

Understanding Archaea-type Ether Lipids –  
searching for new derivatives in Bacteria and Archaea,  
functional studies on Geranylgeranylglyceryl Phosphate Synthase



DISSERTATION

ZUR ERLANGUNG DES DOKTORGRADES  
DER NATURWISSENSCHAFTEN (DR. RER. NAT.) DER FAKULTÄT FÜR  
BIOLOGIE UND VORKLINISCHE MEDIZIN DER UNIVERSITÄT REGENSBURG

vorgelegt von  
**Cosimo Kropp**  
aus München

Dezember 2021



Das Promotionsgesuch wurde eingereicht am:

21.12.2021

Die Arbeit wurde angeleitet von:

PD DR. PATRICK BABINGER

Unterschrift:

.....

COSIMO KROPP



## Abstract

A trademark of Archaea is their exclusive membrane lipid composition. They possess so-called ether lipids that differ from bacterial ones in three characteristics: (i) isoprenoid chains, commonly C20, are connected to (ii) *sn*-glycerol 1-phosphate (G1P) by (iii) ether linkages. These characteristics are introduced by the geranylgeranylglyceryl phosphate synthase (GGGPS) which catalyzes the condensation of G1P and the isoprenoid. The GGGPS has been considered exclusive for the archaeal domain, therefore it was surprising when homologues were found in *Bacillus subtilis*, followed by the identification of Archaea-type ether lipids. However, these ether lipids are different to their archaeal counterparts: They consist of one C35 isoprenoid chain and instead of addition of a second chain, the glycerol is diacetylated by the *O*-acetyltransferase YvoF. GGGPS enzymes have also been found in Bacteroidetes and in phylogenetic studies the GGGPS enzyme family was clustered into two groups, both containing archaea and bacteria, with dimeric enzymes in group I and dimers as well as hexamers in group II.

The first part of this thesis deals with the characterization of group II GGGPS predecessors, in order to retrace evolution of the GGGPS. For this purpose, ancestral enzymes that were calculated previously in an ancestral sequence reconstruction approach, were examined for their oligomerization state and stability. Results imply that the hexamer must have developed early and that dimeric GGGPS enzymes in this group have evolved from the hexamer. However, the fine-tuning of the flexibilization, which can be observed for some GGGPS enzymes, developed late. It is assumed that this flexibilization might provide high activity in a rigid, stability-ensuring backbone. In addition, characterization of predecessors provided information about the oligomerization state of modern GGGPS enzymes that have not been purified so far.

In the second part, I present a proof-of-concept in which the oligomerization state of the GGGPS was selectively modulated by external input to gain control over enzymatic activity. This was performed for priorly dimerized modern and ancestral GGGPS enzymes. As those dimerized variants exhibited no or very low enzymatic activity, the implementation of a switch for the oligomerization state also allowed switching of activity. Hexameric GGGPS enzymes have a cation- $\pi$  bond in the hexamerization interface between neighboring protomers, which was converted into a molecular switch using chemical rescue and optical control. The former aimed at the cationic residue of the bond and the latter targeted the aromate. Both strategies were used successfully to selectively hexamerize the GGGPS enzymes and thus gain control over enzymatic activity.

The third part presents experiments that aimed to contribute to understanding the function of ether lipid derivatives in Bacteria and Halobacteria (Archaea). Quantification of ether lipids in *B. subtilis* using mass spectrometry (MS) indicated that the proportion is rather small. Furthermore, MS was performed to find ether lipids in Bacteroidetes, which failed, however, we tentatively identified a variety of acetylated diether lipids in Halobacteria inspired by previous studies that discussed an occurrence of YvoF in Halobacteria.



## References of Published Manuscripts

This thesis is composed of the following manuscripts. Permission to use the manuscripts as chapters and to make minor editorial changes was granted under the following licenses: <https://creativecommons.org/licenses/by/4.0/> and <https://creativecommons.org/licenses/by-nc/4.0/>.

- A** **Kropp C**, Straub K, Linde M, Babinger P. 2020. Hexamerization and thermostability emerged very early during geranylgeranylgeranyl phosphate synthase evolution. *Protein Sci.* 30: 583-96
- B** **Kropp C**, Bruckmann A, Babinger P. 2021. Controlling enzymatic activity by modulating the oligomerization state via chemical rescue and optical control. *Chembiochem* 23: e202100490
- C** **Kropp C**, Lipp J, Seisenberger C, Linde M, Hinrichs KU, Babinger P. 2021. Identification of acetylated diether lipids in halophilic Archaea. *Submitted for publication*

In the course of this work, I contributed to further publications that are not part of this thesis:

- D** Straub K, Linde M, **Kropp C**, Blanquart S, Babinger P, Merkl R. 2019. Sequence selection by FitSS4ASR alleviates ancestral sequence reconstruction as exemplified for geranylgeranylgeranyl phosphate synthase. *Biol. Chem.* 400: 367-81





## Personal Contributions

### Publication A

The research was designed by myself and Patrick Babinger with contributions from Mona Linde. Kristina Straub performed ancestral sequence reconstruction and Mona Linde cloned, purified, and characterized the oligomerization state of the ancestral wild-type proteins \_N2, \_N3, \_N4, \_N5, and \_N12. All other experiments were performed by myself. The publication was written by myself and Patrick Babinger with contributions from Kristina Straub.

### Publication B

The research was designed by myself and Patrick Babinger. Astrid Bruckmann performed tryptic digest and mass spectrometry measurements. All other experiments were performed by myself. The manuscript was written by myself and Patrick Babinger with contributions from Astrid Bruckmann.

### Publication C

The research was designed by myself and Patrick Babinger. Christina Seisenberger purified and characterized hvYvoF. Mona Linde supervised experiments of Christina Seisenberger. hvYvoF absorption spectroscopy, strain cultivation and production of total lipid extracts was performed by myself. Ultra-high performance liquid chromatography coupled to mass spectrometry was performed by Julius Lipp and myself. Figure 3 and Figure S6 – S11 were made by Julius Lipp and myself, with major contribution from Julius Lipp. The rest of the figures was made by myself. The manuscript was written by myself and Patrick Babinger with contributions from Julius Lipp.



# Contents

Abstract.....	v
References of Published Manuscripts .....	vii
Personal Contributions.....	ix
Contents.....	xi
List of Figures.....	xv
List of Tables.....	xvii
<b>1 General Introduction.....</b>	<b>1</b>
1.1 The domain of Archaea and their membrane lipids .....	1
1.1.1 Ether lipid synthesis and evolutionary aspects.....	2
1.1.2 Phylogenetic distribution of the GGGPS .....	6
1.1.3 The GGGPS – fold, structure, and stability.....	7
1.2 Protein design and ancestral sequence reconstruction .....	9
1.3 Aim and scope of this work.....	11
1.4 Guide to the following chapters .....	11
<b>2 Hexamerization and thermostability emerged very early during geranylgeranylglyceryl phosphate synthase evolution.....</b>	<b>13</b>
2.1 Abstract.....	14
2.2 Statement for a broader audience .....	14
2.3 Introduction.....	15
2.4 Results.....	17
2.4.1 Ancestral sequence reconstruction of group II GGGPS .....	17
2.4.2 Probably all crenarchaeal GGGPS are hexamers .....	19
2.4.3 Hexamerization emerged very early during group II GGGPS evolution .....	21

2.4.4	Extant dimeric group II GGGPS are secondary dimers.....	23
2.4.5	Thermostability emerged very early during group II GGGPS evolution, flexibilization of the active site late.....	25
<b>2.5</b>	<b>Discussion.....</b>	<b>28</b>
<b>2.6</b>	<b>Materials &amp; Methods.....</b>	<b>30</b>
2.6.1	Ancestral sequence reconstruction of GGGPS group II sequences.....	30
2.6.2	Cloning and site-directed mutagenesis.....	30
2.6.3	Production and purification of recombinant proteins.....	31
2.6.4	Characterization of oligomerization state of recombinant proteins.....	32
2.6.5	CD spectroscopy.....	32
2.6.6	DSC.....	32
2.6.7	nanoDSF.....	32
2.6.8	Classification of thermal transitions as T <sub>1</sub> (partial denaturation) or T <sub>2</sub> (complete denaturation).....	33
2.6.9	Radiometric GGGPS activity assays.....	33
2.6.10	Heat inactivation.....	34
2.6.11	ITC.....	34
<b>2.7</b>	<b>Acknowledgements.....</b>	<b>34</b>
<b>2.8</b>	<b>Supporting Information.....</b>	<b>34</b>
<b>3</b>	<b>Controlling enzymatic activity by modulating the oligomerization state via chemical rescue and optical control.....</b>	<b>49</b>
<b>3.1</b>	<b>Abstract.....</b>	<b>50</b>
<b>3.2</b>	<b>Introduction.....</b>	<b>50</b>
<b>3.3</b>	<b>Results &amp; Discussion.....</b>	<b>53</b>
3.3.1	Cation- $\pi$ interaction as molecular switch.....	53
3.3.2	Activation of catalytic activity by switching the oligomerization state.....	54
3.3.3	Physico-biochemical analysis of the inactive and reactivated variants.....	57
<b>3.4</b>	<b>Conclusion.....</b>	<b>59</b>
<b>3.5</b>	<b>Materials &amp; Methods.....</b>	<b>60</b>
3.5.1	Cloning and site-directed mutagenesis.....	60
3.5.2	Production and purification of proteins.....	60
3.5.3	Selective activation of enzymatic activity.....	61
3.5.4	Characterization of the oligomerization state of proteins.....	61
3.5.5	MW determination by SLS.....	61

---

3.5.6	Steady-State Enzyme kinetics .....	62
3.5.7	CD spectroscopy .....	62
3.5.8	nanoDSF .....	62
3.5.9	ITC .....	63
3.5.10	Tryptic digest and MS .....	63
<b>3.6</b>	<b>Acknowledgements .....</b>	<b>64</b>
<b>3.7</b>	<b>Supporting Information .....</b>	<b>65</b>
<b>4</b>	<b>Spectrometric analysis of ether lipids in Bacteria and Archaea .....</b>	<b>77</b>
<b>4.1</b>	<b>Archaea-type ether lipids in Firmicutes and Bacteroidetes.....</b>	<b>77</b>
4.1.1	Summary.....	77
4.1.2	Introduction.....	77
4.1.3	Results & Discussion.....	78
4.1.4	Conclusion.....	84
4.1.5	Materials.....	85
4.1.6	Methods.....	87
4.1.7	Digital supplement.....	88
<b>4.2</b>	<b>Identification of acetylated diether lipids in halophilic Archaea .....</b>	<b>89</b>
4.2.1	Abstract.....	90
4.2.2	Introduction.....	90
4.2.3	Results & Discussion.....	92
4.2.4	Conclusion.....	96
4.2.5	Materials & Methods.....	97
4.2.6	Acknowledgements.....	100
4.2.7	Supporting Information.....	101
<b>5</b>	<b>Comprehensive Summary, Discussion, and Outlook .....</b>	<b>113</b>
5.1	Comprehensive Summary & Discussion .....	113
5.2	Comprehensive Outlook .....	119
	<b>List of Abbreviations.....</b>	<b>127</b>
	<b>Bibliography.....</b>	<b>129</b>
	<b>Acknowledgements .....</b>	<b>141</b>



## List of Figures

Figure 1.1. The three-domain model in the tree of life presentation.....	1
Figure 1.2. Comparison of archaeal ether lipids and ester lipids from Bacteria and Eukarya.....	2
Figure 1.3. Separation of Archaea and Bacteria from the cenancestor due to the “lipid divide”	3
Figure 1.4. Biosynthesis of Archaea-type ether lipids in <i>B. subtilis</i> . .....	5
Figure 1.5. Schematic presentation of the phylogenetic tree of the GGGPS enzyme family.....	6
Figure 1.6. Structural features of the GGGPS. ....	7
Figure 1.7. Schematic presentation of an ancestral sequence reconstruction (ASR) generated tree. .....	10
Figure 2.1. Simplified phylogenetic tree of the GGGPS enzyme family.....	16
Figure 2.2. Simplified phylogenetic tree used for ASR of GGGPS group II predecessors. ....	18
Figure 2.3. Analysis of GGGPS variants with histidine as “aromatic anchor”. .....	20
Figure 2.4. Analytical SEC of ancestral GGGPS proteins. ....	22
Figure 2.5. Analytical SEC of AncGGGPS2_N4_IF_n12 and AncGGGPS2_N12_IF_n4. ....	23
Figure 2.6. Analytical SEC of AncGGGPS2_N73. ....	24
Figure 2.7. Heat inactivation of mtGGGPS and AncGGGPS2_N12.....	27
Figure 3.1. Properties of the GGGPS hexamer. ....	52
Figure 3.2. Switch conceptualization.....	54
Figure 3.3. Oligomerization states of GGGPS variants.....	55
Figure 4.1. RP-UHPLC-ESI-MS analysis of total lipid extracts (TLE) from <i>B. subtilis</i> wild-type and mutant variants.....	81
Figure 4.2. Search criteria for UHPLC-MS analysis of Bacteroidetes TLEs. ....	83
Figure 4.3. Biosynthesis of ether lipids in Bacteria and Archaea.....	91
Figure 4.4. Test for GGG-specific acetyltransferase activity of hvYvoF.....	92
Figure 4.5. RP-UHPLC-ESI-MS analysis of a <i>H. volcanii</i> TLE.....	94
Figure 5.1. Graphical summary and links between the projects.....	114
Figure 5.2. Analysis of mfGGGPS, N29, and N28. ....	121

**Supplementary Figures**

Figure S2.1. Phylogenetic tree deduced from the manually curated sequence set used for ASR of GGGPS group II predecessors.....	40
Figure S2.2. Activity assay of GGGPS variants used in this work.....	41
Figure S2.3. Structural integrity of GGGPS variants characterized in this study.....	42
Figure S2.4. Analytical SEC with varying concentrations of AncGGGPS2_N2 and AncGGGPS2_N3.....	43
Figure S2.5. Sequence alignment of AncGGGPS2_N4 and AncGGGPS2_N12.....	44
Figure S2.6. Analysis of G1P binding to AncGGGPS2_N4 variants by isothermal titration calorimetry.....	45
Figure S2.7. Analytical SEC of rsGGGPS, fsGGGPS, mrGGGPS and nvGGGPS.....	46
Figure S2.8. Thermal stability of GGGPS proteins, followed by nanoDSF and DSC.....	47
Figure S2.9. Purity of GGGPS variants characterized in this study.....	48
Figure S3.1. Analytical SEC of N4 and N12.....	67
Figure S3.2. Steady-state kinetic measurement of AncGGGPS_N12.....	68
Figure S3.3. Activity test of GGGPS variants N4 and RGGGPS <sub>OC</sub> _W139ONBY <sup>hv</sup> .....	68
Figure S3.4. Chemical structures of modified amino acids used in this study.....	69
Figure S3.5. LC-MS/MS analysis of GGGPS variants.....	70
Figure S3.6. Reactivation of RGGGPS <sub>CKr</sub> _K144C.....	71
Figure S3.7. Steady-state kinetic measurements.....	72
Figure S3.8. Structural integrity of GGGPS variants characterized in this study.....	73
Figure S3.9. Analysis of substrate binding to GGGPS variants.....	74
Figure S3.10. Thermal stability of GGGPS variants followed by nanoDSF.....	75
Figure S4.1. Pairwise sequence alignment of YvoF from <i>B. subtilis</i> (bsYvoF) and from <i>H. volcanii</i> (hvYvoF).....	102
Figure S4.2. Purification of hvYvoF by IMAC.....	102
Figure S4.3. Absorption spectrum of hvYvoF.....	103
Figure S4.4. Thermal stability of hvYvoF, measured by nanoDSF.....	104
Figure S4.5. Test of salt dependent activity of hvYvoF.....	105
Figure S4.6. RP-UHPLC-ESI-MS analysis of a <i>H. salinarum</i> TLE.....	106
Figure S4.7. Detection of saturated and unsaturated ARs.....	107
Figure S4.8. Detection of saturated and unsaturated Ac-ARs.....	108
Figure S4.9. MS analysis of a <i>H. volcanii</i> TLE.....	109
Figure S4.10. Detection of saturated and unsaturated Ac-AGs.....	110
Figure S4.11. Detection of saturated and unsaturated AGs.....	111



## List of Tables

Table 2.1. Temperatures of thermal transitions of different GGGPS proteins as followed by nanoDSF and DSC. <sup>a</sup> .....	26
Table 3.1. Catalytic parameters of GGGPS variants in inactive and active states. <sup>a</sup> .....	57
Table 4.1. Culture conditions of strains used for UHPLC-MS analysis. <sup>a</sup> .....	79
Table 4.2. Oligonucleotides for Bacteroidetes strain identification. ....	85
Table 4.3. Strains used in this project. ....	85
Table 4.4. Medium 67 for <i>F. johnsoniae</i> and <i>C. pinensis</i> cultivation .....	86
Table 4.5. Medium 7 for <i>S. linguale</i> cultivation.....	87
Table 4.6. Modified Medium 514 for <i>Z. profunda</i> cultivation .....	87
Table 4.7. Identified ether lipids in Halobacteria. ....	95

## Supplementary Tables

Table S2.1. Experimentally determined oligomerization states of GGGPS enzymes. ....	35
Table S2.2. GGGPS variants characterized in this study with their purification yields after expression in <i>E. coli</i> . ....	37
Table S2.3. Primer sequences used for site-directed mutagenesis.....	38
Table S3.1. Oligomerization states and MWs of GGGPS complexes. <sup>a</sup> .....	65
Table S3.2. Oligonucleotide sequences for production of the substitutions by site-directed mutagenesis. ....	66
Table S4.1. Primers used for amplification and cloning of <i>hvyvoF</i> into pTA1228 vector. <sup>a</sup> .....	101

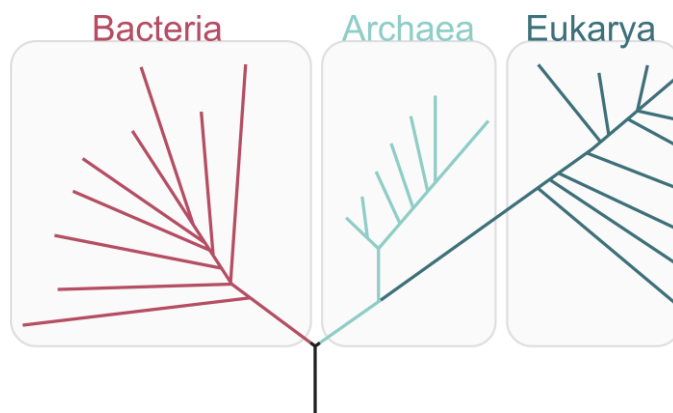


# Chapter 1

## 1 General Introduction

### 1.1 The domain of Archaea and their membrane lipids

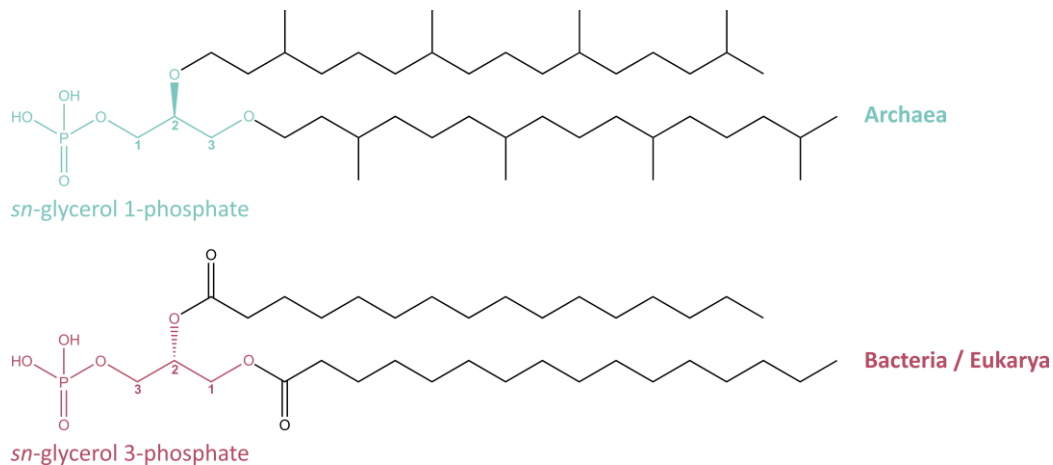
Organismic life can be populated into subgroups following certain attributes. The most widespread presentation is the so-called tree of life, in which life is clustered into three domains: Bacteria, Archaea, and Eukarya (Figure 1.1; Woese et al 1990). This three-domain model was calculated based on 16S rRNA sequence analysis. Until this groundbreaking study, archaea were considered as part of the bacterial domain, however, genotypic as well as phenotypic investigations proposed to classify Archaea as a domain of its own. Since then, Archaea have been considered a separate primary lineage (Woese & Gupta 1981).



**Figure 1.1.** The three-domain model in the tree of life presentation. Bacteria are colored red, Archaea cyan, and Eukarya anthracite. The figure is adapted from Woese et al (1990).

The domain of Archaea has been attracting the interest of science, as these organisms are known to colonize areas hostile to life. These extreme environmental conditions include high temperatures, halophilic surroundings, or very acidic pH conditions. Consequently, archaea show interesting unique biochemical properties, like highly diverged ribosomal RNAs, or exclusive metabolic cofactors (Woese et al 1978). Another distinguishable attribute that stands out in Archaea is their membrane composition as they have unique lipid types for surviving under these harsh conditions. In Bacteria and Eukarya, the membrane consists of ester lipids, where

*sn*-glycerol 3-phosphate (G3P) is esterified with two fatty acids (Lombard et al 2012b). In archaeal membranes, however, the main membrane components are ether lipids (Lombard et al 2012b, Woese et al 1978). These are quite different to ester lipids with respect to stereochemistry and composition. Most commonly, polyisoprenyl moieties with 20 carbon atoms (C<sub>20</sub>) are connected to *sn*-glycerol 1-phosphate (G1P) by ether linkage (Figure 1.2; Kates 1993, Koga et al 1993, Matsumi et al 2011, Wächtershäuser 2003).



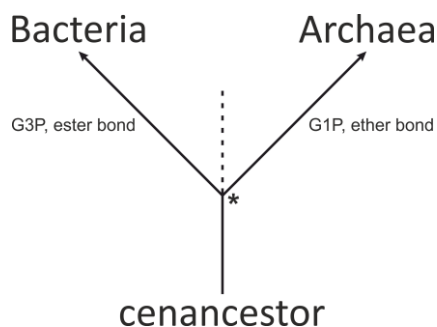
**Figure 1.2.** Comparison of archaeal ether lipids and ester lipids from Bacteria and Eukarya. In ether lipids, *sn*-glycerol 1-phosphate (G1P, cyan) is linked to isoprenoid chains via ether linkage, while ester lipids possess *sn*-glycerol 3-phosphate (G3P, red) esterified with fatty acids.

In all domains, these core lipids are later on modified with polar head groups. This is done by producing an activated intermediate by attaching cytidine diphosphate (CDP) to the glycerol under elimination of the phosphate. CDP is then replaced by various headgroups resulting in typical membrane lipids, such as archaetidylethanolamine or archaetidylglycerol in archaea and phosphatidylethanolamine or phosphatidylglycerol in bacteria (Moldoveanu & Kates 1988).

### 1.1.1 Ether lipid synthesis and evolutionary aspects

The three specific attributes of ether lipids – G1P stereochemistry, ether linkage, isoprenoid chains – are introduced by two stereospecific enzymes in Archaea. G1P is derived from dihydroxyacetone phosphate (DHAP), which is a metabolite in glycolysis. DHAP is converted to G1P by the stereospecific glycerol 1-phosphate dehydrogenase (G1PDH). Afterwards, the stereospecific geranylgeranylgeranyl phosphate synthase (GGGPS) condenses G1P and an isoprenoid, usually geranylgeranyl pyrophosphate (GGPP, C<sub>20</sub>), which produces (*S*)-3-*O*-geranylgeranylgeranyl phosphate (GGGP). A second isoprenoid chain is then added by the digeranylgeranylgeranyl phosphate synthase (DGGGPS) (Koga & Morii 2007).

These vast differences between the lipids of Archaea and Bacteria are summarized under the term “lipid divide” and it is precisely because of these why the emergence of the enzymes introducing them (G1PDH and GGGPS) is considered a branch point in early evolution of Archaea and Bacteria (Figure 1.3; Boucher 2007, Boucher et al 2004, Glansdorff et al 2008, Koga 2011, Koga & Morii 2007, Lombard et al 2012c, Payandeh & Pai 2007, Pereto et al 2004).

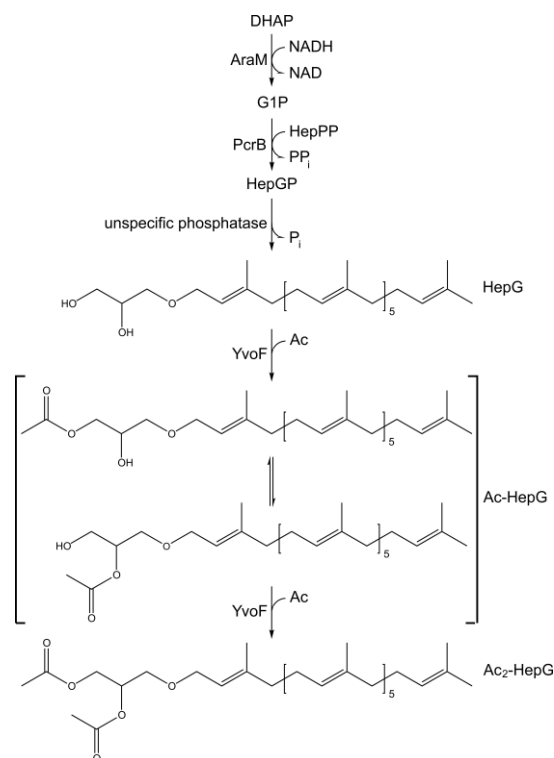


**Figure 1.3.** Separation of Archaea and Bacteria from the cenancestor due to the “lipid divide”. Separation was initiated by the emergence of glycerol 1-phosphate dehydrogenase (G1PDH) and geranylgeranyl glyceryl phosphate synthase (GGGPS) (marked as “\*”), being responsible for the unique membrane lipids. The dotted line represents a fictional timeline, which will not be further defined here, and the presentation ignores the question, which domain radiated first from the cenancestor. The figure is adapted from Koga et al (1998) and Payandeh & Pai (2007).

The cause for the “lipid divide” remains one of the fundamental unresolved questions in evolutionary biology. Nowadays, many theories hypothesize that Archaea and Bacteria must have evolved directly from the cenancestor, also called “last universal common ancestor” (LUCA) (Woese et al 1990, Woese et al 1978) and are therefore designated as primary domains (Koga et al 1998, Lange et al 2000). There is increasing evidence that LUCA was a complex organism containing many important genes and biosynthesis routes. However, there is disagreement about its membrane composition, with a few theories dominating. To begin with, some postulated that LUCA was acellular without a real membrane (Koga et al 1998). The most parsimonious and widespread theory however describes the LUCA with a heterochiral membrane consisting of G3P and G1P-based lipids together with fatty acids and isoprenoids (Lombard et al 2012b, Pereto et al 2004). During evolution a “lipid divide” event occurred, which separated Archaea from Bacteria by separating their respective lipid types. Many robust data indicate that LUCA was a hyperthermophilic organism (Boussau et al 2008), which provides an explanation for the “lipid divide” as archaea needed to maintain hyperthermophilic resistance. They adapted their membrane composition to do so by establishing the stable ether lipids as main membrane component. It was hypothesized that heterochiral membranes are in some way unstable, which might have driven the “lipid divide” towards homochiral membranes (Koga 2011, Wächtershäuser 2003). Meanwhile, research has shown that heterochiral membranes are stable

(Shimada & Yamagishi 2011) and in some cases even more stable with respect to environmental stress (Caforio et al 2018). But despite successful generation of mixed membranes in the lab, organisms with heterochiral membranes remained elusive. But recently, certain Euryarchaeota (e.g. the genus *Lokiarchaeum*) (Villanueva et al 2017) and bacteria (the Fibrobacteres-Chlorobi-Bacteroidetes (FCB) group) (Villanueva et al 2021) were described to harbor the genetic potential to synthesize archaeal ether-linked as well as bacterial/eukaryotic ester lipids. As a consequence from these findings, the “lipid divide” between the domains becomes less clear-cut than previously assumed, however, it supports the theory of a LUCA with heterochiral membranes. Since a selective advantage of homochiral membranes has not yet been discovered, there are now considerations that claim that it was not the lipids themselves, but the membrane proteins that are embedded in the lipid layers that led to the “lipid divide”. The reasoning behind this is that membrane proteins interact with the adjacent lipids and that mixed lipid surrounding would cause maladaptive protein-lipid interactions (Sojo 2019). However, this contradicts the recent tentative discovery of organisms with mixed membranes (Villanueva et al 2017, Villanueva et al 2021).

The uniqueness of the G1PDH and GGGPS to Archaea was irrefutable and G1P was assumed to not exist outside the archaeal domain in context of membrane lipids. It was only described in bacteria as part of lipoteichoic acids of Gram-positives (Neuhaus & Baddiley 2003, Taron et al 1983), in phosphatidylglycerols (Itabashi & Kuksis 1997) and in lipopolysaccharides (Perepelov et al 2006). Here, the special stereochemistry of G1P is only generated during synthesis of the entire molecule by “flipping”. Astonishingly, recent studies discovered G1PDH and GGGPS homologues in some bacteria as well. In *Bacillus subtilis*, *in silico* analysis suggested AraM to be the corresponding G1PDH-like enzyme, which was later confirmed *in vitro* and *in vivo*. Phylogenetic analyses showed a patchy distribution of AraM among Bacteria, indicating horizontal gene transfer (HGT) from the archaeal domain (Guldan et al 2008). Subsequently, the GGGPS-like enzyme, PcrB, was found in *B. subtilis* and characterized successfully as heptaprenylglyceryl phosphate synthase (HepGPS): It catalyzes condensation of G1P and C35 isoprenoid chains (heptaprenyl pyrophosphate, HepPP), resulting in the heptaprenylglyceryl phosphate (HepGP) ether lipid (Figure 1.4; Guldan et al 2011).

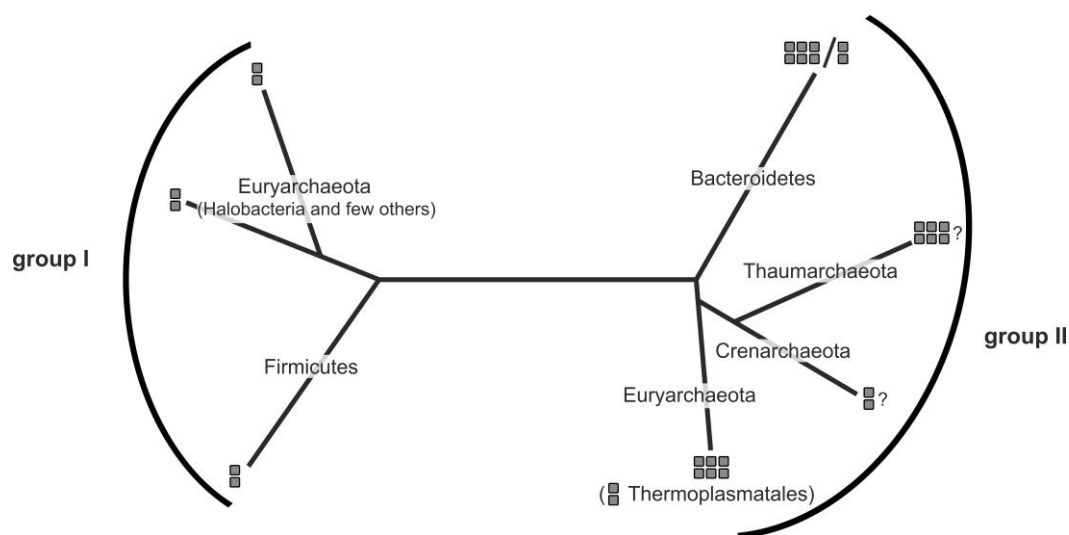


**Figure 1.4.** Biosynthesis of Archaea-type ether lipids in *B. subtilis*. Dihydroxyacetone phosphate (DHAP) is derived from glycolysis and processed by AraM (G1PDH of *B. subtilis*) to G1P. The GGGPS-like enzyme, PcrB, then condenses G1P and heptaprenyl pyrophosphate (HepPP), generating heptaprenylglyceryl phosphate (HepGP), which is afterwards dephosphorylated by unspecific phosphatases to heptaprenylglyceryl (HepG). The *O*-acetyltransferase YvoF then mono- (Ac-HepG) and diacetylates the lipid (Ac<sub>2</sub>-HepG). Ac, acetyl group; NAD, nicotinamide adenine dinucleotide; Pi, phosphate; PPi, diphosphate. The figure is adapted from Guldan et al (2011) and Linde et al (2016).

Unexpectedly, contrary to archaeal ether lipids, no second polyprenyl moiety is linked to HepGP, which is actually dephosphorylated and acetylated (Guldan et al 2011, Linde et al 2016). It is postulated that dephosphorylation is performed in an unspecific way, either by the alkaline phosphatase PhoB or by any other alkaline phosphatase, depending on their respective expression levels. For acetylation of HepG, YvoF has been identified to be the catalyzing enzyme (Figure 1.4). YvoF belongs to the *O*-acetyltransferases and is said to be peripherally membrane associated. This specific cellular localization could help to achieve substrate specificity and membrane incorporation of the product (Linde et al 2016).

### 1.1.2 Phylogenetic distribution of the GGGPS

After discovery of Archaea-type ether lipids (also referred to as “ether lipids”) in *B. subtilis*, comprehensive analysis was conducted to get insight into the phylogenetic distribution of GGGPS(-like) enzymes among Archaea and Bacteria. As expected, a GGGPS gene was found in almost all Archaea, except for Nanoarchaeota (Peterhoff et al 2014), which is coherent with the proposed key role of the GGGPS in the evolution of Archaea (Glansdorff et al 2008, Payandeh & Pai 2007). The GGGPS enzyme family is clustered into two separate phylogenetic groups, group I and group II (Boucher et al 2004, Guldan et al 2011, Nemoto et al 2003, Payandeh & Pai 2007). This was confirmed in a phylogenetic tree as the GGGPS enzyme family clearly disassembled into two distinct groups: Group I consisted of some Euryarchaeota and the Firmicutes and group II was assembled by most of the Archaea, namely Thaum-, Cren-, and Euryarchaeota, and the Bacteroidetes (Figure 1.5; Boucher et al 2004, Guldan et al 2011, Nemoto et al 2003, Payandeh & Pai 2007, Peterhoff et al 2014).

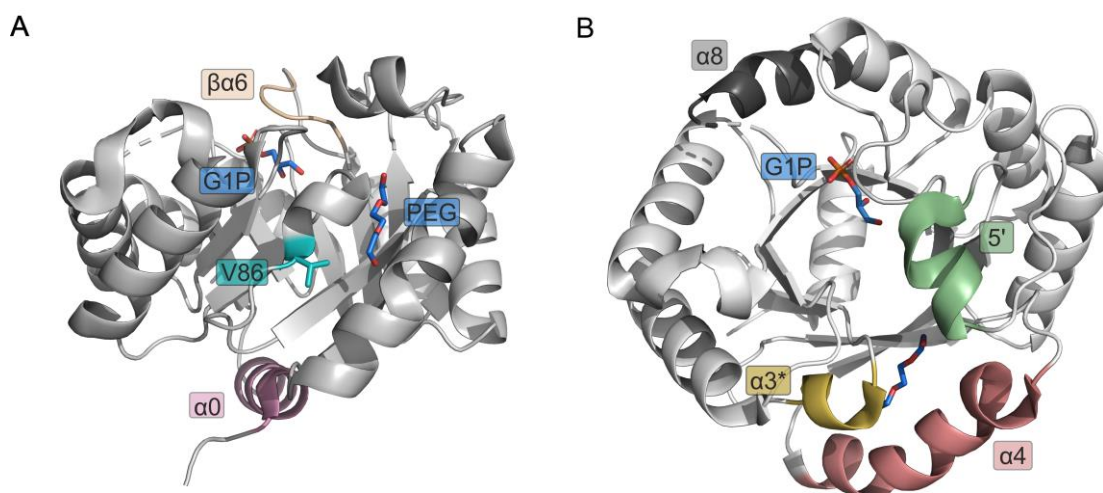


**Figure 1.5.** Schematic presentation of the phylogenetic tree of the GGGPS enzyme family. The enzyme family is separated into group I and group II, both containing Archaea and Bacteria. Oligomerization states of characterized GGGPS representatives of the respective phyla are indicated by gray rectangles (2 = dimeric / 6 = hexameric). Oligomerization states that were deduced from sequence analysis but are lacking *in vitro* confirmation are marked with “?”. The figure is adapted from Peterhoff et al (2014) and Kropp et al (2020).

Two characteristics stand out that distinguish between the groups. (i) The enzymes show differences in substrate specificity, as bacterial group I GGGPS enzymes accept longer polyprenyl substrates (C35), whereas archaeal group I and II enzymes metabolize polyprenyl substrates of C20 length. Somewhat unexpected do bacterial group II GGGPS enzymes not exhibit HepGPS (C35) but GGGPS (C20) activity (Peterhoff et al 2014). Length of the polyprenyl substrate is controlled by a so-called “hydrocarbon ruler”, which resembles a hydrophobic binding pocket



(Payandeh et al 2006, Ren et al 2013) and harbors a “limiter residue”, a bulky sidechain that is located at the end of the pocket. This bulky sidechain is placed differently in group I and II enzymes and additionally, whereas group I enzymes use aromatic residues as limiter, in group II this function is fulfilled by sterically challenging hydrophobic amino acids (Figure 1.6A; Guldan et al 2011, Peterhoff et al 2014). (ii) The second distinguishable characteristic is the oligomerization state. So far characterized group I enzymes have been shown to be exclusively dimeric, whereas group II was found to possess dimeric and hexameric representatives (Badger et al 2005, Blank et al 2020, Guldan et al 2011, Linde et al 2018, Nemoto et al 2003, Payandeh et al 2006, Peterhoff et al 2014, Ren et al 2013). *In vitro* identification of the oligomerization of GGGPS enzymes is still missing for some phyla, e.g. for Thaumarchaeota and Crenarchaeota. For Thaumarchaeota a hexameric and for Crenarchaeota a dimeric association state was predicted based on sequence analysis of conserved residues in the hexamerization interface (Figure 1.5; Peterhoff et al 2014).



**Figure 1.6.** Structural features of the GGGPS. *Methanothermobacter thermautotrophicus* GGGPS (mtGGGPS, PDB ID 4mm1) is shown crystallized with G1P and polyethylene glycol (PEG). The PEG is bound to the hydrophobic binding pocket for the polyprenyl substrate. (A) Side view with highlighted limiter residue (here valine, V86), helix  $\alpha 0$  and loop  $\beta\alpha 6$ . (B) Top view with highlighted  $\alpha 5'$ ,  $\alpha 3^*$  modification and helix  $\alpha 8$ .

### 1.1.3 The GGGPS – fold, structure, and stability

GGGPS enzymes belong to the highly diverse family of prenyltransferases and adopt an ancient and widespread fold, the so-called  $(\beta\alpha)_8$ -barrel or TIM-barrel fold, named after its first discovery in the triosephosphate isomerase (TIM) (Banner et al 1975, Sterner & Höcker 2005). The TIM-barrel fold of GGGPS enzymes was confirmed in multiple crystal structures (Blank et al 2020, Nemoto et al 2019, Payandeh et al 2006, Peterhoff et al 2014, Ren et al 2013, Ren et al 2012). The canonical TIM-barrel consists of eight units, each of which is composed of one  $\beta$ -strand and

an  $\alpha$ -helix (Sternier & Höcker 2005). The GGGPS TIM-barrel however shows some quite unique properties (Figure 1.6A and B). A “plug” is formed by an additional  $\alpha$ -helix, called  $\alpha_0$ , which is built by residues preceding sheet  $\beta_1$  at the N-terminus of the proteins. This “plug” is enriched in hydrophobic and basic amino acids and thus might play a role in membrane association (Payandeh et al 2006). Another modification is the substitution of helix  $\alpha_3$  through a flexible loop, called  $\alpha_3^*$ , which is assigned to function as a “swinging door” to allow binding of the hydrophobic polyprenyl substrate. The hydrophobic cavity for binding is further formed by an additional  $\alpha$ -helix, which is inserted between helix  $\alpha_4$  and helix  $\alpha_5$  and is referred to as helix  $\alpha_5'$  (Payandeh et al 2006, Ren et al 2013). Helices  $\alpha_4$  and  $\alpha_5'$  constitute the homo-oligomerization interface (Payandeh et al 2006, Peterhoff et al 2014, Ren et al 2013).

It has been shown that some GGGPS enzymes show two transition points in thermal denaturation experiments: The first transition point ( $T_1$ ) is associated with only partial unfolding or rearrangement of structure elements, that causes loss of enzymatic activity when the enzyme is heated above its  $T_1$ . The second transition point ( $T_2$ ) is linked to complete denaturation, which happens at high temperatures for most GGGPS variants, suggesting a stable TIM-barrel core (Linde et al 2018). This agrees with the emergence of the GGGPS in a (hyper-)thermophilic environment. Recent work has shown that the denaturing transition point ( $T_2$ ) is hardly influenced when disrupting a hexameric GGGPS to lower oligomerization states (dimers, monomers) by mutation. The additional transition point,  $T_1$ , however shifts quite drastically to lower temperatures. Molecular dynamics (MD) simulation implied that certain structure elements are quite flexible, causing this first thermal transition point. Especially helix  $\alpha_4$ , helix  $\alpha_5$ , loop  $\beta\alpha_6$  and helix  $\alpha_8$  (Figure 1.6) were affected by the high temperature used in the MD simulation (Linde et al 2018). Except for helix  $\alpha_8$ , these elements participate in substrate binding (Payandeh et al 2006, Peterhoff et al 2014, Peterhoff et al 2012). The reason behind this putative partial flexibility is however not yet completely understood. In general, (hyper-)thermostable proteins have to handle a flexibility – stability trade-off (Karshikoff et al 2015, Sternier & Liebl 2001), i.e., they need to compromise between high stability and sufficient flexibility for enzymatic function. The first is achieved by tight packing and many intramolecular interactions, which make the protein rigid. The second is important for fast substrate binding, catalysis and product release, which is why certain structure elements are kept more flexible. It is hypothesized that these elements are important for the fine-tuning of the flexibility – stability trade-off and might rearrange or partially unfold in heating experiments, causing the additional transition point ( $T_1$ ) in some GGGPS enzymes. Occurrence of the additional thermal transition has only been found for group II enzymes so far. It is not restricted to hexameric enzymes as it was also found in dimeric enzymes of mesophilic organisms, e.g. for the GGGPS from *Flavobacterium johnsoniae* (Bacteroidetes) (Linde et al 2018).

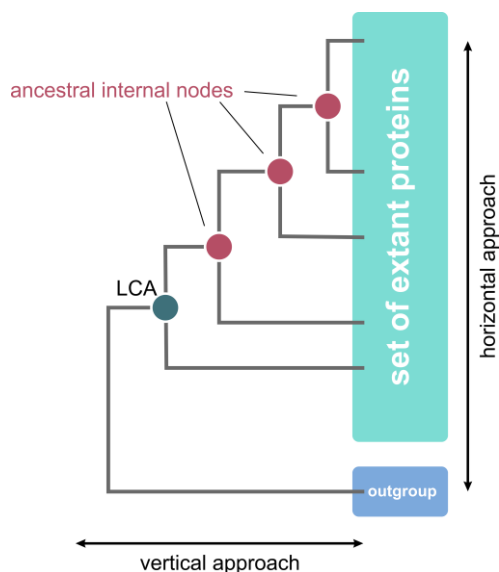
## 1.2 Protein design and ancestral sequence reconstruction

Enzymes are specialists in catalyzing reactions under their preferred conditions. To do so, they adopt characteristic secondary and tertiary structures, and association states. But are characteristics such as oligomerization or substrate specificity actually set in stone? And are there possibilities to understand certain properties such as this additional thermal transition point molecularly and evolutionarily? In fact, this is what protein engineering is made for. Protein engineering is a quite young discipline and describes the process of developing or manipulating proteins, which is quite valuable not only for research but also for industrial application. For instance, it is used to create proteins with increased stability regarding heat, pH, or organic solvents as in washing powder.

Protein engineering is essential for research as it improves understanding of proteins by clarifying the role of amino acid residues and entire structural units. Within its framework, a commonly used strategy is the so-called protein design. Protein design covers the whole spectrum, from the introduction of single mutations to the insertion or deletion of several residues in a protein sequence by rational means. Hence, it is inevitably linked to understanding protein structure and can target the modification of fold, catalytic function, target molecule binding, or protein-protein interaction. Consequently, creating a protein with novel traits involves a substantial amount of *a priori* rational input. For a long time, collecting *a priori* knowledge was done without computing, meanwhile, there are computational methods that support protein design tasks. The most famous software is Rosetta (Leaver-Fay et al 2011). It helps the researcher to acquire *a priori* knowledge and to evaluate introduced mutations. Rosetta can predict protein structures (DiMaio et al 2011) and it can evaluate by scoring structures (Alford et al 2017), or by calculating protein interactions in docking simulations (Sircar et al 2010). At the moment, protein design is still vaguely defined in literature. One definition that encompasses all attributes, is that it is a rational creation of proteins or enzymes with novel behavior, with the purpose to advance understanding of protein function (Marshall et al 2019). Most of the studies that are concerned with the design of proteins use modern enzymes as starting point. A widely used computational approach to increase the sequence space for protein design tasks is the so-called ancestral sequence reconstruction (ASR).

ASR is a sequence-based protein design approach. It was developed in the 1980s and is used to reconstruct predecessors of protein families (Hanson-Smith et al 2010, Thornton 2004). In brief, for ASR a set of modern protein sequences is acquired, which is then used to generate a multiple sequence alignment that supplies the basis for a phylogenetic tree. Based on that alignment, predecessors are calculated by means of a (e.g.) maximum likelihood approach for each residue position, creating a tree with internal nodes that represent the ancestors. The oldest one, the last common ancestor (LCA), can only be determined confidently when an outgroup is used, which is assembled of sequences that are less similar to the enzyme family of interest (Figure 1.7; Merkl

& Sterner 2016). Eventually, the sequences of predecessors “just” need to be cloned, and after gene expression, the proteins can theoretically be purified and characterized.



**Figure 1.7.** Schematic presentation of an ancestral sequence reconstruction (ASR) generated tree. Starting from a set of modern proteins (cyan), ancestral protein sequences are derived, which are represented as internal nodes (red). The oldest predecessor is called last common ancestor (LCA, dark green), which can only be determined through usage of an outgroup (blue).

ASR is to follow the evolution of a trait of interest and can theoretically be applied to all enzyme families with suitable sequence datasets available. Recent ASR-based studies dealt with thermostability and specificity of enzyme predecessors (Wheeler et al 2016), or the reductase chemistry of thioredoxin enzymes (Perez-Jimenez et al 2011). There is a plethora of reconstructions of enzyme families that provide a look at the evolution of certain characteristics, and in theory it is possible to go back to the time of LUCA by means of ASR. Its outcome is of course not infallible, but it does provide a clue.

As mentioned, ASR calculated predecessor enzymes can be candidates for protein design tasks, as they increase the sequence space for collecting *a priori* knowledge. Extant proteins only reflect a population at one point in time and comparison of these is called horizontal approach. ASR adds another dimension, as one can now compare proteins observed at different time points in evolution, which is called vertical approach (Figure 1.7). This can be quite beneficial as important residues can be identified rationally by analyzing mutations chronologically (Harms & Thornton 2010). This applies because a reconstructed sequence differs from the next one only by a few mutations. Predecessors often are easier to investigate as they are more thermostable than the modern ones (Wheeler et al 2016), which makes them interesting for industry, too.

### 1.3 Aim and scope of this work

A first aim of this work was to understand evolution of group II GGGPS enzymes in Archaea and Bacteria. For this purpose, ASR reconstructed progenitors were characterized and examined primarily for thermal stability ( $T_1$  and  $T_2$ ) and their oligomerization state in the context of evolution.

The oligomerization state and enzymatic activity were further targeted in the next project. The aim was to introduce a molecular switch into dimerized GGGPS enzymes in order to selectively induce hexamer formation by external input. This proof-of-concept study was set out to gain selective control over enzymatic activity.

Another goal was to collect spectrometric data about ether lipids in Bacteria as well as in Halobacteria (Archaea). To this end, analytic quantification of ether lipids in *B. subtilis* was conducted. Spectrometric methods were further intended to identify ether lipids in Bacteroidetes and acetylated diether lipids in Halobacteria.

### 1.4 Guide to the following chapters

Chapter 2 consists of the publication “**Hexamerization and thermostability emerged very early during geranylgeranylglyceryl phosphate synthase evolution**”, which deals with the evolution of group II GGGPS enzymes. Data indicate that hexameric oligomerization, as well as thermostability, might have evolved early in group II enzymes whereas the fine-tuning of the flexibility – stability trade-off, which occurs in some extant GGGPS enzymes, must have evolved late in evolution.

Chapter 3 constitutes the publication “**Controlling enzymatic activity by modulating the oligomerization state via chemical rescue and optical control**”. Here, an ancestral GGGPS variant as well as a modern enzyme were modified to control hexamerization and concomitantly enzymatic activity. For this, a crucial cation- $\pi$  bond in the hexamerization interface was exploited as molecular switch by establishing chemical rescue and optical control. By using both strategies, hexamerization and consequently enzymatic activity was controlled successfully upon external input.

Chapter 4 is assembled of non-published additional studies and a publication. “**Archaea-type ether lipids in Firmicutes and Bacteroidetes**” (section 4.1) describes experiments for first-time quantification of ether lipids in *B. subtilis* and the attempt to identify ether lipid derivatives in four Bacteroidetes strains, respectively. Analogously, *Haloferax volcanii* and *Halobacterium salinarum* were analyzed successfully for the occurrence of so far unknown acetylated diether lipids described in the publication “**Identification of acetylated diether lipids in halophilic Archaea**” (section 4.2).



## Chapter 2

### 2 Hexamerization and thermostability emerged very early during geranylgeranylgeranyl glyceryl phosphate synthase evolution

Cosimo Kropp<sup>[1]</sup>, Kristina Straub<sup>[1]</sup>, Mona Linde<sup>[1]</sup> and Patrick Babinger<sup>[1]\*</sup>

<sup>[1]</sup> Institute of Biophysics and Physical Biochemistry, Regensburg Center for Biochemistry, University of Regensburg  
93040 Regensburg, Germany

\*Corresponding Author: Patrick Babinger

E-mail: patrick.babinger@ur.de

Tel.: +49 941 943 1634; Fax: +49 941 943 2813

## 2.1 Abstract

A large number of archaea live in hyperthermophilic environments. In consequence, their proteins need to adopt to these harsh conditions, including the enzymes that catalyze the synthesis of their membrane ether lipids. The enzyme that catalyzes the formation of the first ether bond in these lipids, GGGPS, exists as a hexamer in many hyperthermophilic archaea, and a recent study suggested that hexamerization serves for a fine tuning of the flexibility – stability trade-off under hyperthermophilic conditions. We have recently reconstructed the sequences of ancestral group II GGGPS enzymes and now present a detailed biochemical characterization of nine of these predecessors, which allowed us to trace back the evolution of hexameric GGGPS and to draw conclusions about the properties of extant GGGPS branches that were not accessible to experiments up to now. Almost all ancestral GGGPS proteins formed hexamers, which demonstrates that hexamerization is even more widespread among the GGGPS family than previously assumed. Furthermore, all experimentally studied ancestral proteins showed high thermostability. Our results indicate that the hexameric oligomerization state and thermostability were present very early during the evolution of group II GGGPS, while the fine tuning of the flexibility – stability trade-off developed very late, independent of the emergence of hexamerization.

## 2.2 Statement for a broader audience

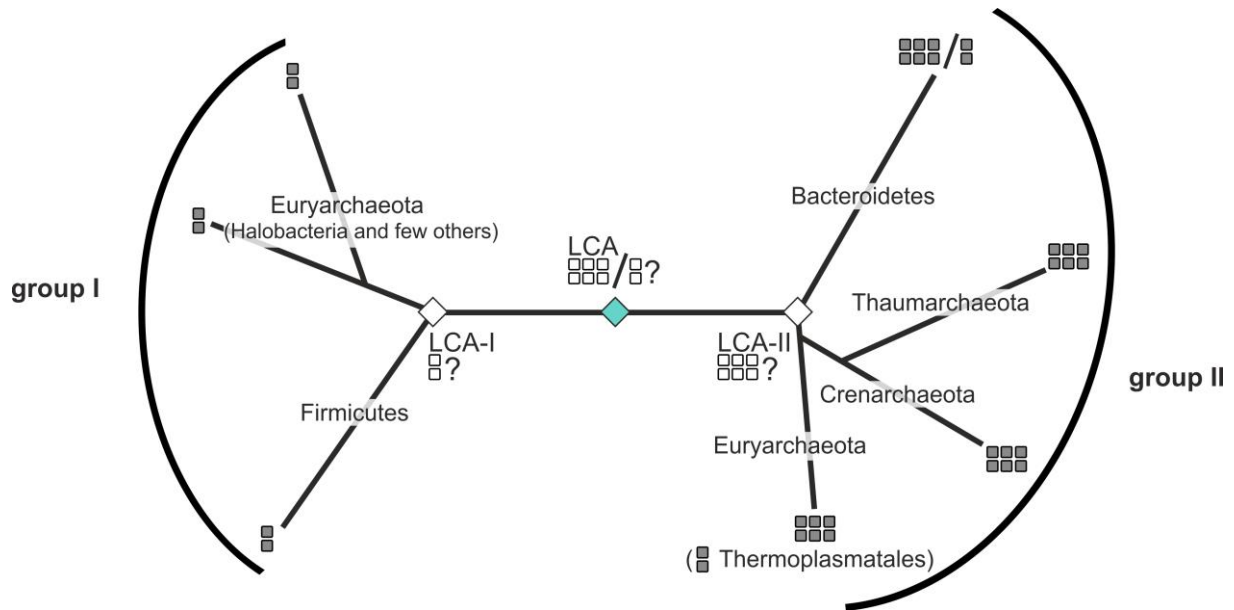
Organisms that live in hot environments need thermostable enzymes, but due to their rigidity, such enzymes frequently are less active. We have resurrected ancestral representatives of the GGGPS to study the emergence of its thermostability. We found that thermostability occurred very early in evolution, but only extant representatives of the enzyme show a flexibilization of the active site, which allows higher activity.



## 2.3 Introduction

The membrane lipids of Archaea have a quite different composition than those of Bacteria and Eukaryotes. While the latter consist of fatty acids that are linked to a G3P backbone by ester bonds, archaeal membrane lipids are ether lipids. Their backbone is a G1P, to which polyprenyl moieties are linked (Jain et al 2014). The formation of the first ether bond is catalyzed by the enzyme GGGPS, which usually transfers a geranylgeranyl residue (C<sub>20</sub>; four isoprene units) to the G1P. GGGPS is very interesting in the context of evolution, because it is a key enzyme in the context of this so-called “lipid divide” between the kingdoms of Archaea on the one hand, and Bacteria and Eukarya on the other (Villanueva et al 2017). Recent studies have unexpectedly revealed that GGGPS-like enzymes also occur in some bacteria. There, ether lipids with either four (Bacteroidetes) or seven (Firmicutes) isoprene units are synthesized (Guldan et al 2011, Guldan et al 2008, Peterhoff et al 2014), but their biological function is still unknown. In Firmicutes, they are further processed by acetylation (Guldan et al 2011, Linde et al 2016), while a recent study suggests that bacteria of the FCB group might synthesize prototypical archaeal ether lipids (Villanueva et al 2021).

The GGGPS enzyme family can be divided into two groups, both comprising archaeal and bacterial members (Figure 2.1). While the enzymes of Firmicutes and of a few Euryarchaeota, among them all Halobacteria, belong to group I, most Archaea and the Bacteroidetes possess a group II enzyme. In the recent years, the oligomerization states and structures of many GGGPS representatives have been experimentally determined (as summarized and referenced in Table S2.1). Up to now, only dimeric variants have been discovered within group I, while most group II enzymes have been described as homohexamers. Recently, we have shown that the  $(\beta\alpha)_8$ -barrel backbone of several GGGPS enzymes (mtGGGPS, tkGGGPS, taGGGPS, slGGGPS, fjGGGPS; for full names see Table S2.1) is structurally stable at hyperthermophilic conditions, independent of the oligomerization state. However, some GGGPS like mtGGGPS or fjGGGPS show a higher flexibility of the active site, going along which higher activity. Hexamerization stabilizes this portion of the protein, which otherwise would collapse at elevated temperatures (Linde et al 2018).



**Figure 2.1.** Simplified phylogenetic tree of the GGGPS enzyme family. The oligomerization states of a number of representatives have been experimentally identified (Table S2.1) and are indicated by 2 or 6 gray rectangles (dimeric / hexameric). The oligomerization states of Crenarchaeota and Thaumarchaeota were elucidated within this work. LCA, last common ancestor; LCA-I / LCA-II, LCA of group I / group II. Putative oligomerization states of the LCAs, as discussed within this work, are indicated by white rectangles.

The hexamer is formed as a “trimer of dimers”, where the conformation of the dimeric building blocks is identical to those of the natively dimeric GGGPS enzymes (Linde et al 2018, Peterhoff et al 2014). The conformation of the hexamer is stabilized by three individual contact interfaces between the protomers: The “dimer module interface” is symmetric and is present in dimeric GGGPS as well. The hexamer is stabilized by an additional symmetric interface between the dimer modules which we called “interconnecting interface”, plus an asymmetric interface that we called “ring interface”, because it stabilizes rings of three protomers each. As a hallmark of hexameric GGGPS, the presence of an aromatic amino acid (“aromatic anchor”) within the ring interface has been identified. It forms a cation- $\pi$  interaction with a lysine or arginine from the opposing protomer that is supposed to be essential for hexamerization. However, several group II sequences contain a histidine at the position of the aromatic anchor, particularly a large group of crenarchaeal sequences, which led to the assumption that they might be dimers (Peterhoff et al 2014). The corresponding proteins were not amenable to *in vitro* studies up to now, so that their true oligomerization state remained unclear.

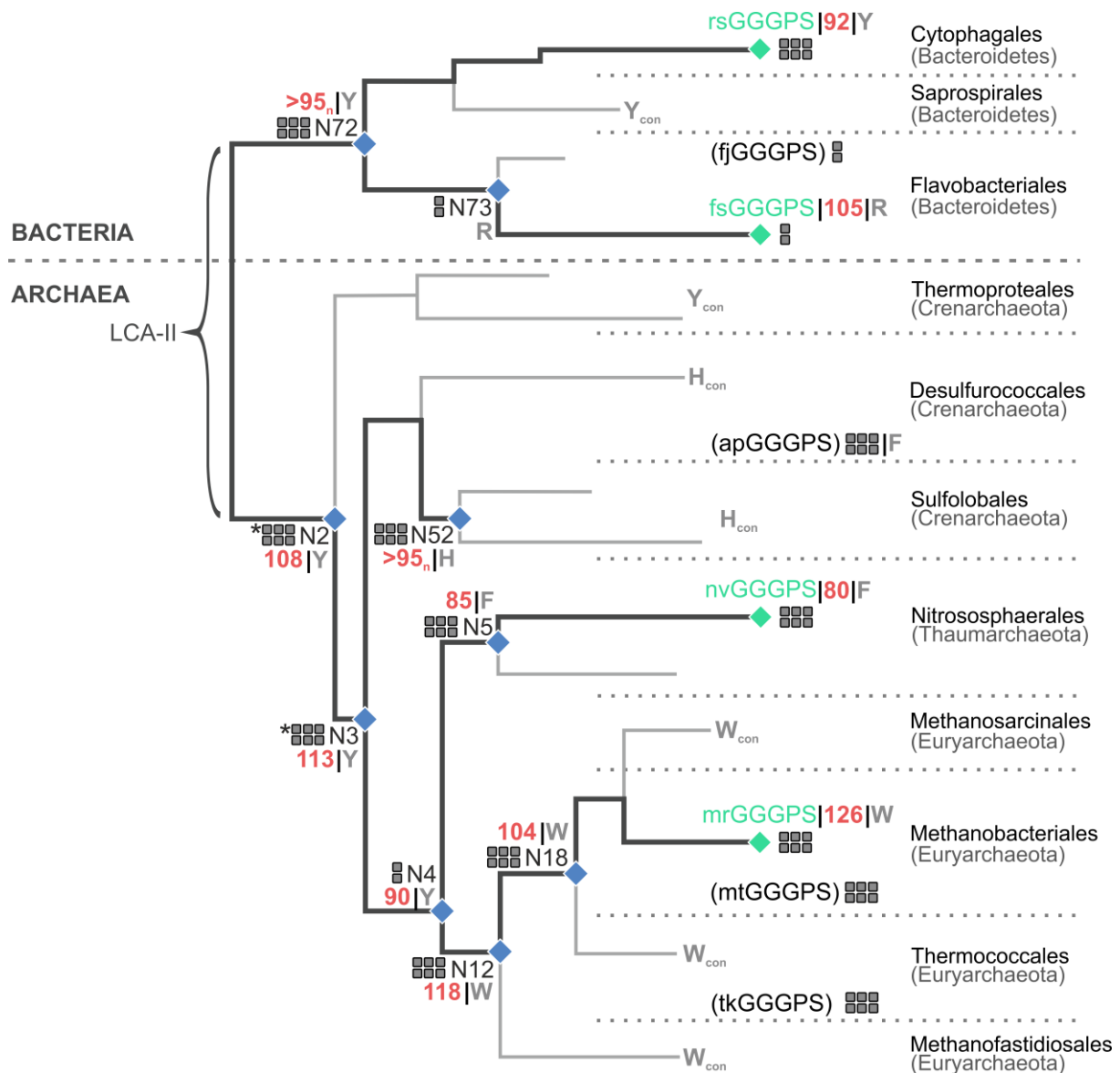
ASR is a powerful method to deduce environmental conditions in the Precambrian era and to study the evolution of proteins in terms of their structural stability as well as their interaction specificities with ligands and other proteins (Hochberg & Thornton 2017, Merkl & Sterner 2016, Pillai et al 2020). We recently have reconstructed the sequences of ancestral GGGPS enzymes (Straub et al 2019) and now provide a comprehensive *in vitro* characterization of nine of these

predecessors to trace back the evolution of hexameric GGGPS. We were able to successfully purify and study crenarchaeal predecessors with a His at the above mentioned “anchor position” and show here that they also form hexamers. Furthermore, we tested the ancestral GGGPS proteins for their thermostability. Our results indicate that hexamerization and thermostability were already present at the very beginning of group II GGGPS evolution, while the above-mentioned fine-tuning of active site flexibility developed late and independent of the emergence of hexamerization. Furthermore, we conclude that the few extant dimeric representatives among group II (Thermoplasmatales, Flavobacteriales) emerged by secondary events.

## 2.4 Results

### 2.4.1 Ancestral sequence reconstruction of group II GGGPS

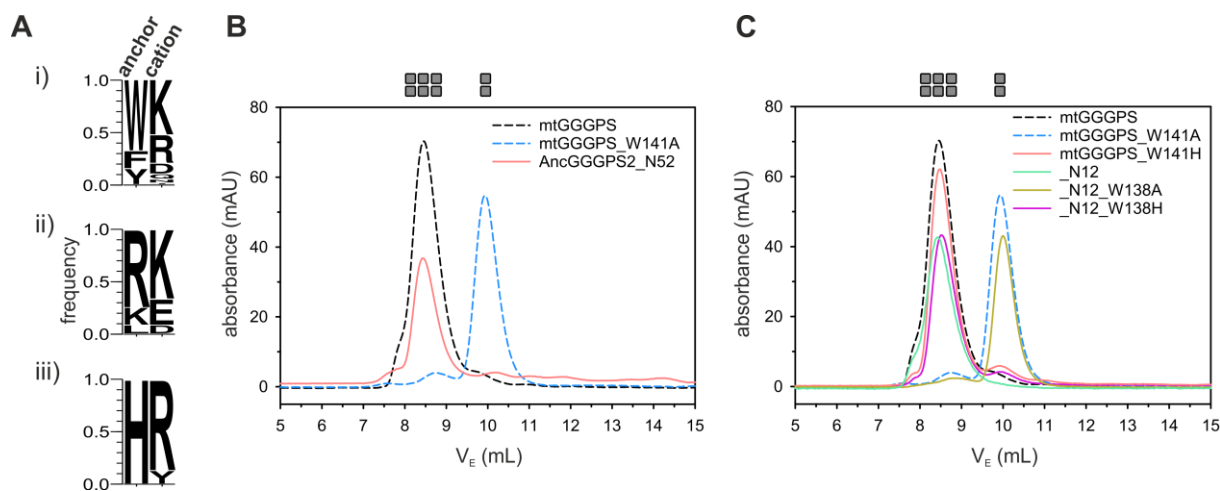
Ancestral sequence reconstruction of GGGPS has been performed to evaluate FitSS4ASR, a tool to prepare representative sequence sets for ASR, and is described in detail elsewhere (Straub et al 2019). We initially strived to perform ASR with a sequence set comprising both GGGPS groups I and II. However, this turned out to be not possible, because the length of the edge that interconnects the nodes representing the ancestors of groups I and II was too long ( $> 4$  substitutions per site) (Straub et al 2019). We then concentrated on the more interesting group II with its diverse oligomerization states. For practical reasons, all experiments presented here are based on the ASR from the manually curated sequence set as described earlier (Straub et al 2019), comprising 87 extant GGGPS group II variants. A simplified and annotated phylogenetic tree deduced from these sequences is shown in Figure 2.2; the complete tree with all nodes and leaves is shown in Figure S2.1. The names and properties of all GGGPS proteins used in this study are summarized in Table S2.1 and Table S2.1. We did not evaluate the properties of the LCA of all group II enzymes (LCA-II, N1), because it would require to use an outgroup to reconstruct a confident sequence for it. We tested several group I sequences as outgroup, but this resulted in non-robust trees. We use the same nomenclature to address the reconstructed proteins as introduced earlier (Straub et al 2019), “AncGGGPS2\_N..” stands for “ancestor of group II GGGPS” at node N.. .



**Figure 2.2.** Simplified phylogenetic tree used for ASR of GGGPS group II predecessors. Nodes representing ancestral sequences that were characterized experimentally are marked with blue dots and are labelled (N..). Leaves representing extant sequences that were characterized experimentally and that contributed to the calculation of the tree are shown with their names (cf. Table S2.1) and labelled in green. Protein names in brackets stand for representative proteins of the respective phyla which have been characterized previously (cf. Table S2.1), but whose sequences have not been used to calculate the tree. Oligomerization states of ancestral and extant proteins are indicated by gray rectangles, dimer-hexamer equilibria are marked by a \*. The identity of the amino acid located at the “aromatic anchor” position is given as a gray letter. For branches with no studied extant proteins the most conserved residue at that position is given (index con). In red numbers, the transition temperatures  $T_2$  of thermal denaturation ( $^{\circ}\text{C}$ ) are given. The  $T_2$  values result from differential scanning calorimetry (DSC) experiments (no index), or from nano differential scanning fluorimetry (nanoDSF) experiments (index n; Table 2.1). Branches that denote evolutionary paths with nodes representing GGGPS sequences that have been characterized within this study are shown in thick lines. Branches with nodes without experimental data are indicated by thin lines or are omitted for clarity. The full tree is shown in Figure S2.1. Close to the leaves, the names of the orders and phyla of the extant representatives are given.

#### 2.4.2 Probably all crenarchaeal GGGPS are hexamers

We have postulated in a previous study that an aromatic residue (“aromatic anchor”; Trp, Phe, Tyr) in the so called “ring interface” is a hallmark of hexameric GGGPS (Peterhoff et al 2014). This residue forms a cation- $\pi$  interaction with a lysine or arginine of the opposite protomer, and when the aromatic residue is mutated to an alanine, the hexamer falls apart into dimers. An analysis of group II GGGPS sequences by calculation of sequence similarity networks (SSN) revealed that in most sequence clusters this aromatic residue is highly conserved, except in two of them. One of the two clusters representing bacterial group II GGGPS (represented by the branch “Flavobacteriales” in Figure 2.2) has charged residues at this position (Figure 2.3A ii), and accordingly, extant representatives of this cluster (fjGGGPS, zpGGGPS; Table S2.1) have been demonstrated to be dimeric (Peterhoff et al 2014). The other sequence cluster without the aromatic residue is represented by the branches “Desulfurococcales” and “Sulfolobales” in Figure 2.2. In these sequences, a histidine is highly conserved at this position (Figure 2.3A iii). Extant proteins of these branches have not been accessible by heterologous expression up to now except hexameric apGGGPS, which is quite diverse in sequence and has a Phe at the anchor position (Peterhoff et al 2014). We now expressed synthetic genes of ancestral GGGPS and were successful in purifying a predecessor protein of the “Sulfolobales” branch, AncGGGPS2\_N52 (Figure 2.2, Table S2.1, Table S2.2), which also has a His at the anchor position. AncGGGPS2\_N52 displays high activity in an enzymatic assay (Figure S2.2) and shows a well-defined structure in circular dichroism (CD) spectroscopy (Figure S2.3A). Analytical size exclusion chromatography (SEC) revealed that AncGGGPS2\_N52 forms a hexamer (Figure 2.3B), indicating that His might replace the conventional aromatic residues Trp, Phe, Tyr as the “aromatic anchor”. In fact, His is less frequently observed in cation- $\pi$  interactions and usually rather plays the cation role in its protonated form, but it can also act as the aromatic motif in such interactions (Cauet et al 2005, Kumar et al 2018, Liao et al 2013). In the Sulfolobales sequences, the position of the cationic interaction partner is occupied by a conserved Arg (Figure 2.3A iii).



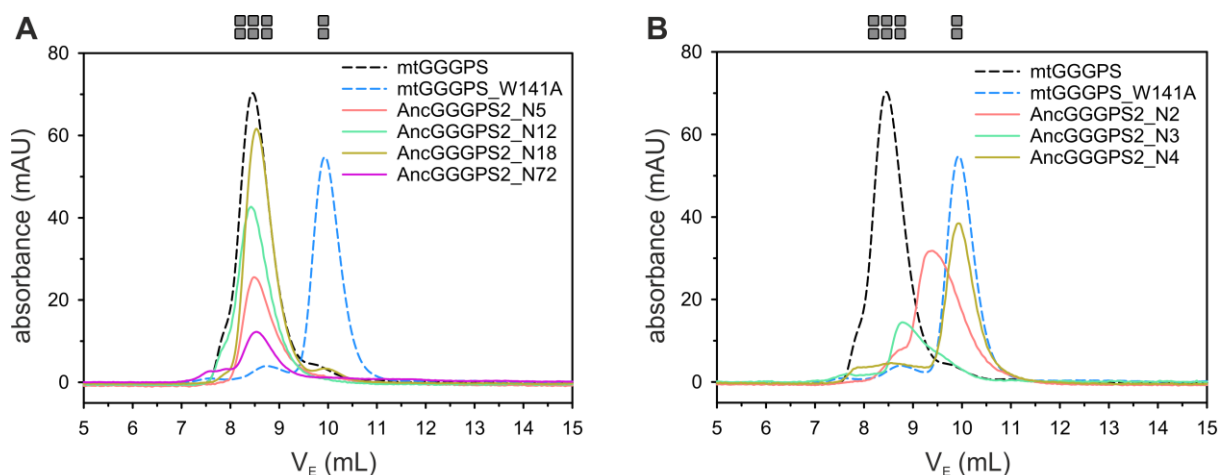
**Figure 2.3.** Analysis of GGGPS variants with histidine as “aromatic anchor”. Relative frequency of amino acid residues at the positions that form a cation- $\pi$  interaction in hexameric GGGPS variants. Anchor, “aromatic anchor” residue; cation, cationic counterpart. i) Sequences from all extant proteins used in ASR except Flavobacteriales, Sulfolobales and Desulfurococcales (presumed hexamers; Figure 2.2, Figure S2.1); ii) sequences from Flavobacteriales, iii) sequences from Sulfolobales and Desulfurococcales. (B) Analytical size exclusion chromatography (SEC) of AncGGGPS2\_N52. (C) Analytical SEC of the variants mtGGGPS\_W141H, AncGGGPS2\_N12, AncGGGPS2\_N12\_W138H and AncGGGPS2\_N12\_W138A. The proteins (40  $\mu$ M subunit concentration) were applied to a S75 analytical column equilibrated with 50 mM potassium phosphate, pH 7.5, 300 mM KCl. Elution was performed at a flow rate of 0.5 mL/min and followed by measuring the absorbance at 280 nm, which was plotted against the elution volume. mtGGGPS\_wt and mtGGGPS\_W141A (dashed lines) served as references for the hexameric and dimeric oligomerization states. The oligomerization state is indicated by gray symbols. The oligomerization state of AncGGGPS2\_N12 has already been determined in a previous study (Straub et al 2019).

We have previously shown for mtGGGPS that a mutation of the “aromatic anchor” to alanine (mtGGGPS\_W141A; serving as reference in Figure 2.3) breaks the hexamer down into dimers (Peterhoff et al 2014). To demonstrate that histidine does in fact play a similar role for hexamerization like the common aromatic residues do, we first introduced the corresponding His-to-Ala exchange into AncGGGPS2\_N52, but the resulting protein strongly tended to aggregation and could not be used for further experiments. As an alternative experiment, we changed the Trp in hexameric mtGGGPS\_wt to His. The resulting variant mtGGGPS\_W141H still eluted as a hexamer in analytical SEC (Figure 2.3C), which supports that the His can still form a cation- $\pi$  interaction with the opposing cation (a Lys in mtGGGPS). Additionally, we randomly chose the hexameric ancestral protein AncGGGPS2\_N12 (described later in detail) and introduced the W138H exchange, which again did not affect the oligomerization state (Figure 2.3C). In contrast, the hexamer was disrupted in the variant AncGGGPS2\_N12\_W138A. We conclude that histidine with its heterocyclic aromatic ring also functions as an “aromatic anchor” to stabilize the hexamer, and suspect that all extant sulfolobal GGGPS proteins are hexamers as well. In conclusion, almost all group II GGGPS enzymes must be considered to be hexameric, except the bacterial proteins from Flavobacteriales and a very small number of archaeal proteins, mainly

from Thermoplasmatales (Blank et al 2020, Nemoto et al 2003, Peterhoff et al 2014), as summarized in Table S2.1.

### 2.4.3 Hexamerization emerged very early during group II GGGPS evolution

To study the time point of emergence of hexamerization during group II GGGPS evolution, our rationale was to follow the evolutionary paths from the well-characterized extant proteins fjGGGPS (dimeric; Flavobacteriales) and mtGGGPS (hexameric; Methanobacteriales) to LCA-II (Figure 2.2). We expressed synthetic genes encoding seven ancestral GGGPS proteins in *E. coli*, purified the proteins and analyzed their oligomerization states by SEC (Figure 2.4). The results are summarized in Table S2.1 and Figure 2.2 (gray symbols). Additionally, activity assays (Figure S2.2) were performed, and the structural integrity of the proteins was analyzed by CD spectroscopy (Figure S2.3B). AncGGGPS2\_N5, \_N12, \_N18 and \_N72 turned out to be active proteins and hexameric (Figure 2.4A). AncGGGPS2\_N2 and \_N3 are also active proteins, but eluted at a volume between the dimeric and hexameric controls at the initially used protein concentration of 40  $\mu$ M (Figure 2.4B). SEC experiments with a series of different protein concentrations revealed a dimer-hexamer equilibrium for these variants with tendency to the hexamer under the given experimental conditions (Figure S2.4). Qualitatively, these variants can still be considered as hexamers, and they also possess all characteristics of the hexamer-specific contact interfaces like the “aromatic anchor” and the cation. A recent study on the robustness of reconstructed ancestral protein functions has revealed that quantitative markers of function - like dissociation constants - may vary, while qualitative conclusions on reconstructed protein functions are robust (Eick et al 2016). Only AncGGGPS2\_N4 elutes as a dimer in SEC (Figure 2.4B), although it also contains the “aromatic anchor” and the cation. Furthermore, this protein is inactive (Figure S2.2). As discussed below, this suggests that the dimeric nature of AncGGGPS2\_N4 might be an artifact due to sequence reconstruction. From the evolutionary point of view, there is no rationale why \_N4 should be a dimer, while \_N3, \_N12 and \_N5 are hexamers. In summary, these results indicate that hexamerization must have occurred very early in group II GGGPS evolution, but because we could not reconstruct a robust sequence for the group II LCA (LCA-II; as discussed above), we cannot clearly define whether LCA-II already has been hexameric or not (Figure 2.1).

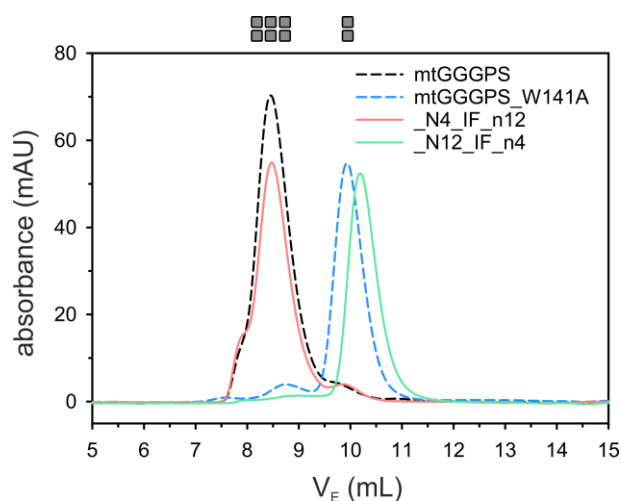


**Figure 2.4.** Analytical SEC of ancestral GGGPS proteins. (A) Hexameric proteins. (B) Proteins that are dimeric or form dimer-hexamer equilibria at the given concentration. The identities of the proteins are given in the legend. The proteins (40  $\mu$ M, subunit concentration) were applied to a S75 analytical column equilibrated with 50 mM potassium phosphate, pH 7.5, 300 mM KCl. Elution was performed at a flow rate of 0.5 mL/min and followed by measuring the absorbance at 280 nm, which was plotted against the elution volume. mtGGGPS\_wt and mtGGGPS\_W141A (dashed lines) served as references for the hexameric and dimeric oligomerization states. The oligomerization state is indicated by gray symbols. The oligomerization states of AncGGGPS2\_N4, \_N5 and \_N12 have already been determined in a previous study (Straub et al 2019).

To shed light on the impairments of AncGGGPS2\_N4, we studied this protein in more detail. We have shown previously that a single mutation (W141A) that disrupts the mtGGGPS hexamer, also drastically worsens the catalytic efficiency ( $k_{\text{cat}}/K_M$  (G1P) 380x lower,  $K_M$  (G1P) 50x higher), although the mutated position is not directly associated with the active site. Natively dimeric GGGPS like fjGGGPS are highly active, however (Linde et al 2018, Peterhoff et al 2014). A similar correlation between catalytic activity and oligomerization state of modified N4 variants would support the hypothesis that \_N4 has a misbalanced hexamerization interface due to artefacts in sequence reconstruction. Along these lines, we transplanted the ring interface of hexameric AncGGGPS\_N12 into AncGGGPS2\_N4 (Figure S2.5). The resulting variant AncGGGPS\_N4\_IF\_n12 eluted as a hexamer in analytical SEC (Figure 2.5) and showed GGGPS activity (Figure S2.2). As a reverse experiment, we transplanted the ring interface from AncGGGPS2\_N4 to AncGGGPS\_N12. The resulting variant AncGGGPS\_N12\_IF\_n4 eluted as a dimer in analytical SEC (Figure 2.5) and showed no GGGPS activity (Figure S2.2). Furthermore, isothermal titration calorimetry (ITC) experiments indicated that AncGGGPS2\_N4 cannot bind the substrate G1P any more (Figure S2.6), which is tantamount to the severely increased  $K_M$  (G1P) in mtGGGPS\_W141A. In summary, these results strongly support that the dimeric nature of AncGGGPS2\_N4 is an artefact from ASR due to a misbalanced ring interface. Furthermore, additional experiments indicate that the interplay of ring interface and interconnecting interface might be impaired as well in AncGGGPS2\_N4 (data not shown). We have previously shown that both interfaces cooperate to keep the hexamer stable



(Linde et al 2018). Ancestral sequence reconstruction, however, calculates each amino acid position independently from other positions. Therefore, mutual interactions of individual residues in an interface, and interactions of the different interfaces in context of oligomerization are difficult to reconstruct, which may occasionally lead to such artifacts. Such effects might also be the reason for the observed dimer-hexamer equilibrium in AncGGGPS2\_N2 and \_N3. In general, these considerations support the robustness of the reconstruction of the hexameric state of the ancestral GGGPS proteins. It is obvious that it is much more probable to disturb the interface balance and therefore break the hexamer, than to convert a dimer to a hexamer by undesired reconstruction artifacts.



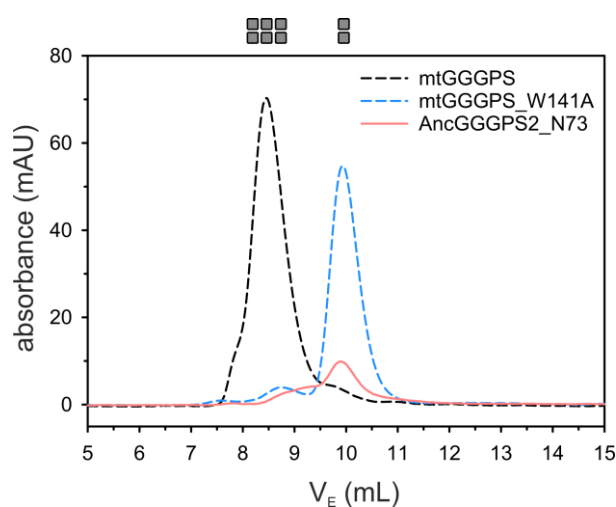
**Figure 2.5.** Analytical SEC of AncGGGPS2\_N4\_IF\_n12 and AncGGGPS2\_N12\_IF\_n4. The proteins (40  $\mu$ M, subunit concentration) were applied to a S75 analytical column equilibrated with 50 mM potassium phosphate, pH 7.5, 300 mM KCl. Elution was performed at a flow rate of 0.5 mL/min and followed by measuring the absorbance at 280 nm, which was plotted against the elution volume. mtGGGPS\_wt and mtGGGPS\_W141A (dashed lines) served as references for the hexameric and dimeric oligomerization states. The oligomerization state is indicated by gray symbols.

Additionally, we were able to purify and study the first extant protein from Thaumarchaeota, nvGGGPS, which shows a well-defined structure in CD spectroscopy, is enzymatically active, and is also a hexamer (Figure 2.1, Table S2.2, Figure S2.2, Figure S2.3D, Figure S2.7).

#### 2.4.4 Extant dimeric group II GGGPS are secondary dimers

To our knowledge, group II GGGPS enzymes from only two phyla have been reported to be dimeric based on experimental data (Table S2.2). These are the bacterial Flavobacteriales (fjGGGPS, zpGGGPS (Peterhoff et al 2014); fsGGGPS (this work)) and the archaeal Thermoplasmatales (taGGGPS (Nemoto et al 2019, Nemoto et al 2003, Peterhoff et al 2014);

tvGGGPS (Blank et al 2020)). To study the emergence of these dimeric group II GGGPS, we expressed synthetic genes for predecessors of the Flavobacteriales and analyzed the purified proteins. AncGGGPS2\_N72 is hexameric, as described above. In contrast, AncGGGPS2\_N73, the LCA of the Flavobacteriales GGGPS, eluted as a dimer in analytical SEC (Figure 2.6). The protein showed a well-defined structure in CD spectroscopy (Figure S2.3C), but only poor activity (Figure S2.2). This may indicate an impairment of this protein as discussed above for AncGGGPS2\_N4, but AncGGGPS2\_N73 has an Arg at the “aromatic anchor” position and a Lys at the position of the cationic partner, like most extant Flavobacteriales enzymes (Peterhoff et al 2014). This supports that the experimentally obtained dimeric oligomerization state of AncGGGPS2\_N73 is not an artifact.



**Figure 2.6.** Analytical SEC of AncGGGPS2\_N73. The proteins (22  $\mu$ M, subunit concentration) were applied to a S75 analytical column equilibrated with 50 mM potassium phosphate, pH 7.5, 300 mM KCl. Elution was performed at a flow rate of 0.5 mL/min and followed by measuring the absorbance at 280 nm, which was plotted against the elution volume. AncGGGPS2\_N73 tends to form aggregates and has a low extinction coefficient, resulting in a quite low peak. mtGGGPS\_wt and mtGGGPS\_W141A (dashed lines) served as references for the hexameric and dimeric oligomerization states. The oligomerization state is indicated by gray symbols.

Similarly, the sequences of Thermoplasmatales have an Arg at the “aromatic anchor” position. GGGPS sequences from Thermoplasmatales are missing in our ASR, because their somewhat diverse sequences were filtered out within the process of identification of wandering sequences to obtain robust ASR results (Straub et al 2019). An alternative ASR of GGGPS sequences including Thermoplasmatales has been published recently (Blank et al 2020). Based on computational data only, the authors postulate a convergent evolution for the hexamerization of group II GGGPS in the different phyla, which goes along with dimeric predecessors at the deep nodes in the phylogenetic tree. In contrast, our experimental results strongly support that hexamerization occurred very early in evolution, close to the group II LCA. Our data rather

implicate that the occurrence of dimeric GGGPS among the group II proteins is a secondary and convergent effect, because these proteins are direct descendants of hexameric predecessors (Figure 2.2). A HGT of a dimeric GGGPS from group I is very unlikely, because the Flavobacteriales and Thermoplasmatales GGGPS sequences are much more diverse from group I sequences than from any group II sequence. The biological reason for this change to a dimeric oligomerization state remains unclear, however, and needs further elucidation.

#### **2.4.5 Thermostability emerged very early during group II GGGPS evolution, flexibilization of the active site late**

Transition temperatures from thermal denaturation experiments (differential scanning calorimetry (DSC), nano differential scanning fluorimetry (nanoDSF)) of many extant GGGPS enzymes have been obtained in this and in a previous study (Figure 2.2, Table 2.1). Notably, the temperatures of complete denaturation ( $T_2$ ) are all in the hyperthermophilic (up to 127 °C) or at least thermophilic range, although there are proteins from mesophilic species among them (fjGGGPS, fsGGGPS, rsGGGPS, slGGGPS, nvGGGPS). We have previously found that some GGGPS proteins show an additional thermal transition in DSC or nanoDSF experiments ( $T_1$ ). This transition is hardly visible in CD spectroscopic analysis, indicating that it is not associated with a significant change of secondary structure, but goes along with enzymatic inactivation (Linde et al 2018). For some GGGPS proteins, these  $T_1$  temperatures are in the upper mesophilic range only. Our interpretation of these results was that the protein scaffold of all GGGPS proteins is very thermostable and rigid, whereas the active site region is destabilized in some GGGPS variants to allow a higher flexibility and therefore higher activity at lower temperatures (Linde et al 2018). A localized flexibilization of the active site within a relatively rigid protein scaffold is not unique to GGGPS, but has been frequently observed in enzymes from psychrophilic species to allow high turnover rates at low temperatures (D'Amico et al 2003, Socan et al 2020).

To investigate the emergence of thermostability among group II GGGPS, we have now determined transition temperatures of thermal denaturation of a large number of reconstructed ancestral and extant GGGPS (Figure 2.2, Table 2.1). Remarkably, the  $T_2$  temperatures of complete denaturation for all ancestral GGGPS proteins are in the hyperthermophilic range. An exception is AncGGGPS2\_N73, which tends to aggregation and where  $T_2$  could not be determined confidently. The representatives at the deepest nodes, AncGGGPS2\_N2 and AncGGGPS3\_N3 have  $T_2$  temperatures of 108 °C and 113 °C, respectively. This clearly supports a hyperthermophilic nature of the group II GGGPS LCA. The flexibilization of the active site, going along with the emergence of the  $T_1$  transition, obviously has emerged in some clades only. Besides the hyperthermophilic mtGGGPS and mrGGGPS, only the representatives from mesophilic Bacteroidetes (fsGGGPS, fjGGGPS) showed a  $T_1$  transition. Our results indicate that

the flexibilization emerged quite late in evolution, because none of the analyzed predecessors showed a  $T_1$  transition.

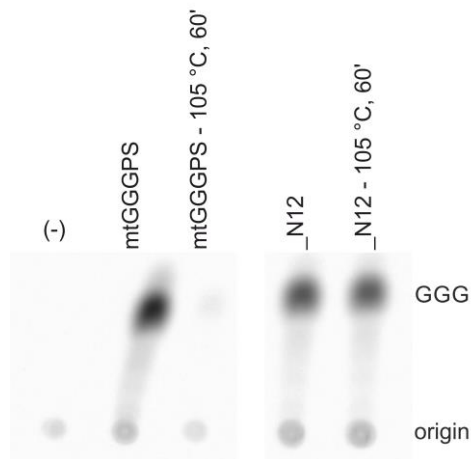
**Table 2.1.** Temperatures of thermal transitions of different GGGPS proteins as followed by nanoDSF and DSC.<sup>a</sup>

protein	nanoDSF		DSC	
	$T_1$ (°C)	$T_2$ (°C)	$T_1$ (°C)	$T_2$ (°C)
AncGGGPS2_N2	-	>95	-	107.8±0.1
AncGGGPS2_N3	-	>95	-	113.4±0.5
AncGGGPS2_N4	-	88.4±0.0	-	89.2±0.3
AncGGGPS2_N5	-	81.0±0.3	-	85.2±0.2
AncGGGPS2_N12	-	>95	-	118.4±0.2
AncGGGPS2_N18	-	>95	-	103.5±0.0
AncGGGPS2_N52	-	>95	n.d.	n.d.
AncGGGPS2_N72	-	>95	n.d.	n.d.
AncGGGPS2_N73		58.4±0.1 <sup>b</sup>	n.d.	n.d.
rsGGGPS	-	88.8±0.2	-	91.8±0.6
fsGGGPS	38.8±0.0	>95	39.6±0.3	104.9±0.3
mrGGGPS	>95	>95	96.4±0.1	126.3±0.5
nvGGGPS	-	78.2±0.1	-	80.3±1.4
mtGGGPS	>95	>95	102.5±0.5	127.0±0.5
tkGGGPS	n.d.	n.d.	-	115.0±0.0
taGGGPS	n.d.	n.d.	-	76.0±1.4
slGGGPS	n.d.	n.d.	-	83.1±0.9
fjGGGPS	53.2±0.0	>95	54.1±0.1	110.0±0.1

<sup>a</sup> The first block shows ancestral proteins, the second block proteins studied in this work, the raw data are shown in Figure S2.8. The third block shows proteins that have been investigated earlier (Linde et al 2018). The given temperatures are the mean of duplicates, standard deviations are given. Note that mtGGGPS shows two  $T_1$  transitions ( $T_1$ ,  $T_1^*$ ; (Linde et al 2018)). Only the transition leading to inactivation is given. -,  $T_1$  does not exist; n.d., not determined.

<sup>b</sup> The experimentally obtained data do not allow to decide whether the observed transition is due to a partial ( $T_1$ ) or complete denaturation ( $T_2$ ). See methods section for details.

To exclude that the reconstructed ancestral GGGPS proteins may not be active at elevated temperatures due to reconstruction artefacts, we analyzed the activity of the most stable ancestral GGGPS in our study, AncGGGPS2\_N12 ( $T_2 = 118\text{ }^\circ\text{C}$ ; no  $T_1$ ). One of its direct descendants is the well-studied mtGGGPS ( $T_1 = 102.5\text{ }^\circ\text{C}$ ,  $T_2 = 127\text{ }^\circ\text{C}$ ), which showed an apparent half life time of 16.5 min at  $100\text{ }^\circ\text{C}$  in heat inactivation studies (Linde et al 2018). We incubated both mtGGGPS and AncGGGPS2\_N12 for 60 min at  $105\text{ }^\circ\text{C}$  under pressure in the DSC instrument and tested their activity before and after this incubation step (Figure 2.7). mtGGGPS completely lost its activity under these conditions, while the apparent activity of AncGGGPS2\_N12 was unchanged. This means that AncGGGPS2\_N12 is even more thermostable than mtGGGPS and supports that all ancestral group II GGGPS enzymes may have been active at extreme temperatures



**Figure 2.7.** Heat inactivation of mtGGGPS and AncGGGPS2\_N12. Proteins were incubated for 60 min at  $105\text{ }^\circ\text{C}$ . Heat-treated and untreated samples ( $1\text{ }\mu\text{M}$  subunit concentration) were incubated with the substrates  $^{14}\text{C}$ -G1P and GGPP for 1.5 h at  $40\text{ }^\circ\text{C}$ . The generated products were extracted, dephosphorylated to allow for subsequent separation by thin layer chromatography (TLC), and visualized by autoradiography. As negative control (-), no enzyme was added. The origin of the chromatography and the reaction product geranylgeranylgeranyl (GGG) are marked.

## 2.5 Discussion

Recent studies have shown that many extant representatives of group II GGGPS proteins have a hexameric oligomerization state. A key element to stabilize the hexamer is a cation- $\pi$  interaction in the so-called ring interface, usually formed by a Trp, Tyr or Phe together with a Lys or Arg at the opposing protein surface (Linde et al 2018, Peterhoff et al 2014). A large number of Crenarchaeota, however, have a conserved His at the position of the aromatic moiety, and their GGGPS proteins were not accessible to experiments up to now. Using ancestral sequence reconstruction, we have now characterized a large number of GGGPS predecessors with respect to their oligomerization state, thermostability and activity. Our results demonstrate that the crenarchaeal proteins form hexamers, too. This means that almost all group II GGGPS, and consequently the GGGPS of almost all archaea, have an hexameric oligomerization state. Furthermore, almost all GGGPS predecessors form hexamers, among them those at the deepest nodes of the ASR tree. This makes it very likely that hexameric GGGPS developed close to the emergence of the group II LCA (LCA-II; Figure 2.1), which itself could not be reconstructed.

In contrast to the group II GGGPS proteins, all studied representatives of group I proteins are reported to be dimers. An exception is hsGGGPS, which eluted as a monomer in SEC, but unpublished results support that this might be an artefact due to the high salt requirements of this halobacterial protein that could not be satisfied during SEC analysis (Peterhoff et al 2014). Because of the divergence of group I and group II GGGPS sequences, we were not able to calculate a combined ASR tree of groups I and II and their common ancestor sequence (Straub et al 2019). However, the exclusive presence of dimers among group I makes it very likely that the LCA of group I (LCA-I; Figure 2.1) also was a dimer. Under these assumptions (dimeric LCA-I, hexameric LCA-II), the LCA of all GGGPS proteins might have been either dimeric or hexameric. The analysis of large datasets of oligomeric structures has revealed as a common rule that the larger an oligomerization interface is, the earlier it has emerged during the evolution of the complex (Levy et al 2008, Venkatakrisnan et al 2010). Furthermore, symmetric interfaces are more likely to emerge during evolution, because they are easier to generate (Villar et al 2009). In GGGPS, the symmetric dimer module interface is by far the largest (1163 Å<sup>2</sup> in mtGGGPS), while the key interface for hexamerization, the ring interface, is much smaller (256 Å<sup>2</sup> in mtGGGPS) and asymmetric. These facts and the principle of parsimony speak for a dimeric nature of the LCA, and the need for thermostabilization may have driven the evolution towards hexamerization in group II GGGPS by a cyclic trimerization of the dimeric building block. We assume that the few dimeric group II GGGPS like in Flavobacteriales and Thermoplasmatales have emerged convergently as secondary events, because they descend from hexameric predecessors.

Ancestral sequence reconstruction also allowed us to shed light on the evolution of thermostability among group II GGGPS. All reconstructed GGGPS predecessors are highly thermostable. As extreme example, AncGGGPS2\_N12 only denatures at temperatures above

118 °C and is still fully active after 60 min incubation at 105 °C. We have recently presented data which suggest that hexamerization of GGGPS supports thermostability in a rather indirect manner (Linde et al 2018). Even as a monomeric module, the GGGPS protein scaffold is highly thermostable in terms of the temperature of complete denaturation. However, such rigid enzymes tend to be less catalytically active (Karshikoff et al 2015). Similarly to what has been observed for cold-adapted enzymes to allow high turnover rates at low temperatures (D'Amico et al 2003, Socan et al 2020), we have previously found a localized flexibilization of the active site of some GGGPS representatives, which goes along with two transitions in thermal denaturation experiments. The first transition is associated with loss of activity and collapse of the active site, while the second transition marks the complete denaturation of the protein. We could demonstrate that hexamerization elevates the temperature of the first transition, and we assumed that oligomerization serves for a fine tuning of the balance between flexibility and stability in hexameric GGGPS (Linde et al 2018). Ancestral sequence reconstruction now revealed that the flexibilization of the active site, apparent as the occurrence of two transitions in thermal denaturation experiments, obviously emerged quite late in evolution, because only extant GGGPS enzymes showed this phenomenon. This means that the primary stimulus for the emergence of hexamerization was not the need to stabilize the active site, but must have been different, such as a general need for thermostability. The flexibilization of the active site has obviously emerged independently of hexamerization in convergent processes, because it is also present in dimeric fjGGGPS. Its stabilization by hexamerization must rather be regarded as a secondary benefit, that made use of the already existing higher oligomeric structure. Consequently, ancestral GGGPS enzymes must have used another still unknown, probably less effective way to balance catalytic activity and stability, like some extant dimeric GGGPS from hyperthermophilic species still do.

## 2.6 Materials & Methods

### 2.6.1 Ancestral sequence reconstruction of GGGPS group II sequences

Ancestral sequence reconstruction (ASR) has been described in detail elsewhere (Straub et al 2019). Briefly, the 87 GGGPS group II sequences that have been manually selected by using several filter algorithms in order to obtain high posterior probabilities on the edges of the tree, branch lengths  $< 1.0$  mutations per site and a species diversity as high as possible. Filter algorithms as they are the basis of FitS4ASR (Straub et al 2019) were used in a manual manner which identify wandering sequences by both RogueNaRok (Aberer et al 2013) and the described heatmap approach. The resulting sequences are from Crenarchaeota (23 sequences), Euryarchaeota (40 s.), Thaumarchaeota (8 s.), and Bacteroidetes (16 s.). The tree needed for reconstruction was computed by using PhyloBayes (Lartillot et al 2009), utilizing a site-heterogeneous CAT model. Four independent MCMC samplings of length 50,000 were launched to ensure convergence. After discarding the first 6000 trees as burn-in, the remaining trees of two MCMC chains - chosen by the quality criteria maximum discrepancy ( $< 0.1$ ; 0.01699) and minimum effective size ( $> 100$ ; 2769) (score for a good run; received score) - were concatenated to deduce a consensus tree. By means of FastML (Ashkenazy et al 2012), the ancestral sequences were reconstructed with the substitution model JTT and a gamma distribution. The parameter “probability cutoff”, which helps to prefer ancestral indel over character, was set to 0.8 in order to adjust the length of the reconstructed sequences to the length of the recent sequences, compensating a bias towards longer than true ancestors (Vialle et al 2018). The most probable sequence was determined for each internal node of the tree. A FASTA file with all extant and ancestral sequences as well as a file containing the posterior probabilities for all sites in the ancestral sequences are provided as supplementary material.

### 2.6.2 Cloning and site-directed mutagenesis

As summarized in Table S2.2, twelve ancestral GGGPS variants and four extant GGGPS proteins were produced and purified. Gene sequences for the reconstructed GGGPS variants (AncGGGPS2\_N2, AncGGGPS2\_N3, AncGGGPS2\_N4, AncGGGPS2\_N5, AncGGGPS2\_N12, AncGGGPS2\_N18, AncGGGPS2\_N52, AncGGGPS2\_N72, AncGGGPS2\_N73, AncGGGPS2\_N4\_IF\_n12, AncGGGPS2\_N12\_IF\_n4) and extant GGGPS proteins of *Roseivirga seohaensis* (rsGGGPS), *Flavobacterium saliperosum* (fsGGGPS), *Methanobrevibacter ruminatum* (mrGGGPS) and *Nitrososphaera viennensis* (nvGGGPS) were optimized in their codon usage for expression in *E. coli* and ordered as GeneArt™ Strings™ DNA fragments from Thermo Fisher Scientific. BsaI restriction sites were incorporated at the 5' and 3'-ends of the DNA fragments to allow cloning into a modified pET21a expression vector (Rohweder et al 2018), providing a C-terminal hexahistidine (His)<sub>6</sub> tag. AncGGGPS2\_N73 was cloned into a modified pMAL-c5T expression vector (Rohweder et al 2018), providing a C-



terminal maltose binding protein (MBP) fusion with a thrombin cleavage site. Other variants of AncGGGPS2\_N12 and mtGGGPS were generated by QuickChange mutagenesis (Zheng et al 2004) with oligonucleotides as given in Table S2.3. All constructs were verified by sequencing. The wild-type mtGGGPS and the mtGGGPS\_W141A variant have been cloned previously into pET21a (Peterhoff et al 2014). The sequence numbering of mtGGGPS used in this study refers to EMBL ENA entry AAB85058 and pdb-id 4mm1, which have an N-terminal three amino acid extension compared to UniProt entry O26652.

### 2.6.3 Production and purification of recombinant proteins

Heterologous gene expression was performed in the *E. coli* strain BL21-Gold(DE3) (Agilent Technologies). The transformed cells were grown at 37 °C in LB containing ampicillin (150 µg ml<sup>-1</sup>). When OD<sub>600</sub> reached 0.6-0.8, expression was induced by adding 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG). Growth was continued overnight at 20 °C, the cells were harvested by centrifugation, resuspended in 50 mM potassium phosphate, pH 7.5, 300 mM KCl, 10 mM imidazole and disrupted by sonication. The His-tagged proteins were purified from the clarified cell extract by metal chelate affinity chromatography. An ÄKTApurifier system with a HisTrap FF crude column (5 ml, GE Healthcare) was used, and a linear gradient of imidazole (10–500 mM) in 50 mM potassium phosphate, pH 7.5, 300 mM KCl was applied to elute the protein. Interfering imidazole and salt were removed from the purified proteins by dialysis against 50 mM potassium phosphate, pH 7.5 at 4 °C. The dialysis buffer for rsGGGPS additionally contained 300 mM KCl. For the MBP fused construct, the dialysis buffer was supplemented with thrombin (2 U mL<sup>-1</sup> final concentration) for cleavage. MBP and thrombin were removed by subsequent preparative SEC on a Highload<sup>TM</sup> 26/600 Superdex<sup>TM</sup> S75 pg column (GE Healthcare). The column was equilibrated with 50 mM potassium phosphate, pH 7.5, 300 mM KCl and was run at a flow rate of 1.5 ml min<sup>-1</sup>. Protein concentrations were determined by absorbance spectroscopy in the UV range; if the extinction coefficient of the variant was too low, the Bradford assay was used. The molar extinction coefficients  $\epsilon_{280}$  and the molecular weight (MW) were calculated from the amino acid sequence by means of ProtParam (Gasteiger et al 2005). Purified protein was dropped into liquid nitrogen and stored at -80 °C. If high protein concentrations were needed, the samples were concentrated by ultrafiltration (Amicon Ultra-15, mwco 10 kDa, Merck Millipore). The purification yields of all proteins are listed in Table S2.2, purity is shown in Figure S2.9.

#### 2.6.4 Characterization of oligomerization state of recombinant proteins

All variants were characterized by SEC experiments. Standard SEC experiments were performed on a Superdex S75 column (GE Healthcare), which was operated with 50 mM potassium phosphate, pH 7.5, 300 mM KCl at a flow rate of 0.5 ml min<sup>-1</sup>. 100 µl of protein with a subunit concentration of 22-40 µM was applied. All SEC runs were performed at room temperature in an air-conditioned room (approx. 22 °C).

#### 2.6.5 CD spectroscopy

Spectra of 6 µM protein (subunit concentration) were recorded at a scan rate of 200 nm min<sup>-1</sup> in degassed 50 mM potassium phosphate, pH 7.5 with a response time of 0.5 sec from 190 to 260 nm in a JASCO J-815 circular dichroism spectrometer (d = 0.1 cm) at 25 °C.

#### 2.6.6 DSC

40 µM protein (subunit concentration) was heated in degassed 50 mM potassium phosphate, pH 7.5 from 30 °C to 130 °C at a ramp rate of 1 K min<sup>-1</sup> in a VP-DSC differential scanning microcalorimeter (MicroCal, Malvern Instruments) with fixed reference and sample cell (0.511 mL each). The change in heat capacity with raising temperature was recorded. The proper equilibration of the calorimeter was ascertained by performing several buffer-buffer baselines. Overpressure was applied to prevent boiling above 100 °C. Where possible, DSC transitions were baseline corrected and fitted with a non-two state model using the software supplied by the manufacturer. The apparent midpoint temperature ( $T_{\text{Mapp}}$ ) of the irreversible unfolding transition was determined as an operational measure of protein stability. Experiments were done in duplicates.

#### 2.6.7 nanoDSF

40 µM protein (subunit concentration) was heated in degassed 50 mM potassium phosphate, pH 7.5 from 20 °C to 95 °C at a ramp rate of 1 K min<sup>-1</sup> in a Prometheus NT.48 instrument (NanoTemper Technologies GmbH; access provided by 2bind GmbH). The excitation power at 280 nm was 20-40%. Emission spectra were measured at 330 and 350 nm. The change in the ratio of the fluorescence signal at 350 nm to 330 nm with raising temperature was followed. Fluorescence transitions were fitted by the program supplied by the manufacturer and the apparent midpoint temperature ( $T_{\text{Mapp}}$ ) of the irreversible unfolding transition was determined as an operational measure of protein stability. The experiments were done in duplicates.

### 2.6.8 Classification of thermal transitions as $T_1$ (partial denaturation) or $T_2$ (complete denaturation)

The existence of two individual thermal transitions, where the first (at temperature  $T_1$ ) is associated with a loss of enzymatic activity, but with a partial denaturation of the protein only, while the second (at a higher temperature  $T_2$ ) goes along with a complete denaturation, has been unambiguously demonstrated by a combination of DSC, nanoDSF and CD-spectroscopic experiments as well as enzymatic activity tests for the two proteins mtGGGPS and fjGGGPS (Linde et al 2018). In DSC experiments, the proteins are heated up to 130 °C. If only a single transition is visible, it is therefore very likely that this transition goes along with a complete denaturation of the protein and can be addressed as a  $T_2$  transition. This has been demonstrated experimentally by experiments with taGGGPS and slGGGPS (Linde et al 2018). In nanoDSF experiments, however, a complete denaturation cannot be achieved for highly thermostable proteins, because this would exceed the upper temperature limit of the instrument (95 °C). This makes it difficult to conclude whether a transition observed in nanoDSF is due to a partial ( $T_1$ ) or a complete ( $T_2$ ) denaturation, if only a single transition is visible. Although a decrease of the  $F_{350}/F_{330}$  ratio points to a  $T_1$  transition in most cases and an increase of the ratio to a  $T_2$  transition (Linde et al (2018) and this study), we performed DSC experiments wherever possible to clarify the situation. For the three proteins AncGGGPS2\_N52, \_N72 and \_N73, DSC experiments were not possible due to technical reasons. In case of AncGGGPS2\_N52 and \_N72, no transition could be observed in nanoDSF. We assume that the  $T_2$  transition for these proteins is above 95 °C. In case of AncGGGPS2\_N73, a transition was observed in nanoDSF at 58.4 °C with a decrease in  $F_{350}/F_{330}$  ratio, which would speak for a  $T_1$  transition, but the protein strongly tended to aggregation and was difficult to handle. Therefore it is impossible to decide whether this transition is due to a partial or complete denaturation.

### 2.6.9 Radiometric GGGPS activity assays

$^{14}\text{C}$ -G1P was synthesized as described by Guldan et al (Guldan et al 2011). To test the activity of purified GGGPS enzymes, 1  $\mu\text{M}$  of protein (subunit concentration) was incubated with 20  $\mu\text{M}$  GGPP and 20  $\mu\text{M}$   $^{14}\text{C}$ -G1P (302 nCi) in 10 mM  $\text{MgCl}_2$ , 0.2 % Tween80, 50 mM Tris-HCl, pH 8.0, in a total volume of 100  $\mu\text{L}$  for 1.5 h at 40 °C. The products were dephosphorylated by adding 1 U calf intestinal alkaline phosphatase (CIP) (New England Biolabs) for 30 min at 40 °C and extracted according to the method of Bligh and Dyer (Bligh & Dyer 1959) as modified by Kates (Kates 1986). The solvent was evaporated to dryness in a rotary evaporator and the products were analyzed by thin layer chromatography (TLC) on Silica 60 plates, developed in ethyl acetate/hexane 1:1 (v/v), and visualized with a phosphorimager system (PerkinElmer Life Sciences).

### 2.6.10 Heat inactivation

For observing residual activity after irreversible thermal inactivation, the isoscan function of the VP-DSC differential scanning microcalorimeter (MicroCal, Malvern Instruments) was used with a subsequent radiometric GGGPS activity assay. All enzymes were adjusted to a subunit concentration of 40  $\mu\text{M}$  in degassed 50 mM potassium phosphate, pH 7.5. The enzymes were incubated at 105 °C for 60 min in the DSC instrument to allow exact temperature control of the sample. The samples were chilled on ice and centrifuged at 16.100 g for 5 min at 4 °C. The residual activity of the supernatant was assayed in the presence of geranylgeranyl pyrophosphate (GGPP) and G1P in a radiometric GGGPS activity assay, as described above.

### 2.6.11 ITC

To follow G1P binding to GGGPS variants, a MicroCal PEAQ-ITC microcalorimeter (Malvern Instruments) with a cell volume of 280  $\mu\text{L}$  was used. Degassed protein (100  $\mu\text{M}$  subunit concentration) and G1P (1 mM) solutions were prepared from the identical buffer batch (50 mM Tris, pH 8.0, 10 mM  $\text{MgCl}_2$ ). The G1P solution was titrated to the protein in 2  $\mu\text{L}$  aliquots for a total of 18 injections at 2.5 min intervals at 25 °C. During the course of the titration, the reaction mixture was continuously stirred. As controls, titrations of buffer with buffer, protein solution with buffer and buffer with ligand solution were performed. Each titration experiment was baseline corrected, and the change of heat per injection was calculated by integrating the area under each peak using the Origin software provided by MicroCal. The experimentally observed signals for the ligand binding experiment were corrected for the signals of the control experiments, a  $\Delta\text{H}$  versus molar ratio plot was generated and a fit curve was plotted with the “single set of identical sites” option provided by the software.

## 2.7 Acknowledgements

We thank 2bind GmbH for access to the Prometheus NT.48 instrument (NanoTemper Technologies). We are grateful to Christiane Endres, Sonja Fuchs, Sabine Laberer and Jeannette Ueckert for technical assistance. We thank Rainer Merkl and Reinhard Sterner for helpful discussions and critical reading of the manuscript.

## 2.8 Supporting Information

A FASTA file with all extant and ancestral sequences as well as a file containing the posterior probabilities for all sites in the ancestral sequences are provided on the journal’s website. The files can be found at <https://onlinelibrary.wiley.com/doi/full/10.1002/pro.4016>.

**Table S2.1.** Experimentally determined oligomerization states of GGGPS enzymes.

GGGPS group	species	protein name	method	oligomerization state	reference
I	<i>Halobacterium salinarum</i>	hsGGGPS	SEC, SLS	(Monomer) <sup>a</sup>	(Peterhoff et al 2014)
	<i>Archaeoglobus fulgidus</i>	afGGGPS	Crystal structure, SEC, SLS	Dimer	(Guldan et al 2011, Payandeh et al 2006)
	<i>Bacillus subtilis</i>	bsPcrB	Crystal structure, SEC, SLS	Dimer	(Badger et al 2005, Guldan et al 2011, Peterhoff et al 2014, Ren et al 2013)
	<i>Geobacillus kaustophilus</i>	gkPcrB	Crystal structure, SEC, SLS	Dimer	(Guldan et al 2011, Peterhoff et al 2014)
	<i>Staphylococcus aureus</i>	saPcrB	Crystal structure, SEC, SLS	Dimer	(Guldan et al 2011, Peterhoff et al 2014, Ren et al 2013)
	<i>Listeria monocytogenes</i>	lmGGGPS	SEC	Dimer	(Guldan et al 2011, Peterhoff et al 2014)
II	<i>Roseivirga seohaensis</i>	rsGGGPS	SEC	Hexamer	This work (Figure S2.7)
	<i>Flavobacterium saliperosum</i>	fsGGGPSa	SEC	Dimer	This work (Figure S2.7)
	<i>Methanobrevibacter ruminantium</i>	mrGGGPS	SEC	Hexamer	This work (Figure S2.7)
	<i>Nitrososphaera viennensis</i>	nvGGGPS	SEC	Hexamer	This work (Figure S2.7)
	<i>Methanothermobacter thermautotrophicus</i>	mtGGGPS	Crystal structure, SEC, SLS	Hexamer	(Linde et al 2018, Peterhoff et al 2014)
	<i>Flavobacterium johnsoniae</i>	fjGGGPS	Crystal structure, SEC, SLS	Dimer	(Peterhoff et al 2014)
	<i>Zunongwangia profunda</i>	zpGGGPS	SEC, SLS	Dimer	(Peterhoff et al 2014)
	<i>Chitinophaga pinensis</i>	cpGGGPS	SEC, SLS	Hexamer	(Peterhoff et al 2014)
	<i>Spirosoma linguale</i>	slGGGPS	SEC, SLS	Hexamer	(Peterhoff et al 2014)

<i>Thermococcus kodakaraensis</i>	tkGGGPS	SEC, SLS	Hexamer	(Peterhoff et al 2014)
<i>Aeropyrum pernix</i>	apGGGPS	SEC, SLS	Hexamer	(Peterhoff et al 2014)
<i>Thermoplasma acidophilum</i>	taGGGPS	Crystal structure, SEC, SLS	Dimer	(Peterhoff et al 2014) (Nemoto et al 2003) (Nemoto et al 2019) (Blank et al 2020)
<i>Thermoplasma volcanium</i>	tvGGGPS	Crystal structure	Dimer	(Blank et al 2020)

Anc II	AncGGGPS2_N2	SEC	Hexamer	This work (Figure 2.4)
	AncGGGPS2_N3	SEC	Hexamer	This work (Figure 2.4)
	AncGGGPS2_N4	SEC	Dimer	(Straub et al 2019)
	AncGGGPS2_N5	SEC	Hexamer	(Straub et al 2019)
	AncGGGPS2_N12	SEC	Hexamer	(Straub et al 2019)
	AncGGGPS2_N18	SEC	Hexamer	This work (Figure 2.4)
	AncGGGPS2_N52	SEC	Hexamer	This work (Figure 2.3)
	AncGGGPS2_N72	SEC	Hexamer	This work (Figure 2.4)
	AncGGGPS2_N73	SEC	Dimer	This work (Figure 2.6)

<sup>a</sup> The apparent monomeric oligomerization state of hsGGGPS is very likely an artifact due to the high salt requirements of this protein that could not be provided during SEC analysis (Peterhoff et al 2014). The GGGPS protein from *Halogeometricum borinquense* (HbGGGPS-I, UniProt entry E4NSJ9) is currently under our investigation and elutes as a dimer in SEC/SLS experiments in 500 mM NaCl (data not shown).

**Table S2.2.** GGGPS variants characterized in this study with their purification yields after expression in *E. coli*.

species	protein name	yield (mg/L culture volume)
	AncGGGPS2_N2	4
	AncGGGPS2_N3	37
	AncGGGPS2_N4	7.1
	AncGGGPS2_N5	2.5
	AncGGGPS2_N12	22.6
	AncGGGPS2_N18	27.6
	AncGGGPS2_N52	1.9
	AncGGGPS2_N72	0.7
	AncGGGPS2_N73	0.08
	AncGGGPS2_N4_IF_n12	59.7
	AncGGGPS2_N12_IF_n4	45.4
	AncGGGPS2_N12_W138H	61
	AncGGGPS2_N12_W138A	41.5
<i>Roseivirga seohaensis</i>	rsGGGPS	13
<i>Flavobacterium saliperosum</i>	fsGGGPS	3.5
<i>Methanobrevibacter ruminantium</i>	mrGGGPS	3.5
<i>Nitrososphaera viennensis</i>	nvGGGPS	39.5
<i>Methanothermobacter thermautotrophicus</i>	mtGGGPS <sup>a</sup>	5
<i>Methanothermobacter thermautotrophicus</i>	mtGGGPS_W141A <sup>a, b</sup>	41 <sup>b</sup>
<i>Methanothermobacter thermautotrophicus</i>	mtGGGPS_W141H	18.1

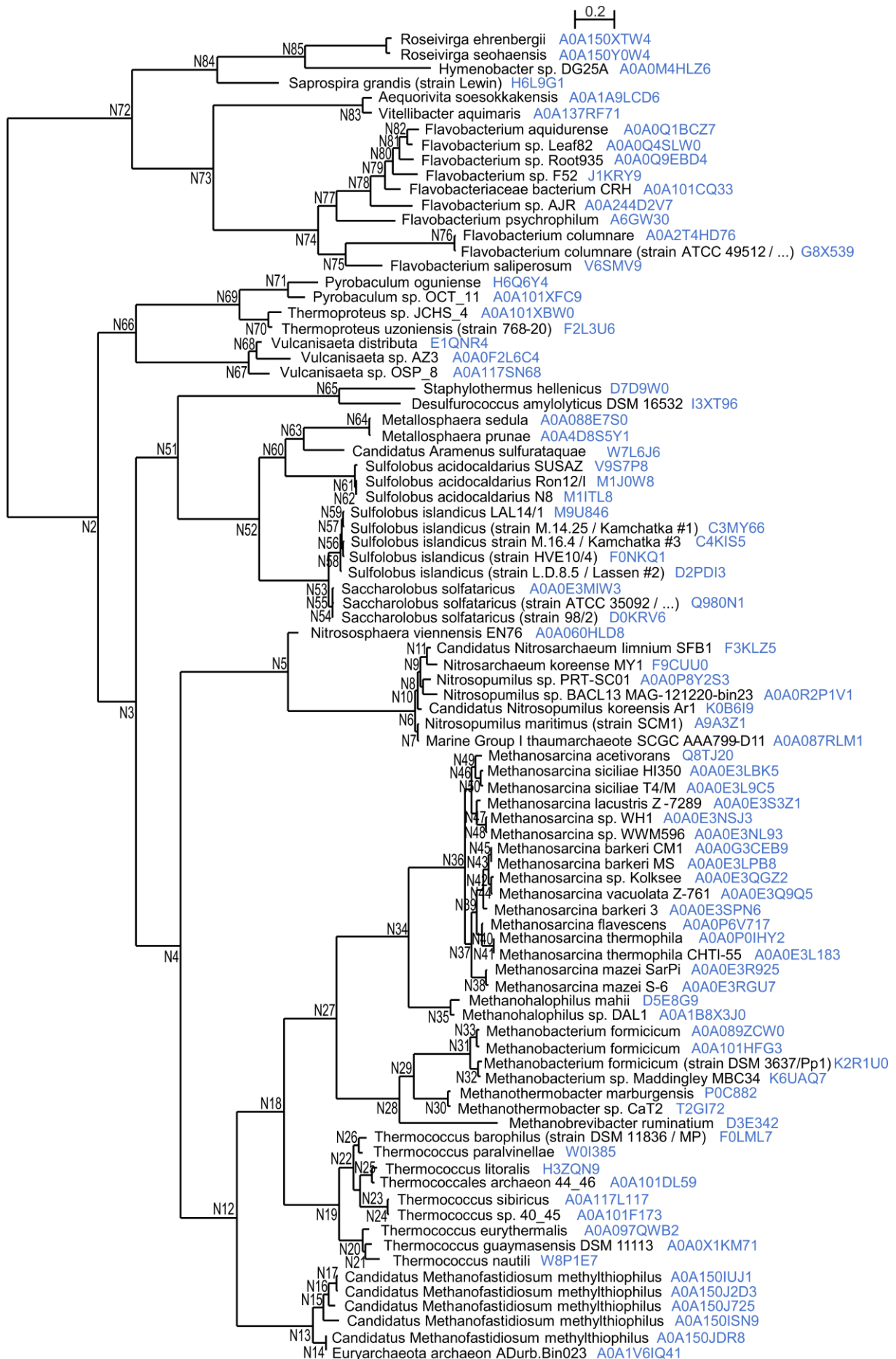
<sup>a</sup> The genes of these variants have been cloned previously (Peterhoff et al 2014).

<sup>b</sup> This variant has been purified previously (Peterhoff et al 2014).

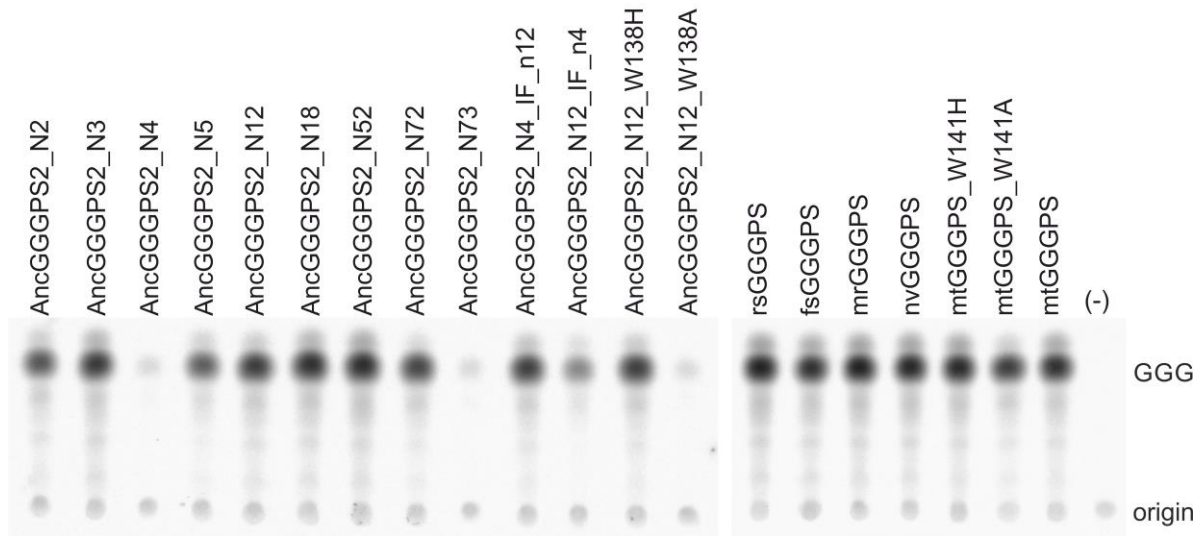
**Table S2.3.** Primer sequences used for site-directed mutagenesis.

variant	5'-primer	3'-primer
AncGGGPS2_N12_W138H	ACGTTGGTGATGCAAAACCTATTCCG	GACCAACGGTTTCACCCGGTTC
AncGGGPS2_N12_W138A	GCGGTTGGTGATGCAAAAC	ACCAACGGTTTCACCCG
mtGGGPS_W141H	CACGTCGGTGACACCAAGC	TCCAACCGTACCCCCTG

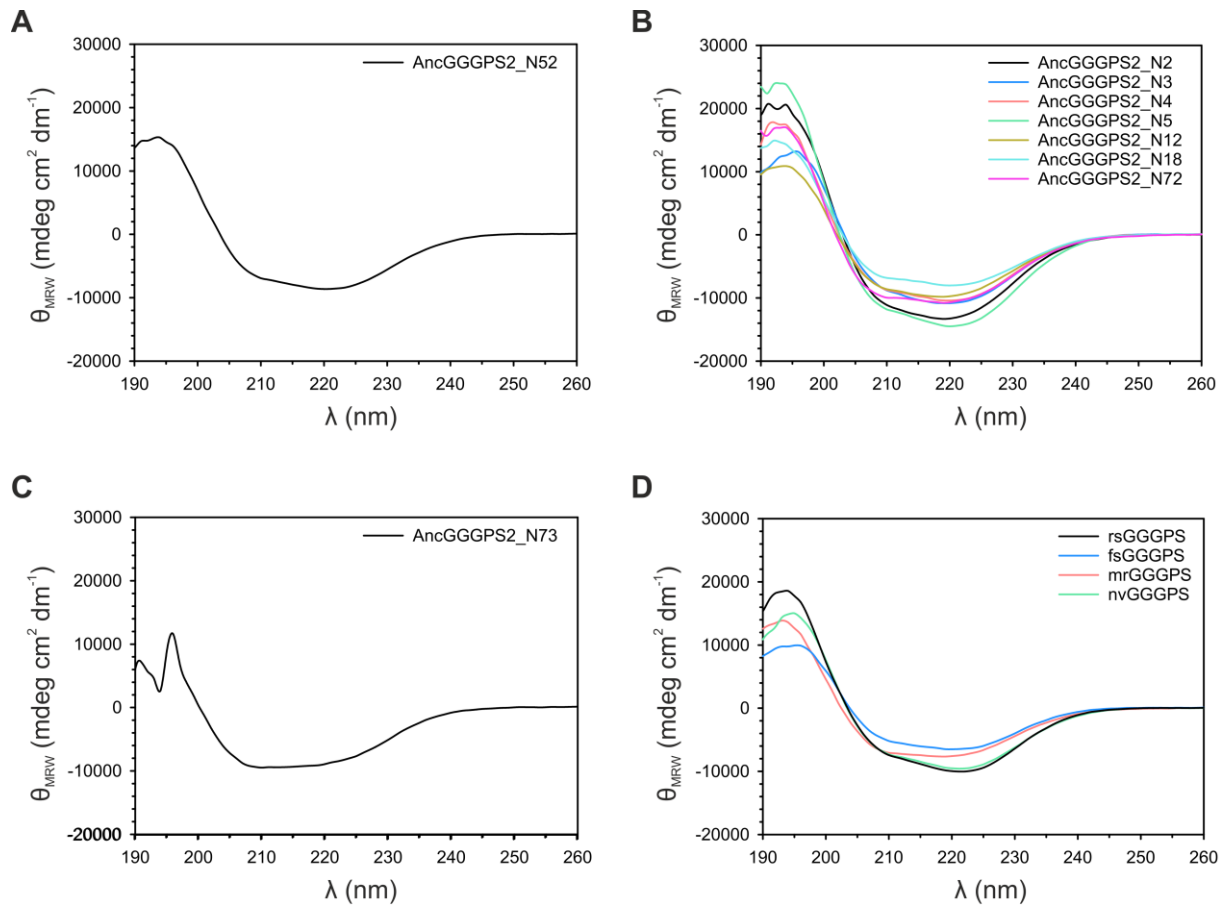




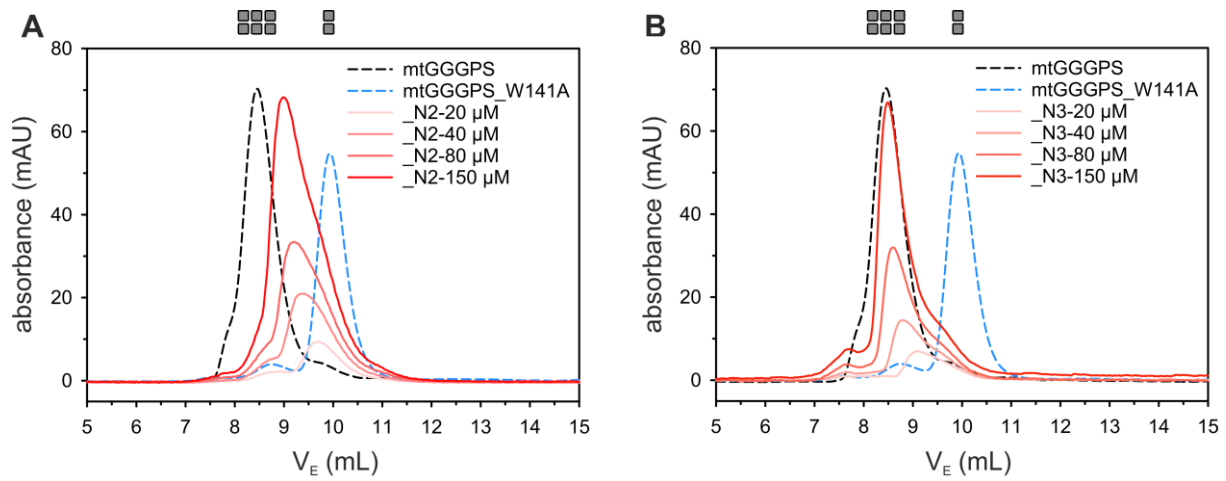
**Figure S2.1.** Phylogenetic tree deduced from the manually curated sequence set used for ASR of GGGPS group II predecessors. PhyloBayes (Lartillot et al 2009) was used to calculate the tree. The length of the horizontal bar corresponds to 0.2 substitutions per site. Nodes "N.." correspond to evolutionary intermediates. The tree has been calculated in a previous study (Straub et al 2019). The full species and strain names as well as UniprotKB accession numbers are given for each extant sequence. A FASTA file with all extant and ancestral sequences as well as a file containing the posterior probabilities for all sites in the ancestral sequences are provided as supplementary material.



**Figure S2.2.** Activity assay of GGGPS variants used in this work. The substrate  $^{14}\text{C}$ -G1P was incubated with GGPP and  $1\ \mu\text{M}$  of the GGGPS variant (subunit concentration) for 1.5 h at  $40\ ^\circ\text{C}$ . The generated products were extracted, dephosphorylated to allow for subsequent separation by TLC, and visualized by autoradiography. As negative control (-), no enzyme was added to the assay. The origin of the chromatography and the reaction product GGG are marked. The samples were run on different TLC plates each, as indicated by the vertical white gap.



**Figure S2.3.** Structural integrity of GGGPS variants characterized in this study. Far-UV CD spectra of ancestral and extant GGGPS variants (6  $\mu\text{M}$ , subunit concentration) were recorded in 50 mM potassium phosphate, pH 7.5 from 190 nm to 260 nm ( $d = 1$  mm) at room temperature. The CD spectra are sorted in the order they are mentioned in the main text (A-D). The identities of the proteins are given in the legends.



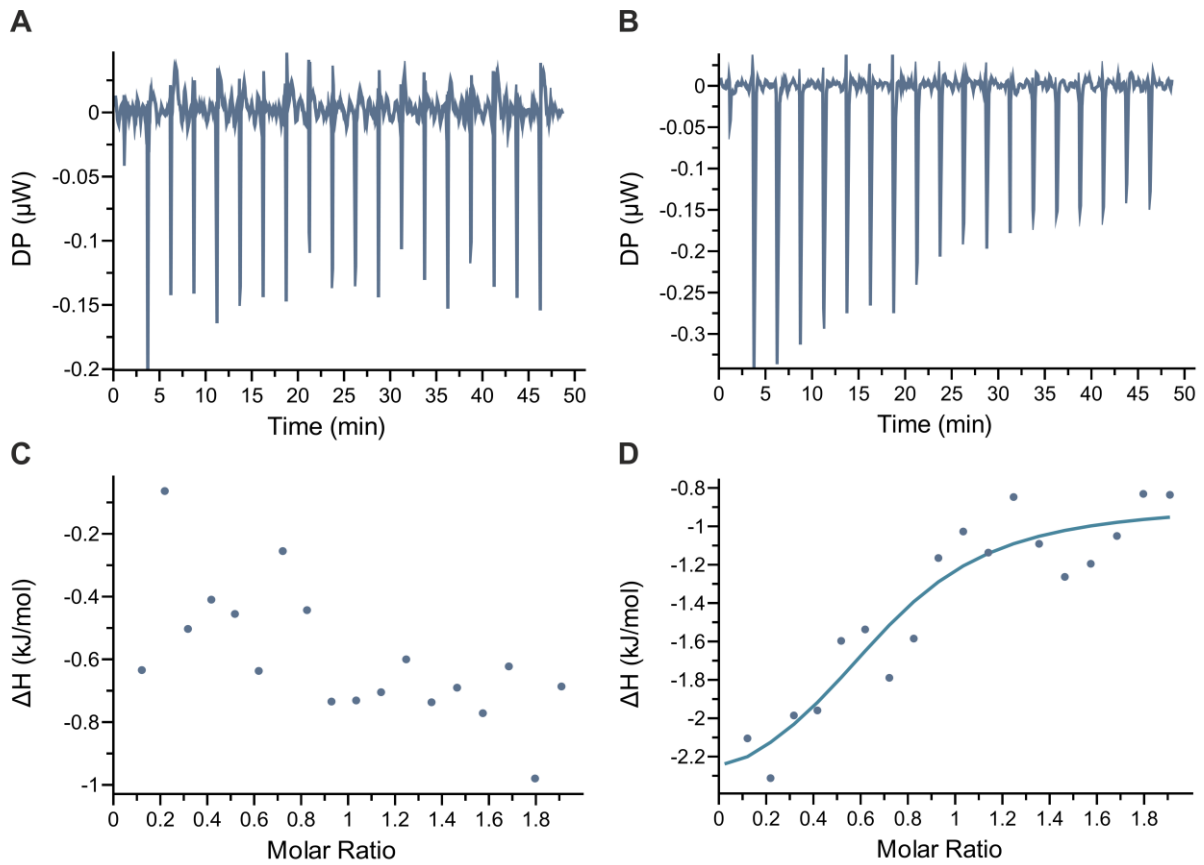
**Figure S2.4.** Analytical SEC with varying concentrations of AncGGGPS2\_N2 and AncGGGPS2\_N3. (A) AncGGGPS2\_N2 and (B) AncGGGPS2\_N3 were applied at four different subunit concentrations (as given in the legend) to a S75 analytical column equilibrated with 50 mM potassium phosphate, pH 7.5, 300 mM KCl. Elution was performed at a flow rate of 0.5 ml/min, followed by measuring the absorbance at 280 nm, which was plotted against the elution volume. mtGGGPS\_wt and mtGGGPS\_W141A (dashed lines) served as references for the hexameric and dimeric oligomerization states. The oligomerization state is indicated by gray symbols.

**AncGGGPS2\_N4**

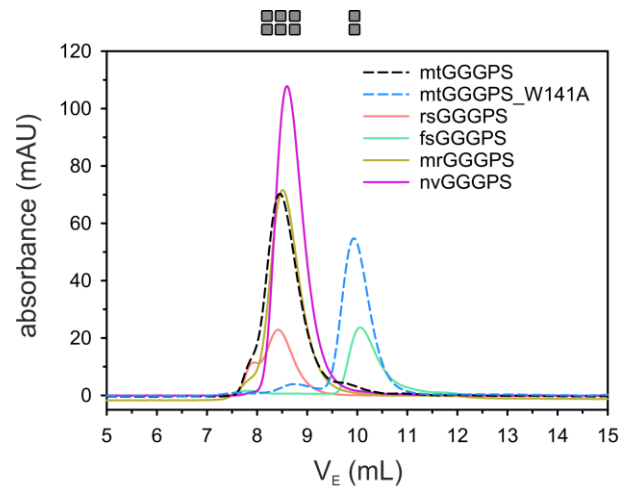
121- E P I P T A Y L I V G D G T T V G **Y** V G N A **R** P I P Y H K P E I A V A Y A -158  
122- E P I P T A Y L I V E P G E T V G **W** V G D A **K** P I P R H K P E I A A A Y A -157

**AncGGGPS2\_N12**

**Figure S2.5.** Sequence alignment of AncGGGPS2\_N4 and AncGGGPS2\_N12. Only the region around the ring interface is shown. The transplanted region is highlighted in blue, the non-swapped rest of the ring interface highlighted in beige. The positions of the “aromatic anchor” (a) and the cation (c) are marked.

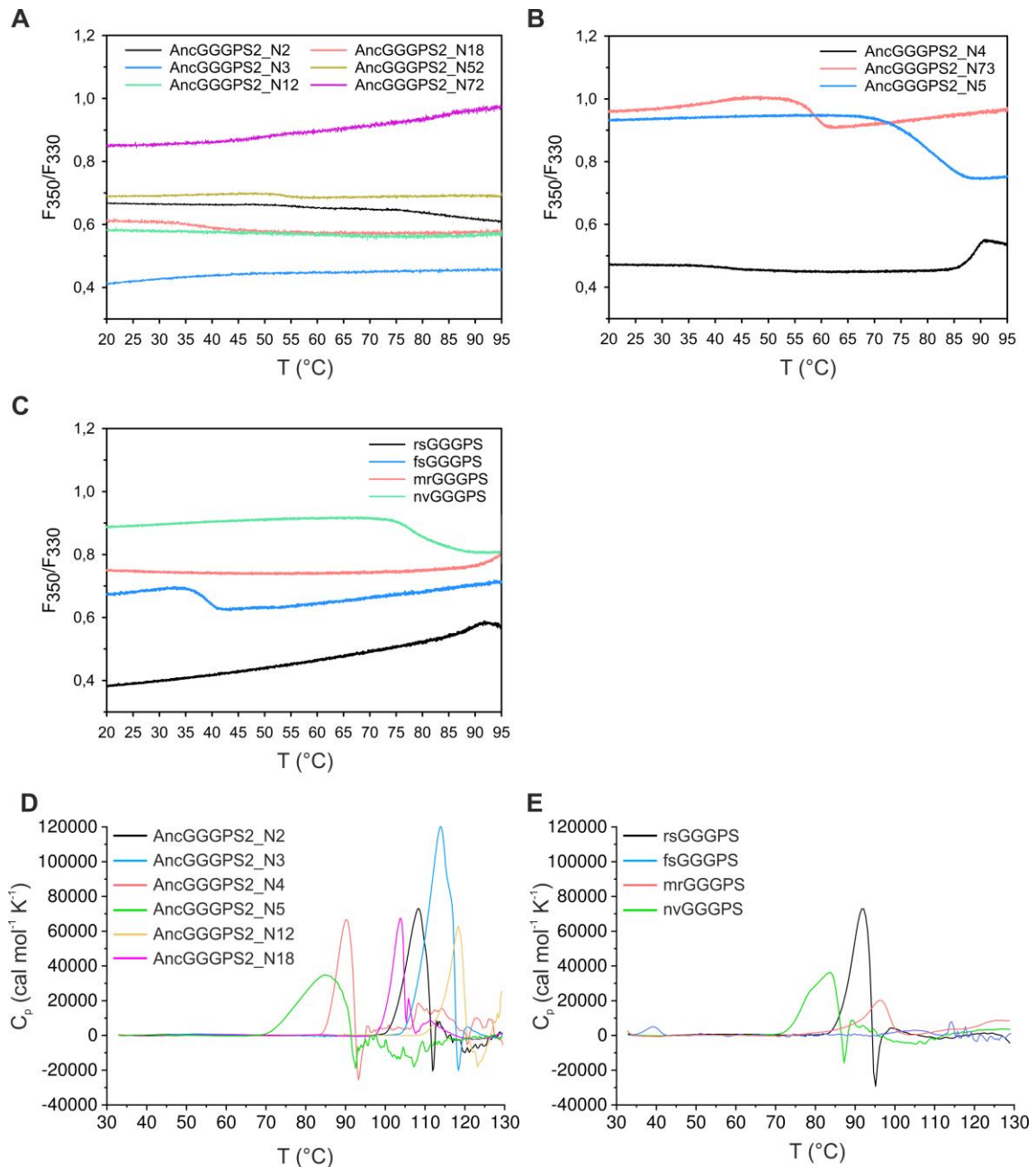


**Figure S2.6.** Analysis of G1P binding to AncGGGPS2\_N4 variants by isothermal titration calorimetry. (A/C) AncGGGPS2\_N4. (B/D) AncGGGPS2\_N4\_IF\_n12. (A) and (B) show the raw data of the titration experiments, in (C) and (D) the derived  $\Delta H$  for each titration step is plotted against the molar ratio G1P:protein. The observed change in  $\Delta H$  for AncGGGPS2\_N4\_IF\_n12 is too weak to reliably deduce binding parameters.

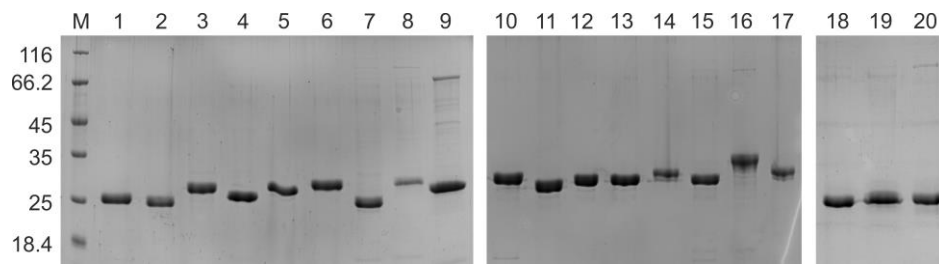


**Figure S2.7.** Analytical SEC of rsGGGPS, fsGGGPS, mrGGGPS and nvGGGPS. The proteins (40  $\mu$ M, subunit concentration) were applied to a S75 analytical column equilibrated with 50 mM potassium phosphate, pH 7.5, 300 mM KCl. Elution was performed at a flow rate of 0.5 mL/min and followed by measuring the absorbance at 280 nm, which was plotted against the elution volume. mtGGGPS\_wt and mtGGGPS\_W141A (dashed lines) served as references for the hexameric and dimeric oligomerization states. The oligomerization state is indicated by gray symbols.





**Figure S2.8.** Thermal stability of GGGPS proteins, followed by nanoDSF and DSC. (A) nanoDSF analysis of ancestral GGGPS proteins with an apparent  $T_2 > 95$  °C. (B) nanoDSF analysis of ancestral proteins with an observable thermal transition. (C) nanoDSF analysis of extant GGGPS proteins. The change in the ratio of the fluorescence emission at 350 and 330 nm of 40  $\mu$ M protein (subunit concentration) in 50 mM potassium phosphate, pH 7.5 (supplemented with 300 mM KCl for AncGGGPS2\_N73 and rsGGGPS) was monitored from 20 to 95 °C at a scan rate of 1 K/min. We have shown previously that a decrease in the  $F_{350}/F_{330}$  ratio points to a partial denaturation only (transition  $T_1$ ), which is followed by a second transition ( $T_2$ ) at higher temperatures, going along with complete denaturation of the protein (Linde et al 2018). Due to the limited temperature range of the instrument,  $T_2$  usually cannot be seen in nanoDSF analysis. (D) DSC analysis of ancestral GGGPS enzymes. (E) DSC analysis of extant GGGPS enzymes. Changes in the heat capacity of 40  $\mu$ M protein (subunit concentration) in 50 mM potassium phosphate, pH 7.5 were monitored from 30 to 130 °C at a scan rate of 1 K/min. The identities of the proteins are given in the legend. Repetitive thermal denaturation experiments revealed that the thermal transitions observed by DSC are mostly irreversible (data not shown).



**Figure S2.9.** Purity of GGGPS variants characterized in this study. SDS-PAGE analysis (12.5% polyacrylamide) of the purified GGGPS variants (2.5  $\mu$ g protein each). Size marker (M), MW in kDa. Ancestral GGGPS variants: (1) AncGGGPS2\_N2, (2) AncGGGPS2\_N3, (3) AncGGGPS2\_N4, (4) AncGGGPS2\_N5, (5) AncGGGPS2\_N12, (6) AncGGGPS2\_N18, (7) AncGGGPS2\_N52, (8) AncGGGPS2\_N72, (9) AncGGGPS2\_N73, (10) AncGGGPS2\_N4\_IF\_n12, (11) AncGGGPS2\_N12\_IF\_n4, (12) AncGGGPS2\_N12\_W138H, (13) AncGGGPS2\_N12\_W138A. Extant GGGPS variants: (14) rsGGGPS, (15) fsGGGPS, (16) mrGGGPS, (17) nvGGGPS, (18) mtGGGPS, (19) mtGGGPS\_W141H, (20) mtGGGPS\_W141A. The samples were run on three separate SDS gels.

# Chapter 3

## 3 Controlling enzymatic activity by modulating the oligomerization state via chemical rescue and optical control

Cosimo Kropp<sup>[1]</sup>, Astrid Bruckmann<sup>[2]</sup> and Patrick Babinger<sup>[1]\*</sup>

<sup>[1]</sup> Institute of Biophysics and Physical Biochemistry, Regensburg Center for Biochemistry, University of Regensburg  
93040 Regensburg, Germany

<sup>[2]</sup> Institute of Biochemistry, Genetics and Microbiology, Regensburg Center for Biochemistry, University of Regensburg  
93040 Regensburg, Germany

\*Corresponding Author: Patrick Babinger

E-mail: [patrick.babinger@ur.de](mailto:patrick.babinger@ur.de)

Tel.: +49 941 943 1634; Fax: +49 941 943 2813

### 3.1 Abstract

Selective switching of enzymatic activity has been a longstanding goal in synthetic biology. Drastic changes in activity upon mutational manipulation of the oligomerization state of enzymes have frequently been reported in literature, but scarcely exploited for switching. Using geranylgeranyl glyceryl phosphate synthase as a model, we demonstrate that catalytic activity can be efficiently controlled by exogenous modulation of the association state. We introduced a lysine-to-cysteine mutation, leading to the breakdown of the active hexamer into dimers with impaired catalytic efficiency. Addition of bromoethylamine (BrEtAm) chemically rescued the enzyme by restoring hexamerization and activity. As an alternative method, we incorporated the photo-sensitive unnatural amino acid (UAA) *o*-nitrobenzyl-*O*-tyrosine (ONBY) into the hexamerization interface. This again led to inactive dimers, but the hexameric state and activity could be recovered by UV-light induced cleavage of ONBY. For both approaches, we obtained switching factors of larger than 350-fold, which compares favorably with previously reported activity changes that were caused by site-directed mutagenesis.

### 3.2 Introduction

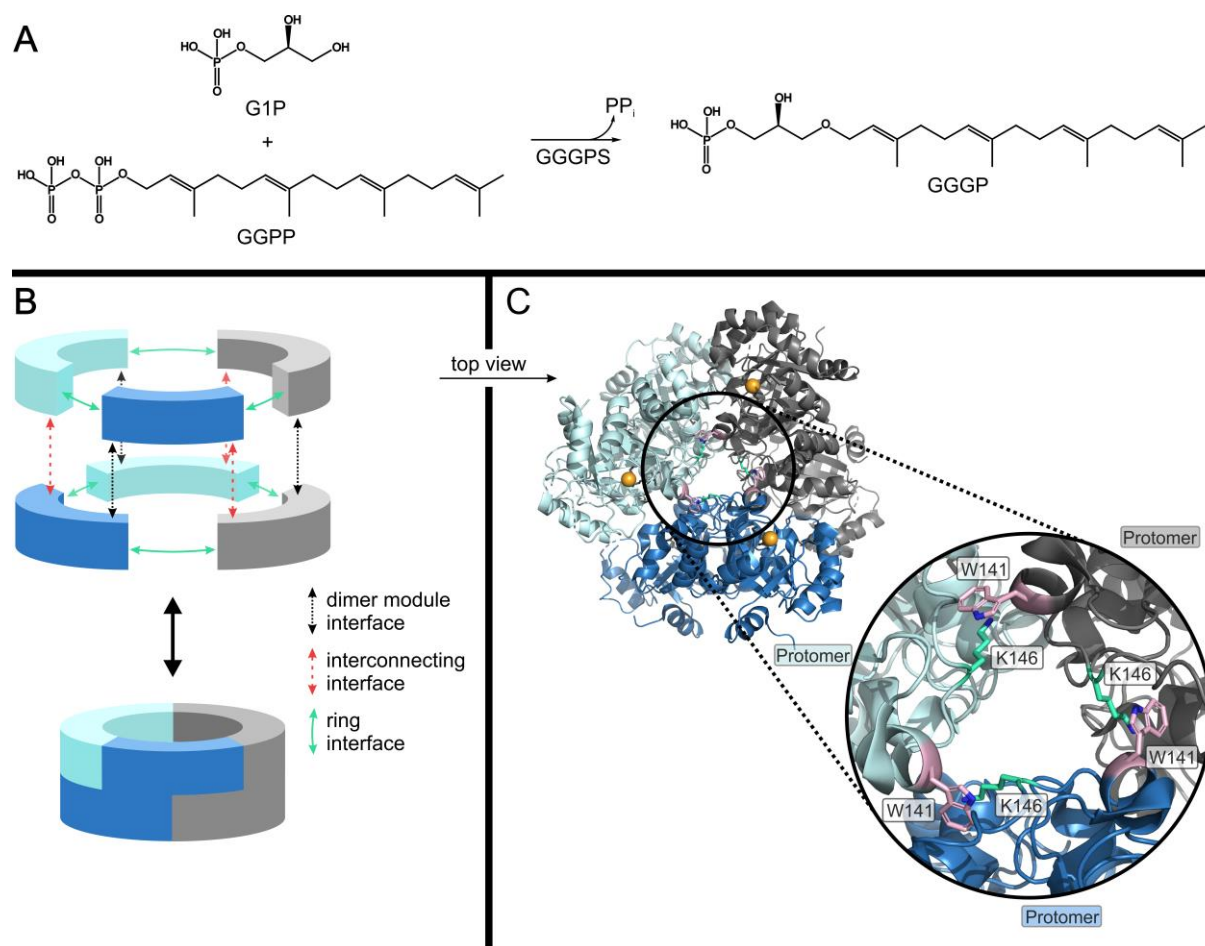
Selective control over enzymatic activity is a longstanding goal in synthetic biology. Especially in the field of biocatalysis, the design of controllable enzymes has been pushed forward in recent years. Engineered enzymatic biocatalysts will replace current production methods of high-value chemicals for an increased economical and resource efficient production and a decreased environmental impact. Selective control of enzymatic activity for instance helps to organize enzymes in cascades (Schmidt-Dannert & Lopez-Gallego 2016, Woodley 2019).

Control over enzymatic activity can be achieved by a wide variety of techniques such as the implantation of molecular switches. There are various methods to generate such switches, among them chemical rescue and optochemical tools. For so-called chemical rescue, an essential amino acid residue is mutated to render the enzyme less active or inactive. Upon addition of small exogenous compounds to the purified protein, activity is restored. The compound thereby mimics the mutated residue – in other words, the enzyme becomes dependent on a cofactor. Since its first introduction (Toney & Kirsch 1989), several *in vitro* studies using chemical rescue have been performed (Peracchi 2008). Alternative and nowadays more widespread strategies use optochemical tools, which can be subdivided into the overlapping fields of optogenetics (Repina et al 2017) and methods that directly control the target protein (Ankenbruck et al 2018). For the latter methods, most commonly an UAA is incorporated into the protein, which is encoded by a reprogrammed stop codon and delivered by a modified tRNA plus aminoacyl-tRNA synthetase. The sterically challenging properties or other features of the UAA render the enzyme less active. Upon light exposure, the photo-sensitive UAA is, for example, degraded to the canonical residue it was incorporated for and the protein becomes functional again.

Both for chemical rescue and optical control of enzymes, the most common approaches to switch activity are either based on direct modification of catalytically relevant residues (Hemphill et al 2013, Olucha et al 2012, Toney & Kirsch 1992) or on indirect blocking of the active site. For example, incorporation of bulky UAAs can impede the access of the substrate to the binding pocket (Wang et al 2019), or introduced and then rescued mutations negatively affect the structural integrity or stability of the protein (Eriksson et al 1992). Such structure-based approaches often exploit allosteric effects (Deckert et al 2012, Kneuttinger et al 2019). Although such approaches seem to be straight forward, it is frequently difficult to achieve activity switching factors  $> 10 - 100$ .

Allostery often includes the interaction of two or more subunits within an enzyme, and many enzymes are at least dimeric proteins. Although the oligomerization state of proteins has already been successfully manipulated by optical control or chemical rescue to control transcription factors (Engelke et al 2014, Xia et al 2013) or to impede virus-host interactions (Erickson et al 2017), this approach has been scarcely used to directly switch the activity of enzymes. However, it has frequently been reported that the mutational disruption of a functional oligomer is accompanied by a severe decrease of activity or catalytic efficiency, with factors of  $10^2$  over  $10^5$  to even unlimited, in case the broken complex is completely inactive. Among them are, for example, peroxiredoxins (Parsonage et al 2005), dihydroorotate dehydrogenase (Ottosen et al 2002), glutathione transferase (Thompson et al 2006), tryptophan synthase (Leopoldseder et al 2006) or glutamine amidotransferases (Sammelmann et al 2019). Especially high switching factors can be achieved in cases where substrate channeling between the subunits is disrupted, or when two subunits form the active site in concert (Yang et al 2019).

As a proof-of-principle, we set out to switch the activity of GGGPS by modulating its oligomerization state using either chemical rescue or optical control. GGGPS catalyzes the condensation of G1P and GGPP (Figure 3.1A), which is a key step in the biosynthesis of the typical archaeal membrane ether lipids (Boucher et al 2004, Peterhoff et al 2014). Hexameric oligomerization is widely spread among this enzyme family (Figure 3.1B, C), but can easily be disturbed by introducing mutations at the subunit interfaces. The hexamer then disassembles into dimers (Kropp et al 2020, Peterhoff et al 2014), which is accompanied by a 380-fold decrease in catalytic efficiency (Linde et al 2018). We made use of this effect and now disturbed the hexamer either by mutating an essential Lys in the hexamerization interface, which can be chemically rescued in the purified protein, or by inserting the UAA ONBY instead of an essential aromate, which can be optically degraded to a Tyr. With both strategies, we obtained similar switching factors like previously observed in mutational studies with GGGPS, but now by manipulating the purified protein by exogenic means.



**Figure 3.1.** Properties of the GGGPS hexamer. (A) Catalyzed reaction of GGGPS. (B) Schematic visualization of the GGGPS structure in exploded and assembled view. Each building block represents a GGGPS monomer. A dimer module is indicated by matching colors. The figure is adapted from preceding work (Linde et al 2018). (C) Crystal structure of mtGGGPS (PDB ID 4mm1) with co-crystallized G1P substrate. The ring interface is marked by the black circle. The hot spot residues in the ring interface, W141 (pink) and K146 (green), that form the cation- $\pi$  bond between adjacent protomers are shown as sticks. The phosphate of the G1P in the upper three protomers is shown as an orange sphere.

### 3.3 Results & Discussion

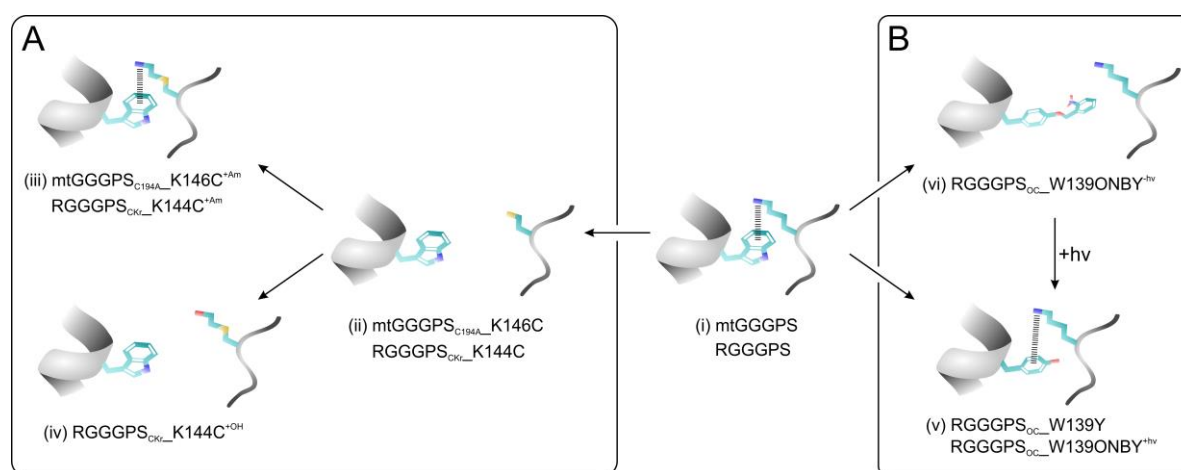
We selected the mtGGGPS as model enzyme for our proof-of-concept study, which allowed us to base on previous results (Kropp et al 2020, Linde et al 2018, Peterhoff et al 2014, Straub et al 2019). In the hexameric mtGGGPS, complex formation is mediated by three distinct interfaces: the dimer module interface, the interconnecting interface and the ring interface (Figure 3.1B). The dimer module interface connects two monomers to a dimer. The other two interfaces assemble the hexamer as a trimer of dimers. A crucial interaction in the ring interface between adjacent dimer modules is a cation- $\pi$  bond, which is an interaction of an aromatic residue in the one protomer (Trp in mtGGGPS) and a cationic residue in the other (Lys in mtGGGPS; Figure 3.1C). Previous studies have confirmed the importance of the cation- $\pi$  bond by mutation of either the Trp or the Lys residue. The mutated GGGPS variants disassembled into dimers and showed a massively decreased affinity for the G1P substrate ( $\sim 75x$ ) as well as a moderate decrease of the turnover number ( $\sim 5x$ ), resulting in a 380x decrease in catalytic efficiency (Linde et al 2018).

In a subsequent study, we resurrected evolutionary predecessors of mtGGGPS by ancestral sequence reconstruction (Kropp et al 2020). We identified two sequential GGGPS ancestors that were located on the same evolutionary path, called AncGGGPS2\_N4 (N4) and AncGGGPS2\_N12 (N12). N4 was dimeric and non-functional, in contrast to hexameric and active N12 (Figure S3.1, Figure S3.2, Figure S3.3, Table S3.1). We managed to hexamerize N4 by transplanting the ring interface from N12 into N4. The resulting variant AncGGGPS2\_N4\_IF\_n12 (N4\_IF\_n12), which differs only in five contact interface residues from N4 (Kropp et al 2020), showed comparable activity as N12 (Figure S3.7E). Due to the inactivity of N4, N4\_IF\_n12 promised the possibility to establish an infinite switching factor by modulating its oligomerization state using a single mutation. Hence, in addition to mtGGGPS, we selected N4\_IF\_n12 as a second study object. For simplicity, we call this variant “RGGGPS” (rescued GGGPS) from now on.

#### 3.3.1 Cation- $\pi$ interaction as molecular switch

Cation- $\pi$  interactions are widespread among oligomers. They occur in the subunit contact interfaces from about 50% of all protein complexes and significantly contribute to their stabilization (Crowley & Golovin 2005). Cation- $\pi$  interactions are accessible to manipulation by both chemical rescue or optical control. The aromatic residue can be rescued by non-covalent binding of indole after its mutation to a small residue (Deckert et al 2012). The cationic counterpart, if it is a Lys, it can be rescued by BrEtAm after its mutation to Cys, forming a covalently linked Lys-mimetic residue (Olucha et al 2012). Alternatively, both interacting residues can be replaced by sterically demanding photocaged UAAs: the aromate by ONBY, which can be photo-decaged to Tyr, the cationic counterpart by photocaged Lys (Courtney & Deiters 2018).

We followed both approaches, which are depicted in Figure 3.2A (chemical rescue) and Figure 3.2B (photo-control), respectively. In mtGGGPS, the cation- $\pi$  interaction is formed by the residues W141 and K146, in RGGGPS by W139 and K144 (Figure 3.2(i)). To begin with, we pursued chemical rescue for mtGGGPS, which first required to remove a native Cys to avoid inadvertent effects (introduced mutation C194A). The cation- $\pi$  interaction was then disrupted by mutating the Lys to a Cys (K146C). Similarly, we proceeded with RGGGPS and introduced the analogous mutation K144C (Figure 3.2(ii)). After incubation with BrEtAm (Figure 3.2(iii)), the cation- $\pi$  bond was intended to be re-established. As control, RGGGPS<sub>CKr\_K144C</sub> was supplemented with bromoethanol (BrEtOH). BrEtOH generates an isosteric alcohol instead of an amine after Cys modification and should therefore not be able to rescue the cation- $\pi$  interaction (Figure 3.2(iv)). Optical control was only implemented for RGGGPS by ONBY incorporation instead of the aromate (W139ONBY; Figure 3.2(v)). This inactive variant is intended to be reactivated by light (+hv; Figure 3.2(vi)), which degrades ONBY to Y. As control, the Trp was directly mutated to Tyr (W139Y; Figure 3.2(vi)).



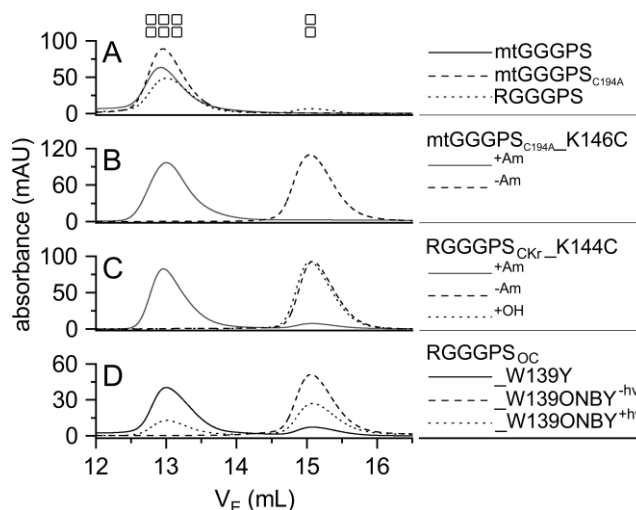
**Figure 3.2.** Switch conceptualization. Starting from mtGGGPS or RGGGPS with a Trp-Lys cation- $\pi$  interaction, two strategies were implemented. (A) chemical rescue, (B) optical control. For detailed description, see the text. An intact cation- $\pi$  interaction is symbolized by black bars between the aromatic and the cationic moiety. The indices mean: Am, bromoethylamine; OH, bromoethanol; CKr, Chemical lysine (K) rescue; OC, optical control; C194A, auxiliary mutation. Detailed chemical structures of the modified amino acids are shown in Figure S3.4.

### 3.3.2 Activation of catalytic activity by switching the oligomerization state

We expressed the variants depicted in Figure 3.2 in *E. coli* and purified the proteins by immobilized metal chelate affinity chromatography (IMAC). A preparative SEC purification step was added to remove residual contaminating proteins of higher and lower MW, which facilitated subsequent analysis of switched oligomerization states. Before and after treating the proteins for chemical rescue or with light, we subjected them to analytical SEC in combination with static



light scattering (SLS) to determine the oligomerization states of the complexes and to obtain exact molecular masses. As expected, wild type mtGGGPS and mtGGGPS<sub>C194A</sub> as well as RGGGPS eluted as hexamers (Figure 3.3A). For chemical rescue, mtGGGPS<sub>C194A</sub>\_K146C and RGGGPS<sub>CKr</sub>\_K144C were incubated over night with 20 mM BrEtAm (<sup>+Am</sup>) or without (<sup>-Am</sup>). The Lys-to-Cys mutation led to exclusively dimeric proteins in both mtGGGPS<sub>C194A</sub>\_K146C<sup>+Am</sup> and RGGGPS<sub>CKr</sub>\_K144C<sup>+Am</sup>. For both proteins, BrEtAm rescued the hexameric oligomerization state to almost 100% (Figure 3.3B, C). The presence of the Lys-mimetic residue (Lys- $\gamma$ -S) in RGGGPS<sub>CKr</sub>\_K144C<sup>+Am</sup> was verified by tryptic digest coupled with mass spectrometry (MS) (Figure S3.5A). The control RGGGPS<sub>CKr</sub>\_K144C<sup>+OH</sup>, which was incubated with BrEtOH instead of BrEtAm, remained dimeric (Figure 3.3C). All oligomerization states and the molecular masses as derived by SLS are summarized in Table S3.1.



**Figure 3.3.** Oligomerization states of GGGPS variants. The denoted proteins (40  $\mu$ M subunit concentration) were applied to a S200 10/300 GL analytical column, which was equilibrated with 50 mM potassium phosphate, pH 7.5, 300 mM KCl. Elution was performed at a flow rate of 0.4 ml min<sup>-1</sup>, followed by measuring the absorbance at 280 nm, and plotted against the elution volume. The derived oligomerization states are indicated by rectangles (2 = dimer, 6 = hexamer) and listed in Table S3.1.

Next, we analyzed the GGGPS variants that were designed for optical control. Because it is a common issue that protein preparations with an integrated UAA contain wild-type protein contaminations, the preparative SEC purification step can additionally support the homogenous preparation of correctly produced protein, as the incorporation of the UAA changes its oligomerization state, according to our conceptual design. But for RGGGPS<sub>OC</sub>\_W139ONBY<sup>hv</sup>, incorporation worked very well, an elution peak corresponding to an hexameric fraction was not detectable in preparative SEC (data not shown). Consequently, the protein eluted as a pure dimer in subsequent analytical SEC, indicating a complete disruption of the hexamer due to the incorporated ONBY (Figure 3.3D). Incorporation of ONBY instead of W139 was verified by

tryptic digest coupled with MS (Figure S3.5B). After the protein was decaged by irradiation at 365 nm for 3 min (RGGGPS<sub>OC</sub>\_W139ONBY<sup>hv</sup>), approximately 30% of it (as estimated from the peak integrals) eluted as a hexamer in SEC, which indicated that the hexamer could be reconstituted as projected. Because ONBY is decaged by light to Tyr, we created the variant RGGGPS<sub>OC</sub>\_W139Y as control, which elutes almost completely as hexamer under the given experimental conditions (Figure 3.3D).

To test whether the switch in oligomerization state affected enzymatic activity, we tested the variants with a photometric assay that detects the pyrophosphate which is liberated in the GGGPS reaction (Figure 3.1A). As expected from *a priori* knowledge from N4 (cf. first paragraph of results; Figure S3.3), RGGGPS<sub>CKr</sub>\_K144C<sup>+Am</sup> showed no activity at all when we incubated the enzyme with saturating concentrations of G1P for 24 h (Figure S3.6, white triangles). In contrast, after addition of BrEtAm, RGGGPS<sub>CKr</sub>\_K144C<sup>+Am</sup> showed rising activity with time. Full restoration of activity was reached after ~30 min and stayed unchanged even after 24 h of incubation with BrEtAm (Figure S3.6, black circles).

To determine activity switching factors of the enzymes upon chemical rescue and optical control, we determined the steady-state kinetic parameters of the variants (Figure S3.7, Table 3.1). The variant mtGGGPS<sub>C194A</sub>, which was created as “wild-type reference” to avoid inadvertent effects in the chemical rescue experiment due to the native Cys, showed similar catalytic parameters like mtGGGPS. In comparison, the variant mtGGGPS<sub>C194A</sub>\_K146C<sup>+Am</sup> with disrupted cation- $\pi$  interaction and dimeric oligomerization state showed an 8x decrease in  $k_{\text{cat}}$ , a 19x increase in  $K_M$  and consequently a 160x decrease in  $k_{\text{cat}}/K_M$ . Within the range of experimental fluctuation, this is consistent with the previously published kinetic data for hexameric mtGGGPS and the dimeric mutant mtGGGPS\_W141A (5x decrease in  $k_{\text{cat}}$ , 75x increase in  $K_M$ , 380x decrease in  $k_{\text{cat}}/K_M$ ) (Linde et al 2018). After chemical rescue with BrEtAm (mtGGGPS<sub>C194A</sub>\_K146C<sup>+Am</sup>), activity was fully restored. The rescued variant even showed a slightly elevated activity, corresponding to an activation factor of 350x in terms of  $k_{\text{cat}}/K_M$ . Because we strived to achieve high switching factors, we analyzed the variant RGGGPS, which is about 25x less active than the wild-type mtGGGPS, in terms of  $k_{\text{cat}}/K_M$ . As discussed above, RGGGPS<sub>CKr</sub>\_K144C<sup>+Am</sup> showed no detectable activity anymore. After chemical rescue with BrEtAm (RGGGPS<sub>CKr</sub>\_K144C<sup>+Am</sup>), activity was fully restored again, which is tantamount to an infinite switching factor.

Analogously, we investigated reactivation upon ONBY decaging to Tyr in the photo-sensitive variant RGGGPS<sub>OC</sub>\_W139ONBY. Again, RGGGPS<sub>OC</sub>\_W139ONBY<sup>hv</sup> was non-functional (Figure S3.3). Upon decaging, 27% of catalytic efficiency were restored in comparison to RGGGPS. This is congruent with the portion of about 30% hexameric protein after light treatment (Figure 3.3D). The reason for this incomplete restoration of the active hexamer most likely is that frequently, a large portion of ONBY is irreversibly reduced in cells and thus becomes uncleavable (Kneuttinger et al 2019). We could detect the presence of the reduced form by tryptic

digest coupled with MS also in our preparation (Figure S3.5C). The control RGGGPS<sub>OC\_W139Y</sub> showed almost identical catalytic parameters as RGGGPS.

**Table 3.1.** Catalytic parameters of GGGPS variants in inactive and active states.<sup>a</sup>

protein	$k_{\text{cat}}$ ( $\text{s}^{-1}$ )	$K_{\text{M}}$ ( $\mu\text{M}$ )	$k_{\text{cat}}/K_{\text{M}}$ ( $\text{s}^{-1} \text{M}^{-1}$ )
mtGGGPS	$2.3 (\pm 0.05) \times 10^{-1}$	$2.0 \pm 0.3$	$11.5 (\pm 1.7) \times 10^4$
mtGGGPS <sub>C194A</sub>	$2.9 (\pm 0.1) \times 10^{-1}$	$3.3 \pm 0.6$	$8.8 (\pm 1.8) \times 10^4$
mtGGGPS <sub>C194A_K146C<sup>+Am</sup></sub>	$2.3 (\pm 0.04) \times 10^{-1}$	$1.2 \pm 0.1$	$19.2 (\pm 2.4) \times 10^4$
mtGGGPS <sub>C194A_K146C<sup>-Am</sup></sub>	$3.5 (\pm 0.07) \times 10^{-2}$	$63.4 \pm 5.8$	$5.5 (\pm 0.6) \times 10^2$
RGGGPS	$4.4 (\pm 0.2) \times 10^{-2}$	$14.2 \pm 2.2$	$3.6 (\pm 0.7) \times 10^3$
RGGGPS <sub>CKr_K144C<sup>+Am</sup></sub>	$3.7 (\pm 0.05) \times 10^{-2}$	$6.7 \pm 0.5$	$5.8 (\pm 0.5) \times 10^3$
RGGGPS <sub>CKr_K144C<sup>-Am</sup></sub>	-	-	-
RGGGPS <sub>OC_W139Y</sub>	$3.8 (\pm 0.09) \times 10^{-2}$	$10.2 \pm 1.2$	$3.7 (\pm 0.5) \times 10^3$
RGGGPS <sub>OC_W139ONBY<sup>+hv</sup></sub>	$2.2 (\pm 0.07) \times 10^{-2}$	$22.5 \pm 3.0$	$0.99 (\pm 0.2) \times 10^3$
RGGGPS <sub>OC_W139ONBY<sup>-hv</sup></sub>	-	-	-

<sup>a</sup> kinetic parameters were determined at 40 °C with a photometric assay for phosphate detection in duplicates and fitting the Michaelis–Menten equation to the data (Figure S3.7). Standard deviations are given. -, no analyzable activity in the presence of 500 nM protein, 250  $\mu\text{M}$  G1P and 11  $\mu\text{M}$  GGPP.

### 3.3.3 Physico-biochemical analysis of the inactive and reactivated variants

The structural integrity of all variants used in the study was monitored by CD spectroscopy. All spectra indicated a well-defined secondary structure with no significant differences between active, inactive and reactivated state (Figure S3.8). This supports that the disruption of the hexamer is not associated with larger secondary structure rearrangements, but mainly results from the destruction of the cation- $\pi$  interaction. Nevertheless, we assume that the significant loss of catalytic efficiency upon hexamer disruption is caused by small structural rearrangements that affect substrate binding. The oligomerization interface with the disturbed cation- $\pi$  interaction is not immediately neighboring the active site (Figure 3.1C; Linde et al 2018, Peterhoff et al 2014), but especially  $K_{\text{M}}$  is drastically impaired in the dimeric variants (Table 3.1). To shed light on this, we analyzed binding of the substrate G1P to RGGGPS<sub>CKr\_K144C<sup>-Am</sup></sub> and RGGGPS<sub>CKr\_K144C<sup>+Am</sup></sub> by ITC (Figure S3.9). Although a dissociation constant cannot be derived from the raw data due to the high protein concentrations that were necessary to obtain good signals, the data clearly indicate a severely impaired binding of G1P to RGGGPS<sub>CKr\_K144C<sup>-Am</sup></sub> which can be rescued by addition of BrEtAm.

We have shown previously that modulation of the oligomerization state of mtGGGPS influences the thermal stability of the protein. While the overall fold of mtGGGPS is extremely thermostable, activity is lost in the dimeric variants at much lower temperature than in the hexamer, and we assume this is due to an extra stabilization of the active site in the hexamer (Linde et al 2018). We tested by nanoDSF whether this is also the case for the inactive and rescued mtGGGPS variants. mtGGGPS<sub>C194A</sub> shows a very high thermal stability with a melting temperature >95 °C like mtGGGPS (Figure S3.10A; no visible peak in the first derivative plot of nanoDSF). Like previously reported for dimeric mtGGGPS\_W141A (Linde et al 2018), mtGGGPS<sub>C194A</sub>\_K146C<sup>-Am</sup> shows a reduced thermal stability with a transition at approx. 62 °C, which can be completely restored to wild-type level in mtGGGPS<sub>C194A</sub>\_K146C<sup>+Am</sup> (Figure S3.10B). The tested RGGGPS variants all show very similar thermal denaturation in nanoDSF with a transition at 88-92 °C (Figure S3.10C, D). Due to the strong fluorescence of ONBY, RGGGPC<sub>OC</sub>\_W139ONBY could not be analyzed by nanoDSF.

In summary, these results support that activity switching of GGGPS by modulating its oligomerization state is accompanied by changes in the physico-biochemical properties of the enzyme that can be robustly monitored and controlled. As expected, some structural properties like thermal stability are impaired in some inactivated variants, but can be restored to wild-type level upon exogenous reactivation.

### 3.4 Conclusion

Our proof-of-concept study demonstrates that catalytic activity of our model enzyme, GGGPS, can be readily and efficiently regulated *in vitro* by changing the association state via chemical rescue or photo-switching. It is obvious that chemical rescue with BrEtAm is not a suitable approach for *in vivo* applications due to its toxicity, and also incorporation of UAAs remains challenging for use in living cell systems. But both methods have been well established over the years and might work well for production of switchable enzymes for the *in vitro* synthesis of chemicals, such as in one-pot reactions. Our central aim, however, was to demonstrate that the control of the oligomerization state by introducing molecular switches might be a widely applicable method to control enzymatic activity of homo- or heteromeric enzyme complexes, since drastic changes of activity upon oligomer disruption are common and frequently reported in the literature. Importantly, switching activity by reversibly dissociating an oligomer using externally added chemical compounds or UAAs has several advantages compared to the direct modification of the active site. For example, it is often straightforward to dissociate oligomers by single mutations, e.g. by incorporation of bulky photocaged amino acids or by disrupting cation- $\pi$  interactions. The latter are common in contact interfaces and represent ideal targets to establish a switch, because both interaction partners are accessible to chemical rescue and optical control. Furthermore, the change in oligomerization state is predictable, rather easy to detect and to be analyzed by SEC, and it allows high-quality purification of either switched state for *in vitro* use of enzymes.

We additionally used the benefit of ASR in our study, which usually creates a bundle of homologous proteins with diverse features concerning activity. A main advantage of ASR-generated protein variants is that they are frequently structurally very stable, not rarely even more stable than the extant representatives (Wheeler et al 2016). This allowed us either to select highly active variants in the on state (the extant wild-type variants) with a good activity switching factor (>100x) to the off state, or variants that are completely inactive in the off state albeit with less activity in the on state, just as the application conditions would require.

## 3.5 Materials & Methods

### 3.5.1 Cloning and site-directed mutagenesis

All mutated variants as described in Table S3.1 were generated by QuickChange mutagenesis (Zheng et al 2004) with oligonucleotides listed in Table S3.2. For ONBY incorporation, a stop codon point mutation (TAG = amber) was introduced into RGGGPS (previously called AncGGGPS2\_N4\_IF\_n12 (Kropp et al 2020)). To confirm the successful mutation, all genes were sequenced entirely. RGGGPS has been cloned previously (Kropp et al 2020), as well as mtGGGPS (Peterhoff et al 2014). The sequence numbering for mtGGGPS used in this study refers to EMBL ENA entry AAB85058.

### 3.5.2 Production and purification of proteins

Proteins were produced by heterologous gene expression in BL21-Gold(DE3) *E. coli* cells (Agilent Technologies). Transformed cells were grown at 37 °C in LB medium containing ampicillin (150 µg ml<sup>-1</sup>) until an OD<sub>600</sub> of 0.6. Expression was induced by adding 1 mM IPTG and growth was continued overnight at 20 °C. Cells were harvested by centrifugation and suspended in 50 mM Tris, pH 8.0, 300 mM NaCl, 10 mM imidazole. For the chemical rescue variants (RGGGPS<sub>CKr</sub>\_K144C, mtGGGPS\_K146C), 10 mM 2-mercaptoethanol was added to the lysis buffer. Cells were disrupted by sonication and the His-tagged proteins were purified from the clarified cell extract by IMAC using an ÄKTApurifier system with a HisTrap FF crude column (5 mL, Cytiva). Proteins were eluted by a linear gradient of imidazole (10–500 mM) in 50 mM Tris, pH 8.0, 300 mM NaCl. Contaminating proteins of higher and lower MW, imidazole and salt were removed by subsequent preparative SEC on a Highload<sup>TM</sup> 26/600 Superdex<sup>TM</sup> S200 pg column (Cytiva) at a flow rate of 1.5 ml min<sup>-1</sup>. The column was equilibrated with 50 mM Tris, pH 8.0. 1 mM Tris(2-carboxyethyl)phosphine was added for the chemical rescue variants.

The photo-sensitive variant, RGGGPS<sub>OC</sub>\_W139ONBY, was also produced by heterologous gene expression in BL21-Gold(DE3) *E. coli* cells (Agilent Technologies). Chemically competent cells were co-transformed with the expression vector, carrying the desired TAG codon for ONBY incorporation at position W139, and pEVOL\_ONBY. Transformed cells were grown at 37 °C in 6 L LB medium containing ampicillin (150 µg ml<sup>-1</sup>) and chloramphenicol (30 µg ml<sup>-1</sup>) until an OD<sub>600</sub> of 0.6. Cells were harvested by centrifugation at room temperature and suspended in 600 mL terrific broth medium. Cell growth was continued at 37 °C to an OD<sub>600</sub> of 10. Protein production and incorporation of ONBY was induced by addition of 1 mM ONBY, 0.02% L-arabinose and 0.5 mM IPTG. Protein production was performed overnight at 20 °C. Cells were harvested by centrifugation, suspended in 50 mM Tris, pH 8.0, 300 mM NaCl, 10 mM imidazole and disrupted by sonication. The His-tagged protein was purified from cell extract by IMAC in 50 mM Tris, pH 8.0, 300 mM NaCl, applying a linear gradient of imidazole to elute the protein,

as described above. Higher oligomers, imidazole and salt were removed by preparative SEC. The column was equilibrated with 50 mM Tris, pH 8.0 and 300 mM NaCl and the system was operated at a flow rate of 1.5 ml min<sup>-1</sup>. All purification steps and all systems were operated in the dark.

Protein concentrations were determined by absorbance spectroscopy or with Bradford assay in case of the photo-sensitive variant. The molar extinction coefficients  $\epsilon_{280}$  and the MW were calculated from the amino acid sequence using ProtParam (Gasteiger et al 2005). Proteins were dropped into liquid nitrogen and stored at -80 °C. High concentrations were achieved by concentrating the sample in ultrafiltration units (Amicon Ultra-15, 10 kDa WMCO, Merck KGaA).

### 3.5.3 Selective activation of enzymatic activity

Chemical rescue of RGGGPS<sub>CKr\_K144C</sub> and mtGGGPS<sub>C194A\_K146C</sub> was conducted by supplementing 40  $\mu$ M protein with 20 mM BrEtAm or BrEtOH (final concentrations) in 50 mM Tris, pH 8.0. Both chemicals were purchased from Merck KGaA. Samples were incubated, when not stated otherwise, for 24 h at 40 °C and shaking (300 rpm). Prior to ITC and nanoDSF experiments, buffer solution was exchanged to remove excess BrEtAm using ultrafiltration units (Amicon Ultra-15, 30 kDa WMCO, Merck KGaA).

Decaging of ONBY was achieved by irradiating the protein mixture in a 1.5 mL reaction tube for 3 min at 365 nm using a high-power LED (LED Engin, Osram; settings: 700 mA and 16 V). This was done directly before subjecting the protein to the analytical methods in the respective buffer solutions and at the protein concentrations needed for analysis.

### 3.5.4 Characterization of the oligomerization state of proteins

Oligomerization states were determined by SEC experiments, using a calibrated Superdex 200 Increase 10/300 GL column (Cytiva), which was operated in 50 mM potassium phosphate, pH 7.5, 300 mM potassium chloride at a flow rate of 0.4 ml min<sup>-1</sup> at room temperature (approx. 23 °C). 100  $\mu$ l of protein with a subunit concentration of 40  $\mu$ M was applied.

### 3.5.5 MW determination by SLS

SLS was performed for MW calculation. 40  $\mu$ M (subunit concentration) protein was applied in a volume of 50  $\mu$ L. A Superdex 200 Increase 10/300 GL column (Cytiva) was operated on an ÄKTAmicro system (Cytiva) in combination with a Viscotek TDA 305 triple detector array (Malvern) including right-angle light scattering and refractive index detectors. The system was operated in degassed buffer (50 mM potassium phosphate, pH 7.5, 300 mM KCl) at a flow rate

of 0.3 ml min<sup>-1</sup> at room temperature (approx. 23 °C). Data was analyzed using the OmniSec software (Viscotek, version 4.7.0; Malvern).

### 3.5.6 Steady-State Enzyme kinetics

The kinetic parameters of the proteins were determined in a G1P dependent continuous enzyme-coupled assay for phosphate detection (Suarez et al 2012). The assay mixture was composed of 50 mM Tris, pH 8.0, 10 mM MgCl<sub>2</sub>, 0.2% Tween80, 11 μM GGPP, 1.25 mM inosine, 0.027 U mL<sup>-1</sup> *E. coli* pyrophosphatase, 0.25 U mL<sup>-1</sup> bacterial purine nucleoside phosphorylase, and 2.5 U mL<sup>-1</sup> microbial xanthine oxidase (all enzymes were obtained from Sigma-Aldrich). G1P (0.75-1200 μM) was mixed with the assay mixture in a total volume of 200 μL and incubated at 40 °C. The reaction was started by addition of the enzyme (100 nM for mtGGGPS variants and N12, 500 nM for N4 and RGGGPS variants) and followed at 293 nm using a Jasco V650 spectrophotometer using a 1 cm cuvette. The ε of uric acid was considered equal to 12.6 × 10<sup>3</sup> M<sup>-1</sup> cm<sup>-1</sup> at 293 nm. The reaction velocities were calculated from the initial slopes and the protein concentration. Kinetic constants were deduced by fitting the Michaelis-Menten equation to the data from duplicate measurements using SigmaPlot 13.0. Kinetic and statistic parameters were calculated using the “XY replicate” feature of SigmaPlot.

### 3.5.7 CD spectroscopy

Proteins were diluted to 6 μM in 50 mM potassium phosphate, pH 7.5. CD spectra were recorded from 180-260 nm with a response time of 0.5 sec at a scan rate of 50 nm min<sup>-1</sup> and 25 °C in a JASCO J-815 spectrophotometer using a 0.1 cm cuvette. Data was normalized to obtain the mean residue ellipticity, as described by literature (Kelly et al 2005).

### 3.5.8 nanoDSF

nanoDSF was performed using an excitation power at 280 nm of 20% for mtGGGPS variants and 60% for RGGGPS variants. 20 μM (mtGGGPS variants) or 40 μM (RGGGPS variants) protein (subunit concentration) was heated in 50 mM potassium phosphate, pH 7.5 from 20 °C to 95 °C at a ramp rate of 1 K min<sup>-1</sup> in a Prometheus NT.48 instrument (NanoTemper Technologies GmbH; access provided by 2bind GmbH). Emission was measured at 330 and 350 nm. The change in the ratio of the fluorescence signal at 350 nm to 330 nm with raising temperature was followed. Fluorescence data were fitted by the program supplied by the manufacturer and the apparent midpoint temperature ( $T_{Mapp}$ ) was determined as an operational measure of protein stability. Results are shown as the first derivative of the fluorescence ratio with respect to temperature. Measurements were done in triplicates, which overlapped perfectly.



### 3.5.9 ITC

To investigate G1P binding to the chemically rescued RGGGPS<sub>CKr\_K144C</sub> variant, a MicroCal PEAQ-ITC microcalorimeter (Malvern Instruments) was used. Degassed protein (150  $\mu$ M subunit concentration) was added to the analyte cell (280  $\mu$ L) and a G1P (1 mM) solution was prepared from the identical buffer batch the protein was prepared in (50 mM Tris, pH 8.0, 10 mM MgCl<sub>2</sub>). The G1P solution was titrated to the protein in 2  $\mu$ L aliquots for a total of 18 injections at 2.5 min intervals at 25 °C during continuous stirring. Titrations of buffer with buffer, protein solution with buffer and buffer with ligand solution were performed as controls. Each titration was baseline corrected. The experimentally observed signals for ligand binding experiment were corrected for the signals of the control experiments. The differential power (DP) between reference and sample cell, necessary to maintain the temperature difference at zero, was plotted against time.

### 3.5.10 Tryptic digest and MS

RGGGPS<sub>CKr\_K144C<sup>+Am</sup></sub> and RGGGPS<sub>OC\_W139ONBY<sup>hv</sup></sub> were run on a 12.5% SDS-PAGE gel and were stained with Coomassie G250 (SimplyBlue SafeStain, Lifetech). Protein bands were cut out and subsequently washed with 50 mM NH<sub>4</sub>HCO<sub>3</sub>, a 50 mM NH<sub>4</sub>HCO<sub>3</sub>/acetonitrile mixture (3:1), and a 50 mM NH<sub>4</sub>HCO<sub>3</sub>/acetonitrile mixture (1:1) and eventually lyophilized. After a reduction/alkylation treatment of cysteines and additional washing steps, proteins were digested in-gel by trypsin (Trypsin Gold, MS grade, Promega) overnight at 37 °C. The resulting peptides were extracted with 50 mM NH<sub>4</sub>HCO<sub>3</sub> and 50 mM NH<sub>4</sub>HCO<sub>3</sub> in 50% acetonitrile. After lyophilization, peptides were reconstituted in 20 mL of 1% trifluoroacetic acid and separated by reversed-phase (RP) chromatography. An UltiMate 3000 RSLCnano System (Thermo Fisher Scientific, Dreieich, Germany) equipped with a C18 Acclaim Pepmap100 preconcentration column [100  $\mu$ m (inside diameter)  $\times$  20 mm, Thermo Fisher Scientific] and an Acclaim Pepmap100 C18 nano column [75  $\mu$ m (inside diameter)  $\times$  250 mm, Thermo Fisher Scientific] was operated at a flow rate of 300 nL min<sup>-1</sup> and a 60 min linear gradient of 4% to 40% acetonitrile in 0.1% formic acid. The liquid chromatograph was online-coupled to a maXis plus UHR-QTOF System (Bruker Daltonics) via a CaptiveSpray nanoflow electrospray source. Acquisition of MS/MS spectra after CID fragmentation was performed in data-dependent mode at a resolution of 60000. The precursor scan rate was 2 Hz processing a mass range between m/z 175 and 2000. A dynamic method with a fixed cycle time of 3 s was applied via the Compass 1.7 acquisition and processing software (Bruker Daltonics). Prior to database searching with Protein Scape 3.1.3 (Bruker Daltonics) connected to Mascot 2.5.1 (Matrix Science), raw data were processed in Data Analysis 4.2 (Bruker Daltonics). A customized database comprising the sequences of the RGGGPS, RGGGPS<sub>CKr\_K144C<sup>-Am</sup></sub>, RGGGPS<sub>OC\_W139Y</sub> and RGGGPS<sub>OC\_W139ONBY<sup>hv</sup></sub> proteins as well as common contaminants was used for a database search with the following parameters: enzyme specificity trypsin with two missed cleavages allowed, precursor tolerance of

10 ppm, and MS/MS tolerance of 0.04 Da. As general variable modifications were included: deamidation of asparagine and glutamine, oxidation of methionine, and carbamidomethylation or propionamide modification of cysteine. BrEtAm induced modification of cysteine resulting in the lysine mimetic Lys- $\gamma$ -S, was detected as a customized variable modification of lysine. ONBY was detected as 2-nitrobenzyl modification, reduced ONBY as 2-aminobenzyl modification of tyrosine. MS/MS spectra of the specific modifications were inspected manually.

### **3.6 Acknowledgements**

We thank Reinhard Sterner for critical reading of the manuscript, Caroline Hiefinger and Andrea Kneuttinger for helpful discussions, and Christiane Endres, Sonja Fuchs, Sabine Laberer and Jeannette Ueckert for technical support. We are grateful to 2bind GmbH for access to the Prometheus NT.48 instrument (NanoTemper Technologies).

### 3.7 Supporting Information

**Table S3.1.** Oligomerization states and MWs of GGGPS complexes.<sup>a</sup>

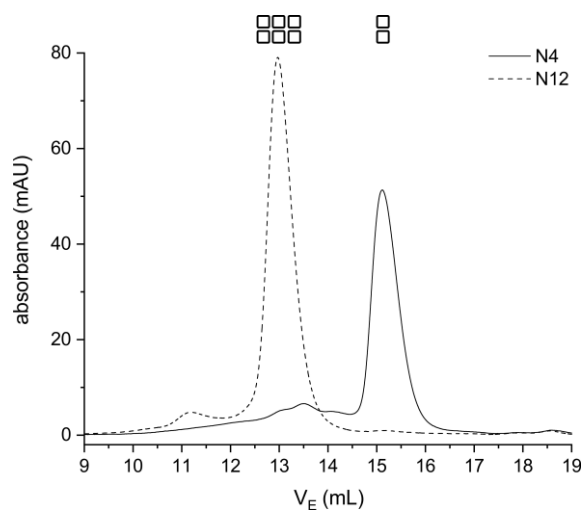
protein [peak] <sup>b</sup>	V <sub>E</sub> (mL)	MW <sup>calc</sup> (kDa)	SLS MW <sup>app</sup> (kDa)	oligo. state
mtGGGPS	12.9	159.15	151.5	hexamer
mtGGGPS <sub>C194A</sub>	12.9	158.96	162.2	hexamer
mtGGGPS <sub>C194A_K146C<sup>Am</sup></sub>	15.1	52.94	55.0	dimer
mtGGGPS <sub>C194A_K146C<sup>+Am</sup></sub>	13.0	160.0	178.3	hexamer
N4	15.1	53.55	57.1	dimer
N12	13.0	160.38	157.6	hexamer
N4_IF_n12 (=RGGGPS)	13.0	161.48	158.6	hexamer
RGGGPS <sub>CKr_K144C<sup>Am</sup></sub>	15.2	53.64	54.8	dimer
RGGGPS <sub>CKr_K144C<sup>+Am</sup></sub> [1]	13.1	162.15	167.0	hexamer
RGGGPS <sub>CKr_K144C<sup>+Am</sup></sub> [2]	15.2	53.64	54.8	dimer
RGGGPS <sub>CKr_K144C<sup>+OH</sup></sub>	15.1	53.64	56.5	dimer
RGGGPS <sub>OC_W139Y</sub> [1]	13.0	160.94	152.5	hexamer
RGGGPS <sub>OC_W193Y</sub> [2]	15.1	53.65	48.9	dimer
RGGGPS <sub>OC_W139ONBY<sup>-hv</sup></sub>	15.1	53.95	50.3	dimer
RGGGPS <sub>OC_W139ONBY<sup>+hv</sup></sub> [1]	13.0	160.94	149.7	hexamer
RGGGPS <sub>OC_W139ONBY<sup>+hv</sup></sub> [2]	15.1	53.95	50.3	dimer

<sup>a</sup> The elution profiles are shown in Figure 3.3 and Figure S3.1. V<sub>E</sub>, elution volume obtained in analytical gel filtration chromatography; MW<sup>calc</sup>, MW calculated from the amino acid sequence for dimer or hexamer; SLS MW<sup>app</sup>, apparent MW as deduced from SLS data; oligo.state, deduced oligomerization state. Characterization of mtGGGPS, N4, N12, and N4\_IF\_n12 has been performed previously (Kropp et al 2020) but has been repeated for this publication.

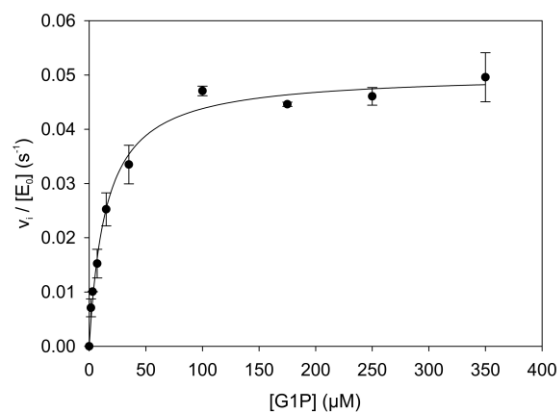
<sup>b</sup> When two peaks eluted, peaks are labelled as [1] and [2].

**Table S3.2.** Oligonucleotide sequences for production of the substitutions by site-directed mutagenesis.

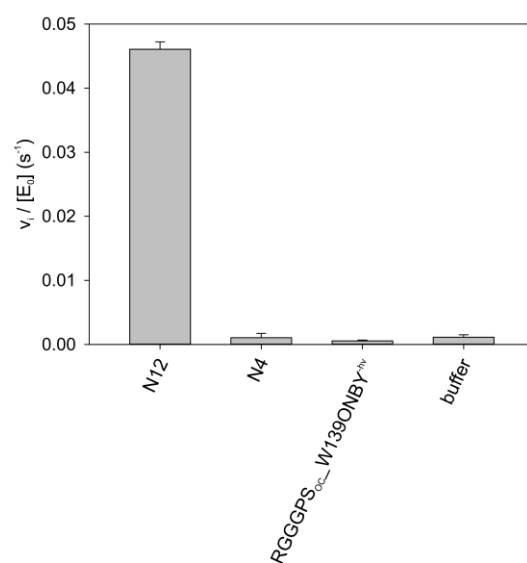
name	forward primer	reverse primer
mtGGGPS_C194A	CCACAGATCAGATACTCATAGTTGGCG	CTCGCTTAACAAGGGCTATCATCTC
mtGGGPS_K146C	CCGGTCCCAGGAACAAACC	GCAGGTGTCACCGACCCATCC
RGGGPS_K144C	CCTATTCGCGTCATAAACCG	GCATGCATCACCAACCCAACC
RGGGPS_W139Y	GTTGGTGATGCAAACCTATTCCG	ATAACCAACGGTTTCACCCGGTTC
RGGGPS_W139amber	GTTGGTGATGCAAACCTATTCCGC	CTAACCAACGGTTTCACCCGGTTC



**Figure S3.1.** Analytical SEC of N4 and N12. 40  $\mu\text{M}$  of protein (subunit concentration) was applied to a S200 10/300 GL analytical column, which has been equilibrated with 50 mM potassium phosphate, pH 7.5, 300 mM KCl. Elution was performed at a flow rate of 0.4 ml min<sup>-1</sup>, followed by measuring the absorbance at 280 nm and plotted against the elution volume. The derived oligomerization state is indicated by symbols. SEC with N4 and N12 have been performed previously (Kropp et al 2020) but have been repeated for this publication.



**Figure S3.2.** Steady-state kinetic measurement of AncGGGPS\_N12. Saturation curves were recorded at 40 °C using a photometric assay, as described in Materials and Methods. The experiment was done in duplicates, the error bars show standard deviations. Kinetic constants were obtained by fitting the Michaelis-Menten equation to the data using SigmaPlot 13.0. The derived parameters are:  $k_{cat}$ :  $0.05 \pm 0.001$  s<sup>-1</sup>,  $K_M$ :  $14.7 \pm 1.6$  μM,  $k_{cat}/K_M$ :  $0.34 \times 10^4$  s<sup>-1</sup> M<sup>-1</sup>.



**Figure S3.3.** Activity test of GGGPS variants N4 and RGGGPS<sub>0C</sub>\_W139ONBY<sup>-hv</sup>. The variants were assayed at 40 °C at fixed concentrations of 250 μM G1P and 11 μM GGPP, as described for steady-state kinetic measurements. N12 served as reference, buffer was used as control. Note that due to phosphate contaminations in the different components of the assay, there is a detectable background activity in the buffer control that resembles the detected activity in N4 and RGGGPS<sub>0C</sub>\_W139ONBY<sup>-hv</sup>.

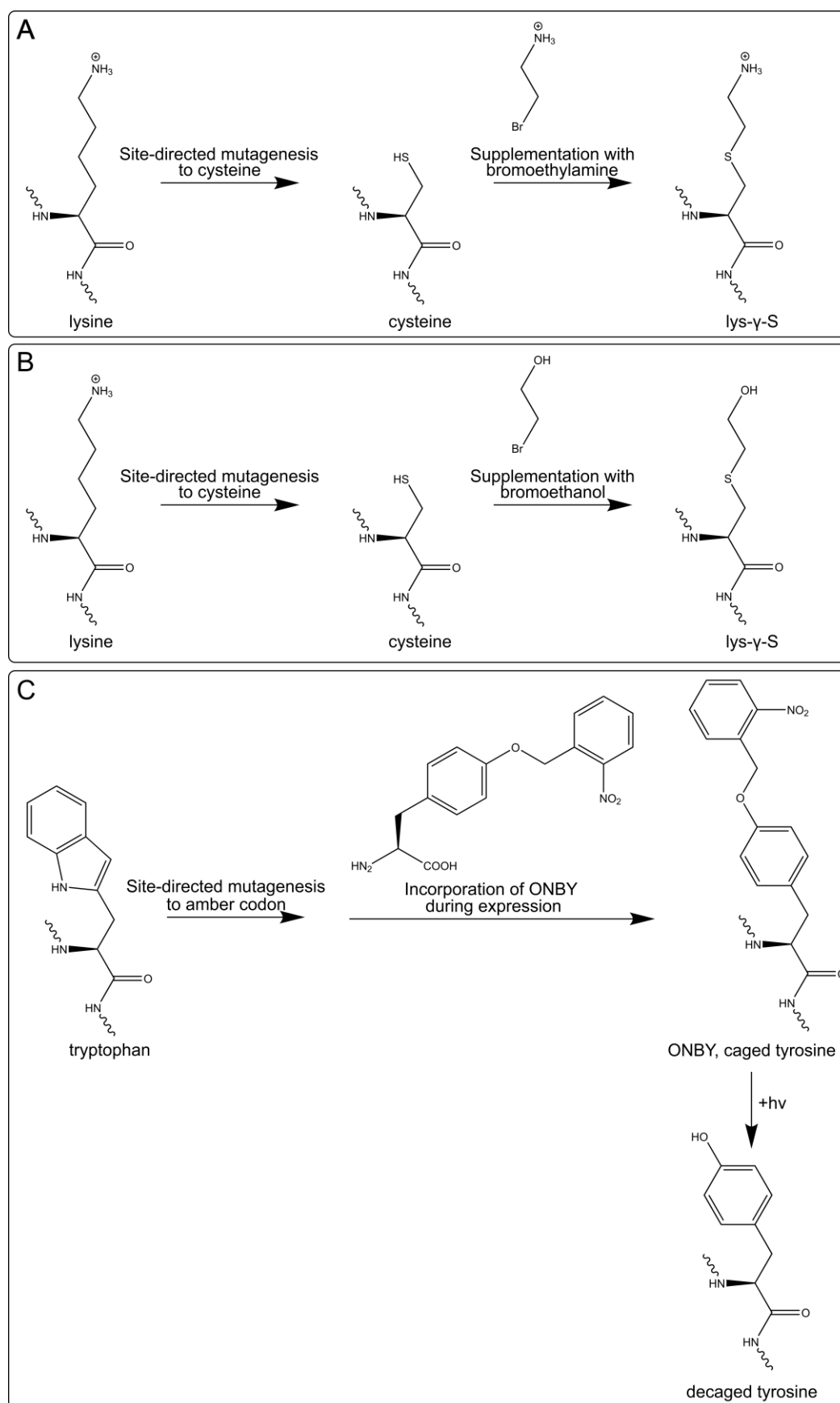
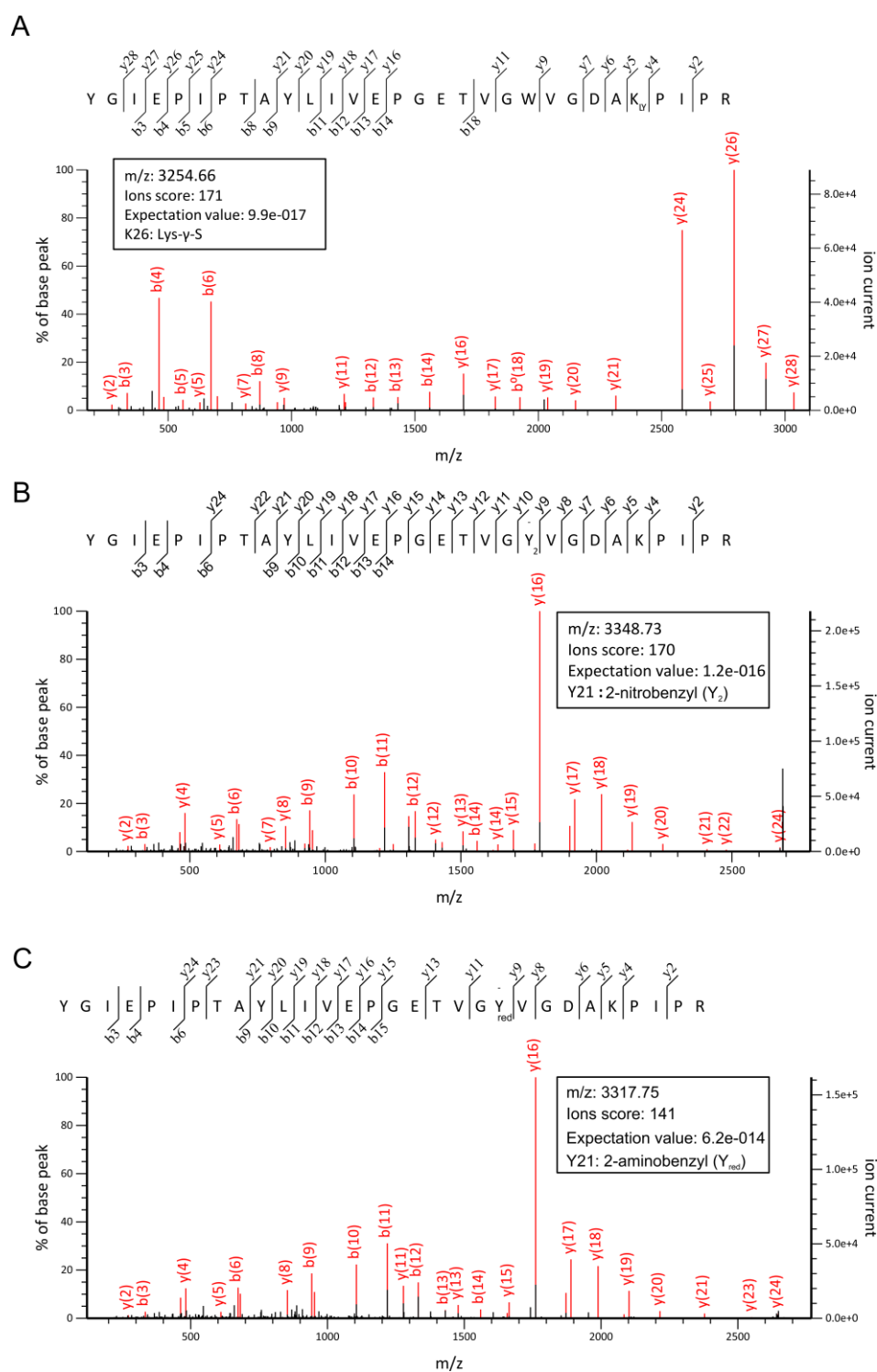
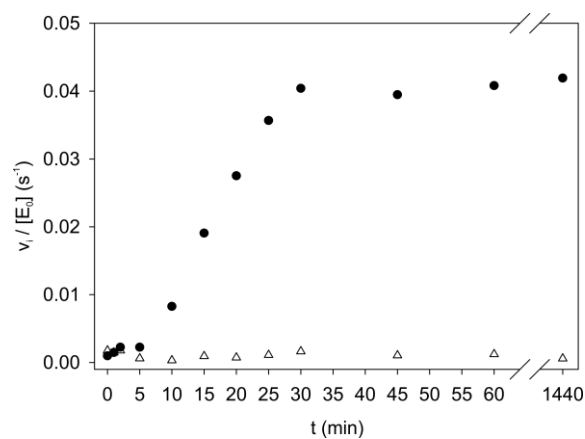


Figure S3.4. Chemical structures of modified amino acids used in this study.

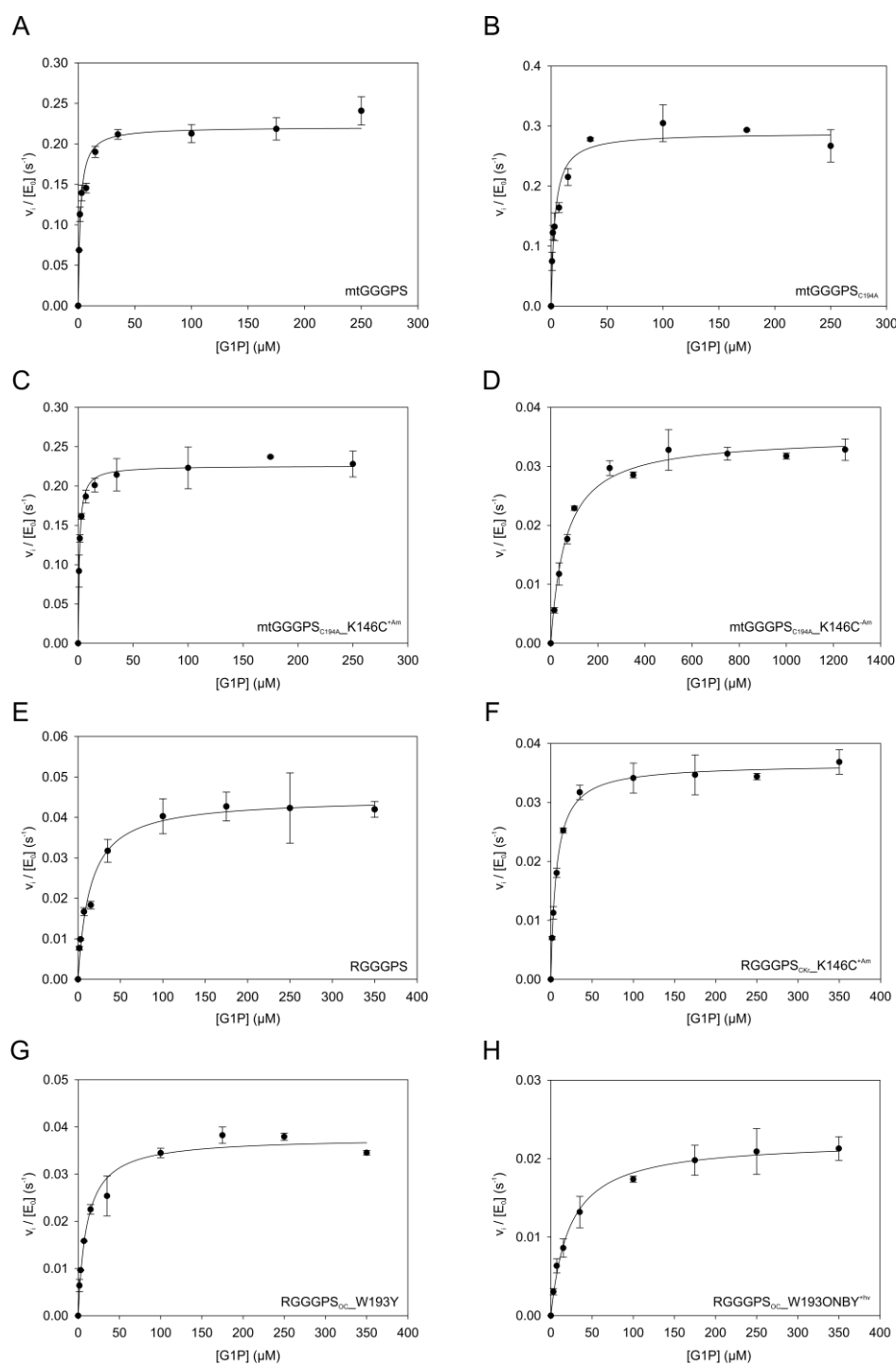


**Figure S3.5.** LC-MS/MS analysis of GGGPS variants. MS/MS spectra of tryptic peptides confirming the localization and identity of (A) Lys- $\gamma$ -S in RGGGPS<sub>CKR</sub>-K144<sup>+Am</sup>, as well as (B) ONBY (2-nitrobenzyl-*O*-tyrosine) and (C) reduced ONBY (2-aminobenzyl-*O*-tyrosine) in RGGGPS<sub>OC</sub>-W139ONBY<sup>hv</sup>. Fragment ions of the  $\gamma$ - and  $b$ - series (red peaks) indicate in (A) the presence of Lys- $\gamma$ -S at position K144 (designated as “K<sub>LY</sub>”), in (B) the presence of ONBY at position Y139 (designated as “Y<sub>2</sub>”) and in (C) the presence of reduced ONBY at the same position (designated as “Y<sub>red</sub>”). Black peaks represent contaminating fragments. The peaks were normalized to the highest signal (100%).

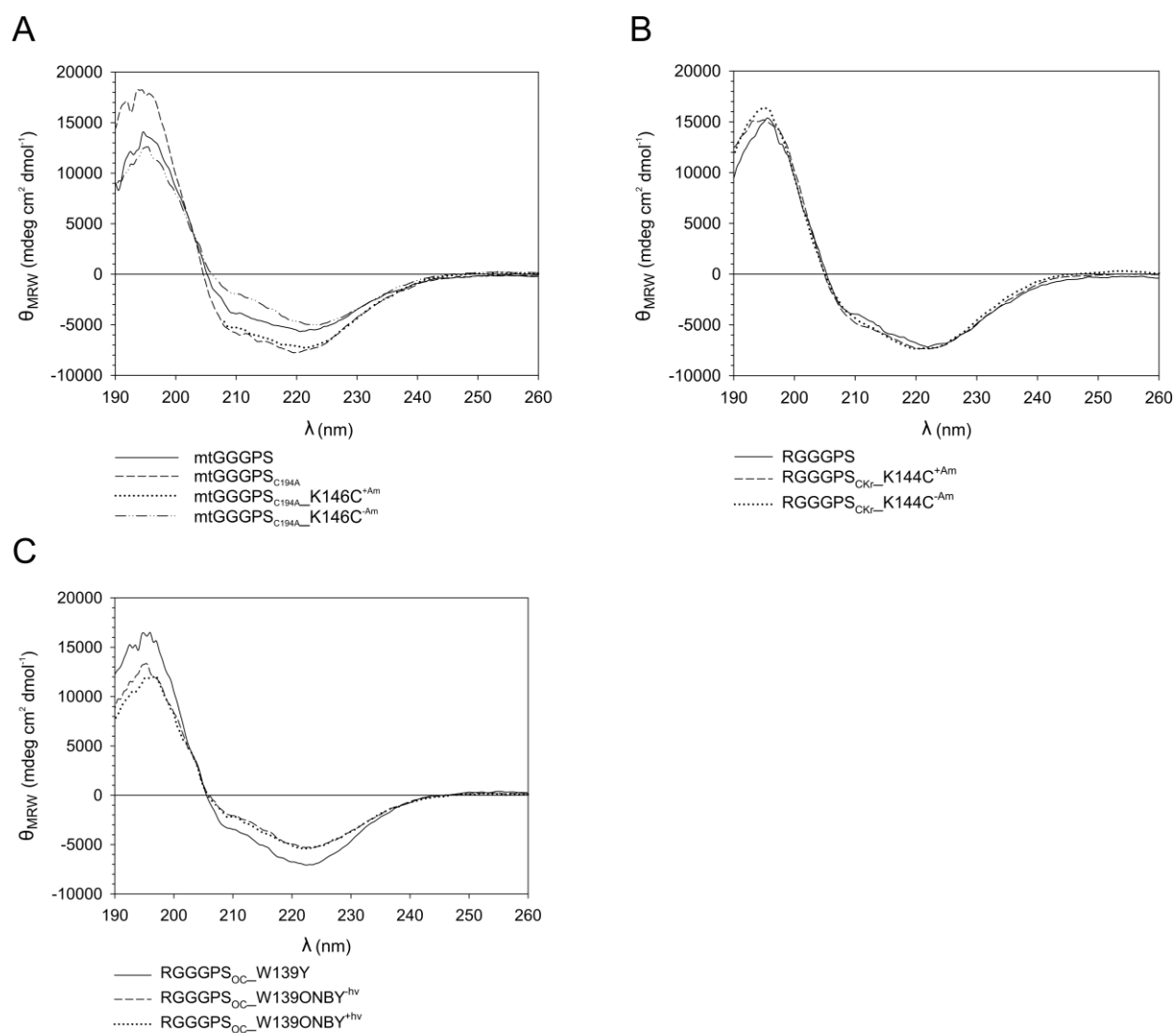




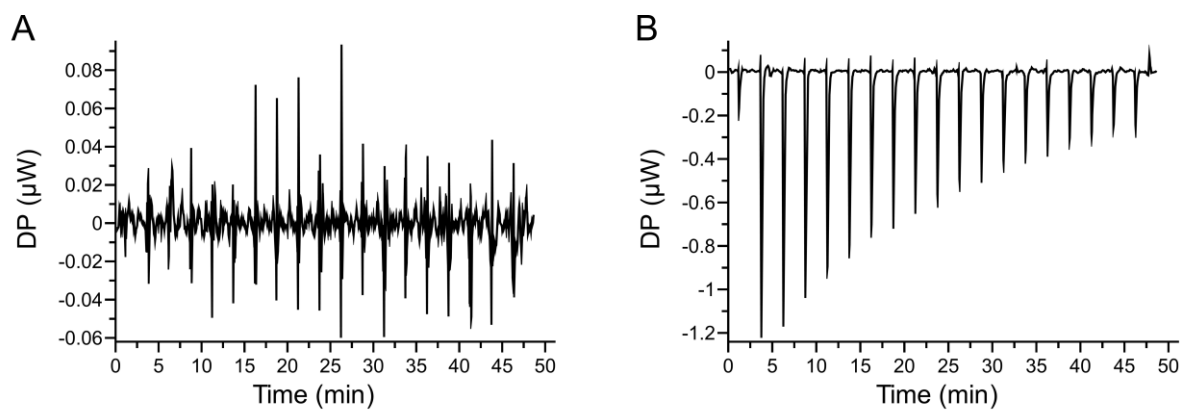
**Figure S3.6.** Reactivation of RGGGPS<sub>CKr</sub>\_K144C. Enzymatic activity was assayed at 40 °C as described for steady-state kinetic measurements at fixed concentrations of 250 μM G1P and 11 μM GGPP. RGGGPS<sub>CKr</sub>\_K144C was incubated without (white triangles) and with 20 mM BrEtAm (black circles) and samples were taken after different times ( $t = 0, 1, 2, 5, 10, 15, 20, 25, 30, 45, 60,$  and 1440 min).



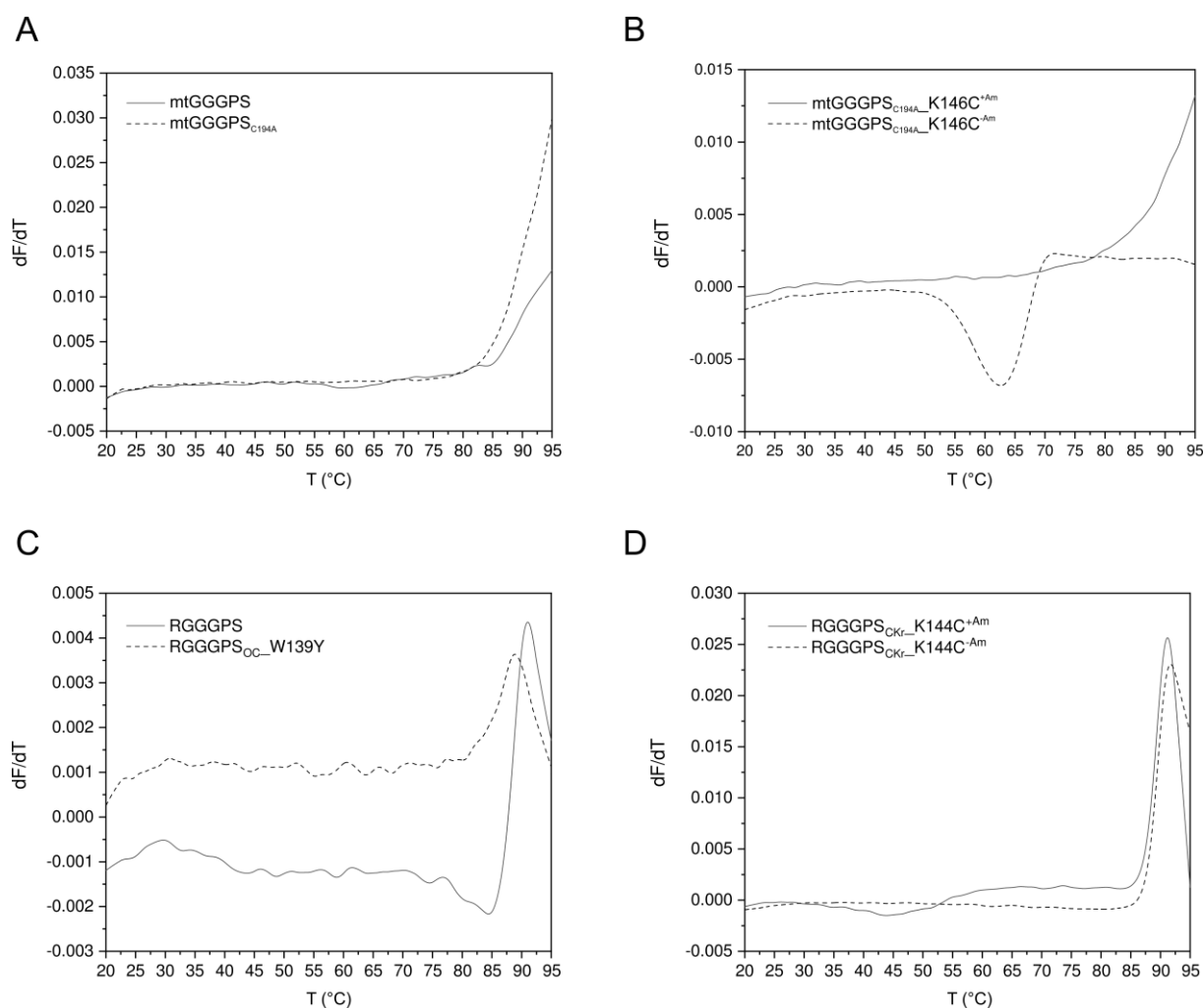
**Figure S3.7.** Steady-state kinetic measurements of (A) mtGGGPS wild-type, (B) mtGGGPS<sub>C194A</sub>, (C) mtGGGPS<sub>C194A\_K146C<sup>+Am</sup></sub>, (D) mtGGGPS<sub>C194A\_K146C<sup>-Am</sup></sub>, (E) RGGGPS (= N4\_IF\_N12), (F) RGGGPS<sub>CKr\_K144C<sup>+Am</sup></sub>, (G) RGGGPS<sub>OC\_W193Y</sub> and (H) RGGGPS<sub>OC\_W193ONBY<sup>+hv</sup></sub>. Saturation curves were recorded at 40 °C using a photometric assay, as described in Materials and Methods. The experiment was done in duplicates, the error bars show standard deviations. Kinetic constants were obtained by fitting the Michaelis-Menten equation to the data using SigmaPlot 13.0. The derived parameters  $k_{cat}$ ,  $K_M$ , and  $k_{cat}/K_M$  are listed in Table 3.1.



**Figure S3.8.** Structural integrity of GGGPS variants characterized in this study. Far-UV CD spectra of all variants (6  $\mu$ M, subunit concentration) were recorded in 10 mM potassium phosphate, pH 7.5 from 190 nm to 260 nm ( $d = 1$  mm) at 25 °C.



**Figure S3.9.** Analysis of substrate binding to GGGPS variants. The binding of the substrate G1P to (A) RGGGPS<sub>CKr\_K144C<sup>Am</sup></sub> and (B) RGGGPS<sub>CKr\_K144C<sup>+Am</sup></sub> was monitored by ITC. G1P was stepwise titrated in 2.5 min intervals to the proteins. The DP between reference and sample cell is plotted against the time.



**Figure S3.10.** Thermal stability of GGGPS variants followed by nanoDSF. The change in the ratio of the fluorescence emission at 350 and 330 nm of (A and B) 20  $\mu\text{M}$  or (C and D) 40  $\mu\text{M}$  protein (subunit concentration) in 50 mM potassium phosphate, pH 7.5 was monitored from 20 to 95  $^{\circ}\text{C}$  at a scan rate of 1  $\text{K min}^{-1}$ , and the first derivative of the fluorescence ratio with respect to temperature was plotted against temperature. Extrema in the curve point to thermal transitions. Note that there is a minimum in the mtGGGPS variant (resulting from a decrease in the  $F_{350}/F_{330}$  ratio), but maxima in the RGGGPS variants (resulting from an increase in the  $F_{350}/F_{330}$  ratio), because the change in exposure to water upon denaturation is different for those variants (Linde et al 2018).



# Chapter 4

## 4 Spectrometric analysis of ether lipids in Bacteria and Archaea

### 4.1 Archaea-type ether lipids in Firmicutes and Bacteroidetes

#### 4.1.1 Summary

Recently, Archaea-type ether lipids have been described in the phylum of Firmicutes, namely in *B. subtilis*. The function of these, however, remains unclear. Interestingly, ether lipids in *B. subtilis* look slightly different as they only contain one C35 isoprenoid chain and are acetylated by an enzyme called YvoF. As discussed in section 4.2, YvoF has also been identified in Halobacteria (Archaea) by phylogenetic analysis, but acetylated lipids have not been known for archaea so far. For understanding the function of ether lipids in bacteria, it is essential to gain knowledge about the quantities as well as to identify ether lipids in further bacterial organisms. To this end, we analyzed different species by ultra-high performance liquid chromatography (UHPLC) coupled to MS. We used *B. subtilis* lipid extracts to quantify the ether lipid amount, and our results indicate that the relative ether lipid content is far below 1% compared to the main “canonical” phospholipid classes. We furthermore attempted identification of ether lipids in Bacteroidetes strains via UHPLC-MS, but we were not able to find any related components.

#### 4.1.2 Introduction

Archaea and Bacteria may have evolved from a LUCA with a heterochiral membrane consisting of G3P and G1P-based lipids with fatty acids and isoprenoids. In a separation event, progenitors of the two primary lineages with homochiral membranes emerged. The “lipid divide” describes the clear-cut differences between homochiral membranes of Archaea and Bacteria: Archaeal membranes are composed of G1P-based ether lipids with isoprenoids, whereas Bacteria possess G3P-based lipids esterified with fatty acids. For a long time, G1P-based ether lipids as well as the synthesizing enzymes, G1PDH and GGGPS, were considered exclusive to archaea. Recently, AraM (G1PDH) was discovered and confirmed to synthesize G1P in *B. subtilis* (Guldan et al 2008) followed by the description of PcrB as GGGPS-like enzyme condensing G1P and the C35-isoprenoid HepPP (Guldan et al 2011). GGGPS enzymes were also discovered in Bacteroidetes

and GGGPS activity with C20-isoprenoids was confirmed *in vitro* (this work (Chapter 2) and in Peterhoff et al (2014)). In contrast to archaea, ether lipids in *B. subtilis* become dephosphorylated by unspecific phosphatases and then diacetylated by the *O*-acetyltransferase YvoF in ensuing reactions (Figure 1.4; Guldán et al 2011, Linde et al 2016). So far, the function of ether lipids in bacteria as well as the quantities have remained elusive.

In Bacteroidetes, the situation is even less clear, since no G1PDH has been found so far in these species. Recent studies hypothesize that these organisms are able to build heterochiral membranes consisting of bacterial ester and archaeal ether lipids (Villanueva et al 2021). Nevertheless, analytic evidence of ether lipids in this phylum is still missing. Interestingly, phylogenetic studies revealed that YvoF homologues are absent in Bacteroidetes but exist in Halobacteria (Linde et al 2016).

### 4.1.3 Results & Discussion

#### 4.1.3.1 Growth of strains and preparation of lipid extracts

Our aim was to quantify all ether lipid derivatives in *B. subtilis* to get an idea about the amounts of these compounds in this species. Furthermore, we pushed identification of ether lipids in Bacteroidetes, which is still missing so far. To this end, *B. subtilis* wild-type,  $\Delta yvoF$ ,  $\Delta pcrB$  and  $\Delta pcrB+pcrB$  strains (for all used strains refer to 4.1.5.2) were cultivated (4.1.6.1), harvested and stored as described under section 4.1.6.3. The wild-type strain was used for quantification, whereas  $\Delta yvoF$ ,  $\Delta pcrB$  and  $\Delta pcrB+pcrB$  strains served as controls. The first two strains should contain no acetylated ether lipids, or no ether lipids at all, respectively, while the last strain overexpresses *pcrB* and therefore should show an elevated level of ether lipids. For experiments with Bacteroidetes, *Flavobacterium johnsoniae*, *Chitinophaga pinensis*, *Spirosoma linguale* and *Zunongwangia profunda* were cultivated (4.1.6.2.1) and strain identity was confirmed afterwards by sequencing (4.1.6.2.2) in order to rule out any strain contamination. To this end, single colonies were produced by streaking and used as template in a colony polymerase chain reaction (PCR) with 16S ribosomal RNA gene primers (Klindworth et al 2013), which is a specific sequence for most organisms. The amplicons were sequenced and entered in a BLAST search to confirm strain identity. The cultures were harvested and stored as described in 4.1.6.3.

Since there is no *a priori* knowledge about the dependence of the ether lipid synthesis on different cultivation conditions in bacteria (neither for Firmicutes nor for Bacteroidetes), conditions for all strains (temperature and medium, 4.1.5.3) were chosen according to the German Collection of Microorganisms and Cell Cultures (DSMZ), referred to as “standard lab cultivation conditions”. For all tested strains, cultivation optimization was performed regarding shaking, filling amount of the incubation flasks, and incubation time to guarantee “optimal” oxygenation and growth. All parameters for all are indicated in Table 4.1. Cultivation of all strains was initiated with a pre-culture (used for inoculation of main cultures to 0.1 OD<sub>600</sub>) and then



continued in main culture settings (cells were harvested when exceeding 4 OD<sub>600</sub>). An exception was *Z. profunda*, which showed no growth after transfer from pre into main culture. Therefore, cultivation was initiated directly from glycerol stock using main culture conditions (Table 4.1).

**Table 4.1.** Culture conditions of strains used for UHPLC-MS analysis.<sup>a</sup>

organism	strain	pc (d)	mc (d)	capacity: filling (L)	Vol (L)	T (°C)	rpm
<i>B. subtilis</i>	wild-type	1	3	3:1	4	37	140
	$\Delta pcrB$	1	3	3:1	4	37	140
	$\Delta pcrB+pcrB$	1	4	3:1	4	37	140
	$\Delta yvoF$	1	2	3:1	4	37	140
<i>F. johnsoniae</i>	wild-type	1	2	3:1	4	30	140
<i>C. pinensis</i>	wild-type	2	2	2:0.5	1	22	140
<i>S. linguale</i>	wild-type	5	3	2:0.5	1	26	140
<i>Z. profunda</i>	wild-type	-	4	2:0.5	1	28	140

<sup>a</sup> Most strains were grown in pre cultures (pc) for the given number of days (full 24 h) (d). Afterwards, they were transferred into main cultures (mc) and growth was continued for the given number of days. The flask capacity to actual filling amount ratio of the incubation flasks was as indicated (capacity:filling). To ensure sufficient yield, the given total main culture volume (Vol) was used. Incubation of pre and main culture was performed at the indicated temperatures (T) whilst shaking (rpm) in a Multitron shaker (INFORS HT, Basel).

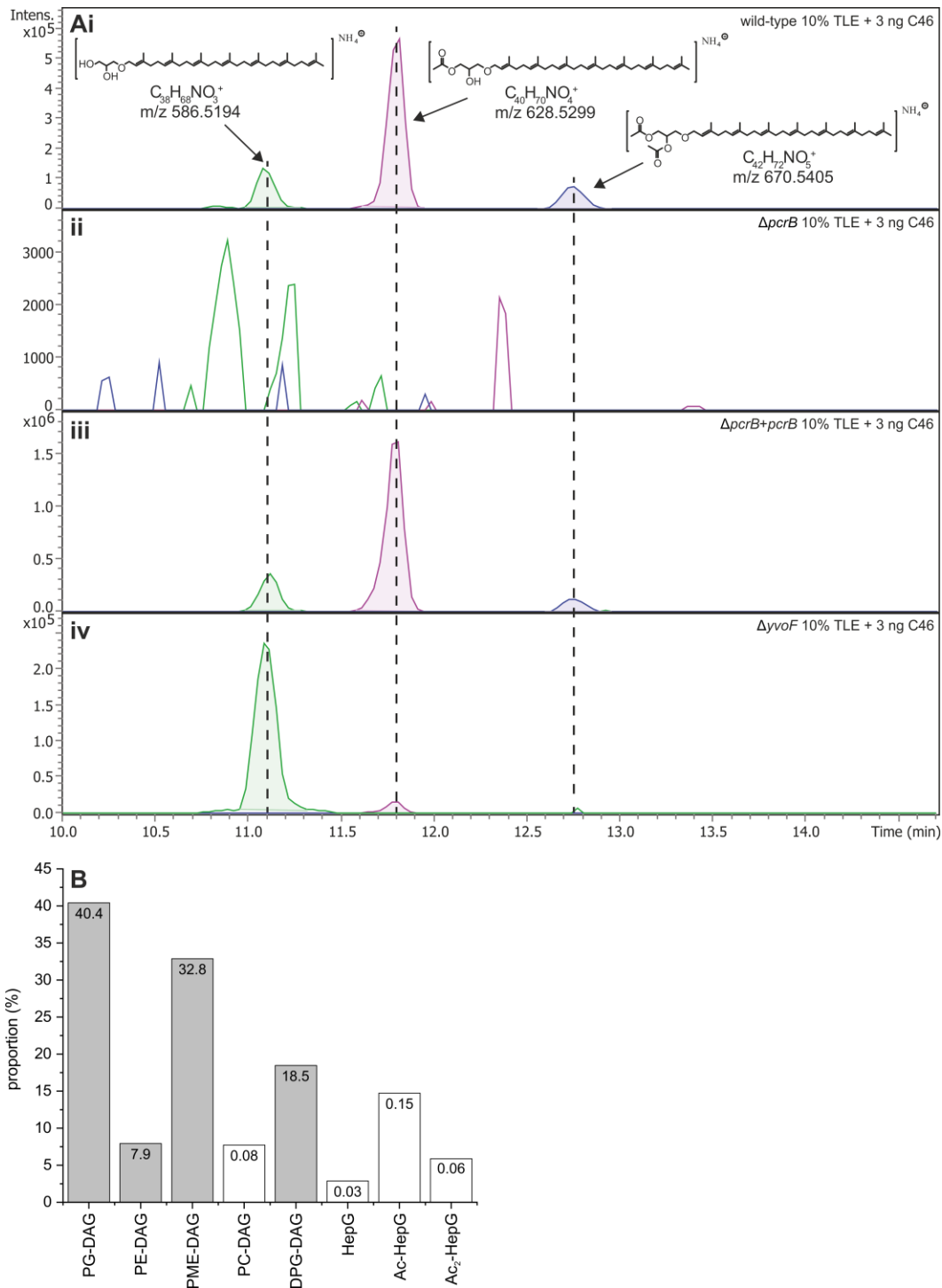
After strain preparation, total lipid extracts (TLE) of all strains were produced. This was done by isolating mainly hydrophobic membrane components via extraction (4.2.5.7). The TLE components were then separated and analyzed by means of reversed phase (RP) UHPLC electrospray ionization MS (RP-UHPLC-ESI-MS, 4.2.5.8). In RP-UHPLC-ESI-MS, molecules are separated by alkyl chain hydrophobicity. This technique separates compounds from of a wide range of polarities and offers very good separation of lipids differing in the degree of unsaturation, cyclization or length of the carbon chain. Thus, lipids with the same headgroup but with slight differences in the hydrophobic chain can be separated effectively. This was considered quite useful for the identification of relatively apolar ether lipids in Bacteroidetes since there was no prior knowledge about the length of the isoprenoid chains.

Production of the TLEs and UHPLC-MS experiments were conducted at the lab of Prof. Dr. Hinrichs, MARUM (Bremen). Transportation of the pelleted cultures took place on dry ice.

#### 4.1.3.2 Quantification of ether lipids in *B. subtilis*

The TLEs were subjected to RP-UHPLC-ESI-MS to separate and analyze the hydrophobic membrane components, such as ether lipids. *B. subtilis* wild-type showed peaks for non-, mono- and diacetylated ether lipids at retention times that match the literature (Figure 4.1A(i)) (Guldan et al 2011). The control strains  $\Delta pcrB$ ,  $\Delta pcrB+pcrB$  and  $\Delta yvoF$  generated results as expected: The  $\Delta pcrB$  strain showed no ether lipids at all as indicated in Figure 4.1A(ii), in  $\Delta pcrB+pcrB$  ether lipids were over-synthesized by a factor of 3-5 estimated from peak height (iii), and the  $\Delta yvoF$  strain only possessed non-acetylated ether lipids (iv).

*B. subtilis* wild-type was examined to quantify the ether lipid fraction in RP-UHPLC-ESI-MS measurements. The *B. subtilis* lipidome consists of ~70% phospholipids and ~30% neutral glucolipids (Seydlova & Svobodova 2008, Willdigg & Helmann 2021) and UHPLC-MS analysis of  $\text{NH}_4^+$  adducts allowed for the identification of the main “canonical” phospholipids (data not shown, retention times matched the literature): phosphatidylglycerols (PG-DAG), phosphatidylethanolamines (PE-DAG), phosphatidyl-(N)-methylethanolamines (PME-DAG), phosphatidylcholines (PC-DAG), and cardiolipins (diphosphatidylglycerol, DPG-DAG). These phospholipids were used as reference to relatively estimate the content of ether lipids (from now on also referred to as monoether lipids) in *B. subtilis*. Because monoether lipids (heptaprenylglyceryl, HepG) and acetylated monoether lipids (Ac- and Ac<sub>2</sub>-HepG) are not commercially available, their mass spectrometric response factors could not be determined. Therefore, the proportion of HepG, Ac- and Ac<sub>2</sub>-HepG compared to the main phospholipids was only estimated roughly, assuming that the monoether lipids and the phospholipids are detected in RP-UHPLC-ESI-MS with the same mass spectrometric response. As derived from the peak heights, the relative content of the monoether lipids would be approximately i) 0.0288% for HepG, (ii) 0.1472% for Ac-HepG, and (iii) 0.0588% for Ac<sub>2</sub>-HepG, whereas the phospholipids accounted for 99,7652%. (Figure 4.1B, refer to 4.2.5.8 for the method and for relative content estimation see digital supplement 4.1.7). The neutral glucolipids that account for ~30% of the lipidome were not considered in this analysis. The experiments were conducted together with Julius Lipp, MARUM (Bremen).

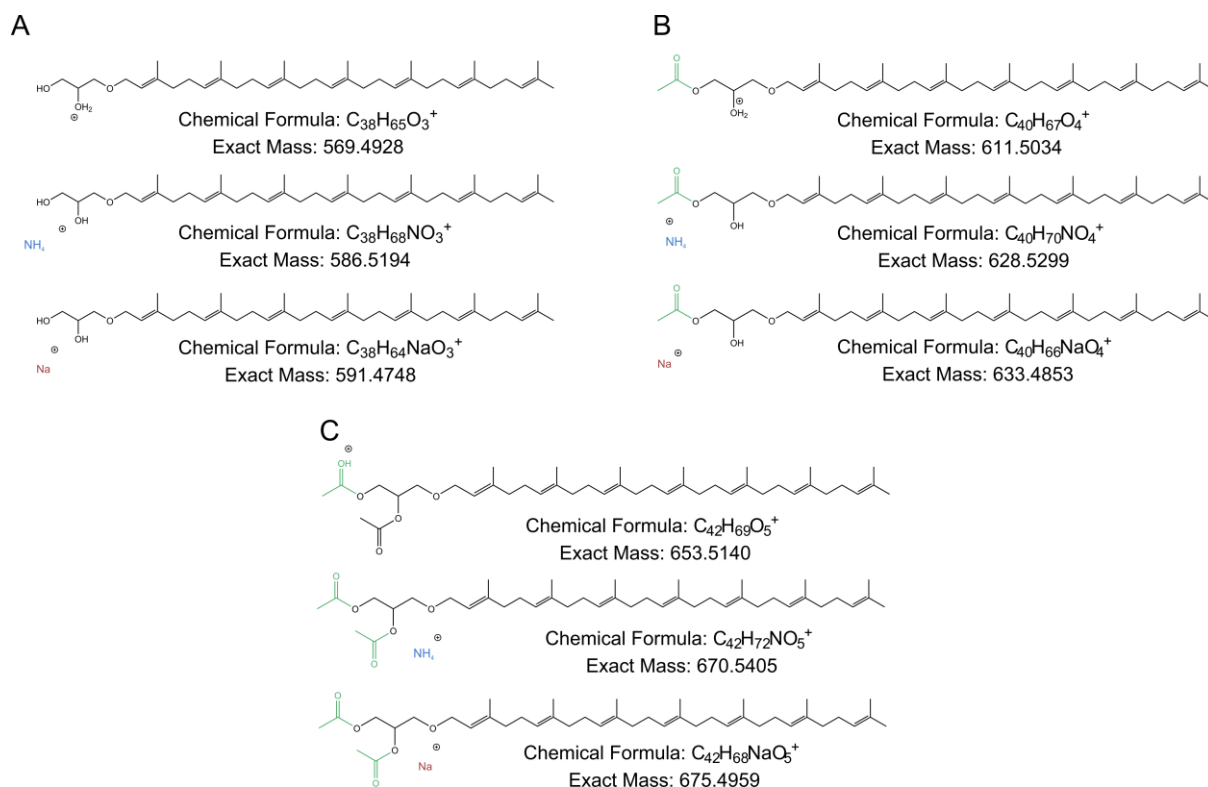


**Figure 4.1.** RP-UHPLC-ESI-MS analysis of total lipid extracts (TLE) from *B. subtilis* wild-type and mutant variants. (A) Superimposed extracted ion chromatograms of  $m/z$  586.5194 (green) representing non-acetylated,  $m/z$  628.5299 (pink) representing monoacetylated, and  $m/z$  670.5405 (blue) representing diacetylated monoether-lipid derivatives in (i) wild-type, (ii)  $\Delta pcrB$ , (iii)  $\Delta pcrB+pcrB$  and (iv)  $\Delta yvoF$  strains. The three derivatives are given as  $NH_4^+$  adducts. (B) Relative proportion of the main phospholipid classes and monoether-lipid derivatives in the TLE from *B. subtilis* wild-type. White bars were multiplied by 100 to render them visible in the figure, the depicted values however are not multiplied. PG-DAG, phosphatidylglycerol; PE-DAG, phosphatidylethanolamine; PME, phosphatidyl-(N)-methylethanolamines; PC-DAG, phosphatidylcholine DAG; DPG-DAG, diphosphatidylglycerol; HepG, heptaprenylglyceryl (monoether-lipid); Ac-HepG, monoacetylated HepG; Ac<sub>2</sub>-HepG, diacetylated HepG. C46 refers to the added standard (for the method refer to 4.2.5.8).

Our experiment indicates that the ether lipid fraction in this organism is rather small compared to the class of the phospholipids at standard lab cultivation conditions. This small but defined amount implies that these components could act as structure-giving membrane lipids for fine-tuning of certain membrane properties, or act as signal molecule. The knockout of  $\Delta\text{perB}$  and  $\Delta\text{yvoF}$  causes diminished cell growth in standard growth conditions which has been observed previously (Guldán 2010) and furthermore, data indicate that biofilm formation is diminished when ether lipid synthesis is absent (conducted in the course of this thesis, data not shown). This implies that the cells are, even though Archaea-type ether lipids only account for such a small amount in the membrane, stressed by the absence of these molecules indicating considerable importance of ether lipids. In this framework, it would be interesting to see whether the ether lipid synthesis level depends on the growth conditions when cultivating *B. subtilis*.

#### 4.1.3.3 Ether lipids in Bacteroidetes

To identify ether lipids in Bacteroidetes, we subjected the TLE of *F. johnsoniae*, *C. pinensis*, *S. linguale* and *Z. profunda* to RP-UHPLC-ESI-MS. Since both the isoprenoid chain length and the degree of unsaturation were unknown, a broad range of potential monoether lipid derivatives was searched after the following criteria: (i) Different isoprenoid chain lengths were searched for (C15, C20, C25, C30, and C35) and since Bacteroidetes GGGPS enzymes belong to group II that specifically turn over C20 but not C35 in *B. subtilis in vivo* experiments (Peterhoff et al 2014), C20 was analyzed in more detail as it was considered completely unsaturated or saturated during the search. (ii) Furthermore, although no YvoF homologue is known to exist in Bacteroidetes, monoether lipids were searched for as non-acetylated, monoacetylated and diacetylated. (iii) At last, the different adduct combinations had to be considered as well, since the compounds can be ionized as protonated ( $\text{H}^+$ ), ammoniated ( $\text{NH}_4^+$ ), or sodiated ( $\text{Na}^+$ ) adducts (for clarification see Figure 4.2)



**Figure 4.2.** Search criteria for UHPLC-MS analysis of Bacteroidetes TLEs. The monoether lipid derivatives were considered with different chain lengths (C15, C20, C25, C30, C35), a C35 isoprenoid length is shown exemplarily. Different adducts ( $H^+$ ,  $NH_4^+$  (blue),  $Na^+$  (red)) were considered without and with derivatization: (A) non-acetylated, (B) monoacetylated, (C) diacetylated. The elemental composition and the exact mass were calculated theoretically and were used as search criterion.

We could not find any monoether lipid components in the four investigated Bacteroidetes strains (data not shown). Several studies have confirmed GGGPS activity of Bacteroidetes strains *in vitro* (this work (Chapter 2) and previously (Linde et al 2018, Peterhoff et al 2014)), but it has remained unclear, how Bacteroidetes can provide themselves with G1P since no G1PDH is encoded in the genome. In a recent study an *E. coli* strain was engineered by establishing genes for the biosynthesis of archaeal diether lipids. The *E. coli* cells showed a heterochiral membrane and were interestingly more resistant against for instance elevated temperature. In the same work, another strain was created that had all genes for Archaea-type diether lipid synthesis except for the G1PDH. Surprisingly, in the absence of the G1PDH, still substantial amounts of diether lipids were synthesized (Caforio et al 2018). This means that *E. coli* possesses an unknown G1P synthesis route, which could also apply for the Bacteroidetes strains investigated here. It remains elusive when this hypothetical G1P synthesis route is activated and ether lipid synthesis occurs, which provides an explanation for why we were not able to find any of them in our MS analysis. Furthermore, metagenomic analysis of the Fibrobacteres–Chlorobi–Bacteroidetes (FCB) group members, which the Bacteroidetes are part of, implies that these organisms could have a DGGGPS encoded and thus be able to synthesize diether lipids, which we did not search for. In

general, it becomes obvious that the mechanisms of ether lipid synthesis in these organisms are not understood yet and need more research.

#### 4.1.4 Conclusion

We quantified the relative ether lipid content compared to the main “canonical” phospholipid classes that account for around 70% of the membrane lipids of *B. subtilis* (Seydlova & Svobodova 2008, Willdig & Helmann 2021). Our data indicated that the relative ether lipid amount is far below 1%. The small but defined amount implies that these compounds fulfill a distinct function within the cell, such as providing fine-tuning of membrane fluidity under different growth conditions. The acetyl group could hereby act as small headgroup changing the polarity of the lipid and by that introduce different functionalities for the non-acetylated or acetylated derivatives. The organism might adjust the synthesis levels of acetylated ether lipids as response to changing environmental conditions. In this context, it is also unknown, whether the non-acetylated derivative only serves as intermediate for the acetylated ether lipids, or if itself can also be considered an “end product”.

Despite  $\Delta pcrB$  and  $\Delta yvoF$  knockout strains showing diminished growth behavior (Guldan 2010), the knockout is not lethal for *B. subtilis*. It would be interesting to see, if a GGGPS knockout in Bacteroidetes has a similar effect or if it actually is lethal for the strains. Of course, lethality of the knockout only realistic if ether lipids are present in the cells, which is controversially discussed in this section, since we were not able to find any. The genetic accessibility of Bacteroidetes strains is limited and creating a knockout is rather difficult, however, methods to genetically modify Bacteroidetes have been developed recently (Rhodes et al 2011, Shrivastava et al 2013). Generation of an ether lipid synthesis knockout strain and subsequent MS analysis of the lipid extract should be considered for future experiments, as it could provide insight into identification of these lipids in this phylum. Metagenomic data indicate that Bacteroidetes might possess an unknown gene for a DGGGPS (Villanueva et al 2021), meaning that representatives of this phylum might be able to synthesize diether lipids, which was so far not considered in our experiments.

On a functional point of view, comparison of ether lipids in Bacteroidetes and Firmicutes becomes quite interesting. Taken all discussed things together, the missing of ether lipids in Bacteroidetes under standard lab cultivation conditions (for whatever reason: systematic error, or biological reason), the putative missing of acetylation by YvoF or the potential occurrence of a DGGGPS in Bacteroidetes implies that ether lipids might exhibit different functions in these two phyla. Consequently, I believe that ether lipids in Firmicutes and Bacteroidetes must be considered and examined separately.

## 4.1.5 Materials

### 4.1.5.1 Oligonucleotides

**Table 4.2.** Oligonucleotides for Bacteroidetes strain identification.

name	sequence (5' to 3')
S-D-Bact-0341-b-S-17	CCTACGGGNGGCWGCAG
S-D-Bact-0785-a-A-21	GACTACHVGGGTATCTAATCC

### 4.1.5.2 Strains

**Table 4.3.** Strains used in this project.

name	description <sup>a</sup>	origin	use
<i>B. subtilis</i> subsp. <i>subtilis</i> str. 168	Marburg; ATCC 6051, tryptophan auxotrophic	(Zeigler et al 2008)	Strain used for UHPLC-MS. In the text referred to as “wild-type”.
<i>B. subtilis</i> subsp. <i>subtilis</i> str. 168 $\Delta$ <i>pcrB</i>	$\Delta$ <i>pcrB</i> , Ery <sup>R</sup>	(Kobayashi et al 2003) Prof. Dr. Hecker (Universität Greifswald)	Strain used for UHPLC-MS. In the text referred to as $\Delta$ <i>pcrB</i> .
<i>B. subtilis</i> subsp. <i>subtilis</i> str. 168 $\Delta$ <i>pcrB</i> + <i>pDG148_</i> <i>bspcrB</i>	$\Delta$ <i>pcrB</i> , Ery <sup>R</sup> , <i>pDG148-StuI</i> mit <i>bspcrB</i> , Kan <sup>R</sup>	(Guldán 2010)	Strain used for UHPLC-MS. In the text referred as $\Delta$ <i>pcrB</i> + <i>pcrB</i> .
<i>B. subtilis</i> subsp. <i>subtilis</i> str. 168 $\Delta$ <i>yvoF</i>	$\Delta$ <i>yvoF</i> , Ery <sup>R</sup>	(Linde et al 2016), National BioResource Project <sup>b</sup>	Strain used for UHPLC-MS. In the text referred to as $\Delta$ <i>yvoF</i> .
<i>F. johnsoniae</i>	ATCC 17061 pectolytic	(Beer 2012), German Collection of Microorganisms and Cell Cultures (DSMZ)	Strain used for UHPLC-MS.
<i>C. pinensis</i>	ATCC 43595 chitin-degrading	(Beer 2012), DSMZ	Strain used for UHPLC-MS.

<i>S. linguale</i>	Claus 1, ATCC 33905	(Beer 2012), DSMZ	Strain used for UHPLC-MS.
<i>Z. profunda</i>	SM-A87	(Beer 2012), DSMZ	Strain used for UHPLC-MS.

<sup>a</sup> genetic nomenclature as stated in the literature (Bachmann 1990)

<sup>b</sup> National BioResource Project from National Institute of Genetics, Microbial Genetics Laboratory in Japan

### 4.1.5.3 Buffers and solutions

#### 4.1.5.3.1 Miscellaneous stock solutions

CaCl <sub>2</sub> x 2 H <sub>2</sub> O	1 M dissolved in H <sub>2</sub> O, filter-sterilized and stored at RT
glucose	20% (w/v) dissolved in H <sub>2</sub> O, filter-sterilized and stored at RT
glycerol	87% (v/v) diluted in H <sub>2</sub> O, autoclaved and stored at RT

#### 4.1.5.3.2 *B. subtilis* solutions and media

erythromycin (1000x)	0.8 mg mL <sup>-1</sup> dissolved in 100% ethanol
kanamycin (1000x)	7.5 mg mL <sup>-1</sup> dissolved in H <sub>2</sub> O
lysogeny broth (LB)	0.5% (w/v) yeast extract 1.0% (w/v) NaCl 2.0% (w/v) tryptone 1.5% (w/v) Bacto-Agar

#### 4.1.5.3.3 Media for Bacteroidetes cultivation

The media listed below are composed as described on the website of DSMZ, if not stated otherwise (German Collection of Microorganisms and Cell Cultures GmbH, <https://www.dsmz.de/collection/catalogue/microorganisms/culture-technology/list-of-media-for-microorganisms>).

**Table 4.4.** Medium 67 for *F. johnsoniae* and *C. pinensis* cultivation

casitone	3 g L <sup>-1</sup>
yeast extract	1 g L <sup>-1</sup>
CaCl <sub>2</sub> x 2 H <sub>2</sub> O <sup>a</sup>	1.36 g L <sup>-1</sup>
Bacto-Agar	15 g L <sup>-1</sup> pH 7.2 (5 M NaOH)

<sup>a</sup> filter-sterilized CaCl<sub>2</sub> x 2 H<sub>2</sub>O is added after autoclaving



**Table 4.5.** Medium 7 for *S. linguale* cultivation

glucose <sup>b</sup>	1 g L <sup>-1</sup>
yeast extract	1 g L <sup>-1</sup>
peptone	1 g L <sup>-1</sup>
Bacto-Agar	15 g L <sup>-1</sup>

<sup>b</sup> filter-sterilized glucose is added after autoclaving

**Table 4.6.** Modified Medium 514 for *Z. profunda* cultivation

NaCl	10 g L <sup>-1</sup>
yeast extract	1 g L <sup>-1</sup>
bacto peptone	5 g L <sup>-1</sup>
sea salts <sup>c</sup>	37 g L <sup>-1</sup>
Bacto-Agar	15 g L <sup>-1</sup>

<sup>c</sup> ordered as sea salt mix from Merck KGaA

#### 4.1.6 Methods

##### 4.1.6.1 Cultivation and storage of *B. subtilis* strains

*B. subtilis* strains (4.1.5.2) were cultivated at 40 °C whilst shaking at 140 rpm (1 L cultures in 3 L flasks) or 170 rpm (5 mL cultures in 50 mL flasks and 100 mL cultures in 500 mL flasks). For cultivation, LB medium was used. For plasmid-harboring strains, the medium was supplemented with 0.8 µg mL<sup>-1</sup> erythromycin using a filter-sterilized, 1000-fold concentrated stock solution. For knockout strains, the medium was supplemented with 7.5 µg mL<sup>-1</sup> kanamycin using a filter-sterilized, 1000-fold concentrated stock solution. To obtain single colonies, the cell suspension was plated on agar plates containing the adequate antibiotics, and incubated over night at 37 °C. Pre-cultures were inoculated from glycerol stocks and the main cultures were inoculated to OD<sub>600</sub> of 0.1. Growth was continued until an OD<sub>600</sub> of 4 was exceeded. For temporary storage, plates and suspensions were sealed with parafilm and stored at 4 °C. For long-term storage, glycerol cultures were prepared. For this purpose, 150 µL overnight cell culture was mixed with 850 µL 87% glycerol, and stored in a sterile screw cap reaction vessel at -80 °C.

##### 4.1.6.2 Experiments with Bacteroidetes strains

###### 4.1.6.2.1 Cultivation and storage

*F. johnsoniae*, *C. pinensis*, *Z. profunda*, *S. linguale* (4.1.5.2) were cultivated whilst shaking at 140 rpm (1 L cultures in 3 L flasks) or 170 rpm (5 mL cultures in 50 mL flasks and 100 mL cultures in 500 mL flasks) at temperatures in the preferred range (30 °C *F. johnsoniae*, 25 °C *C. pinensis*, 28 °C *Z. profunda*, 25 °C *S. linguale*) in their respective media (4.1.5.3.3). Pre-

cultures were inoculated from glycerol stocks and the main cultures were inoculated to OD<sub>600</sub> of 0.1. Growth was continued until an OD<sub>600</sub> of 4 was exceeded. For temporary storage, plates and suspensions were sealed and stored at 4 °C. For long-term storage, glycerol cultures were prepared. For this purpose, 150 µL cell culture was mixed with 850 µL 87% glycerol, and stored in a sterile screw cap reaction vessel at -80 °C.

#### **4.1.6.2.2 Confirmation of strain identity by sequencing**

For evaluation of the identity of strains (*F. johnsoniae*, *C. pinensis*, *Z. profunda*, *S. linguale*) 16S ribosomal RNA gene primers were used for sequencing (Klindworth et al 2013). Strains were grown in their respective medium at their preferred temperature and streaked to obtain single colonies. Single colonies were used as templates using 16S ribosomal RNA gene primers (4.1.5.1) in a colony PCR (final concentration: 1x reaction buffer, 0.5 µM primer (each), 20 nM dNTPs (each), 1 U Phusion® HF polymerase). The amplicons were sequenced and the sequencing results were blasted confirming strain identity.

#### **4.1.6.3 Harvest and storage followed by transportation**

Cell cultures were harvested by centrifugation at 3993 rcf for 20-40 min at 4 °C. Cell pellets were stored at -80 °C in 50 mL reaction tubes. For transportation, the reaction tubes were stored on dry ice.

#### **4.1.6.4 Disposal of cells**

Microbial cultures as well as contaminated equipment were autoclaved for 20 min before disposal.

#### **4.1.7 Digital supplement**

An excel file (BS\_quantification.xlsx) with the dilution series to determine the relative response factors compared to the C46-GTGT standard as well as the relative quantification of ether lipids is provided separately. The file can be found at <https://epub.uni-regensburg.de/publications/eldiss.html>.

## 4.2 Identification of acetylated diether lipids in halophilic Archaea

Cosimo Kropp<sup>[1]</sup>, Julius Lipp<sup>[2]</sup>, Christina Seisenberger<sup>[1],[3]</sup>, Mona Linde<sup>[1],[4]</sup>, Kai-Uwe Hinrichs<sup>[2]</sup>, Patrick Babinger<sup>[1]\*</sup>

<sup>[1]</sup> Institute of Biophysics and Physical Biochemistry, Regensburg Center for Biochemistry, University of Regensburg  
93040 Regensburg, Germany

<sup>[2]</sup> MARUM, Center for Marine Environmental Sciences, University of Bremen  
28359 Bremen, Germany

<sup>[3]</sup> present address: Roche Diagnostics GmbH,  
82377 Penzberg, Germany

<sup>[4]</sup> present address: Boehringer Ingelheim Pharma GmbH & Co. KG.,  
88400 Biberach an der Riß, Germany

\*Corresponding Author: Patrick Babinger

E-mail: patrick.babinger@ur.de

Tel.: +49 941 943 1634; Fax: +49 941 943 2813

### 4.2.1 Abstract

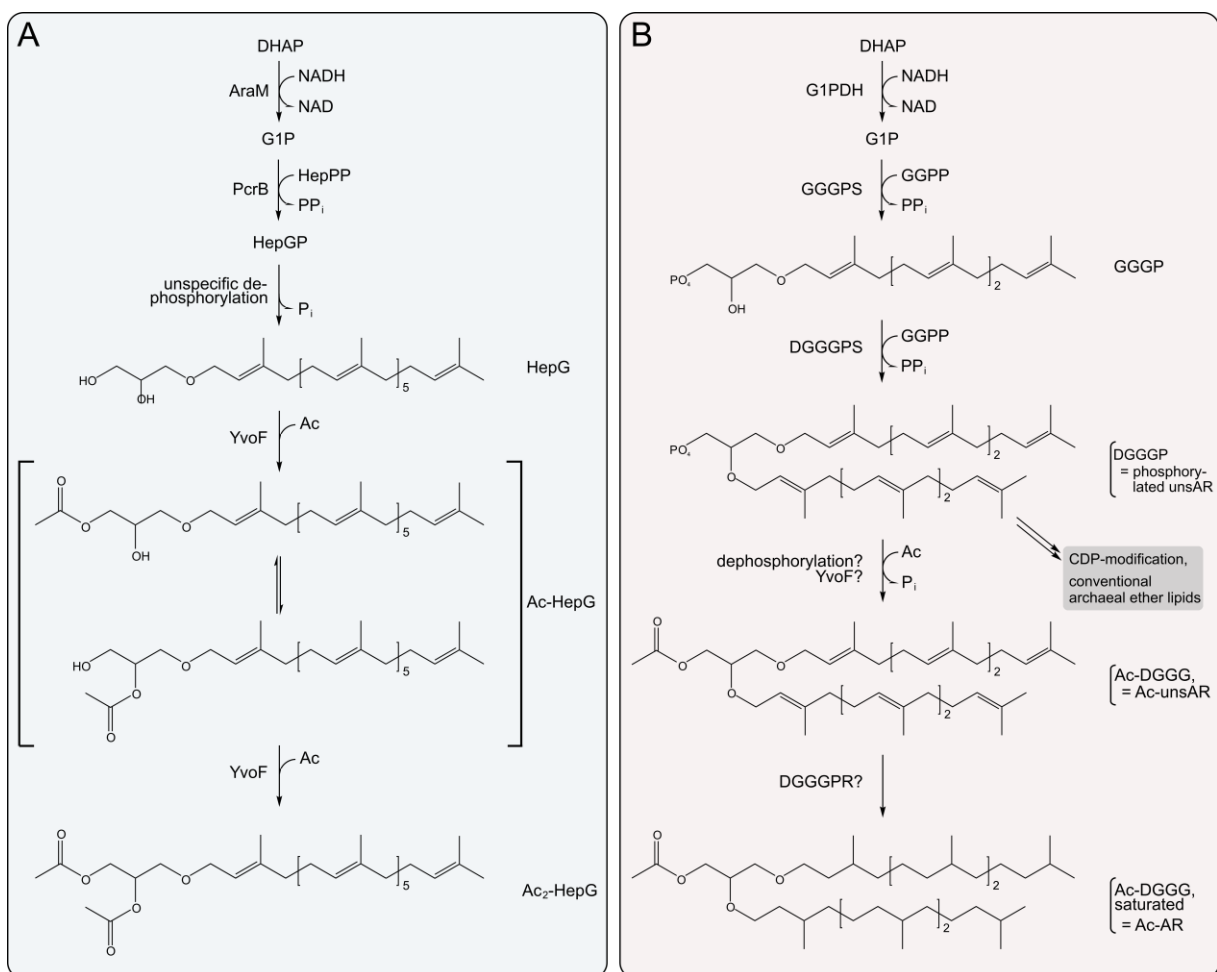
As a hallmark of Archaea, their cell membranes are comprised of ether lipids. However, Archaea-type ether lipids have recently been identified in Bacteria as well, with a somewhat different composition: In Bacillales, G1P is etherified with one C35 isoprenoid chain, which is much longer than the typical C20 chain in Archaea, and instead of a second isoprenoid chain, the product HepGP becomes dephosphorylated and afterwards diacetylated by the O-acetyltransferase YvoF. Interestingly, database searches have revealed YvoF homologues in Halobacteria (Archaea), too. Here, we demonstrate that YvoF from *Haloferax volcanii* can acetylate GGG *in vitro*. Additionally, we present the first-time identification of acetylated diether lipids in *Haloferax volcanii* and *Halobacterium salinarum* by MS. A variety of different acetylated lipids, namely acetylated archaeol and acetylated archaetidylglycerol, were found, indicating that halobacterial YvoF has a broad substrate range. We suppose that the acetyl group might serve to modify the polarity of the lipid headgroup, with still unknown biological effect.

### 4.2.2 Introduction

Membrane lipids of Archaea and Bacteria differ in some respects. In Archaea, their backbone consists of G1P, to which isoprenoids, typically 20 C-atoms in length (C20), are linked via an ether bond (Jain et al 2014). In Bacteria, G3P is esterified with fatty acids. Separation of Archaea and Bacteria during evolution has been postulated to be driven by the emergence of the enzymes synthesizing the membrane lipids, which has been summarized under the term “lipid divide” (Boucher 2007, Boucher et al 2004, Glansdorff et al 2008, Koga 2011, Koga & Morii 2007, Lombard et al 2012c, Payandeh & Pai 2007, Pereto et al 2004, Villanueva et al 2017). For a long time, lipids consisting of isoprenoids ether linked to G1P were considered exclusive for the archaeal domain, until the recent identification of Archaea-type ether lipids in gram-positive *B. subtilis* (Guldan et al 2011, Guldan et al 2008). Correspondingly, homologues to the enzyme that links the first isoprenoid chain to G1P, GGGPS, have been found in Bacillales and also in gram-negative Bacteroidetes (Peterhoff et al 2014). However, the ether lipids found in *B. subtilis* are quite different compared to those in Archaea. While in Archaea commonly a second C20 isoprenoid chain is attached to the glycerol by DGGGPS, ether lipids in *B. subtilis* possess one C35 chain (Guldan et al 2008, Peterhoff et al 2014). The glycerol moiety furthermore becomes dephosphorylated and acetylated, which is catalyzed by the acetyl-CoA-dependent O-acetyltransferase YvoF, resulting in mono- or diacetylated derivatives (Figure 4.3A; Guldan et al 2011, Linde et al 2016). Their biological function in Bacteria is still an enigma.

Database searches have indicated that YvoF homologues exist in many Halobacteria as well (Linde et al 2016), which was quite surprising as acetylated ether lipids have not been known to exist in Archaea so far. Halobacterial YvoF sequences share an over-all identity of about 50% among each other, and of about 35-45% with bacterial YvoF representatives; however, at the

acetyl-CoA binding site, the similarity is significantly higher (Figure S4.1). Since a confirmation of the activity of YvoF from Halobacteria as well as an identification of acetylated ether lipids within this phylogenetic clade has still been missing, we purified YvoF from *Haloferax volcanii* (hvYvoF) and verified acetyltransferase activity with bacterial C20 monoether lipids as substrate. We subsequently performed MS to tackle identification of acetylated ether lipids in extracts from *H. volcanii* and *H. salinarum*. Our results let us postulate a pathway for the synthesis of acetylated diether lipids in Halobacteria (Figure 4.3B). We also discovered further acetylated phospholipids in the two halobacterial strains and discuss their occurrence in the context of YvoF.



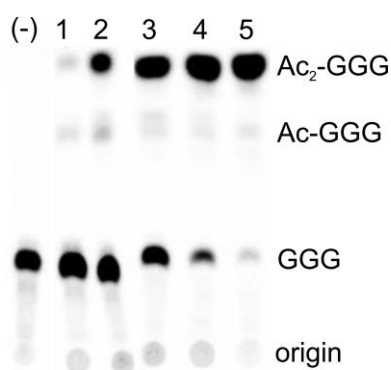
**Figure 4.3.** Biosynthesis of ether lipids in Bacteria and Archaea. (A) Biosynthetic pathway as discovered in *B. subtilis*. The figure is adapted from (Linde et al 2016). The acetyl moiety in Ac-HepG can putatively swap between the hydroxyl groups by acyl migration. (B) Postulated biosynthetic pathway of acetylated ether lipids in Halobacteria. AR, archaeol (corresponding to 2,3-diphytanyl-sn-glycerol); CDP, cytidine diphosphate; DGGGP, digeranylgeranylgeranyl phosphate (corresponding to phosphorylated unsaturated archaeol (unsAR)); DGGGPR, digeranylgeranylgeranyl phosphate reductase; DGGGPS, digeranylgeranylgeranyl phosphate synthase; G1PDH, glycerol 1-phosphate dehydrogenase; GGG, geranylgeranylgeranyl; GGGP, geranylgeranylgeranyl phosphate; GGGPS, geranylgeranylgeranyl phosphate synthase; GGPP, geranylgeranyl pyrophosphate; uns, unsaturated.

## 4.2.3 Results & Discussion

### 4.2.3.1 Activity of *H. volcanii* YvoF *in vitro*

We produced YvoF from *H. volcanii* (hvYvoF) in a *H. volcanii* expression system (Allers et al 2010). To this end, we amplified the gene from genomic DNA and cloned it into the pTA1228 expression vector. After transformation of chemically competent *H. volcanii* cells, expression was induced. Subsequent purification of hvYvoF was performed by IMAC and controlled by SDS-PAGE (Figure S4.2). The purified protein solution was colored pink, indicating that carotenoids such as lycopene and bacterioruberin, which are abundant in *H. volcanii* cells (Ronnekleiv 1995), were co-purified with hvYvoF. We assume that due to their isoprenoid nature, carotenoids bind to hvYvoF in the binding pocket of its isoprenoid substrate. In accordance with this assumption, the UV-Vis spectrum of purified hvYvoF in the range 450-550 nm is characteristic of bacterioruberin (Figure S4.3; Dummer et al 2011). In thermal denaturation experiments by means of nanoDSF, hvYvoF showed cooperative transition curves (Figure S4.4A). As expected for a protein from a halophilic organism, hvYvoF was significantly stabilized by increasing salt concentrations. The apparent midpoint temperature of the unfolding transition raised by about 27 K when NaCl was increased from 0 mM to 2000 mM (Figure S4.4B).

To verify acetyltransferase activity, we performed an activity assay with  $^{14}\text{C}$ -labelled substrate as described previously (Linde et al 2016). In brief, we synthesized radiolabeled GGGP ( $^{14}\text{C}$ -GGGP), which is the Archaea-type ether lipid as identified in *B. subtilis*, but with a C20 instead of a C35 isoprenoid.  $^{14}\text{C}$ -GGGP was incubated with CIP, generating  $^{14}\text{C}$ -GGG, and hvYvoF as well as acetyl-CoA as donor for the acetyl group. The reaction was stopped by chloroform addition at different points of time. Afterwards, the products were extracted, separated by TLC and visualized with a phosphoimager system (Figure 4.4).



**Figure 4.4.** Test for GGG-specific acetyltransferase activity of hvYvoF. hvYvoF at a concentration of 1  $\mu\text{M}$  was incubated with 0.25 mM acetyl-CoA and  $^{14}\text{C}$ -GGG in 50 mM Tris-HCl, pH 8.0, 10 mM  $\text{MgCl}_2$ , 0.2% Tween80 for different periods of time at 40  $^\circ\text{C}$ : (1) 0.1 min, (2) 2 min, (3) 10 min, (4) 30 min, (5) 120 min. A negative control (-) was performed without added hvYvoF. Reaction products were extracted, separated by TLC, and visualized by autoradiography. The origin of chromatography as well as spots of GGG, single acetylated Ac-GGG, and double acetylated  $\text{Ac}_2$ -GGG are marked. The figure is combined from sections of Figure S4.5.

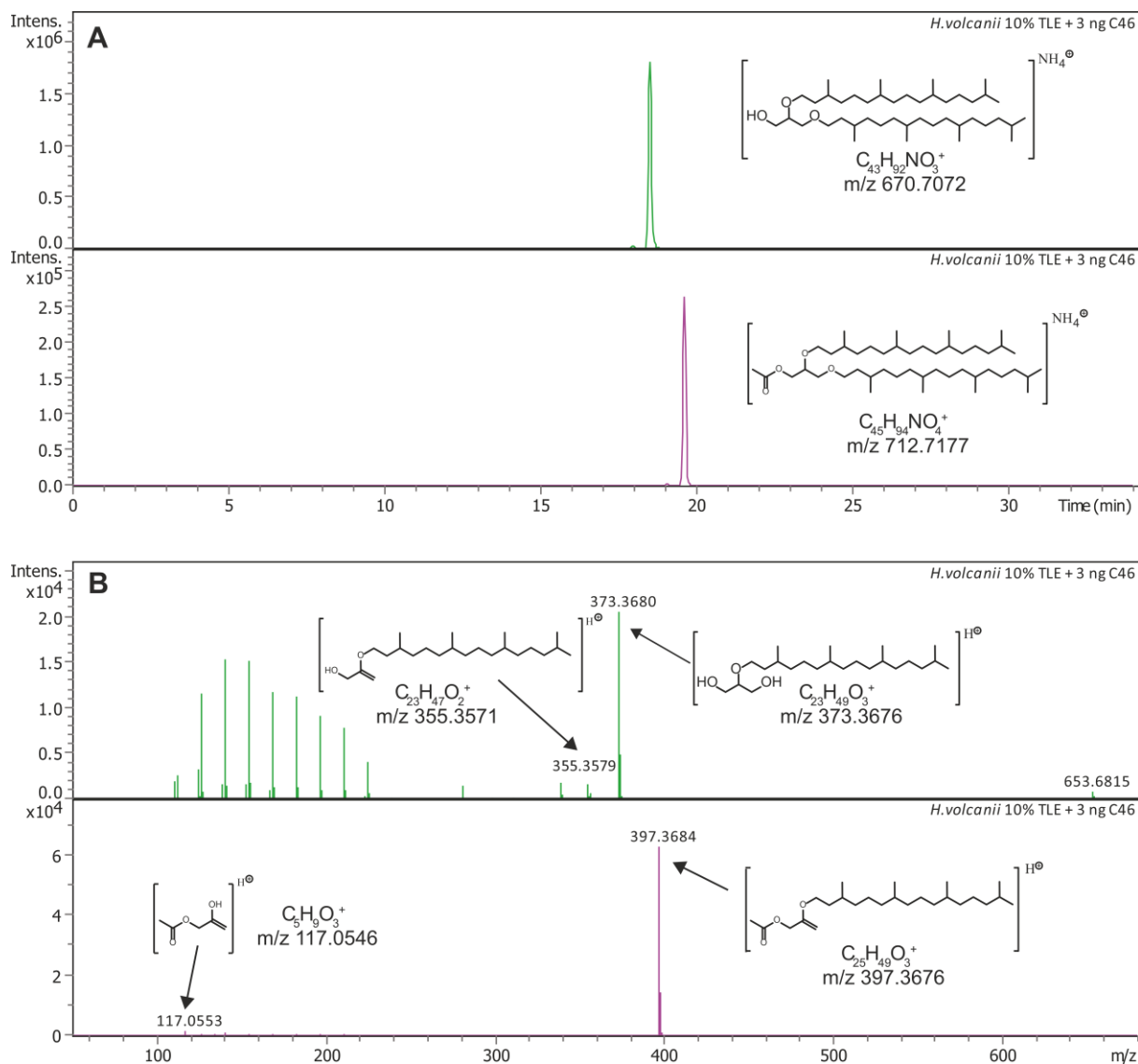
TLC analysis showed that GGG was gradually diacetylated (to Ac<sub>2</sub>-GGG) by hvYvoF with increasing incubation time. The monoacetylated intermediate Ac-GGG was scarcely detected. Diacetylation such as previously observed in *B. subtilis* (Figure 4.3A) is possible in this case, as two hydroxyl groups are available for acetylation when using GGG as substrate. In Archaea however, the second hydroxyl group is normally occupied by another isoprenoid chain, which is also seen later in MS experiments with *in vivo* samples. Because hvYvoF is a protein from a halophilic organism, we tested the dependence of activity upon increasing salt concentration in the assay (Figure S4.5). As expected, the activity was significantly increased in the presence of salt, which became most clearly visible in the faster decrease of the substrate GGG over the incubation time.

#### 4.2.3.2 Identification of acetylated diether lipids in Halobacteria

To identify what kinds of acetylated ether lipids occur *in vivo* in Halobacteria, we strived to analyze the lipids of two well characterized halobacterial species, *H. volcanii* H1424 and *H. salinarum* ATCC 700922. To this end, both strains were cultivated in the appropriate medium, cells were harvested, and TLEs were produced. The TLEs were then analyzed via RP-UHPLC coupled to ESI-MS. First, we screened the extract for acetylated mono- and dietherified glycerols. Figure 4.5 exemplarily shows the UHPLC elution profile and MS analysis for the *H. volcanii* extract, the data for *H. salinarum* are shown in Figure S4.6. The results are summarized in Table 4.7.

The fragment mass spectra indicate that besides regular archaeol (AR, corresponding to saturated DGGG or 2,3-diphytanyl-*sn*-glycerol; cf. Table 4.7), identified based on the characteristic fragment of  $m/z$  373.3676 from loss of one C<sub>20</sub> isoprenoid chain as phytene (cf. Liu & Cropp 2012), acetylated archaeol (Ac-AR) was present in the TLEs of both *H. volcanii* and *H. salinarum*. Ac-AR was tentatively identified based on a characteristic fragment at 397.37 Da expected for a structure where one C<sub>20</sub> isoprenoid chain including the OH-group has been lost, and a fragment at 117.05 Da representing an acetylated glycerol derivative (see structural formulas in Figure 4.5B and Figure S4.6B). In addition to their fully saturated molecules, ARs and Ac-ARs were also detected with one to eight unsaturations (Figure S4.7 and Figure S4.8) in both organisms. As expected for RP chromatography, the less polar Ac-ARs elute approximately 1 min later compared to their non-acetylated counterparts, and additional unsaturations reduce the retention time relative to the saturated counterparts (cf. Wörmer et al (2015)). Because acetylated diether lipids are not commercially available, their mass spectrometric response factors cannot be determined and the proportion of Ac-AR can only be estimated roughly, assuming that AR and Ac-AR are detected in RP-UHPLC-ESI-MS with the same mass spectrometric response. As derived from the peak heights (Figure 4.5, Figure S4.6),

the content of fully saturated Ac-AR would be approximately 10% compared to the amount of AR in the *H. volcanii* TLE, and approximately 5% in the *H. salinarum* TLE.



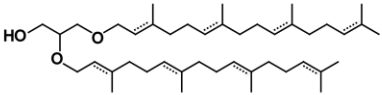
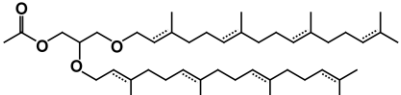
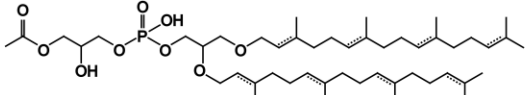
**Figure 4.5.** RP-UHPLC-ESI-MS analysis of a *H. volcanii* TLE. (A) Extracted ion chromatograms of m/z 670.7072 $\pm$ 0.01 representing AR (upper panel) and of m/z 712.7177 $\pm$ 0.01 representing Ac-AR (bottom panel). (B) Fragment mass spectra of AR (upper panel) and Ac-AR (lower panel). Tentative structures and elemental formulas of diagnostic peaks are shown. C46 refers to the added standard (cf. section 4.2.5.8).

Interestingly, screening the UHPLC-MS data for other compounds containing acetylated glycerol using the characteristic fragment at 117.05 Da (cf. Figure 4.5B) revealed additional acetylated lipids, namely archaetidylglycerols (Ac-AG) (Figure S4.9, Figure S4.10, Table 4.7). Considering the fact that archaetidylglycerol (AG) and AG methylphosphate (AG-PCH<sub>3</sub>) are the main polar lipids of *H. volcanii* (Sprott et al 2003), this finding seems reasonable. In *H. volcanii* we could



find almost all possible unsaturated structures of AG with the exception of compounds with 7 and 8 double bonds (Figure S4.10A), while in *H. salinarum* we could only identify saturated Ac-AG (Figure S4.10B). Unsaturated compounds were absent in this species. For AG, we found all possible unsaturated structures in both strains (Figure S4.11). As a rough quantitative estimation under the preconditions as stated above, the content of fully saturated Ac-AG would be approximately 40% compared to the amount of AG in the *H. volcanii* TLE, as derived from the peak heights, and approximately 30% in the *H. salinarum* TLE.

**Table 4.7.** Identified ether lipids in Halobacteria.

compound	structural formula <sup>1</sup>	theo. mass <sup>2</sup>
archaeol (AR)		670.7072
acetylated archaeol (Ac-AR)		712.7177
acetylated archaetidylglycerol (Ac-AG)		866.7218

<sup>a</sup> Compounds were found with up to 8 unsaturations, which is indicated by dashed double bonds.

<sup>b</sup> The theoretical (theo.) mass is calculated for the NH<sub>4</sub><sup>+</sup> adducts and is given for the saturated molecules only.

Although we searched the RP-UHPLC-ESI-MS results thoroughly, we did not find any signals of mono- or diacetylated monoether lipid derivatives similar to those found in *B. subtilis* (Figure 4.3A), even though hvYvoF is able to acetylate monoether lipids as suggested by our experimental data (Figure 4.4). In contrast to *B. subtilis*, Halobacteria possess a DGGGPS, which could rapidly process monoether to diether lipids and therefore be the reason for the absence of acetylated monoether lipids in the investigated strains under the given conditions.

Since headgroup modification and saturation of the isoprenoid chains normally happens after attachment of the CDP to the lipid, there must be an additional synthesis branch for Ac-AR, which we propose as presented in Figure 4.3B. Instead of those modifications that lead to the conventional archaeal ether lipids, we propose that DGGGP (= phosphorylated unsaturated AR (unsAR)) is dephosphorylated and subsequently acetylated resulting in Ac-unsAR. Afterwards, partial or complete saturation of the polyprenyl moiety may take place. Alternatively, halobacterial YvoF could be a promiscuous enzyme that accepts both saturated and unsaturated

AR lipids as substrate. The existence of acetylated AG derivatives supports this and indicates that YvoF might also act on readily processed lipids with a phosphoglycerol head group. YvoF might primarily acetylate the most abundant ether lipids with free hydroxyl groups at a glycerol in the cell without certain substrate specificity. The hypothesis of YvoF being a promiscuous enzyme is in agreement with previous findings that many *O*-acetyltransferases show a broad substrate range. The closest relatives to YvoF, maltose acetyltransferase (MAT) and galactoside *O*-acetyltransferase, are known to acetylate a variety of sugars with different efficiencies (Lo Leggio et al 2003). Similarly, *B. subtilis* YvoF can acetylate maltose, and MAT can acetylate GGG *in vitro* (Linde et al 2016).

#### 4.2.4 Conclusion

Our results demonstrate that halobacterial YvoF exhibits GGG acetyltransferase activity *in vitro*, just like its homologue from *B. subtilis* (Linde et al 2016). This matches our tentative mass spectrometric identification of Ac-AR, Ac-AG and various unsaturated derivatives thereof in *H. volcanii* and *H. salinarum*. We suggest that YvoF might be responsible for acetylating these lipids in both species (the sequence identity of their YvoF protein sequences is 67%), and that acetylated ether lipids might also occur in many other Halobacteria that possess a YvoF homologue.

As outlined in our proposed pathway (Figure 4.3B), our findings suggest that YvoF catalyzes acetylation of diether lipids in Halobacteria after the second isoprenoid chain has been added by DGGGPS, because despite of a thorough search, we did not find any acetylated monoether lipids. Most likely, acetylation happens in parallel to the reduction of the isoprenoid moieties by DGGGPR and the attachment of glycerol as polar head group, as we found a series of different partially unsaturated Ac-AR and Ac-AG derivatives. We suppose that the acetyl group might serve alone as a small headgroup that reduces the polarity of the glycerol core of AR or serves to modify the polarity of the glycerol headgroup in case of AG. It remains elusive how acetylation is in balance with the attachment of other head groups which cannot be acetylated.

According to a previous study (Sprott et al 2003), the main polar membrane lipids of *H. volcanii* are AG-PCH<sub>3</sub> (44% of total lipid) and AG (35%), followed by sulfated glycolipids (14%), archaeal cardiolipin (5%) and archaetidic acid (2%). As a rough estimation for *H. volcanii*, the content of fully saturated Ac-AR was about 10% of that of the detected amount of AR, and the amount of Ac-AG was 40% of that of AG. Based on these assumptions, a total content of acetylated ether lipids in a low percentage range seems reasonable and might be similar in *H. salinarum*.

Since YvoF does not occur in other Archaea than Halobacteria, it probably migrated from the Bacillales into the archaeal domain of life or vice versa by HGT, and due to the significant amounts of acetylated lipids we detected, they probably fulfill a distinct function. The biological

role of the acetylated monoether lipids that have been found in Bacillales is still unclear, and the discovery of acetylated diether lipids in the archaeal domain of life makes it even more interesting to study them in more detail.

## 4.2.5 Materials & Methods

### 4.2.5.1 Halobacterial strains and culture media

For lipid analysis, *Haloferax volcanii* H1424 (Stroud et al 2012) and *Halobacterium salinarum* ATCC 700922 (Ng et al 2000) were used. hvYvoF was produced in *H. volcanii* H1424. Both strains were obtained from Sébastien Ferreira-Cerca, University of Regensburg.

*H. volcanii* was grown in salt rich hv-YPC medium at 42 °C. For non plasmid-harboring strains, the medium was supplemented with 40 µg mL<sup>-1</sup> thymidine and for plasmid-harboring strains with 75 µg mL<sup>-1</sup> kanamycin (Allers et al 2004, Guy et al 2006). *H. salinarum* was grown in medium for extreme halophiles at 42 °C (Cline et al 1989).

### 4.2.5.2 Cloning

hvyvoF was amplified by PCR from *H. volcanii* H1424 genomic DNA. The primers used for amplification and cloning are given in Table S4.1. The produced fragments were cloned via NdeI/XhoI into the pTA1228 expression vector (Allers et al 2010) providing a C-terminal hexahistidine (His)<sub>6</sub> tag.

### 4.2.5.3 Transformation procedures

Transformation of *E. coli* for cloning was performed using a standard protocol for chemically competent cells (Sambrook 1989). Production of competent cells of *H. volcanii* and polyethylene mediated transformation was performed as described by Cline et al (1989). The used solutions and media were produced accordingly.

### 4.2.5.4 Production and purification of recombinant hvYvoF

Heterologous gene expression was performed in *H. volcanii* H1424 (Allers et al 2010). To this end, 81.7 mg L-tryptophan were dissolved in 360 mL hv-YPC medium (1.11 mM final concentration) by shaking at 180 rpm and 42 °C in 3 L flasks. The medium was inoculated with an overnight culture (40 mL from the pre-culture diluted to 1 OD<sub>600</sub>) of transformed *H. volcanii* cells to 0.1 OD<sub>600</sub> and shaking was continued until 0.5 OD<sub>600</sub> was reached. Gene expression was induced a second time by adding 2 mM L-tryptophan (final concentration) in 18% salt water and incubation was continued for 16 h. Cells were harvested by centrifugation, resuspended in 40 mL

50 mM Tris, pH 8.0, 600 mM NaCl, 10 mM imidazole and disrupted by sonication until the suspension was non-turbid anymore. Cell debris was removed by centrifugation. The His-tagged protein was purified from the clarified cell extract by IMAC. An ÄKTApurifier system with a HisTrap FF crude column (5 mL, Cytiva) was used, and a linear gradient of imidazole (10–500 mM) in 50 mM Tris, pH 8.0, 600 mM NaCl was applied to elute the protein. To remove interfering salts and imidazole, the protein was subjected to dialysis against 50 mM Tris, pH 8.0, 600 mM NaCl. Protein concentrations were determined by absorbance spectroscopy using a Jasco V650 spectrophotometer. The molar extinction coefficients  $\epsilon_{280}$  and the MW were calculated from the amino acid sequence by means of ProtParam (Gasteiger et al 2005). Purified protein was dropped into liquid nitrogen and stored at -80 °C. UV-Vis spectra for analysis of bound ligands were recorded using 10  $\mu$ M protein (final concentration) in 50 mM Tris, pH 8.0, 600 mM NaCl (200-600 nm; response time 0.96 sec; scan rate 40 nm min<sup>-1</sup>; bandwidth 2 nm).

#### 4.2.5.5 nanoDSF

hvYvoF was heated at a final subunit concentration of 20  $\mu$ M in 50 mM Tris, pH 8.0 buffer from 20 to 95 °C (1 K min<sup>-1</sup>), in a Prometheus NT.48 instrument (NanoTemper Technologies GmbH; access provided by 2bind; 10% excitation power at 280 nm). The change in the ratio of the fluorescence signal at 350 nm to 330 nm with raising temperature was monitored and the fluorescence transitions were evaluated using the program supplied by the manufacturer. The apparent melting point temperature ( $T_M$ ) of the irreversible unfolding transition was determined as an operational measure of thermal protein stability. Experiments were done in duplicates, which overlapped perfectly.

#### 4.2.5.6 <sup>14</sup>C-GGG-dependent activity assay

<sup>14</sup>C-GGGP was synthesized as described previously (Linde et al 2016), without the dephosphorylation step. To test activity of the purified hvYvoF enzyme, 1 U CIP, 1  $\mu$ M hvYvoF, 0.25 mM acetyl-CoA (Sigma) and 2.5  $\mu$ M synthesized <sup>14</sup>C-GGG (37.5 nCi) were incubated in 50 mM Tris, pH 8.0, 10 mM MgCl<sub>2</sub>, 0.2% Tween80 at 40 °C. For time-dependent activity, the reaction was stopped at certain time points by addition of chloroform. The products were extracted according to the method of Bligh and Dyer (Bligh & Dyer 1959) as modified by Kates (Kates 1986). The extract was analyzed on Silica 60 plates developed in ethyl acetate/hexane 1:1 (v/v), and visualized with a phosphorimager system (PerkinElmer Life Sciences).

#### 4.2.5.7 Production of TLE

*H. volcanii* and *H. salinarum* were grown at 42 °C. Cells were harvested by centrifugation and stored at -80 °C. Lipid extraction was performed in a one-step procedure based on Wörmer et al (2015). 100 mg of each strain were washed into a 50 mL teflon vessel using 20 mL of Bligh & Dyer extraction mix A containing methanol (MeOH)/dichloromethane (DCM)/PO<sub>4</sub> buffer in a ratio of 2:1:0.8 (v/v/v). The vessel was sealed, mixed thoroughly by shaking and subjected to sonication in an ultrasonic bath for 10 min. Afterwards, the mixture was transferred into a 100 mL separating funnel and 3 mL DCM and 6.3 mL H<sub>2</sub>O were added. After intensive shaking, the separating funnel was vented and left for phase separation. Upon completion of separation, the organic bottom layer was transferred into a 40 mL evaporation vessel. The remaining aqueous phase in the separating funnel was washed with 5 mL DCM and left for phase separation. The organic phase was carefully drawn off and added into the evaporation vessel. The washing step was repeated for a total of three times. Finally, the aqueous phase was discarded and the separating funnel was washed with DCM and MeOH. The organic phase was now transferred to the funnel and washed three times with 5 mL purified H<sub>2</sub>O in the same way as described for the previous washing procedure. After the third repetition, the organic phase, now representing the TLE, was put into an evaporator (TurboVap, Biotage) in a water bath set to 40 °C and the solvent was evaporated to dryness under a nitrogen stream. The sample was re-dissolved in solvent containing DCM/MeOH in a ratio of 5:1, transferred into 4 mL vials and evaporated under N<sub>2</sub> flow at 40 °C (Merck Supelco). The TLE was stored at -20 °C until further use.

#### 4.2.5.8 UHPLC-RP-ESI-MS

The TLE was dissolved in 2 mL of DCM/MeOH (5:1, v/v) and 400 µL (20% of the TLE) were transferred into a new vial. C46-GTGT standard (Huguet et al 2006) was added to a final amount of 6 ng. The mix was evaporated under a gentle N<sub>2</sub> flow and re-dissolved in 20 µL MeOH/DCM (9:1, v/v). From that, 10 µL were injected (equals 10% of the TLE and 3 ng standard) for analysis in RP-UHPLC-ESI-MS.

For analysis of the intact polar lipids, the prepared TLE dilution was separated by UHPLC on a Dionex ultimate 3000RS system (Thermo Fisher Scientific) coupled to a Bruker quadrupole time-of-flight mass spectrometer (QTOF-MS; Bruker Daltonics, Bremen, Germany). At the end of each analysis a tune mix solution was injected via a 20 µL loop to ensure mass calibration for each analysis and additionally a lock mass calibration was applied to correct each mass spectrum. The resulting mass accuracy was better than 3 ppm. In order to separate the compounds mainly by alkyl chain hydrophobicity, the system was operated using an Acquity BEH C18 column (1.7 µm, 2.1 x 150 mm) (Waters, Eschborn, Germany). The flow rate was set to 0.4 mL min<sup>-1</sup>, column temperature at 65 °C and a gradient program using RP-ESI mobile phase A and B buffer was programmed according to literature (Wörmer et al 2015, Wörmer et al 2013). Output data

was analyzed using software supplied by the manufacturer (DataAnalysis 5.0, Bruker Daltonics, Bremen, Germany).

#### **4.2.6 Acknowledgements**

We thank 2bind GmbH for access to the Prometheus NT.48 instrument (NanoTemper Technologies). We also thank Christiane Endres, Sonja Fuchs, Sabine Laberer and Jeannette Ueckert for technical assistance. We are grateful to Reinhard Sterner for critical reading of the manuscript and to Sébastien Ferreira-Cerca for providing the halobacterial strains and the expression vector.

## 4.2.7 Supporting Information

**Table S4.1.** Primers used for amplification and cloning of *hvyvoF* into pTA1228 vector.<sup>a</sup>

primer	Sequence (5' -> 3')
<i>hvyvoF_NdeI_fwd</i>	CTCC <b>ATAT</b> GACCGACGACGATGCCGCACCG
<i>hvyvoF_XhoI_rev</i>	TGCG <b>CTCGAG</b> CGCTCCGTCAGTCTCGTCGC

<sup>a</sup> Restriction sites are printed in bold letters.

```

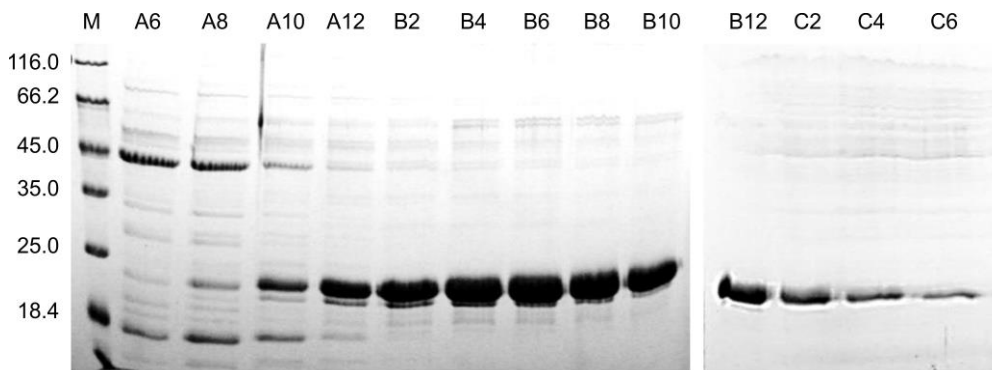
# Aligned_sequences: 2
# 1: bsYvoF
# 2: hvYvoF
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 188
# Identity:      67/188 (35.6%)
# Similarity:   102/188 (54.3%)
# Gaps:         24/188 (12.8%)
# Score: 319.0
#
#
#=====
bsYvoF            1  -----MRKTD---HPVSG-ANSLWHVYQTVPFVKVVKNFIVI      34
                    .:|:| |   | |.:|  .|||.:.|.|.:.|..|.:.|:|:.|:|.:.|:|.:.|
hvYvoF            1  MTDDDAAPADA-EARRHRDIQHHPTPGPRNSLQYWTDAKPVWRVMLNYVFV    50

bsYvoF            35  QIARYTPFIGMKNWLYRTFLRMKVGGKQTSFALMMPDIMPFPEKISVGTNT    84
                    .:|:|..|:|:|:|..|:|. .:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|.:.|:|.:.|:|.:.|
hvYvoF            51  LVARIAPSLKLRNWLRR-IGVTVGRSVWGLEATPDVFWPDLVTIEDDA      99

bsYvoF            85  IIGYNTTILAHEYLIHEYRIGKVLIGDEVMI GANTTILPGVKIGDGAVVS    134
                    |||:|:|.|||.|||.|||.|||.||:|:|:|:|.|||.|||:|:|:|:|:|:|:|:|.|:|:|:|:|
hvYvoF            100 IIGYDSVILCHEFLQDEYRTGEVVVGERAMIGAKATILPGVRIEGAQVA     149

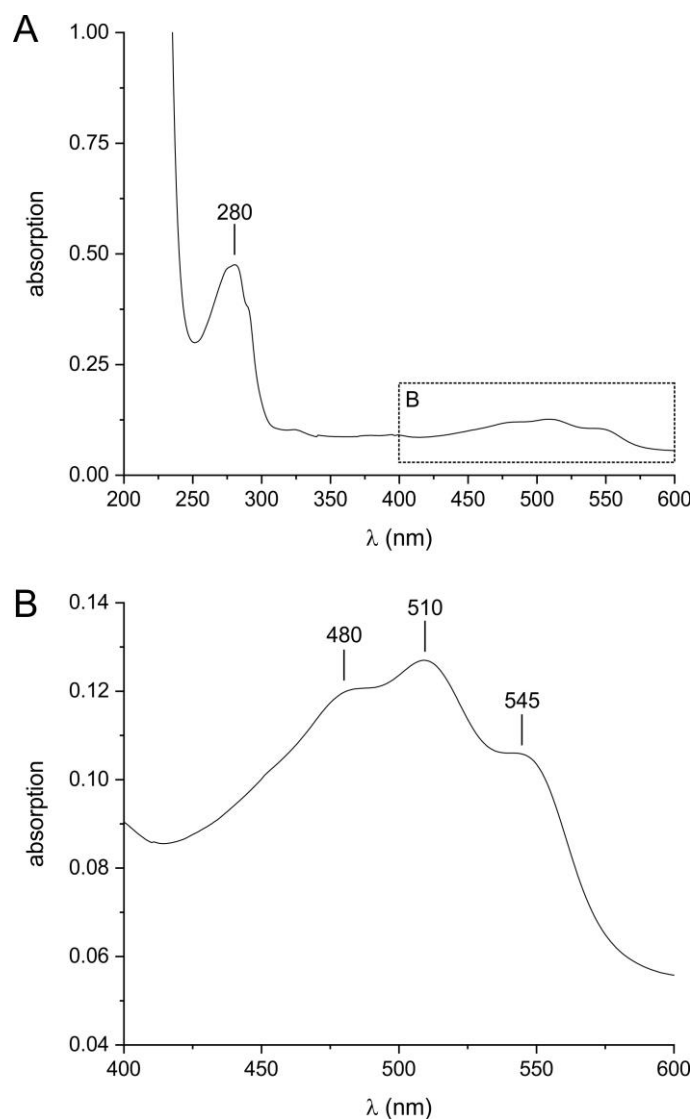
bsYvoF            135  AGTLVHKDVPDGA FVGGNPMRIIYTKEEMQERLKKS AE      172
                    |:|:|:|:|:|:|:|:|:|.|:|.|:|.||:|:|.:.|:|.:.|:|.:.|
hvYvoF            150  ANSLVTRDVPAGETVAGV PARPMGDETDGA-----      180
    
```

**Figure S4.1.** Pairwise sequence alignment of YvoF from *B. subtilis* (bsYvoF) and from *H. volcanii* (hvYvoF). The alignment was created using the Needleman-Wunsch algorithm. The numbers of identical and similar amino acids are given in the header. The acetyl-CoA binding site in bsYvoF (residues 114-140 according to GenBank entry KIX83267.1) is marked with a black bar.

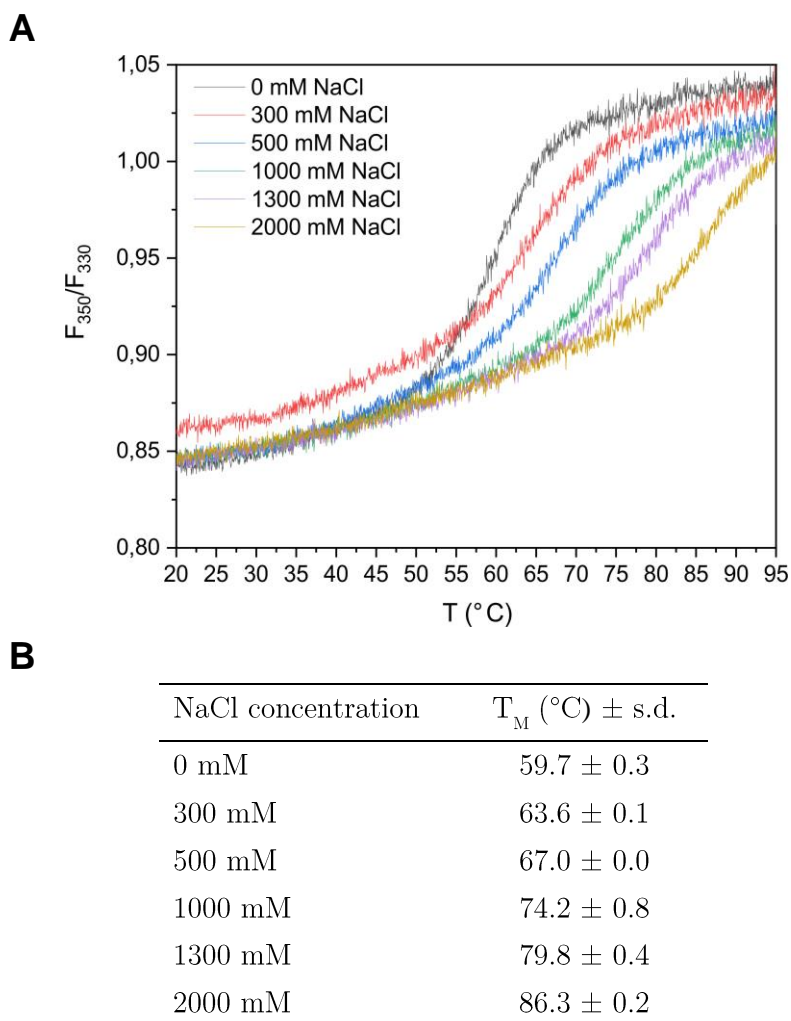


**Figure S4.2.** Purification of hvYvoF by IMAC. Fractions with a volume of 5 ml were collected during chromatography, and 30 µL of every second fraction were analyzed on a 13.5% SDS-PAGE gel. Fractions A12-C6 were pooled and used for further experiments. Size marker (M), MW in kDa. The calculated molecular mass of hvYvoF is 20.5 kDa.

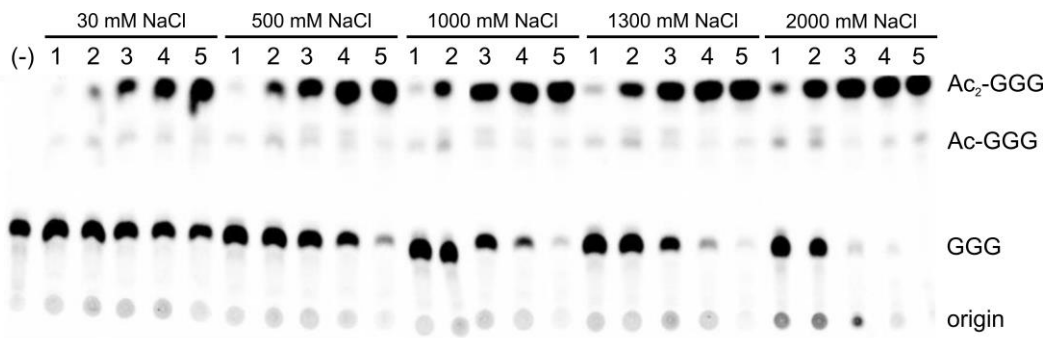




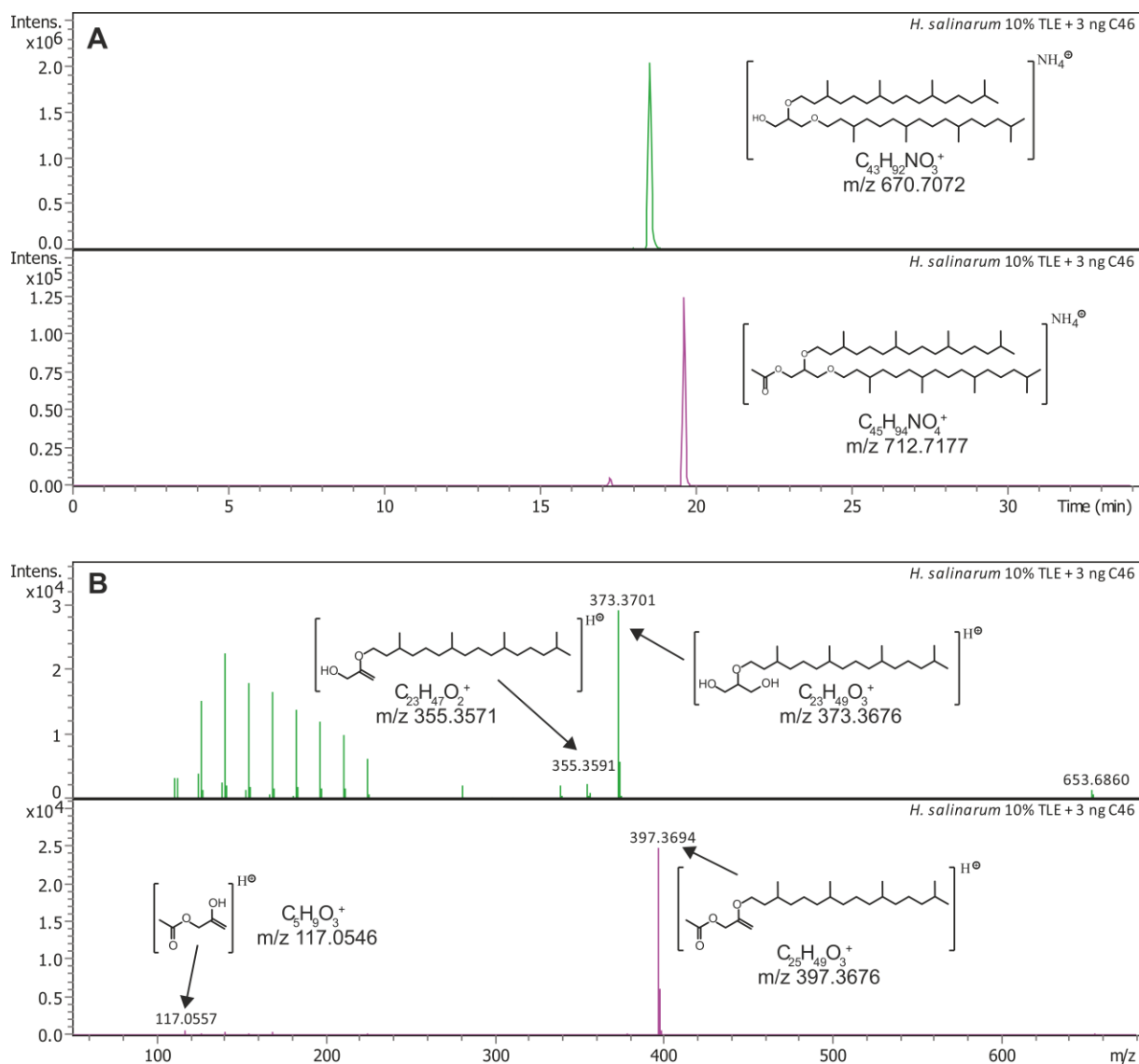
**Figure S4.3.** Absorption spectrum of hvYvoF. The spectrum was recorded using 10  $\mu$ M hvYvoF in 50 mM Tris-HCl, pH 8.0, 600 mM NaCl from (A) 200 nm to 600 nm and from (B) 400 nm to 600 nm at 25  $^{\circ}$ C ( $d = 0.4$  cm). The box in (A) marks the section shown in (B). The wavelengths of the absorption maxima are indicated.



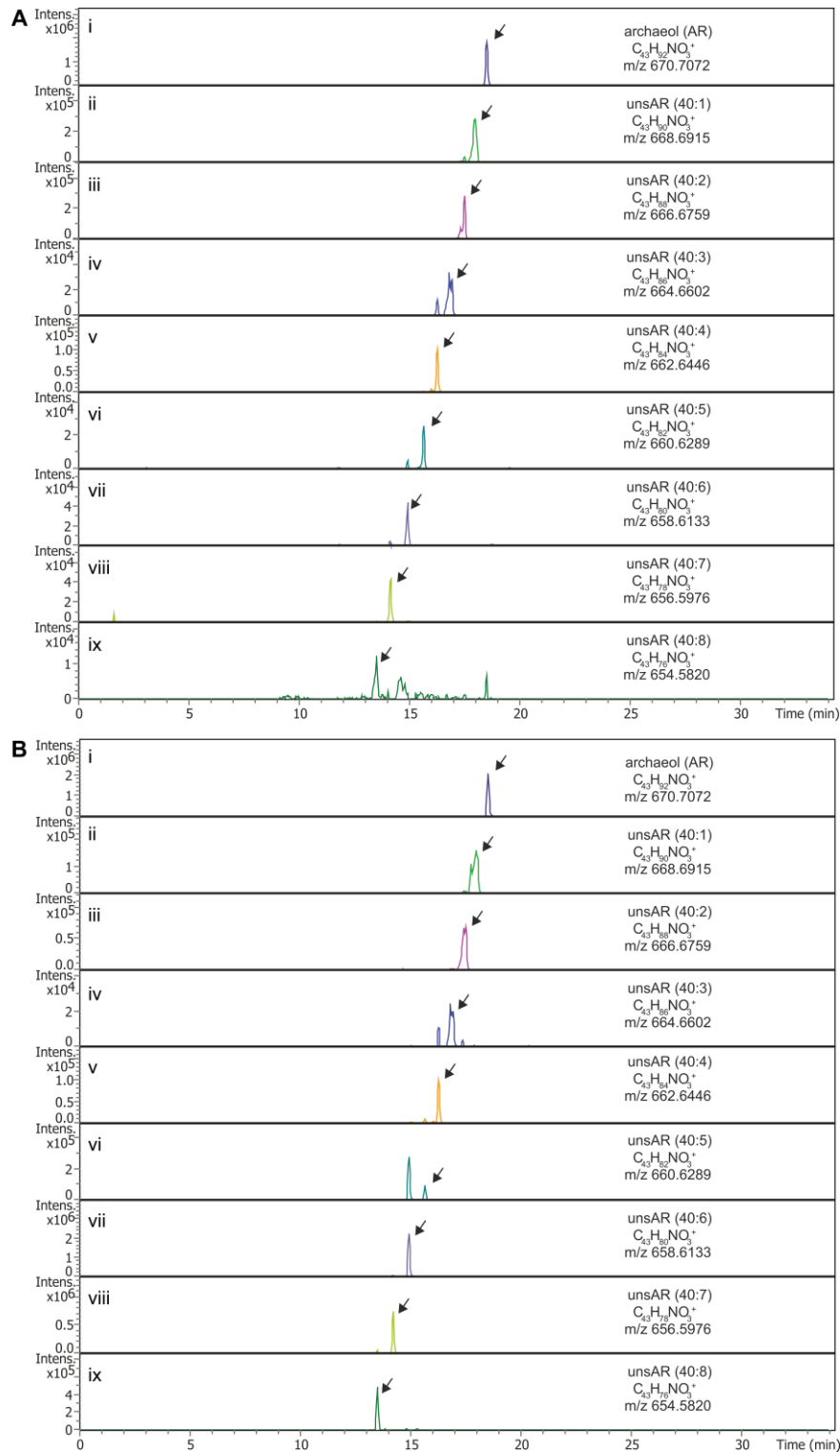
**Figure S4.4.** Thermal stability of hvYvoF, measured by nanoDSF. (A) The change in the ratio of fluorescence emission of 20  $\mu\text{M}$  protein at 350 and 330 nm was monitored from 20 to 95  $^{\circ}\text{C}$  at a scan rate of 1  $\text{K min}^{-1}$ . The protein was dissolved in 50 mM Tris-HCl pH 8.0 and supplemented with varying NaCl concentrations (as indicated in the legend) (B) Apparent melting temperatures ( $T_m$ ) were derived as the inflection points of the curves from duplicates, the standard deviation is given.



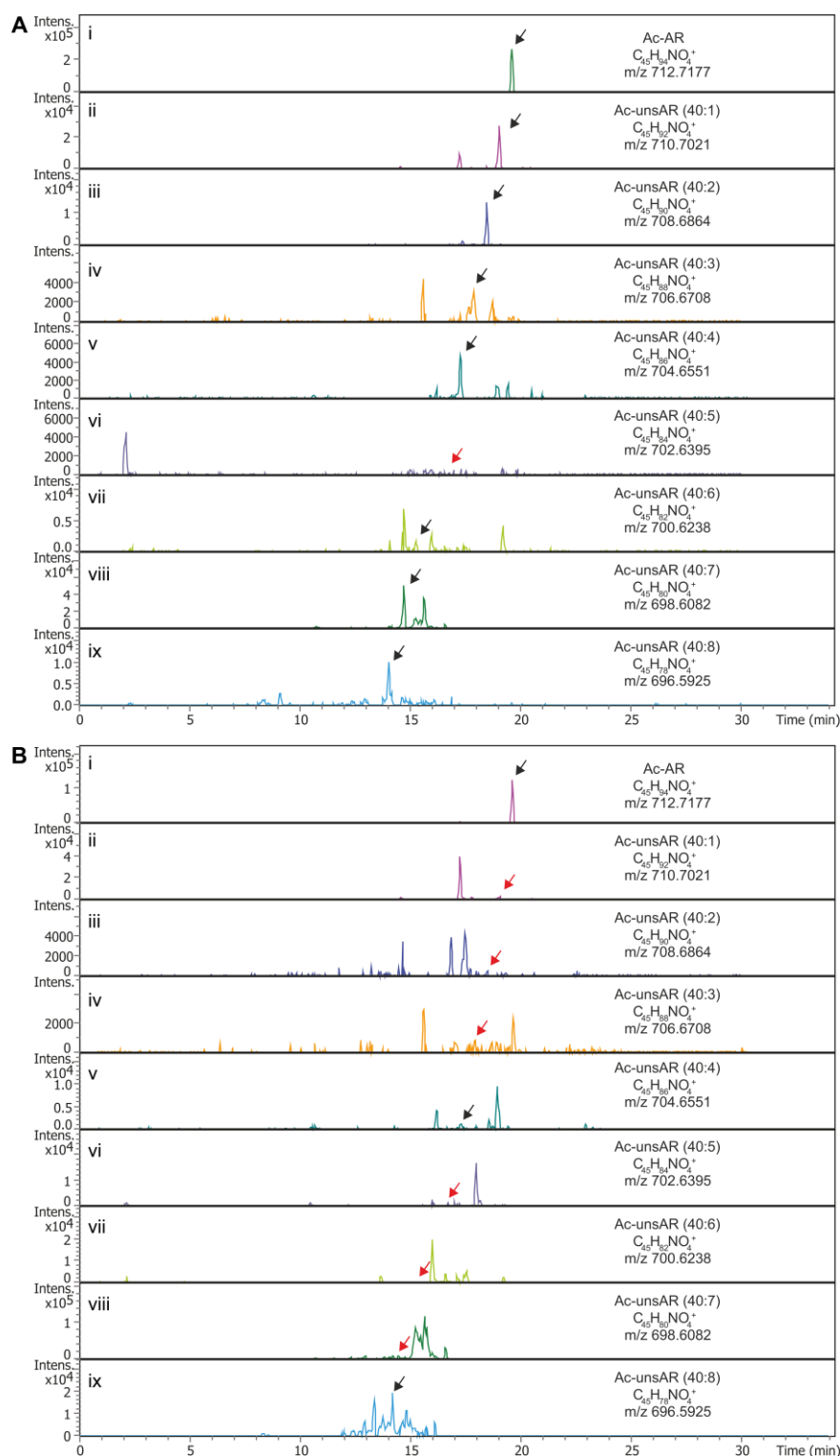
**Figure S4.5.** Test of salt dependent activity of hvYvoF. 1  $\mu\text{M}$  hvYvoF was incubated with 0.25 mM acetyl-CoA and  $^{14}\text{C}$ -GGG in 50 mM Tris-HCl, pH 8.0, 10 mM  $\text{MgCl}_2$ , 0.2% Tween80 and 0-2000 mM NaCl at 40  $^\circ\text{C}$  for different periods of time at 40  $^\circ\text{C}$ : (1) 0.1 min, (2) 2 min, (3) 10 min, (4) 30 min, (5) 120 min. The concentration of NaCl in the reaction buffer was varied as indicated. A negative control (-) was performed without added hvYvoF. Reaction products were extracted, separated by TLC, and visualized by autoradiography. The origin of chromatography as well as spots of GGG, single acetylated Ac-GGG, and double acetylated Ac<sub>2</sub>-GGG are marked. The figure is assembled from two separate TLC runs.



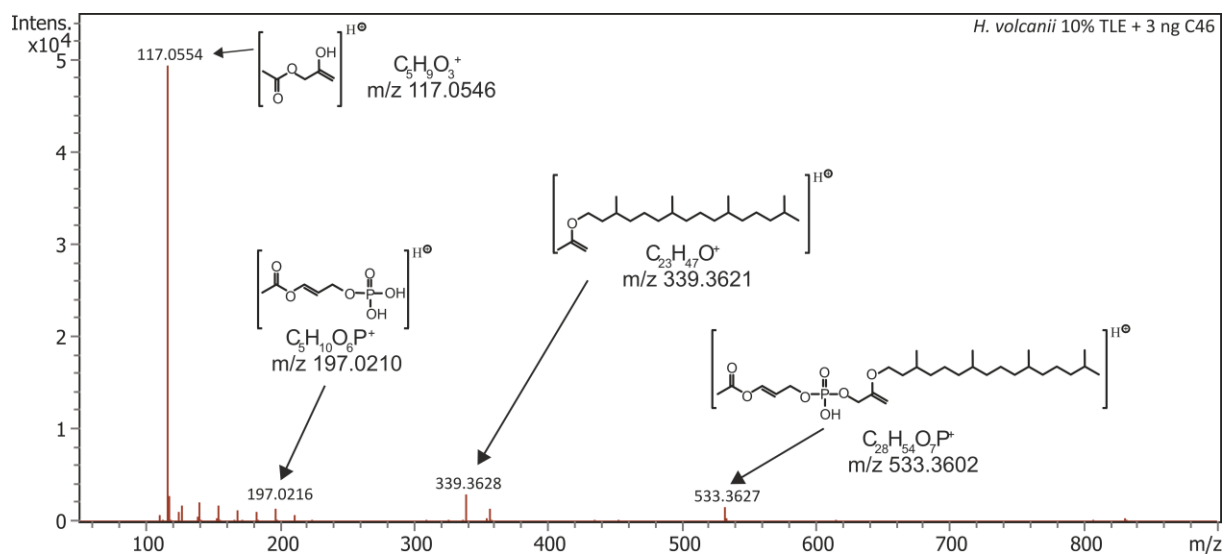
**Figure S4.6.** RP-UHPLC-ESI-MS analysis of a *H. salinarum* TLE. (A) Extracted ion chromatograms of m/z 670.7072 $\pm$ 0.01 representing AR (upper panel) and of m/z 712.7177 $\pm$ 0.01 representing Ac-AR (bottom panel). (B) Fragment mass spectra of AR (upper panel) and Ac-AR (lower panel). Tentative structures and elemental formulas of diagnostic peaks are shown.



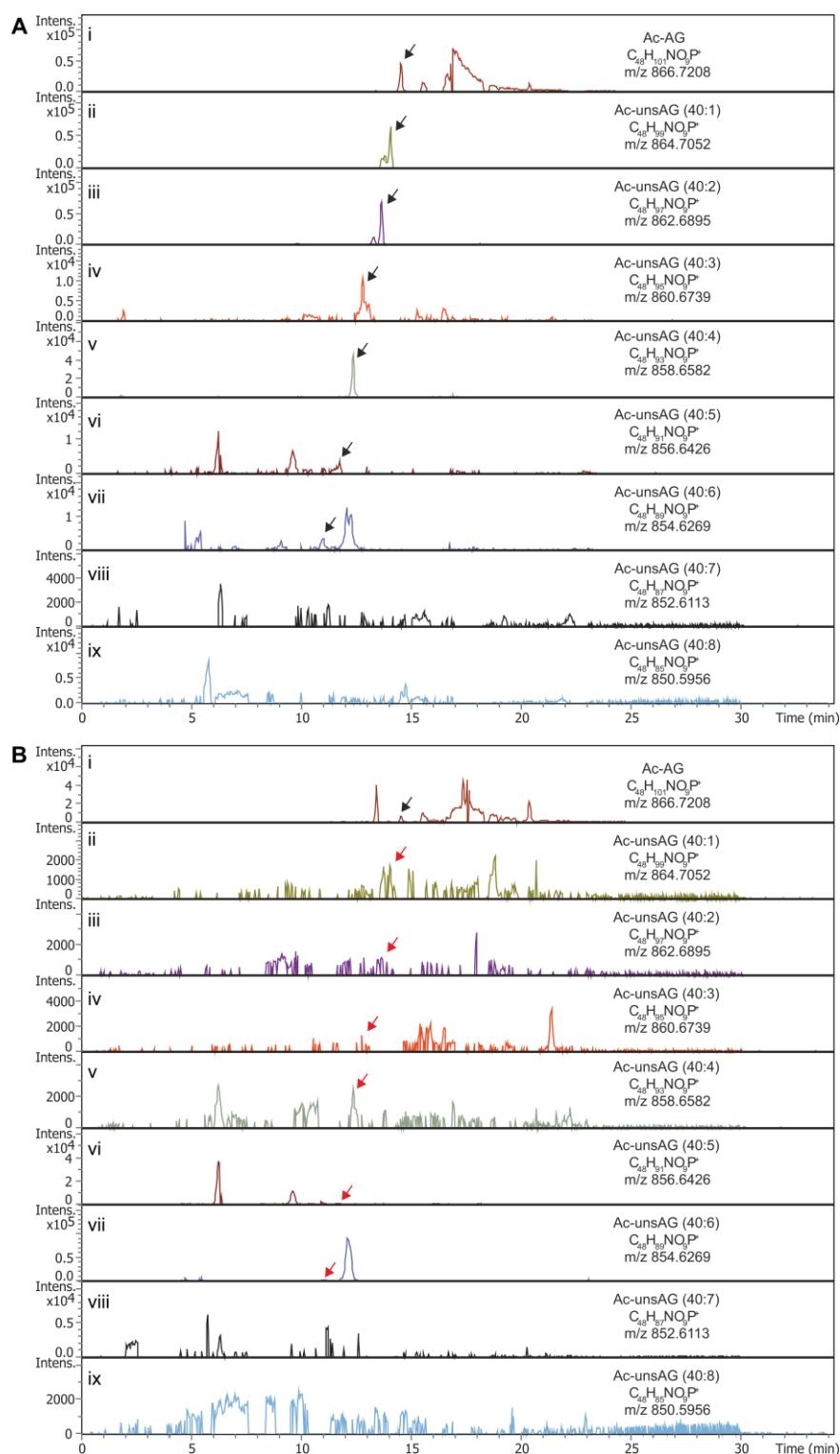
**Figure S4.7.** Detection of saturated and unsaturated ARs. Extracted ion chromatograms (0.01 Da width) of AR analogs with 1–8 unsaturations (panel i–ix; numbers in parentheses denote number of carbons in both isoprenoid chains and number of unsaturations in both chains) in (A) *H. volcanii* and (B) *H. salinarum*. Black arrows point to the respective detected chromatographic peaks, minor peaks at earlier retention times represent isotopologues with two  $^{13}C$  atoms in unsAR with one more unsaturation or other compounds within the extracted mass window.



**Figure S4.8.** Detection of saturated and unsaturated Ac-ARs. Extracted ion chromatograms (0.1 Da width) of Ac-AR analogs with 1-8 unsaturations (panel i-ix; numbers in parentheses denote number of carbons in both isoprenoid chains and number of unsaturations in both chains) in (A) *H. volcanii* and (B) *H. salinarum*. Black arrows point to the respective detected chromatographic peaks, whereas red arrows indicate the expected retention time for compounds that could not be detected. Minor peaks at earlier retention times represent isotopologues with two  $^{13}C$  atoms in acetylated unsAR with one more unsaturation or other compounds within the extracted mass window.

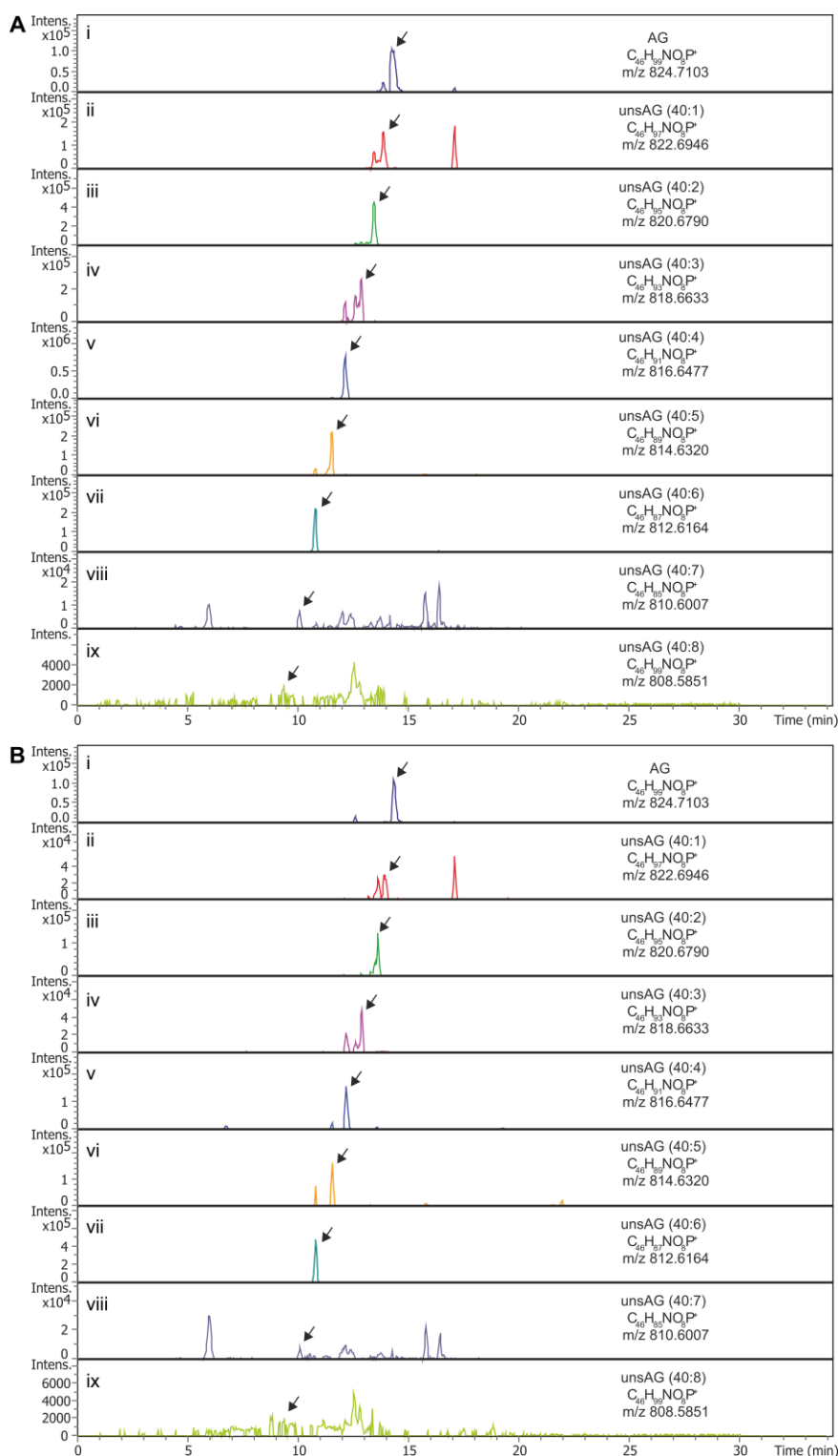


**Figure S4.9.** MS analysis of a *H. volcanii* TLE. Fragment mass spectrum of Ac-AG. Tentative structures and elemental formulas of diagnostic fragment peaks are shown.



**Figure S4.10.** Detection of saturated and unsaturated Ac-AGs. Extracted ion chromatograms (0.1 Da width) of Ac-AG analogs with 1-8 unsaturations (panel i-ix; numbers in parentheses denote number of carbons in both isoprenoid chains and number of unsaturations in both chains) in *H. volcanii* (A) and *H. salinarum* (B). Black arrows point to the respective detected chromatographic peaks, whereas red arrows show expected retention time for compounds that could not be detected. Minor peaks at earlier retention times represent isotopologues with two  $^{13}C$  atoms in acetylated unsAR with one more unsaturation or other compounds within the extracted mass window.





**Figure S4.11.** Detection of saturated and unsaturated AGs. Extracted ion chromatograms (0.1 Da width) of AG analogs with 1-8 unsaturations (panel i-ix; numbers in parentheses denote number of carbons in both isoprenoid chains and number of unsaturations in both chains) in *H. volcanii* (A) and *H. salinarum* (B). Black arrows point to the respective detected chromatographic peaks, whereas red arrows show expected retention time for compounds that could not be detected. Minor peaks at earlier retention times represent isotopologues with two  $^{13}C$  atoms in acetylated unsAR with one more unsaturation or other compounds within the extracted mass window.



# Chapter 5

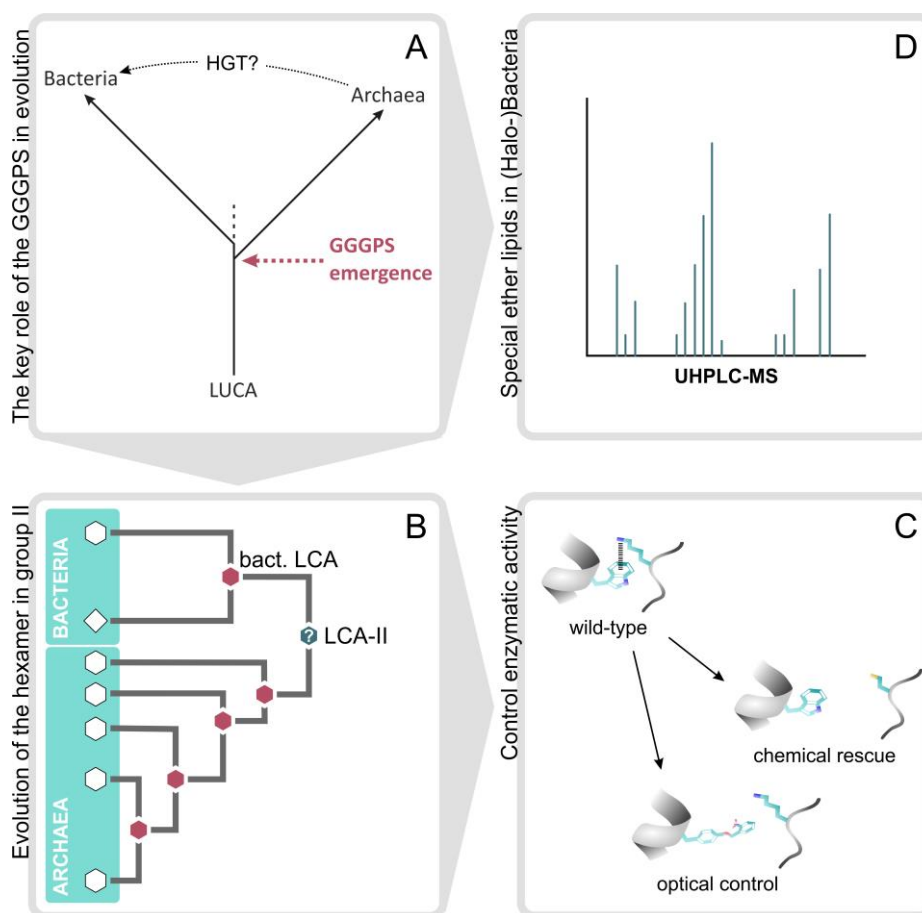
## 5 Comprehensive Summary, Discussion, and Outlook

### 5.1 Comprehensive Summary & Discussion

Ether lipids are a trademark of Archaea and could have been a driving force in speciation of Bacteria and Archaea during evolution. Starting from a last universal common ancestor (LUCA) cell with a heterochiral membrane consisting of G3P and G1P-based lipids together with fatty acids and isoprenoids, bacterial progenitors with G3P-based ester lipids and archaeal progenitors with G1P-based ether lipids evolved (Wächtershäuser 2003). These vast differences in the membrane lipids of Bacteria and Archaea are summarized as the “lipid divide”. It is hypothesized that Archaea were the first primary lineage to emerge (Wächtershäuser 2003), which makes the enzymes synthesizing these ether lipids, such as the GGGPS, interesting, as their first appearance is considered a branch point in evolution (Figure 5.1A; Boucher 2007, Boucher et al 2004, Glansdorff et al 2008, Koga 2011, Koga & Morii 2007, Lombard et al 2012c, Payandeh & Pai 2007, Pereto et al 2004).

#### Considerations from an enzyme-based view

Despite the exclusiveness of ether lipids for the archaeal domain, homologues of the synthesizing enzymes have been found in bacteria as well (Guldan et al 2011, Guldan et al 2008). Either the “lipid divide” event is less clear-cut than thought, or the genes migrated through horizontal gene transfer (HGT) into the bacterial domain, which matches the patchy distribution among Bacteria. Comprehensive work on the phylogenetic distribution of the GGGPS shows that this enzyme family is clustered into two groups, I and II, with both groups harboring archaea and bacteria. Enzymes in group I are solely dimeric, whereas group II enzymes are hexameric or dimeric. On molecular level, a crucial interaction to maintain the hexamer is a cation- $\pi$  bond consisting of an aromate and a cationic counterpart that is located in the ring interface (hexamerization interface, Figure 3.1) (Peterhoff et al 2014).



**Figure 5.1.** Graphical summary and links between the projects. (A) Emergence of the GGGPS is considered an evolutionary branch point in separation of Archaea and Bacteria. Despite being exclusive for the archaeal domain, GGGPS homologues and ether lipids have been found in bacteria too (Chapter 1). (B) Group II GGGPS predecessors have been characterized to find out more about the evolution of the oligomerization state and thermostability, as well as the horizontal gene transfer (HGT) into bacteria (Chapter 2). (C) One of the ancestral enzymes as well as a modern GGGPS were used in a subsequent protein design approach, which aimed at controlling enzymatic activity by switching the association state (Chapter 3). (D) The third aim was to contribute to understanding the function of ether lipids in bacteria. To this end spectrometric measurements in some bacterial and halobacterial species were conducted (Chapter 4). LUCA, last common universal ancestor; LCA-II, last common ancestor of group II GGGPS enzymes; bact. LCA, bacterial LCA of group II GGGPS enzymes; UHPLC-MS, ultra-high performance liquid chromatography coupled to mass spectrometry.

Chapter 2 deals with the evolution of group II GGGPS enzymes in an ancestral sequence reconstruction (ASR) based approach (Figure 5.1B). ASR allows for analysis of proteins where extant representatives are difficult to handle. This was shown lately for a representative of the thioredoxins that was mostly insoluble *in vitro* after heterologous expression in *E. coli* (Gamiz-Arco et al 2021). The authors of this study used computational tools to find back-to-ancestor mutations and by that achieved rescue of inefficient heterologous expression. Enhanced solubility of ancestors was advantageous in our case as well. By characterizing a crenarchaeal GGGPS predecessor (N52), we found out that crenarchaeal GGGPS enzymes might be hexameric, as we

could provide first *in vitro* data for the oligomerization of a crenarchaeal GGGPS. Purification of modern crenarchaeal enzymes was attempted in the past (Peterhoff et al 2014) and in this work (data not shown) but failed due to insolubility. Therefore, it has been proposed that crenarchaeal GGGPS enzymes are dimers as they possess a His on the aromate position in the ring interface, which made it unclear, if the cation- $\pi$  bond would be formed, which is crucial for hexamer formation. Predecessor N52 has a His at the aromatic position and clearly showed hexameric oligomerization (Figure 2.4, Chapter 2). From this, we concluded that all GGGPS enzymes with His and a cationic counterpart at the position of the cation- $\pi$  bond could be hexamers. Consequently, almost all group II enzymes are hexamers. This again demonstrates that ancestral enzymes often are more robust with respect to purification and stability.

The HGT event of the group II GGGPS into the bacterial domain must have occurred at an early stage of evolution. A hexameric GGGPS variant from an archaeal predecessor putatively entered the progenitor of today's Bacteroidetes, because the last common ancestor (LCA) of the bacterial GGGPS enzymes (called N72, indicated in Figure 5.1B), eluted as hexamer in our experiments. As outlined in Chapter 2, we believe that the stimulus for the early emergence of the hexamer was due to stability reasons. Nowadays, the hexamer in a few group II enzymes might fulfill a more sophisticated function, since modern dimeric and hexameric variants are equally thermostable and exhibit high denaturation temperatures. For some modern GGGPS enzymes, hexameric oligomerization seems to further stabilize certain enzyme parts, such as the active site. It is hypothesized that a partial flexibilization of the active site occurs, leading to the appearance of an additional thermal transition point in denaturation experiments (Linde et al (2018) and this work (Chapter 2)). This additional transition point (called  $T_1$ ) is not associated with denaturation (that would be  $T_2$ ), but with partial unfolding of certain structural elements upon heating. It is postulated that these elements are kept more flexible to allow fine-tuning of the flexibility – stability trade-off in GGGPS enzymes. The need for a flexibilization of the active site is consistent with previous studies, hypothesizing that enzymes that have evolved to rigid backbones to achieve higher thermal stability are generally less active (Karshikoff et al 2015). However, because this flexibilization would make the enzyme more sensitive to high temperatures (exceeding  $T_1$  temperature is associated with loss of enzymatic activity; this work and Linde et al (2018)), it must be stabilized by the hexamer in some thermophilic organisms. This has been shown for the hyperthermostable mtGGGPS, whose  $T_1$  shifts to lower temperatures when the hexamer is decomposed to lower homo-oligomers (Linde et al 2018). It remains questionable, why so far only enzymes from Methanobacteriales (e.g., mtGGGPS) have been found to exhibit an additional transition point. Either other hexamers from thermophiles do not have this flexibilization or their  $T_1$  is not visible in experiments.

In the past, the ambient temperature was probably much higher than nowadays. Therefore, ancestral organisms adapted to the hot environment by stabilizing their enzymes, including the GGGPS, by evolving a rigid structure backbone. This was evidenced by high melting points for

ancestral variants (Table 2.1). Fittingly, ancestral enzymes were less active than modern ones (e.g. N12  $k_{\text{cat}}/K_M \sim 34\text{x}$  lower compared to mtGGGPS; Table 3.1 and Figure S3.2), which might be related to low flexibility in the rigid TIM-barrel backbone. The high overall stability likely remained as an artifact in modern mesophilic variants, such as the GGGPS from *F. saliperosum*, which is indicated by the high melting points in DSC and nanoDSF measurements (Table 2.1). Contrary to the analyzed ancestral GGGPS variants, many modern mesophilic representatives most certainly use partial flexibilization as a strategy to achieve high enzymatic activity in a rigid TIM-barrel backbone. Past studies support this theory, as it was shown that rigidifying mutations that were introduced into a heat labile enzyme far from the active site, resulted in a decrease in overall flexibility and catalytic efficiency (D'Amico et al 2001). It might be interesting to analyze whether further stabilization of this partial flexibility can be observed when a mesophilic, dimeric enzyme that is known to exhibit a  $T_1$ , is hexamerized by mutational means.

It can be concluded that nowadays in few thermophilic organisms the hexamer might be important to provide high stability in a hot environment by stabilizing the flexible active site through intermolecular interactions between monomers. Dimeric enzymes of thermophiles may have established an alternative strategy, which still needs to be uncovered. For mesophilic organisms on the other hand, the hexamer is no longer important, since the active site does not need to be thermostable, which is why the hexamer was lost over time in some species. The partial flexibilization is however critical for ensuring high activity in an overall stable backbone, which remained as a remnant of evolution.

In the third chapter, I describe how the oligomerization state can be modified to selectively control enzymatic activity using hexameric GGGPS representatives. This was achieved by exploiting a cation- $\pi$  bond that is located in the ring interface and is indispensable for hexamer maintenance. The cation- $\pi$  forming residues were used to establish a molecular switch for a modern and a reconstructed ancestral enzyme (introduced in Chapter 2) by implementing two strategies: chemical rescue and optical control (Figure 5.1C). The former was quite uncomplicated regarding protein purification, enzyme handling and efficiency of hexamer reformation and thus activity recovery. However, the used exogenous compound had to be applied in 500-fold excess compared to the protein and its toxicity interferes with *in vivo* applications. Optical control on the other hand represents a minimally invasive approach (Szymanski et al 2013). But working with light-sensitive enzymes is much more difficult in terms of purification and handling, and hexamer as well as activity restoration efficiency was decreased. Both strategies have advantages and disadvantages and eventually both could be established to gain selective control over enzymatic activity by inducing hexamer formation via external input.

Activation of GGGPS enzymes upon induction of hexamer formation was only possible because hexameric enzymes show decreased enzymatic activity when disassembled to dimers. The underlying mechanism of a hexameric enzyme experiencing such a deterioration is poorly understood so far, however, data from this thesis (Figure S2.6, Figure S3.9, Table 3.1) and from

previous studies (Linde et al 2018) indicate that binding of the G1P substrate is drastically worsened as the  $K_M$  increases considerably. Since the active site is remote from the interfaces and should not be influenced directly by complex disassembly, some communication between the protomers might happen through the hexamerization interfaces that induces correct conformation of the G1P binding site. When the communication path that possibly propagates through certain residues to the active site is interrupted by disassembly of the enzyme, G1P affinity is worsened, which renders the enzyme less active. Overall, Chapter 3 demonstrates how protein design can help to better understand fundamental attributes of enzymes and how these can be used for enabling selective control over certain properties.

### Considerations from a product-based view

To learn more about Archaea-type ether lipids, as well as the evolution of Archaea and Bacteria, it is important to also consider the lipid product itself. This is especially interesting in context of the “lipid divide”. Archaea-type ether lipids have been found in *B. subtilis* (Firmicutes), however, these are slightly different from those in archaea: Their isoprenoid chain length is C35 compared to C20 chain length in archaea, and instead of addition of a second chain resulting in a diether lipid, the two hydroxyl groups become acetylated in Firmicutes by an enzyme called YvoF (Guldán et al 2011, Linde et al 2016). The physiological function of ether lipids in bacteria is still not understood.

We performed mass spectrometric analysis of Archaea-type ether lipids in archaea and in bacteria as described in Chapter 4 (Figure 5.1D). Our quantification results of ether lipids in *B. subtilis* show that the relative content of ether lipids compared with the main “canonical” phospholipid classes is well below 1% (section 4.1). We therefore hypothesize that ether lipids possess structure-giving attributes and may serve, for example, to fine-tune membrane fluidity, or that they may function as signaling molecules. The latter would be interesting in particular because it would mean that ether lipids in bacteria have structurally and above all, functionally diverged from those in archaea. A recent study has demonstrated that an *E. coli* strain with a heterochiral membrane consisting of canonical bacterial ester and archaeal diether lipids is more resistant against elevated temperatures (Caforio et al 2018). Following these findings, it could be argued that *B. subtilis* could increase ether lipid synthesis in response to elevated environmental temperatures to fine-tune the stabilization of its membrane. Under certain growth conditions, such as in biofilms, intercellular communication plays an important role (Branda et al 2005, Lasa 2006, Whitchurch et al 2002). Along this line, I performed biofilm related experiments with *B. subtilis* in the course of this thesis. However, data indicated no involvement of ether lipids in the formation of biofilm (data not shown).

It has been found that monoether lipids in *B. subtilis* become acetylated at the glycerol (Linde et al 2016), the reason for the acetylation, however, remains speculative. It could serve as a small

head group to adapt the membrane in response to changes in the environment. The discovery of YvoF in Halobacteria (Linde et al 2016) and the tentative identification of acetylated diether lipids in *H. volcanii* and *H. salinarum* (this work, section 4.2) suggests that YvoF must have migrated from bacteria to archaea or vice versa. Since acetylation of lipids is not known to occur in other archaeal species, this modification might in fact serve a distinct function – perhaps the same as in *B. subtilis*. It could confer additional stability to the membrane against stress factors, as discussed above, or alter the spatial polarity of the membrane for the formation of membrane domains. According to the literature, eukaryotic cells can establish temporary spatial polarity during the differentiation process, for example (Putnam 1995). In Halobacteria, acetylated ether lipids rather not act as signaling molecules because their relative content is quite high, as we found in our studies (acetylated archaeol is 5-10% of the amounts of non-acetylated archaeol and acetylated archaetidylglycerols are 30-40% of the amounts of non-acetylated archaetidylglycerol). The reason for the acetylation of quite substantial amounts of diether lipids in Halobacteria as well as its function in Firmicutes remains an enigma. However, tentative identification of acetylated lipids in archaea provides an opportunity to study acetylated mono-/diether lipids more comprehensively in bacteria and archaea.

We speculate that YvoF in Halobacteria might be able to acetylate different kinds of lipids. This raises the question of whether bacterial YvoF (e.g., from *B. subtilis*) also exhibits promiscuous activity and is able to acetylate different lipids in the cell, which is suggested by previous studies showing that YvoF from *B. subtilis* is able to acetylate maltose (Linde et al 2016). The discovery of additional acetylated lipids in *B. subtilis* could help to understand the function of this modification. Phylogenetic analysis revealed that YvoF is absent in Bacteroidetes, implying that in these organisms ether lipids might serve a different function than the acetylated derivatives discovered in Firmicutes.

Our MS experiments suggest that in Bacteroidetes no monoether lipids are present (section 4.1). An explanation why we did not find monoether lipid components in the strains studied, could be that Bacteroidetes synthesize them only under certain growth conditions. Identification of ether lipids in these species would be compelling in the context of the “lipid divide”, as it is postulated that some of these organisms are able to synthesize heterochiral membranes (Villanueva et al 2021), like some Euryarchaeota (Villanueva et al 2017). According to this study, some members of the Fibrobacteres-Chlorobi-Bacteroidetes (FCB) group superphylum may harbor a complete archaea-like membrane lipid biosynthetic pathway. In a metagenomic analysis, genes encoding conventional bacterial membrane synthesis plus putative genes encoding the archaeal lipid biosynthetic pathway were detected in a deep-sea representative of the FCB group. To verify the activity of the enzymes encoded by the detected genes, the putative GGGPS was produced heterologously in *E. coli* and purified for analysis: The enzyme showed typical GGGPS activity. Subsequently, the GGGPS and DGGGPS genes of this species were co-expressed in *E. coli*, resulting in cells with heterochiral membranes (Villanueva et al 2021). This further justifies



search for ether lipids in these organisms, however, identification of ether lipids in Bacteroidetes (or other FCB group member) is still pending and the biosynthesis remains to be understood.

In this context, the absence of a G1PDH enzyme, which is responsible for G1P synthesis, in Bacteroidetes is particularly striking. This may indicate that these organisms are capable of synthesizing Archaea-type ether lipids with G3P. Indeed, G3P promiscuity of GGGPS enzymes has been reported in previous studies (Caforio et al 2018, Chen et al 1993, Zhang & Poulter 1993), with only an eightfold increased  $K_M$  for G3P over G1P in one case. Recently, however, it has been discussed that bacteria must have an alternative pathway for G1P synthesis, as a genetically engineered *E. coli* that has all the genes for diether lipid synthesis except for a G1PDH is still capable of synthesizing substantial amounts of G1P-based diether lipids (Caforio et al 2018). This means that when searching for ether lipids in Bacteroidetes, different cultivation conditions should be tried, as well as experiments to understand the unknown G1P synthesis. In addition, because of tentative identification of a DGGGPS in members of the FCB group in the metagenomic study (Villanueva et al 2021), one should consider the possibility that these organisms synthesize dietherified lipids. This would further distinguish Bacteroidetes from Firmicutes in terms of Archaea-type ether lipids, as dietherified G1P-based lipids with C20 isoprenoid chains (Firmicutes have monoether lipids with C35 chains (Peterhoff et al 2014)) would exactly match archaeal diether lipids. Archaeal diether lipids are thought to be robust against, for example, oxidative stress and elevated temperatures (Koga 2012). Compared to monoether lipids, diether lipids would be more effective against these stress factors due to their second isoprenoid chain, which is why it stands to reason that Bacteroidetes could use diether lipids to stabilize their membrane against aforementioned stress factors.

In summary, these organisms might contradict the clear-cut “lipid divide” as they seem to be able to synthesize original archaeal diether lipids, but Bacteroidetes (and the entire FCB group) could also be considered chimeric forms with a LUCA-like membrane composition and might reveal LUCA-like mechanisms for acquiring and maintaining heterochiral membranes. This would be fascinating, considering that the membranes of most organisms have become homochiral during evolution.

## 5.2 Comprehensive Outlook

### The oligomerization state of the LCA of group I and II

GGGPS enzymes are clustered into two major groups, with group I enzymes being exclusively dimeric and group II consisting of mainly hexameric and some dimeric enzymes (this work and previous studies (Boucher et al 2004, Nemoto et al 2003, Payandeh & Pai 2007, Peterhoff et al 2014)). For group II, we characterized reconstructed predecessors. Our results suggest that the LCA of group II (LCA-II; Figure 2.1) most probably was hexameric, thus, the hexamer evolved extremely early. ASR for the whole GGGPS enzyme family (group I + group II) was attempted

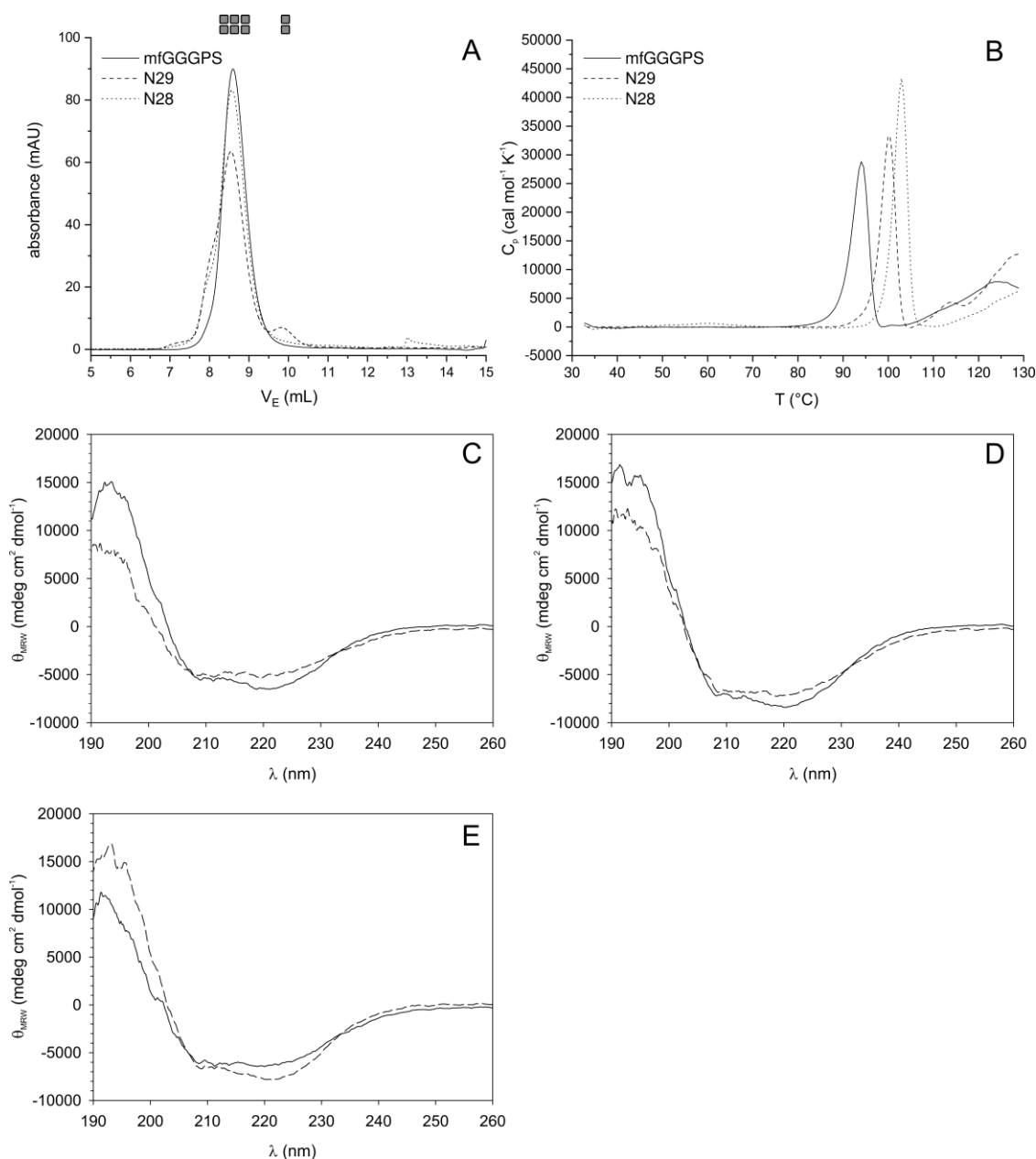
in the past but failed due to the low sequence identity between the groups resulting in a non-robust tree. As outlined in Chapter 2, the most parsimonious scenario is that the LCA of group I (LCA-I; Figure 2.1) was dimeric, due to the exclusive presence of dimers among group I, as well as the hierarchy of interface evolution: The bigger the interface (the dimer module interface is the biggest of all GGGPS interfaces) the earlier it has developed during evolution of the complex (Levy et al 2008, Venkatakrishnan et al 2010). Furthermore, symmetric interfaces, like the dimer module interface, are said to emerge more likely during evolution, because they are easier to generate (Villar et al 2009). All that implies that the group I LCA was dimeric. Nevertheless, characterization of ancestral group I GGGPS enzymes could reveal interesting facts about the evolution of the group I dimer and could confirm the above-mentioned theory of a dimeric LCA-I. To this end, ASR of group I enzymes could be done, followed by cloning, gene expression, and purification and characterization of the predecessors.

### **ASR adds another dimension to understanding the additional thermal transition**

ASR not only helps to trace certain characteristics back in evolution, it also adds an additional dimension for comparison of homologous proteins by creating the basis for the so-called vertical approach (cf. Figure 1.7). Some GGGPS enzymes show two thermal transition points ( $T_1$  and  $T_2$ ).  $T_2$  marks the denaturing temperature, whereas  $T_1$  is caused by partial unfolding of structure elements and exceeding the  $T_1$  in heating experiments can result in inactivation of the enzyme as described in Chapter 2 and in Linde et al (2018). The cause for the additional thermal transition point ( $T_1$ ) could be flexible structure elements, that have not been localized yet *in vitro*. Only MD simulation data is available suggesting that helix  $\alpha_4$ , helix  $\alpha_5$ , loop  $\beta\alpha_6$  and helix  $\alpha_8$  are involved (Linde et al 2018). The involved elements could be identified by characterizing more group II GGGPS predecessors with respect to occurrence two thermal transitions. Important residues can then be filtered out rationally, by analyzing mutations happening between the internal nodes chronologically in the ASR-generated tree (Harms & Thornton 2010). We already analyzed several ancestral GGGPS for occurrence of two thermal transition points, but they showed no  $T_1$  (Figure S2.8). Since we postulate in Chapter 2 that the fine-tuning of the flexibility – stability trade-off evolved late, it might be interesting to expand this search to more recent predecessors of modern GGGPS enzymes known to exhibit a  $T_1$ .

I have already identified two hexameric  $T_1$ -possessing group II GGGPS enzymes that are part of the ASR dataset used in Chapter 2: (i) mrGGGPS (cf. Figure S2.7, for  $T_1$  refer to Figure S2.8E) and (ii) *Methanobacterium formicium* GGGPS (mfGGGPS; Figure 5.2A, B). mrGGGPS was already characterized within Chapter 2. For mfGGGPS, the DSC measurement indicates a  $T_1$  at 94 °C (unpublished data). I then selected interesting predecessors, N29 and N28 (cf. Figure S2.1), along the path from mfGGGPS and mrGGGPS, cloned and purified them (for the method refer to 2.6.2 and 2.6.3) followed by analytical SEC and DSC (2.6.4 and 2.6.6). Both eluted as hexamers (Figure 5.2A, elution volume matches hexamers characterized in Chapter 2 (cf. e.g.,

Figure S2.7)) and exhibited a  $T_1$  in DSC measurements (Figure 5.2B; N29  $T_1$  100 °C, N28  $T_1$  103 °C).



**Figure 5.2.** Analysis of mfGGGPS, N29, and N28. (A) Analytical SEC. The proteins (40  $\mu$ M, subunit concentration) were applied to a S75 analytical column equilibrated with 50 mM potassium phosphate, pH 7.5, 300 mM KCl. Elution was performed at a flow rate of 0.5 mL min<sup>-1</sup> and followed by measuring the absorbance at 280 nm, which was plotted against the elution volume ( $V_E$ ). The derived oligomerization states are indicated by rectangles (2 = dimer, 6 = hexamer). (B) DSC analysis. Changes in the heat capacity of 40  $\mu$ M protein (subunit concentration) in 50 mM potassium phosphate, pH 7.5 were monitored from 30 to 130 °C at a scan rate of 1 K/min. (C-E) Far-UV CD spectra. (C) mfGGGPS, (D) N28 and (E) N29 (6  $\mu$ M, subunit concentration) were recorded in 10 mM potassium phosphate, pH 7.5 from 190 nm to 260 nm ( $d = 1$  mm) at 25 °C (solid line) and 100 °C (dashed line).

In order to exclude the transition points seen to be denaturing, CD spectra of all three proteins were recorded at 25 °C and at 100 °C (Figure 5.2C-E). Comparison of the spectra at low and high temperatures reveals no differences. Even though the transition points seen in the DSC measurements are close to 100 °C, the CD spectra at elevated temperatures clearly indicate that the proteins are still folded and that the observed transition points most likely reflect  $T_1$  transitions.

Next, more predecessors should be investigated for occurrence of an additional transition point. The two progenitors where the “ $T_1$ -to-no- $T_1$  transition” happens can then be compared extensively on sequence level to identify differences in the amino acid sequence to localize the flexible structure elements that cause the additional transition. I have already cloned and purified the next predecessor, N27. N27 is followed by N18, which, as we know from data discussed in Chapter 2 (Figure S2.1, Figure S2.8D), has no additional transition point. Therefore, chances are high that between these two internal nodes, the “ $T_1$ -to-no- $T_1$  transition” occurs.

### **Subunit communication between GGGPS protomers**

In GGGPS(-like) enzymes, homo-oligomeric complex formation was found to influence enzymatic activity, either by enhancing substrate specificity (Peterhoff et al 2012) or by enabling high substrate affinity ( $K_M$ ) (Linde et al 2018). The latter was exploited in Chapter 3 for creating an enzyme switch that allowed to switch on activity of an hexameric GGGPS by modulating its association state. In brief, this was achieved by dimerizing hexameric GGGPS enzymes, which consequently led to loss or decrease of enzymatic activity due to a drastic increase of the  $K_M$  for G1P (Table 3.1, Figure S3.9). This project emphasized that communication might take place between the protomers, since enzymatic activity is decreased massively upon disassembly of the hexamer even though the active sites are remote from the interfaces (Figure 3.1C). Perhaps the communication works similarly to allosteric activation: upon binding of the other protomers (= complex formation), correct conformation of the G1P binding site might be induced. The basis of this study was an ancestral GGGPS enzyme introduced in Chapter 2, called N4, which was dimeric and showed no activity (cf. Figure S2.2 and Figure S3.3). I found out, that the enzyme was unable to bind G1P (Figure S2.6). This ancestral representative can now be used for comparative sequence analysis with other active GGGPS variants to identify residues important for the communication between the protomers. N12, introduced in Chapter 2, provides the perfect counterpart for the sequence comparison as it shows 91% sequence identity to N4, and is still active when dimerized by mutation of the cationic counterpart forming the cation- $\pi$  bond, which is sitting in the ring interface (preliminary data, not shown). By that, residues important for activity in the dimerized N12 could be identified and introduced into N4. The goal would be to design an active dimeric N4. Similarly, mutation of the by that identified residues in a hexameric variant, such as wild-type N12, will hopefully generate an inactive or less active variant, confirming that the communication pathway has been disrupted and thus identified. In

this way, one could learn something about the fundamental relationship between complex formation and activity in GGGPS enzymes and perhaps draw conclusions about the evolution of a communication pathway in GGGPS hexamers and dimers.

### Identification of ether lipids in Bacteroidetes

In our experiments described in Chapter 4, we did not find monoether lipids in the Bacteroidetes strains studied. Several *in vitro* studies have already confirmed GGGPS activity of Bacteroidetes' enzymes (e.g. this work, Linde et al (2018) and Peterhoff et al (2014)). Strikingly, no G1PDH is encoded in the genome of these organisms, although recent metagenomic studies suggest that Bacteroidetes are indeed capable of synthesizing ether lipids (Villanueva et al 2021). Therefore, Bacteroidetes require an alternative G1P source or synthetic pathway. The existence of an alternative pathway was suggested for *E. coli*, as a strain was engineered to contain all genes necessary for the synthesis of diether lipids, with the exception of the G1PDH gene. Remarkably, despite the absence of the G1PDH, the cells had a substantial amount of diether lipids (Caforio et al 2018). This suggests that *E. coli* does have an unknown G1P source, which could also be true for Bacteroidetes, making the detection of ether lipids in these species even more challenging, as it is not known under what circumstances G1P synthesis is activated.

To address identification in future experiments, Bacteroidetes strains could be grown in small culture volumes and fed either labeled  $^{13}\text{C}$ - or  $^{14}\text{C}$ -G1P mixed with unlabeled G1P to increase concentration without using too much expensive labeled G1P. The  $^{13}\text{C}$ - or  $^{14}\text{C}$ -G1P could then be used to search for labeled ether lipids in MS experiments. This would facilitate the search, since only the signal from the labeled G1P needs to be detected. This would fix the problem we had of not knowing what to look for. Of course, this only works if the strains actually synthesize G1P-based ether lipids and if the cells actually take up the G1P. Indeed, experiments from the past (Beer 2012) and from this work (data not shown) indicate that G1P is not taken up into the cells. Briefly, cells were grown under standard lab cultivation conditions. The medium was supplemented with substantial amounts of  $^{14}\text{C}$ -G1P. A lipid extract was produced and analyzed via TLC. However, no spots on the TLC plate could be observed, indicating that no uptake had occurred (Beer 2012, this work (data not shown)). Furthermore, investigation of Bacteroidetes GGGPS substrate specificity in terms of isoprenoid chain length and G1P/G3P promiscuity could be worth a try to further specify the search in MS experiments.

Although production of Bacteroidetes knockout strains is complex due to limited genetic accessibility (cf. Rhodes et al (2011) and Shrivastava et al (2013)), this might be worth trying, as the GGGPS knockout strain could be used in growth or MS experiments for comparison with the wild-type lipid profile. However, one of the most important experiments should be to culture the strains under a variety of cultivation conditions to potentially trigger ether lipid synthesis and analyze them in MS. As outlined in section 5.1, ether lipid synthesis might only be activated

under certain environmental conditions, such as elevated temperatures. In addition, diether lipids should be considered in all analyses, as these organisms might have an unknown DGGGPS (outlined in section 5.1, too).

### **The function of acetylated ether lipids in Bacilli (Firmicutes) and Halobacteria**

The function of ether lipids in bacteria remains elusive so far. As discussed in Chapter 4 and in section 5.1, I believe that ether lipids serve different functions in Firmicutes and Bacteroidetes, and therefore this section focuses only on *B. subtilis* and Halobacteria, since we have tentatively identified acetylated diether lipids in two halobacterial species.

In Chapter 4 (section 4.1), we quantified the relative content of ether lipids in *B. subtilis* compared to the main “canonical” phospholipids. Our results indicate that the proportion of ether lipids is very low (well below 1% compared to the phospholipids). It is possible that *B. subtilis* synthesizes ether lipids to a greater extent under certain environmental conditions. Therefore, the wild-type strain should be cultured under different conditions (temperature, salt, nutrition in general, oxygenation, etc.) and then the total lipid extracts (TLE) of the cultured cells should be analyzed by thin layer chromatography (TLC) or UHPLC-MS for increased ether lipid levels. Altered synthesis levels depending on different cultivation conditions could reveal a physiological function of ether lipids. Since acetylated ether lipids are specific for Firmicutes and Halobacteria, these experiments should also be performed for halobacterial strains. In addition, a *H. volcanii*  $\Delta yvoF$  knockout strain could be generated and studied phenotypically in growth experiments or analytically by UHPLC-MS. It is possible that the strain compensates for the lack of acetylation by up-regulating other components, or that the cells show a distinct phenotype in growth experiments that provides functional information for the acetylation. The results may help to determine a function of ether lipids in bacteria in general. Elucidation of the function may also help to understand the “lipid divide” by providing new general properties of archaeal ether lipids that may have supported the “lipid divide”.

### **Reenacting the “lipid divide”**

How and, above all, why the “lipid divide” occurred, remains the most fundamental question in evolutionary biology that has not yet been answered. It has long been suggested that a homochiral membrane is more stable than a heterochiral one, which is why it established itself as the dominant phenotype in evolution. This is now being discussed controversially, since artificially created heterochiral vesicles, which were produced in test tube, do not have any disadvantages in terms of stability compared to homochiral ones. Furthermore, an engineered *E. coli* with heterochiral membrane showed higher resistance to elevated temperatures (Caforio et al 2018, Shimada & Yamagishi 2011). It would be interesting to see if the “lipid divide” can be reenacted

in the lab to investigate if and how fast the lipid types separate in living organisms. For this, one could take an archaeal strain that is genetically easily accessible (cf. Leigh et al (2011)) and equip it with the most important genes for ester lipid synthesis in addition to its native ether lipid biosynthetic pathway. This should theoretically be feasible because fatty acid (Lombard et al 2012a) and G3P synthesis (Williams et al 2017) are already present in archaea. Alternatively, the already engineered *E. coli* mentioned above could be used for reenacting the “lipid divide”. There are naturally occurring archaea and bacteria that are suspected of being able to build heterochiral membranes, which was proposed after metagenomic analysis of members of the Euryarchaeota and of the FCB group superphylum (Villanueva et al 2017, Villanueva et al 2021). Intensive study of these and the engineered *E. coli* (Caforio et al 2018, Yokoi et al 2012) should definitely be done beforehand in order to collect *a priori* knowledge.

The modified strain (containing biosynthesis for both ester and ether lipids) could then be incubated at high growth temperatures to mimic the high environmental temperature in the past and could repeatedly be transferred to new nutrient media in several rounds of selection. Perhaps it is useful to increase the mutation rate by, for example, irradiation. It is possible that the cells throw out the foreign genes for the respective lipid synthesis. But the organism might also go through a small “lipid divide” in which the cells could downregulate the respective promoters to generate homochiral membranes – perhaps two cell populations emerge, one with bacterial, the other with archaeal membrane lipids. This could be assessed by sequencing and by MS. The experiment would be done without artificial selection pressure, since it is discussed whether the selection pressure towards homochiral membranes is present within the cell (Koga 2011, Sojo 2019, Wächtershäuser 2003). This project is obviously still highly theoretical, however, reenacting the “lipid divide” would be incredibly interesting.





## List of Abbreviations

Ac	acetyl...
AG	archaetidylglycerol
AG-PCH <sub>3</sub>	AG methylphosphate
AncGGGPS_ <u>N..</u> / <u>_N..</u> / N..	reconstructed ancestral variant of group II GGGPS at node N..
APCI	atmospheric pressure chemical ionization
AR	archaeol
ASR	ancestral sequence reconstruction
BrEtAm	bromoethylamine
BrEtOH	bromoethanol
CD	circular dichroism
CDP	cytidine diphosphate
CIP	calf intestinal phosphatase
DCM	dichloromethane
DGGGPS	digeranylgeranylgeranyl glyceryl phosphate synthase
DHAP	dihydroxyacetone phosphate
DP	differential power
DPG-DAG	diphosphatidylglycerol
DSC	differential scanning calorimetry
DSMZ	German Collection of Microorganisms and Cell Cultures
ESI	electrospray ionization
ether lipid / monoether lipid	used for “Archaea-type ether lipid”
FCB group	Fibrobacteres–Chlorobi–Bacteroidetes group
G1P	<i>sn</i> -glycerol 1-phosphate
G1PDH	glycerol 1-phosphate dehydrogenase
G3P	<i>sn</i> -glycerol 3-phosphate
G3PDH	glycerol 3-phosphate dehydrogenase
GGG	geranylgeranylgeranyl glycerol
GGGP	(S)-3- <i>O</i> -geranylgeranylgeranyl glyceryl phosphate
GGGPS	geranylgeranylgeranyl glyceryl phosphate synthase
..GGGPS (e.g. mtGGGPS)	extant GGGPS enzymes, listed in Table S2.1
GGPP	geranylgeranyl pyrophosphate
HGT	horizontal gene transfer
HepGP	heptaprenylglyceryl phosphate
HepPP	heptaprenyl pyrophosphate
HepGPS	heptaprenylglyceryl phosphate synthase
(His) <sub>6</sub> tag	hexahistidine tag

IMAC	immobilized metal chelate affinity chromatography
IPTG	isopropyl $\beta$ -D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
LCA	last common ancestor
LUCA	last universal common ancestor
MAT	maltose acetyltransferase
MBP	maltose binding protein
MD	molecular dynamics
MeOH	methanol
mfGGGPS	<i>Methanobacterium formicium</i> GGGPS
MS	mass spectrometry
MW	molecular weight
NAD	nicotinamide adenine dinucleotide
nanoDSF	nano differential scanning fluorimetry
NP	normal phase
ONBY	<i>o</i> -nitrobenzyl- <i>O</i> -tyrosine
P <sub>i</sub>	phosphate
PC-DAG	phosphatidylcholine
PCR	polymerase chain reaction
PE-DAG	phosphatidylethanolamine
PEG	polyethylene glycol
PG-DAG	phosphatidylglycerol
PME-DAG	phosphatidyl-(N)-methylethanolamines
PP <sub>i</sub>	diphosphate
RGGGPS	rescued GGGPS
RP	reversed phase
SEC	size exclusion chromatography
SLS	static light scattering
SSN	sequence similarity network
TIM	triosephosphate isomerase
TLC	thin layer chromatography
TLE	total lipid extract
UAA	unnatural amino acid
UHPLC	ultra-high performance liquid chromatography
uns	unsaturated ...

## Bibliography

- Aberer AJ, Krompass D, Stamatakis A. 2013. Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. *Syst. Biol.* 62: 162-6
- Alford RF, Leaver-Fay A, Jeliazkov JR, O'Meara MJ, DiMaio FP, et al. 2017. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* 13: 3031-48
- Allers T, Barak S, Liddell S, Wardell K, Mevarech M. 2010. Improved strains and plasmid vectors for conditional overexpression of His-tagged proteins in *Haloferax volcanii*. *Appl. Environ. Microbiol.* 76: 1759-69
- Allers T, Ngo HP, Mevarech M, Lloyd RG. 2004. Development of additional selectable markers for the halophilic Archaeon *Haloferax volcanii* based on the leuB and trpA genes. *Appl. Environ. Microbiol.* 70: 943-53
- Ankenbruck N, Courtney T, Naro Y, Deiters A. 2018. Optochemical Control of Biological Processes in Cells and Animals. *Angew. Chem. Int. Ed.* 57: 2768-98
- Ashkenazy H, Penn O, Doron-Faigenboim A, Cohen O, Cannarozzi G, et al. 2012. FastML: a web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Res.* 40: W580-4
- Bachmann BJ. 1990. Linkage map of *Escherichia coli* K-12, edition 8. *Microbiol Rev.* 54: 130-97
- Badger J, Sauder JM, Adams JM, Antonysamy S, Bain K, et al. 2005. Structural analysis of a set of proteins resulting from a bacterial genomics project. *Proteins* 60: 787-96
- Banner DW, Bloomer AC, Petsko GA, Phillips DC, Pogson CI, et al. 1975. Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 angstrom resolution using amino acid sequence data. *Nature* 255: 609-14
- Beer B. 2012. *Archaeen-typische Lipide in Bakterien: Vergleichende Charakteristiken von Prenyltransferasen*. Master thesis. Universität Regensburg, Regensburg. 96 pp.
- Blank PN, Barnett AA, Ronnebaum TA, Alderfer KE, Gillott BN, et al. 2020. Structural studies of geranylgeranylglycerol phosphate synthase, a prenyltransferase found in thermophilic Euryarchaeota. *Acta Crystallogr. D Struct. Biol.* 76: 542-57
- Bligh EG, Dyer WJ. 1959. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* 37: 911-7

- Boucher Y. 2007. Lipids: biosynthesis, function, and evolution. In *Archaea, Molecular and Cellular Biology*, ed. R Cavicchioli. Washington, D.C.: ASM Press
- Boucher Y, Kamekura M, Doolittle WF. 2004. Origins and evolution of isoprenoid lipid biosynthesis in archaea. *Mol. Microbiol.* 52: 515-27
- Boussau B, Blanquart S, Necsulea A, Lartillot N, Gouy M. 2008. Parallel adaptations to high temperatures in the Archaean eon. *Nature* 456: 942-5
- Branda SS, Vik S, Friedman L, Kolter R. 2005. Biofilms: the matrix revisited. *Trends Microbiol* 13: 20-6
- Caforio A, Siliakus MF, Exterkate M, Jain S, Jumde VR, et al. 2018. Converting *Escherichia coli* into an archaeobacterium with a hybrid heterochiral membrane. *Proc. Natl. Acad. Sci. U. S. A.* 115: 3704-09
- Cauet E, Rooman M, Wintjens R, Lievin J, Biot C. 2005. Histidine-Aromatic Interactions in Proteins and Protein-Ligand Complexes: Quantum Chemical Study of X-ray and Model Structures. *J. Chem. Theory Comput.* 1: 472-83
- Chen A, Zhang D, Poulter CD. 1993. (S)-geranylgeranylgeranyl glyceryl phosphate synthase. Purification and characterization of the first pathway-specific enzyme in archaeobacterial membrane lipid biosynthesis. *J. Biol. Chem.* 268: 21701-5
- Cline SW, Lam WL, Charlebois RL, Schalkwyk LC, Doolittle WF. 1989. Transformation-Methods for Halophilic Archaeobacteria. *Can. J. Microbiol.* 35: 148-52
- Courtney T, Deiters A. 2018. Recent advances in the optical control of protein function through genetic code expansion. *Curr. Opin. Chem. Biol.* 46: 99-107
- Crowley PB, Golovin A. 2005. Cation-pi interactions in protein-protein interfaces. *Proteins* 59: 231-9
- D'Amico S, Gerday C, Feller G. 2001. Structural determinants of cold adaptation and stability in a large protein. *J. Biol. Chem.* 276: 25791-6
- D'Amico S, Marx JC, Gerday C, Feller G. 2003. Activity-stability relationships in extremophilic enzymes. *J. Biol. Chem.* 278: 7891-6
- Deckert K, Budiardjo SJ, Brunner LC, Lovell S, Karanicolas J. 2012. Designing allosteric control into enzymes by chemical rescue of structure. *J. Am. Chem. Soc.* 134: 10055-60

- DiMaio F, Leaver-Fay A, Bradley P, Baker D, Andre I. 2011. Modeling symmetric macromolecular structures in Rosetta3. *PLoS One* 6: e20450
- Dummer AM, Bonsall JC, Cihla JB, Lawry SM, Johnson GC, Peck RF. 2011. Bacterioopsin-mediated regulation of bacterioruberin biosynthesis in *Halobacterium salinarum*. *J. Bacteriol.* 193: 5658-67
- Eick GN, Bridgham JT, Anderson DP, Harms MJ, Thornton JW. 2016. Robustness of Reconstructed Ancestral Protein Functions to Statistical Uncertainty. *Molecular Biology and Evolution*. 34(2): 247-61.
- Engelke H, Chou C, Uprety R, Jess P, Deiters A. 2014. Control of protein function through optochemical translocation. *ACS Synth. Biol.* 3: 731-6
- Erickson SB, Mukherjee R, Kelemen RE, Wrobel CJ, Cao X, Chatterjee A. 2017. Precise Photoremovable Perturbation of a Virus-Host Interaction. *Angew. Chem. Int. Ed.* 56: 4234-37
- Eriksson AE, Baase WA, Wozniak JA, Matthews BW. 1992. A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. *Nature* 355: 371-3
- Gamiz-Arco G, Risso VA, Gaucher EA, Gavira JA, Naganathan AN, et al. 2021. Combining Ancestral Reconstruction with Folding-Landscape Simulations to Engineer Heterologous Protein Expression. *J. Mol. Biol.* Volume 433: 167321
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, et al. 2005. Protein identification and analysis tools on the ExpASY server. In *The proteomics protocols handbook*, ed. JM Walker, pp. xviii, 988 p. Totowa, N.J.: Humana Press
- Glansdorff N, Xu Y, Labedan B. 2008. The Last Universal Common Ancestor: emergence, constitution and genetic legacy of an elusive forerunner. *Biol. Direct* 3: 29
- Guldan H. 2010. *Nachweis Archaea-typischer Lipide in Bacteria über die Aufklärung der Funktion von AraM und PcrB aus Bacillus Subtilis*. Dissertation thesis. Universität Regensburg, Regensburg
- Guldan H, Matysik FM, Bocola M, Sterner R, Babinger P. 2011. Functional assignment of an enzyme that catalyzes the synthesis of an Archaea-type ether lipid in Bacteria. *Angew. Chem. Int. Ed. Engl.* 50: 8188-91
- Guldan H, Sterner R, Babinger P. 2008. Identification and characterization of a bacterial glycerol-1-phosphate dehydrogenase: Ni<sup>2+</sup>-dependent AraM from *Bacillus subtilis*. *Biochemistry* 47: 7376-84

- Guy CP, Haldenby S, Brindley A, Walsh DA, Briggs GS, et al. 2006. Interactions of RadB, a DNA repair protein in archaea, with DNA and ATP. *Journal of Molecular Biology* 358: 46-56
- Hanson-Smith V, Kolaczowski B, Thornton JW. 2010. Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. *Mol Biol Evol* 27: 1988-99
- Harms MJ, Thornton JW. 2010. Analyzing protein structure and function using ancestral gene reconstruction. *Curr. Opin. Struc. Biol.* 20: 360-66
- Hemphill J, Chou C, Chin JW, Deiters A. 2013. Genetically encoded light-activated transcription for spatiotemporal control of gene expression and gene silencing in mammalian cells. *J. Am. Chem. Soc.* 135: 13433-9
- Hochberg GKA, Thornton JW. 2017. Reconstructing Ancient Proteins to Understand the Causes of Structure and Function. *Annu. Rev. Biophys.* 46: 247-69
- Huguet C, Hopmans EC, Febo-Ayala W, Thompson DH, Damste JSS, Schouten S. 2006. An improved method to determine the absolute abundance of glycerol dibiphytanyl glycerol tetraether lipids. *Organic Geochemistry* 37: 1036-41
- Itabashi Y, Kuksis A. 1997. Reassessment of stereochemical configuration of natural phosphatidylglycerols by chiral-phase high-performance liquid chromatography and electro spray mass spectrometry. *Anal. Biochem.* 254: 49-56
- Jain S, Caforio A, Driessen AJ. 2014. Biosynthesis of archaeal membrane ether lipids. *Front. Microbiol.* 5: 641
- Karshikoff A, Nilsson L, Ladenstein R. 2015. Rigidity versus flexibility: the dilemma of understanding protein thermal stability. *FEBS J.* 282: 3899-917
- Kates M. 1986. Techniques of lipidology: isolation, analysis, and identification of lipids. In *Laboratory techniques in biochemistry and molecular biology vol. 3, part 2*, ed. RH Burdon, PH van Knippenberg, pp. 1-464. Amsterdam: Elsevier
- Kates M. 1993. Membrane lipids in Archaea. In *The biochemistry of archaea*, ed. M Kates, DJ Kushner, AT Matheson, pp. 261-95. Amsterdam: Elsevier
- Kelly SM, Jess TJ, Price NC. 2005. How to study proteins by circular dichroism. *Bba-Proteins Proteom* 1751: 119-39
- Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, et al. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41: e1

- Kneuttinger AC, Zwisele S, Straub K, Bruckmann A, Busch F, et al. 2019. Light-Regulation of Tryptophan Synthase by Combining Protein Design and Enzymology. *Int. J. Mol. Sci.* 20
- Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, et al. 2003. Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci. U. S. A.* 100: 4678-83
- Koga Y. 2011. Early Evolution of Membrane Lipids: How did the Lipid Divide Occur? *J Mol Evol* 72: 274-82
- Koga Y. 2012. Thermal Adaptation of the Archaeal and Bacterial Lipid Membranes. *Archaea-an International Microbiological Journal* 2012: 1-6
- Koga Y, Kyuragi T, Nishihara M, Sone N. 1998. Did archaeal and bacterial cells arise independently from noncellular precursors? A hypothesis stating that the advent of membrane phospholipid with enantiomeric glycerophosphate backbones caused the separation of the two lines of descent. *J Mol Evol* 46: 54-63
- Koga Y, Morii H. 2007. Biosynthesis of ether-type polar lipids in archaea and evolutionary considerations. *Microbiol. Mol. Biol. Rev.* 71: 97-120
- Koga Y, Nishihara M, Morii H, Akagawa-Matsushita M. 1993. Ether polar lipids of methanogenic bacteria: structures, comparative aspects, and biosyntheses. *Microbiol. Rev.* 57: 164-82
- Kropp C, Straub K, Linde M, Babinger P. 2020. Hexamerization and thermostability emerged very early during geranylgeranylglyceryl phosphate synthase evolution. *Protein Sci.* 30: 583-96
- Kumar K, Woo SM, Siu T, Cortopassi WA, Duarte F, Paton RS. 2018. Cation- $\pi$  interactions in protein-ligand binding: theory and data-mining reveal different roles for lysine and arginine. *Chem. Sci.* 9: 2655-65
- Lange BM, Rujan T, Martin W, Croteau R. 2000. Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc. Natl. Acad. Sci. U. S. A.* 97: 13172-7
- Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25: 2286-8
- Lasa I. 2006. Towards the identification of the common features of bacterial biofilm development. *Int Microbiol* 9: 21-8
- Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, et al. 2011. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* 487: 545-74

- Leigh JA, Albers SV, Atomi H, Allers T. 2011. Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales. *Fems Microbiology Reviews* 35: 577-608
- Leopoldseder S, Hettwer S, Sterner R. 2006. Evolution of multi-enzyme complexes: the case of tryptophan synthase. *Biochemistry* 45: 14111-9
- Levy ED, Boeri Erba E, Robinson CV, Teichmann SA. 2008. Assembly reflects evolution of protein complexes. *Nature* 453: 1262-5
- Liao SM, Du QS, Meng JZ, Pang ZW, Huang RB. 2013. The multiple roles of histidine in protein interactions. *Chem. Cent. J.* 7: 44
- Linde M, Heyn K, Merkl R, Sterner R, Babinger P. 2018. Hexamerization of Geranylgeranylglyceryl Phosphate Synthase Ensures Structural Integrity and Catalytic Activity at High Temperatures. *Biochemistry* 57: 2335-48
- Linde M, Peterhoff D, Sterner R, Babinger P. 2016. Identification and characterization of heptaprenylglyceryl phosphate processing enzymes in *Bacillus subtilis*. *J. Biol. Chem.* 291: 14861-70
- Liu J, Cropp TA. 2012. Experimental methods for scanning unnatural amino acid mutagenesis. *Methods Mol. Biol.* 794: 187-97
- Lo Leggio L, Dal Degan F, Poulsen P, Andersen SM, Larsen S. 2003. The structure and specificity of *Escherichia coli* maltose acetyltransferase give new insight into the LacA family of acyltransferases. *Biochemistry* 42: 5225-35
- Lombard J, Lopez-Garcia P, Moreira D. 2012a. An ACP-independent fatty acid synthesis pathway in archaea: implications for the origin of phospholipids. *Mol Biol Evol* 29: 3261-5
- Lombard J, Lopez-Garcia P, Moreira D. 2012b. The early evolution of lipid membranes and the three domains of life. *Nat. Rev. Microbiol.* 10: 507-15
- Lombard J, Lopez-Garcia P, Moreira D. 2012c. Phylogenomic investigation of phospholipid synthesis in archaea. *Archaea* 2012: 630910
- Marshall LR, Zozulia O, Lengyel-Zhand Z, Korendovych IV. 2019. Minimalist *de novo* Design of Protein Catalysts. *ACS Catal.* 9: 9265-75
- Matsumi R, Atomi H, Driessen AJ, van der Oost J. 2011. Isoprenoid biosynthesis in Archaea - biochemical and evolutionary implications. *Res. Microbiol.* 162: 39-52



- Merkl R, Sterner R. 2016. Ancestral protein reconstruction: techniques and applications. *Biol. Chem.* 397: 1-21
- Moldoveanu N, Kates M. 1988. Biosynthetic studies of the polar lipids of *Halobacterium cutirubrum*. Formation of isoprenyl ether intermediates. *Biochimica et biophysica acta*: 164 - 82
- Nemoto N, Miyazono KI, Tanokura M, Yamagishi A. 2019. Crystal structure of (S)-3-O-geranylgeranylglyceryl phosphate synthase from *Thermoplasma acidophilum* in complex with the substrate sn-glycerol 1-phosphate. *Acta Crystallogr. F Struct. Biol. Commun.* 75: 470-79
- Nemoto N, Oshima T, Yamagishi A. 2003. Purification and characterization of geranylgeranylglyceryl phosphate synthase from a thermoacidophilic archaeon, *Thermoplasma acidophilum*. *J. Biochem. (Tokyo)* 133: 651-7
- Neuhaus FC, Baddiley J. 2003. A continuum of anionic charge: structures and functions of D-alanyl-teichoic acids in gram-positive bacteria. *Microbiol. Mol. Biol. Rev.* 67: 686-723
- Ng WV, Kennedy SP, Mahairas GG, Berquist B, Pan M, et al. 2000. Genome sequence of *Halobacterium species* NRC-1. *P. Natl. Acad. Sci. U. S. A.* 97: 12176-81
- Olucha J, Meneely KM, Lamb AL. 2012. Modification of residue 42 of the active site loop with a lysine-mimetic side chain rescues isochorismate-pyruvate lyase activity in *Pseudomonas aeruginosa* PchB. *Biochemistry* 51: 7525-32
- Ottosen MB, Bjornberg O, Norager S, Larsen S, Palfey BA, Jensen KF. 2002. The dimeric dihydroorotate dehydrogenase A from *Lactococcus lactis* dissociates reversibly into inactive monomers. *Protein Sci.* 11: 2575-83
- Parsonage D, Youngblood DS, Sarma GN, Wood ZA, Karplus PA, Poole LB. 2005. Analysis of the link between enzymatic activity and oligomeric state in AhpC, a bacterial peroxiredoxin. *Biochemistry* 44: 10583-92
- Payandeh J, Fujihashi M, Gillon W, Pai EF. 2006. The crystal structure of (S)-3-O-geranylgeranylglyceryl phosphate synthase reveals an ancient fold for an ancient enzyme. *J. Biol. Chem.* 281: 6070-8
- Payandeh J, Pai EF. 2007. Enzyme-driven speciation: crystallizing Archaea via lipid capture. *J Mol Evol* 64: 364-74
- Peracchi A. 2008. How (and Why) to Revive a Dead Enzyme: The Power of Chemical Rescue. *Current Chemical Biology* 2: 32-49

- Perepelov AV, Wang Q, Senchenkova SN, Shevelev SD, Zhao G, et al. 2006. Structure of a teichoic acid-like O-polysaccharide of *Escherichia coli* O29. *Carbohydr. Res.* 341: 2176-80
- Pereto J, Lopez-Garcia P, Moreira D. 2004. Ancestral lipid biosynthesis and early membrane evolution. *Trends Biochem. Sci.* 29: 469-77
- Perez-Jimenez R, Ingles-Prieto A, Zhao ZM, Sanchez-Romero I, Alegre-Cebollada J, et al. 2011. Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. *Nat. Struct. Mol. Biol.* 18: 592-6
- Peterhoff D, Beer B, Rajendran C, Kumpula EP, Kapetaniou E, et al. 2014. A comprehensive analysis of the geranylgeranylglyceryl phosphate synthase enzyme family identifies novel members and reveals mechanisms of substrate specificity and quaternary structure organization. *Mol. Microbiol.* 92: 885-99
- Peterhoff D, Zellner H, Guldan H, Merkl R, Sterner R, Babinger P. 2012. Dimerization determines substrate specificity of a bacterial prenyltransferase. *Chembiochem* 13: 1297-303
- Pillai AS, Chandler SA, Liu Y, Signore AV, Cortez-Romero CR, et al. 2020. Origin of complexity in haemoglobin evolution. *Nature* 581: 480-85
- Putnam RW. 1995. Polarity of Cell Membranes. In *Cell Physiology Source Book*, ed. N Sperelakis, pp. Pages 230-42: Academic Press
- Ren F, Feng X, Ko TP, Huang CH, Hu Y, et al. 2013. Insights into TIM-barrel prenyl transferase mechanisms: crystal structures of PcrB from *Bacillus subtilis* and *Staphylococcus aureus*. *Chembiochem* 14: 195-9
- Ren F, Ko TP, Feng X, Huang CH, Chan HC, et al. 2012. Insights into the mechanism of the antibiotic-synthesizing enzyme MoeO5 from crystal structures of different complexes. *Angew. Chem. Int. Ed. Engl.* 51: 4157-60
- Repina NA, Rosenbloom A, Mukherjee A, Schaffer DV, Kane RS. 2017. At Light Speed: Advances in Optogenetic Systems for Regulating Cell Signaling and Behavior. *Annu. Rev. Chem. Biomol. Eng.* 8: 13-39
- Rhodes RG, Pucker HG, McBride MJ. 2011. Development and use of a gene deletion strategy for *Flavobacterium johnsoniae* to identify the redundant gliding motility genes remF, remG, remH, and remI. *J. Bacteriol.* 193: 2418-28
- Rohweder B, Semmelmann F, Endres C, Sterner R. 2018. Standardized cloning vectors for protein production and generation of large gene libraries in *Escherichia coli*. *Biotechniques* 64: 24-26

- Ronnekleiv M. 1995. Bacterial carotenoids 53\* C50-carotenoids 23; carotenoids of *Haloferox volcanii* versus other halophilic bacteria. *Biochemical Systematics and Ecology* 23: 627-34
- Sambrook J, Frisch, E. E., Maniatis, T. 1989. Molecular cloning: a laboratory manual. New York: Cold Spring Harbour
- Schmidt-Dannert C, Lopez-Gallego F. 2016. A roadmap for biocatalysis - functional and spatial orchestration of enzyme cascades. *Microb. Biotechnol.* 9: 601-9
- Semmelmann F, Hupfeld E, Heizinger L, Merkl R, Sterner R. 2019. A Fold-Independent Interface Residue Is Crucial for Complex Formation and Allosteric Signaling in Class I Glutamine Amidotransferases. *Biochemistry* 58: 2584-88
- Seydlova G, Svobodova J. 2008. Development of membrane lipids in the surfactin producer *Bacillus subtilis*. *Folia Microbiol (Praha)* 53: 303-7
- Shimada H, Yamagishi A. 2011. Stability of heterochiral hybrid membrane made of bacterial sn-G3P lipids and archaeal sn-G1P lipids. *Biochemistry* 50: 4114-20
- Shrivastava A, Johnston JJ, van Baaren JM, McBride MJ. 2013. *Flavobacterium johnsoniae* GldK, GldL, GldM, and SprA Are Required for Secretion of the Cell Surface Gliding Motility Adhesins SprB and RemA. *Journal of Bacteriology* 195: 3201-12
- Sircar A, Chaudhury S, Kilambi KP, Berrondo M, Gray JJ. 2010. A generalized approach to sampling backbone conformations with RosettaDock for CAPRI rounds 13-19. *Proteins* 78: 3115-23
- Socan J, Purg M, Aqvist J. 2020. Computer simulations explain the anomalous temperature optimum in a cold-adapted enzyme. *Nat. Commun.* 11: 2644
- Sojo V. 2019. Why the Lipid Divide? Membrane Proteins as Drivers of the Split between the Lipids of the Three Domains of Life. *Bioessays* 41: 1800251-7
- Sprott GD, Larocque S, Cadotte N, Dicaire CJ, McGee M, Brisson JR. 2003. Novel polar lipids of halophilic eubacterium *Planococcus* H8 and archaeon *Haloferox volcanii*. *Biochim. Biophys. Acta* 1633: 179-88
- Sterner R, Höcker B. 2005. Catalytic versatility, stability, and evolution of the (ba)<sub>8</sub>-barrel enzyme fold. *Chemical reviews* 105: 4038-55
- Sterner R, Liebl W. 2001. Thermophilic adaptation of proteins. *Crit. Rev. Biochem. Mol. Biol.* 36: 39-106

- Straub K, Linde M, Kropp C, Blanquart S, Babinger P, Merkl R. 2019. Sequence selection by FitSS4ASR alleviates ancestral sequence reconstruction as exemplified for geranylgeranylgeranyl glyceryl phosphate synthase. *Biol. Chem.* 400: 367-81
- Stroud A, Liddell S, Allers T. 2012. Genetic and Biochemical Identification of a Novel Single-Stranded DNA-Binding Complex in *Haloferax volcanii*. *Front. Microbiol.* 3: 224
- Suarez ASG, Stefan A, Lemma S, Conte E, Hochkoeppler A. 2012. Continuous enzyme-coupled assay of phosphate- or pyrophosphate-releasing enzymes. *Biotechniques* 53: 99-103
- Szymanski W, Beierle JM, Kistemaker HAV, Velema WA, Feringa BL. 2013. Reversible Photocontrol of Biological Systems by the Incorporation of Molecular Photoswitches. *Chemical Reviews* 113: 6114-78
- Taron DJ, Childs WC, 3rd, Neuhaus FC. 1983. Biosynthesis of D-alanyl-lipoteichoic acid: role of diglyceride kinase in the synthesis of phosphatidylglycerol for chain elongation. *J. Bacteriol.* 154: 1110-6
- Thompson LC, Walters J, Burke J, Parsons JF, Armstrong RN, Dirr HW. 2006. Double mutation at the subunit interface of glutathione transferase rGSTM1-1 results in a stable, folded monomer. *Biochemistry* 45: 2267-73
- Thornton JW. 2004. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.* 5: 366-75
- Toney MD, Kirsch JF. 1989. Direct Bronsted analysis of the restoration of activity to a mutant enzyme by exogenous amines. *Science* 243: 1485-8
- Toney MD, Kirsch JF. 1992. Bronsted analysis of aspartate aminotransferase via exogenous catalysis of reactions of an inactive mutant. *Protein Sci.* 1: 107-19
- Venkatakrishnan AJ, Levy ED, Teichmann SA. 2010. Homomeric protein complexes: evolution and assembly. *Biochem. Soc. Trans.* 38: 879-82
- Vialle RA, Tamuri AU, Goldman N. 2018. Alignment Modulates Ancestral Sequence Reconstruction Accuracy. *Mol Biol Evol* 35: 1783-97
- Villanueva L, Schouten S, Damste JS. 2017. Phylogenomic analysis of lipid biosynthetic genes of Archaea shed light on the 'lipid divide'. *Environ. Microbiol.* 19: 54-69
- Villanueva L, von Meijenfeldt FAB, Westbye AB, Yadav S, Hopmans EC, et al. 2021. Bridging the membrane lipid divide: bacteria of the FCB group superphylum have the potential to synthesize archaeal ether lipids. *ISME J.* 15: 168-82

- Villar G, Wilber AW, Williamson AJ, Thiara P, Doye JP, et al. 2009. Self-assembly and evolution of homomeric protein complexes. *Phys. Rev. Lett.* 102: 118106
- Wächtershäuser G. 2003. From pre-cells to Eukarya--a tale of two lipids. *Mol. Microbiol.* 47: 13-22
- Wang J, Liu Y, Liu Y, Zheng S, Wang X, et al. 2019. Time-resolved protein activation by proximal decaging in living systems. *Nature* 569: 509-13
- Wheeler LC, Lim SA, Marqusee S, Harms MJ. 2016. The thermostability and specificity of ancient proteins. *Curr. Opin. Struct. Biol.* 38: 37-43
- Whitchurch CB, Tolker-Nielsen T, Ragas PC, Mattick JS. 2002. Extracellular DNA required for bacterial biofilm formation. *Science* 295: 1487
- Willdigg JR, Helmann JD. 2021. Mini Review: Bacterial Membrane Composition and Its Modulation in Response to Stress. *Front. Mol. Biosci.* 8
- Williams TJ, Allen M, Tschitschko B, Cavicchioli R. 2017. Glycerol metabolism of haloarchaea. *Environ. Microbiol.* 19: 864-77
- Woese CR, Gupta R. 1981. Are archaebacteria merely derived 'prokaryotes'? *Nature* 289: 95-6
- Woese CR, Kandler O, Wheelis ML. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U. S. A.* 87: 4576-9
- Woese CR, Magrum LJ, Fox GE. 1978. Archaeobacteria. *J Mol Evol* 11: 245-51
- Woodley JM. 2019. Accelerating the implementation of biocatalysis in industry. *Appl. Microbiol. Biotechnol.* 103: 4733-39
- Wörmer L, Lipp JS, Hinrichs KU. 2015. Comprehensive Analysis of Microbial Lipids in Environmental Samples Through HPLC-MS Protocols In *Hydrocarbon and lipid microbiology protocols*, pp. 289-317. Berlin, Heidelberg: Springer
- Wörmer L, Lipp JS, Schröder JM, Hinrichs KU. 2013. Application of two new LC-ESI-MS methods for improved detection of intact polar lipids (IPLs) in environmental samples. *Organic Geochemistry* 59: 10-21
- Xia Y, DiPrimio N, Keppel TR, Vo B, Fraser K, et al. 2013. The designability of protein switches by chemical rescue of structure: mechanisms of inactivation and reactivation. *J. Am. Chem. Soc.* 135: 18840-9

- Yang Y, Davis I, Matsui T, Rubalcava I, Liu A. 2019. Quaternary structure of alpha-amino-beta-carboxymuconate-semialdehyde decarboxylase (ACMSD) controls its activity. *J. Biol. Chem.* 294: 11609-21
- Yokoi T, Isobe K, Yoshimura T, Hemmi H. 2012. Archaeal Phospholipid Biosynthetic Pathway Reconstructed in *Escherichia coli*. *Archaea-an International Microbiological Journal* 2012
- Zeigler DR, Pragai Z, Rodriguez S, Chevreux B, Muffler A, et al. 2008. The origins of 168, W23, and other *Bacillus subtilis* legacy strains. *J. Bacteriol.* 190: 6983-95
- Zhang DL, Poulter CD. 1993. Biosynthesis of Archaeobacterial Lipids in Halobacterium-Halobium and Methanobacterium-Thermoautotrophicum. *J. Org. Chem.* 58: 3919-22
- Zheng L, Baumann U, Reymond JL. 2004. An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res.* 32: e115

## Acknowledgements

I would like to thank PD Dr. Patrick Babinger for supervising my doctoral thesis. Our collaboration was characterized by a good atmosphere and was ultimately extremely productive. I thank Prof. Dr. Reinhard Sterner that his door was always open for me. I gratefully acknowledge his awareness for creating a productive and trustful working atmosphere.

I would like to thank for the successful collaboration with Dr. Julius Lipp and Prof. Dr. Kai-Uwe Hinrichs from MARUM (Bremen). Dr. Julius Lipp and myself carried out the UHPLC-MS experiments described in Chapters 4. Together we quantified ether lipids in *B. subtilis* (at this point I would like to thank you for your help with Figure 4.1A) and discovered acetylated diether lipids in *H. volcanii* and *H. salinarum*. I thank you for the support, the helpful discussions and for the cordially reception. I felt very comfortable at your lab.

Many thanks to my colleagues. I had a really good time with you guys and the atmosphere was always great.

A “huge thank you” goes to our technical assistants Sonja Fuchs, Christiane Endres, Jeanette Ueckert and Sabine Laberer. I had a lot of fun with you guys and you gave me fantastic support in my projects. Sonja Fuchs supported me with full commitment, especially for the publication in Chapter 3.

“Thank you” to Carolin Apfel and Kinga Ay, who take care of RIGeL with great commitment. I had a great time with you and always enjoyed coming to your office. Together we have taken the excursion group to a new level.

Of course, I would also like to thank the students I supervised. Not only have I learned a lot through and with you, some have stayed with me as friends. Therefore “thanks” to Sebastian Pirner (Intern 2018, MA 2019), Anna Lena Schmidt (BA 2019, Intern 2021), Sebastian Schindler (BA 2020), Johannes Wilhelm (MA 2021), Zubeir El Ahmad (Intern 2021) und Maximilian Hoffmann-Becking (BA 2021).

I thank my friends for their friendship, which I hope will last for a long time. My thanks also go to my mother and Addi, who have always supported me in every way and have always been my biggest fans. Thank you!

Lastly, I would like to thank Zora for her love and support. You make me better!