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


- Planned and co-designed experiments, executed experimental work of His6-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


- 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


- 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


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

**Book chapter (contributed)**

# Table of contents

Abstract .......................................................................................................................................................... vii

1 Introduction ............................................................................................................................................... 1

1.1 Enzyme search and biocatalytic metabolite phosphorylation .............................................................. 1

1.2 Phosphoryl donors for industrial biocatalytic conversions ..................................................................... 3

1.3 Functional metagenomics and mining for biocatalysts ............................................................................ 5

1.4 Challenges and prospects of function-based metagenomics ................................................................... 6

1.5 Cell-free protein expression and its potential for functional screening ............................................... 10

1.6 Aims of the study ................................................................................................................................... 13

2 Establishing selective screening method and functional search for β-glycerophosphate-involving activities ................................................................................................................................. 14

3 Biocatalytic asymmetric phosphorylation catalyzed by recombinant glycerate-2-kinase ............................... 31

4 Bioreaction engineering leading to efficient synthesis of L-glyceraldehyde-3-phosphate ......................... 48

5 Recombinant RNA Polymerase from *Geobacillus* sp. GHH01 as tool for rapid generation of metagenomic RNAs using *in vitro* technologies ............................................................................. 58

6 Discussion ................................................................................................................................................ 77

6.1 Function-based approaches to search for β-glycerophosphate-involving activities ............................... 78

6.2 D-glycerate-2-phosphate is sustainably synthesized using glycerate-2-kinase coupled with energy regeneration system .................................................................................................................. 80

6.3 Instability of glyceraldehyde-3-phosphate led to the search for kinases from extreme acidophilic Archaea ................................................................................................................................. 82

6.4 Both *in vitro* and *in vivo* reconstitution approaches resulted active recombinant RNA polymerase .................................................................................................................................................. 83

6.5 *Geobacillus* RNAP has remarkable working and storage stability profiles .... 85

6.6 *Geobacillus* RNAP transcribes metagenomic DNAs on innate promoters with a potential to be used for cell-free metagenomics platform ........................................................................ 86

7 References ................................................................................................................................................ 88

Acknowledgements ....................................................................................................................................... 103
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-phosphotransferase activity was challenged, functional search of β-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-
Abstract

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 \textit{in vitro} transcription system based on a recombinant RNA polymerase (RNAP) from \textit{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 \textit{Geobacillus} RNAP has a significant comparative advantage over the tested commercial \textit{Escherichia coli} RNAP in terms of both the transcript generated and other important features. The \textit{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,

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).
et al., 2013, Knietisch, 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*,
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


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 \textit{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 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. Kinfu¹, Jennifer Chow¹, Getachew S. Molla², Andreas Liese²,
Roland Wohlgemuth³ and Wolfgang R. Streit*¹

¹Department of Microbiology and Biotechnology, Biocenter Klein Flottbek, University of Hamburg, Ohnhorststr. 18, D-22609 Hamburg, Germany; ²Institute of Technical Biocatalysis, Hamburg University of Technology, Hamburg, Germany; ³Sigma-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
Establishing selective screening method and functional search of β-glycerophosphate ... 

Introduction


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 β-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
Establishing selective screening method and functional search of β-glycerophosphate 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 β-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 β-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.

<table>
<thead>
<tr>
<th>Strains &amp; isolates</th>
<th>Features</th>
<th>Reference/Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Streptococcus pneumonia</em> 23F</td>
<td>Wild-type strain</td>
<td>(Ramirez and Tomasz, 1998); (Wang, <em>et al.</em>, 2010)</td>
</tr>
<tr>
<td><em>E. coli</em> EPI300</td>
<td>F−, mcrA Δ(mrr-hsdRMS-mcrBC), Φ80lacZΔM15, ΔlacX74, recA1, endA1, araD139, Δ(ara, leu)7697, galU, galK, λ−, rpsL, nupG, trfA, tonA, dhfr</td>
<td>Epicentre (Madion, WI, USA)</td>
</tr>
<tr>
<td><em>E. coli</em> DH5α</td>
<td>F− Q80 lacZΔM15 Δ(lacZYA-argF)U169 recA1 endA1 hsdR17(rk−, mk+) phoA supE44 λ− thi1 gyrA96 relA1</td>
<td>Life Technologies (Frankfurt, Germany)</td>
</tr>
<tr>
<td><em>E. coli</em> BL21(DE3)</td>
<td>F−ompT hsdS6( rB− mB−) gal dcm (DE3)</td>
<td>Novagen/Merck (Darmstadt, Germany)</td>
</tr>
<tr>
<td>Four different isolates</td>
<td>Isolated from enrichment cultures and able to utilize β-glycerophosphate as C-source</td>
<td>This work</td>
</tr>
</tbody>
</table>
Table 2: Plasmids and constructs.

<table>
<thead>
<tr>
<th>Plasmids and constructs</th>
<th>Traits</th>
<th>Reference/Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pDrive</td>
<td>3.85 kb TA-cloning vector, oriEc, Plac/lacZ, AmpR, KanR, T7-promotor</td>
<td>QiAGEN (Hilden, Germany)</td>
</tr>
<tr>
<td>pET-28a(+)</td>
<td>5.37 kb expression vector, lacI, KanR, T7-promotor, N- and C-terminal His6-tag coding sequence</td>
<td>Novagen/Merck (Darmstadt, Germany)</td>
</tr>
<tr>
<td>pCC1FOS™</td>
<td>8.139 kb, ChlR, redF, oriV, ori2, repE, parA, parB, parC, cos, loxP, lacZ, T7-promoter</td>
<td>Epicentre (Madison, WI, USA)</td>
</tr>
<tr>
<td>pDrive::Gtp3</td>
<td>pDrive carrying gtp3 gene</td>
<td>This work</td>
</tr>
<tr>
<td>pET-28a::Gtp3</td>
<td>pET-28a (+) carrying gtp3 gene</td>
<td>This work</td>
</tr>
<tr>
<td>Fosmid clones</td>
<td>About 8000 fosmid clones constructed from enrichment isolates in PCC1FOS</td>
<td>This work</td>
</tr>
</tbody>
</table>

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 β-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 MgSO₄ (1 M), and 100 µl of CaCl₂ (1 M). CaCl₂ 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 (B₁₂), Leucine, yeast extract, and casamino acid.

M9 salt solution (5x) contains 85 g of Na₂HPO₄·2H₂O, 30 g of KH₂PO₄, 2.5 g of NaCl, 5 g of NH₄Cl dissolved in 1 L deionized water and autoclaved (Sambrook, 2001). Trace element stock solution (100X) contains 1 g of EDTA, 166 mg of FeCl₃·6H₂O, 16.8 mg of ZnCl₂, 2.6 mg of CuCl₂·2H₂O, 2 mg CoCl₂·2H₂O, 2 mg H₃BO₃, 0.32 mg MnCl₂·4H₂O (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. MgSO₄ and CaCl₂ were separately prepared in 1 M stock each and autoclaved. Vitamin stock solution (1000x) has components of 10 mg Ca-pantothenate, cyanocobalaminine (B₁₂), nicotinic acid, pyridoxal-HCl (B₆), 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

<table>
<thead>
<tr>
<th>Antibiotics/supplements</th>
<th>Solvent</th>
<th>Stock solution</th>
<th>Working concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chloramphenicol</td>
<td>EtOH</td>
<td>50 mg/ml</td>
<td>50 μg/ml</td>
</tr>
<tr>
<td>Kanamycin</td>
<td>H₂O</td>
<td>25 mg/ml</td>
<td>25 μg/ml</td>
</tr>
<tr>
<td>IPTG</td>
<td>H₂O</td>
<td>1 M</td>
<td>0.1 mM</td>
</tr>
</tbody>
</table>
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 \( \beta \)-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 \( \beta \)-glycerophosphate in liquid cultures was also checked using the TLC-based detection method.

Bacterial cultivation for \( \beta \)-glycerophosphate utilization experiments and fosmid library screening was performed in M9 medium supplemented with \( \beta \)-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
Establishing selective screening method and functional search of \( \beta \)-glycerophosphate …

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 \textit{gtp3} gene**

The glycerol-2-phosphotransferase gene was amplified from genomic DNA of \textit{S. pneumoniae} using \textit{Gtp3\_Ndel\_for} (CATATGAAATTGACAAATAGAGTTGAT) and \textit{Gtp3\_HindIII\_rev} (AAGCTTGACAATTCCTTTCCACATTTTC) primer pairs. The gene was subcloned into pET28a vector from pDrive::Gtp3 clone after digesting both with \textit{Ndel} and \textit{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, \textit{et al.}, 2003). Plasmid DNA construct with the right insert sequence was then heat-shock transformed into \textit{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
Establishing selective screening method and functional search of β-glycerophosphate …

four β-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 β-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 (OD600) 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$_2$PO$_4$·2H$_2$O 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
Establishing selective screening method and functional search of β-glycerophosphate…

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$_2$PO$_4$·2H$_2$O 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$_2$PO$_4$·2H$_2$O 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 β-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 β-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).
Establishing selective screening method and functional search of β-glycerophosphate …

Kinase activity assays were performed in 50 mM reaction buffer (HEPES, PIPES, tris-HCl, phosphate buffers were tested) containing 30mM glycerol, 15mM ATP, 10mM MgCl₂. The reaction mixtures were incubated at 37°C up to 1 hour monitoring possible formation of β-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 His₆-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 β-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 β-glycerophosphate from glycerol as substrate. Despite all the attempts, there was no detectable activity.
Establishing selective screening method and functional search of β-glycerophosphate employing the newly established TLC-based assay method (Fig. 1C). In addition to \textit{gtp3}, possible promiscuity of glycerate-2-kinase from \textit{Thermotoga maritima} was tested for the phosphorylation of glycerol to β-glycerophosphate under different parameters although no activity could be detected.

![Figure 1: Ion-exchange purification of protein and activity assay.](image)

(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 β-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 β- and α-glycerophosphates with close properties is required. In this study, TLC-based method for screening of β-glycerophosphate-involving activities was established. The α- and β-glycerophosphate isomers were separated from the rest of possible reaction components by employing either silica or cellulose plates and a mixture of methanol,
Establishing selective screening method and functional search of β-glycerophosphate …

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 β-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 β-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 β-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
Establishing selective screening method and functional search of β-glycerophosphate …

shown from 2\textsuperscript{nd} to 5\textsuperscript{th} 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 β-glycerophosphate applied onto each spot.

**Enrichment of environmental samples and construction of fosmid libraries**

The presence of β-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 β-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 β-glycerophosphate either. The cellular uptake of β-glycerophosphate is described to be via *ugp*-encoded transporter the same way to the α-isomer (Yang, *et al.*, 2009).
Establishing selective screening method and functional search of β-glycerophosphate …

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 β-glycerophosphate only (◊), 10 mM of α-glycerophosphate (□), α- and β-glycerophosphate mix each 5 mM (△), and α- and β-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 β-glycerophosphate as the only carbon source was done for functional searches. Interestingly, enrichment cultures resulted isolates which can utilize β-glycerophosphate as carbon source (Fig. 4A and B). TLC-base assay also verified that cell growth was concomitant to depletion of β-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.
Establishing selective screening method and functional search of β-glycerophosphate …

Figure 4: Enrichment of environmental samples using defined medium for β-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 β-glycerophosphate. (C) TLC analysis of β-glycerophosphate depletion during cell growth. Lanes 1 to 4 are β-glycerophosphate utilizing isolates. Lane 5 is for the same medium incubated with E. coli cells and lane 6 is the standard β-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 β-glycerophosphate utilizing environmental isolates were screened based on growth using minimal growth medium with β-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).
Establishing selective screening method and functional search of $\beta$-glycerophosphate …

**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 regiospecificity 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:

Biocatalytic Asymmetric Phosphorylation Catalyzed by Recombinant Glycerate-2-Kinase


The efficient synthesis of pure δ-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 31P NMR and showed excellent enantioselectivity towards phosphorylation of the δ-enantiomer of glyceric acid. This straightforward phosphorylation reaction and subsequent product isolation enabled the preparation of enantiomerically pure δ-glycerate-2-phosphate. This phosphorylation reaction, using recombinant glycerate-2-kinase, yielded δ-glycerate-2-phosphate in fewer reaction steps and with higher purity than chemical routes.

Introduction

δ-Glyceric acid 2-phosphate (Scheme 1) is a key metabolite in central carbon metabolism, glycolysis/glucoseogenesis, 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.

The glycolytic intermediate between δ-glycerate-3-phosphate and phosphoenolpyruvate was first isolated from yeast and characterized as a crystalline δ-glycerate-2-phosphate barium salt by Meyerhof and Kiesling.[7] An efficient, robust, and scalable route for the preparation of a soluble form of pure δ-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 δ-glycerate-2-phosphate was described by Ballou and Fischer.[8] 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 (δ-galactose) to the final product. Although a shorter synthesis starting from δ-glucoconactone was later described,[9] the bottleneck 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.[10] The enzymatic phosphorylation of glyceric acid is also of clinical relevance for patients with δ-glyceric aciduria, a rare inborn error of serine and fructose metabolism, as mutations in the GYVT7 gene have been revealed as the cause of δ-glycerate kinase deficiency and δ-glyceric aciduria.[10]

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 δ-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,[11] Escherichia coli,[11] Pyrococcus horikoshii,[11] Picrophilus torridus,[11] Thermoplasma acidophilum,[11] Sulfolobus tokodai,[11] and Hypomicrobium methylivorans,[11] which are able to catalyze the formation of glycerate-2-phosphate, were investigated, detailed enzymatic information on the key enzymes, glycerate-2-kinase, was reported for the glycerate kinase of the hyperthermophilic archaean Thermoproteus tenax.[11] From the comparison of the kinetic and biochemical properties of the characterized class II glycerate kinases of T. tenax, P. torridus, T. acidophilum, and
H. methyllovorans, the glycerate-2-kinase from Thermotoga maritima was selected due to its favorable biocatalytic properties and structure-function characterization.  

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 phosphorolyations 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 $\alpha$- and $\epsilon$-glycerate as substrate. Nearly 100% conversion of $\alpha$-glycerate to $\epsilon$-glycerate-2-phosphate was observed. With pure $\epsilon$-glycerate as substrate, no $\epsilon$-glycerate-2-phosphate was formed, whereas ox-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 $\alpha$-glyceric acid, $\epsilon$-glyceric acid, and $\epsilon$-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 $\epsilon$-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...
Biocatalytic asymmetric phosphorylation catalyzed by recombinant glycerate-2-kinase

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% NaN₃ (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 α-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 α-glyceric acid-2-phosphate. The stability of α-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 α-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 1H and 31P 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 α-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 α-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 by a 0/1, enantiomeric ratio of 98.3:1.7 for the product α-glyceric acid-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 31P NMR reaction monitoring,[10] as both product formation and PEP donor consumption could be measured simultaneously. From the time course of biocatalytic O-phosphorylation of α-glyceric acid in the 2-position, followed by 31P 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-
spending 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 glyceraldehyde-2-phosphate kinase was obtained. A $K_{\text{m}}$ value of $(0.13 \pm 0.03)$ mM to glyceraldehyde, a $K_{\text{cat}}$ value of $(5.4 \pm 0.7) \times 10^{-2}$ s$^{-1}$, and a $k_{\text{cat}}/K_{\text{m}}$ of $(4.0 \pm 0.05) \times 10^{3}$ M$^{-1}$ s$^{-1}$ was determined. The glyceraldehyde-2-phosphate catalyzed phosphorylation of the $\alpha$- and $\epsilon$-enantiomers of glyceraldehyde, as well as its racemic form, were directly investigated by NMR. These experiments clearly demonstrated the power and selectivity of biocatalytic phosphorylation. Only the $\alpha$-glyceraldehyde was accepted as a substrate for this recombinant glyceraldehyde-2-phosphate kinase, whereas the $\epsilon$-glyceraldehyde was not phosphorylated to any extent. This straightforward and highly selective phosphorylation reaction is therefore very valuable for the enzymatic resolution of racemic glyceraldehyde or for the synthesis of $\alpha$-glyceraldehyde-2-phosphate by phosphorylation of $\alpha$-glyceraldehyde in the 2-position.

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

The excellent enantioselectivity and the first crystal structure$^{[10]}$ of a representative of the glyceraldehyde-2-phosphate 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 $\alpha$-glyceraldehydes and functional isosteres 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 $\alpha$-glyceraldehyde and the elucidation of the mechanism of the enzyme function require further studies and a 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$_2$O at room temperature on a Bruker Avance III 600 MHz spectrometer equipped with a BBFO probe head with z-gradient by using 600.2 MHz for $^1$H and 150.9 MHz for $^{13}$C. The NMR kinetics of enzymatic phosphorylations of $\alpha$-glyceraldehyde, $\epsilon$-glyceraldehyde, and $\alpha$-glyceraldehyde was measured by acquisition every 15 min of $^1$H NMR and $^{31}$P NMR spectra. $T_1 = 0$ (red, bottom) was measured approxi-

mately 2 min after addition of the enzymes. Conversion was tracked by the decrease in the PEP signals ($^1$H: 5.0–5.4 ppm; $^{31}$P: −0.9 ppm) and the increase in product signals ($^1$H: 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. Uncoat-
ed 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 glyceraldehyde-2-phosphate gene:** The glyceraldehyde gene from T. maritima was amplified by PCR using primer pairs G2K-EcoRI_F (5’-gaatct AGTGTAT CCTGAATC CTG TGG and G2K-HindIII_R (5’-aagct CATAG GATGAC GCCTAT TAC) and cloned into the pDrive cloning vector. The gene was then excised from the plasmid withendonucleases EcoRI and HindIII and subcloned into pMAL-C2X in the same translational reading frame as the maltE 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’-TGATAC AAGGAG GCCAGT G), both binding within a vector sequence upstream (within the maltE 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 ampicillin. Three percent of the cells grown overnight were transferred into the same fresh medium, further incubated until the OD$_{540}$ reached 0.6, and induced with isopropyl $\beta$-1-thio-galacto-

pyranoside (IPTG, 0.3 mM) at 17°C for 18 h for maximum soluble protein expression. Cells were harvested by centrifugation (5000g, 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 Na$_2$ATP). A French press was used to break apart the cells and, after centrifugation (16000g, 60 min, 4°C), the supernatant was loaded onto a pre-equilibrated column packed with amylase 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 glyceraldehyde-2-phosphate protein, the MBP was cleaved from the fusion protein with factor Xa protease (New England Bio-

labs) overnight on ice. The protein was further concentrated and subsequently dialyzed on Vivaspin columns (Sartorius) and then quantified with Roti-Quant solution (Carl Roth), according to the manufacturer’s protocol.

**Kinetic measurements of glyceraldehyde-2-phosphate from T. maritima:**

Kinetic determinations of the glyceraldehyde-2-phosphate 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), $\alpha$-lactate dehydrogenase (1.5 U), and $\alpha$-glyceraldehyde 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 glyceraldehyde-2-phosphate (0.2 µg). Apparent values of $K_{\text{cat}}$ and $K_{\text{m}}$ 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.
Biocatalytic asymmetric phosphorylation catalyzed by recombinant glycerate-2-kinase

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

Enzyme reactions

Enzymatic syntheses of α,β-, β- and γ-glyceric acid 2-phosphate: αL-, β- or γ-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 D2O (1.7 mL) with MgCl2 (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. 1H NMR and 31P NMR spectra were measured every 15 min.

Gram-scale preparation of α-glycerate-2-phosphate catalyzed by recombinant glycerate-2-kinase: α-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 MgCl2 solution (10 mL, 200 mL). The pH was adjusted to 7.0, pyruvate kinase (1000 U in 1000 μL) and 110 μL glycerate-2-kinase (7.1 unit/mg) 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 10 000 MW Centrifloc filter membrane. Calcium acetate monohydrate (28.2 mmol 3 equiv) was added to the resulting reaction solution containing α-glycerate-2-phosphate (3.3 g) to precipitate the α-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 50WX 8H+ ion exchange resin (150 mL). The solution of α-glycerate-2-phosphate lithium salt was evaporated to give 1.255 g white solid product in 72% yield and excellent purity. 1H NMR (D2O, 600 MHz): δ = 4.48 (dd, J = 8.6, 5.4, 3.1 Hz, 1H), 3.89 (dd, J = 11.9, 3.1, 1.1H), 3.82 ppm (dd, J = 11.9, 5.4, 3.1 Hz, 1H); 13C NMR (D2O, 151 MHz): δ = 177.4 (d, J = 5.6 Hz), 76.7 (d, J = 5.5 Hz), 64.2 ppm (d, J = 3.6 Hz); 31P NMR (D2O, 162 MHz): δ = -2.2 ppm (s, 1P); MS: m/z calc for C4H9O7P2: 185.0 (M-H+), found: 185.0; capillary electrophoresis: enantiomeric ratio αL/αD = 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 · α-glycerate-2-phosphate · enzymatic phosphorylation · glycerate-2-kinase


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

Biocatalytic Asymmetric Phosphorylation Catalyzed by Recombinant Glycerate-2-Kinase


cbic_201700201_sm_misssellaneous_information.pdf
Biocatalytic asymmetric phosphorylation catalyzed by recombinant glycerate-2-kinase

Supporting Information

1) Capillary zone electrophoresis

**D-Glycerate-2-phosphate lithium salt**

**PRINZIP**

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

**GERAETEINSTELLUNGEN**

Detektor: PDA-Detektor
Detektionswellenlänge: 233 nm (indirekt)
Referenzwellenlänge: keine
Temperatur (Kapillare): 25 °C
Kapillare: unbeschichtete Kapillare,
75 um Innendurchmesser,
120 cm Laenge, 110 cm bis Detektor

**REAGENTZIEN**

- 95283 WASSER
- 72082 Natriumhydroxid-Loesung (1 M)
- 94747 Vancomycin HCl
- CExof(TM) anions 5 Kit (analis)

Anstelle der angegebenen Reagenzien durfen andere gleichwertige Reagenzien verwendet werden.

**LOESUNGEN**

a) Elektrophoresepuffer (EP): Accelerator aus CExof anions 5 Kit
b) Conditioner (Cond): Initiator aus CExof 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**

B1:
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**

(anil5_120.met)

Druck/Strom Zeit inlet outlet

Rinse-Pressure 20 psi 1 min B1:D1 BO:C1
Rinse-Pressure 20 psi 1 min B1:B1 BO:C1
Inject-Pressure 0.5 psi 5 sec SI:A1 BO:C1
Inject-Pressure 0.1 psi 10 sec SI:B2 BO:C1
Separate Volt. -30 kV 45 min B1:A1 BO:A1 1 min Ramp
Rinse-Pressure 40 psi 1.5 min B1:C1 BO:C1
Rinse-Pressure 40 psi 1.5 min B1:C2 BO:C1

**AUSWERTUNG**

Korrigierte Flächenprozente.
Hintergrundsignale nicht integrieren.
### PDA - 233nm Results

<table>
<thead>
<tr>
<th>Pk #</th>
<th>Migration Time</th>
<th>Area</th>
<th>Corrected Area</th>
<th>Corrected Area Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.567</td>
<td>8545</td>
<td>1483</td>
<td>2.646</td>
</tr>
<tr>
<td>2</td>
<td>11.913</td>
<td>547</td>
<td>84</td>
<td>0.150</td>
</tr>
<tr>
<td>3</td>
<td>17.925</td>
<td>528832</td>
<td>54088</td>
<td>96.534</td>
</tr>
<tr>
<td>4</td>
<td>23.271</td>
<td>4764</td>
<td>375</td>
<td>0.670</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td></td>
<td><strong>542688</strong></td>
<td><strong>56030</strong></td>
<td><strong>100.000</strong></td>
</tr>
</tbody>
</table>
Biocatalytic asymmetric phosphorylation catalyzed by recombinant glycerate-2-kinase

CEER

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

GERÄETEINSTELLUNGEN
Detektor: PDA-Detektor
Detektionswellenlänge: 233 nm (indirekt)
Referenzwellenlänge: keine
Temperatur (Kapillare): 25°C
Kapillare: unbeschichtete Kapillare,
75 µm Innendurchmesser,
120 cm Länge, 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.

LOFSUNGEN
a) Elektrophoresepuffer (EP): Accelerator aus CEOFix anions 5 Kit
b) Conditioner (Cond): Initiator aus CEOFix anions 5 Kit
c) Vancomycin-Lösung (Vanc): 60 mg Vancomycin in 1.0 mL EP lösen (~40 mM).
d) Probelsung: Wenige mg der Probe werden zu 5 mg/mL in Wasser gelöst.

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 Flächenprozente der Singale für D- und fuer L-Enantiomer
bestimmen:
M(t) D-2-Phosphoglycerinsäure ~22.3 min
M(t) L-2-Phosphoglycerinsäure ~23.3 min
Biocatalytic asymmetric phosphorylation catalyzed by recombinant glycerate-2-kinase

<table>
<thead>
<tr>
<th>PDA - 233nm Results</th>
<th>Pk #</th>
<th>Migration Time</th>
<th>Area</th>
<th>Corrected Area</th>
<th>Corrected Area Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>20.071</td>
<td>273719</td>
<td>25002</td>
<td>98.325</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>20.221</td>
<td>4699</td>
<td>426</td>
<td>1.675</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td></td>
<td>278418</td>
<td>25428</td>
<td>100.000</td>
</tr>
</tbody>
</table>
2) NMR

D-Glycerate-2-phosphate lithium salt

$^1$H-NMR
Biocatalytic asymmetric phosphorylation catalyzed by recombinant glycerate-2-kinase

$^{13}$C-NMR
Biocatalytic asymmetric phosphorylation catalyzed by recombinant glyceraldehyde-3-phosphate dehydrogenase

**FULL PAPER**

$^{31}$P-NMR
Biocatalytic asymmetric phosphorylation catalyzed by recombinant glycerate-2-kinase

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

<table>
<thead>
<tr>
<th>Gradient</th>
<th>Time</th>
<th>Flow</th>
<th>%A</th>
<th>%B</th>
<th>%C</th>
<th>%D</th>
<th>Curve</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.01</td>
<td>0.30</td>
<td>95.0</td>
<td>0.0</td>
<td>5.0</td>
<td>0.0</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>1.20</td>
<td>0.30</td>
<td>95.0</td>
<td>0.0</td>
<td>5.0</td>
<td>0.0</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>10.00</td>
<td>0.30</td>
<td>65.0</td>
<td>0.0</td>
<td>35.0</td>
<td>0.0</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>20.00</td>
<td>0.30</td>
<td>65.0</td>
<td>0.0</td>
<td>35.0</td>
<td>0.0</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>21.00</td>
<td>0.30</td>
<td>5.0</td>
<td>0.0</td>
<td>95.0</td>
<td>0.0</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>30.00</td>
<td>0.30</td>
<td>5.0</td>
<td>0.0</td>
<td>95.0</td>
<td>0.0</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>30.10</td>
<td>0.30</td>
<td>95.0</td>
<td>0.0</td>
<td>5.0</td>
<td>0.0</td>
<td>6</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Column:</th>
<th>Column-Oven:</th>
</tr>
</thead>
</table>
| 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) |

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

<table>
<thead>
<tr>
<th>Gas Flow</th>
<th>Temperatures in °C</th>
</tr>
</thead>
</table>
| Desolvation (L/h): 600  
Cone (L/h): 60 | Desolvation: 400  
Source: 140 |

<table>
<thead>
<tr>
<th>Voltages</th>
<th>Analyser:</th>
</tr>
</thead>
</table>
| 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
Biocatalytic asymmetric phosphorylation catalyzed by recombinant glycerate-2-kinase
Bioreaction engineering leading to efficient synthesis of L-glyceraldehyde-3-phosphate

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

Getachew S. Molla¹, Birhanu M. Kinfu², Jennifer Chow², Wolfgang R. Streit², Roland Wohlgemuth³, and Andreas Liese¹

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

Published in:

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

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

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 gerceraldehyde 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\(_6\)) suppresses the enzyme activity; otherwise PP\(_6\) 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 synthesis for the L-GAP.

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 stereocrystallographic 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 phosphorylated 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 3-phosphoryl-glycerol-3-phosphate using an NADPH dependent aldose reductase as a complement.
of triose phosphate isomerase (TIM) deficiency that enters into gluconeogenesis metabolic pathway [12]. Thus, practical synthetic methodology for the preparation of enantio-pure 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 glyceraldehyde (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 MgCl₂ 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 work as well as numerical performance evaluation of different reactor types for an optimized glyceraldehyde 3-phosphate catalyzed synthesis of L-GAP.

2 Materials and methods

2.1 Materials

Adenosine-5’-triphosphate (ATP) disodium salt, adenine-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 (PF₆n), glyceraldehyde from Cellulomonas sp., glyceraldehyde from E. coli and glyceraldehyde from Streptomyces canus were purchased from Sigma Aldrich GmbH (Buchs, Switzerland). Calcium chloride, magnesium chloride hexahydrate, phosphoric acid, potassium dichromate, 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 Eurokast-H column (300 mm x 8 mm, Knauer) with 5 mM H₂SO₄ 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 10.3 ± 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 x 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 pH of 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 ADP are instable at pH below 2 while at pH above 4 the enzyme could not be fully deactivated.

The operational stabilities of glyceraldehyde 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 MgCl₂ at pH 7.5, 0.5 and 8.5, and 25, 30 and 45°C. Remaining activities were measured routinely for a month. For substrate screening the activity of glyceraldehyde 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 MgCl₂ prepared in 100 mM TEA buffer, pH 8, at 30°C. All reactions were started by the addition of 0.05 mL of glyceraldehyde kinase from Cellulomonas sp. (1.2 mg/mL) prepared in 100 mM TEA buffer, pH 8, and 1.45 mL of 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, bacth reactions were carried out using substrate solutions of fixed 15 mM ATP and 15 mM MgCl₂ 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 MgCl₂ to ATP molar ratio, the activity of glyceraldehyde kinase from Cellulomonas sp. was assayed as a function of MgCl₂ concentration, while maintaining fixed ATP concentrations. Two sets of enzyme activity assays were carried out as a function of MgCl₂ 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 glyceraldehyde kinase from Cellulomonas sp. (1.2 mg/mL) prepared in 100 mM TEA buffer, pH 8, and 1.45 mL of the substrate solutions.
pre-incubated at 30°C. Additionally, the effect of Mg²⁺ concentration on the reaction thermodynamics was examined by performing a reaction using 30 mM L-GA, 30 mM ATP and 500 mM Mg²⁺ 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 Mg²⁺ 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 Mg²⁺ and 0.5 mM (0.5 mM) and 5 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 equations were applied in order to determine reaction kinetic parameters.

The inhibition of glycerol kinase from Cellulomonas sp. by ATP, m-methylglyoxal, DL-GAP, Ca²⁺, inorganic hexametaphosphate (PP₆) and inorganic orthophosphate (P₆) was investigated by measuring enzyme activities as a function of component concentration. Reactions were carried out using substrate solutions of 30 mM L-GA, 30 mM ATP and 21 mM Mg²⁺, prepared in 100 mM TEA buffer, pH 8, at 30°C. Moreover, the enzyme activity was examined as a function of PP₆ and Ca²⁺ concentrations at different Mg²⁺ to ATP molar ratios in order to evaluate the effect of complexation between Mg²⁺ and PP₆ as well as Ca²⁺ and ATP, respectively. Investigation of the effect of Ca²⁺ or other dialvalent 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)/(PP₆) mediated in situ ATP regeneration (results not included in this article) the effect of PP₆ 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 Streptomycetes 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.

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.

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 Mg²⁺ and (0 mM, 3 mM) ADP in 100 mM TEA buffer, pH 8, and 30°C.

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 Mg²⁺ and (0 mM, 3 mM) ADP in 100 mM TEA buffer, pH 8, and 30°C.

<table>
<thead>
<tr>
<th>Reaction components</th>
<th>Kᵢₜ (mM)</th>
<th>Kᵢ (mM)</th>
<th>vₘₚ (μM/μg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP at 21 mM Mg²⁺</td>
<td>0.8 ± 0.05</td>
<td>–</td>
<td>110 ± 3.90</td>
</tr>
<tr>
<td>ATP at 50 mM Mg²⁺</td>
<td>1.90</td>
<td>–</td>
<td>110 ± 3.90</td>
</tr>
<tr>
<td>L-Glyceraldehyde</td>
<td>4 ± 0.06</td>
<td>–</td>
<td>110 ± 3.90</td>
</tr>
<tr>
<td>ADP at 21 mM Mg²⁺</td>
<td>–</td>
<td>1.4 ± 0.18</td>
<td>–</td>
</tr>
<tr>
<td>ADP at 50 mM Mg²⁺</td>
<td>–</td>
<td>3.30</td>
<td>–</td>
</tr>
</tbody>
</table>

Figure 1. Activity of glycerol kinase as a function of the Mg²⁺ to ATP molar ratio: 30 mM L-GA, 70 mM ATP, (0 mM to 400 mM) Mg²⁺, 0.02 mg/mL glycerol kinase from Cellulomonas sp. in 100 mM TEA buffer, pH 8, and 30°C.
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 $3 \times 10^{-6}$, $2.8 \times 10^{-5}$ and $1.2 \times 10^{-3}$ s$^{-1}$ as well as half-lives of 8.35, 6.88 and 0.15 h, respectively.

### 3.3 Reaction kinetics model development and validation

Development of a biocatalytic reaction kinetics model is a 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, 28] 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 supresses the enzyme activity to a non-zero asymptotic value.

Despite being thermodynamically favorable with change in reaction Gibbs free energy ($\Delta G$) 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 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 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
Bioreaction engineering leading to efficient synthesis of L-glyceraldehyde-3-phosphate

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\(_6\) 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\(_6\) suppresses the enzyme activity; whereas, at a higher Mg\(^{2+}\) to ATP molar ratio than the optimum value PP\(_6\) enhances the enzyme activity, until the excess amount of Mg\(^{2+}\) is titrated by PP\(_6\). The inhibition and activation of glycerol kinase by PP\(_6\) at different Mg\(^{2+}\) to ATP molar ratio is due to the complexation of PP\(_6\) 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.

\[
V = V_{\text{max}} \times \frac{[\text{ATP}]}{K_{\text{M,ATP}} \times \left(1 + \frac{[\text{ADP}]}{K_{\text{I,ADP}}} \right) + [\text{ATP}]} \times \frac{[\text{L-GA}]}{K_{\text{M,L-GA}} + [\text{L-GA}]}
\]  

(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, respectively, composed of conversion and reaction terms. [S\(_1\)] and [S] represent influx and efflux substrate concentration, respectively. [P\(_i\)] 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.

\[
\frac{d[\text{ATP}]}{dt} = \frac{d[\text{L-GA}]}{dt} = \left[\text{Glycerol kinase}\right] \times v_2 
\]  

(2)

\[
\frac{d[\text{L-GAP}]}{dt} = \left[\text{Glycerol kinase}\right] \times v_2 \times e^{-0.10t} 
\]  

(3)

\[
\frac{d[\text{ATP}]}{dt} = \frac{d[\text{L-GA}]}{dt} = \left[\frac{[\text{S}]-[\text{S}]}{t}\right] \left[\text{Glycerol kinase}\right] v_2 
\]  

(4)

\[
\frac{d[\text{L-GAP}]}{dt} = \left[\frac{[\text{P}]-[\text{P}]}{t}\right] + \left[\text{Glycerol kinase}\right] v_2 \times e^{-0.10t} 
\]  

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

For glyceraldehyde 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 performance 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 glyceraldehyde 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 2.3 \(\times\) 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 \(\times\) 10\(^{-2}\) s\(^{-1}\).

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

Equal concentration of C{\textsuperscript{2}+} exerts different percentage of inhibition at different Mg{\textsuperscript{2+}} to ATP molar ratios due to mass action on the complexation equilibrium between Ca-ATP and Mg-ATP complexes. If the Mg{\textsuperscript{2+}} to ATP molar ratio is set to be at the optimum value or below, PP,6 suppresses the enzyme activity; otherwise PP,6 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/ to 1/ g kg\textsuperscript{-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 BMFB (German Federal Ministry of Education and Research) cluster “Biocatalysis 2021” (project number: 0316056) 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 Lisee 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


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

© 2016 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
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.

<table>
<thead>
<tr>
<th>Glycerol kinase source</th>
<th>Activity (U/mg)</th>
<th>Half-life (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em></td>
<td>15.0</td>
<td>3.2</td>
</tr>
<tr>
<td><em>Cellulomonas</em> sp.</td>
<td>24.0</td>
<td>No significant loss of the enzyme activity was detected in a month of incubation period</td>
</tr>
<tr>
<td><em>Streptomyces canus</em></td>
<td>18.5</td>
<td>21.0</td>
</tr>
</tbody>
</table>

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

Birhanu M. Kinfu¹, Maike Köster¹, Mareike Janus¹, Volkan Besirlioglu², Michael Roggenbuck³, Richard Meurer², Ljubica Vojcic², Martin Borchert³, Ulrich Schwaneberg², Jennifer Chow¹ and Wolfgang R. Streit¹

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

Published in:

Biotechnology and Bioengineering, 114(12): 2739-2752. doi:10.1002/bit.26436
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, Ohrhorstr. Hamburg, Germany
\(^2\) Institute of Biotechnology, RWTH Aachen, Aachen, Germany
\(^3\) Novozymes A/S, Bagsvaerd, Denmark

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

**Funding information**
Deutsches Akademisches Austauschdienst, Grant number: DAAD (57076385); Danish Innovationsfonden + BMBF, Grant number: MetaCat (ERA-IB-14-03001A1565A); Bundesministerium für Bildung und Forschung (BMBF), Grant number: 1VDT-V\(^*\) (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 housekeeping sigma factor and two transcription elongation factors were cloned and over expressed as His\(_\text{6}\)-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\(^\circ\)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, Sweat, 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...
Recombinant RNAP from *Geobacillus* sp. GHH01 as tool for rapid generation of ...
Recombinant RNAP from *Geobacillus* sp. GH01 as tool for rapid generation of ...
### TABLE 3 Plasmids and constructs employed in this study

<table>
<thead>
<tr>
<th>Plasmids and constructs</th>
<th>Features</th>
<th>Reference/source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBluescript II SK (+)</td>
<td>3.0 kb phagemid vector, lacZ, bla, P7, P3</td>
<td>Stratagene, La Jolla (CA)</td>
</tr>
<tr>
<td>pDrive</td>
<td>3.85 kb TA-cloning vector, oriE, PlacIaZ, AmpR, KanR, T7-promoter</td>
<td>QIAGEN (Hilden, Germany)</td>
</tr>
<tr>
<td>pET-21a(+)</td>
<td>5.44 kb Expression vector, lac, AmpR, T7-promoter, C- His6-tag coding sequence</td>
<td>Novagen/Merck (Darmstadt, Germany)</td>
</tr>
<tr>
<td>pET-28a(+)</td>
<td>5.37 kb Expression vector, lac, KanR, T7-promoter, N- and C-terminal His6-tag coding sequence</td>
<td>Novagen/Merck (Darmstadt, Germany)</td>
</tr>
<tr>
<td>pBRR1MCS-5</td>
<td>4.77 kb broad host range expression vector, rep, mob, lacZ, GmR</td>
<td>Kovach et al. (1995)</td>
</tr>
<tr>
<td>pET-28a(+)::rpoA</td>
<td>N-His6-tagged α-subunit of GH01 RNAP</td>
<td>This work</td>
</tr>
<tr>
<td>pET-28a(+)::rpoB</td>
<td>N-His6-tagged β-subunit of GH01 RNAP</td>
<td>This work</td>
</tr>
<tr>
<td>pET-28a(+)::rpoC</td>
<td>N-His6-tagged β’-subunit of GH01 RNAP</td>
<td>This work</td>
</tr>
<tr>
<td>pET-28a(+)::rpoD</td>
<td>N-His6-tagged ω-subunit of GH01 RNAP</td>
<td>This work</td>
</tr>
<tr>
<td>pET-21a(+)::rpoA</td>
<td>α-subunit of GH01 RNAP</td>
<td>This work</td>
</tr>
<tr>
<td>pET-21a(+)::rpoB</td>
<td>β-subunit of GH01 RNAP</td>
<td>This work</td>
</tr>
<tr>
<td>pET-21a(+)::rpoC</td>
<td>β’-subunit of GH01 RNAP</td>
<td>This work</td>
</tr>
<tr>
<td>pET-21a(+)::rpoE</td>
<td>6-subunit of GH01 RNAP</td>
<td>This work</td>
</tr>
<tr>
<td>pET-21a(+)::rpoZ</td>
<td>ω-subunit of GH01 RNAP</td>
<td>This work</td>
</tr>
<tr>
<td>pET-21a(+)::rpoD</td>
<td>ω-β-ω-ω-β subunits, His6-tag on β’-subunit</td>
<td>This work</td>
</tr>
<tr>
<td>pET-28a(+)::greA</td>
<td>N-His6-tagged transcription elongation factor GreA</td>
<td>This work</td>
</tr>
<tr>
<td>pET-28a::Mfd</td>
<td>N-His6-tagged transcription elongation factor Mfd</td>
<td>This work</td>
</tr>
<tr>
<td>pGFP cloning vector</td>
<td>3.35 kb, AmpR, lac, Expressed eGFP tag</td>
<td>CLONTECH Laboratories, Palo Alto (CA 94303-4230), GenBank: U76561.1</td>
</tr>
<tr>
<td>pBRR1MCS-5: eGFP</td>
<td>eGFP under control of lac-promotor</td>
<td>This work</td>
</tr>
<tr>
<td>pCC1FOS 108</td>
<td>Fosmid harboring metagenomic DNA including LrpF</td>
<td>Rabausch et al. (2013)</td>
</tr>
</tbody>
</table>

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, 8, 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 His6-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 gam2 2(DE3) host cells. Expression was then induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) at an approximate OD600 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 x g and 4°C or for larger volumes using a Sorvall RC6+ centrifuge (rotor F105–6 x 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 His6-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 NaH2PO4, 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 MgCl2 to a final concentration of 2 mM.
Recombinant RNAP from *Geobacillus* sp. GHH01 as tool for rapid generation of...

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 (HiTrap columns, GE Healthcare, Freiburg, Germany) preequilibrated with lysis buffer and purified using a fast protein liquid chromatography (FPLC) system. Bond protein was washed with 5 column volumes (CV) of wash buffer (50 mM NaH₂PO₄ : 2H₂O 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 NaH₂PO₄ : 2H₂O 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 σ⁵₂-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 (β and ω) were separated from the majority of E.coli proteins through heat-precipitation. For this, the cell pellet was resuspended in 5 mM 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
Recombinant RNAP from *Geobacillus* sp. GHH01 as tool for rapid generation of ...
Recombinant RNAP from Geobacillus sp. GHH01 as tool for rapid generation of …

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 MgCl₂, 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, 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 (nanofluor) technology (Prometheus NI-AS, Nanoimper Technologies, Munich, Germany). In this thermo-reconstitution approach, purified subunits were mixed in reconstitution buffer with a ratio of 2:8:4:2:2 (α:β:δ:ω:σ) 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 (RNAPc).

RNAPc obtained using both reconstitution methods were further assembled to the respective RNAP holozymes (RNAPh) by addition of purified σ70-factor in an equimolar concentration and incubating at 37°C for 1 hr. Both reconstituted RNAPc and RNAPh 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 (cRNAPc) was purified using two chromatographic steps. His-tagged cRNAPc and non-assembled β’ subunit were co-purified using HiTrap columns as described above. Protein-containing fractions were pooled and either purified further via anion-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 cRNAPc were identified by SDS–PAGE analysis and transcription assays. Buffer was exchanged to 1× 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 VIVASPIN®6 concentrator as described above. Similarly, the core enzyme was further assembled to holozyme by addition of purified σ70-factor in an equimolar concentration and incubation at 37°C for 1 hr.

2.6 In vitro transcription

In vitro transcription reactions contained 40 mM Tris-HCl pH 8.0, 20 mM MgCl₂, 2 mM spermidine, 5 mM DTT, rNTP mix (rATP, rGTP, rUTP, rTTP, 4 mM each), 0.1 µM thermostable inorganic pyrophosphatase (New England Biolabs GmbH, Frankfurt Main, Germany), and 1–2 µM RiboLock RNase Inhibitor (ThermoFisher Scientific GmbH, Schwerte, Germany). If not mentioned otherwise, a polymerase concentration of 0.25 µg/µl and RNAP holozyme reconstituted from His-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 Soil 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 GroEL and FMD from Geobacillus sp. GHH001 to in 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.

Thermotolerance of the RNAPc 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 (Ilmerberger et al., 2014), a biogas reactor (Gülleret et al., 2016), and a microalgae culture (Krohn-Molt et al., 2013) as well as gDNAs from E. coli, Bacillus subtilis, Propanobacterium acidopropioni, Psenbacillus 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 μg/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, pEGFP-PvC); (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 lip4-1 including a putative upstream promoter-region as well as additional downstream nucleotides. Transcription was performed using co-expressed core enzyme (cRNAP) with α-32P-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 (α,β,β',ω,ω') forming the core enzyme (RNAPc) (Figure 1). Genes coding for RNAP subunits (rpoA, rpoB, rpoC, rpoE, rpoD) and the housekeeping α-32P-factor (rpoD) were cloned as described in section 2 into two different expression vectors (pET21α[+] and pET28a[+]) for heterologous expression of untagged or His6-tagged proteins, respectively (Figures 1a and 1b). With the N-terminal His6-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 α-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-21α:rpocCBZAE. The first gene of this construct, the rpoC coding for the β'-subunit, was fused with an N-terminal His6-tag for RNAP 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 cRNAPc.

Purification of in vitro co-expressed RNAPc was achieved with a two-step chromatographic approach. As a first step, His6-tag based chromatographic purification resulted in fractions containing non-assembly His6-β' along with cRNAPc. Pure cRNAPc 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 cRNAPc 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 His6-β' subunit. The presence of active polymerase was verified by testing the fractions for transcription activities (Figure 53). Purified transcription elongation factors GreA and MID proteins were obtained the same way as Hs8-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 RNAPc and (ii) co-expression of the five RNAP-subunits in a single vector construct and simultaneous in vivo assembly of cRNAPc followed by chromatographic purification (section 3.1). RNAPc 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 52a). In addition to the urea-based denaturation and subsequent renaturation, a new RNAPc assembly method referred to 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 RNAPc-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 52b). RNAP core enzymes assembled with both reconstitution methods were either used directly for transcription assay or further incubated with α-32P-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. GHH01 on the amount of RNA generated was examined. Accordingly, addition of GreA improved transcript quantity significantly whereas MID did not have a detectable effect (Figure 52C). Depending on the template DNA used for transcription, the amounts of produced RNA was approximately in a range of 92-316.4 μg RNA/μg His6-tagged RNAPc 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 His6-tagged and all untagged subunits, the His6-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 RNAPc 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 52B).
Recombinant RNAP from *Geobacillus* sp. GHH01 as tool for rapid generation of …

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 ancestral 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 ancestral genomic DNA from *P. torridus* with a yield of 23.5 ± 4.9 ng/µl RNA. The amount of mRNA synthesized with *Geobacillus* RNAP, was in most cases significantly (up to three times) higher compared to the transcript produced by employing commercial *E. coli* RNAP, 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.
and was independent on the presence or absence of the recombinantly fused poly-histidine protein tag. The α-subunit and α70-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 6-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 6- and α-subunits followed by β'-subunit, α70-factor and β-subunit (see Table 4). Accordingly, a maximum non-denaturing temperature of 50°C was chosen for assembly of RNAPα with thermost-reconstitution.

Further tests were performed to investigate thermostability of the assembled RNAPα complexes. Comparative analysis of in vitro transcription profiles with commercial E. coli RNAPα indicated that RNAPα from Geobacillus is stable and active over a broad temperature range under the conditions tested. Geobacillus RNAPα 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 RNAPα 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 (Eikkers 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
Recombinant RNAP from *Geobacillus* sp. GHH01 as tool for rapid generation of ...
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 σ54-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 RNAP1 were also up to three times higher compared to commercial E. coli RNAP1 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 (Esynunia & Kulbachinsky, 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-I8 network program MetaCat and through the BMBF-cooperation project ivDIT-V². Deutsche Akademischer Austauschdienst (DAAD) is gratefully acknowledged for the scholarship provided to B.M.K.

REFERENCES


Recombinant RNAP from *Geobacillus* sp. GHH01 as tool for rapid generation of …


SUPPORTING INFORMATION

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

Recombinant RNAP from *Geobacillus* sp. GHH01 as tool for rapid generation of …

Supplementary Material for DOI:10.1002/bit.26436

Table S1: Expression conditions for individual RNAP subunits, co-assembled RNAP<sub>c</sub> 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.

<table>
<thead>
<tr>
<th>Plasmid constructs</th>
<th><em>E. coli</em> expression strain type</th>
<th>IPTG (mM)</th>
<th>T (°C)</th>
<th>Induction time (hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pET21a(+)::rpoA</td>
<td>BL21(DE3)</td>
<td>1.0</td>
<td>28</td>
<td>1</td>
</tr>
<tr>
<td>pET21a(+)::rpoB</td>
<td>BL21(DE3)</td>
<td>1.0</td>
<td>37</td>
<td>3</td>
</tr>
<tr>
<td>pET21a(+)::rpoC</td>
<td>BL21(DE3)</td>
<td>1.0</td>
<td>22</td>
<td>3</td>
</tr>
<tr>
<td>pET21a(+)::rpoE</td>
<td>BL21(DE3)</td>
<td>1.0</td>
<td>37</td>
<td>6</td>
</tr>
<tr>
<td>pET21a(+)::rpoZ</td>
<td>Rosetta gami 2</td>
<td>0.5</td>
<td>37</td>
<td>17</td>
</tr>
<tr>
<td>pET21a(+)::rpoD</td>
<td><em>no detectable expression</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pET28a(+)::rpoA</td>
<td>BL21(DE3)</td>
<td>0.4</td>
<td>37</td>
<td>6</td>
</tr>
<tr>
<td>pET28a(+)::rpoB</td>
<td>Rosetta gami 2</td>
<td>0.4</td>
<td>22</td>
<td>17</td>
</tr>
<tr>
<td>pET28a(+)::rpoC</td>
<td>BL21(DE3)</td>
<td>0.4</td>
<td>22</td>
<td>6</td>
</tr>
<tr>
<td>pET28a(+)::rpoE</td>
<td>BL21(DE3)</td>
<td>0.4</td>
<td>37</td>
<td>6</td>
</tr>
<tr>
<td>pET28a(+)::rpoZ</td>
<td>BL21(DE3)</td>
<td>0.4</td>
<td>22</td>
<td>6</td>
</tr>
<tr>
<td>pET28a(+)::rpoD</td>
<td>Rosetta gami 2</td>
<td>0.4</td>
<td>22</td>
<td>17</td>
</tr>
<tr>
<td>pET21a(+)::rpoCBZAE</td>
<td>Rosetta gami 2</td>
<td>0.5</td>
<td>37</td>
<td>17</td>
</tr>
<tr>
<td>pET-28a::greA</td>
<td>BL21(DE3)</td>
<td>0.5</td>
<td>37</td>
<td>17</td>
</tr>
<tr>
<td>pET-28a::Mfd</td>
<td>BL21(DE3)</td>
<td>0.5</td>
<td>22</td>
<td>17</td>
</tr>
</tbody>
</table>
Recombinant RNAP from *Geobacillus* sp. GHH01 as tool for rapid generation of …

**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 His6-tag coding sequence for chromatographic purification of co-expressed RNAP (cRNAP). RNAPc and RNAPh stand for RNAP core and holoenzyme respectively.
Recombinant RNAP from *Geobacillus* sp. GHH01 as tool for rapid generation of ...

**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.
Recombinant RNAP from *Geobacillus* sp. GHH01 as tool for rapid generation of …

**FIGURE S3:** Activity test of co-expressed RNAP<sub>C</sub>. Multiple possible cRNAP<sub>C</sub>-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, Wolfgang 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 β-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 β-glycerophosphate forming enzyme via multiple approaches. In sequence databases, there is an 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 β-glycerophosphate side chain (Wang, et al., 2010). In this study, the gene was expressed and purified as His6-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 β-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 β-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 β-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 β-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 β-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 β-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 glyceralic 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 fused 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 within 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
Discussion

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


References


References


References


References


References


References


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.

አተንሸንስ!