid in the 2-position, followed by ^{31}P NMR as shown in Figure 5, the reaction conditions for complete conversion and optimal process design could be derived.

Conclusion

Glycerate-2-kinase from *T. maritima* was selected as a biocatalyst for the phosphorylation of glyceric acid, and the corre-

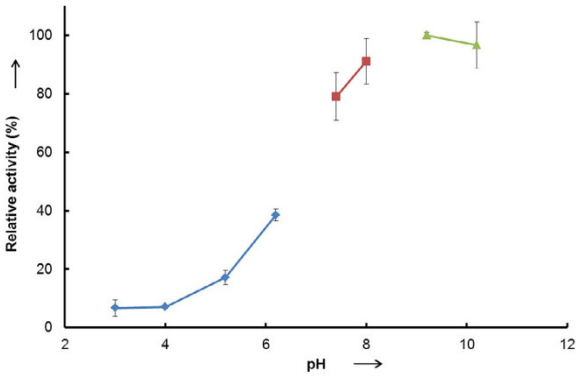


Figure 4. Glycerate-2-kinase activity as a function of pH value. Citrate buffer (pH 3.0, 4.0, 5.2, and 6.2 (◆), 50 mM Tris·HCl buffer pH 7.4 and 8.0 (■) and sodium carbonate–bicarbonate buffer (pH 9.2 and 10.2 (▲), 50 mM).

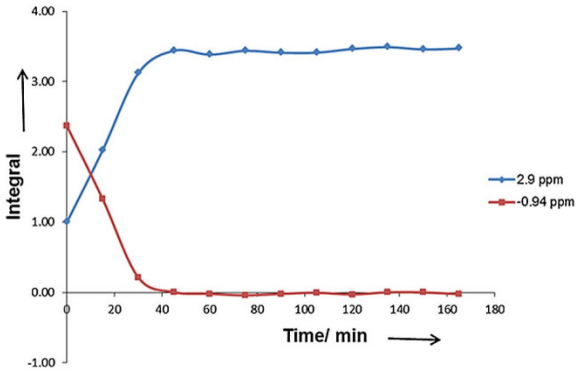


Figure 5. Kinetics of formation of D-glycerate-2-phosphate (◆) and consumption of phosphoenolpyruvate (■) derived from ^{31}P NMR signals.

ChemPubSoc Europe

CHEMBIOCHEM Full Papers

sponding gene was expressed in *E. coli* as a fusion protein with MBP to allow easy affinity purification. After cleaving MBP with factor Xa protease, a highly active and stable glycerate-2-kinase was obtained. A K_m value of (0.13 ± 0.03) mM to glyceric acid, a k_{cat} value of $(5.4 \pm 0.7) \times 10^{-2}\ s^{-1}$, and a k_{cat}/K_m of $(4.0 \pm 0.05) \times 10^4\ M^{-1} s^{-1}$ was determined. The glycerate-2-kinase-catalyzed phosphorylation of the D- and L-enantiomers of glyceric acid, as well as its racemic form, were directly investigated by NMR. These experiments clearly demonstrated the power and selectivity of biocatalytic phosphorylation. Only the D-glyceric acid was accepted as a substrate for this recombinant glycerate-2-kinase, whereas the L-glyceric acid was not phosphorylated to any extent. This straightforward and highly selective phosphorylation reaction is therefore very valuable for the enzymatic resolution of racemic glyceric acid or for the synthesis of D-glycerate-2-phosphate by phosphorylation of D-glyceric acid in the 2-position.

The glycerate-2-kinase-catalyzed phosphorylation of D-glyceric acid and subsequent product isolation enabled the preparation of the key metabolite D-glycerate-2-phosphate in enantiomerically pure form. This biocatalytic phosphorylation reaction with recombinant glycerate-2-kinase is also very sustainable, as it yields D-glycerate-2-phosphate in a one-step reaction and with higher purity than chemical routes.

The excellent enantioselectivity and the first crystal structure[16] of a representative of the glycerate-2-kinase family provide exciting opportunities for extending this work to new phosphorylated compounds which, until now, were only accessible through lengthy chemical routes, such as substituted D-glycerate analogues and functional isosteric and negatively charged substrates able to dock properly into the active site. Although highly conserved amino acid residues from the Rossmann-like domain and the C-terminal domain have been identified and contain one highly conserved basic residue, a detailed understanding of the interaction with the negatively charged D-glycerate and the elucidation of the mechanism of the enzyme function require further studies and higher resolution.

Experimental Section

Unless otherwise stated, all chemicals were of analytical grade and were provided by Sigma–Aldrich.

NMR spectroscopy: NMR spectra were measured in D_2O at room temperature on a Bruker Avance III 600 MHz spectrometer equipped with a BBO probe head with *z*-gradient by using 600.2 MHz for ^1H and 150.9 MHz for ^{13}C. The NMR kinetics of enzymatic phosphorylations of DL-glyceric acid, L-glyceric acid, and D-glyceric acid was measured by acquisition every 15 min of ^1H NMR and ^{31}P NMR spectra. $T = 0$ (red, bottom) was measured approximately 2 min after addition of the enzymes. Conversion was tracked by the decrease in the PEP signals (^1H: 5.0–5.4 ppm; ^{31}P: −0.9 ppm) and the increase in product signals (^1H: 4.4 ppm; ^{31}P: 2.4 ppm).

Capillary electrophoresis: Capillary zone electrophoresis with EOF-reversal (dynamic coating) was performed with a PDA detector and indirect detection at a wavelength of 233 nm and at 25 C. Uncoated capillaries with 75 µm inner diameter, 120 cm length, and 110 cm distance to the detector were used. The accelerator from the CEofix Anions 5 kit (Analis) was used as an electrophoresis buffer, the initiator from the CEofix Anions 5 kit as was used as a conditioner, and a vancomycin solution (40 mM) was used as a chiral selector.

Amplification and cloning of the glycerate-2-kinase gene: The glycerate kinase gene from *T. maritima* was amplified by PCR by using primer pairs G2K-EcoRI_F (5′-gaattc ATGTTT GATCCT GAATCC TTG and G2K-HindIII_R (5′-aagctt CTAGAC GATGAG GCCTAT TATC) and cloned into the pDrive cloning vector. The gene was then excised from the plasmid with endonucleases EcoRI and HindIII and subcloned into pMAL-C2X in the same translational reading frame as the *malE* gene for a N-terminal MBP-fused expression. The clone was verified by sequencing with primers mal2E (5′-AGCTGG CAAAAG AGTTCC) and M13 forward (5′-TGTAAA ACGACG GCCAGT G), both binding within a vector sequence upstream (within the *malE* gene) and downstream of the insert. The fusion plasmid clone was finally transformed by using heat shock into chemically competent *E. coli* BL21(DE3) host cells for expression.

Culture conditions, protein expression, and purification: *E. coli* BL21(DE3) expression host cells carrying the correct fusion clone were grown overnight at 37 °C in rich medium (10 g tryptone, 5 g yeast extract and 5 g NaCl) with 2% glucose and 100 µg mL^{-1} ampicillin). Three percent of the cells grown overnight were transferred into the same fresh medium, further incubated until the OD_{600} reached 0.6, and induced with isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.3 mM) at 17 °C for 18 h for maximum soluble protein expression. Cells were harvested by centrifugation (5000 *g*, 20 min, 4 °C) and resuspended in column buffer (20 mM Tris·Cl pH 7.4, 0.2 Mm NaCl, 1 mM EDTA, 20 mM 2-ME and 1 mM NaN_3). A French press was used to break apart the cells and, after centrifugation (16000 *g*, 60 min, 4 °C), the supernatant was loaded onto a pre-equilibrated column packed with amylose resin. After being washed with 12 column volumes of column buffer, the fusion protein was eluted with maltose (10 mM) in the same buffer.

To obtain tag-free glycerate-2-kinase protein, the MBP was cleaved from the fusion protein with factor Xa protease (New England Biolabs) overnight on ice. The protein was further concentrated and simultaneously dialyzed on Vivaspin columns (Sartorius) and then quantified with Roti-Quant solution (Carl Roth), according to the manufacturer's protocol.

Kinetic measurements of glycerate-2-kinase from *T. maritima*: Kinetic determinations of the glycerate-2-kinase were performed by using a continuous coupled assay based on the oxidation of NADH, in which the decrease in absorbance at 340 nm was monitored by using a Synergy HT multimode microplate reader with Gen5 software. The assay mixture contained NADH (0.3 mM), ATP (1.5 mM), $MgSO_4$ (10 mM), PEP (1.5 mM), pyruvate kinase (1.5 U), L-lactate dehydrogenase (1.5 U), and D-glyceric acid concentrations varying from 0.1 to 1.4 mM in Tris buffer pH 7.4 (50 mM) to a volume of 200 µL. The reactions were carried out in triplicate at 37 °C after adding purified glycerate 2-kinase (0.2 µg). Apparent values of k_{cat} and K_m were determined by applying the Michaelis–Menten kinetic model. A discontinuous assay was used, however, for measuring the enzyme activity over pH conditions ranging from pH 3.0 to 10.2. The phosphorylation reaction took place separately at 37 °C for 10 min and was immediately placed on ice. The pH of assay reactions was adjusted by adding predetermined volumes of NaOH or HCl immediately before coupling to the indicator reaction, which had a fivefold excess of pyruvate kinase and lactate

ChemPubSoc Europe

CHEMBIOCHEM Full Papers

dehydrogenase. The amount of ADP formed was then measured by monitoring oxidation of NADH, as above.

Enzyme reactions

Enzymatic syntheses of DL-, D- and L-glyceric acid 2-phosphate: DL-, D- or L-glyceric acid (10 mg, 94 μmol, 1.0 equiv), ATP (2.6 mg, 5 μmol, 0.05 equiv), and PEP (17.5 mg, 85 μmol, 0.9 equiv) were dissolved in D_2O (1.7 mL) with $MgCl_2$ (10 mM). The pH was adjusted to 7.0. Subsequently, pyruvate kinase (50 U in 50 μL) and the recombinant glycerate-2-kinase (250 μL) were added. The enzymatic reaction was observed by NMR spectroscopy. ^{1}H NMR and ^{31}P NMR spectra were measured every 15 min.

Gram-scale preparation of D-glycerate-2-phosphate catalyzed by recombinant glycerate-2-kinase: D-Glyceric acid (1.0 g, 9.4 mmol, 1.0 equiv), ATP (260 mg, 0.5 mmol, 0.05 equiv), and PEP (1.75 g, 8.5 mmol, 0.9 equiv) were dissolved in $MgCl_2$ solution (10 mM, 200 mL). The pH was adjusted to 7.0, pyruvate kinase (1000 U in 1000 μL) and 1100 μL glycerate-2-kinase (*T. maritima*) were added, and the mixture was stirred gently. The conversion of the enzymatic reaction was observed by NMR and reached completion after 24 h. Subsequently, the enzyme was removed by filtration through a 10000 MW Centricon filter membrane. Calcium acetate monohydrate (28.2 mmol, 3 equiv) was added to the resulting reaction solution containing D-glycerate 2-phosphate (3.3 g) to precipitate the D-glycerate 2-phosphate as its calcium salt and allow harvesting by filtration. This calcium salt was finally converted into the corresponding lithium salt by using lithium-conditioned Dowex 50W X8 H+ ion exchange resin (150 mL). The solution of D-glycerate 2-phosphate lithium salt was evaporated to give 1.255 g white solid product in 72% yield and excellent purity. ^{1}H NMR (D_2O, 600 MHz): $\delta = 4.48$ (ddd, $J = 8.6$, 5.4, 3.1 Hz, 1H), 3.89 (dd, $J = 11.9$, 3.1, 1H), 3.82 ppm (dd, $J = 11.9$, 5.4 Hz, 1H); ^{13}C NMR (D_2O, 151 MHz): $\delta =$ 177.4 (d, $J = 5.6$ Hz), 76.7 (d, $J = 5.5$ Hz), 64.2 ppm (d, $J = 3.6$ Hz); ^{31}P NMR (D_2O, 162 MHz): $\delta = 2.2$ ppm (s, 1P); MS: *m/z* calcd for $C_3H_7O_7P^-$: 185.0 $[M-H]^+$, found: 185.0; capillary electrophoresis: enantiomeric ratio D/L = 98.3:1.7

Acknowledgements

We would like to thank Dr. Eduard Frick for capillary electrophoresis analyses and the German Federal Ministry of Education and Research (BMBF) for the support of project P28 under the Biocatalysis2021 cluster. The German Academic Exchange Service (DAAD) is gratefully acknowledged for the grant supporting B.M.K.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: asymmetric biocatalysis · chiral metabolite · D-glycerate-2-phosphate · enzymatic phosphorylation · glycerate-2-kinase

[1] O. Meyerhof, W. Kiessling, *Biochem. Z.* **1935**, *276*, 239-.
[2] a) V. Pancholi, *Cell. Mol. Life Sci.* **2001**, *58*, 902–920; b) M. Kaufmann, P. Bartholmes, *Caries Res.* **1992**, *26*, 110–116; c) Y. Wu, C. Wang, S. Lin, M. Wu, L. Han, C. Tian, X. Zhang, J. Zang, *Acta Crystallogr. Sect. D* **2015**, *71*, 2457–2470; d) J. A. Kornblatt, V. Quiros, M. J. Kornblatt, *PLoS One* **2015**, *10*, e0135754; e) B. C. Figueiredo, A. A. Da'dara, S. C. Oliveira, P. J. Skelly, *PLoS Pathog.* **2015**, *11*, e1005335; f) V. Hannaert, M.-A. Albert, D. J. Rigden, M. T. da Silva Giotto, O. Thiemann, R. C. Garratt, J. Van Roy, F. R. Opperdoes, P. A. M. Michels, *Eur. J. Biochem.* **2003**, *270*, 3205–3213.
[3] C. E. Ballou, H. O. L. Fischer, *J. Am. Chem. Soc.* **1954**, *76*, 3188–3193.
[4] F. C. Hartman, F. Wold, *Biochim. Biophys. Acta Gen. Subj.* **1967**, *141*, 445–453.
[5] P. A. Sims, G. H. Reed, *J. Mol. Catal. B* **2005**, *32*, 77–81.
[6] J. O. Sass, K. Fischer, R. Wang, E. Christensen, S. Scholl-Bürgi, R. Chang, K. Kapelari, M. Walter, *Hum. Mutat.* **2010**, *31*, 1280–1285.
[7] a) R. Wohlgemuth, A. Liese, W. Streit, *Trends Biotechnol.* **2017**, *35*, 452–465; b) D. Gauss, B. Schönenberger, G. S. Molla, B. M. Kinfu, J. Chow, A. Liese, W. Streit, R. Wohlgemuth in *Applied Biocatalysis: From Fundamental Science to Industrial Applications* (Eds.:A. Liese, L. Hilterhaus, U. Kettling, G. Antranikian), Wiley-VCH, Weinheim, **2016**, pp. 147–177.
[8] W. Lamprecht, F. Heinz, T. Diamartstein, *Zeitschrift Physiolog. Chem.* **1962**, *328*, 204–206.
[9] L. Zelcbuch, M. Razo-Mejia, E. Herz, S. Yahav, N. Antonovsky, H. Kroytoro, R. Milo, A. Bar-Even, *PLoS One* **2015**, *10*, e0122957.
[10] B. Liu, Y. Hong, L. Wu, Z. Li, J. Ni, D. Sheng, Y. Shen, *Extremophiles* **2007**, *11*, 733–739.
[11] a) M. Reher, M. Bott, P. Schönheit, *FEMS Microbiol. Lett.* **2006**, *259*, 113–119; b) M. Reher, P. Schönheit, *FEBS Lett.* **2006**, *580*, 1198–1204.
[12] M. Noh, J. H. Jung, S. B. Lee, *Biotechnol. Bioprocess Eng.* **2006**, *11*, 344–350.
[13] B. Liu, L. Wu, T. Liu, Y. Hong, Y. Shen, J. Ni, *Biotechnol. Lett.* **2009**, *31*, 1937–1941.
[14] T. Yoshida, K. Fukuta, T. Mitsunaga, H. Yamada, Y. Izumi, *Eur. J. Biochem.* **1992**, *210*, 849–854.
[15] D. Kehrer, H. Ahmed, H. Brinkmann, B. Siebers, *BMC Genomics* **2007**, *8*, 301.
[16] R. Schwarzenbacher, D. McMullan, S. S. Krishna, Q. Xu, M. D. Miller, J. M. Canaves, M. A. Elsliger, R. Floyd, S. K. Grzechnik, L. Jaroszewski, H. E. Klock, E. Koesema, J. S. Kovarik, A. Kreusch, P. Kuhn, T. M. McPhillips, A. T. Morse, K. Quijano, G. Sparggon, R. C. Stevens, H. van den Bedem, G. Wolf, K. O. Hodgson, J. Woooley, A. M. Deacon, A. Godzik, S. A. Lesley, I. A. Wilson, *Proteins Struct. Funct. Bioinf.* **2006**, *65*, 243–248.
[17] C. Yang, D. A. Rodionov, I. A. Rodionova, Y. Li, A. L. Osterman, *J. Bacteriol.* **2008**, *190*, 1773–1782.
[18] a) R. Matsumi, C. Hellriegel, B. Schoenenberger, T. Milesi, J. van der Oost, R. Wohlgemuth, *RSC Adv.* **2014**, *4*, 12989–12994; b) D. Gauss, B. Schoenenberger, R. Wohlgemuth, *Carbohydr. Res.* **2014**, *389*, 18–24.

Manuscript received: April 11, 2017
Accepted manuscript online: April 18, 2017
Version of record online: June 19, 2017

CHEMBIOCHEM

Supporting Information

Biocatalytic Asymmetric Phosphorylation Catalyzed by Recombinant Glycerate-2-Kinase

Norman Hardt,[a] Birhanu M. Kinfu,[b] Jennifer Chow,[b] Bernhard Schoenenberger,[a] Wolfgang R. Streit,*[b] Markus Obkircher,[a] and Roland Wohlgemuth*[a]

cbic_201700201_sm_miscellaneous_information.pdf

Supporting Information

1) Capillary zone electrophoresis

D-Glycerate-2-phosphate lithium salt

PRINZIP
CZE, EOF-Umkehr (dynam. Beschichtung), indirekte Detektion

GERAETEEINSTELLUNGEN
Detektor: PDA-Detektor
Detektionswellenlaenge: 233 nm (indirekt)
Referenzwellenlaenge: keine
Temperatur (Kapillare): 25 C
Kapillare: unbeschichtete Kapillare,
75 um Innendurchmesser,
120 cm Laenge, 110 cm bis Detektor
REAGENZIEN
- 95283 WASSER
- 72082 Natriumhydroxid-Loesung (1 M)
- 94747 Vancomycin HCl
- CEofix(TM) anions 5 Kit (analis)
Anstelle der angegebenen Reagenzien duerfen andere gleichwertige
Reagenzien verwendet werden.

LOESUNGEN
a) Elektrophoresepuffer (EP): Accelerator aus CEofix anions 5 Kit
b) Conditioner (Cond): Initiator aus CEofix anions 5 Kit
c) Vancomycin-Loesung (Vanc): 60 mg Vancomycin in 1.0 mL EP loesen (~40 mM).
d) Probeloesung: Wenig Probe einwiegen (~5 mg), zu 5 mg/mL in Wasser loesen.

POSITIONIERUNG
BI:
A B C D E F
leer 3
95283 95283 2
EP EP 72082 Cond 1

BO:
A B C D E F
EP leer leer 1

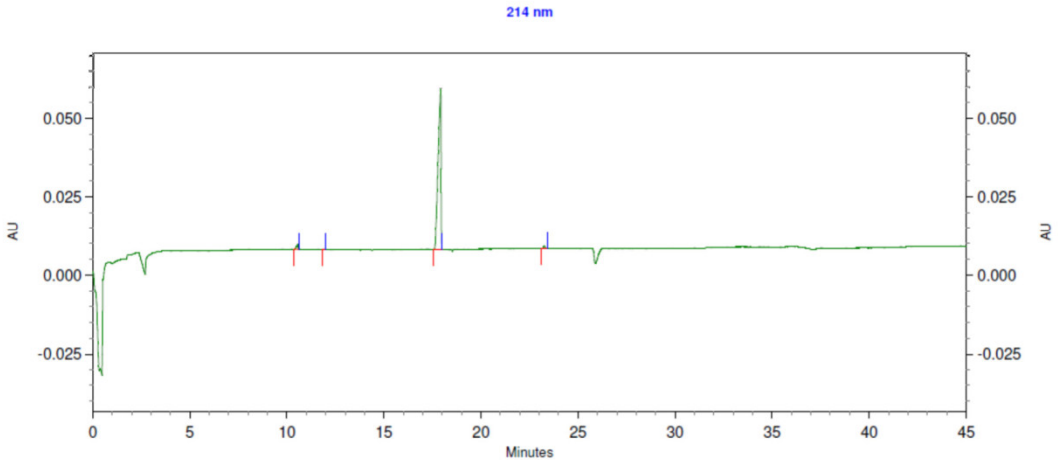
PROGRAMM
(ani5_120.met)
Druck/Strom Zeit inlet outlet

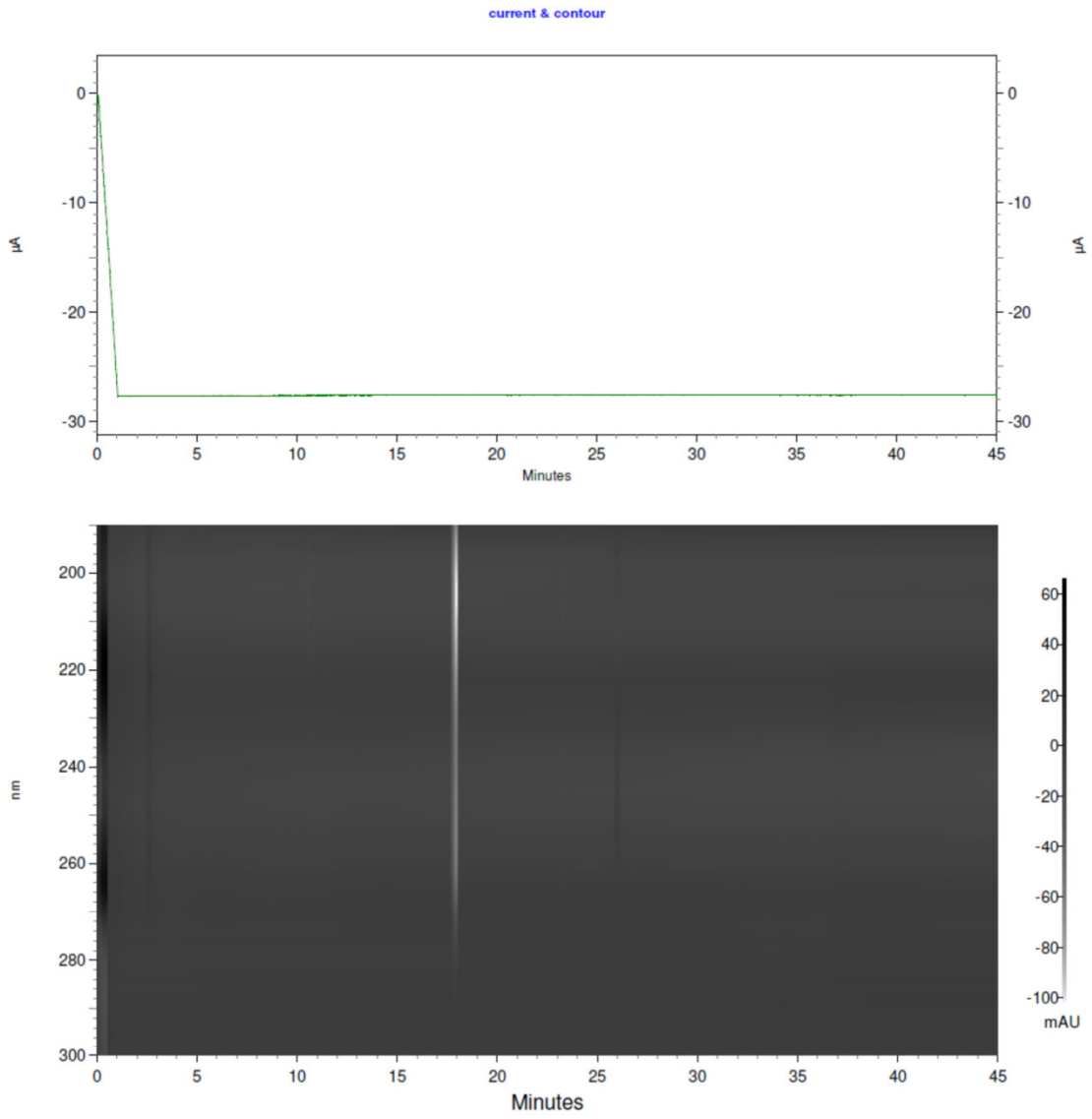
Rinse-Pressure 20 psi 1 min BI:D1 BO:C1
Rinse-Pressure 20 psi 1 min BI:B1 BO:C1
Inject-Pressure 0.5 psi 5 sec SI:A1 BO:C1
Inject-Pressure 0.1 psi 10 sec BI:D2 BO:C1
Separate Volt. -30 kV 45 min BI:A1 BO:A1 1 min Ramp
Rinse-Pressure 40 psi 1.5 min BI:C1 BO:C1
Rinse-Pressure 40 psi 1.5 min BI:C2 BO:C1

AUSWERTUNG
Korrigierte Flaechenprozente.
Hintergrundsingale nicht integrieren.

PDA - 233nm
Results

Pk #	Migration Time	Area	Corrected Area	Corrected Area Percent
1	10.567	8545	1483	2.646
2	11.913	547	84	0.150
3	17.925	528832	54088	96.534
4	23.271	4764	375	0.670
Totals		542688	56030	100.000




current & contour
µA
0
-10
-20
-30
0
5
10
15
20
25
30
35
40
45
Minutes
nm
200
220
240
260
280
300
60
40
20
0
-20
-40
-60
-80
-100
mAU
Minutes

CEER

PRINZIP
CZE, EOF-Umkehr (dynam. Beschichtung), Vancomycin als chiraler Selektor,
2-Zonen-Fuellung der Kapillare (Einlassseite Vancomycin=chirale Trennzone;
Auslassseite EP ohne Vancomycin), indirekte Detektion

GERAETEEINSTELLUNGEN
Detektor: PDA-Detektor
Detektionswellenlaenge: 233 nm (indirekt)
Referenzwellenlaenge: keine
Temperatur (Kapillare): 25 C
Kapillare: unbeschichtete Kapillare,
75 um Innendurchmesser,
120 cm Laenge, 110 cm bis Detektor
REAGENZIEN
- 95283 WASSER
- 72082 Natriumhydroxid-Loesung (1 M)
- 94747 Vancomycin HCl
- CEofix(TM) anions 5 Kit (analis)
Anstelle der angegebenen Reagenzien duerfen andere gleichwertige
Reagenzien verwendet werden.

LOESUNGEN
a) Elektrophoresepuffer (EP): Accelerator aus CEofix anions 5 Kit
b) Conditioner (Cond): Initiator aus CEofix anions 5 Kit
c) Vancomycin-Loesung (Vanc): 60 mg Vancomycin in 1.0 mL EP loesen (~40 mM).
d) Probeloesung: Wenige mg der Probe werden zu 5 mg/mL in Wasser geloest.

POSITIONIERUNG
BI:
A B C D E F
leer 3
95283 95283 2
EP EP 72082 Cond Vanc 1

BO:
A B C D E F
EP leer leer 1

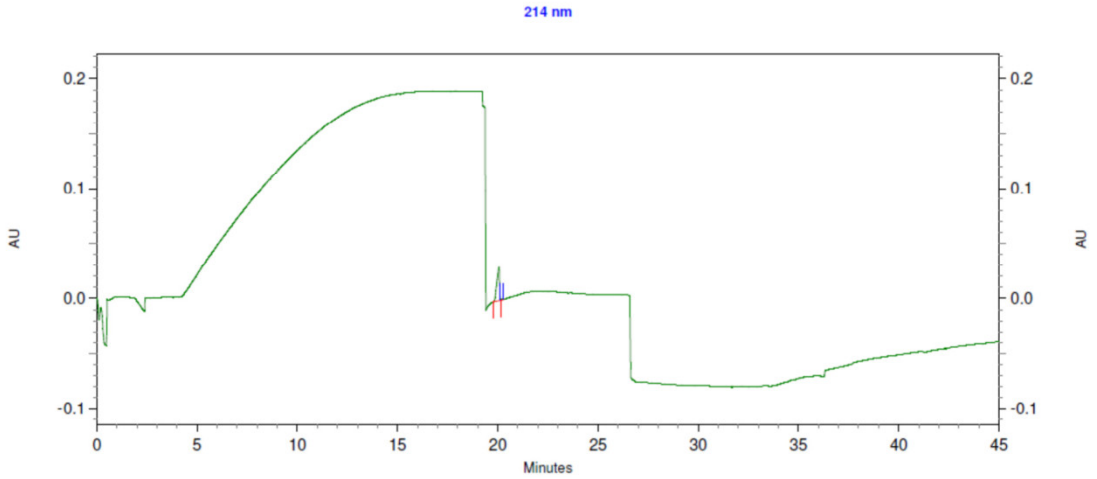
PROGRAMM
(vanc_120.met)
Druck/Strom Zeit inlet outlet

Rinse-Pressure 20 psi 1 min BI:D1 BO:C1
Rinse-Pressure 20 psi 1 min BI:B1 BO:C1
Rinse-Pressure 20 psi .6 min BI:E1 BO:C1
Inject-Pressure 0.5 psi 5 sec SI:A1 BO:C1
Inject-Pressure 0.1 psi 10 sec BI:D2 BO:C1
Separate Volt. -30 kV 45 min BI:A1 BO:A1 1 min Ramp
Rinse-Pressure 40 psi 1.5 min BI:C1 BO:C1
Rinse-Pressure 40 psi 1.5 min BI:C2 BO:C1

AUSWERTUNG
Korrigierte Flaechenprozente der Singale fuer D- und fuer L-Enantiomer
bestimmen:
M(t) D-2-Phosphoglycerinsaeure ~22.3 min
M(t) L-2-Phosphoglycerinsaeure ~23.3 min

PDA - 233nm
Results

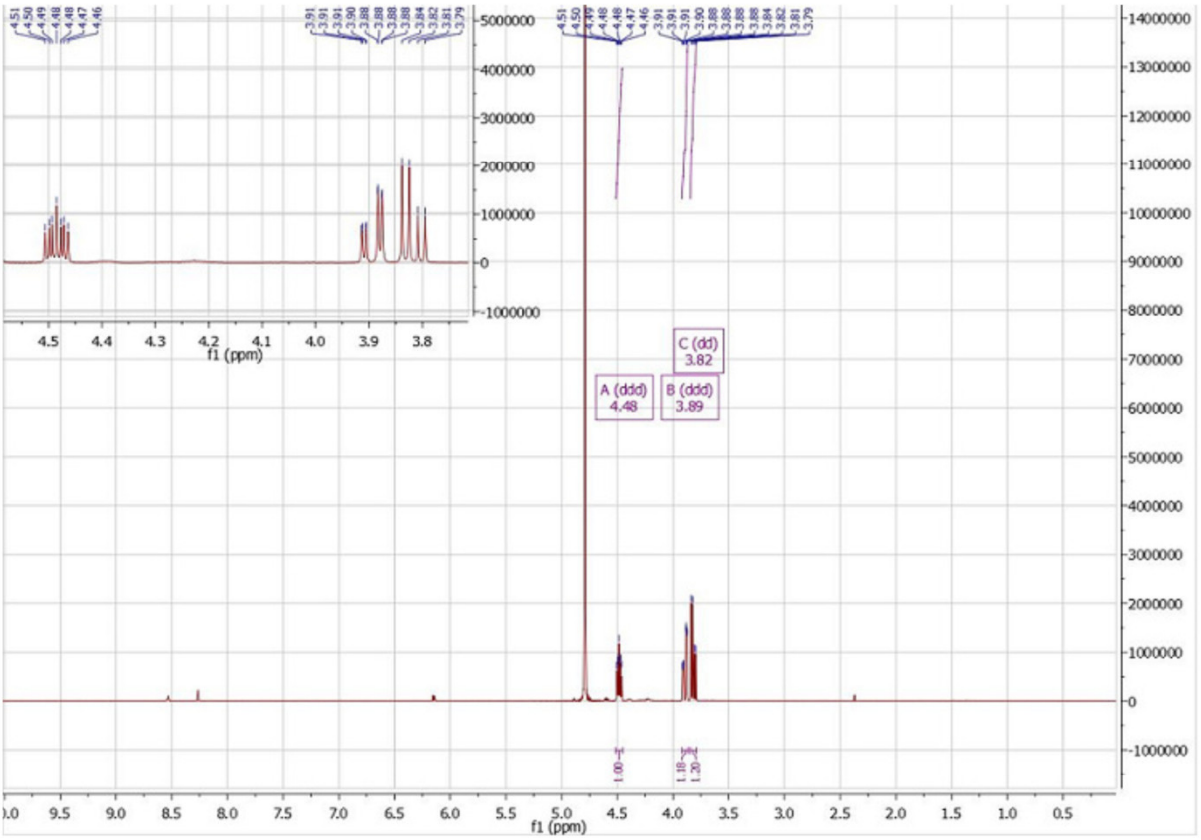
Pk #	Migration Time	Area	Corrected Area	Corrected Area Percent
1	20.071	273719	25002	98.325
2	20.221	4699	426	1.675
Totals		278418	25428	100.000



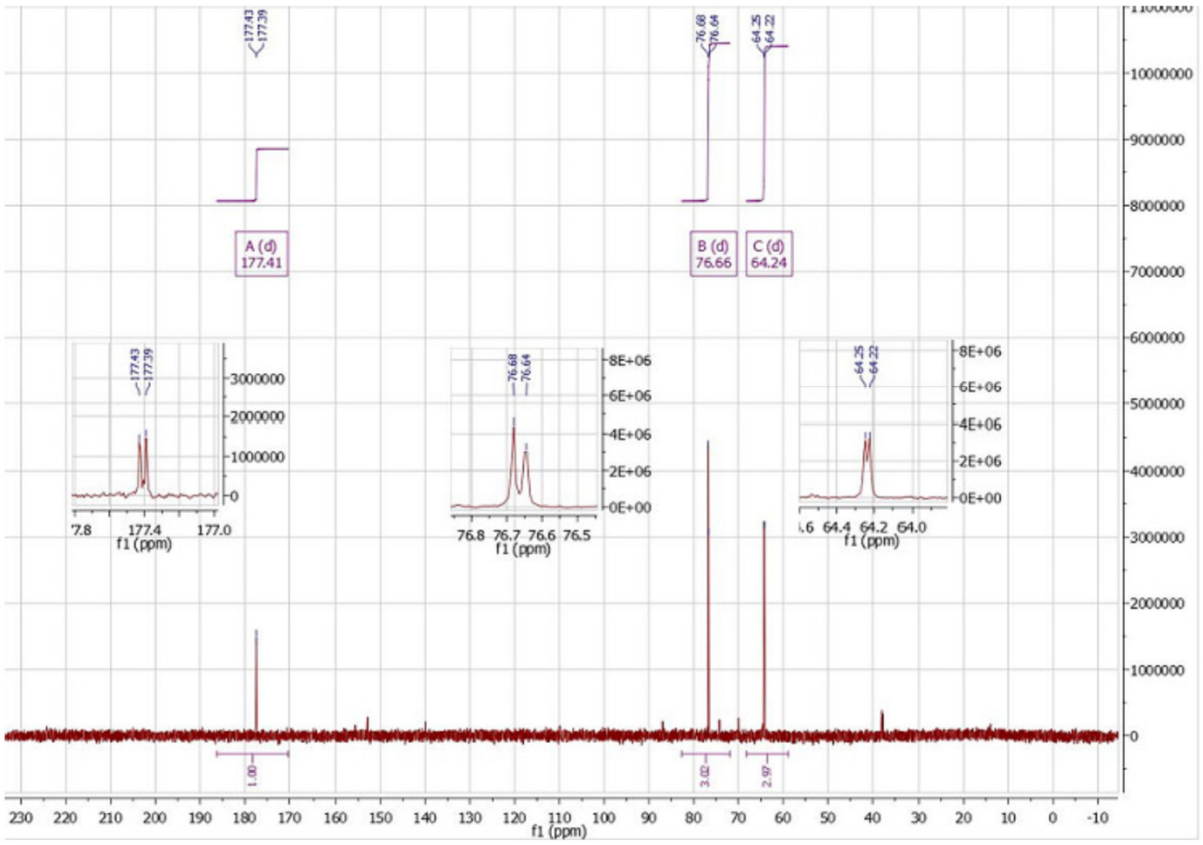
2) NMR

D-Glycerate-2-phosphate lithium salt

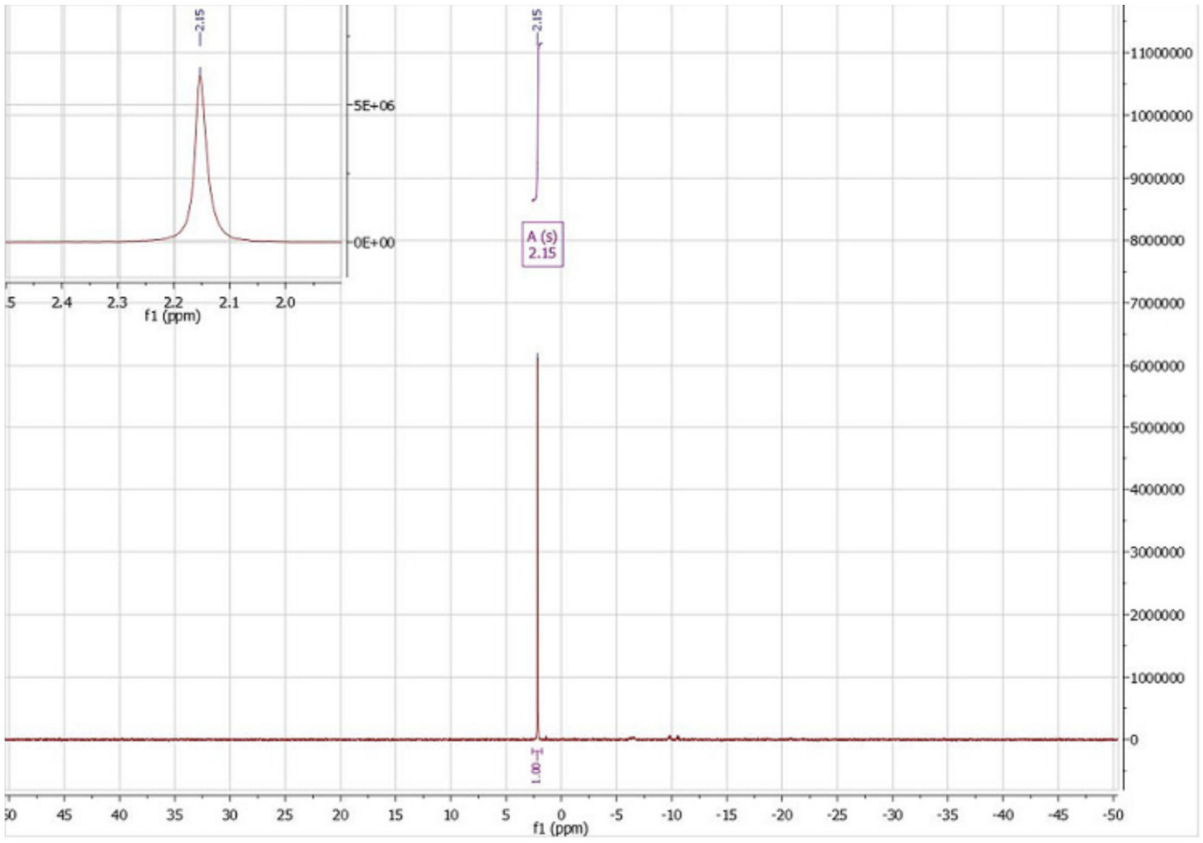
^{1}H-NMR



^{13}C-NMR


177.43
177.39
76.68
76.64
64.35
64.22
A (d)
177.41
B (d)
76.66
C (d)
64.24
f1 (ppm)
1.00
3.02
2.97

^{31}P-NMR


2.15
5E+06
0E+00
2.5 2.4 2.3 2.2 2.1 2.0
f1 (ppm)
2.15
A (s)
2.15
1.00
11000000
10000000
9000000
8000000
7000000
6000000
5000000
4000000
3000000
2000000
1000000
0
50 45 40 35 30 25 20 15 10 5 0 -5 -10 -15 -20 -25 -30 -35 -40 -45 -50
f1 (ppm)

3) LC-MS

LC-MS System:
LC: Waters alliance 2695 / SN: M97SM4963M
Column Oven: Waters 2695 / SN: C00SMH064M
MS: Waters micromass ZQ / SN: LAA802

LC :

Syringe Pump Flow (µl/min): 5	Equilibration Time (mins): 10

Gradient:

	Time	Flow	%A	%B	%C	%D	Curve
1	0.01	0.30	95.0	0.0	5.0	0.0	6
2	1.20	0.30	95.0	0.0	5.0	0.0	6
3	10.00	0.30	65.0	0.0	35.0	0.0	6
4	20.00	0.30	65.0	0.0	35.0	0.0	6
5	21.00	0.30	5.0	0.0	95.0	0.0	6
6	30.00	0.30	5.0	0.0	95.0	0.0	6
7	30.10	0.30	95.0	0.0	5.0	0.0	6

Solvent A: Tributylamine-Buffer pH 4.95
Solvent B: Acetonitrile
Solvent C: Methanol
Solvent D: Methanol

Column:	**Column-Oven:**
Ascentis Express C18, 2.1 x 50 mm, 2.7 µm, PN: 53822-U, SN: USMD003132	Column Temperature (°C): 35 (± 5°C) Sample Temperature (°C): 25 (± 5°C)

UV/VIS: 3D Data Collection from 210 nm to 600 nm

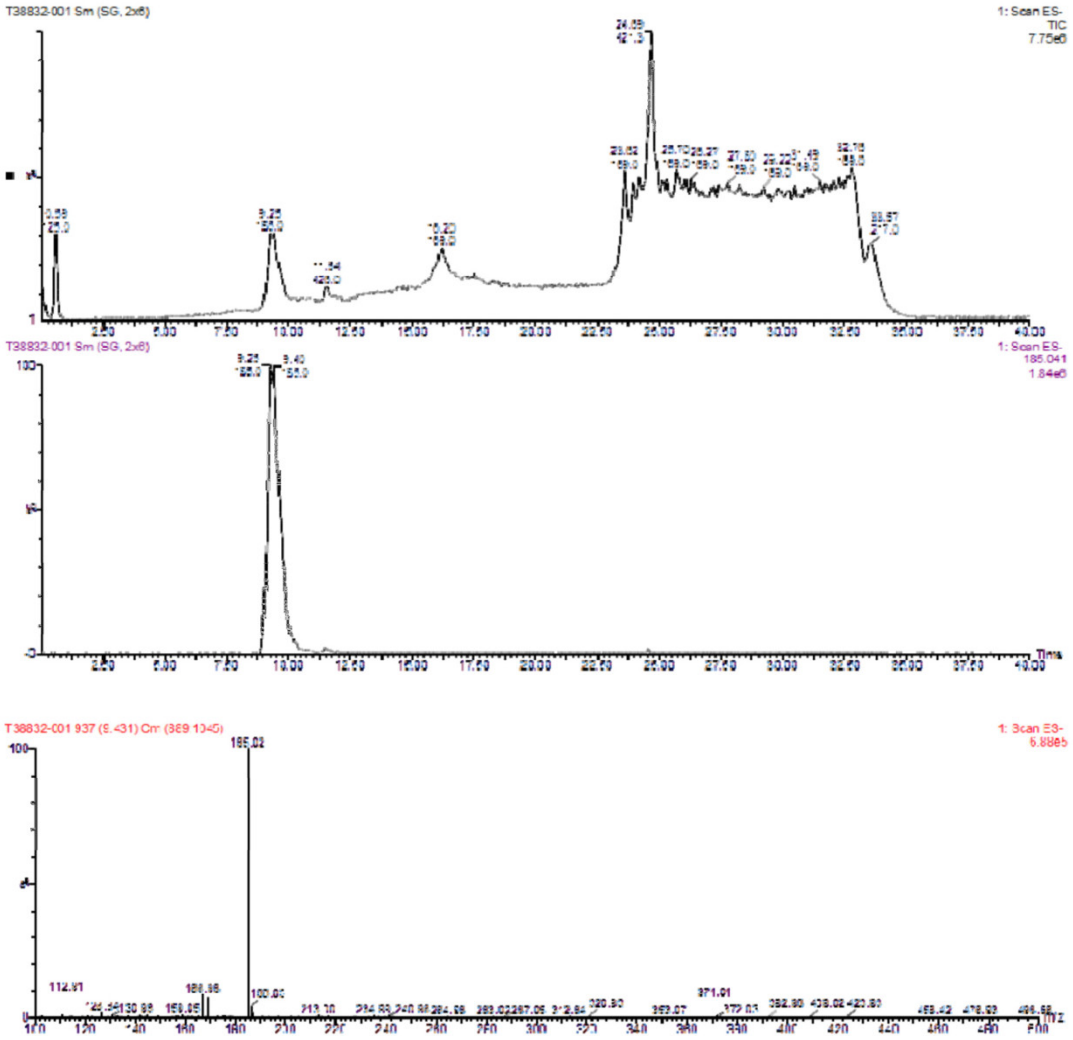
MS :

MS Scan, Time 0 to 40, Mass 100 to 500 ES-, CV 20

Gas Flow:	**Temperatures in °C**
Desolvation (L/h): 600 Cone (L/h): 60	Desolvation: 400 Source: 140

Voltages:	**Analyser:**
Capillary (kV): 2.5 Cone (V): 25 Extractor (V): 4 RF Lens (V): 0.5	LM Resolution: 15 HM Resolution: 15 Ion Energy: 0.5 Multiplier: 650

Solvent A: **Tributylamine-Buffer pH 4.95:**
2382.4 µl Tributylamin + 859.4 µl Acetic acid + 1000 ml Water


T38832-001 Sm (SG, 2x6)
1: Scan ES-
TIC
T38832-001 Sm (SG, 2x6)
1: Scan ES-
185.041
T38832-001 937 (9.431) Cm (369:1045)
1: Scan ES-
185.02
112.91
100.00
371.01

4 Bioreaction engineering leading to efficient synthesis of L-glyceraldehyde-3-phosphate

Getachew S. Molla[1], Birhanu M. Kinfu[2], Jennifer Chow[2], Wolfgang R. Streit[2], Roland Wohlgemuth[3], and Andreas Liese[1]

[1]Institute of Technical Biocatalysis, Hamburg University of Technology, Hamburg, Germany; [2]Department of Microbiology and Biotechnology, University of Hamburg, Hamburg, Germany; [3]Sigma-Aldrich Chemie GmbH, Member of Merck Group, Buchs, Switzerland

Published in:

Biotechnology Journal, 12(3): 1600625 (1-7). doi:10.1002/biot.201600625

Biotechnology Journal
www.biotechnology-journal.com

Biotechnol. J. 2017, 12, 1600625

DOI 10.1002/biot.201600625

Research Article

Bioreaction engineering leading to efficient synthesis of L-glyceraldehyde-3-phosphate

Getachew S. Molla[1], Birhanu M. Kinfu[2], Jennifer Chow[2], Wolfgang Streit[2], Roland Wohlgemuth[3] and Andreas Liese[1]

[1] Institute of Technical Biocatalysis, Hamburg University of Technology, Hamburg, Germany
[2] Department of Microbiology and Biotechnology, University of Hamburg, Hamburg, Germany
[3] Sigma-Aldrich Chemie GmbH, Member of Merck Group, Buchs, Switzerland

Enantiopure L-glyceraldehyde-3-phosphate (L-GAP) is a useful building block in natural biological and synthetic processes. A biocatalytic process using glycerol kinase from *Cellulomonas* sp. (EC 2.7.1.30) catalyzed phosphorylation of L-glyceraldehyde (L-GA) by ATP is used for the synthesis of L-GAP. L-GAP has a half-life of 6.86 h under reaction conditions. The activity of this enzyme depends on the Mg^{2+} to ATP molar ratio showing maximum activity at the optimum molar ratio of 0.7. A kinetic model is developed and validated showing a 2D correlation of 99.9% between experimental and numerical data matrices. The enzyme exhibits inhibition by ADP, AMP, methylglyoxal and Ca^{2+}, but not by L-GAP and inorganic orthophosphate. Moreover, equal amount of Ca^{2+} exerts a different degree of inhibition relative to the activity without the addition of Ca^{2+} depending on the Mg^{2+} to ATP molar ratio. If the Mg^{2+} to ATP molar ratio is set to be at the optimum value or less, inorganic hexametaphosphate (PP_i6) suppresses the enzyme activity; otherwise PP_i6 enhances the enzyme activity. Based on reaction engineering parameters such as conversion, selectivity and specific productivity, evaluation of different reactor types reveals that batchwise operation via stirred-tank reactor is the most efficient process for the synthesis of L-GAP.

Received	11 OCT 2016
Revised	29 NOV 2016
Accepted	15 DEC 2016
Accepted article online	19 DEC 2016

Supporting information available online

Keywords: Glycerol kinase · L-Glyceraldehyde-3-phosphate instability · Mg^{2+} to ATP ratio · Reaction kinetics · Reactor simulation

1 Introduction

For the past decades, there has been a remarkable increase of in vitro applications of biocatalysts to synthesize high value-added products [1–3]. Due to their inherent chirality, biocatalysts offer a wide range of in vitro synthetic routes to produce useful enantiopure chiral compounds such as pharmaceuticals, food additives, reagents in biomedical research and standards to develop analytical technologies. The synthesis of enantiopure metabolites is a necessity in the analysis of enzyme functions as well as for detailed stereochemical investigations of metabolic pathways [4]. As phosphorylated metabolites are thereby of major importance and their chemical synthesis is challenging due to limitations in asymmetric phosphorylation methods, the large diversity of phosphorylating enzymes offers great opportunities for the application of biocatalytic phosphorylation [5].

Phosphorylated metabolites such as L-glyceraldehyde-3-phosphate (L-GAP) are synthetically useful to design in vitro single-pot cascade biocatalytic reaction sequences using enzymes such as aldolases or transketolases [6–10]. L-GAP has been described to have a bactericidal effect on *Escherichia coli* while D-GAP does not have the same effect [11]. On the other hand, *E. coli* (YghZ) enantioselectively transforms L-GAP to *sn*-glycerol-3-phosphate using an NADPH dependent aldose reductase as a complement

Correspondence: Prof. Dr. Andreas Liese, Institute of Technical Biocatalysis, Hamburg University of Technology (TUHH), Denickestr. 15, 21073 Hamburg, Germany
E-mail: liese@tuhh.de

Abbreviations: AMP, adenosine-5'-monophosphate; **ADP**, adenosine-5'-diphosphate; **ATP**, adenosine-5'-triphosphate; **DAD**, diode array detector; **DHAP**, dihydroxyacetone phosphate; **DSP**, downstream processing; **L-GA**, L-glyceraldehyde; **L-GAP**, L-glyceraldehyde-3-phosphate; **PPi6**, inorganic hexametaphosphate; **TIM**, triose phosphate isomerase

Biotechnology Journal
www.biotechnology-journal.com

Biotechnol. J. 2017, 12, 1600625

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

of triose phosphate isomerase (TIM) deficiency that enters into gluconeogenesis metabolic pathway [12]. Thus, practical synthetic methodology for the preparation of enantiopure L-GAP is of a major advantage for metabolic elucidation purposes. Several biocatalytic reaction systems for the synthesis of L-GAP have been described using glycerol kinase (EC 2.7.1.30) catalyzed kinetic resolution of DL-glyceraldehyde (DL-GA) or phosphorylation of L-glyceraldehyde (L-GA) [13–16].

A comprehensive bioreaction engineering study is required in order to design an efficient practical scale biocatalytic process for the synthesis of L-GAP. Therefore, the present paper describes detailed reaction engineering analyses including substrate and enzyme screening, the effect of Mg^{2+} to ATP molar ratio on the catalytic properties of the enzyme as well as on the reaction thermodynamics, stability of L-GAP under conditions of the reaction, reaction and process kinetics models development and experimental as well as numerical performance evaluation of different reactor types for an optimized glycerol kinase catalyzed synthesis of L-GAP.

2 Materials and methods

2.1 Materials

Adenosine-5'-triphosphate (ATP) disodium salt, adenosine-5'-diphosphate (ADP) sodium salt, adenosine-5'-monophosphate (AMP) disodium salt, DL-glyceraldehyde (DL-GA), D-glyceraldehyde (D-GA), L-glyceraldehyde (L-GA), methylglyoxal, trichloroacetic acid (TCA), ammonium acetate, sodium hexametaphosphate (PP_6n), glycerol kinase from *Cellulomonas* sp., glycerol kinase from *E. coli* and glycerol kinase from *Streptomyces canus* were purchased from Sigma-Aldrich GmbH (Buchs, Switzerland). Calcium chloride, magnesium chloride hexahydrate, phosphoric acid, potassium dihydrogen phosphate, sulfuric acid, triethanolamine and acetonitrile HPLC-gradient grade were purchased from Carl Roth GmbH (Karlsruhe, Germany). All chemicals and solvents were used without further purification.

2.2 Methods

The concentrations of DL-GA, D-GA, L-GA and DL-GAP were analyzed by HPLC (Agilent 1100, Hewlett Packard) on a Eurokat-H column (300 mm × 8 mm, Knauer) with 5 mM H_2SO_4 as eluent at a flow rate of 0.5 mL/min and 75°C, using a refractive index detector at 35°C. Typical retention times were 15.3 ± 0.1 min for DL-GA and 9.8 ± 0.2 min for DL-GAP. All reactions and enzyme activity assays which were analyzed by measuring the concentrations of DL-GA, D-GA, L-GA and DL-GAP were quenched by the addition of 10% w/v TCA stop reagent at a ratio of 1:1 (sample: stop reagent, v/v), followed by vigorous mixing. The concentrations of ATP, ADP and AMP were analyzed by HPLC (Agilent 1100, Hewlett Packard), using a Nucleodur® HILIC column (250 mm × 4 mm, Macherey-Nagel) and acetonitrile: 100 mM ammonium acetate in aqueous solution (70:30 v/v) pH 5.3 as eluent at a flow rate of 0.8 mL/min and 25°C, detected by using a diode array detector (DAD) at 259 nm. Typical retention times for AMP, ADP and ATP were 6.2 ± 0.1 min, 9 ± 0.1 min and 13 ± 0.3 min, respectively. All reactions and enzyme activity assays which were analyzed by measuring the concentrations of ATP, ADP and AMP, were quenched by the addition of 74 mM phosphate buffer pH 1.8 stop reagent at a ratio of 1:1 (sample: stop buffer, v/v), followed by vigorous mixing that generated a final HPLC sample of pH 2.8. After the addition of the stop buffer it was important to maintain the final pH of HPLC samples between 2.5 and 3.5, because ATP and APD are instable at pH below 2 while at pH above 4 the enzyme could not be fully deactivated.

The operational stabilities of glycerol kinases from *Cellulomonas* sp., *E. coli* and *Streptomyces canus* were examined by incubating in 100 mM TEA buffer containing 100 mM ATP and 100 mM Mg^{2+} at pH 7.5, 8.5 and 9.5, and 25, 30 and 45°C. Remaining activities were measured routinely for a month. For substrate screening the activity of glycerol kinase from *Cellulomonas* sp. was assayed with regard to racemic DL-GA, D-GA and L-GA. Reactions for the activity assays were carried out using substrate solution of 25 mM D-GA or 25 mM L-GA or 25 mM DL-GA, 25 mM ATP and 30 mM of Mg^{2+} prepared in 100 mM TEA buffer, pH 8, at 30°C. All reactions were started by the addition of 0.05 mL of glycerol kinase from *Cellulomonas* sp. (1.2 mg/mL) prepared in 100 mM TEA buffer, pH 8, into 1.45 mL of the substrate solutions pre-incubated at 30°C. In order to estimate the required excess amount of ATP to achieve full conversion of L-GA in DL-GA due to the side enzymatic hydrolysis of ATP in the presence of D-GA, batch reactions were carried out using substrate solutions of fixed 15 mM ATP and 15 mM Mg^{2+} at 3 mM, 7 mM and 15 mM of DL-GA prepared in 100 mM TEA buffer, pH 8, at 30°C. All reactions were started by using the same procedure as used for the activity assay. The stability of DL-GAP was examined by incubating 40 mM of DL-GAP prepared in 100 mM TEA buffer, pH 8, at 25, 30 and 60°C.

To determine an optimum Mg^{2+} to ATP molar ratio, the activity of glycerol kinase from *Cellulomonas* sp. was assayed as a function of Mg^{2+} concentration, while maintaining fixed ATP concentrations. Two sets of enzyme activity assays were carried out as a function of Mg^{2+} concentrations up to 200 mM at two different fixed concentrations (30 mM and 70 mM) of ATP and 50 mM L-GA prepared in 100 mM TEA buffer, pH 8, and 30°C. All reactions were started by adding 0.05 mL of glycerol kinase from *Cellulomonas* sp. (1.2 mg/mL), prepared in 100 mM TEA buffer, pH 8, into 1.45 mL of the substrate solutions

Biotechnology Journal
www.biotechnology-journal.com

Biotechnol. J. 2017, 12, 1600625

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

pre-incubated at 30°C. Additionally, the effect of Mg^{2+} concentration on the reaction thermodynamics was examined by performing a reaction using 30 mM L-GA, 30 mM ATP and 500 mM Mg^{2+}, prepared in 100 mM TEA buffer, pH 8, at 30°C. In order to develop a reaction kinetic model, initial rate measurements as a function of concentrations of L-GA and ATP were carried out at a fixed 21 mM and 50 mM Mg^{2+} prepared in 100 mM TEA buffer, pH 8, at 30°C. Several sets of initial rate measurements as function of ATP concentration at fixed 50 mM L-GA, 21 mM and 50 mM Mg^{2+} and various concentrations (0 mM, 5 mM) of ADP were carried out in order to examine the inhibition type of glycerol kinase from *Cellulomonas* sp. by ADP and to determine the inhibition constant. Linear representations of Michaelis–Menten enzyme kinetic equation were applied in order to determine reaction kinetic parameters.

The inhibition of glycerol kinase from *Cellulomonas* sp. by AMP, methylglyoxal, DL-GAP, Ca^{2+}, inorganic hexametaphosphate (PP_i6) and inorganic orthophosphate (P_i) was investigated by measuring enzyme activities as function of component concentration. Reactions were carried out using substrate solutions of 30 mM L-GA, 30 mM ATP and 21 mM Mg^{2+}, prepared in 100 mM TEA buffer, pH 8, at 30°C. Moreover, the enzyme activity was examined as a function of PP_i6 and Ca^{2+} concentrations at different Mg^{2+} to ATP molar ratios in order to evaluate the effect of complexation between Mg^{2+} and PP_i6 as well as Ca^{2+} and ATP, respectively. Investigation of the effect of Ca^{2+} or other divalent metal ions on the activity of glycerol kinase is crucial in order to choose ATP salt of non-influencing counter cation or to perform an appropriate pretreatment. As a strategy to apply polyphosphate kinase (PPK)/PP_in mediated in situ ATP regeneration (results not included in this article) the effect of PP_i6 on the activity of glycerol kinase was examined.

3 Results

3.1 Enzyme and substrate screening

Screening of glycerol kinases from *Cellulomonas* sp., *E. coli* and *Streptomyces canus* has shown that glycerol kinase from *Cellulomonas* sp. exhibited the highest activity and stability (data shown in Supporting information, Table 1), leading to the selection of glycerol kinase from *Cellulomonas* sp. for L-GAP synthesis. Glycerol kinase catalyzes enantioselective phosphorylation of L-GA (data shown in Supporting information, Fig. 1); therefore, L-GAP can be synthesized either by kinetic resolution using racemic DL-GA or by phosphorylation of using enantiopure L-GA starting materials. However, using DL-GA as a starting material has a limitation, because glycerol kinase catalyzes hydrolysis of ATP to ADP and P_i in the presence of D-GA without effective phosphorylation of D-GA [16]. In an aqueous solution, the carbonyl functional group of D-GA exists in a reversible hydrated geminal-diol form. Therefore, in binding with the active site of glycerol kinase instead of the terminal –OH group as in the case of L-GA, one of the geminal-diols acts as a phosphoryl group acceptor that produces hemiacetal phosphate. The hemiacetal phosphate is labile and undergoes splitting-off the phosphate moiety [16–17]. Due to the enzymatic hydrolysis of ATP induced by the presence of D-GA, a 10-fold stoichiometric excess of ATP with respect to L-GA in DL-GA was required to achieve a full conversion of L-GA, when 3 mM DL-GA was used as a starting material. Consequently, phosphorylation using enantiopure L-GA as a starting material was selected for the synthesis of L-GAP.

Table 1. Kinetic parameters for the phosphorylation of L-glyceraldehyde with ATP catalyzed by glycerol kinase from *Cellulomonas* sp.: (0 mM, 50 mM) L-GA, (0 mM, 70 mM) ATP, 21 mM and 50 mM Mg^{2+} and (0 mM, 5 mM) ADP in 100 mM TEA buffer, pH 8, and 30°C

Reaction components	K_m (mM)	K_i (mM)	ν_{max} (U/mg)
ATP at 21 mM Mg^{2+}	0.8 ± 0.05	–	110 ± 3.90
ATP at 50 mM Mg^{2+}	1.90	–	110 ± 3.90
L-Glyceraldehyde	4 ± 0.06	–	110 ± 3.90
ADP at 21 mM Mg^{2+}	–	1.4 ± 0.18	–
ADP at 50 mM Mg^{2+}	–	3.30	–

3.2 Product instability

Triosephosphate metabolites such as dihydroxyacetone phosphate (DHAP) and DL-GAP are unstable at neutral and alkaline pH conditions [8, 15, 18–21]. The instability of L-GAP at nearly neutral pH conditions is a critical factor for a biocatalytic process development using glycerol

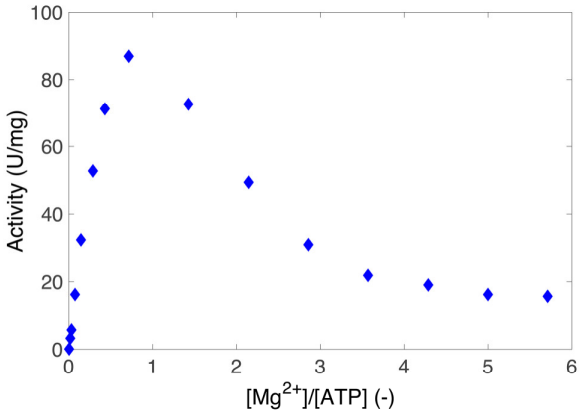


Figure 1. Activity of glycerol kinase as a function of the Mg^{2+} to ATP molar ratio: 30 mM L-GA, 70 mM ATP, (0 mM to 400 mM) Mg^{2+}, 0.02 mg/mL glycerol kinase from *Cellulomonas* sp. in 100 mM TEA buffer, pH 8, and 30°C

Biotechnology Journal
www.biotechnology-journal.com

Biotechnol. J. 2017, 12, 1600625

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

kinase from *Cellulomonas* sp., because the enzyme shows no activity at pH below 4, while L-GAP is stable at pH below 4. The rate of decomposition of DL-GAP in 100 mM TEA buffer, pH 8, at 25, 30 and 60°C can be defined by first order kinetic with rate constants of 3×10^{-5}, 2.8×10^{-5} and 1.2×10^{-3} s^{-1} as well as half-lives of 8.35, 6.86 and 0.15 h, respectively.

3.3 Reaction kinetics model development and validation

Development of a biocatalytic reaction kinetics model is a major tool to understand reaction mechanism and to rationally design an efficient bioprocess [22–24]. Glycerol kinase from *Cellulomonas* sp. essentially requires the cofactors ATP as phosphoryl donor and a bivalent metal ion, preferably Mg^{2+}. The interaction of Mg^{2+} and ATP forms Mg-ATP complexes of various physical and chemical features depending on Mg^{2+} to ATP molar ratio [25]. Thus the effect of Mg^{2+} to ATP molar ratio on the kinetics and thermodynamics of glycerol kinase catalyzed phosphorylation of L-GA were examined.

3.3.1 *Effects of Mg^{2+} to ATP ratio on reaction kinetics and thermodynamics*

For the synthesis of L-GAP using glycerol kinase catalyzed phosphorylation of L-GA, most often, equimolar [15, 26] or excess moles of Mg^{2+} relative to ATP have been applied [13, 16]. In this study the activity of glycerol kinase from *Cellulomonas* sp. was examined as a function of the Mg^{2+}/ATP molar ratio. As depicted in Fig. 1, glycerol kinase from *Cellulomonas* sp. shows no activity without the addition of Mg^{2+} and maximum activity at the Mg^{2+}/ATP molar ratio of 0.7. A subsequent increase of the Mg^{2+} to ATP molar ratio higher than the optimum suppresses the enzyme activity to a non-zero asymptotic value.

Despite being thermodynamically favorable with change in reaction Gibbs free energy ($\Delta_r G$) of -16.6 kJ/mol at pH 8 and ionic strength of 0.6 M [27], the effect of Mg^{2+} concentration on the equilibrium conversion for the phosphorylation of L-GA was examined by establishing a 23.7-fold higher Mg^{2+} to ATP molar ratio than the optimum Mg^{2+} to ATP molar ratio. Results showed that the equilibrium conversion is not affected; however, the enzyme activity decreases by a factor of 22 relative to its activity at the optimum Mg^{2+} to ATP molar ratio.

3.3.2 *Kinetic model development and validation*

Initial rate measurements as a function of concentrations of L-GA and ATP were carried out at fixed concentrations of 21 mM and 50 mM Mg^{2+}, establishing the optimum Mg^{2+} to ATP molar ratio of 0.7 at 30 mM and 71.43 mM ATP, respectively. The enzyme exhibits Michaelis–Menten kinetic behavior with respect to both substrates. As anticipated for initial rate measurements at 21 mM Mg^{2+} and at concentration of ATP higher than 30 mM, glycerol

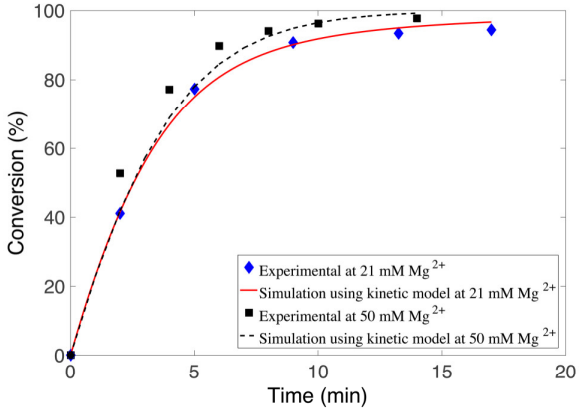


Figure 2. Correlation of experimental conversion data determined by HPLC measurement of ATP concentration and numerically simulated conversion data using the kinetic model for kinetic parameters at 21 and 50 mM Mg^{2+}: (1) for kinetic parameters at 21 mM Mg^{2+}: 30 mM L-glyceraldehyde, 30 mM ATP, 21 mM Mg^{2+}, 0.08 mg/mL glycerol kinase from *Cellulomonas* sp. in 100 mM TEA buffer pH 8 and 30°C and (2) for kinetic parameters at 50 mM Mg^{2+}: 70 mM L-glyceraldehyde, 40 mM ATP, 50 mM Mg^{2+}, 0.1 mg/mL glycerol kinase from *Cellulomonas* sp. in 100 mM TEA buffer, pH 8, and 30°C

kinase exhibits a pseudo-ATP surplus inhibition due to the drop of Mg^{2+} to ATP molar ratio below the optimum value (i.e. 0.7). This pseudo-ATP surplus inhibition can be circumvented by changing the concentration of Mg^{2+} that adjusts the Mg^{2+} to ATP molar ratio as no inhibition was observed at 50 mM Mg^{2+} until 70 mM ATP. Moreover, as given in Table 1 the enzyme shows different affinity constants (K_m) at 21 mM and 50 mM Mg^{2+}, because of different resulting concentrations Mg-ATP complex configurations. Examination of glycerol kinase from *Cellulomonas* sp. inhibition revealed that ADP, AMP, methylglyoxal and Ca^{2+} have been shown to be inhibitors, while DL-GAP and P_i are not (data shown in the Supporting information, Fig. 2). The inhibition of glycerol kinase by ADP as well as AMP and methylglyoxal is due to the competition with ATP and L-GA, respectively, for binding at the enzyme's active site.

The inhibition by Ca^{2+} is due to the formation of an enzymatically inactive Ca-ATP. Furthermore, the extent of glycerol kinase inhibition by Ca^{2+} was assayed at the optimum Mg^{2+} to ATP molar ratio (0.7), less than the optimum Mg^{2+} to ATP molar ratio (0.12) and higher than the optimum Mg^{2+} to ATP molar ratio (2.1). In all cases glycerol kinase displays inhibition by Ca^{2+}. However, the percentage of inhibition exerted by the same amount of Ca^{2+} relative to the activity without the addition of Ca^{2+} depends on the Mg^{2+} to ATP molar ratio. For example, 15 mM Ca^{2+} impedes 76.0, 55.3, and 29.7% of glycerol kinase inhibition at Mg^{2+} to ATP molar ratio of 0.12, 0.7 and 2.1, respectively. The different percentage of glycerol kinase inhibition by the same amount of Ca^{2+} at different

Biotechnology Journal
www.biotechnology-journal.com

Biotechnol. J. 2017, 12, 1600625

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Mg^{2+} to ATP molar ratios can be explained by mass action on the complexation equilibrium between Ca-ATP and Mg-ATP complexes. Addition of equal amount of Ca^{2+} yields different amounts of enzymatically inactive Ca-ATP complexes at different Mg^{2+} to ATP molar ratios; larger amount of Ca-ATP complex at less Mg^{2+} to ATP molar ratio. Likewise, the effect of PP_i6 on the activity of glycerol kinase depends on the magnitude of the Mg^{2+} to ATP molar ratio. If the Mg^{2+} to ATP molar ratio is set to be at the optimum value or below, PP_i6 suppresses the enzyme activity; whereas, at a higher Mg^{2+} to ATP molar ratio than the optimum value PP_i6 enhances the enzyme activity, until the excess amount of Mg^{2+} is titrated by PP_i6. The inhibition and activation of glycerol kinase by PP_i6 at different Mg^{2+} to ATP molar ratio is due to the complexation of PP_i6 with Mg^{2+}. Relevant reaction kinetic parameters and their value used for kinetics model development are listed in Table 1.

Eq. (1) shows the reaction kinetics equation defined by double substrate Michaelis–Menten kinetics including competitive product inhibition by ADP.

$$v = v_{max} \times \frac{[ATP]}{K_{m,ATP} \times \left(1 + \frac{[ADP]}{K_{i,ADP}}\right) + [ATP]} \times \frac{[L\text{-}GA]}{K_{m,L\text{-}GA} + [L\text{-}GA]} \quad (1)$$

The reaction kinetic model was validated by simulating the time courses of several batch reactions at different starting substrate concentrations. Sets of numerical simulations were performed using the same kinetic model but kinetics parameters at 21 and 50 mM Mg^{2+}. Evaluations of the experimental data and numerically simulated data matrices generated 2D correlation coefficients of 99.9% and 99.1% for kinetics parameters at 21 mM and 50 mM Mg^{2+}, respectively. Graphical correlations between experimental and numerically simulated conversions as function of reaction time are shown in Fig. 2 for batch reactions.

3.4 Process development and evaluation

An evaluation of different reactor types such as stirred tank reactor (STR), continuously operated stirred tank reactor (CSTR) and two- and five-stages cascade continuously operated stirred tank reactors was performed. Reaction engineering parameters such as conversion, selectivity and specific productivity were used as evaluation criteria of the reactors. Eq. (2) and Eq. (3) show differential equations for the simulation of substrate and product concentrations for a batch-wise mode of operation, respectively. Eq. (4) and Eq. (5) show steady-state differential equations for the simulation of substrate and product concentrations for a continuous mode of operation,

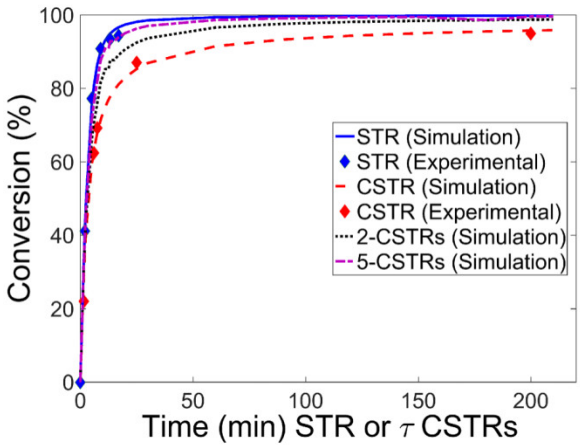


Figure 3. Comparison of STR, CSTR and two- and five-stages cascade CSTRs based on conversion as a function of reaction time for STR or residence time for the continuous operations for the synthesis of L-GAP: 30 mM L-GA, 30 mM ATP, 21 mM Mg^{2+}, 0.08 mg/mL glycerol kinase from *Cellulomonas* sp. in 100 mM TEA buffer, pH 8, and 30°C

respectively, composed of convection and reaction terms. $[S]_0$ and [S] represent influx and efflux substrate concentration, respectively. $[P]_0$ and [P] are influx product concentration, which is negligible, and efflux product concentration, respectively. As can be seen, Eq. (3) and Eq. (5) include the decomposition kinetics of L-GAP at the reaction conditions assuming the unit of reaction time and residence time in hours.

$$\frac{\partial[ATP]}{\partial t} = \frac{\partial[L\text{-}GA]}{\partial t} = [\text{Glycerol kinase}](-v_2) \quad (2)$$

$$\frac{\partial[L\text{-}GAP]}{\partial t} = [\text{Glycerol kinase}](v_2) \times e^{-0.101t} \quad (3)$$

$$\frac{\partial[ATP]}{\partial t} = \frac{\partial[L\text{-}GA]}{\partial t} = \frac{([S]o\text{-}[S])}{\tau} - [\text{Glycerol kinase}]v_2 \quad (4)$$

$$\frac{\partial[L\text{-}GAP]}{\partial t} = \left(\frac{([P]o\text{-}[P])}{\tau} + [\text{Glycerol kinase}]v_2\right) \times e^{-0.101\tau} \quad (5)$$

Conversion is defined as the number of converted L-GA molecules per starting L-GA molecules. An equal amount of enzyme was used for all the reactors to compare conversion. As can be seen in Fig. 3, the best performance can be achieved via STR. The cascade configurations of CSTR show better performance than CSTR allone. The higher the number of stages, the better the performance approximates to STR: 42-fold higher for CSTR, five-fold higher for 2-CSTRs and two-fold higher for 5-CSTRs amounts of the enzyme are required to achieve the same 98.5% conversion as STR at the same time.

Selectivity is defined as the number of synthesized L-GAP molecules per numbers of converted L-GA mole-

Biotechnology Journal
www.biotechnology-journal.com

Biotechnol. J. 2017, 12, 1600625

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

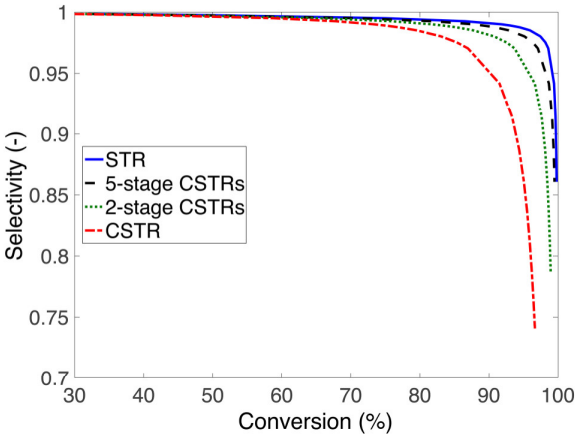


Figure 4. Comparison of STR, CSTR and two- and five-stages cascade CSTRs based on selectivity as a function conversion for the synthesis of L-GAP: 30 mM L-glyceraldehyde, 30 mM ATP, 21 mM Mg^{2+}, 0.08 mg/mL glycerol kinase from *Cellulomonas* sp. in 100 mM TEA buffer, pH 8, and 30°C

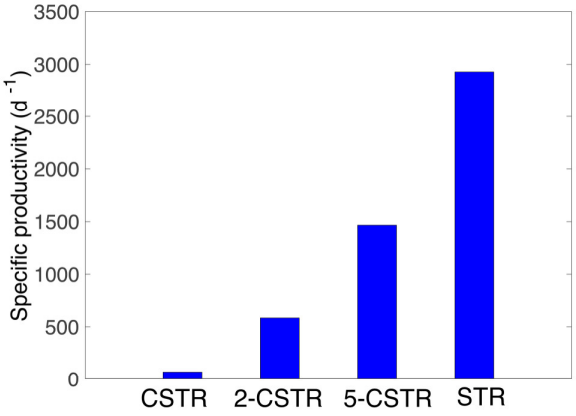


Figure 5. Comparison of STR, CSTR and two- and five stages cascade CSTRs based on specific productivity for the synthesis of L-GAP: 30 mM L-GA, 30 mM ATP, 21 mM Mg^{2+}, 0.08 mg/mL glycerol kinase from *Cellulomonas* sp. in 100 mM TEA buffer, pH 8, and 30°C

cules. For glycerol kinase catalyzed synthesis of L-GAP, the overall selectivity decreases exponentially with time due to the subsequent first order decay of L-GAP under the reaction conditions. The selectivity performances of the reactors were evaluated as a function of conversion applying equal amounts of enzyme. As can be seen in Fig. 4, the STR shows higher selectivity performance than the continuous operations; moreover, the selectivity of CSTR can be enhanced by establishing a cascade thereof, which approximates to a plug flow reactor with increasing number of CSTRs in the cascade.

Specific productivity is defined as the mass of synthesized L-GAP per unit mass of enzyme and unit time. Evaluation of reactor performance based on specific productivity is of major importance as the specific productivity in terms of process economy can be fine-tuned by multiplying reaction time and applied amount of enzyme. However, to decrease the cost of enzyme an increase of reaction time is required that affects process selectivity due to the decay of L-GAP. A fixed conversion of 98.5% was considered in order to evaluate the performance of the reactors based on specific productivity. STR offers the highest specific productivity, while CSTR performs least (Fig. 5) due to the low steady state reaction rate in CSTR that could be compensated by applying a high amount of enzyme while operating for a high number of residence times.

Biocatalyst consumption is defined as the unit mass of glycerol kinase consumed per unit mass of L-GAP synthesized. Recently, the synthesis of L-GAP via STR has been reported using the same reaction system with the biocatalyst consumption of 19.7 g kg^{-1} [15]. Applying the optimized conditions that were determined here, the biocatalyst consumption for the same reaction time and reactor type can be significantly reduced to 1.7 g kg^{-1}. Furthermore, downstream processing (DSP) based on the adsorption of the co-product ADP using packed bed activated carbon and precipitation of L-GAP was applied.

4 Discussion

Detailed reaction engineering aspects for the biocatalytic synthesis of L-GAP have been presented. Phosphorylation of enantiopure L-GA was selected instead of kinetic resolution of racemic DL-GA due to the enzymatic hydrolysis of ATP induced by D-GA. The half-life of L-GAP is 6.86 h under reaction conditions and its decomposition rate can be defined by first order kinetics with the rate constant of 2.8×10^{-5} s^{-1}. L-GAP decomposes faster at higher temperature, at 60°C its half-life is reduced to 0.15 h and the first order decay constant is increased of 1.2×10^{-3} s^{-1}.

The activity of glycerol kinase from *Cellulomonas* sp. depends on the Mg^{2+} to ATP molar ratio and exhibits a maximum at the optimum molar ratio of 0.7. Increasing the Mg^{2+} to ATP molar ratio suppresses the enzyme activity to a non-zero asymptotic value but does not affect the equilibrium conversion. An appropriate reaction kinetics model was developed following Michaelis–Menten kinetic behavior with respect to L-GA and ATP taking into account a competitive inhibition by the co-product ADP. Moreover, the enzyme kinetics with respect to ATP depends on Mg^{2+} concentration that exhibits pseudo-ATP surplus inhibition, if Mg^{2+} to ATP molar ratio reaches below the optimum value. The enzyme shows different K_m values with respect to ATP at different fixed concentrations of Mg^{2+}. Model validation showed 2D correlation coefficients of 99.9% and 99.1% for kinetic parameters at 21 and 50 mM Mg^{2+}, respectively between experimental

Biotechnology Journal
www.biotechnology-journal.com

Biotechnol. J. 2017, 12, 1600625

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

data and numerical data matrices. The enzyme is inhibited by AMP, methylglyoxal and Ca^{2+}, but not by L-GAP and P_i.

Equal concentration of Ca^{2+} exerts different percentage of inhibition at different Mg^{2+} to ATP molar ratios due to mass action on the complexation equilibrium between Ca-ATP and Mg-ATP complexes. If the Mg^{2+} to ATP molar ratio is set to be at the optimum value or below, PP_i6 suppresses the enzyme activity; otherwise PP_i6 enhances the enzyme activity. Different reactor types were simulated and evaluated based on the parameters including conversion, selectivity and specific productivity. STR offers the best performance in all these parameters, while CSTR shows the least performance. The STR is therefore the most suitable reactor type for the biocatalytic synthesis of L-GAP, whereby the biocatalyst consumption could be significantly reduced from 19.7 to 1.7 g kg^{-1} by means of reaction engineering. If a continuously operated reactor is aimed at a plug flow reactor as the reaction engineering equivalent STR has to be chosen.

The authors thank the BMBF (German Federal Ministry of Education and Research) cluster "Biocatalysis 2021" (project number: 0316055) for the financial support.

The authors declare no financial or commercial conflict of interests.

Getachew S. Molla contributed in planning and performing the experiments, analyzing the data and writing the article. Prof. Dr. Andreas Liese and Dr. Roland Wohlgemuth contributed in analyzing the data as well as writing the article. Prof. Dr. Wolfgang Streit, Dr. Jennifer Chow and Birhanu M. Kinfu contributed in results discussion.

5 References

[1] Buchholz, K., Kasche, V., Bornscheuer, U. T., *Biocatalysts and Enzyme Technology*, Wiley-VCH, Weinheim 2005, pp. 448.
[2] Faber, K., *Biotransformations In Organic Chemistry: A Textbook*, 6th Edn., Springer-Verlag, Berlin, Heidelberg 2011, pp. 423.
[3] Liese, A., Seelbach, K., Wandrey, C., *Industrial Biotransformations*. Wiley-VCH, Weinheim, New York 2000, pp 423.
[4] Wohlgemuth, R., Tools and ingredients for the biocatalytic synthesis of metabolites. *Biotechnol. J.* 2009, *4*, 1253–1265.
[5] Gauss, D., Schoenenberger, B., Molla, G. S., Kinfu B. M. et al., Biocatalytic Phosphorylation of Metabolites, in: Liese, A., Hilterhaus, L., Kettling, U., Antranikian, G. (Eds.), *Applied Biocatalysis – From Fundamental Science to Industrial Applications* , Wiley-VCH, Weinheim 2016, pp. 147–177.
[6] Bednarski, M. D., Simon, E. S., Bischofberger, N., Fessner, W. D. et al., Rabbit muscle aldolase as a catalyst in organic synthesis. *J. Am. Chem. Soc.* 1989, *111*, 627–635.
[7] Fessner, W.-D., Sinerius, G., Synthesis of dihydroxyacetone phosphate (and isosteric analogues) by enzymatic oxidation; Sugars from glycerol. *Angew. Chem. Int. Ed.*. 1994, *33*, 209–212.
[8] Molla, G. S., Wohlgemuth, R., Liese, A. One-pot enzymatic reaction sequence for the syntheses of D-glyceraldehyde 3-phosphate and L-glycerol 3-phosphate. *J. Mol. Catal. B: Enzymatic* 2016, *124*, 77–82.
[9] Sánchez-Moreno, I., Hélaine, V., Poupard, N., Charmantray, F. et al., One-pot cascade reactions using fructose-6-phosphate aldolase: Efficient synthesis of D-arabinose 5-phosphate, D-fructose 6-phosphate and analogues. *Adv. Synth. Catal.* 2012, *354*, 1725–1730.
[10] Schoevaart, R., van Rantwijk, F., Sheldon, R. A., A four-step enzymatic cascade for the one-pot synthesis of non-natural carbohydrates from glycerol. *J. Org. Chem.* 2000, *65*, 6940–6943.
[11] Tang, C. T., Engel, R., Tropp, B. E., L-Glyceraldehyde 3-phosphate, a bactericidal agent. *Antimicrob Agents Chemother* 1977, *11*, 147–153.
[12] Desai, K. K., Miller, B. G., A metabolic bypass of the triosephosphate isomerase reaction. *Biochemistry* 2008, *47*, 7983–7985.
[13] Adelman, R. C., Brox, L., Krulwich, T. A., Preparation of radioactive L-glyceraldehyde 3-phosphate. *Anal. Biochem.* 1969, *32*, 258–262.
[14] Bublitz, C., Kennedy, E. P., Synthesis of phosphatides in isolated mitochondria. III. The enzymatic phosphorylation of glycerol. *J. Biol. Chem.* 1954, *211*, 951–961.
[15] Gauss, D., Schönenberger, B., Wohlgemuth, R., Chemical and enzymatic methodologies for the synthesis of enantiomerically pure glyceraldehyde 3-phosphates. *Carbohydr. Res.* 2014, *389*, 18–24.
[16] Hayashi, S. I., Lin, E. C., Purification and properties of glycerol kinase from *Escherichia coli*. *J. Biol. Chem.* 1967, *242*, 1030–1035.
[17] Janson, C. A., Cleland, W. W., The kinetic mechanism of glycerokinase. *J. Biol. Chem.* 1974, *249*, 2562–2566.
[18] Bonsignore, A., Leoncini, G., Siri, A., Ricci, D., Kinetic behaviour of glyceraldehyde 3-phosphate conversion into methylglyoxal. *Ital. J. Biochem.* 1973, *22*, 131–140.
[19] Pompliano, D. L., Peyman, A., Knowles, J. R., Stabilization of a reaction intermediate as a catalytic device: Definition of the functional role of the flexible loop in triosephosphate isomerase. *Biochemistry* 1990, *29*, 3186–3194.
[20] Richard, J. P., Mechanism for the formation of methylglyoxal from triose-phosphates. *Biochem. Soc. Trans.* 1993, *21*, 549–553.
[21] Richard, J. P., Restoring a metabolic pathway. *ACS Chem. Biol.* 2008, *3*, 605–607.
[22] Biselli, M., Kragl, U., Wandrey, C., Reaction engineering for enzyme-catalyzed biotransformations, in: Drauz, K., Waldmann, H. (Eds.), *Enzyme Catalysis in Organic Synthesis*, Wiley-VCH Verlag GmbH, Weinheim 2002, pp. 185–257.
[23] Brass, J., Hoeks, F., Rohner, M., Application of modelling techniques for the improvement of industrial bioprocesses. *J. Biotechnol.* 1997, *59*, 63–72.
[24] Vasic-Racki, D., Bongs, J., Schörken, U., Sprenger, G. A. et al., Modeling of reaction kinetics for reactor selection in the case of L-erythrulose synthesis. *Bioprocess Biosyst. Eng.* 2003, *25*, 285–290.
[25] Glonek, T., ^{31}P-NMR of Mg-ATP in dilute solutions: Complexation and exchange. *Int. J. Biochem.* 1992, *24*, 1533–1559.
[26] Wong, C. H., Whitesides, G. M., Synthesis of sugars by aldolase-catalyzed condensation reactions. *J. Org. Chem.* 1983, *48*, 3199–3205.
[27] Flamholz, A., Noor, E., Bar-Even, A., Milo, R., eQuilibrator – the biochemical thermodynamics calculator. *Nucleic Acids Res.* 2012, *40 (Database issue)*, D770–D775.

Biotechnology Journal

Supporting Information for DOI 10.1002/biot.201600625

Bioreaction engineering leading to efficient synthesis of L-glyceraldehyde-3-phosphate

Getachew S. Molla, Birhanu M. Kinfu, Jennifer Chow, Wolfgang Streit, Roland Wohlgemuth, Andreas Liese

Table 1: Activities and half-lives of glycerol kinase from *E.* coli, *Cellulomonas* sp. and *Streptomyces canus*. Incubation conditions are 100 mM ATP 100 mM Mg^{2+} in 100 mM Tris-HCl buffer, pH 8.5 and 25°C. Reaction conditions for the activity assays are 50 mM ATP, 50 mM glycerol, 100 mM Mg^{2+}, in 100 mM Tris-HCl buffer, pH 8.5, and 25°C.

Glycerol kinase source	Activity (U/mg)	Half-life (days)
E. coli	15.0	3.2
***Cellulomonas* sp.**	24.0	No significant loss of the enzyme activity was detected in a month of incubation period
Streptomyces canus	18.5	21.0

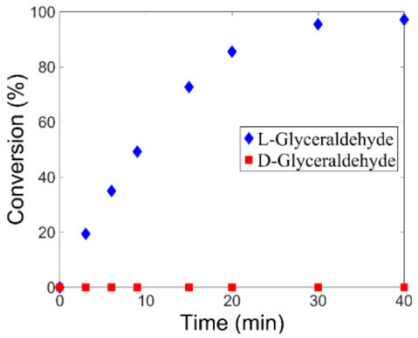


Figure 1: Phosphorylation of D- and L-glyceraldehyde catalyzed by glycerol kinase from *Cellulomonas* sp.; reaction conditions: 50 mM ATP, 30 mM D- and L-glyceraldehyde, 35 mM Mg^{2+} and 0.02 mg/ml glycerol kinase from *Cellulomonas* sp. in 100 mM TEA buffer, pH 8, and 25°C

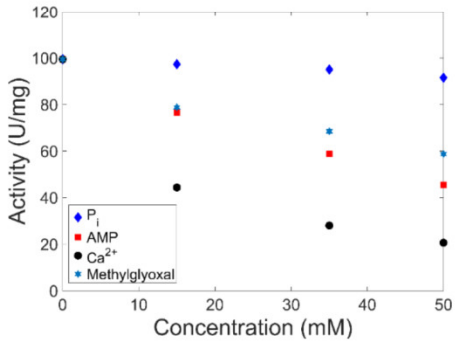


Figure 2: Inhibition of glycerol kinase from *Cellulomonas* sp. by inorganic orthophosphate (P_i), adenosine monophosphate (AMP), calcium ion (Ca^{2+}), methylglyoxal; reaction conditions: 50 mM ATP, 50 mM L-glyceraldehyde, 35 mM Mg^{2+}, 0.08 mg/ml glycerol kinase from *Cellulomonas* sp. in 100 mM TEA buffer, pH 8, and 30°C

5 Recombinant RNA Polymerase from *Geobacillus* sp. GHH01 as tool for rapid generation of metagenomic RNAs using *in vitro* technologies

Birhanu M. Kinfu[1], Maike Köster[1], Mareike Janus[1], Volkan Besirlioglu[2], Michael Roggenbuck[3], Richard Meurer[2], Ljubica Vojcic[2] Martin Borchert[3], Ulrich Schwaneberg[2], Jennifer Chow[1] and Wolfgang R. Streit[1]

[1]Microbiology and Biotechnology, Biocenter Klein Flottbek, University of Hamburg, Ohnhorststr. 18, D-22609 Hamburg, Germany; [2]Institute of Biotechnology, RWTH Aachen, Worringerweg 3, D-52074 Aachen, Germany; [3]Novozymes A/S, Krogshøjvej 36, 2880 Bagsværd, Denmark

Published in:

Biotechnology and Bioengineering, 114(12): 2739-2752. doi:10.1002/bit.26436

Received: 12 June 2017 | Revised: 9 August 2017 | Accepted: 21 August 2017
DOI: 10.1002/bit.26436

ARTICLE

WILEY BIOTECHNOLOGY & BIOENGINEERING

Recombinant RNA Polymerase from *Geobacillus* sp. GHH01 as tool for rapid generation of metagenomic RNAs using in vitro technologies

Birhanu M. Kinfu[1] | Maike Jahnke[1] | Mareike Janus[1] | Volkan Besirlioglu[2] | Michael Roggenbuck[3] | Richard Meurer[2] | Ljubica Vojcic[2] | Martin Borchert[3] | Ulrich Schwaneberg[2] | Jennifer Chow[1] | Wolfgang R. Streit[1]

[1] Microbiology and Biotechnology, Biocenter Klein Flottbek, University of Hamburg, Ohnhorststr, Hamburg, Germany

[2] Institute of Biotechnology, RWTH Aachen, Aachen, Germany

[3] Novozymes A/S, Bagsværd, Denmark

Correspondence
Wolfgang R. Streit, Department of Microbiology and Biotechnology, Biocenter Klein Flottbek, University of Hamburg, Ohnhorststr 18, D-22609 Hamburg, Germany.
Email: wolfgang.streit@uni-hamburg.de

Funding information
Deutscher Akademischer Austauschdienst, Grant number: DAAD (57076385); Danish Innovationsfonden + BMBF, Grant number: MetaCat (ERA-IB-14-030031A565A); Bundesministerium für Bildung und Forschung (BMBF), Grant number: iVDT-V^2 (031A571A)

Abstract

The exciting promises of functional metagenomics for the efficient discovery of novel biomolecules from nature are often hindered by factors associated with expression hosts. Aiming to shift functional metagenomics to a host independent innovative system, we here report on the cloning, heterologous expression, and reconstitution of an RNA polymerase (RNAP) from the thermophilic *Geobacillus* sp. GHH01 and in vitro transcription thereafter. The five genes coding for RNAP subunits, a house keeping sigma factor and two transcription elongation factors were cloned and over expressed as His_6-tagged and/ or tag-free proteins. Purified subunits were reconstituted into a functional polymerase through either the classical method of denaturation and subsequent renaturation or through a new resource and time efficient thermo-reconstitution method which takes advantage of the subunits' temperature stability. Additionally, all subunits were cloned into a single vector system for a co-expression and in vivo reconstitution to the RNAP core enzyme. Both the core and holoenzyme form of the RNAP exhibited a robust transcription activity and were stable up to a temperature of 55°C close to their fullest activity. The *Geobacillus* RNAP showed a remarkable in vitro transcription profile recognizing DNA template sequences of diverse bacteria and archaea as well as metagenomic samples. Coupled with a subsequent in vitro translation step, this recombinant transcription system could allow a new, clone-free, and functional metagenomic screening approach.

KEYWORDS
cell-free metagenomics, *Geobacillus* sp. GHH01, in vitro metagenomics, in vitro transcription, metagenomics, recombinant RNA polymerase

1 | INTRODUCTION

Since the beginning of the metagenome era, numerous researchers have focused on a function- or sequence-based screening for novel enzymes or other valuable biomolecules (e.g., Culligan, Sleator, Marchesi, & Hill, 2014; DeCastro, Rodriguez-Belmonte, & Gonzalez-Siso, 2016; Ferrer et al., 2016; Iqbal, Feng, & Brady, 2012; Streit, Daniel, & Jaeger, 2004; Uchiyama & Miyazaki, 2009). Even though many biocatalysts have been identified using function-based approaches, the construction of metagenomic libraries, functional screening and the recovery of positive clones that

carry the desired gene sequences on an insert can be a cumbersome and time-consuming procedure. Function-based metagenomics is often hampered by many problems starting with the extraction of environmental DNA in adequate amounts, purity, and integrity. Once the library is established, the outcome of the functional screening approach itself can be limited due to differences in codon usage, poor protein expression, incorrect folding, or miss-folding and secretion of heterologous proteins (Ekkers, Cretoiu, Kielak, & Elsas, 2012; Kudla, Murray, Tollervey, & Plotkin, 2009; Lam, Cheng, Engel, Neufeld, & Charles, 2015; Steele, Jaeger, Daniel, & Streit, 2009; Uchiyama & Miyazaki, 2009). Despite these well-known difficulties, a function-based search can be superior to sequence-based metagenomics as it offers the advantage of finding truly novel enzymes rather than variants of already known enzymes with sequences related to each other. Furthermore, it ensures the identification of full-length genes and functional proteins and gives a firsthand estimate on the enzyme activities even during the screening procedure. Thus, there is a large interest for the use, development, and improvement of function-based metagenome screening technologies (Leis, Angelov, & Liebl, 2013; Liebl et al., 2014; Tuffin, Anderson, Heath, & Cowan, 2009).

Since *Escherichia coli* is the host that is mostly used for the construction of metagenome libraries, several studies have been published on addressing expression problems observed in this host by constructing improved vector systems (e.g., Liebl et al., 2014; Troeschel, Drepper, Leggewie, Streit, & Jaeger, 2010). Furthermore, other host strains have been successfully developed to overcome some of the problems associated with the poor folding and secretion of foreign proteins in *E. coli*. Alternative hosts used in function-based metagenomics are for example *Pseudomonas putida*, *Streptomyces lividans*, *Sinorhizobium meliloti*, *Rhizobium leguminosarum*, *Desulfovibiro* sp., *Streptomyces* sp., *Bacillus*, and *Saccharomyces* (Cheng, Pinnell, Engel, Neufeld, & Charles, 2014; Courtois et al., 2003; Li, Wexler, Richardson, Bond, & Johnston, 2005; Martinez et al., 2004; Rousset et al., 1998; Schallmey et al., 2011; Wang et al., 2006; Wexler, Bond, Richardson, & Johnston, 2005). Other mesophilic and Gram-negative hosts that have been shown to accept metagenome DNA include *Salmonella enterica*, *Serratia marcescens*, *Vibrio vulnificus*, and *Enterobacter nimipressuralis* (Kakirde et al., 2011). Moreover, *Ralstonia metallidurans* (Maroti et al., 2009) was employed and only recently *Thermus thermophilus* was established as an alternative host to gain access to proteins derived from thermophiles (Leis et al., 2015; Liebl et al., 2014).

While these are all recent and successful examples for the development of alternative metagenomic hosts, none of these strains will finally be a solution to overcome heterologous expression problems per se. In general, it is assumed that *E. coli* expresses about 40 % of the foreign genes into proteins which can be recovered by function-based searches. This assumption is based on calculations taking the size of the insert and copy number into account (Gabor, Alkema, & Janssen, 2004). Experimental data transcribing genes derived from *Pseudomonas aeruginosa* and *Haemophilus influenza* in *E. coli* further suggest that its σ-factors are capable to recognize about 50% of foreign promoters (Warren et al., 2008) with rather low phylogenetic distance. Recent studies indicate that one possibility to partly overcome this transcriptional limitation may be by expressing additional σ-factors from different species in *E. coli* (Gaida et al., 2015; Liebl et al., 2014).

In this work, we suggest for the first time the idea of cell-free metagenomics which will bypass complex cell regulation and toxicity-related problems in cloning and expression hosts. This method can easily be customized toward desired applications. Cell-free expression systems in general, employing either whole cell extracts or recombinant components, have been reported for various applications (Hiroyuki et al., 2010; Körfer, Pitzler, Vojcic, Martinez, & Schwaneberg, 2016; Pardee et al., 2016; Shimizu et al., 2001; Tuckey, Asahara, Zhou, & Chong, 2014; Zemella, Thoring, Hoffmeister, & Kubick, 2015). The first and critical step in this process is the transcription in which the RNA polymerase recognizes specific DNA sequences (i.e., promotors). The required promotor sequences (often P_tac, P-lac, or P_T7 are usually introduced during cloning using the promotors supplied with the cloning vectors.

Cell-free expression systems based on bacteriophage- and mainly T7-RNAPs are highly selective for DNA templates containing viral promoter sequences. This extreme specificity limits the systems applicability in transcribing template DNAs with other promoter regions and recognition factors. On the other hand, crude cell extract-based expression systems developed from bacterial strains have common drawbacks as they contain traces of undesirable cellular components resulting in unwanted background activities during screenings as well as degradation of template DNA, generated transcript, or synthesized protein.

As a first step toward overcoming problems associated with low transcription rates of metagenomic DNA, we have developed a heat stable and robust in vitro transcription system for the rapid generation of metagenomic RNAs from diverse DNA samples. For this, we have employed a recombinant RNAP from the recently isolated thermophilic *Geobacillus* sp. strain GHH01 (Wiegand et al., 2013). The recombinant enzyme appears to be highly efficient and superior compared to the tested RNAPs with respect to the amount and quality of the generated metagenome-derived RNAs which are suitable for further use in in vitro translation assays.

2 | MATERIAL AND METHODS

2.1 | Bacterial strains, primers, and plasmids

Bacterial strains, oligonucleotides used for amplification with respective flanking restriction sites, plasmids employed for cloning, and constructs are listed in Tables 1–3 and Figure 1a. Primers were synthesized by Eurofins MWG Operon (Ebersberg, Germany). If not mentioned otherwise, *E. coli* clones were grown at 37°C in LB medium (1 % tryptone/peptone, 0.5% yeast extract, 0.5 % NaCl) supplemented with appropriate antibiotics. *Geobacillus* sp. strain GHH01 was grown at 55°C in Standard I medium (0.78% peptone from meat, 0.78% peptone from casein, 0.28% yeast extract, 0.56% NaCl, 0.1% glucose).

TABLE 1 Bacterial strains used in this study

Bacterial strains	Traits	Reference/ source
Geobacillus sp. GHH01	Wild type isolate	Wiegand et al. (2013)
E. coli DH5α	F-φ80 *lacZ*ΔM15 Δ(*lacZYA-argF*) U169 *recA*1 *endA*1 *hsdR*17(rk⁻, mk⁺) *phoA supE*44 λ⁻thi⁻1 *gyrA*96 *relA*1	Life Technologies (Frankfurt, Germany)
E. coli BL21 (DE3)	F⁻*omp ThsdS*$_B$(r$_B$⁻m$_B$⁻) *gal dcm* (DE3)	Novagen/ Merck (Darmstadt, Germany)
E. coli Rosetta-gami 2 (DE3)	Δ(*ara–leu*)7697 Δ*lacX*74 Δ*phoA PvuII phoR araD*139 *ahpC galE galK rpsL* (DE3) F'[*lac*⁺*lacI*^{q} *pro*] *gor*522::Tn10 *trxB* pRARE2[3] (CamR, StrR, TetR)[4]	Novagen/ Merck (Darmstadt, Germany)

2.2 | Molecular cloning strategies and expression of RNAP subunits

We followed two parallel expression strategies in order to obtain active RNAP. First, we cloned and expressed each subunit individually with or without poly-histidine (His$_6$)-tag. Second, we constructed an expression vector carrying all five RNAP genes together to allow simultaneous expression and in vivo assembly (Figures 1b and 1c and Table 3). General overview of the working strategy is outlined in Figure S1.

The RNAP core enzyme (RNAP$_C$) subunits α, β, β′, δ, and ω coding genes *rpoA*, *rpoB*, *rpoC*, *rpoE*, and *rpoZ*, respectively, and the house-keeping σ^{70}-factor coding gene *rpoD* were all individually amplified from the genome of *Geobacillus* sp. GHH01 using high fidelity phusion polymerase (*ThermoFisher Scientific GmbH*, Schwerte, Germany). Primer pairs carrying restriction sites compatible for cloning into pET21a(+) and pET28a(+) (*Novagen* via *Merck Millipore*, Darmstadt, Germany) for untagged and N-terminally His$_6$-tagged proteins,

TABLE 2 Primers used and their respective melting temperatures [Tm (°C)]

Primer name	Sequence (5′–3′)	Tm (°C)
M13 forward	TTGTAAAACGACGGCCAGTG	57.3
M13 reverse	GGAAACAGCTATGACCATGA	55.3
1492R	TACGGYTACCTTGTTACGAC	56.3
616V	AGAGTTTGATYMTGGCTCAG	55.3
T7 Prom	TAATACGACTCACTATAGGG	53.2
T7 Term	GCTAGTTATTGCTCAGCGG	56.7
rpoAGH-NdeI_F	TCGCATATGATTGAAATTGAAAAGCCGAAAATTG	67.0
rpoAGH-EcoRI_R	TCTGAATTCTTAGTCGTCTTTGCGCAGGCTGAG	72.0
rpoBGH-BamHI_F	GCAGGATCCATGACAGGCCGACTAGTTCAA	69.5
rpoBGH-XhoI_R	AGCGCTCGAGTTATTCTTTCGTCACCGCATC	69.5
rpoCGH-BamHI_F	GTCGGATCCATGCTGGATGTCAATAAATTTG	65.5
rpoCGH-XhoI_R	ACTGCTCGAGTTATTTGGAAGACACCGTGTC	68.2
rpoEGH-NdeI_F	CATATGAGCCTGCAGCAGCAATACTCG	66.5
rpoEGH-HindIII_R	AAGCTTTTATTCGTCTTCGTCCGGCTCTTCG	68.2
rpoZGH-NdeI_F	CATATGACGATGCTGTATCCTTCCATCGATTT	65.6
rpoZGH-EcoRI_R	GAATTCTTATTTTTCCTCCTCCACTAATTC	61.3
rpoDGH-NheI_F	GCTAGCATGGCTGAAAAACCAGCCCAATCAAAG	69.5
rpoDGH-HindIII_R	AAGCTTTCACTCTAAAAAGTCTTTCAGCCGTTTG	65.9
GreAGH-BamHI_F	AAGGGATCCATGGCGAACGAAAAGC	64.6
GreAGH-XhoI_R	GGGCTCGAGTTATTTCACAGCCACGATT	66.6
MfdGH-BamHI_F	AGTGCGGATCCATGCTTTCGTTG	62.4
MfdGH-XhoI_R	CGTCTCGAGTTATGCCGTCACC	64.0
eGFP PCRtemp_rv	CATCATGCCGTTTGTGATGGCTTCC	62.7
eGFP_rv	ATGCGCCCAATACGCAAACCGCCTC	69.0
eGFP_fw	GGCCGCTCTAGAACTAGTGGATCCC	65.4
3_2c_12_−212for	ATTTCGCCGAAGCGACTGTGAC	62.3
3_2c_12_ +51rev	TCATAGGGCCGGACAATCCTAC	60.5

Restriction sites are underlined.

TABLE 3 Plasmids and constructs employed in this study

Plasmids and constructs	Features	Reference/source
pBluescript II SK (+)	3.0 kb phagemid vector, *lacZ*, *bla*, P_{T7}, P_{T3}	Stratagene, La Jolla (CA)
pDrive	3.85 kb TA-cloning vector, *oriEc*, Plac*lacZ*, AmpR, KanR, T7-promotor	QIAGEN (Hilden, Germany)
pET-21a(+)	5.44 kb Expression vector, *lacI*, AmpR, T7-promotor, C- His_6-tag coding sequence	Novagen/Merck (Darmstadt, Germany)
pET-28a(+)	5.37 kb Expression vector, *lacI*, KanR, T7-promotor, N- and C-terminal His_6-tag coding sequence	Novagen/Merck (Darmstadt, Germany)
pBBR1MCS-5	4.77 kb broad host range expression vector, *rep*, *mob*, *lacZ*, GmR	Kovach et al. (1995)
pET-28a(+)::*rpoA*	N-His_6-tagged α-subunit of GHH01 RNAP	This work
pET-28a(+)::*rpoB*	N-His_6-tagged β-subunit of GHH01 RNAP	This work
pET-28a(+)::*rpoC*	N-His_6-tagged β'-subunit of GHH01 RNAP	This work
pET-28a(+)::*rpoE*	N-His_6-tagged δ-subunit of GHH01 RNAP	This work
pET-28a(+)::*rpoZ*	N-His_6-tagged ω-subunit of GHH01 RNAP	This work
pET-28a(+)::*rpoD*	N-His_6-tagged σ^{70}-factor of GHH01 RNAP	This work
pET-21a(+)::*rpoA*	α-subunit of GHH01 RNAP	This work
pET-21a(+)::*rpoB*	β-subunit of GHH01 RNAP	This work
pET-21a(+)::*rpoC*	β'-subunit of GHH01 RNAP	This work
pET-21a(+)::*rpoE*	δ-subunit of GHH01 RNAP	This work
pET-21a(+)::*rpoZ*	ω-subunit of GHH01 RNAP	This work
pET-21a(+)::*rpoD*	σ^{70}-factor of GHH01 RNAP	This work
pET-21a(+):: *rpoCBZAE*	β'βωαδ-subunits, His_6-tag on β'-subunit	This work
pET-28a(+)::*greA*	N-His_6-tagged transcription elongation factor GreA	This work
pET-28a::*MfD*	N-His_6-tagged transcription elongation factor MfD	This work
peGFP cloning vector	3.35 kb, AmpR, *lac*, Expresses eGFP tag	CLONTECH Laboratories, Palo Alto (CA 94303-4230), GenBank: U76561.1
pBBR1MCS-5:: *eGFP*	*eGFP* under control of *lac*-promotor	This work
pCC1FOS 108	Fosmid harboring metagenomic DNA including *Lip41*	Rabausch et al. (2013)

respectively were used. Amplified gene products with intact stop codon were either subcloned from pDrive clones or directly cloned into the pET expression vectors. All constructs were verified by automated DNA sequencing. For co-expression of all five subunits, *rpoA*, *B*, *C*, *E*, and *Z* were subcloned from either pET21a(+) or pET28a(+)-single gene constructs into pET21a(+) through a series of restriction and ligation processes. *RpoC* is fused with His_6-tag coding sequence for purification of the core enzyme using ion exchange chromatography (Figure 1c).

Plasmid constructs were heat-shock transformed into chemically competent *E. coli* BL21(DE3) or Rosetta gami 2 (DE3) host cells. Expression was then induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) at an approximate OD_{600} of 0.6. Detailed expression conditions (such as *E. coli host* strain, expression temperature, and incubation time) can be found in Table S1. After expression, cells were harvested using a Falcon centrifuge 5804R (rotor A-4-44, Eppendorf, Hamburg, Germany) for 20 min at 4,500 × g and 4°C or for larger volumes using a Sorvall RC6+ centrifuge (rotor F10S–6 × 500y, Thermo scientific, Braunschweig, Germany) for 20 min at 5,000 rpm and 4°C. Cell pellets were either directly processed or stored at −20°C until protein purification.

2.3 | Purification of single subunits and transcription factors

All His_6-tagged subunits were expressed in soluble form and purified under native conditions. For this, the respective *E. coli* cell pellet was resuspended in 5 ml lysis buffer (50 mM $NaH_2PO_4 \cdot 2H_2O$ pH 8.0, 300 mM NaCl, 20 mM imidazole, 0.1 mM phenylmethylsulfonyl fluoride [PMSF], 1 mM dithiothreitol [DTT]) per 250 ml *E. coli* expression culture. Cells were lysed using French pressure cell press at 1,300 psi and incubated afterwards with 20 units DNase for 30 min at 37°C supplemented with $MgCl_2$ to a final concentration of 2 mM.

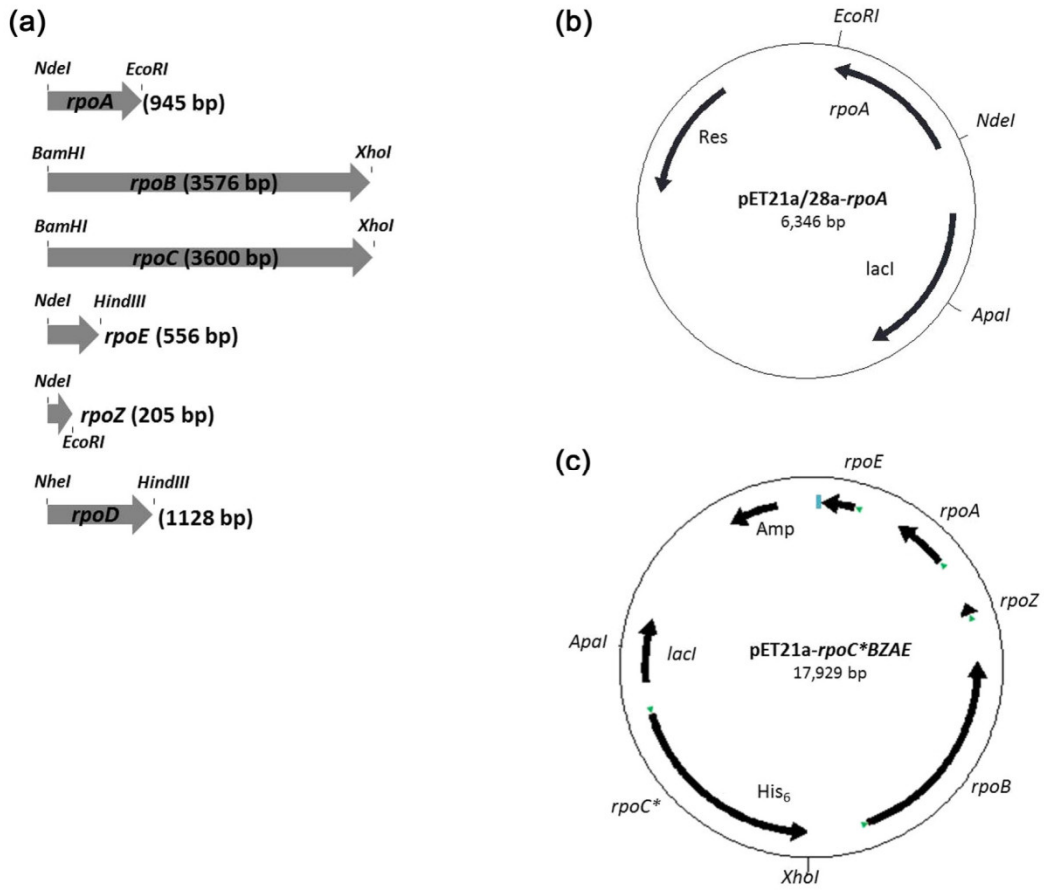


FIGURE 1 Plasmid constructs for expression of the *rpo* genes. (a) *Rpo* genes with intact stop codons were amplified from genomic DNA of *Geobacillus* sp. GHH01. Sizes of genes and restriction flanks used for cloning are indicated. (b) *Rpo* genes were cloned into pET21a(+) and pET28(+) vectors for expression. Here, a mixed vector map of both is shown exemplarily for *rpoA*. Vector pET21a(+) contains the bla gene for ampicillin resistance (Res) and pET28a(+) carries the Kan coding sequence for kanamycin selection. Furthermore, subunit proteins expressed in pET28a(+) carry a N-terminal His_6-tag. (c) For co-expression of $RNAP_C$ subunits, all five genes coding for the subunits were cloned into pET21a(+) each under control of its own T7-promoter (small arrowheads) and labeled as pET21a-*rpoC*BZAE*. The T7-terminator is located after *rpoE* (blue box). *RpoC* is expressed with an N-terminal His_6-tag (*)

The mixture was centrifuged at 16,100 × *g* and 4°C for at least 1 hr. The clear cell lysate was then loaded onto a Ni-chelating NTA-column (HisTrap columns, *GE Healthcare*, Freiburg, Germany) preequilibrated with lysis buffer and purified using a fast protein liquid chromatography (FPLC) system. Bound protein was washed with 5 column volumes (CV) of wash buffer (50 mM $NaH_2PO_4 \cdot 2H_2O$ pH 8.0, 500 mM NaCl, 40 mM imidazole, 1% glycerol, 0.1 mM PMSF, 1 mM DTT) and finally eluted using elution buffer (50 mM $NaH_2PO_4 \cdot 2H_2O$ pH 8.0, 300 mM NaCl, 250 mM to 500 mM imidazole, 0.1 mM PMSF, 1 mM DTT). All buffers used for the FPLC system were filtered, degassed and cooled prior to use. Afterwards, buffer was exchanged to 2x storage buffer (80 mM Tris-HCl pH 7.9, 0.4 M KCl, 2 mM EDTA, 2 mM DTT) by at least three rounds of fivefold volume dilution and concentration in a VIVASPIN® 6 concentrator with molecular weight cut-off (MWCO) suitable to each protein size. Accordingly, MWCO of 5,000 was used for concentration of ω- and δ-subunits, 10,000 for α-subunit, 30,000 for σ^{70}-factor, and 50,000 for β- and β′-subunits (see Figures 2a and 2b for approximate molecular weight of each subunit). The protein solution was adjusted with 1x storage buffer to a desired concentration and supplemented with glycerol to a final concentration of 50% (v/v) glycerol and finally stored in aliquots at −20°C. The success of purification was monitored by doing SDS–PAGE analysis.

Untagged RNAP subunits were crudely purified via two approaches depending on solubility of the proteins. Soluble subunits (δ and ω) were separated from the majority of *E. coli* proteins through heat-precipitation. For this, the cell pellet was resuspended in 5 ml lysis buffer no. 2 (40 mM Tris-HCl pH 7.9, 0.3 M KCl, 10 mM EDTA) per 250 ml *E. coli* liquid culture supplemented with 0.1 mM PMSF and 1 mM DTT. Cells were lysed with ultrasonication on ice for 1 min (duty cycle: 0.5, amplitude: 50%) alternating with a 1 min break for five cycles. The lysate was cleared by at least 1 hr of centrifugation at 16,100 × *g* and 4°C. The supernatant was then incubated at 80°C for 35 min for δ-subunit and 70°C for 12 min for ω-subunit purification. Precipitated proteins were removed by another round of centrifugation (16,100 × *g*, for at least 1 hr at 4°C). Buffer

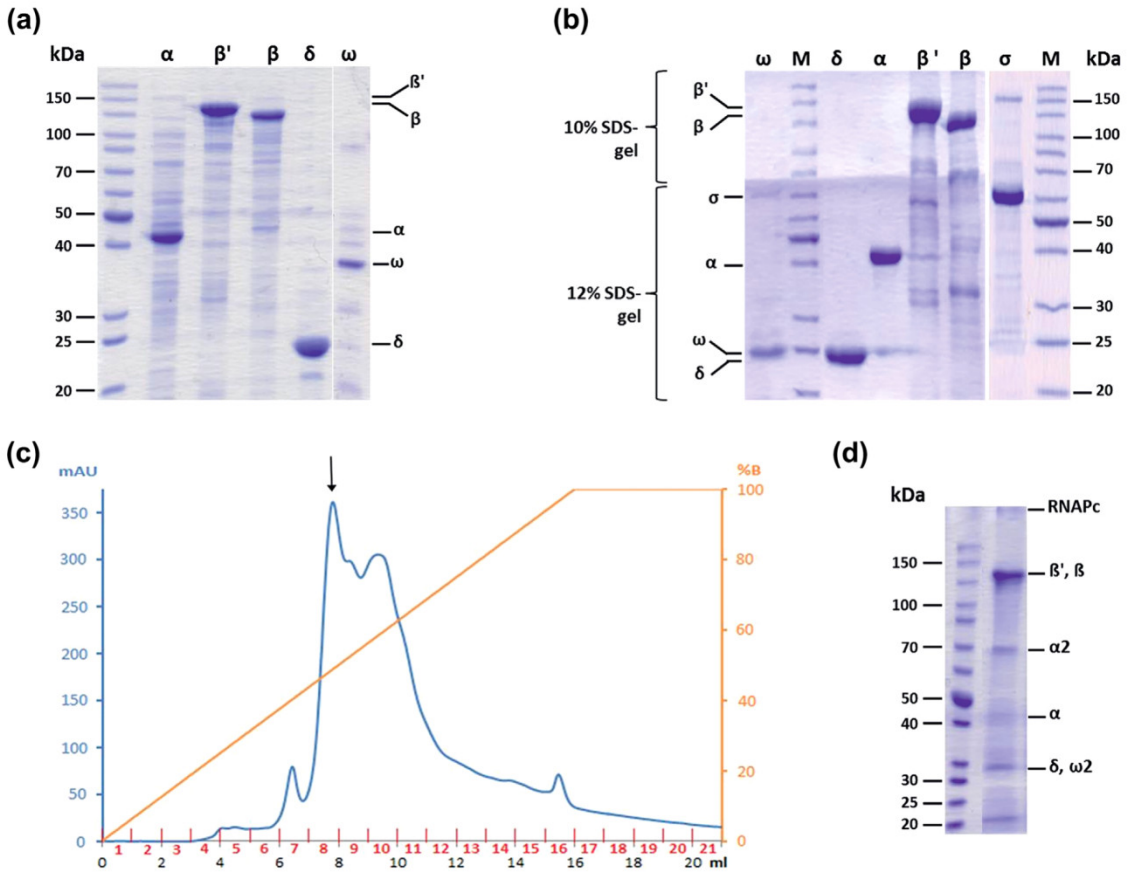


FIGURE 2 Purification of single subunits and co-expressed $RNAP_C$. (a) Untagged subunits were crudely purified and separated in a 15% (right, for ω-subunit) or 12% (left, for the rest of the subunits) SDS-polyacrylamide gel, respectively. Subunits α, β and β′ were solubilized from inclusion bodies with 6 M urea. Subunits δ and ω were purified via heat-treatment. (b) His_6-tagged subunits and a σ^{70}-factor were purified chromatographically using HisTrap column. Samples were separated in a two step-gradient gel of 12% (bottom) and 10% (top). (c) In vivo assembled $RNAP_C$ and non-assembled His_6-β′ were separated by anionic exchange chromatography. Elution was performed with a linear salt gradient from 0.1 M NaCl (0% B) to 1 M NaCl (100% B). The $RNAP_C$-containing peak is indicated with an arrow. (d) Denaturing SDS–PAGE analysis of pure co-expressed $RNAP_C$. Individual subunits and still intact $RNAP_C$ are indicated. About 4 µg total protein was applied per lane

exchange, protein concentration, analysis and storage were done as described above for the His_6-tagged subunits. Subunits found in inclusion bodies (α, β, and β′) were prepared for in vitro reconstitution by washing them several times with 0.2% Na-deoxycholate containing lysis buffer no. 2 following the protocol of Kuzndedelov and Severinov (2009) and were stored in aliquots at −20°C until further use. Protein concentration was quantified using Bradford assay with BSA as a standard (Bradford, Pohle, Gunther, & Mehlenbacher, 1942).

Transcription elongation factors *GreA* and *MfD* from *Geobacillus* sp. GHH01 were cloned into pET28a(+), expressed in *E. coli* BL21 (DE3) and purified the same way as His_6-tagged RNAP subunits.

2.4 | In vitro reconstitution of recombinant RNAP

Classical reconstitution of core enzyme ($RNAP_C$), subunit composition α2ββ′ωδ, was accomplished by mixing all subunits under denaturing conditions and subsequently removing the denaturant through gradient dialysis. For this, we followed a slightly modified protocol based on the method described by Kuznedelov and Severinov (2009). Two parallel assembly reactions, one for exclusively His_6-tagged and the other for the untagged subunits, were performed to evaluate the effect of His_6-tag on the overall transcription process. Untagged subunits found in inclusion bodies were first solubilized with 6 M urea. For this, Na-deoxycholate-washed pellets (see above) were resuspended in 200 µl denaturation buffer (50 mM Tris-HCl pH 7.9, 6 M urea, 10 mM $MgCl_2$, 10 µM $ZnCl_2$, 10% glycerol, 1 mM EDTA, 10 mM DTT) and incubated on ice for at least 30 min vortexing every 10 min. Afterwards, the soluble fractions were collected by centrifugation (30 min, 16,100 × g, 4°C) and used for in vitro reconstitution.

All five subunits were mixed in denaturation buffer in a defined molar ratio of 2:8:4:2:2 (α:β:β′:δ:ω) for both untagged and His_6-tagged subunits. The total protein concentration was adjusted to 0.5 mg/ml by

adding denaturation buffer. The reconstitution mix was then dialyzed stepwise at 4°C against up to 250 volumes of reconstitution buffer (50 mM Tris-HCl pH 7.9, 200 mM KCl, 10 mM $MgCl_2$, 10 µM $ZnCl_2$, 10% glycerol, 1 mM EDTA, 1 mM DTT) reducing the urea concentration from 6 M to 3 M to 0 M with every buffer exchange. After dialysis, potential protein precipitates were removed by centrifugation (10 min, 16,100 × *g*, 4°C).

To further improve feasibility and circumvent the time-consuming process of the classical reconstitution method, we developed a new approach for in vitro reconstitution of $RNAP_C$ which takes advantage of the heat-stability of *Geobacillus* proteins. In order to determine a maximum nondenaturant temperature to all subunits, heat stability profiles of each soluble subunit proteins were established by incubating at different temperatures and analyzing residual soluble protein using SDS–PAGE analysis (Table 4). Stability was further verified using differential scanning fluorimetry (nanoDSF) technology (Prometheus NT.48, *NanoTemper Technologies*, Munich, Germany). In this thermo-reconstitution approach, purified subunit proteins were mixed in reconstitution buffer with a ratio of 2:8:4:2:2 (α:β:β′:δ:ω) and adjusted to a final protein concentration of 0.5 mg/ml with reconstitution buffer. The mixture was then incubated at 50°C for 10 min to form RNAP core enzyme ($RNAP_C$).

$RNAP_C$ obtained using both reconstitution methods were further assembled to the respective RNAP holoenzymes ($RNAP_H$) by addition of purified σ^{70}-factor in an equimolar concentration and incubating at 37°C for 1 hr. Both reconstituted $RNAP_C$ and $RNAP_H$ were directly tested for transcription activity (see section 3.2). For long term storage, reconstitution buffer was changed to storage buffer by at least three rounds of fivefold dilution and concentration using VIVASPIN® 6 concentrator with MWCO of 50,000. The enzyme was then stored at 4°C.

2.5 | Purification of recombinant RNAP with co-expressed subunits

Co-expressed RNA polymerase core enzyme ($cRNAP_C$) was purified using two chromatographic steps. His_6-tagged $cRNAP_C$ and non-assembled β′ subunit were co-purified using HisTrap columns as described above. Protein-containing fractions were pooled and either purified further via anionic exchange chromatography (AEC) or buffer exchanged and protein concentration was performed as described below. For AEC pooled fractions were diluted 1:5 with TGE-buffer (40 mM Tris-HCl pH 7.4, 5% [v/v] glycerol, 1 mM EDTA, 1 mM β-mercaptoethanol [2-ME]). The protein solution was applied to a MonoQ column (*GE Healthcare*, Freiburg, Germany) equilibrated with TGE-buffer containing 100 mM NaCl. The column was washed with 10 CV TGE-buffer containing 100 mM NaCl. Elution was performed with a linear salt-gradient of 0.1–1 M NaCl in TGE-buffer collecting 1 ml fractions. Fractions containing pure $cRNAP_C$ were identified by SDS–PAGE analysis and transcription assays. Buffer was exchanged to 1x storage buffer or buffer A (10 mM Tris-acetate pH 8.2, 14 mM MgOAc, 60 mM potassium glutamate, 2 mM DTT) by at least three rounds of fivefold dilution and concentration using a VIVASPIN®6 concentrator as described above. Similarly, the core enzyme was further assembled to holoenzyme by addition of purified σ^{70}-factor in an equimolar concentration and incubation at 37°C for 1 hr.

TABLE 4 Thermostability of different *Geobacillus* RNA polymerase subunits

	Heat-treatment and SDS–PAGE analysis		nanoDSF
Subunit	$T°_{max}$ tested [°C]	time[min]	$T°_{max}$ [°C]
His_6-α	65	>90	79
His_6-β	50	10	53
His_6-β'	55	15	72
His_6-δ	80	> 90	89
His_6-σ^{70}	65	>80	71
His_6-ω	70	30	n.d.
ω	70	20–30	80
δ	80	>120	n.d.

$T°_{max}$ is the maximum temperature tested and time is duration of incubation at the respective temperatures used for the stability test. N.d. is for "not determined."

2.6 | In vitro transcription

In vitro transcription reactions contained 40 mM Tris-HCl pH 8.0, 20 mM $MgCl_2$, 2 mM spermidine, 5 mM DTT, rNTP mix (rATP, rCTP, rGTP, rUTP, 4 mM each), 0.1 µ/µl thermostable inorganic pyrophosphatase (*New England Biolabs GmbH*, Frankfurt am Main, Germany), and 1–2 µ/µl RiboLock RNase inhibitor (*ThermoFisher Scientific GmbH*, Schwerte, Germany). If not mentioned otherwise, a polymerase concentration of 0.25 µg/µl and RNAP holoenzyme reconstituted from His_6-tagged subunits was used. General transcription activity was assayed with template gDNA from *Geobacillus* sp. GHH01 isolated with the peqGOLD Bacterial DNA Kit (*PEQLAB Biotechnologie GmbH*, Erlangen, Germany). Metagenomic DNAs were extracted using the QIAamp DNA Stool kit from Qiagen (Hilden, Germany). A standard concentration of 0.5–1.0 µg template DNA was applied per 20 µl reaction mix and incubated at 37°C for 2 hr. The effect of transcription factors on in vitro transcription activity was examined by adding recombinantly expressed and purified GreA and MfD from *Geobacillus* sp. GHH001 in to transcription reaction mixtures. The outcome of transcription assays was monitored by electrophoresis in a 1.2% (w/v) agarose gel with 0.7% (v/v) formaldehyde.

Thermostability of the $RNAP_H$ was tested incubating the reaction mix at temperatures between 37 and 75°C. Activity was monitored by quantifying RNA in high sensitivity assays using Qubit 3.0 fluorometer (Life Technologies, ThermoFisher Scientific Inc.) according to the manufacturer's instructions.

To investigate the transcription coverage of the polymerase, different template DNAs other than genomic DNA (gDNA) from *Geobacillus* sp. GHH01 were employed. Accordingly, metagenomic DNAs from elephant feces (Ilmberger et al., 2014), a biogas reactor (Güllert et al., 2016), and a microalgae culture (Krohn-Molt et al., 2013) as well as gDNAs from *E. coli*, *Bacillus subtilis*, *Propionibacterium acidipropionici*, *Paenibacillus* sp., and *Picrophilus torridus* were used as templates for further transcription assays. A commercial *E. coli* RNAP

holoenzyme from New England Biolabs was used according to manufacturer's recommendations to compare transcription profiles. To determine the total amount of synthesized RNA, the template DNA was removed by addition of DNase (2 μ/μg template DNA) and incubated at 37°C for 20 min. Afterwards, RNA was purified using the SureClean Plus solution (*Bioline*, Luckenwalde, Germany) or the RNA clean & concentrator kit by Zymo Research Europe-Germany (Freiburg, Germany). Finally, RNA concentration was measured spectrophotometrically using Nanodrop ND-2000 (PEQLAB Biotechnologie GmbH, Erlangen, Germany).

In preparation of a first-hand-established translation system coupled with in vitro transcription, we also examined the transcription performance of the recombinant RNAP on circular plasmid DNA and linear PCR products as template. We tested three different template DNAs which were (i) a circular vector harboring *eGFP* under the control of a lac promoter (peGFP cloning vector, peGFPcv); (ii) a PCR product of *eGFP* including an upstream lac promoter and additional downstream nucleotides; and (iii) a PCR product of a metagenome-derived lipase *lip41* including a putative upstream promoter-region as well as additional downstream nucleotides. Transcription was performed using co-expressed core enzyme (cRNAP$_C$) with σ^{70}-factor added separately to the reaction master mix.

3 | RESULTS

3.1 | Heterologous expression and purification of subunits and RNAP core enzyme

The RNAP of *Geobacillus* consists of five different subunits ($\alpha_2\beta\beta'\delta\omega$) forming the core enzyme (RNAP$_C$) (Figure 1). Genes coding for RNAP subunits (*rpoA*, *rpoB*, *rpoC*, *rpoE*, *rpoZ*) and the house-keeping σ^{70}-factor (*rpoD*) were cloned as described in section 2 into two different expression vectors (pET21a[+] and pET28a[+]) for heterologous expression of untagged or His$_6$-tagged proteins, respectively (Figures 1a and 1b). With the N-terminal His$_6$-tag, all subunits were expressed as soluble proteins and purified using affinity chromatography. Untagged δ- and ω- subunits were located in the soluble fraction and crudely purified using heat treatment, whereas the α-, β-, and β′-subunits were found in inclusion bodies and purified using repeated washing with 0.2% Na-deoxycholate containing lysis buffer no. 2 (Figures 2a and b). There was no detectable expression of the untagged σ-factor.

In parallel, a plasmid clone with pET21a(+)-vector-backbone was constructed encoding all subunits of the RNAP core enzyme, each under control of its own T7-promoter (Figure 1C). The plasmid was designated as pET-21a::rpoCBZAE. The first gene of this construct, the *rpoC* coding for the β'-subunit, was fused with an N-terminal His$_6$-tag for RNAP$_C$ purification. Simultaneous co-expression of all five genes for 17 hr in *E. coli* allowed the *Geobacillus* RNA polymerase subunits to assemble in vivo to cRNAP$_C$.

Purification of in vivo co-expressed RNAP$_C$ was achieved with a two-step chromatographic approach. As a first step, His$_6$-tag based chromatographic purification resulted in fractions containing non-assembled His$_6$-β′ along with cRNAP$_C$. Pure cRNAP$_C$ was obtained in the second purification step by anion exchange chromatography (AEC) using MonoQ column. The corresponding chromatogram is depicted in Figure 2c. Proteins were eluted in three peaks whereby the first peak contained pre- and nonassembled subunits. The cRNAP$_C$ was detected in the second peak and eluted at approximately 460 mM NaCl. In an SDS–PAGE analysis of this peak, all five subunits can be detected in a reasonable ratio as it occurs in the core enzyme (Figure 2d). The following fractions contained mainly free His$_6$-β′ subunit. The presence of active polymerase was verified by testing the fractions for transcription activities (Figure S3). Purified transcription elongation factors GreA and MfD proteins were obtained the same way as His$_6$-tagged RNAP subunits.

3.2 | Reconstitution and transcriptional activity

The recombinant RNAP core enzyme was obtained through (i) reconstitution from individually overexpressed subunits in vitro to RNAP$_C$ and (ii) co-expression of the five RNAP-subunits in a single vector construct and simultaneous in vivo assembly of cRNAP$_C$ followed by chromatographic purification (section 3.1). RNAP$_C$ from individually expressed and purified subunits was assembled using the classical reconstitution method as described in the section 2. The transcription activities of assembled core and holoenzymes were also compared (Figure S2a). In addition to the urea-based denaturation and subsequent renaturation, a new RNAP$_C$ assembly method referred as thermo-reconstitution, where the various subunits were mixed at equimolar concentrations and incubated at 50°C for 10 min, was established. This method shortened the RNAP$_C$ assembly time from 17 hr to just 10 min of incubation avoiding the need of extended dialysis and the use of large volume of buffer, yet, yielding comparably active polymerase (Figure S2b). RNAP core enzymes assembled with both reconstitution methods were either used directly for transcription assay or further incubated with σ^{70}-factor to form the respective holoenzymes.

Both reconstitution methods resulted in functional RNAPs. Transcription activity was primarily investigated using gDNA from *Geobacillus* as template (Figure 3). As part of an effort to enhance the efficiency of transcription, the effect of two elongation factors from *Geobacillus* sp. GH001 on the amount of RNA generated was examined. Accordingly, addition of GreA improved transcript quantity significantly whereas MfD did not have a detectable effect (Figure S2C). Depending on the template DNA used for transcription, the amount of produced RNA was approximately in a range of 92–316.4 μg RNA/μg His$_6$-tagged RNAP$_H$ and the amount of synthesized RNA correlated with the amount of polymerase added (Figure 3 and Table 5). Comparing the transcription product of RNAP assembled from all His$_6$-tagged and all untagged subunits, the His$_6$-tag showed no measurable negative effect on the overall transcription activity (Figures 3a and 3b). Poly-histidine-tag has rather allowed to achieve larger quantities of subunit proteins in their native state with higher purity. As a result, thermo-reconstitution of RNAP$_C$ was made possible as well as enhanced reconstitution efficiency as evidence by increased transcription activity (Figures 3a and 3b). Transcription activity was also decreased when employing polymerase reconstituted from all soluble subunit proteins except β-subunit from inclusion bodies (Figure S2B).

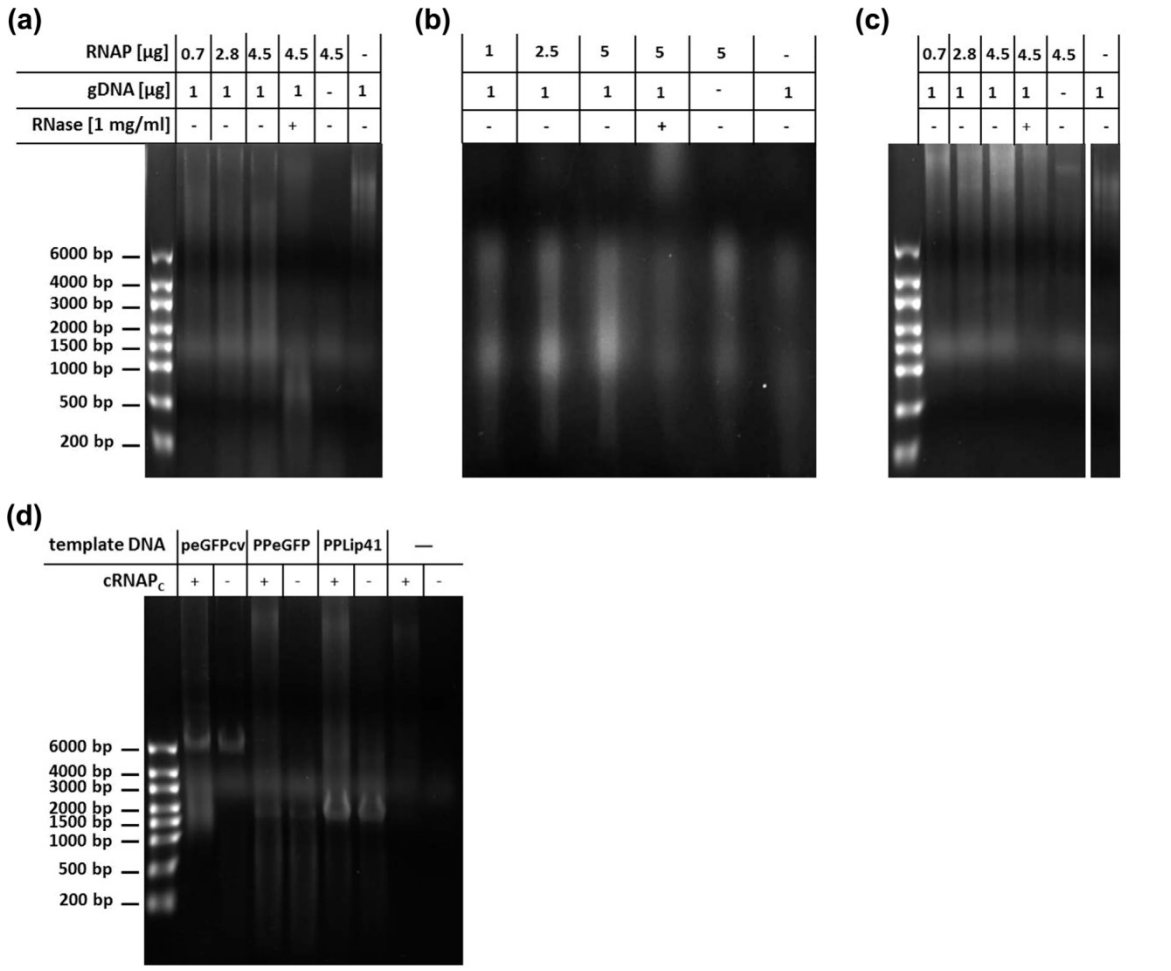


FIGURE 3 In vitro transcription activity of *Geobacillus* RNAP. (a) Activity of RNAP reconstituted from un-tagged subunits. (b) Activity of RNAP reconstituted from His_6-tagged soluble subunits. (c) Activity of co-expressed and in vivo reconstituted RNAP (cRNAP). All RNAPs employed above were in their holoenzyme forms assembled with the house keeping σ^{70}-factor. Genomic DNA (gDNA) from *Geobacillus* sp. GHH01 was used as template for transcription. (d) Transcripts of peGFPcv plasmid DNA construct and PCR product (PP) of *eGFP* and metagenome-derived *Lip41* lipase genes using cRNAP core enzyme ($cRNAP_C$). RNA was run in 1.2% agarose gels containing 0.7% (v/v) formaldehyde and stained with ethidium bromide included in the loading dye (*ThermoFisher Scientific GmbH*, Schwerte, Germany)

To test the promiscuity of the transcription and hence the applicability of *Geobacillus* RNAP for cell-free metagenomics, we investigated its activity on meta/genomic DNAs derived from variety of microorganisms and metagenome samples. We employed three different metagenome samples, an archaeal and five bacterial genomic DNAs as template for RNA production using *Geobacillus* and *E. coli* RNAPs (Table 5 and Figure 4). *Geobacillus* RNAP synthesized up to 65.4 ± 6.9 ng/µl metagenomic mRNA and 79.1 ± 9.2 ng/µl mRNA from individual genomic DNA templates. Interestingly, it also transcribed archaeal genomic DNA from *P. torridus* with a yield of 23.5 ± 4.9 ng/µl RNA. The amount of mRNA synthesized with *Geobacillus* $RNAP_H$ was in most cases significantly (up to three times) higher compared to the transcript produced by employing commercial *E. coli* $RNAP_H$ under comparable conditions (Table 5). T7 RNAP was left out as there was no detectable transcription activity on DNA templates with innate promoters.

Transcription reactions employing defined genes as template DNAs comprised a circular plasmid construct containing an *eGFP* gene or linear PCR products with *eGFP* and a *lip41* gene from a metagenomic lipase. All reactions clearly resulted in the production of mRNAs (Figure 3d). The assembled RNAP enzyme was also stable for several months stored at 4°C, while freezing at −20°C with glycerol led to a dramatic decrease in transcription activity within few weeks.

3.3 | Thermostability of subunits and assembled RNAP

Temperature stability of purified soluble recombinant RNAP subunits was assessed by heat incubation at different temperatures followed by SDS–PAGE analysis of residual protein and using nanoDSF method. Interestingly, the thermostability of the subunits differed significantly

TABLE 5 Quantification of RNA yields transcribed using RNA polymerases from *Geobacillus* sp. GHH01 and *E. coli*

	Yield (ng/µl RNA)	
Template DNA source	***Geobacillus* $RNAP_H$**	***E. coli* $RNAP_H$**
Elephant feces metagenome (Ilmberger et al., 2014)	32.7 + 3.9	23.5 + 4
Biogas reactor metagenome (Güllert et al., 2016)	65.4 + 6.9	26.1 + 0.3
Bacterial metagenome of microalgae (Krohn-Molt et al., 2013)	43.8 + 0.5	14.8 + 1.9
Geobacillus sp. GHH01	79.1 + 9.2	43.6 + 8.1
Geobacillus sp. GHH01	52.9*	–
E. coli	52.2 + 9.1	77.6 + 7.2
Bacillus subtilis	42.9 + 9.1	32.3 + 6.4
Propionibacterium acidipropionici	23 + 3.5	9.7 + 1.6
Paenibacillus sp.	29.1 + 6.2	19.8 + 2.9
Picrophilus torridus	23.5 + 4.9	30.4 + 0.8
Without template DNA	4.2 + 1.1	5.8 + 1

Assays were performed in 20 µl reaction volumes at 37°C for 2 hr in triplicates. Transcription with commercial *E. coli* RNAP holoenzyme was carried out according to manufacturer's instructions. (*) Transcription was performed using *GHH01* $RNAP_C$ while $RNAP_H$ was used for all other transcription reactions.

and was independent on the presence or absence of the recombinantly fused poly-histidine protein tag. The α-subunit and σ^{70}-factor were stable at 65°C for approximately 80–90 min and the ω-subunit was stable at 70°C for 30 min. The highest thermostability was observed for the δ-subunit which resisted well at 80°C for at least 2 hr. The β- and β′-subunits showed a thermostability of approximately 10–15 min at temperatures of 50 and 55°C, respectively. Thermostability data generated by nanoDSF confirmed their stability profiles showing that most stable subunits were δ- and α-subunits followed by β′-subunit, σ^{70}-factor and β-subunit (see Table 4). Accordingly, a maximum non-denaturing temperature of 50°C was chosen for assembly of $RNAP_C$ with thermos-reconstitution.

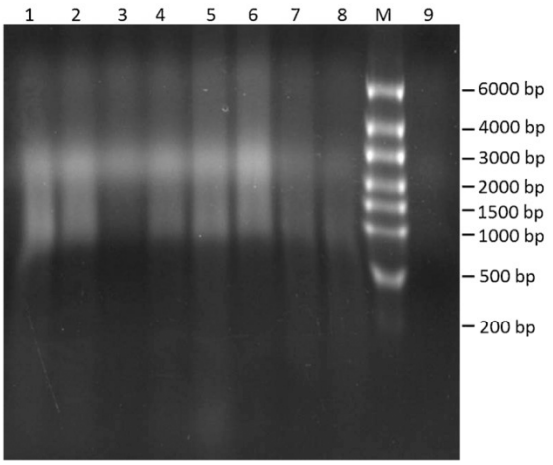


FIGURE 4 Transcription potential of *Geobacillus* RNAP on meta/genomic DNA templates. Transcription reactions were carried out using genomic, metagenomic, and mock metagenomic DNA templates. Lane 1, 2, and 3 show transcripts produced from metagenomic DNA templates isolated from elephant feces, a biogas reactor and bacteria associated with microalgae, respectively. Lane 5 shows a transcription on mock metagenomic DNA consisting of 10 different microbial sources. Lane 4, 6, and 7 show transcriptions on *E. coli*, *Geobacillus* sp. GHH01 and *P. torridus* gDNAs, respectively. Lane 8 shows the transcription on *Geobacillus* gDNA with $RNAP_C$ while all other transcriptions were done using $RNAP_H$. Lane 9 shows a transcription reaction without template DNA and M is for Riboruler high range RNA ladder (*ThermoFisher Scientific GmbH*, Schwerte, Germany). RNA was run in 1.2% agarose gels containing 0.7% (v/v) formaldehyde and stained with ethidium bromide included in the loading dye (*ThermoFisher Scientific GmbH*, Schwerte, Germany)

Further tests were performed to investigate thermostability of the assembled $RNAP_H$ complexes. Comparative analysis of in vitro transcription profiles with commercial *E. coli* $RNAP_H$ indicated that $RNAP_H$ from *Geobacillus* is stable and active over a broad temperature range under the conditions tested. *Geobacillus* $RNAP_H$ transcribed effectively at temperatures from 37°C to up to 55°C maintaining its maximum relative activity (Figure 5a) while dramatic loss of *E. coli* $RNAP_H$ activity was observed from 45°C and above (Figure 5b).

4 | DISCUSSION

Functional metagenome screening is hampered by many problems such as low transcription of the desired target genes, poor translation, and/or incorrect folding of heterologous proteins or the inability of the host to secrete the recombinant enzymes in an active form (Ekkers et al., 2012; Lam et al., 2015; Steele et al., 2009; Uchiyama & Miyazaki, 2009). In addition, classical functional screenings are very time consuming and large numbers of metagenome library clones must be assayed before few positive clones can be detected (Culligan et al., 2014; Gabor et al., 2004). These clones even need to be further validated by subcloning into suitable expression vectors which again requires additional time and face problems associated with poor protein expression or folding and lack of appropriate secretion systems in expression hosts. To address these problems, we have developed a recombinant in vitro and in vivo assembled thermostable RNA polymerase derived from *Geobacillus* sp. GHH01. We have shown

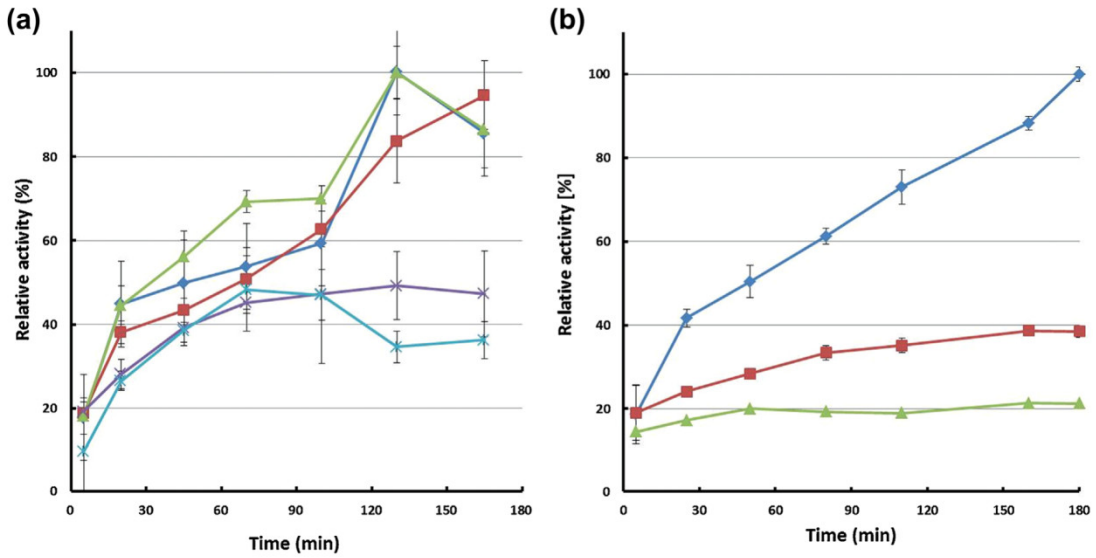


FIGURE 5 Temperature stability and activity profiles of *Geobacillus* $RNAP_H$ (b) and *E. coli* $RNAP_H$ (a). Transcription reactions were carried out at 37°C (◆), 45°C (■), 55°C (▲), 65°C (×), and 75°C (*). High quality *Geobacillus* and *E. coli* gDNAs were employed for the respective polymerase reactions. RNA quantification was done in high-sensitivity assays using Qubit 3.0 fluorometer

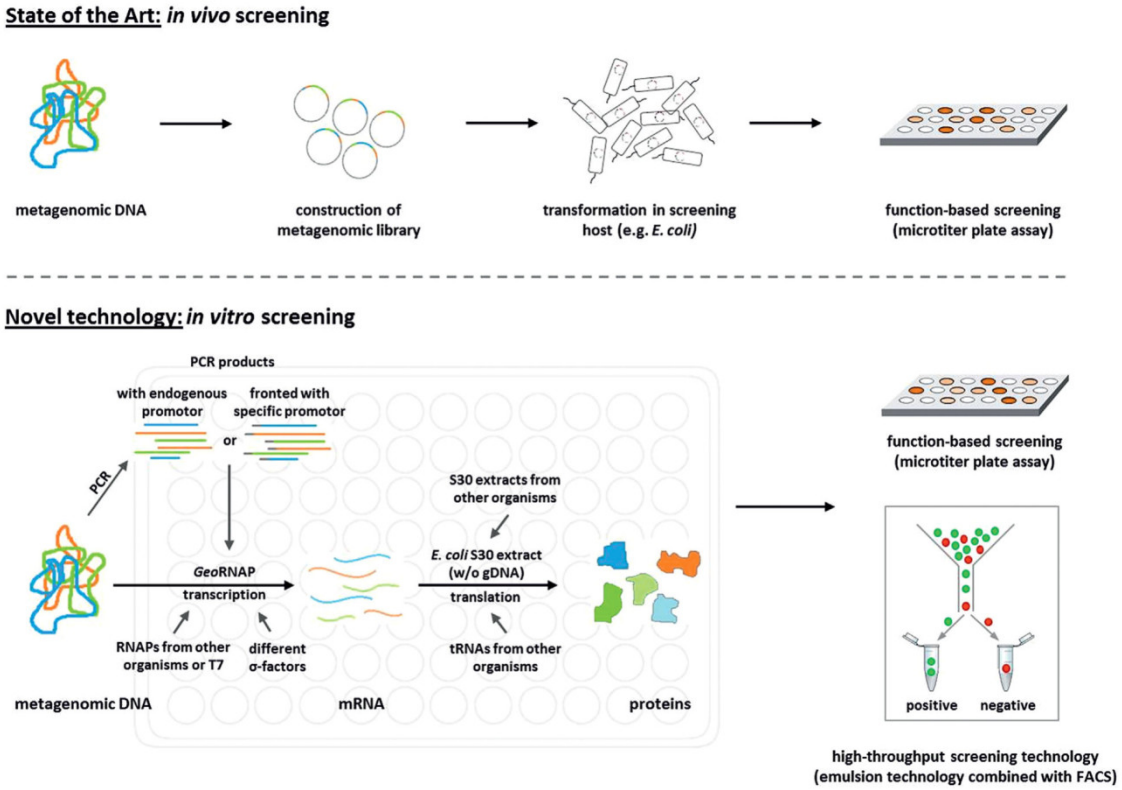


FIGURE 6 Outline of a newly proposed cell-free functional metagenomic approach in comparison to classical in vivo functional screening. In addition to being time efficient, the new method offers multiple possibilities of fine tuning with more control to the system. This may include using mixed RNA polymerases, transcription factors, and translation mixtures (S30 extracts) of various origins to increase the chance of transcription/expression coverage. It can further be performed in in vitro compartmentalization combined with high-throughput screening technologies

that this RNAP is suitable for direct transcription of DNA templates derived from single organisms and metagenomes which can be further coupled to translation mixtures for cell-free activity screening.

Our data indicate that the RNAP has a robust transcription activity and produces mRNA from various DNA templates including archaea and diverse bacteria. In vitro activity of the core enzyme is largely improved upon further assembly with a σ^{70}-factor which suggests its indispensable DNA binding role. Transcription process is also known to be greatly affected by elongation factors in multiple ways (Borukhov, Lee, & Laptenko, 2005). In this work, GreA and MfD elongation factors were employed with the aim to boost transcription activity. The addition of GreA to transcription reactions significantly increased yield of transcripts possibly by resolving RNAP stalling and increasing its recycling rate (Kusuya, Kurokawa, Ishikawa, Ogasawara, & Oshima, 2011; Stepanova et al., 2007). Interestingly, another transcription elongation factor MfD did not have a visible effect in terms of transcript amount generated.

Another important feature of this polymerase is its thermostability. The *Geobacillus* RNA polymerase was stable at elevated temperatures up to 55°C maintaining its highest relative activity. Since this RNAP was derived from a thermostable microorganism (i.e., *Geobacillus* sp.), it is not too surprising that the enzyme was active at higher temperatures. We speculate that this will be of advantage for further use of the enzyme during mRNA synthesis from a wide variety of organisms and metagenomes. The moderate thermostability will be beneficial with respect to both handling and storage of the enzyme. What is impressive is that the thermostability of RNAP subunits allowed their effective assembly into a functional polymerase core enzyme using a new thermo-reconstitution method. The assembly occurs within 10 min of incubation at 50°C compared to the long and resource intensive classical reconstitution method. A co-expression of all subunits in a single plasmid allowing in vivo assembly is also an effective alternative which simplifies the process of protein purification from individual subunits to just the assembled polymerase.

More interestingly, this polymerase can directly transcribe DNA templates of diverse origins including archaeal, bacterial, and metagenomic DNAs as it recognizes a broad variety of promoters. This is considered to be a key feature of the polymerase as broader transcription coverage means greater potential in resolving transcription biases toward host system. The obtained transcription yields from linear DNA templates applying the recombinant $RNAP_H$ were also up to three times higher compared to commercial *E. coli* $RNAP_H$ transcripts (Table 5). This implies higher amounts of protein after a subsequent in vitro translation. In addition to the commercial *E. coli* RNAP we have employed, studies on other polymerases exist, but they primarily address recombinant enzyme preparation and structural characterization, and understanding of the transcription process in vitro employing DNA with predefined promoters (Esyunina & Kulbachinskiy, 2015; Herrera-Asmat et al., 2017; Kuznedelov & Severinov, 2009; Kuznedelov, Minakhin, & Severinov, 2003; Yang & Lewis, 2008). To our knowledge, this is the first report on a direct in vitro transcription of meta/genomic DNA templates using a recombinant RNAP capable of recognizing native promoter sequences.

In light of the challenges associated with the classical functional metagenomics, existing cell-free expression systems and the already reported polymerases, the recombinant *Geobacillus* RNA polymerase has multiple advantages. The most important one lays in its promiscuous transcription activity which produces mRNAs from a variety of DNA sources with native promoters. Secondly, its thermostability is crucial during operation and storage. At last, the polymerase is reconstituted from purified subunits which results in minimal background activities in contrast to polymerases from cell extract preparations.

Development of a robust and stable RNA polymerase for the transcription of metagenomic DNAs is a vital step towards establishing a cell-free metagenomics as outlined in Figure 6. To overcome the multitude of challenges associated with host strains used for library construction and/or expression of recombinant proteins in functional metagenome searches, we suggest developing a cell-free metagenomics approach that will allow a host independent screening and expression of the desired proteins. While this is certainly a very worthwhile task, it perhaps takes several years to develop such a system in its complete and effective form. Within this framework, the current manuscript describes a first step toward the establishment of such a cell-free functional metagenomics platform.

ACKNOWLEDGMENTS

This work was in part funded by the German Ministry of Education and Research "Bundesministerium für Bildung und Forschung" (BMBF) and the Danish Innovationsfonden within the ERA-IB network program MetaCat and through the BMBF-cooperation project iVDT-V^2. Deutsche Akademischer Austauschdienst (DAAD) is gratefully acknowledged for the scholarship provided to B.M.K.

REFERENCES

Borukhov, S., Lee, J., & Laptenko, O. (2005). Bacterial transcription elongation factors: New insights into molecular mechanism of action. *Molecular Microbiology*, *55*(5), 1315–1324.

Bradford, P., Pohle, W. D., Gunther, J. K., & Mehlenbacher, V. C. (1942). Determination of glycerol by oxidation with periodic acid. *Oil & Soap*, *19*(11), 189–193.

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.

Courtois, S., Cappellano, C. M., Ball, M., Francou, F. X., Normand, P., Helynck, G., … Pernodet, J. L. (2003). Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Applied and Environmental Microbiology*, *69*(1), 49–55.

Culligan, E. P., Sleator, R. D., Marchesi, J. R., & Hill, C. (2014). Metagenomics and novel gene discovery: Promise and potential for novel therapeutics. *Virulence*, *5*(3), 399–412.

DeCastro, M. E., Rodriguez-Belmonte, E., & Gonzalez-Siso, M. I. (2016). Metagenomics of thermophiles with a focus on discovery of novel thermozymes. *Frontiers Microbiology*, *7*(1521), 1521.

Ekkers, D. M., Cretoiu, M. S., Kielak, A. M., & Elsas, J. D. (2012). The great screen anomaly-a new frontier in product discovery through functional metagenomics. *Applied Microbiology and Biotechnology*, *93*(3), 1005–1020.

Esyunina, D. M., & Kulbachinskiy, A. V. (2015). Purification and characterization of recombinant *Deinococcus radiodurans* RNA polymerase. *Biochemistry (Moscow)*, *80*(10), 1271–1279.

Ferrer, M., Martinez-Martinez, 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. *Microbiology Biotechnology*, *9*(1), 22–34.

Gabor, E. M., Alkema, W. B., & Janssen, D. B. (2004). Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environmental Microbiology*, *6*(9), 879–886.

Gaida, S. M., Sandoval, N. R., Nicolaou, S. A., Chen, Y., Venkataramanan, K. P., & Papoutsakis, E. T. (2015). Expression of heterologous sigma factors enables functional screening of metagenomic and heterologous genomic libraries. *Nature Communications*, *6*, 7045.

Güllert, S., Fischer, M. A., Turaev, D., Noebauer, B., Ilmberger, N., Wemheuer, B., … Streit, W. R. (2016). Deep metagenome and metatranscriptome analyses of microbial communities affiliated with an industrial biogas fermenter, a cow rumen, and elephant feces reveal major differences in carbohydrate hydrolysis strategies. *Biotechnology for Biofuels*, *9*(1), 121.

Herrera-Asmat, O., Lubkowska, L., Kashlev, M., Bustamante, C. J., Guerra, D. G., & Kireeva, M. L. (2017). Production and characterization of a highly pure RNA polymerase holoenzyme from *Mycobacterium tuberculosis*. *Protein Expression and Purification*, *134*, 1–10.

Hiroyuki, O., Takashi, K., Shimizu, Y., & Takuya, U. (2010). A highly controllable reconstituted cell-free system—A breakthrough in protein synthesis research. *Current Pharmaceutical Biotechnology*, *11*(3), 267–271.

Ilmberger, N., Gullert, S., Dannenberg, J., Rabausch, U., Torres, J., Wemheuer, B., … Streit, W. R. (2014). A comparative metagenome survey of the fecal microbiota of a breast- and a plant-fed Asian elephant reveals an unexpectedly high diversity of glycoside hydrolase family enzymes. *PLoS ONE*, *9*(9), e106707.

Iqbal, H. A., Feng, Z., & Brady, S. F. (2012). Biocatalysts and small molecule products from metagenomic studies. *Current Opinion in Chemical Biology*, *16*(1–2), 109–116.

Kakirde, K. S., Wild, J., Godiska, R., Mead, D. A., Wiggins, A. G., Goodman, R. M., … Liles, M. R. (2011). Gram negative shuttle BAC vector for heterologous expression of metagenomic libraries. *Gene*, *475*(2), 57–62.

Körfer, G., Pitzler, C., Vojcic, L., Martinez, R., & Schwaneberg, U. (2016). In vitro flow cytometry-based screening platform for cellulase engineering. *Scientific Reports*, *6*, 26128.

Kovach, M. E., Elzer, P. H., Steven Hill, D., Robertson, G. T., Farris, M. A., Roop Ii, R. M., & Peterson, K. M. (1995). Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene*, *166*(1), 175–176.

Krohn-Molt, I., Wemheuer, B., Alawi, M., Poehlein, A., Güllert, S., Schmeisser, C., … Streit, W. R.(2013). Metagenome survey of a multispecies and alga-associated biofilm revealed key elements of bacterial-algal interactions in photobioreactors. *Applied and Environmental Microbiology*, *79*(20), 6196–6206.

Kudla, G., Murray, A. W., Tollervey, D., & Plotkin, J. B. (2009). Coding-sequence determinants of gene expression in *Escherichia coli*. *Science*, *324*(5924), 255–258.

Kusuya, Y., Kurokawa, K., Ishikawa, S., Ogasawara, N., & Oshima, T. (2011). Transcription factor GreA contributes to resolving promoter-proximal pausing of RNA polymerase in *Bacillus subtilis* cells. *Journal of Bacteriology*, *193*(12), 3090–3099.

Kuznedelov, K., & Severinov, K. (2009). Recombinant bacterial RNA polymerase: Preparation and applications. *Methods*, *47*(1), 44–52.

Kuznedelov, K., Minakhin, L., & Severinov, K. (2003). Preparation and characterization of recombinant *Thermus aquaticus* RNA polymerase. *Methods in Enzymology*, *370*, 94–108.

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*(1196).

Leis, B., Angelov, A., & Liebl, W. (2013). Screening and expression of genes from metagenomes. *Advances in Applied Microbiology*, *83*, 1–68.

Leis, B., Angelov, A., Mientus, M., Li, H., Pham, V. T. T., Lauinger, B., … Liebl, W. (2015). Identification of novel esterase-active enzymes from hot environments by use of the host bacterium *Thermus thermophilus*. *Frontiers in Microbiology*, *6*, 275.

Li, Y., Wexler, M., Richardson, D. J., Bond, P. L., & Johnston, A. W. (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*(12), 1927–1936.

Liebl, W., Angelov, A., Juergensen, J., Chow, J., Loeschcke, A., Drepper, T., … Jaeger, K. E. (2014). Alternative hosts for functional (meta)genome analysis. *Applied Microbiology and Biotechnology*, *98*(19), 8099–8109.

Maroti, G., Tong, Y., Yooseph, S., Baden-Tillson, H., Smith, H. O., Kovacs, K. L., … Xu, Q. (2009). Discovery of [NiFe] hydrogenase genes in metagenomic DNA: Cloning and heterologous expression in *Thiocapsa roseopersicina*. *Applied and Environmental Mi crobiology*, *75*(18), 5821–5830.

Martinez, A., Kolvek, S. J., Yip, C. L., 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*(4), 2452–2463.

Pardee, K., Slomovic, S., Nguyen P. Q., Lee J. W., Donghia, N., Burrill, D., … Collins, J. J. (2016). Portable, on-demand biomolecular manufacturing. *Cell*, *167*(1), 248–259.

Rabausch, U., Juergensen, J., Ilmberger, N., Böhnke, S., Fischer, S., Schubach, B., … Streit, W. R. (2013). Functional screening of metagenome and genome libraries for detection of novel flavonoid-modifying enzymes. *Applied and Environmental Microbiology*, *79*(15), 4551–4563.

Rousset, M., Casalot, L., Rapp-Giles, B. J., Dermoun, Z., de Philip, P., Belaich, J. P., & Wall, J. D. (1998). New shuttle vectors for the introduction of cloned DNA in *Desulfovibrio*. *Plasmid*, *39*(2), 114–122.

Schallmey, M., Ly, A., Wang, C., Meglei, G., Voget, S., Streit, W. R., … Charles, T. C. (2011). Harvesting of novel polyhydroxyalkanaote (PHA) synthase encoding genes from a soil metagenome library using phenotypic screening. *FEMS Microbiology Letters*, *321*(2), 150–156.

Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., & Ueda, T. (2001). Cell-free translation reconstituted with purified components. *Nature Biotechnology*, *19*(8), 751–755.

Steele, H. L., Jaeger, K. E., Daniel, R., & Streit, W. R. (2009). Advances in recovery of novel biocatalysts from metagenomes. *Journal of Molecular Microbiology and Biotechnology*, *16*(1-2), 25–37.

Stepanova, E., Lee, J., Ozerova, M., Semenova, E., Datsenko, K., Wanner, B. L., … Borukhov, S. (2007). Analysis of promoter targets for *Escherichia coli* transcription elongation factor GreA in vivo and In vitro. *Journal of Bacteriology*, *189*(24), 8772–8785.

Streit, W., Daniel, R., & Jaeger, K. (2004). Prospecting for biocatalysts and drugs in the genomes of non-cultured microorganisms. *Current Opinion in Biotechnology*, *15*(4), 285–290.

Troeschel, S. C., Drepper, T., Leggewie, C., Streit, W. R., & Jaeger, K.-E. (2010). Novel tools for the functional expression of metagenomic DNA. *Methods in Molecular Biology (Clifton. N.J.)*, *668*, 117–139.

Tuckey, C., Asahara, H., Zhou, Y., Chong, S. (2014). Protein synthesis using a reconstituted cell-free system. *Current Protocols in Molecular Biology*. Edited by Frederick M. Ausubel. [et al.]. *108*:16.31.1–16.31.22.

Tuffin, M., Anderson, D., Heath, C., & Cowan, D. A. (2009). Metagenomic gene discovery: How far have we moved into novel sequence space? *Biotechnology Journal*, *4*(12), 1671–1683.

Uchiyama, T., & Miyazaki, K. (2009). Functional metagenomics for enzyme discovery: Challenges to efficient screening. *Current Opinion Biotechnology*, *20*(6), 616–622.

Wang, C., Meek, D. J., Panchal, P., Boruvka, N., Archibald, F. S., Driscoll, B. T., & Charles, T. C. (2006). Isolation of poly-3-hydroxybutyrate metabolism genes from complex microbial communities by phenotypic complementation of bacterial mutants. *Applied and Environmental Microbiology*, *72*(1), 384–391.

Warren, R. L., Freeman, J. D., Levesque, R. C., Smailus, D. E., Flibotte, S., & Holt, R. A. (2008). Transcription of foreign DNA in *Escherichia coli*. *Genome Research*, *18*(11), 1798–1805.

Wexler, M., Bond, P. L., Richardson, D. J., & Johnston, A. W. (2005). A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase. *Environmental Microbiology*, *7*(12), 1917–1926.

Wiegand, S., Rabausch, U., Chow, J., Daniel, R., Streit, W. R., & Liesegang, H. (2013). Complete genome sequence of *Geobacillus* sp. strain GHH01, a thermophilic lipase-Secreting bacterium. *Genome Announcements*, *1*(2), e0009213.

Yang, X., & Lewis, P. J. (2008). Overproduction and purification of recombinant *Bacillus subtilis* RNA polymerase. *Protein Expression and Purification*, *59*(1), 86–93.

Zemella, A., Thoring, L., Hoffmeister, C., & Kubick, S. (2015). Cell-free protein synthesis: Pros and cons of prokaryotic and eukaryotic systems. *Chembiochem*, *16*(17), 2420–2431.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Kinfu BM, Jahnke M, Janus M, et al. Recombinant RNA Polymerase from *Geobacillus* sp. GHH01 as tool for rapid generation of metagenomic RNAs using in vitro technologies. *Biotechnology and Bioengineering*. 2017;114:2739–2752. https://doi.org/10.1002/bit.26436

Supplementary Material for DOI:10.1002/bit.26436

Table S1: Expression conditions for individual RNAP subunits, co-assembled $RNAP_C$ and transcription factors. The expression conditions for each plasmid construct are listed below. The *E. coli* expression host strain used, the IPTG concentration and temperature (T) for induction as well as the incubation time after induction are described. The amount of expressed soluble protein in some cases was highly dependent on culture volume.

Plasmid constructs	*E. coli* expression strain type	IPTG (mM)	T (°C)	Induction time (hr)
pET21a(+)::*rpoA*	BL21(DE3)	1.0	28	1
pET21a(+)::*rpoB*	BL21(DE3)	1.0	37	3
pET21a(+)::*rpoC*	BL21(DE3)	1.0	22	3
pET21a(+)::*rpoE*	BL21(DE3)	1.0	37	6
pET21a(+)::*rpoZ*	Rosetta gami 2	0.5	37	17
pET21a(+)::*rpoD*	*no detectable expression*			
pET28a(+)::*rpoA*	BL21(DE3)	0.4	37	6
pET28a(+)::*rpoB*	Rosetta gami 2	0.4	22	17
pET28a(+)::*rpoC*	BL21(DE3)	0.4	22	6
pET28a(+)::*rpoE*	BL21(DE3)	0.4	37	6
pET28a(+)::*rpoZ*	BL21(DE3)	0.4	22	6
pET28a(+)::*rpoD*	Rosetta gami 2	0.4	22	17
pET21a(+)::*rpoCBZAE*	Rosetta gami 2	0.5	37	17
pET-28a::*greA*	BL21(DE3)	0.5	37	17
pET-28a::*MfD*	BL21(DE3)	0.5	22	17

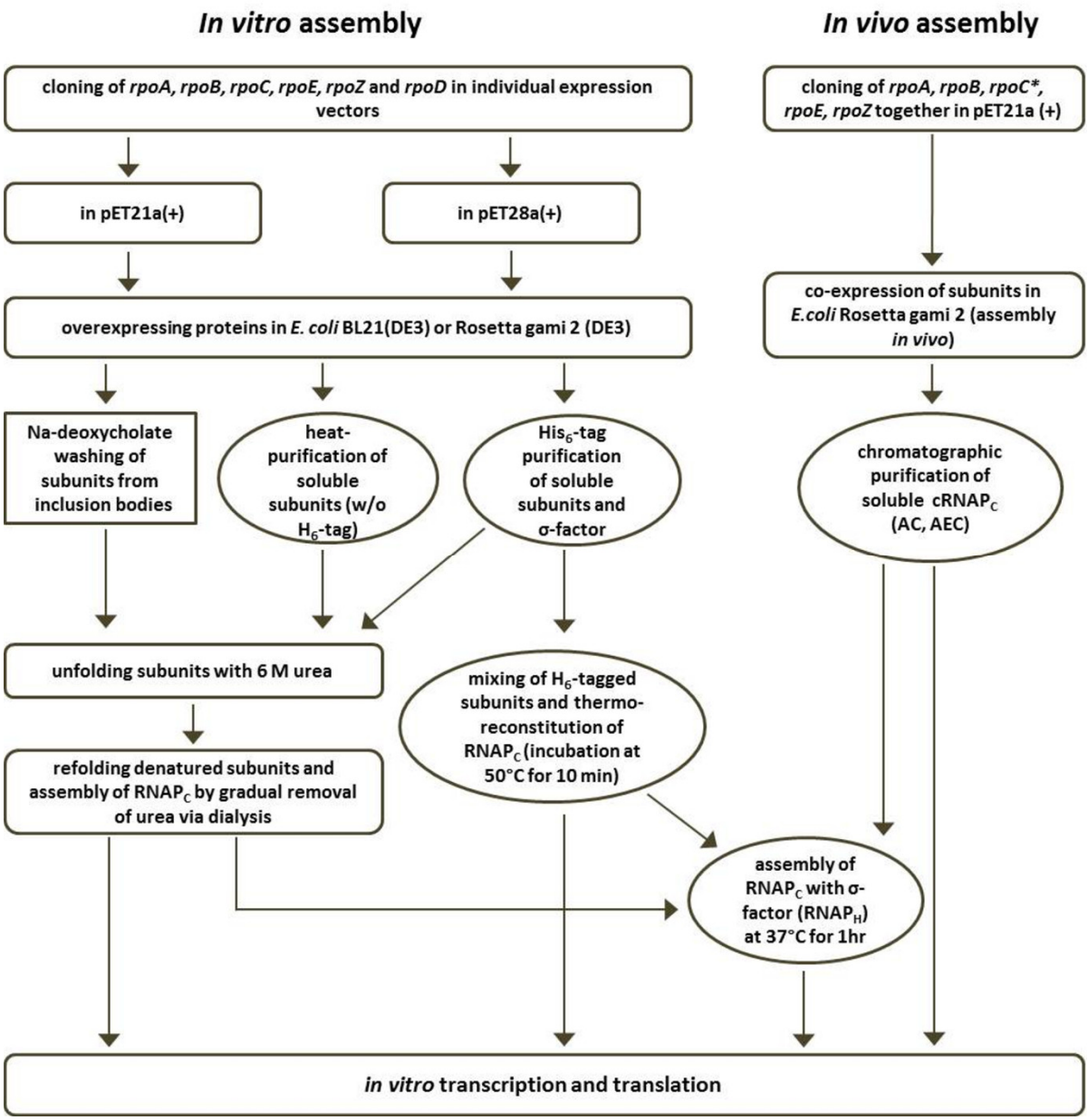


FIGURE S1: General outline of working strategy. Cloning, over-expression and purification of subunits, and RNAP reconstitution (left part); and cloning all subunits in a single vector system, co-expression and purification of *in vivo* assembled RNAP (right part). *RpoC** is cloned with His_6-tag coding sequence for chromatographic purification of co-expressed RNAP (cRNAP). $RNAP_C$ and $RNAP_H$ stand for RNAP core and holoenzyme respectively.

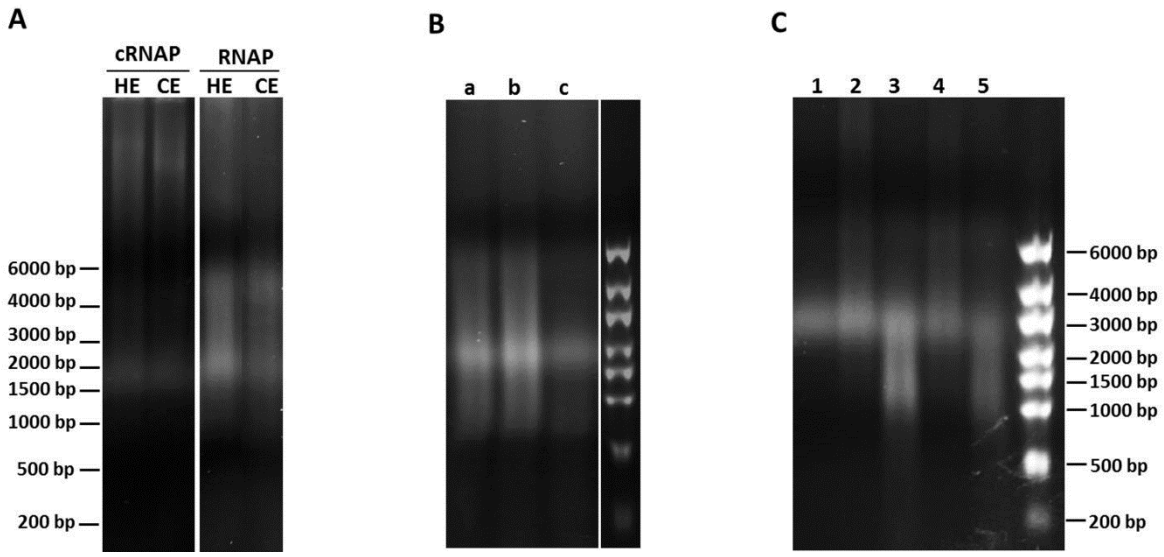


FIGURE S2: Reconstitution and activity test of RNAP using *Geobacillus* template gDNA. (A) Transcription activity of the core (CE) and holo (HE)-enzymes of co-expressed cRNAP and *in vitro* reconstituted RNAP. (B) Thermo-reconstitution (a) as an alternative method for RNAP reconstitution compared to the urea-reconstituted RNAP from all soluble subunit proteins (b). Transcription activity of polymerase reconstituted from soluble subunits α, β', δ and ω but an insoluble β-subunit protein (c). (C) Effect of elongation factors on transcription activity. 1: Negative control without DNA, 2: transcription without transcription factors, 3: transcription with GreA, 4: transcription with MfD and 5: both GreA and MfD were added to the transcription reaction. Transcription activity was analyzed in 1.2% agarose gel containing 0.7 % (v/v) formaldehyde.

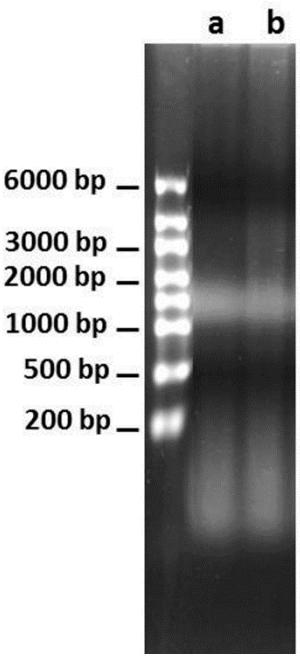


FIGURE S3: Activity test of co-expressed RNAP$_C$. Multiple possible cRNAP$_C$-containing AEC-fractions were tested for transcription activity on gDNA template from *Geobacillus* sp. GHH01. Fraction 8 (b) is the most active fraction along with a less active fraction 7 (a). Activity was assayed by analyzing the produced RNA in 1.2% agarose gel containing 0.7 % (v/v) formaldehyde.

6 Discussion

In this study, functional search efforts of selected phosphotransferases from databases and environmental samples have been performed. Some of these target enzymes are rarely found in nature (or reported) and some others are difficult-to-express proteins. The search for phosphotransferases is based on considerations such as regio- and stereo-specificity, possible promiscuity on substrate analogs, and stability nature of extremozymes. In addition to successes in biocatalytic phosphorylation of the selected glycerol derivatives, functional search of phosphotransferases was largely challenged by protein expression problems which led to another effort aimed to establish host-independent transcription system for cell-free metagenomics platform and expression of difficult-to-express enzymes.

The focus on transcription is based on the transcriptome analyses of *E. coli* host carrying different metagenome fosmids revealing that it is, in fact, a decisive limiting step in functional metagenomics (Jürgensen, 2014, Liebl*, et al.*, 2014). This is because the host machinery is significantly biased where a decreased or absence of transcription of metagenomic DNA originated from phylogenetically distant organisms could be seen. It is also known that only about 40% of metagenome-derived genes in average can be readily expressed in *E. coli* (Gabor*, et al.*, 2004). This can further be explained in improved transcription of metagenome-derived genes in two different accounts when an extra sigma factor from another bacterium is integrated into *E. coli* (Gaida*, et al.*, 2015, Wofgang R. Streit, unpublished). This interestingly explains also how flexible natural systems can be and possibilities of adapting the transcription machineries and their components in an *in vitro* set up so that they can be used for cell-free applications bypassing host-related expression limitations.

6.1 Function-based approaches to search for β-glycerophosphate-involving activities

The high demand of ß-glycerophosphate, which is one of the main target phosphorylated metabolite products of this study, in diverse applications and the complete lack of enzymes for its biocatalytic synthesis have led to functional search efforts for ß-glycerophosphate forming enzyme via multiple approaches. In sequence databases, there is unsolved puzzle in the annotation of the *Cps* locus of *S. pneumoniae* known to include putative glycerol-2-phosphotransferase activity (Morona*, et al.*, 1999, Ramirez and Tomasz, 1998, Van Selm*, et al.*, 2003). Annotated differently in different serotypes of the bacterium, *gtp3* is claimed to be involved in the biosynthesis of CDP-2-glycerol which is the precursor required for addition of ß-glycerophosphate side chain (Wang*, et al.*, 2010). In this study, the gene was expressed and purified as His_6-tagged protein and was subject to activity assays under various reaction conditions although no detectable activity was recovered based on TLC assay. Even though the protein was claimed to be active in a previous experimental study, it was never fully characterized and is reported to be inactive without the presence of a coupling enzyme in the CDP-2-glycerol biosynthetic pathway (Wang*, et al.*, 2010).

In order to set further functional search alternatives, TLC-based activity detection method was established where samples were pre-treated with periodic acid for selective oxidation of compounds with vicinal hydroxyl groups such as glycerol and α-glycerophosphate but not the β-isomer (Leva and Rapoport, 1943). Considering the rare presence of ß-glycerophosphate in nature (or at least is seldom reported) and lack of described biosynthetic mechanisms, a top-down functional

search approach from environmental samples after enrichment on ß-glycerophosphate is pursued here.

In studies conducted to investigate alternative phosphorous sources in nature and their uptake, it is described that the highly regulated phosphate acting enzymes mainly phosphatases and phosphate transport systems are activated during phosphorus starvation (Matsuhisa*, et al.*, 1995, Torriani, 1960, Yang*, et al.*, 2009). The cellular uptake of ß-glycerophosphate is also described to be via Ugp transporter the same way as the α-isomer (Yang*, et al.*, 2009). When growth of *E. coli* EPI300 cells was tested on minimal medium supplemented with ß-glycerophosphate as a sole carbon source, no cell growth was recorded. Not surprisingly, there was normal cell growth when the α-glycerophosphate isomer was supplemented instead. Growth experiments on mixture of the two isomers did not show any shift or metabolic induction towards utilization of ß-glycerophosphate either.

Despite the rare metabolic presence reported, bacterial isolates able to utilize β-glycerophosphate as carbon source were isolated after enrichment of environmental samples. The depletion of the substrate was also verified using the newly established TLC-base assay. Fosmid libraries from selected isolates resulted eight positive hits which complemented the lacking ß-glycerophosphate utilization property in the initial screening although the activity could not be sufficiently established during further steps of re-complementation experiments after extraction of the fosmid DNAs. We believe this is most likely due to problems associated with expression bias, as it is also evidenced by multiple other studies involving functional meta/genomics approach (Uchiyama and Miyazaki, 2009).

It is widely described that expression problems are serious challenges for functional enzyme search efforts mainly due to difference in codon usage leading to biased transcription, poor protein expression, protein misfolding or lack of appropriate secretion systems (Ekkers*, et al.*, 2012, Kudla*, et al.*, 2009, Lam*, et al.*, 2015, Steele*, et al.*, 2009, Uchiyama and Miyazaki, 2009). In chapter 5 of this dissertation, a recombinant RNAP-based transcription system is established with an aim to set cell-free metagenomics platform which can potentially bypass host-related expression limitations during enzyme search in general and phosphotransferases in particular. Applying the use of such a system once fully established is, therefore, highly recommended. Alternatively, directed mutagenesis of existing enzymes which phosphorylate other substrate analogs such as glycerate-2-kinase at the desired carbon position can be a feasible alternative to modify the active site configuration that could fit the hydroxymethyl group of glycerol instead of the carboxyl group in glyceric acid.

6.2 D-glycerate-2-phosphate is sustainably synthesized using glycerate-2-kinase coupled with energy regeneration system

The initial reason for selecting glycerate-2-kinase from *T. maritima* was to investigate its activity towards other substrate analogs particularly to phosphorylate glycerol at its 2nd carbon position. Nonetheless, its use in the sustainable biocatalytic synthesis of D-glycerate-2-phosphate is discussed here. The over-expression of glycerate-2-kinase fused with maltose binding protein (MBP) yielded large amount of recombinant protein with high purity in *E. coli*. The activity of the kinase remains unaffected even in its fused state and the removal of MBP is not required after fusion cleavage as well. The stability of the enzyme is excellent which can also be suitable for slower reaction conversion processes depending on desired bioprocess mode.

The enzyme can easily be stored at 4°C for several months with 0.02% NaN_3 (w/v) without loss of its activity and at -20°C with 50% glycerol for long-term storage.

In addition to its highly thermostable property (Yang*, et al.*, 2008), the enzymes can maintain appreciable amount of its activity at 37°C which allows a direct process coupling with pyruvate kinase-based energy regeneration system. A complete substrate phosphorylation scheme was developed by simultaneously monitoring product formation and co-substrate phosphoenolpyruvate consumption using ^{31}P-NMR. The technique has enabled a simplified and effective way of monitoring several other phosphorylation reactions and analysis of phosphorylated biomolecules (Gauss*, et al.*, 2014, Matsumi*, et al.*, 2014). It is attributed to its suitability in analyzing especially metabolites with chiral properties. Accordingly, enzymatic phosphorylation using glycerate-2-kinase from *T. maritima* yielded an excellent enantiomeric purity of the D-glycerate-2-phosphate product which was further analyzed by capillary zone electrophoresis as well. The D- and L-glycerate-2-phosphate enantiomeric ratio was confirmed as 98.3:1.7 suggesting a highly pure product.

Sustainability in industrial production processes is one of the key features for an effective production process. This is usually challenged by the fast depletion of energy sources during the reaction and unable for the desired conversion to continue as the production scheme requires. Here, ATP regeneration coupled to the glycerate-2-kinase catalyzing phosphorylation enables a sustainable production of D-glycerate-2-phosphate which is effectively established at gram scale synthesis of the product.

6.3 Instability of glyceraldehyde-3-phosphate led to the search for kinases from extreme acidophilic Archaea

The synthesis of enantiopure D- and L-glyceraldehyde 3-phosphates has been much of biotechnology interest. Enzymatic syntheses approaches have been more complicated due to reasons of substrate incompatibility and product instability. The incompatibility of D-glyceraldehyde with the most common phosphoryl donor ATP increases the synthesis cost as only enantiopure L-glyceraldehyde substrate must be used for direct ATP-dependent enzymatic phosphorylation. This can potentially be addressed by examining alternative phosphoryl donors. The most decisive challenge is the hydrolysis of glyceraldehyde-3-phosphates above pH of 4.0, limiting the range of available enzymes that can potentially be used for biocatalytic phosphorylation of glyceraldehyde. The options here are either to search for acid-stable enzymes from extremophiles which can fit to reaction conditions for maximum product stability or to optimize parameters and process engineering with the currently known glycerol-3-kinase enzymes which are active largely around physiological pH.

There have been few successes in expression of extremozymes from acidophilic organisms which are able to grow as low as negative pHs. Examples of such enzymes include glycoamylases, α-glucosidase, α-mannosidase, tryhalose synthase enzymes stable up to pH of 1.4 from the extreme acidophilic sister species, *Picrophilus torridus* and *P. oshimae* (Angelov*, et al.*, 2006, Chen*, et al.*, 2006, Fütterer*, et al.*, 2004, Schepers*, et al.*, 2006, Serour and Antranikian, 2002). In the present study, putative glycerol-3-kinase coding genes from the two species were selected to be investigated for biocatalytic synthesis of glyceraldehyde 3-phosphate. The heterologous expression of Archaeal proteins, however, is often challenging. Several combinations of vectors and expression hosts (both bacteria and yeast), and

native gene sequence or optimized according to *E. coli* codon usage, resulted expression of proteins only as inclusion bodies. The effort to recover active proteins from the inclusion bodies yielded non-detectable activity. In addition, heterologous expression of another candidate gene from *P. torridus* for biocatalytic phosphorylation of metabolites, pyruvate kinase, did not yield any protein even as inclusion bodies. Therefore, using alternative cell-free expression systems suitable for expressing target protein-coding genes from different organisms can be a real potential. With the *Geobacillus* sp. GHH01 RNA polymerase, it was possible to generate transcript products from *P. torridus* genomic DNA directly.

Another option for phosphorylation of glyceraldehyde is developing kinetic models for existing enzymes that can suit to biosynthesis process requirements. Detailed process reaction engineering for the biocatalytic synthesis of enantiopure L-glyceraldehyde-3-phosphate using glycerol-3-kinase (E.C.2.7.1.30) from *Cellulomonas* sp. was investigated in consideration with the product's instability at the reaction conditions. Kinetics of the enzyme activity with respect to Mg^{2+} to ATP molar ratios, inhibitory properties of reaction component or co-product, simulation of different reactor types were deeply studied and the best possible reaction system is set for biocatalytic synthesis of L-glyceraldehyde-3-phosphate.

6.4 Both *in vitro* and *in vivo* reconstitution approaches resulted active recombinant RNA polymerase

With the ultimate objective of establishing cell-free metagenomics screening platform and expression scheme of difficult-to-express enzymes, this study sets an *in vitro* transcription system based on RNA polymerase from *Geobacillus* sp. GHH01. In order to get the RNA polymerase with its maximum possible transcription efficiency,

different strategies of expression of subunits and reconstitution were followed. All the RNAP reconstitution approaches from the His_6-tagged, untagged, and co-expressed subunits resulted active RNAP, despite with varying efficiency. The main reason of expressing tagged and tag-free proteins was to investigate the possible effect of His_6-tag on transcription activity of the polymerase. Transcription assays employing RNAPs reconstituted from exclusively tagged or tag-free proteins showed that the His_6 does not have a negative effect on transcription. The expression of subunits with His_6-tag has rather improved protein solubility and helped to get more subunit proteins with better quality and quantity with easier purification. This has led to the RNAP with best relative activity where most of the transcription assays in this study have employed it. On the contrary, the expression of untagged subunits generated both soluble and insoluble proteins, complicating protein purification, lower reconstitution efficiency, and therefore making this approach the least preferred.

Co-expressing all the five subunits together in a single vector and reconstitution *in vivo* has also yielded active RNAP. Ion-exchange chromatographic purification employing the His_6-tag sequence fussed with rpoC subunit yielded a mixture of both reconstituted RNAP and free β'-subunit, but largely dominated by the latter. A second purification step using size exclusion chromatography was employed to further purify the RNAP. While this approach is simpler in expression from five different subunits as target proteins in the previous strategies to just the *in vivo* reconstituted RNAP, it has greatly suffered by the diminishing RNAP concentration after subjected to consecutive purification steps.

Among the advantages of the *Geobacillus* RNAP is that the subunits, despite having varying temperature stability profiles, can be reconstituted into active RNAP

after a brief treatment of their mixture at a maximum nondenaturing temperature. This has shortened the lengthy classical reconstitution process from more than 17 hours of dialysis and multiple buffer exchange to just 15 minutes incubation saving both time and resources, for example the use of large amount of urea is avoided. The new method, unlike the denaturation and subsequent renaturation, has an exclusive requirement of the subunit proteins to be in their soluble form.

Other transcription components such as transcription elongation factors are known to play significant role during the process of transcription (Borukhov*, et al.*, 2005). Based on assessment of their *in vivo* role towards maximizing transcription rate, GreA and MfD factors were chosen in this work to investigate their role in *in vitro* setups. Accordingly, the addition of GreA has significantly improved the transcript yield generated while MfD did not have the same effect. The positive effect of GreA may possibly be due to its role in encouraging the polymerase to actively involve in transcription by resolving stalling therefore increasing the recycling rate (Kusuya*, et al.*, 2011, Stepanova*, et al.*, 2007). The range of transcript was also widened from mostly larger transcript sizes in the absence of elongation factors to also smaller up to 1kb fragments with the addition of GreA.

6.5 *Geobacillus* RNAP has remarkable working and storage stability profiles

One of the impressive features of the *Geobacillus* RNAP is its handling and working stability. Attributed to its moderately thermostable source organism, the RNA polymerase has comparative advantage over the other tested RNA polymerases maintaining its maximum activity at elevated temperatures up to 55°C. It was also interesting to notice that the reconstituted RNAP having better thermostability than the least thermostable of all the subunits, i.e. the β-subunit. The general stability

feature of the *Geobacillus* RNAP can have multiple advantages. First, transcription reactions and subsequently translation can take place at varying temperatures with in its wide stability range depending on the nature of target protein, screening method applied, background activity interference, and possible combinations with other transcription systems. Second, the RNAP has easy handling features with little to worry about activity loss during working conditions. Third, it will be a superior candidate if continuous scheme of protein synthesis over longer period of time is sought. Last, the RNAP can be stored in non-freezing temperatures for reasonable amount of time, minimizing activity loss during freeze-and-thaw situations. In this study, the *Geobacillus* RNAP maintained its activity after storage at 4°C for several months.

6.6 *Geobacillus* RNAP transcribes metagenomic DNAs on innate promoters with a potential to be used for cell-free metagenomics platform

RNA polymerases derived from phages are known for is their strict stringency exclusively towards their own promoters in spite of their strong transcription profile. This lack of modularity limits their use in more complex and robust applications where broader transcription coverage of template DNAs with varying recognition factors is required.

Transcription analysis with the *Geobacillus* RNA polymerase was done employing individual genomic DNAs, a mixture of genomic DNAs as a mock metagenome, and real metagenome samples. *Geobacillus* RNAP was interestingly able to generate RNAs from DNA templates originated from ranges of organisms. Not surprisingly, the transcript amounts were variable according to the template DNA used where the most RNA was generated from its own *Geobacillus* sp. GHH01

genomic DNA while the least was from Archaeal DNA template. This feature of flexible transcription on templates of varying origin makes the *Geobacillus* RNAP a potential candidate for ranges of application, especially for the newly hypothesized cell-free metagenomics platform.

There are studies on recombinant RNA polymerases from other bacteria as well such as *E.coli*, *Thermus thermophilus*, *T. aquaticus*, *Deinococcus radiodurans*, *Mycobacterium tuberculosis*, *Bacillus subtilis* (e.g., Esyunina and Kulbachinskiy, 2015, Herrera-Asmat*, et al.*, 2017, Kuznedelov*, et al.*, 2003, Kuznedelov and Severinov, 2009, Vassylyev*, et al.*, 2007, Yang and Lewis, 2008). Most of these studies were aimed to understand the role of RNAP subunits and other transcription elements, optimizing RNAP preparation, mechanisms of transcription process, and structural elucidations of RNAP. This polymerase can also be used in simpler screening setups as well as for expression of difficult-to-express genes which have been challenging using the available expression hosts.

Stable and robust RNA polymerase/s able to generate transcripts from wide range of template DNAs with varying recognition factors can also be combined with advanced cell-like microfluidic compartmentalization and high throughput screening methods to establish cell-free functional metagenomics platform.

7 References

Abderhalden E & Eichwald E (1918) Synthese von optisch-aktiver Glyceryl-phosphorsäure. *Ber Dtsch Chem Ges* 51(2):1308-1312.

Acker MG & Auld DS (2014) Considerations for the design and reporting of enzyme assays in high-throughput screening applications. *Perspectives in Science* 1(1-6):56-73.

Andexer JN & Richter M (2015) Emerging enzymes for ATP regeneration in biocatalytic processes. *ChemBioChem* 16(3):380-386.

Angelov A, Putyrski M & Liebl W (2006) Molecular and biochemical characterization of alpha-glucosidase and alpha-mannosidase and their clustered genes from the thermoacidophilic archaeon Picrophilus torridus. *J Bacteriol* 188(20):7123-7131.

Asano Y, Mihara Y & Yamada H (1999) A novel selective nucleoside phosphorylating enzyme from Morganella morganii. *J Biosci Bioeng* 87(6):732-738.

Ballou CE & Fischer HOL (1954) A new synthesis of 2-phosphoryl-D-glyceric acid. *J Am Chem Soc* 76:3188-3193.

Ballou CE & Fischer HOL (1955) The Synthesis of D-Glyceraldehyde-3-phosphate. *J Am Chem Soc* 77(12):3329-3331.

Bimboim HC & Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research* 7(6):1513-1523.

Boehmwald F, Muñoz P, Flores P & Blamey JM (2016) Functional Screening for the Discovery of New Extremophilic Enzymes. *Biotechnology of Extremophiles:,* Grand Challenges in Biology and Biotechnology, ed Rampelotto PH (Springer International Publishing, Cham), pp 321-350.

Bohnke S & Perner M (2015) A function-based screen for seeking RubisCO active clones from metagenomes: novel enzymes influencing RubisCO activity. *ISME J* 9(3):735-745.

Borukhov S, Lee J & Laptenko O (2005) Bacterial transcription elongation factors: new insights into molecular mechanism of action. *Mol Microbiol* 55(5):1315-1324.

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72(1):248-254.

Bundy BC & Swartz JR (2011) Efficient disulfide bond formation in virus-like particles. *J Biotechnol* 154(4):230-239.

Carlson ED, Gan R, Hodgman CE & Jewett MC (2012) Cell-free protein synthesis: applications come of age. *Biotechnol Adv* 30(5):1185-1194.

Catherine C, Lee KH, Oh SJ & Kim DM (2013) Cell-free platforms for flexible expression and screening of enzymes. *Biotechnol Adv* 31(6):797-803.

Chaleckis R, Ebe M, Pluskal T, Murakami I, Kondoh H & Yanagida M (2014) Unexpected similarities between the Schizosaccharomyces and human blood metabolomes, and novel human metabolites. *Mol Biosyst* 10(10):2538-2551.

Chen YS, Lee GC & Shaw JF (2006) Gene cloning, expression, and biochemical characterization of a recombinant trehalose synthase from *Picrophilus torridus* in *Escherichia coli*. *J Agric Food Chem* 54(19):7098-7104.

Chong S (2014) Overview of cell-free protein synthesis: historic landmarks, commercial systems, and expanding applications. *Curr Protoc Mol Biol* 108:16 30 11-11.

Chow J, Kovacic F, Dall Antonia Y, Krauss U, Fersini F, Schmeisser C, Lauinger B, Bongen P, Pietruszka J, Schmidt M, Menyes I, Bornscheuer UT, Eckstein M, Thum O, Liese A, Mueller-Dieckmann J, Jaeger KE & Streit WR (2012) The metagenome-derived enzymes LipS and LipT increase the diversity of known lipases. *PLoS One* 7(10):e47665.

Colin PY, Kintses B, Gielen F, Miton CM, Fischer G, Mohamed MF, Hyvonen M, Morgavi DP, Janssen DB & Hollfelder F (2015) Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. *Nat Commun* 6:10008.

Couble ML, Farges JC, Bleicher F, Perrat-Mabillon B, Boudeulle M & Magloire H (2000) Odontoblast differentiation of human dental pulp cells in explant cultures. *Calcif Tissue Int* 66(2):129-138.

Cowan DA (2000) Microbial genomes – the untapped resource. *Trends Biotechnol* 18(1):14-16.

Craig JW, Chang FY, Kim JH, Obiajulu SC & Brady SF (2010) Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. *Appl Environ Microbiol* 76(5):1633-1641.

Crans DC, Kazlauskas RJ, Hirschbein BL, Wong CH, Abril O & Whitesides GM (1987) Enzymatic regeneration of adenosine 5'-triphosphate: acetyl phosphate, phosphoenolpyruvate, methoxycarbonyl phosphate, dihydroxyacetone phosphate, 5-phospho-alpha-D-ribosyl pyrophosphate, uridine-5'-diphosphoglucose. *Methods Enzymol* 136:263-280.

Crans DC & Whitesides GM (1985a) Glycerol kinase: substrate specificity. *J Am Chem Soc* 107(24):7008-7018.

Crans DC & Whitesides GM (1985b) Glycerol kinase: synthesis of dihydroxyacetone phosphate, sn-glycerol-3-phosphate, and chiral analogs. *J Am Chem Soc* 107(24):7019-7027.

Culligan EP, Sleator RD, Marchesi JR & Hill C (2014) Metagenomics and novel gene discovery: Promise and potential for novel therapeutics. *Virulence* 5(3):399-412.

DeCastro ME, Rodriguez-Belmonte E & Gonzalez-Siso MI (2016) Metagenomics of thermophiles with a focus on discovery of novel thermozymes. *Front Microbiol* 7(1521):1521.

Ding G-J, Zhu Y-J, Qi C, Sun T-W, Wu J & Chen F (2015) Amorphous calcium phosphate nanowires prepared using beta-glycerophosphate disodium salt as an organic phosphate source by a microwave-assisted hydrothermal method and adsorption of heavy metals in water treatment. *RSC Adv* 5(50):40154-40162.

Ding K, Yang Z, Zhang YL & Xu JZ (2013) Injectable thermosensitive chitosan/beta-glycerophosphate/collagen hydrogel maintains the plasticity of skeletal muscle satellite cells and supports their *in vivo* viability. *Cell Biol Int* 37(9):977-987.

Ekkers DM, Cretoiu MS, Kielak AM & Elsas JD (2012) The great screen anomaly--a new frontier in product discovery through functional metagenomics. *Appl Microbiol Biotechnol* 93(3):1005-1020.

Esyunina DM & Kulbachinskiy AV (2015) Purification and characterization of recombinant *Deinococcus radiodurans* RNA Polymerase. *Biochemistry (Mosc)* 80(10):1271-1279.

Faber K (2011) Biocatalytic Applications. *Biotransformations in Organic Chemistry*, (Springer, Berlin, Heidelberg), pp 31-313.

Ferrer M, Martinez-Martinez M, Bargiela R, Streit WR, Golyshina OV & Golyshin PN (2016) Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microb Biotechnol* 9(1):22-34.

Fischer HOL & Baer E (1941) Preparation and Properties of Optically Active Derivatives of Glycerol. *Chem Rev* 29(2):287-316.

Fujimoto A & Smith RA (1960) Enzymatic phosphorylation of glucose by phosphoramidates. *J Biol Chem* 235(10):PC44-45.

Fujiwara K, Sawamura T, Niwa T, Deyama T, Nomura SM, Taguchi H & Doi N (2017) *In vitro* transcription-translation using bacterial genome as a template to reconstitute intracellular profile. *Nucleic Acids Res* 45(19):11449-11458.

Fütterer O, Angelov A, Liesegang H, Gottschalk G, Schleper C, Schepers B, Dock C, Antranikian G & Liebl W (2004) Genome sequence of *Picrophilus torridus* and its implications for life around pH 0. *Proc Natl Acad Sci USA* 101(24):9091-9096.

Gabor EM, Alkema WB & Janssen DB (2004) Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ Microbiol* 6(9):879-886.

Gaida SM, Sandoval NR, Nicolaou SA, Chen Y, Venkataramanan KP & Papoutsakis ET (2015) Expression of heterologous sigma factors enables functional screening of metagenomic and heterologous genomic libraries. *Nat Commun* 6:7045.

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 Synth Biol* 5(4):344-355.

Gauss D, Schoenenberger B & Wohlgemuth R (2014) Chemical and enzymatic methodologies for the synthesis of enantiomerically pure glyceraldehyde 3-phosphates. *Carbohydr Res* 389:18-24.

Gauss D, Schönenberger B, Molla GS, Kinfu BM, Chow J, Liese A, Streit WR & Wohlgemuth R (2016) Biocatalytic phosphorylation of metabolites. *Applied biocatalysis: from fundamental science to industrial applications*, eds Hilterhaus L, Liese A, Kettling U & Antranikian G (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany), pp 147-177.

Graham JE, Clark ME, Nadler DC, Huffer S, Chokhawala HA, Rowland SE, Blanch HW, Clark DS & Robb FT (2011) Identification and characterization of a multidomain hyperthermophilic cellulase from an archaeal enrichment. *Nat Commun* 2:375.

Gray GR & Barker R (1971) The synthesis of d-glyceraldehyde-3,3-d_2 3-phosphate. *Carbohydr Res* 20(1):31-37.

Handelsman J (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* 68(4):669-685.

Hartman FC & Wold F (1967) A new chemical synthesis of D-tartronic acid semialdehyde phosphate and D-glyceric acid 2-phosphate. *Biochim Biophys Acta* 141(3):445-453.

Herrera-Asmat O, Lubkowska L, Kashlev M, Bustamante CJ, Guerra DG & Kireeva ML (2017) Production and characterization of a highly pure RNA polymerase holoenzyme from Mycobacterium tuberculosis. *Protein Expr Purif* 134:1-10.

Hiles RA & Henderson LM (1972) The partial purification and properties of hydroxylysine kinase from rat liver. *J Biol Chem* 247(3):646-651.

Hirao I, Kanamori T & Ueda T (2009) Cell-free synthesis of proteins with unnatural amino acids. The PURE system and expansion of the genetic code. *Protein Engineering,* Nucleic Acids and Molecular Biology, eds Köhrer C & RajBhandary UL (Springer, Berlin, Heidelberg), pp 271-290.

Hiroyuki O, Takashi K & Yoshihiro Shimizu and Takuya U (2010) A highly controllable reconstituted cell-free system - a breakthrough in protein synthesis research. *Curr Pharm Biotechnol* 11(3):267-271.

Hodgman CE & Jewett MC (2012) Cell-free synthetic biology: thinking outside the cell. *Metab Eng* 14(3):261-269.

Iqbal HA, Feng Z & Brady SF (2012) Biocatalysts and small molecule products from metagenomic studies. *Curr Opin Chem Biol* 16(1-2):109-116.

Isaksson L, Enberg J, Neutze R, Goran Karlsson B & Pedersen A (2012) Expression screening of membrane proteins with cell-free protein synthesis. *Protein Expr Purif* 82(1):218-225.

Jacquiod S, Franqueville L, Cecillon S, Vogel TM & Simonet P (2013) Soil bacterial community shifts after chitin enrichment: an integrative metagenomic approach. *PLoS One* 8(11):e79699.

Jürgensen J (2014) Identification of an unusual GTase from a non-cultivated microorganism and the construction of an improved *E. coli* strain harboring the *rpoD* gene from *C. cellulolyticum* for metagenome searches. PhD dissertation (Universität Hamburg).

Kakirde KS, Wild J, Godiska R, Mead DA, Wiggins AG, Goodman RM, Szybalski W & Liles MR (2011) Gram negative shuttle BAC vector for heterologous expression of metagenomic libraries. *Gene* 475(2):57-62.

Kameda A, Shiba T, Kawazoe Y, Satoh Y, Ihara Y, Munekata M, Ishige K & Noguchi T (2001) A novel ATP regeneration system using polyphosphate-AMP phosphotransferase and polyphosphate kinase. *J Biosci Bioeng* 91(6):557-563.

Karig DK (2017) Cell-free synthetic biology for environmental sensing and remediation. *Curr Opin Biotechnol* 45:69-75.

Katzke N, Knapp A, Loeschcke A, Drepper T & Jaeger KE (2017) Novel tools for the functional expression of metagenomic DNA. *Methods Mol Biol* 1539:159-196.

Kim DM & Swartz JR (2001) Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis. *Biotechnol Bioeng* 74(4):309-316.

Kim HC & Kim DM (2009) Methods for energizing cell-free protein synthesis. *J Biosci Bioeng* 108(1):1-4.

Knietsch A, Waschkowitz T, Bowien S, Henne A & Daniel R (2003) Construction and screening of metagenomic libraries derived from enrichment cultures: generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*. *Appl Environ Microbiol* 69(3):1408-1416.

Knowles JR (1980) Enzyme-catalyzed phosphoryl transfer reactions. *Annu Rev Biochem* 49:877-919.

Körfer G, Pitzler C, Vojcic L, Martinez R & Schwaneberg U (2016) *In vitro* flow cytometry-based screening platform for cellulase engineering. *Sci Rep* 6:26128.

Krawiec K, Kierdaszuk B & Shugar D (2003) Inorganic tripolyphosphate (PPP(i)) as a phosphate donor for human deoxyribonucleoside kinases. *Biochem Biophys Res Commun* 301(1):192-197.

Kudla G, Murray AW, Tollervey D & Plotkin JB (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324(5924):255-258.

Kulkarni Y (2013) Process for the synthesis of beta glycerol phosphate. US patent (US 8,420,848 B2).

Kuruma Y & Ueda T (2015) The PURE system for the cell-free synthesis of membrane proteins. *Nat Protoc* 10(9):1328-1344.

Kusuya Y, Kurokawa K, Ishikawa S, Ogasawara N & Oshima T (2011) Transcription Factor GreA contributes to resolving promoter-proximal pausing of RNA Polymerase in *Bacillus subtilis* cells. *J Bacteriol* 193(12):3090-3099.

Kuznedelov K, Minakhin L & Severinov K (2003) Preparation and characterization of recombinant *Thermus aquaticus* RNA Polymerase. *Methods Enzymol* 370:94-108.

Kuznedelov K & Severinov K (2009) Recombinant bacterial RNA polymerase: preparation and applications. *Methods* 47(1):44-52.

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227(5259):680-685.

Lam KN, Cheng J, Engel K, Neufeld JD & Charles TC (2015) Current and future resources for functional metagenomics. *Front Microbiol* 6(1196):1196.

Leis B, Angelov A, Mientus M, Li HJ, Pham VTT, Lauinger B, Bongen P, Pietruszka J, Goncalves LG, Santos H & Liebl W (2015) Identification of novel esterase-active enzymes from hot environments by use of the host bacterium *Thermus thermophilus*. *Front Microbiol* 6:275.

Leva E & Rapoport S (1943) The determination of acid-soluble phosphoglycerol in liver. *J Biol Chem* 149:47-55.

Lewin A, Lale R & Wentzel A (2017) Expression platforms for functional metagenomics: emerging technology options beyond *Escherichia coli*. *Functional metagenomics: tools and applications*, eds Charles TC, Liles MR & Sessitsch A (Springer International Publishing, Cham), pp 13-44.

Li T, Yeh H, Kim JH & Ryu DD (2001) Performance of batch membrane reactor: Glycerol-3-phosphate synthesis coupled with adenosine triphosphate regeneration. *Biotechnol Bioeng* 74(4):326-334.

Li Y, Wexler M, Richardson DJ, Bond PL & Johnston AW (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. *Environ Microbiol* 7(12):1927-1936.

Liebl W, Angelov A, Juergensen J, Chow J, Loeschcke A, Drepper T, Classen T, Pietruszka J, Ehrenreich A, Streit WR & Jaeger KE (2014) Alternative hosts for functional (meta)genome analysis. *Appl Microbiol Biotechnol* 98(19):8099-8109.

Ling K-H & Lardy HA (1954) Uridine- and inosine-triphosphates as phosphate donors for phosphohexokinase. *J Am Chem Soc* 76(10):2842-2843.

Lu Y (2017) Cell-free synthetic biology: Engineering in an open world. *Synth Syst Biotechnol* 2(1):23-27.

Maroti G, Tong Y, Yooseph S, Baden-Tillson H, Smith HO, Kovacs KL, Frazier M, Venter JC & Xu Q (2009) Discovery of [NiFe] hydrogenase genes in metagenomic DNA: cloning and heterologous expression in Thiocapsa roseopersicina. *Appl Environ Microbiol* 75(18):5821-5830.

Martinez A, Kolvek SJ, Yip CL, Hopke J, Brown KA, MacNeil IA & Osburne MS (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. *Appl Environ Microbiol* 70(4):2452-2463.

Matsuhisa A, Suzuki N, Noda T & Shiba K (1995) Inositol monophosphatase activity from the *Escherichia coli suhB* gene product. *J Bacteriol* 177(1):200-205.

Matsumi R, Hellriegel C, Schoenenberger B, Milesi T, van der Oost J & Wohlgemuth R (2014) Biocatalytic asymmetric phosphorylation of mevalonate. *Rsc Advances* 4(25):12989-12994.

Morona JK, Miller DC, Coffey TJ, Vindurampulle CJ, Spratt BG, Morona R & Paton JC (1999) Molecular and genetic characterization of the capsule biosynthesis locus of *Streptococcus pneumoniae* type 23F. *Microbiology* 145(4):781-789.

Nakamichi Y, Yoshioka A, Kawai S & Murata K (2013) Conferring the ability to utilize inorganic polyphosphate on ATP-specific NAD kinase. *Sci Rep* 3:2632.

Novak-Hofer I & Thomas G (1985) Epidermal growth factor-mediated activation of an S6 kinase in Swiss mouse 3T3 cells. *J Biol Chem* 260(18):10314-10319.

Pardee K, Slomovic S, Nguyen PQ, Lee JW, Donghia N, Burrill D, Ferrante T, McSorley FR, Furuta Y, Vernet A, Lewandowski M, Boddy CN, Joshi NS & Collins JJ (2016) Portable, on-demand biomolecular manufacturing. *Cell* 167(1):248-259 e212.

Park IS, Woo TG, Jeon WY, Park HH, Lee MH, Bae TS & Seol KW (2007) Surface characteristics of titanium anodized in the four different types of electrolyte. *Electrochimica Acta* 53(2):863-870.

Pfennig N & Lippert KD (1966) Über das Vitamin B12-Bedürfnis phototropher Schwefelbakterien. *Archiv für Mikrobiologie* 55(3):245-256.

Rabausch U, Juergensen J, Ilmberger N, Bohnke S, Fischer S, Schubach B, Schulte M & Streit WR (2013) Functional screening of metagenome and genome libraries for detection of novel flavonoid-modifying enzymes. *Appl Environ Microbiol* 79(15):4551-4563.

Ramirez M & Tomasz A (1998) Molecular characterization of the complete 23F capsular polysaccharide locus of *Streptococcus pneumoniae*. *J Bacteriol* 180(19):5273-5278.

Rosenblum G & Cooperman BS (2014) Engine out of the chassis: cell-free protein synthesis and its uses. *FEBS Lett* 588(2):261-268.

Rousset M, Casalot L, Rapp-Giles BJ, Dermoun Z, de Philip P, Belaich JP & Wall JD (1998) New shuttle vectors for the introduction of cloned DNA in *Desulfovibrio*. *Plasmid* 39(2):114-122.

Samazan F, Rokbi B, Seguin D, Telles F, Gautier V, Richarme G, Chevret D, Varela PF, Velours C & Poquet I (2015) Production, secretion and purification of a correctly folded staphylococcal antigen in *Lactococcus lactis*. *Microb Cell Fact* 14(1):104.

Sambrook J (2001) Molecular cloning: a laboratory manual. ed Russell DW (Cold Spring Harbor Laboratory Press, N.Y.).

Sanchez-Moreno I, Bordes I, Castillo R, Ruiz-Pernia JJ, Moliner V & Garcia-Junceda E (2015) Tuning the phosphoryl donor specificity of dihydroxyacetone kinase from ATP to inorganic polyphosphate. An insight from computational studies. *Int J Mol Sci* 16(11):27835-27849.

Santana-Pereira ALR & Liles MR (2017) Challenges and opportunities in discovery of secondary metabolites using a functional metagenomic approach. in *Functional*

Metagenomics: Tools and Applications, eds Charles TC, Liles MR & Sessitsch A (Springer International Publishing, Cham), pp 119-138.

Schallmey M, Ly A, Wang C, Meglei G, Voget S, Streit WR, Driscoll BT & Charles TC (2011) Harvesting of novel polyhydroxyalkanaote (PHA) synthase encoding genes from a soil metagenome library using phenotypic screening. *FEMS Microbiol Lett* 321(2):150-156.

Schepers B, Thiemann V & Antranikian G (2006) Characterization of a novel glucoamylase from the Thermoacidophilic Archaeon *Picrophilus torridus* heterologously expressed in *E. coli*. *Eng Life Sci* 6(3):311-317.

Schinn SM, Broadbent A, Bradley WT & Bundy BC (2016) Protein synthesis directly from PCR: progress and applications of cell-free protein synthesis with linear DNA. *N Biotechnol* 33(4):480-487.

Schoenenberger B, Wszolek A, Milesi T, Brundiek H, Obkircher M & Wohlgemuth R (2017) Synthesis of N_{ω}-phospho-$_{L}$-arginine by biocatalytic phosphorylation of $_{L}$-arginine. *ChemCatChem* 9(1):121-126.

Serianni AS, Pierce J & Barker R (1979) Carbon-13-enriched carbohydrates: preparation of triose, tetrose, and pentose phosphates. *Biochemistry* 18(7):1192-1199.

Serour E & Antranikian G (2002) Novel thermoactive glucoamylases from the thermoacidophilic Archaea *Thermoplasma acidophilum, Picrophilus torridus* and *Picrophilus oshimae*. *Antonie Leeuwenhoek* 81(1-4):73-83.

Shiba T, Tsutsumi K, Ishige K & Noguchi T (2000) Inorganic polyphosphate and polyphosphate kinase: their novel biological functions and applications. *Biochemistry (Mosc)* 65(3):315-323.

Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K & Ueda T (2001) Cell-free translation reconstituted with purified components. *Nat Biotechnol* 19(8):751-755.

Shin J & Noireaux V (2010) Efficient cell-free expression with the endogenous *E. coli* RNA polymerase and sigma factor 70. *J Biol Eng* 4(1):8.

Slonczewski JL, Fujisawa M, Dopson M & Krulwich TA (2009) Cytoplasmic pH measurement and homeostasis in bacteria and archaea. *Adv Microb Physiol* 55:1-79, 317.

Srivastava A, Asahara H, Zhang M, Zhang W, Liu H, Cui S, Jin Q & Chong S (2016) Reconstitution of protein translation of *Mycobacterium* reveals functional conservation and divergence with the gram-negative bacterium *Escherichia coli*. *PLoS One* 11(8):e0162020.

Staden R, Judge DP & Bonfield JK (2003) Analyzing Sequences Using the Staden Package and EMBOSS. *Introduction to Bioinformatics*, eds Krawetz SA & Womble DD (Humana Press, Totowa, NJ), pp 393-410.

Steele HL, Jaeger KE, Daniel R & Streit WR (2009) Advances in recovery of novel biocatalysts from metagenomes. *J Mol Microbiol Biotechnol* 16(1-2):25-37.

Stepanova E, Lee J, Ozerova M, Semenova E, Datsenko K, Wanner BL, Severinov K & Borukhov S (2007) Analysis of promoter targets for *Escherichia coli* transcription elongation factor GreA *in vivo* and *in vitro*. *J Bacteriol* 189(24):8772-8785.

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

Streit WR & Schmitz RA (2004) Metagenomics – the key to the uncultured microbes. *Curr Opin Microbiol* 7(5):492-498.

Surmenev RA, Surmeneva MA & Ivanova AA (2014) Significance of calcium phosphate coatings for the enhancement of new bone osteogenesis-a review. *Acta Biomater* 10(2):557-579.

Szymanska E, Szekalska M, Czarnomysy R, Lavric Z, Srcic S, Miltyk W & Winnicka K (2016) Novel spray dried glycerol-2-phosphate cross-linked chitosan microparticulate vaginal delivery system—development, characterization and cytotoxicity studies. *Mar Drugs* 14(10):174.

Terzaghi BE & Sandine WE (1975) Improved medium for Lactic Streptococci and their Bacteriophages. *Appl Microbiol* 29(6):807-813.

Thoring L, Wustenhagen DA, Borowiak M, Stech M, Sonnabend A & Kubick S (2016) Cell-free systems based on CHO cell lysates: optimization strategies, synthesis of "difficult-to-express" proteins and future perspectives. *PLoS One* 11(9):e0163670.

Torriani A (1960) Influence of inorganic phosphate in the formation of phosphatases by *Escherichia coli*. *Biochim Biophys Acta* 38:460-469.

Towbin H, Staehelin T & Gordon J (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci U S A* 76(9):4350-4354.

Tuckey C, Asahara H, Zhou Y & Chong S (2014) Protein synthesis using a reconstituted cell-free system. *Curr Protoc Mol biol* 108:16.31.11-16.31.22.

Uchiyama T, Abe T, Ikemura T & Watanabe K (2005) Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nat Biotechnol* 23(1):88-93.

Uchiyama T & Miyazaki K (2009) Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr Opin Biotechnol* 20(6):616-622.

Uchiyama T & Miyazaki K (2010) Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes. *Appl Environ Microbiol* 76(21):7029-7035.

Van Selm S, van Cann LM, Kolkman MAB, van der Zeijst BAM & van Putten JPM (2003) Genetic basis for the structural difference between *Streptococcus pneumoniae* serotype 15B and 15C capsular polysaccharides. *Infect Immun* 71(11):6192-6198.

Vassylyev DG, Vassylyeva MN, Perederina A, Tahirov TH & Artsimovitch I (2007) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* 448:157.

Villarreal F & Tan CM (2017) Cell-free systems in the new age of synthetic biology. *Frontiers of Chemical Science and Engineering* 11(1):58-65.

Wang C, Meek DJ, Panchal P, Boruvka N, Archibald FS, Driscoll BT & Charles TC (2006) Isolation of poly-3-hydroxybutyrate metabolism genes from complex microbial communities by phenotypic complementation of bacterial mutants. *Appl Environ Microbiol* 72(1):384-391.

Wang J, Zheng C, Zhang T, Liu Y, Cheng Z, Liu D, Ying H & Niu H (2017) Novel one-pot ATP regeneration system based on three-enzyme cascade for industrial CTP production. *Biotechnol Lett* 39(12):1875-1881.

Wang Q, Xu Y, Perepelov AV, Xiong W, Wei D, Shashkov AS, Knirel YA, Feng L & Wang L (2010) Characterization of the CDP-2-glycerol biosynthetic pathway in *Streptococcus pneumoniae*. *J Bacteriol* 192(20):5506-5514.

Westheimer FH (1987) Why nature chose phosphates. *Science* 235(4793):1173-1178.

Wexler M, Bond PL, Richardson DJ & Johnston AW (2005) A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase. *Environ Microbiol* 7(12):1917-1926.

Whittaker JW (2013) Cell-free protein synthesis: the state of the art. *Biotechnol Lett* 35(2):143-152.

Wiegand S, Rabausch U, Chow J, Daniel R, Streit WR & Liesegang H (2013) Complete genome sequence of *Geobacillus* sp. strain GHH01, a thermophilic lipase-secreting bacterium. *Genome Announc* 1(2):e0009213.

Wilson K (2001) Preparation of Genomic DNA from Bacteria. *Current Protocols in Molecular Biology*, (John Wiley & Sons, Inc.), pp 00:I:02.04:02.04.01–02.04.05.

Wilson MC & Piel J (2013) Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. *Chem Biol* 20(5):636-647.

Wohlgemuth R (2009) Tools and ingredients for the biocatalytic synthesis of metabolites. *Biotechnol J* 4(9):1253-1265.

Wohlgemuth R, Liese A & Streit W (2017) Biocatalytic phosphorylations of metabolites: past, present, and future. *Trends Biotechnol* 35(5):452-465.

Yang C, Rodionov DA, Rodionova IA, Li X & Osterman AL (2008) Glycerate 2-kinase of *Thermotoga maritima* and genomic reconstruction of related metabolic pathways. *J Bacteriol* 190(5):1773-1782.

Yang K, Wang M & Metcalf WW (2009) Uptake of glycerol-2-phosphate via the *ugp*-encoded transporter in *Escherichia coli* K-12. *J Bacteriol* 191(14):4667-4670.

Yang X & Lewis PJ (2008) Overproduction and purification of recombinant *Bacillus subtilis* RNA polymerase. *Protein Expr Purif* 59(1):86-93.

Zemella A, Thoring L, Hoffmeister C & Kubick S (2015) Cell-free protein synthesis: pros and cons of prokaryotic and eukaryotic systems. *ChemBioChem* 16(17):2420-2431.

Zhou Y, Asahara H, Gaucher EA & Chong S (2012) Reconstitution of translation from *Thermus thermophilus* reveals a minimal set of components sufficient for protein synthesis at high temperatures and functional conservation of modern and ancient translation components. *Nucleic Acids Res* 40(16):7932-7945.

Acknowledgements

My deepest gratitude to Prof. Dr. Wolfgang Streit for the wonderful opportunity he provided me to join his group and for his invaluable guidance throughout the study period. I am truly indebted to his positivity and thoughtful advice that gave me the strength to withstand the challenges especially during the early phase of this work.

I extend my heartfelt gratitude to Dr. Jennifer Chow for her unreserved support from day one up until proof-reading of my dissertation. Thank you for patiently listening to my concerns and for extending your support whenever I needed. I am also indebted to Dr. Maike Jahnke for her brilliant support in the *in vitro* polymerase work.

I would like to thank project partners Dr. Roland Wohlgemuth from Sigma Aldrich, Prof. Dr. Andreas Liese and Dr. Getachew S. Molla from Technology University of Hamburg, Prof. Dr. Ulrich Schwaneberg and his team from RWTH Aachen University, and Dr. Martin Borschert and Dr. Michael Roggenbuck from Novozymes for the nice cooperation and invaluable discussions.

I am grateful for the German Academic Exchange Service (DAAD) for financing majority of my stay and the incredible support from its staff. I am also thankful to the German Ministry of Education and Research (BMBF) for the additional financing.

My warmest appreciation for former and current members and students of Microbiology and Biotechnology department for the pleasant work environment, helpful discussions, and all the fun we had in and outside of the lab.

I thank my friends and family whose gracious presence in my life made this possible. Words cannot describe my gratitude to my father and my late mother for their endless love, encouragement, and sacrifices. To my dear wife Amelework and my son Azariah, I cannot thank you enough for your patience, understanding, and support in every way imaginable.

ተመስገ....ን!