Systematic Structural and Functional Characterization of the Family of Geranylgeranylglyceryl Phosphate Synthase-like Enzymes

Dissertation

Zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.) der Fakultät für Biologie und Vorklinische Medizin der Universität Regensburg

vorgelegt von
David Peterhoff
aus Erding
im Jahr 2014

Das Promotionsgesuch wurde eingereicht am:

14.01.2014

Die Arbeit wurde angeleitet von:

Prof. Dr. Reinhard Sterner

Unterschrift:

Index of contents

Abstract

A characteristic difference between the three phylogenetic domains of life is the chemical composition of the lipids forming their cell membranes. In Bacteria and Eukarya phospholipids are based on a *sn*-glycerol-3-phosphate (G3P) which is esterified with two fatty acids, while in phospholipids from Archaea the enantiomer *sn*-glycerol-1-phosphate (G1P) forms two ether-bonds with isoprenoid chains. As a result of these differences, the chemical, physical and biological properties of the archaeal membranes are drastically altered compared to bacterial and eukaryotic membranes. These changed conditions have been considered to be crucial for separation of archaea from a bacteria-like last universal common ancestor (LUCA). Consequently, the two enzymes that provide stereo-specificity in the synthesis of archaeal phospholipids have been regarded as key enzymes in the early evolution of Archaea and have been assumed to be strictly archaea-specific. These are the G1P-providing enzyme, glycerol-1-phosphate dehydrogenase (G1PDH), and the enzyme transferring an activated polyprenyl substrate (mostly geranylgeranyl diphosphate, GGPP) stereo-specifically to G1P, geranylgeranylglyceryl phosphate synthase (GGGPS). In contradiction to the presumed restriction of GGGPS occurrence to the domain of Archaea, genome sequencing data recently revealed a significant distribution of these enzymes among Bacteria. A phylogenetic analysis of GGGPS-like enzymes showed their partition into two distinct groups (group I and II), both comprising archaeal and bacterial sequences. Group I contains bacterial sequences from Gram-positive Firmicutes and group II from Gram-negative Bacteroidetes. While several bacterial and archaeal group I enzymes have been subjects of recent extensive kinetic and structural analyses, only a few archaeal, but no bacterial group II enzymes have been characterized up to now.

The aim of this work was a systematic comparative characterization of representative members from all important subgroups of the family of GGGPS-like enzymes, with a strong focus on the functional assignment of the Bacteroidetes subgroup and the structure-function-relationship in group II enzymes. Therefore, the variability within the family was analyzed in detail by calculating a sequence similarity network, and 17 representative GGGPS-like enzymes were characterized biochemically regarding their catalytic activities and substrate specificities. Three complementary assays demonstrated that all tested archaeal enzymes preferred short chained polyprenyl substrates, certainly being GGPP in most species. The bacterial enzymes from group II showed a preference for short chained polyprenyl substrates comparable to the archaeal enzymes, but the physiological role of the ether lipid product in those species remains an enigma. In contrast, bacterial enzymes from group I preferred long

chained polyprenyl substrates, and a previous study has showed that the *Bacillus subtilis* enzyme produces heptaprenylglyceryl phosphate, which becomes dephosphorylated and acetylated *in vivo*.

Moreover, the first crystal structures of group II archaeal and bacterial enzymes have been solved within the framework of this thesis. These new structures expanded the knowledge on the mechanism of substrate specificity. In all GGGPS-like enzymes, the length of the polyprenylic substrate is measured by a "hydrocarbon ruler" that uses different residues to act as limiters in the binding pocket to provide the different substrate specificities. Interestingly, these limiter residues also differ between the GGPP-specific group I and group II enzymes.

Ongoing ambiguities on the native oligomerization interface of the group I enzyme from *B. subtilis,* PcrB, were resolved by means of a combined approach including computational, biophysical and biochemical methods. Mutagenesis-mediated monomerization identified the native interface among two computationally predicted alternatives. The result was verified by site-specific incorporation of an unnatural amino acid, which allowed for cross-linking when located in the native interface. Interestingly, monomerization did not affect thermal stability of PcrB, but limited the length of accepted polyprenyl diphosphates to three isoprene units in an *in vitro* assay, whereas the native substrate contains seven isoprene entities. A plausible hypothesis how dimerization determines substrate specificity of PcrB by affecting its secondary structure could be provided.

While these experiments confirmed that all group I enzymes and a number of group II enzymes share the same dimeric architecture, it was surprising that a significant portion of archaeal and bacterial group II enzymes revealed to be hexamers, as shown by crystallization and static light scattering experiments. In a rational mutagenesis study, which was aided by the novel structures, an aromatic "anchor" residue was found to be an inevitable prerequisite for hexamerization. By substituting this aromatic residue, the hexamers could be disrupted to dimers in the prototypical conformation. Because these dimeric mutants showed catalytic activities and thermal stabilities comparable to the hexameric wild types, the functional reason for hexamerization remains unclear.

In summary, the presented systematic comparative characterization leads to a deepened understanding of the mechanisms of substrate specificity and stereo-selectivity and a detailed description of the enzyme architecture in the two subgroups of the family of GGGPS-like enzymes.

Kurzfassung der Arbeit

Ein charakteristisches Unterscheidungsmerkmal zwischen den drei Domänen des Lebens ist die Zusammensetzung ihrer Zellmembranen. Die Phospholipide von Bakterien und Eukaryoten basieren auf *sn*-Glycerin-3-Phosphat (G3P), welches mit zwei Fettsäuren verestert ist, während in den Phospholipiden von Archaeen das Enantiomer *sn*-Glycerin-1-Phosphat (G1P) zwei Etherbindungen mit Isoprenolen ausbildet. Folglich unterscheiden sich die chemischen, physikalischen und biologischen Eigenschaften archaeeller Membranen drastisch zu bakteriellen und eukaryotischen Membranen. Man nimmt an, dass diese veränderten Bedingungen entscheidend für die Abspaltung der Archaeen von einem bakterienartigen, letzten gemeinsamen Vorfahren (*last universal common ancestor, LUCA*) waren. Folglich werden die beiden Enzyme, die Stereospezifität in der Synthese der archaeellen Phospholipide vermitteln, als Schlüssel-Enzyme in der frühen Evolution der Archaeen betrachtet und wurden bisher als strikt Archaeen-spezifisch angesehen. Diese sind das G1P-bereitstellende Enzym, Glycerin-1-Phosphat-Dehydrogenase (G1PDH), und die Geranylgeranylglycerylphosphat Synthase (GGGPS), welche ein aktiviertes Polyprenyl-Substrat (meistens Geranylgeranylpyrophosphat, GGPP) stereospezifisch auf G1P überträgt. Entgegen der allgemein angenommenen Beschränkung des Vorkommens der GGGPS auf die Domäne der Archaeen haben Genomsequenzierungsdaten in der letzten Zeit eine signifikante Verbreitung dieser Enzyme unter Bakterien gezeigt. Eine phylogenetische Analyse von GGGPS-artigen Enzymen zeigte ihre Spaltung in zwei distinkte Gruppen (Gruppe I und II), die beide archaeelle und bakterielle Sequenzen umfassen. Gruppe I enthält bakterielle Sequenzen von Gram-positiven Firmicutes und Gruppe II enthält Sequenzen von Gram-negativen Bacteroidetes. Während in jüngster Zeit verschiedene bakterielle und archaeelle Enzyme aus der Gruppe I Gegenstand eingehender kinetischer und struktureller Analysen waren, wurden aus der Gruppe II bisher nur wenige archaeelle und keine bakteriellen Enzyme charakterisiert.
Ziel dieser Arbeit war eine systematische, vergleichende Charakterisierung von repräsentativen Mitgliedern aller wichtigen Untergruppen der Familie der GGGPS-artigen Enzyme, mit einem starken Fokus auf die Funktionsaufklärung der Bacteroidetes-Untergruppe sowie auf die Struktur-Funktions-Beziehungen in Gruppe II-Enzymen. Hierzu wurde die Variabilität in der Familie durch Berechnung eines Sequenzähnlichkeits-Netzwerkes detailliert analysiert, und 17 repräsentative GGGPS-artige Enzyme wurden biochemisch hinsichtlich ihrer katalytischen Aktivität und Substratspezifität charakterisiert. Drei komplementäre Tests zeigten, dass alle archaeellen Enzyme kurzkettige Polyprenyl-Pyrophosphate als Substrat bevorzugten, welches in den meisten Spezies sehr wahrscheinlich

GGPP ist. Die bakteriellen Enzyme aus Gruppe II zeigten ebenso eine Präferenz gegenüber kurzkettigen Polyprenylsubstraten, vergleichbar zu den archaeellen Enzymen. Die physiologische Rolle der Etherlipidprodukte in diesen Spezies bleibt unbekannt. Bakterielle Enzyme der Gruppe I bevorzugten dagegen langkettige Polyprenylsubstrate, und eine vorangehende Studie konnte zeigen, dass das Enzym aus *Bacillus subtilis* Heptaprenylglycerylphosphat produziert, welches *in vivo* dephosphoryliert und acetyliert wird.

Weiterhin wurden im Rahmen dieser Arbeit die ersten Kristallstrukturen von Gruppe II-Enzymen gelöst. Diese neuen Strukturen erweiterten das Wissen über den Mechanismus der Substratspezifität. In allen GGGPS-artigen Enzymen wird die Länge des Polyprenylsubstrates durch einen *„hydrocarbon ruler“* („Kohlenwasserstoff-Lineal“) gemessen, welches verschiedene Residuen in der Bindetasche als *„limiter“* (Begrenzer) nutzt, um die unterschiedlichen Substratspezifitäten zu ermöglichen. Interessanterweise unterscheiden sich diese Residuen auch zwischen GGPP-spezifischen Enzymen der Gruppe I und II.

Mithilfe eines kombinierten Ansatzes aus bioinformatischen, biophysikalischen und biochemischen Methoden konnten andauernde Unklarheiten bezüglich der nativen Oligomerisierungs-Kontaktfläche des Gruppe I Enzyms aus *B. subtilis*, PcrB, geklärt werden. Mittels mutagenesevermittelter Monomerisierung konnte die native Kontaktfläche unter zwei bioinformatisch vorhergesagten Alternativen identifiziert werden. Das Ergebnis wurde durch ortsspezifischen Einbau einer nichtkanonischen Aminosäure verifiziert, welche bei Lokalisierung in der nativen Kontaktfläche ein Quervernetzen der Untereinheiten erlaubte. Interessanterweise beeinflusste die Monomerisierung nicht die thermische Stabilität von PcrB, aber limitierte in einem *in vitro* Test die Länge der akzeptierten Polyprenylpyrophosphate auf drei Isopreneinheiten, während das native Substrat sieben Isopreneinheiten enthält. Eine plausible Hypothese konnte aufgestellt werden, wie die Dimerisierung die Substratspezifität von PcrB durch Beeinflussung von dessen Sekundärstruktur bedingt.

Während diese Experimente bestätigten, dass alle Enzyme der Gruppe I und einige der Gruppe II eine gemeinsame Dimer-Architektur aufweisen, war es überraschend, dass eine erhebliche Anzahl an archaeellen und bakteriellen Gruppe II Enzymen Hexamere bilden, wie durch Kristallisations- und Lichtstreuungsexperimente gezeigt werden konnte. In einer durch die neuen Strukturen unterstützten Mutagenesestudie konnte eine aromatische „Anker“-Residue als unabdingbare Voraussetzung für die Hexamerisierung identifiziert werden. Durch Austausch dieser aromatischen Residue konnten die Hexamere in Dimere mit prototypischer Konfiguration getrennt werden. Da diese Dimer-Mutanten zum Wildtyp vergleichbare

katalytische Aktivitäten und thermische Stabilitäten zeigten, bleibt der funktionelle Grund der Hexamerisierung vorerst unklar.

Zusammenfassend führt die vorgelegte systematische Charakterisierung zu einem vertieften Verständnis der Mechanismen der Substratspezifität und der Stereoselektivität sowie zu einer detaillierten Beschreibung der Enzym-Architektur in den zwei Untergruppen der Familie der GGGPS-artigen Enzyme.

1 Synopsis

1.1 Homology guided functional assignment of proteins

The challenging legacy of the genomics era is a hardly manageable mass of sequence data. As a result of large-scale genome sequencing projects, a huge number of putative genes is deposited in databases, while only few of their products have been characterized yet. In fact, the UniProtKB/TrEMBL protein database statistics (Release 2013_09 of 18-Sep-2013, 42.821.879 entries) show, that only for 0.05 % of its entries the existence of the protein has been proven experimentally. For further 1.91 % of the entries at least expression data (e.g. cDNA(s), RT-PCR or Northern blots) indicates the existence of the protein, and 23 % of the deposited proteins are inferred by homology, meaning that their existence is likely because orthologs exist in related species. The remaining 75 % of the entries are pure predictions, i.e. no evidence at protein, transcript, or homology level exists. The systematic experimental validation of these predictions is one of the biggest challenges of the post genomic era.

To annotate the function of uncharacterized proteins, predictions based on sequence similarity can be helpful. Although such predictions often turned out to be incorrect due to divergent evolution (Gerlt et al. 2000; Friedberg et al. 2006; Schnoes et al. 2009), they are on the other hand reasonable and advantageous when only taken as a rough estimation of the function. Besides this, the quality of the predictions has strongly improved due to novel prediction strategies involving interconnected knowledge (Loewenstein et al. 2009, Radivojac et al. 2013) – a principle which gene ontology initiatives are currently streamlining (Ashburner et al. 2000; Berardini et al. 2010; Gaudet et al. 2011). Hence, a careful homology guided functional assignment approach is a feasible strategy to unravel novel protein functions.

Against this background and due to our special interest in proteins encountering the $(\beta\alpha)_8$-fold, we started to investigate the group of *sn*-glycerol-1-phosphate geranylgeranyltransferase-like (GGGPS-like) enzymes. In an integrative approach combining biochemical and computational methods we opted for a systematic and comparative characterization of this interesting group of enzymes, including the assignment of function to so far uncharacterized members.

1.2 The family of GGGPS-like enzymes

1.2.1 The GGGPS is a key enzyme in the early evolution of Archaea

Membrane lipids define the domains of life

A characteristic difference between the three superkingdoms of life (Bacteria, Archaea and Eukarya) is the chemical composition of the lipids forming their membranes (Kates 1993; Koga et al. 1993; Wächtershäuser 2003; Matsumi et al. 2011). Among the large variety of membrane lipids, the most important lipid chemistry in all domains of life is the class of phospholipids (Cronan 2003; van Meer et al. 2008; Albers et al. 2011). Although exhibiting the same architecture of a polar head group and a long hydrophobic tail, major differences appear when comparing the phospholipids of Bacteria and Eukarya to archaeal phospholipids. In Bacteria and Eukarya, phospholipids are based on a *sn*-glycerol-3-phosphate (G3P) which is esterified with two fatty acids, while in phospholipids from Archaea the enantiomer *sn*-glycerol-1-phosphate (G1P) forms two ether-bonds with isoprenoid chains (Koga et al. 2007; Ulrih et al. 2009; figure 1).

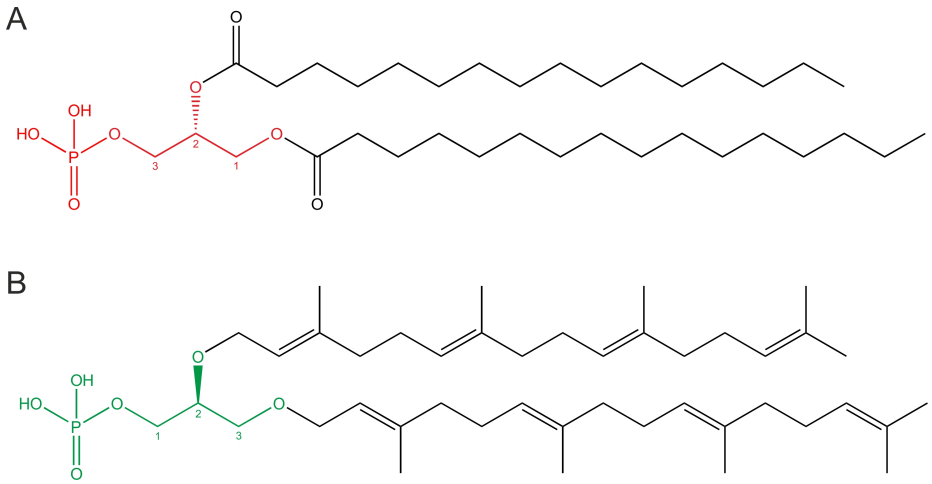


Figure 1: Bacterial and archaeal core phospholipids.
In Bacteria and Eukarya, G3P (red) is bound to fatty acids by ester linkages (A), while in Archaea, G1P (green) is bound to polyprenyl derivatives by ether linkages (B).

A large number of diverse modifications of these core phospholipids exist, and their specific combination in the membrane is characteristic for every organism (De Rosa et al. 1988; Albers et al. 2011). This circumstance guided attempts to phylogenetically classify Archaea based on their membrane lipids (Koga et al. 2008).

In the focus of evolution: key enzymes for the synthesis of membrane lipids

Huge efforts have been made, to elucidate the characteristics and evolution of a primitive ancestor organism (last universal common ancestor, LUCA) at the root of the domains of life, and the mechanisms underlying the separation of the three superkingdoms (Woese 1998; Glansdorff 2000; Glansdorff et al. 2008). A recent model on the evolutionary process of the separation of the Archaea from a bacterial-like LUCA proposes, that changes of the lipid composition of the cell membranes had a strong impact on this speciation event (Payandeh et al. 2007; Lombard et al. 2012; Lombard et al. 2012). On one hand the strongly altered chemical properties of the archaeal membranes are considered to act as a genetic barrier, while on the other hand they might have allowed for the colonization of extreme niches due to their high thermal and chemical stability as well as their impermeability (van de Vossenberg et al. 1998; Ulrih et al. 2009), thereby separating early Archaea from the bacterial genetic pool (Payandeh et al. 2007).

In this context, two enzymes involved in the synthesis of archaeal phospholipids have been widely discussed to act as key enzyme in the early evolution of Archaea (Payandeh et al. 2007; Lombard et al. 2012; Lombard et al. 2012). These are a G1P-providing enzyme and an enzyme transferring an activated polyprenyl substrate stereo-specifically to G1P. In accordance with this hypothesis, all Archaea have a glycerol-1-phosphate dehydrogenase (G1PDH, Interpro family IPR023002), which catalyzes the NADH-dependent reduction of dihydroxyacetone phosphate (DHAP) to G1P, and a *sn*-glycerol-1-phosphate geranylgeranyltransferase, commonly called geranylgeranylglyceryl phosphate synthase (GGGPS), which transfers the polyprenyl moiety of geranylgeranyl diphosphate (GGPP, consisting of 20 C-atoms) stereo-specifically to the C3 hydroxyl group of G1P. Other than GGGPS, the (*S*)-2,3-di-*O*-geranylgeranylglyceryl phosphate synthase (DGGGPS) transfers the second polyprenyl moiety to (*S*)-3-*O*-geranylgeranylglyceryl phosphate (GGGP) without stereo-selectivity. Hence, the chirality of the archaeal membrane is determined by the stereo-specific GGGPS reaction, which consequently has been considered to be the committed step in the synthesis of archaeal membrane lipids (Payandeh et al. 2007). Figure 2 gives an overview of the synthesis pathways of bacterial and archaeal core phospholipids.

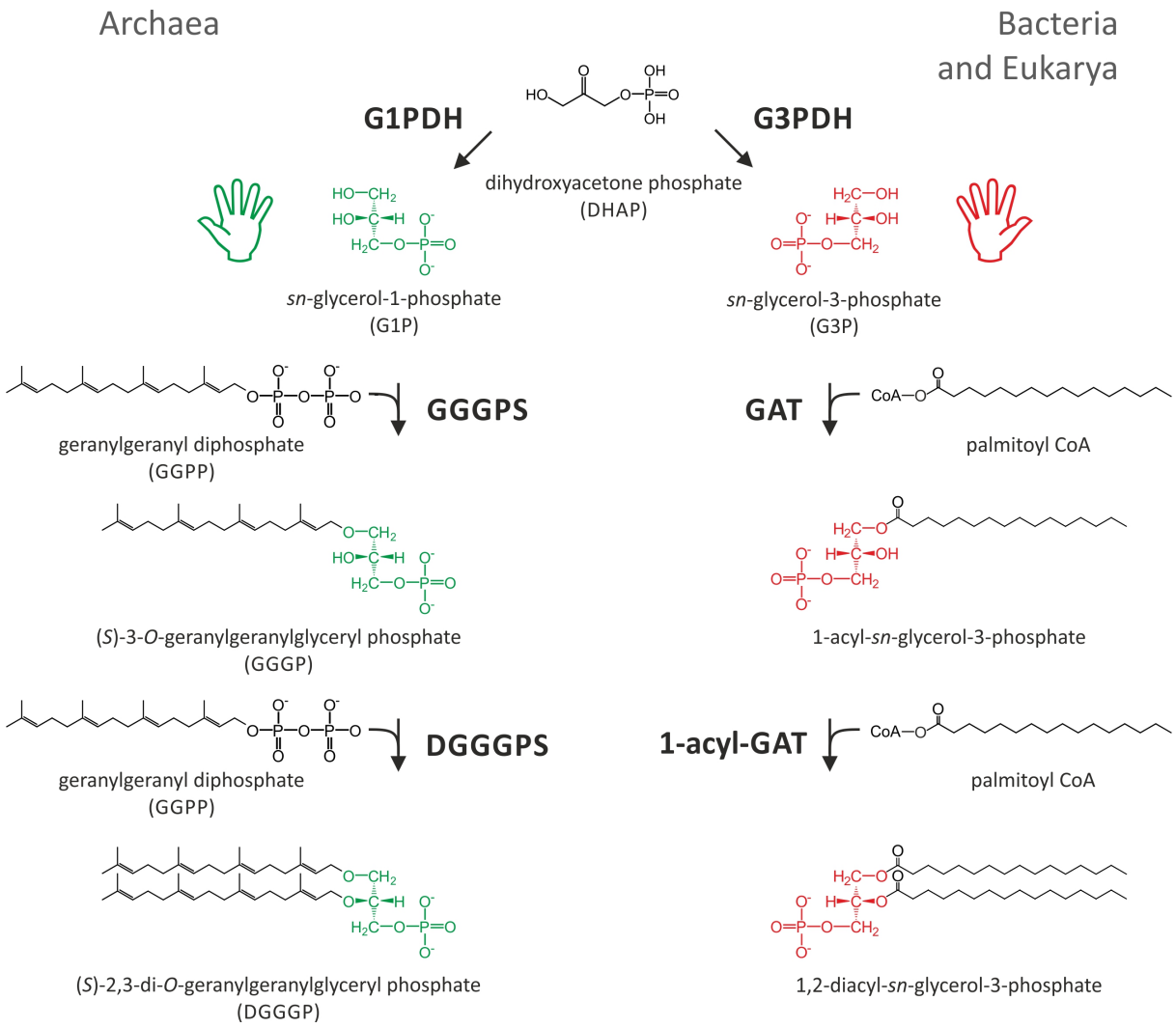


Figure 2: Synthesis pathway of bacterial and archaeal core phospholipids.
G1P (green) in Archaea is provided by a G1PDH, G3P (red) in Bacteria and Eukarya is provided by a glycerol-3-phosphate dehydrogenase (G3PDH). Both enzymes exhibit no structural homology. In Archaea, GGGPS and DGGGPS subsequently transfer the polyprenyl moiety of GGPP to G1P. In Bacteria and Eukarya, a G3P acyltransferase (GAT) and a 1-acyl-G3P acyltransferase (1-acyl-GAT) subsequently transfer the acyl moiety of acyl Coenzyme A (here palmitoyl CoA) to G3P.

Accordingly it has been proposed, that the key enzymes of the synthesis pathway of archaeal membrane compounds, G1PDH and GGGPS, are restricted to the domain of Archaea (Koga et al. 1998; Boucher et al. 2004; Hemmi et al. 2004; Pereto et al. 2004; Payandeh et al. 2007). In contrast to this hypothesis, phylogenetic analyses revealed that the genomes of some Firmicutes and Bacteroidetes encode GGGPS-like proteins (Boucher et al. 2004; Doud et al. 2011; Lombard et al. 2012, for details see 1.2.3). It has recently been shown for Firmicutes that they produce heptaprenylglyceryl phosphate (HepGP) instead of GGGP, which subsequently becomes dephosphorylated and acetylated (Guldan et al. 2011). These GGGPS-like variants from Firmicutes bear for historical reasons the name PcrB (the well characterized

PcrA protein is a DNA helicase, encoded in the same operon, Petit et al. 1998). Whether the occurrence of these enzymes in Bacteria is a consequence of one or more lateral gene transfer events (LGT, Boucher et al. 2003) or a vestigial feature of the early evolution of Archaea from the LUCA, remains to be clarified. However, the existence of ether lipids in such species (Guldan et al. 2011), as well as the presence of a specific product processing phosphorylase and acetyltransferase (Linde 2013), indicate their physiological relevance. Furthermore, it was shown that a bacterial G1PDH (termed AraM) exists in *Bacillus subtilis* (Guldan et al. 2008).

1.2.2 On the reaction catalyzed by GGGPS-like enzymes

A stereo-specific *O*-alkylation reaction

GGGPS-like enzymes catalyze the formation of an ether bond between *sn*-glycerol-1-phosphate and a polyprenylic compound with a varying number of isoprene units (Figure 3) (Chen et al. 1993; Zhang et al. 1993; Soderberg et al. 2001; Nemoto et al. 2003; Payandeh et al. 2006; Guldan et al. 2011). Stereo-specificity of the reaction is enabled by the stereo-specific binding of G1P by the enzyme.

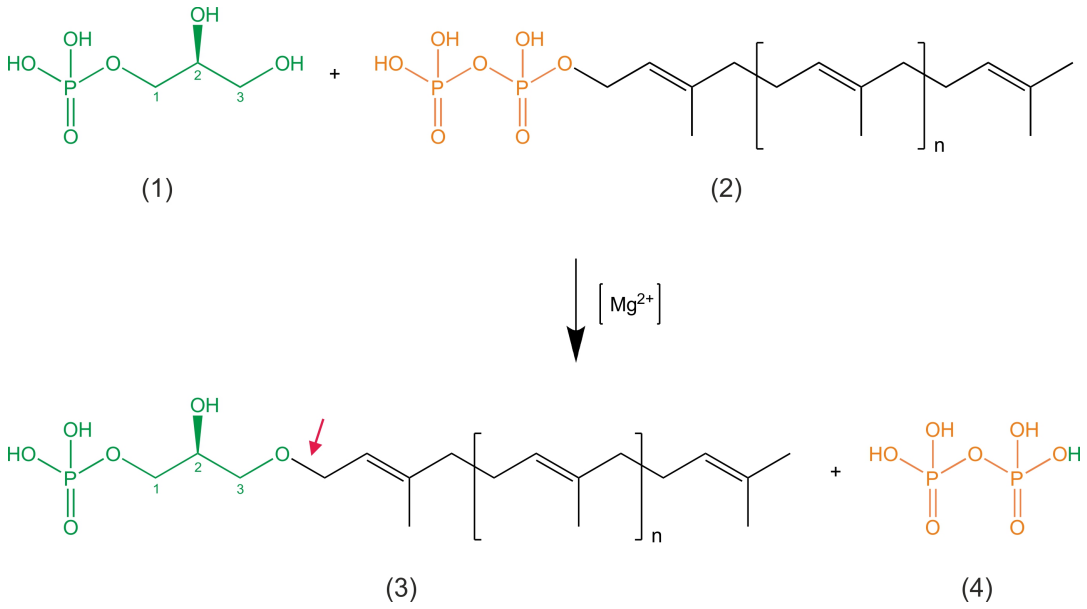


Figure 3: The reaction catalyzed by the family of GGGPS-like enzymes.
GGGPS-like enzymes catalyze the transfer of a polyprenyl diphosphate (2) with a varying number (n+2) of isoprene units to the C3-oxygen of G1P (1). The reaction leads to the formation of an ether bond (red arrow) in the polyprenylglyceryl phosphate product (3), and is energetically driven by the release of diphosphate (4). A magnesium ion is an essential cofactor of the reaction.

In Archaea, where the polyprenyl substrate consists of 4 isoprene units (n=2, GGPP) in most cases, the ether product serves as a precursor of various archaeal cell membrane compounds (Matsumi et al. 2011). In this case the enzyme is called GGGPS (EC 2.5.1.41), being an eponym of the whole family of enzymes. Bacterial GGGPS-like variants from Firmicutes prefer a longer substrate (Guldan et al. 2011) which consists of seven isoprene units (n=5, heptaprenyl diphosphate, HepPP). Consequently these enzymes, which have initially been denominated as "PcrB", are termed heptaprenylglyceryl phosphate synthases (HepGPS, EC 2.5.1.-).

An essential cofactor of the reaction is Mg^{2+}, which binds in complex with the diphosphate moiety of the prenyl substrate (Chen et al. 1993; Soderberg et al. 2001; Nemoto et al. 2003; Payandeh et al. 2006; Ren et al. 2013). In fact, many prenyltransferases depend on divalent cations, including farnesyl diphosphate synthase, undecaprenyl diphosphate synthase and squalene synthase (Christianson 2006).

Two alternative reaction mechanisms are considered for GGGPS-like enzymes

Two alternative catalytic mechanisms have been discussed: electrophilic alkylation and nucleophilic substitution (Chen et al. 1993; Soderberg et al. 2001; Nemoto et al. 2003; Payandeh et al. 2006; Ren et al. 2012; Ren et al. 2013). While during an electrophilic alkylation a highly electrophilic allylic carbocation is formed (geranylgeranyl cation intermediate) which subsequently alkylates G1P (figure 4 A), in a nucleophilic substitution the activated (deprotonated) C3-hydroxyl group oxygen of G1P attacks the C1-atom of the polyprenyl diphosphate and the diphosphate acts as leaving group (figure 4 B). For the GGGPS-like family electrophilic alkylation has been suggested first by Poulter and co-workers (Zhang et al. 1993). To address this problem experimentally, they used three analogous substrates (figure 4 C), of which only two could perform an electrophilic alkylation.

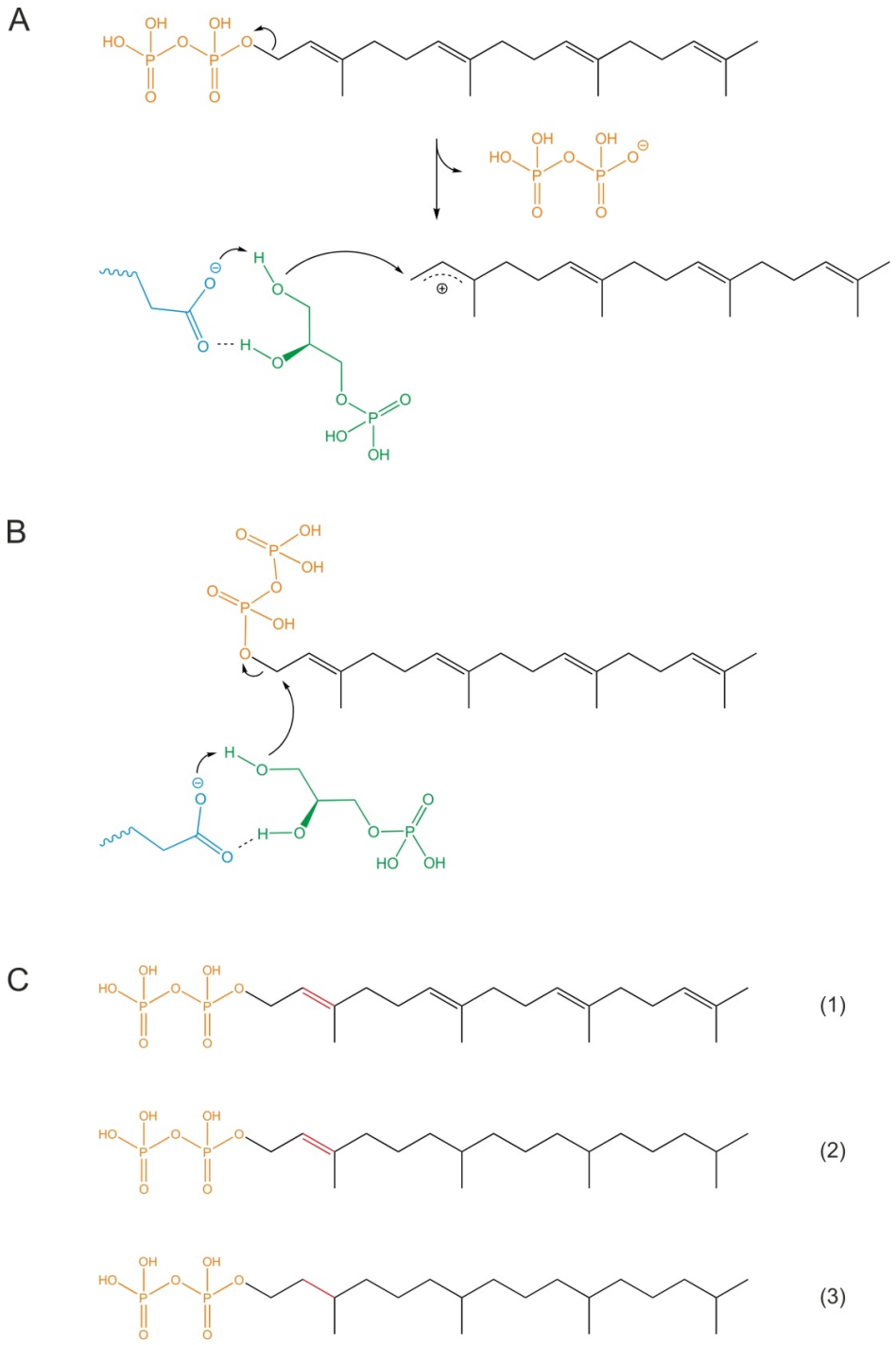


Figure 4: Two possible reaction mechanisms of GGGPS-like enzymes, and substrate analoga to discriminate between them.
A) Electrophilic alkylation mechanism. The diphosphate moiety (orange) leaves the substrate, forming the highly electrophilic allylic carbocation, which subsequently reacts with the C3-oxygen of G1P (green). A catalytic glutamate carboxyl group from the GGGPS (blue, compare 1.2.6) accepts the released proton. B) Nucleophilic substitution mechanism. The catalytic glutamate deprotonates the C3 hydroxyl group. The latter attacks the C1 carbon atom of the polyprenyl substrate which subsequently releases the diphosphate leaving group. C) The three different substrates to discriminate between the proposed alternative mechanisms: GGPP (1), phytyl diphosphate (2) and phytanyl diphosphate (3). The position, where a double bond provides electron density to stabilize the carbocation intermediate of an electrophilic alkylation is given in red.

The native substrate GGPP (figure 4 C (1)), and phytyl diphosphate (figure 4 C (2)) bear an allylic double bond (between the C2 and C3 carbon atom) thereby being able to stabilize the positively charged intermediate during an electrophilic alkylation, while the completely reduced phytanyl diphosphate (figure 4 C (3)) ought to form a highly unstable carbocation during an electrophilic alkylation. In contrast, all three substrates are equally likely to undergo a nucleophilic substitution. The GGGPS (which was provided in a cell free preparation of *Methanothermobacter thermautotrophicus*) only showed activity with the two substrates which are capable of electrophilic alkylation. While radiolabeled phytyl diphosphate was incorporated at 18 % of the maximal rate (GGPP rate), phytanyl diphosphate gave no signal for turnover.

The presumption of an electrophilic alkylation is consistent with the proposals for other prenyltransferases (Poulter et al. 1978; Liang et al. 2002), but nucleophilic substitution reactions have also been described for protein prenyl transferases (Long et al. 2002).

1.2.3 Phylogenetic background of the family of GGGPS-like enzymes

Ambiguities in denomination

The group of enzymes transferring a polyprenyl moiety to G1P is not clearly defined and precisely termed and therefore not represented consistently in the current databases of protein families. The manually curated Pfam database (Punta et al. 2012) subsumes 644 sequences (July 2013, release 27) from Bacteria and Archaea to the “PcrB family” (PF01884). The Interpro database (Hunter et al. 2012) matches 1205 sequences (July 2013, release 43.1) form Bacteria and Archaea to a “Geranylgeranylglyceryl phosphate synthase/Heptaprenylglyceryl phosphate synthase” family (IPR008205) which divides into multiple inconsistent subfamilies (IPR010946, IPR026417, IPR026438). One reason for these ambiguities is the emergence of hundreds of bacterial sequences over the last decade, which undermined the conjecture of consistent substrate specificities among the family members. Due to these ambiguities we use the term “GGGPS-like” enzymes to address all members of the group. An alternative more general and unbiased denomination, legitimate for all members of the family, would be “glyceryl-1-phosphate polyprenyl transferase”.

Phylogenetic relationships in the family of GGGPS-like enzymes

A phylogenetic analysis of the amino acid sequences of GGGPS-like enzymes reveals a partition into two distinct groups (figure 5).

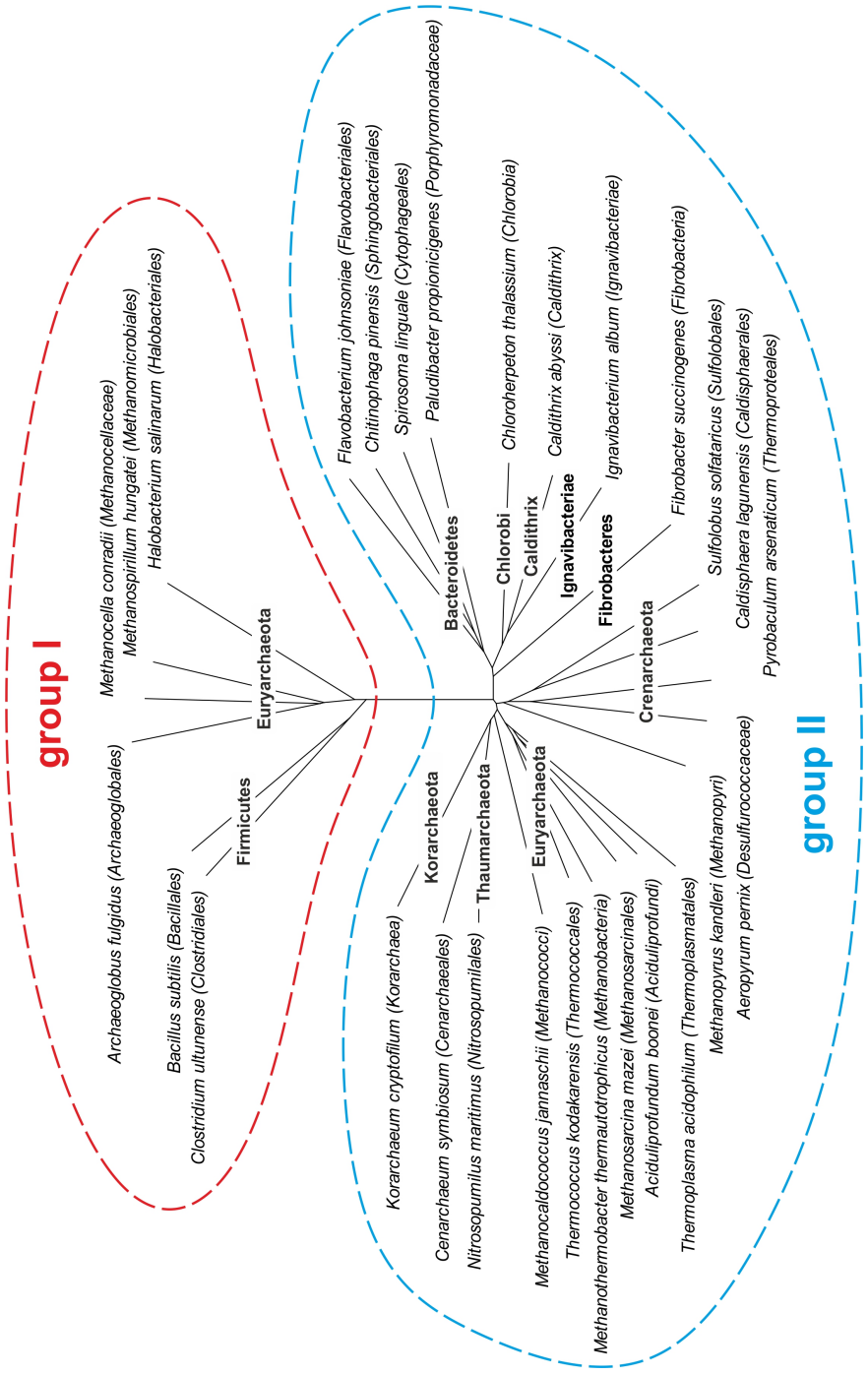


Figure 5: Phylogenetic tree of the family of GGGPS-like enzymes.
Taxa possessing GGGPS-like enzymes were identified using BLAST (Altschul et al. 1990). Representative sequences were selected for every taxonomic order, a multiple sequence alignment was calculated using Clustal Omega (Sievers et al. 2011), and a phylogenetic tree was generated by means of SplitsTree (Huson et al. 2006). Note, that the tree does not reflect the frequency of occurrence among the members of a shown taxon.

Group I includes sequences of species from the phylum Firmicutes (mostly Gram-positive *Bacillales*, group Ib) and some Euryarchaeota (group Ia), while group II assembles sequences from Gram-negative Bacteroidetes, Chlorobi, Caldithrix, Ignavibacteriae, Fibrobacteres (group IIb) and from all archaeal phyla except Nanoachaeota (group IIa), which is consistent with the fact that *Nanoarchaeum equitans* obtains its membrane lipids from its host *Ignicoccus hospitalis* (Jahn et al. 2004). The sequence identities between group I and group II enzymes are less than 20 %, while members belonging to the same group share sequence identities of more than 40 %. While for Firmicutes and Bacteroidetes a broad distribution of sequences among their numerous members can be found, Chlorobi and Ignavibacteriae are only represented by single sequences, although these phyla contain multiple sequenced species. For the phyla Caldithrix and Fibrobacteres only a single genome sequence is available. This suggests that the GGGPS variants of these representatives were obtained recently via LGT (probably from the Bacteroidetes), while Bacteroidetes and Firmicutes received their variants in an earlier stage of evolution from the Archaea (or *vice versa*). A sequence similarity network, which was calculated within the framework of this thesis, illustrates these relationships and reflects the number of sequenced genomes per order (figure 6).

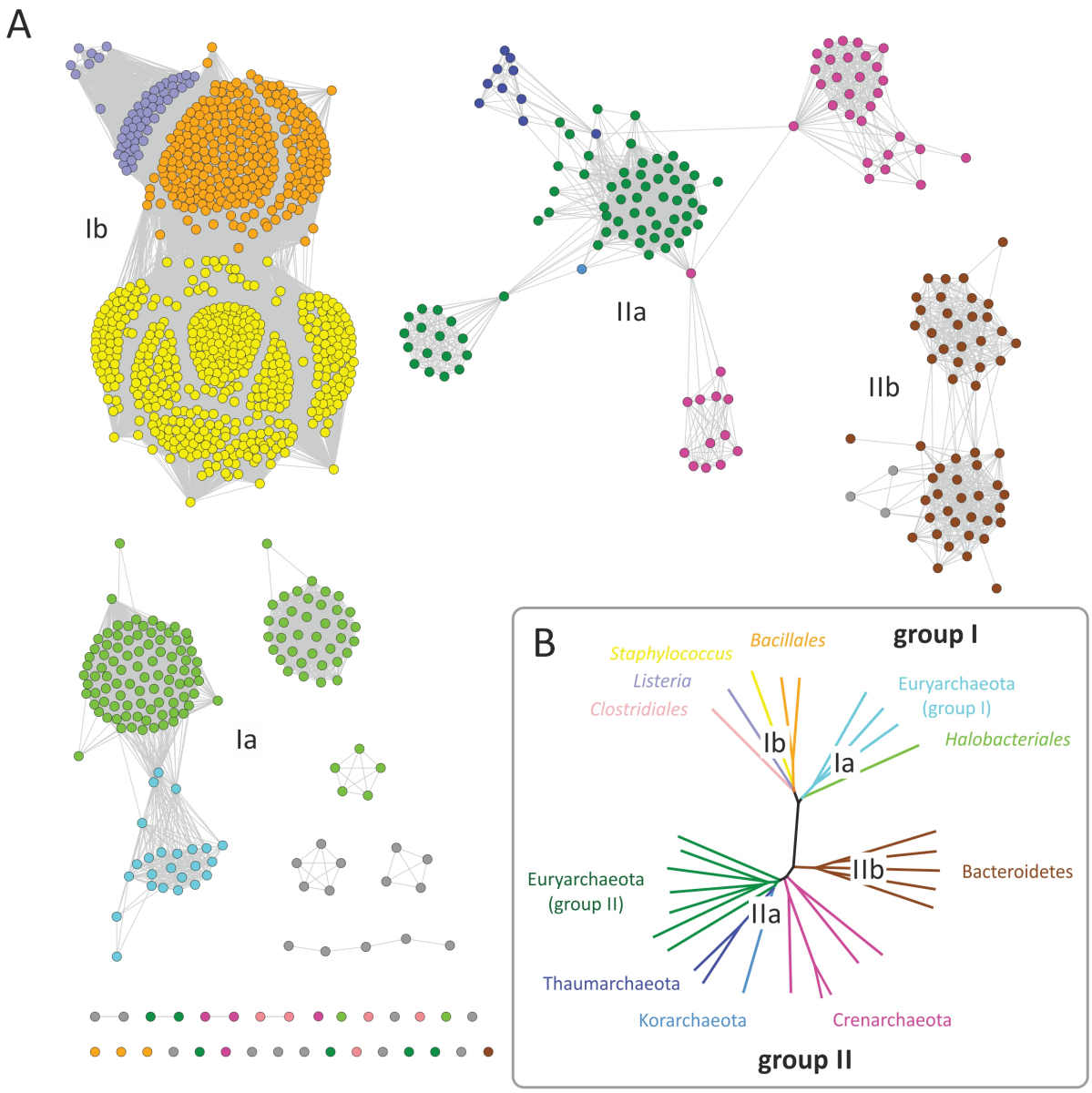


Figure 6: Sequence similarity networks of members of the GGGPS-family.
A) A sequence similarity map, containing 1205 sequences (IPR008205 from InterPro, release 43.1, 7/2013) comprising all currently known GGGPS-like protein sequences, was calculated using the methods developed by Atkinson et al. 2009. A network with an E-value cut-off of 10^{-60} was visualized using the organic layout in Cytoscape 2.8.3 (Smoot et al. 2011). The coloring of the nodes corresponds to B) where the associated phyla and orders are shown (miscellaneous rare occurrences in grey, not shown in B)). B) Phylogenetic tree of representative sequences from the main phyla possessing GGGPS-like enzymes.

In principle, in a sequence similarity network, nodes represent sequences, and edges represent the BLAST E-values from pairwise local alignments. Edges are drawn when a specified BLAST e-value cutoff is outvalued. This gives a two dimensional representation of clustering similarity patterns (Barabasi et al. 2004; Atkinson et al. 2009). Sequence similarity networks have been shown to correlate well with phylogenetic trees (Kalyanaraman et al. 2008; Lukk et al. 2012).

The network of the GGGPS-like enzymes, which contains the 1205 sequences of the IPR008205 family from the Interpro database, reflects the common archaeal taxonomy as well as the phylogeny as deduced from ribosomal protein sequences (Brochier-Armanet et al. 2011; Olsen et al. 1993), and the already mentioned splitting of the family into four subgroups (groups Ia, Ib, IIa, IIb; compare publication B, figure 1). Due to imbalances in the number of sequenced genomes, particularly the group Ib sequences are strongly overrepresented.
Due to a bug in the scripts used for calculating the network (Biocluster, University of Illinois at Urbana-Champaign, http://biocluster.igb.illinois.edu/, as of Sept. 2013), the BLAST hits per sequence, and therefore the number of edges per node, was limited to 250. Tests in Dec. 2013 with a revised script showed that this only affected cluster Ib, which now was much more compact (data not shown). Because all other clusters were completely unaffected, but the limit to 250 hits significantly improved the resolution of subdivisions within cluster Ib, the original version of the network was used within this work.

Sequence similarity networks reveal phylogenetic subgroups with different traits

Interestingly, the clustering in the network visualizes differences in the family, which are not well resolved in a phylogenetic tree. Some specific sequence features lead to a further splitting into subgroups in some orders, which is illustrated in figure 7. First, the Halobacteriales divide in two completely separated groups (IaH1, IaH2), one of the groups being connected to the rest of the sequences of group I Euryarchaeota (IaE). Second, the sequences of the Euryarchaeota from group II split into two subgroups (IIaE1, IIaE2) of which one is connected to the sequences of Crenarchaeota (IIaC1, IIaC2) and Thaumarchaeota (IIaT). The sequence from *Korarchaeum cryptofilum* is located between the two group II Euryarchaeota subgroups. Third the sequences of Crenarchaeota split into two subgroups (IIaC1, IIaC2), which are partially connected by the sequences from group II Euryarchaeota. Fourth, the sequences from Bacteroidetes split into two subgroups (IIb1, IIb2), one of them being connected to the three bacterial sequences from other taxa in group II (IIb3).

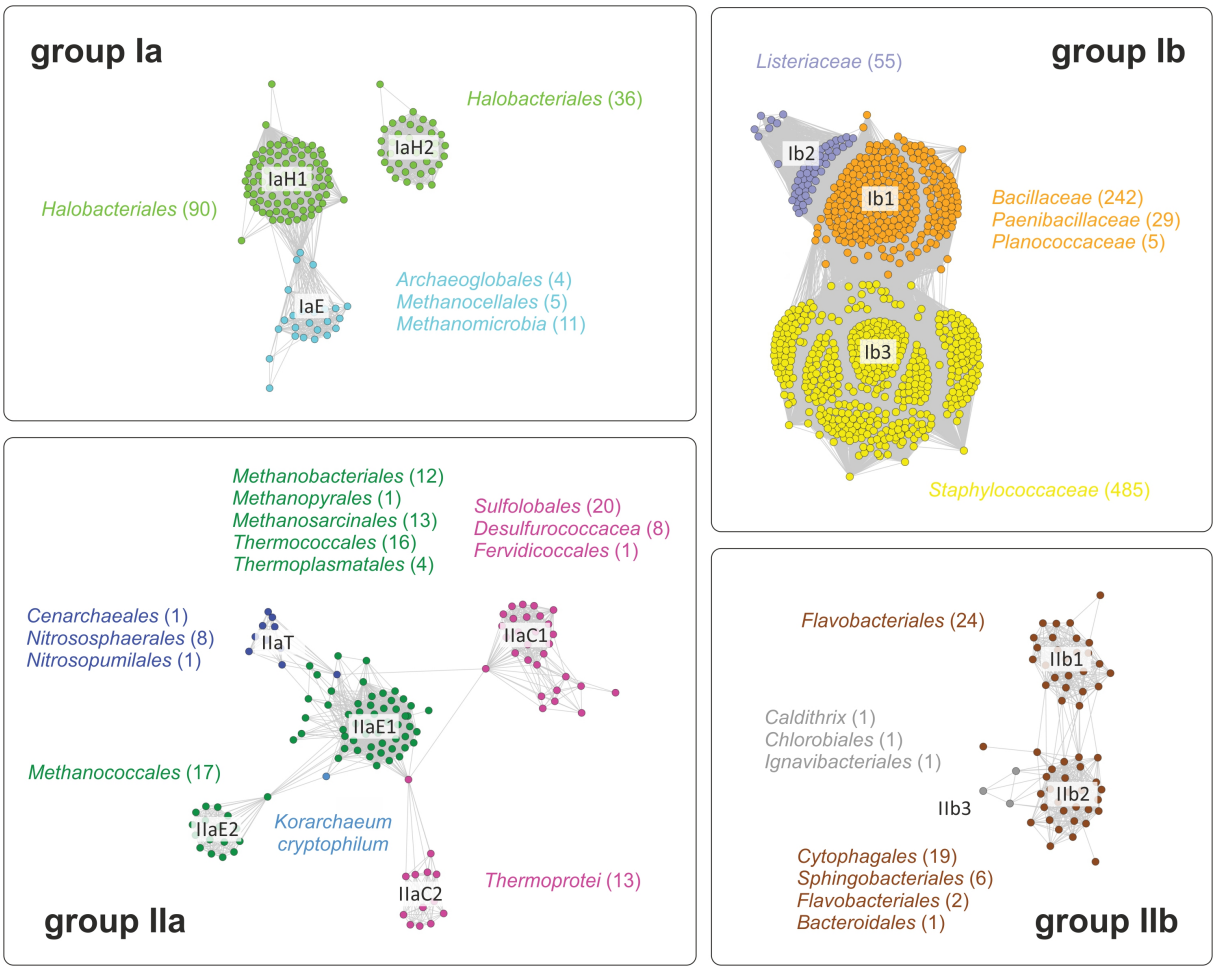


Figure 7: Taxonomic details of the main clusters in the sequence similarity network
All taxonomic orders occurring in a cluster (color of the taxon labels corresponds to the cluster color) are given together with the number of their members (in brackets). Because in case of cluster Ib all members are from the order *Bacillales*, the family names of the clustering members are given instead.

Checking the taxonomic details we found that many Halobacteria possess two different GGGPS enzymes, consistent with former findings (Boucher et al. 2004). It seems that there exist two paralogues of a GGGPS-like enzyme in these species. Interestingly, the two distinct subgroups (IaH1 and IaH2) are formed by these paralogues. The occurrence of a IaH2 type enzyme is in almost all cases connected to the occurrence of an enzyme of type IaH1. Only two Halobacteria strains (*Halovivax asiaticus*, *Halovivax ruber*) contain exclusively a type IaH2 enzyme. Boucher et al. (2004) suggest that the existence of two paralogues in some Halobacteria could be the reason for the occurrence of farnesylgeranylglyceryl phosphate (FGGP) and geranylgeranylglyceryl phosphate (GGGP) since the two enzymes might be the corresponding G1P-prenyl transferases with specificities for C25 and C20 substrates, respectively (Boucher et al. 2004). We resume this discourse in publication B, by discussing how an insertion of 25 amino acids (on average, figure 8), which led to the splitting into the

two halobacterial subgroups, is embedded into the structure of the GGGPS and that it might influence substrate specificity of the enzyme.

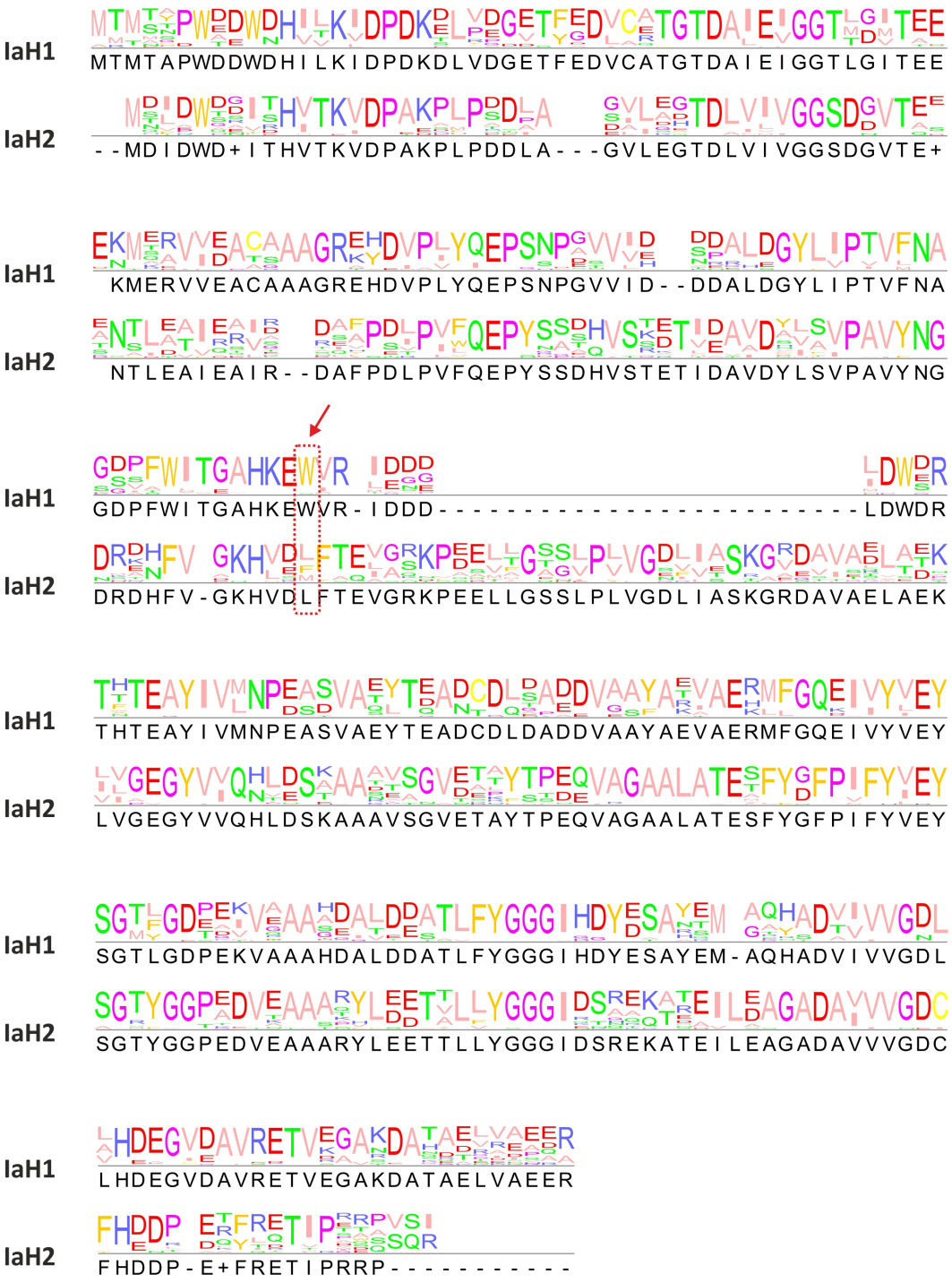


Figure 8: Consensus sequence and conservation logo illustration of the two variants of GGGPS-like enzymes in Halobacteria.
A multiple sequence alignment of the sequences of the two main clusters from Halobacteria in the sequence similarity network (see figure 6 and 7), which represent the two paralogues, was calculated using Clustal Omega (Sievers et al. 2011). The subgroup of sequences possessing the insertion (IaH2) contained 37 sequences while the group without insertion (IaH1) included 86 sequences. The conservation logo (colored, above the line) and the consensus sequence (black, below the line) for the

two subgroups was generated using Jalview (Waterhouse et al. 2009). The residues are colored according to their physicochemical properties (Zappo color code) as follows: Aliphatic/hydrophobic in pink (ILVAM); aromatic in orange (FWY); positive in blue (KRH), negative in red (DE), hydrophilic in green (STNQ), conformationally special in magenta (PG), and cysteine in yellow (C). The position of a proposed substrate length limiting residue, as discussed in 1.2.6 and, publication B is marked by a red arrow and box.

In chapter 1.2.6 evidence is provided, that the splitting of the Bacteroidetes and Crenarchaeal sequences is connected to their different oligomerization states. Another reason for the splitting of the Crenarchaeota into two subgroups could be the occurrence of a highly positively charged N-terminal stretch, which might be important for the interaction with the membrane as it has been proposed in case of the *Archaeoglobus fulgidus* enzyme (Payandeh et al. 2006). The structural basis for the separation of the two clusters of group II Euryarchaeota is currently unclear.

1.2.4 The differing substrate specificities among GGGPS-like enzymes

It has been mentioned before, that different members of the group of GGGPS-like enzymes have two different substrate specificities (see 1.2.2). While three archaeal enzymes (one from group Ia and two from group IIa) have been shown to prefer GGPP (20 C-atoms, Chen et al. 1993; Zhang et al. 1993; Soderberg et al. 2001; Nemoto et al. 2003; Payandeh et al. 2007), bacterial group Ib enzymes prefer HepPP (35 C-atoms, Guldan et al. 2011).

Members of the novel characterized bacterial group IIb exhibit archaeal substrate specificity

In publication B, the substrate specificities of a large subset of members of the GGGPS family were compared, including representatives from the so far uncharacterized group IIb. Three different assays (I, II, III) were used, all based on radioactively labeled substrates.

I) In a series of *in vivo* assays 17 GGGPS-like genes were overexpressed in a *B. subtilis* Δ*pcrB* strain (a *B. subtilis* strain lacking the GGGPS-like gene (*pcrB*), Kobayashi et al. 2003) in presence of radiolabeled G1P (Guldan et al. 2011). The *B. subtilis* Δ*pcrB* strain synthesizes various isoprenyl diphosphates, including farnesyl diphosphate (FPP, 15 C-atoms), GGPP (20 C-atoms), HepPP (35 C-atoms), and undecaprenyl diphosphate (UndPP, 55 C-atoms) by specific short-, medium- and long-chain prenyltransferases (Takahashi et al. 1980; Takahashi et al. 1981; Takahashi et al. 1982). Together with the radiolabeled G1P, the different isoprenyl diphosphates are substrates for the heterologously expressed GGGPS-like enzymes, which produce ether lipids according to their substrate specificity. After the labeling

experiment, the lipids were extracted and analyzed by thin layer chromatography and autoradiography (publication B, figure 2).

II) Purified GGGPS-like enzymes (publication B, figure S2) were subjected to an *in vitro* assay, where we used the (commercially available) short chained substrates geranyl diphosphate (GPP, 10 C-atoms), FPP (15 C-atoms) and GGPP (20 C-atoms) together with radiolabeled G1P (publication B, figure S3).

III) Because *B. subtilis* does not provide substrates of chain length C25, C30 and longer than C40, a coupled two-step *in vitro* assay was established and optimized to monitor the activity of the different variants with such substrates. A mixture of substrates (C20-C40) was synthesized and simultaneously radiolabeled by an octaprenyl diphosphate synthase from *E. coli* (ecOPPS), followed by the conversion of these substrates by the different GGGPS variants in a second step together with G1P (Ren et al. 2013). Effects of detergents on the product chain-length spectrum of polyprenyl diphosphate synthases have been discussed recently (Pan et al. 2013). We therefore tested the effect of some alcohols and detergents on ecOPPS and found that 2-methyl-2-propanol worked best to achieve a broad product spectrum while no interference with the GGGPS reaction is present (figure 9 A). Figure 9 B shows all purified GGGPS-like variants which were characterized with the two-step *in vitro* assay.

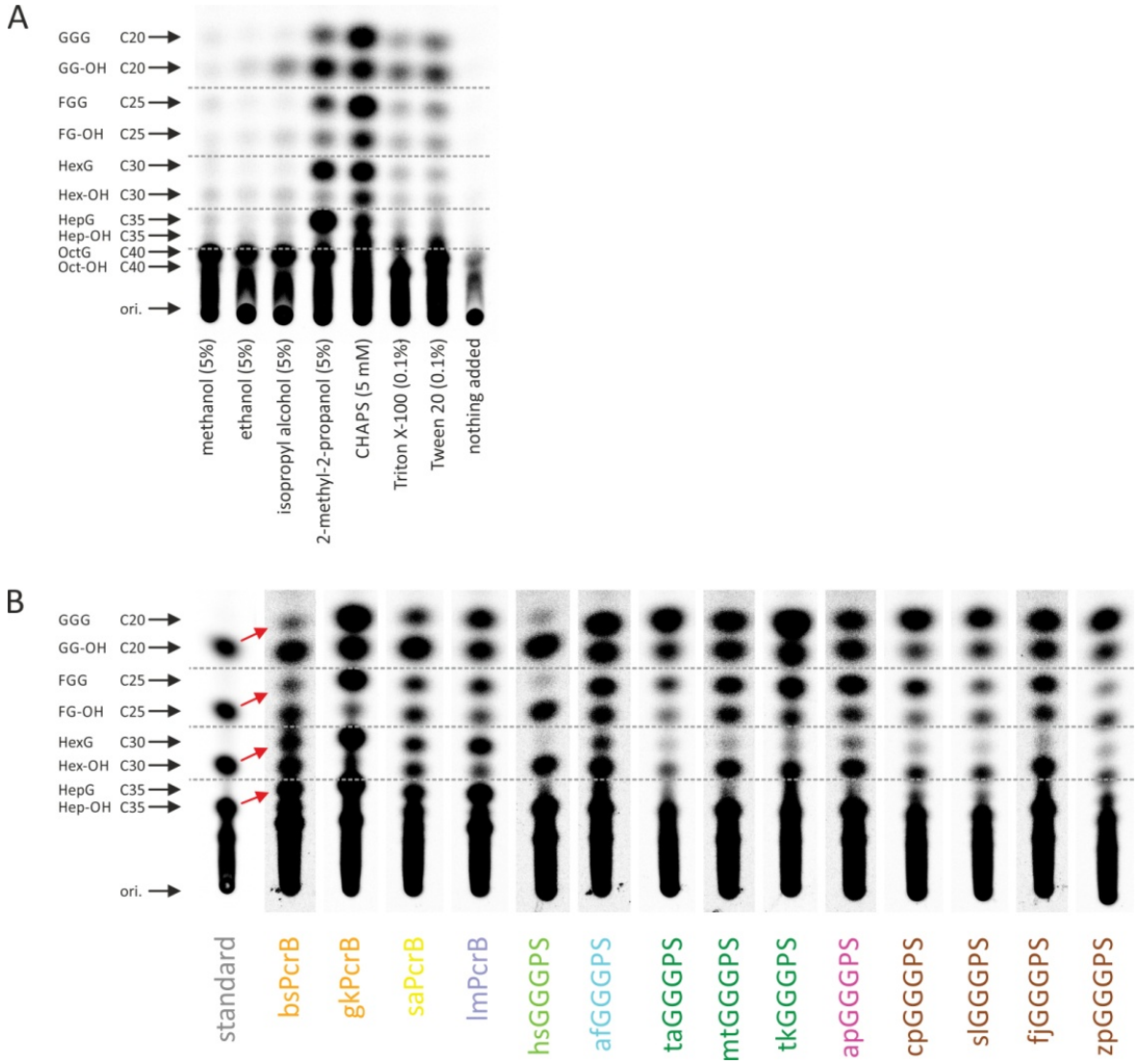


Figure 9: A two-step coupled *in vitro* assay to analyze specificities for C20-C40 polyprenyl diphosphates.
The unlabeled substrates for ecOPPS, GGPP (Sigma-Aldrich) and FPP (Santa Cruz Biotechnology, Inc.) were obtained in methanol/NH_4OH 7:3 (v/v). FPP (25 nmol) and GGPP (5 nmol) were combined, the solvent was removed by a vacuum concentrator, and the substrates were redissolved in 50 µl of reaction mixture, containing 50 mM HEPES (pH 7.5), 50 mM KCl, 0.5 mM $MgCl_2$, 5 % 2-methyl-2-propanol, 1 µM ecOPPS (with N-terminal thioredoxin-fusion) and 100 µM ^{14}C-isopentenyl diphosphate (American Radiolabeled Chemicals, Inc.). The mixture was incubated at 37 °C for 30 min upon slow shaking. In the next step, 15 µM of a GGGPS variant, 4 mM $MgCl_2$ and 20 mM G1P were added, and the slowly shaking mixture was incubated for 6 h at 37 °C. To remove the phosphate of the polyprenyl glycerol phosphates for subsequent TLC analysis, potato acid phosphatase (Sigma-Aldrich) and 100 µL of 40 mM sodium acetate (pH 4.7), 20 % 1-propanol, 0.1 % Triton X-100 was added. After overnight incubation at 37 °C with slow shaking, the lipids were extracted as described before (Bligh et al. 1959; Kates 1986; Guldan et al. 2011). The concentrated extracts were spotted onto a C18 reverse phase HPTLC plate (Merck Millipore) which was subsequently developed in acetone/methanol 2:8 (v/v). The products were visualized by autoradiography. The substrates and products are labeled as follows: GGG, geranylgeranylglycerol; GG-OH, geranylgeraniol; FGG, farnesylgeranylglycerol; FG-OH, farnesylgeraniol; HexG, hexaprenylglycerol; Hex-OH, hexaprenol; HepG, heptaprenylglycerol; Hep-OH, heptaprenol; OctG, octaprenylglycerol; Oct-OH, octaprenol. Red arrows show the band shift due to the conversion of polyprenyl diphosphates to polyprenyl glycerols for bsPcrB. A) Influence of several alcohols and detergents on the assay, using bsPcrB as

GGGPS variant. In the last sample ("nothing added") no detergent or alcohol was added. An equal amount of polyprenyl diphosphate substrates of 20 to 40 carbon atoms is obtained when adding 2-methyl-2-propanol or CHAPS to the reaction mixture. Optimal turnover of the different substrates and G1P by bsPcrB is obtained in the presence of 2-methyl-2-propanol. B) Different GGGPS-like enzymes were incubated with G1P and the radiolabeled mixture of C20-C40 polyprenyl diphosphates produced by ecOPPS (due to insufficient resolution in the HPTLC, only the position C20-C35 is marked by arrows).

In the *in vivo* assay, where the native substrates C20 and C35 and in addition C15 and C55 were available as substrates, group Ib enzymes preferred HepPP, while group II enzymes only used GGPP significantly as substrate. The group Ia variants provided ambiguous signals: while the GGGPS from *Halobacterium salinarum* was inactive the enzyme from *A. fulgidus* seems to be more promiscuous to longer chained substrates compared for group II enzymes (publication B, figure 2). The *in vitro* assay with commercially available substrates (C10, C15, C20) showed all enzymes to accept short chained polyprenyl diphosphates (publication B, figure S3). In addition the two step *in vitro* coupled assay provides C25, C30 and C40 as substrates. In this assay a slight promiscuity towards farnesylgeranyl diphosphate (FGPP, C25) substrate appeared in all group Ia and group II enzymes, while group Ib enzymes accepted all substrates including octaprenyl diphosphate (C40; figure 9 B).

Interestingly, several archaeal organisms possess additional ether lipids based on the alternative polyprenyl building block FGPP (De Rosa et al. 1986; De Rosa et al. 1988; Koga et al. 2005), but *sn*-glycerol-1-phosphate farnesylgeranyl transferases have not been identified yet. The Crenarchaeon *Aeropyrum pernix* displays an exceptional core lipid composition. The membranes of *A. pernix* consist solely of 2,3-di-*O*-sesterterpanyl-*sn*-glycerol (both polyprenyl chains connected to G1P consist of 5 isoprene units; Morii et al. 1999). In chapter 1.2.6 it is discussed how a differing mechanism in substrate length determination could lead to a slight substrate promiscuity of the GGGPS towards longer substrates. In addition, *A. pernix* features a polyprenyl diphosphate synthase of altered specificity (Tachibana et al. 2000), which produces C25 polyprenyl diphosphate.

1.2.5 Structural features of GGGPS-like enzymes

An ancient fold for an ancient enzyme

The $(\beta\alpha)_8$-barrel fold is the most frequent fold among single domain proteins (Sterner et al. 2005). It constitutes 10 % of all proteins with known three-dimensional structure. In the SCOP database, 33 superfamilies are connected to the fold while $(\beta\alpha)_8$-barrels catalyze more than 60 different reactions. Hence it is not surprising that $(\beta\alpha)_8$-barrels are also present in the

large and heterogeneous family of prenyltransferases (Oldfield et al. 2012). Two types of prenyltransferases have been shown to be $(\beta\alpha)_8$-barrels: the GGGPS-like prenyltransferases and MoeO5, an enzyme involved in the biosynthesis of the antibiotic moenomycin, which catalyzes the trans-to-cis isomerization of its substrate farnesyl diphosphate when it is transferred to the second substrate 3-phosphoglycerate (Doud et al. 2011). Considering the spacial extent of the substrates of GGGPS-like enzymes, it is fascinating how the comparatively small $(\beta\alpha)_8$-barrel scaffold can display such a perfect adaptivity. In fact, HepGP (the product of PcrB) in a stretched conformation measures nearly 42 Å, while the average diameter of a $(\beta\alpha)_8$-barrel is with approximately 50 Å only slightly larger. Thus, large sections of the protein are in contact with the substrates.

Structural adaptations

Several crystal structures of three group I enzymes (from *A. fulgidus*, *B. subtilis* and *Staphylococcus aureus*) have been solved in complex with G1P as well as analogues of the hydrophobic prenyl substrate, revealing important catalytic residues and leading to a detailed hypothesis of the mechanism of substrate specificity (Badger et al. 2005; Payandeh et al. 2006; Guldan et al. 2011; Ren et al. 2013). Furthermore, a crystal structure of the closely related prenyltransferase MoeO5 has been solved in complex with its substrates (Ren et al. 2012). In publication B, four additional structures are provided. Two structures are from group II species: one from the group IIa organism *M. thermautotrophicus* and one from the group IIb organism *Flavobacterium johnsoniae*, both solved in complex with G1P. These structures are the first solved structures for group II enzymes. The two other structures are from the group I enzyme from *Geobacillus kaustophilus* (one apo structure and one with its ligand G1P) which is closely related to the PcrB enzyme from *B. subtilis* (58 % sequence identity). Both group II and the group I holo structure (with G1P) bound a polyethylene glycol molecule in their binding pocket for the hydrophobic substrate (originating from the crystallization buffer). The basic data for all published structures of the GGGPS-like enzyme family, including our novel structures (for details see publication B, table S3), is summarized in table 1.

Table 1. Structures solved from the family of GGGPS-like enzymes.

Group	Source organism	Ligands	Res. / Å	pdb-code	Oligom. state	Reference
Ia	*A. fulgidus*	citrate	1.55	2F6U	dimer	(1)
Ia	*A. fulgidus*	G1P, MPD	2.0	2F6X	dimer	(1)
Ib	*B. subtilis*	-	1.9	1VIZ	dimer	(2)
Ib	*B. subtilis*	-	1.54	3VZX	dimer	(3)
Ib	*B. subtilis*	G1P	1.63	3VZY	dimer	(3)
Ib	*B. subtilis*	FsPP	2.02	3VZZ	dimer	(3)
Ib	*B. subtilis*	G1P, FsPP	2.50	3W00	dimer	(3)
Ib	*S. aureus*	PEG	1.54	3W01	dimer	(3)
Ib	*S. aureus*	sulfate	2.98	3W02	dimer	(3)
Ib	*G kaustophilus*	G1P, PEG	2.0	4NAE	dimer	publication B
Ib	*G. kaustophilus*	-	2.0	4NAF	dimer	publication B
IIb	*F. johnsoniae*	G1P, PEG, phosphate	1.52	4JEJ	dimer	publication B
IIa	*M. thermautotrophicus*	G1P, PEG	2.8	4MM1	hexamer	publication B

Grp. = group; Res. = Resolution; Oligom. = Oligomerization; FsPP = *S*-thiolo-farnesyl diphosphate; MPD = (4S)-2-methyl-2,4-pentanediol; PEG = polyethylene glycol; (1) = Payandeh et al. 2006; (2) = Badger et al. 2005; (3) = Ren et al. 2013.

Figure 10 exemplarily shows the GGGPS $(\beta\alpha)_8$-barrel from *F. johnsoniae* with its distinct structural features.

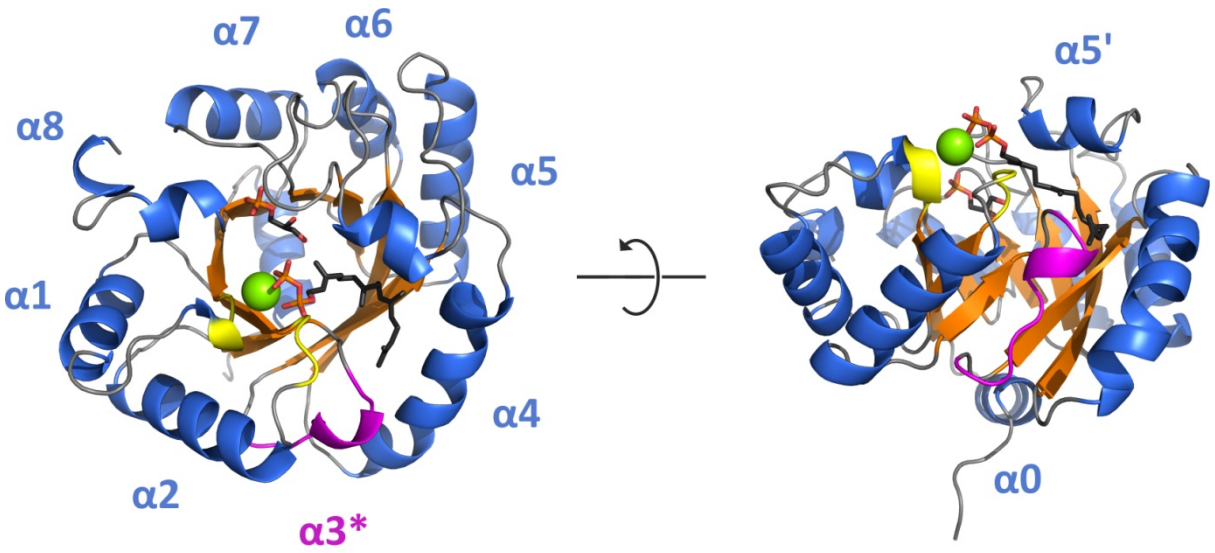


Figure 10: Structural features of the group II GGGPS-like enzyme from *F. johnsoniae* with bound ligands.
Secondary structure elements and G1P (stick representation) of the crystal structure are shown (pdb-code: 4JEJ). GGPP (stick representation) and Mg^{2+} (green sphere) were modeled using YASARA Structure Version 13.4.21 (employing the YAMBER3 force field (Krieger et al. 2004)) based on the position of HepPP in PcrB from *B. subtilis* (Guldan et al. 2011) and the position of Mg^{2+} in MoeO5 (pdb-code: 3VKB). Ligand carbon atoms are given in black, oxygen atoms in red, phosphate atoms in orange. The surface exposed α-helices are given in blue, the central β-sheets in orange, the loops of the Mg^{2+} binding site in yellow, and the flexible "swinging door" helix α3* in magenta. For clarity one protomer of the homodimer is shown. Taken from figure 3 A, publication B.

The overall structure reflects the $(\beta\alpha)_8$-fold with some characteristic variations. As in other closely related $(\beta\alpha)_8$-barrels such as the α-subunit of the tryptophan synthase (TrpA, Hyde et al. 1988), the secondary structure of the GGGPS-like proteins exhibits an additional N-terminal helix α0. This helix has been proposed to mediate the interaction of GGGPS with the membrane, since many members of the family feature a number of positively charged and hydrophobic residues, as a multiple sequence alignment (Payandeh et al. 2006) and our sequence similarity networks show (compare 1.2.3). The most striking difference to the canonical $(\beta\alpha)_8$-barrel fold is the substitution of helix α3 by a long loop (in the following addressed with α3*). This loop has been termed "swinging door" due to its high flexibility (as suggested by high B-factors). It is located at the distal end of the binding pocket and might serve as a gateway for the prenyl substrate (Payandeh et al. 2006). Helix α4 and α5 are somewhat contorted to the outward, which allows them to form the interface with the second protomer of the dimer (Payandeh et al. 2006; Ren et al. 2013; publication B).

Our novel group II structures differ to some extent from the well known group I structures. The most interesting differences appear in the G1P binding site and in the quaternary structure. The hydrophobic part of the G1P binding site is more pronounced in group II

enzymes due to the exchange of a lysine by a leucine (compare figure 11 in chapter 1.2.6 and publication B, figure 4). In fact, the lysine is conserved among group I enzymes, while the leucine is conserved among group II enzymes (Payandeh et al. 2006). The structure of the GGGPS from *M. thermautotrophicus* revealed a hexameric architecture (compare chapter 1.2.6 and publication B, figure 3). In publication B we describe three additional oligomerization interfaces for hexamerization (interface 2, 3a, and 3b), in addition to the standard interface (interface 1) for dimerization (publication B, figure S5). Interestingly the location of the additional interfaces in the hexameric protein is partially identical with the location of the two alternative interfaces of the dimeric protein PcrB from *B. subtilis* which we analyzed publication A. It is tempting to speculate about an evolutionary background of this finding (see chapter 1.2.6).

1.2.6 Structure-function relationships in the family of GGGPS-like enzymes

Two disparate substrate binding sites

The different chemical traits of the two substrates of GGGPS-like enzymes are reflected in their binding sites in the enzyme. While the small hydrophilic G1P binds at the top inner rim of the barrel with its phosphate group attached to the standard phosphate binding motif of the $(\beta\alpha)_8$-fold (Nagano et al. 2002; Payandeh et al. 2006), the long prenyl substrate is bound into a hydrophobic cleft which extends to a tunnel at its end near the above described "swinging door". This "greasy slide" (Payandeh et al. 2006) is basically formed by helix α3*, β-strand β4 and partially β5 as well as helix α4 and α5'.

The diphosphate moiety of the hydrophobic substrate has been discussed to bind in complex with magnesium (Payandeh et al. 2006; Ren et al. 2013). This is supported by the observation that GGGPS-like enzymes are inactivated by the addition of EDTA to the reaction (Chen et al. 1993; Zhang et al. 1993; Nemoto et al. 2003). Furthermore, the same is common in other prenyltransferases like MoeO5 (Ren et al. 2012) or the more distantly related polyprenyl diphosphate synthases (Christianson 2006; Peisajovich et al. 2007; Oldfield et al. 2012). In fact, the active site of the well studied group of polyprenyl synthases contains Mg^{2+} ions. They hold the diphosphates of the substrates in place and are coordinated by aspartates from DXXDD motifs (Oldfield et al. 2012). In comparison, in the GGGPS-like enzymes a conserved aspartate in a negatively charged patch, formed by residues at the rim of the barrel (loops βα1 and βα2), has been shown to be essential for the reaction (Payandeh et al. 2007; Ren et al. 2013). A bound magnesium in the structure of MoeO5 locates in this region when

superimposing it with the available structures of GGGPS-like enzymes. Consequently, this region can be assumed to be the binding site for the diphosphate moiety of the hydrophobic substrate in group I GGGPS-like enzymes.

In publication B we show, based on our novel structures, that this feature is likewise present in group II enzymes (publication B, figure S4). In fact, the negative patch on the protein surface is even more pronounced due to additional negatively charged residues in this region.

Length limitation mechanism for the hydrophobic substrate

Payandeh and co-workers first proposed a mechanism for the length determination of the hydrophobic prenyl substrate similar to the mechanism adopted by the polyprenyl diphosphate synthases (Tarshis et al. 1996; Payandeh et al. 2006). In this model, a bulky amino acid side chain (mostly aromatic) limits the binding pocket at its end by steric hindrance. For the GGGPS-like enzymes this means, that such a "limiter residue" might exist at the end of the "greasy slide" near the "swinging door". For the archaeal group I GGGPS from *A. fulgidus,* Payandeh et al. predicted a tryptophan in this region to undertake the limiter-function which could be confirmed by Guldan et al. (Payandeh et al. 2006; Guldan et al. 2011). Guldan et al. could furthermore show, that the restriction to GGPP as a substrate is revoked in case of the closely related bacterial group I enzyme PcrB from *B. subtilis*, where the tryptophan is substituted by an alanine. Consequently, the enzyme accepts longer substrates. Ren et al. recently suggested a more remotely located tyrosine to carry over the limiter function for the HepPP substrate in *B. subtilis* PcrB (Ren et al. 2013).

Along this line we could identify limiter residues in the two GGGPS variants from group IIa and IIb, whose structures we solved (publication B, figure 5). Since the tryptophan limiter does not exist in group II enzymes, we searched for alternative residues, which could limit the substrate spectrum of those enzymes to a maximal length of 20 C-atoms. We tested several residues by means of alanine exchanges in the group IIb enzyme of *F. johnsoniae.* Subjecting them to the *in vivo* (publication B, figure 5) and the *in vitro* coupled assay (figure 11), we found that one variant (I90A) located in loop α3* exhibited a substrate spectrum shifted towards longer substrates. We obtained the same result, when we tested the analogous variant (V86G) of the enzyme from *M. thermautotrophicus.*

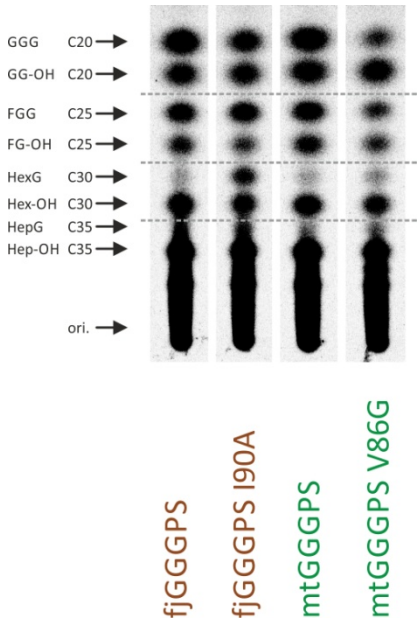


Figure 11: The limiter mutants in the two step coupled *in vitro* assay.
Wild-type GGGPS-like variants from *F. johnsoniae* and *M. thermautotrophicus* and their limiter mutants were incubated with G1P and the radiolabeled mixture of C20-C40 polyprenyl diphosphates produced by ecOPPS. For a detailed description of the protocol compare figure 9.

Interestingly, the identified position is not homologous to the position of the tryptophan limiter from group Ia enzymes, which is located in helix α4. These findings support the hypothesis of two independent LGT events in the evolution of the family of GGGPS-like enzymes (Boucher et al. 2004).

Structural prerequisites for stereo-selective catalysis

Several requirements for the catalysis of the reaction of the GGGPS family are essential and consequently reflected by conserved residues in the proteins. Considering the G1P binding site, two prerequisites have to be taken into account – stereo-selectivity and activation. Stereo-selectivity is obtained by the anchoring of the G1P phosphate group to the $(\beta\alpha)_8$-barrel standard phosphate binding motif (Nagano et al. 2002; Vega et al. 2003; Payandeh et al. 2006), and by the architecture of the binding pocket, especially surrounding the asymmetric C2 carbon atom (figure 12).

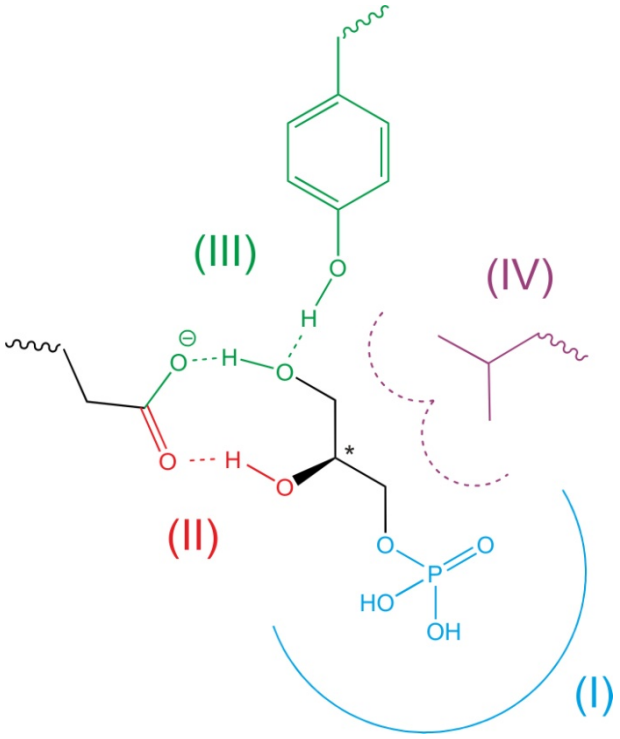


Figure 12: Elements in stereo-selective binding of G1P by GGGPS-like enzymes.
(I) The G1P molecule is anchored with its phosphate group to the standard phosphate binding motif in the $(\beta\alpha)_8$-barrel (blue). (II) A conserved carboxyl side chain (glutamate or aspartate) binds the hydroxy group of the chiral C2 carbon atom. (III) The C3 hydroxy group is bound by the conserved glutamate and a conserved tyrosine. (IV) Binding of the enantiomer G3P is disfavored, partially due to sterical hindrance by the hydrophobic part (violet) of the binding site (illustrated by the side chain of a leucine, as it occurs in group II enzymes).

Aided by our new structures, we propose in publication B, that the hydrophobic part of a lysine (group I enzymes) or the completely hydrophobic side chain of a conserved leucine (group II enzymes) in the active site is an important feature of the mechanism of stereo-selectivity. Positioning and activation of the C3 oxygen for the electrophilic attack by the C1 carbon atom of the polyprenyl diphosphate is accomplished by a conserved glutamate and tyrosine (figure 4, publication B). The rupture of the carbon-oxygen bond in the isoprenoid diphosphate substrate preceding the electrophilic alkylation is most probably assisted by the negatively charged patch as discussed above (compare figure S4, publication B).

The quaternary structure influences substrate specificity

Two different quaternary structures have been described for the family of GGGPS-like enzymes. While all previously characterized group I enzymes apparently exhibit dimeric oligomerization states (Badger et al. 2005; Payandeh et al. 2006; Ren et al. 2013), one group II GGGPS from *M. thermautotrophicus* has been predicted to be a pentamer (Chen et

al. 1993; Soderberg et al. 2001) and another group II GGGPS from *Thermoplasma acidophilum* to be a dimer (Nemoto et al. 2003). Although Pai and co-workers speculated on a mutual interference of the two active sites in the homodimeric GGGPS from *A. fulgidus* (Payandeh et al. 2006), little was published on this matter when we started our investigations on the group of GGGPS-like enzymes.

In publication A we addressed the question of the native dimer configuration of the group Ib enzyme PcrB from *B. subtilis*. This first published structure of a GGGPS-like protein, solved within a structural genomics project (Badger et al. 2005), was deposited (pdb-code: 1VIZ) in an implausible spatial orientation. The predicted interface was, relative to the protein size, small and differed from the dimerization interface of the homologous group Ia enzyme from *A. fulgidus*, which was published one year later (Payandeh et al. 2006). Using the program PresCont (Zellner et al. 2012), which was recently developed in our group, we predicted additional putative interfaces (publication A, figure S1) leading to three possible dimer configurations (publication A, figure 1). We evaluated these configurations using the Rosetta force field (Leaver-Fay et al. 2011). To this end, we basically used the force field to calculate the energetic effects of amino acid exchanges in the interfaces of the different complexes on their stability (publication A, figure S2). The results favored the *A. fulgidus* configuration. Next, we investigated the three predicted alternative configurations by amino acid substitutions which hindered the assembling of the dimer (publication A, figure 1). The dimer was disrupted, as tested by size exclusion chromatography, only in case of mutations in the interface that was homologous to the interface in the GGGPS from *A. fulgidus* (publication A, figure 2). In an alternative experiment, we incorporated an unnatural amino acid into the predicted interfaces (publication A, figure 3 A), which allowed for cross-linking of the two protomers. These experiments confirmed the findings from the monomerization experiments (publication A, figure 3 B and figure S4), thereby proving that the quaternary structure of PcrB from *B. subtilis* corresponds to the one from the *A. fulgidus* enzyme.

Next, we were interested in the significance of oligomer formation. To test a potential impact on stability, we examined the thermal stability of the monomeric proteins. As a result, we found them to be only slightly less stable than wild-type PcrB (publication A, figure S5). In contrast, the spectrum of accepted hydrophobic substrates was dramatically altered for the monomeric variants (publication A, figure 4). They accepted only the short chained substrates GPP (consisting of two isoprene units) or FPP (consisting of three isoprene units) although the native substrate consists of seven isoprene units (Guldan et al. 2011). Comparing structures with modeled substrate and the apo structure, we found significant movements of

secondary structure elements towards the adjoining protomer and postulate that they are necessary for substrate binding (publication A, figure 5). These findings, together with the altered substrate spectrum of the monomeric PcrB variants, provide evidence for an essential impact of oligomerization on activity and substrate specificity in PcrB.

An aromatic "anchor" amino acid mediates hexamerization

While we investigated the effects of dimerization in publication A, we addressed the ongoing question (Soderberg et al. 2001; Payandeh et al. 2006) of the occurrence of higher oligomerization states in the group of GGGPS-like enzymes in publication B. Therefore we analyzed the oligomerization states of 14 variants from all subgroups of the family by light scattering measurements (publication B, table S4). It turned out that most group I enzymes are dimers, except the characterized variant from Halobacteria that gave a signal for a monomer. This might be an experimental artifact due to the special properties of proteins from Halobacteria (Virnekas et al. 1994; Ono et al. 1995; Kayushin et al. 1996).). Group II enzymes are either dimers or hexamers (publication B, figure 1 and figure S1).

Aided by our structure of the hexameric GGGPS variant from *M. thermautotrophicus*, we could pinpoint residues which are important for hexamerization (publication B, figure S5). A tryptophan is essential in hexamer formation in the GGGPS from *M. thermautotrophicus* ("aromatic anchor"). We tested this by mutating the tryptophan to alanine, which lead to a dimeric protein (publication B, table S2 and S4). We furthermore mutated the aromatic amino acids in the homologous position in two other hexameric variants, one from the group IIa organism *Thermococcus kodakaraensis* and one from the group IIb organism *Chitinophaga pinensis* (publication B, table S2 and S4). As before, an alanine in the position led to dimeric proteins. Examining the position in a multiple sequence alignment containing all sequences from the GGGPS-like enzymes from the similarity network (1205 sequences, compare figure 5), we found two groups of amino acids with different traits to be predominant. One group contains a hydrophobic aromatic amino acid (tryptophan, tyrosine or phenylalanine) while the other group contains a positively charged amino acid (arginine, lysine or histidine). Interestingly, in most clusters of the sequence similarity network, the occurrence of either one of these groups is predominant (except for sequences from Halobacteria or group Ib1 sequences, where a broad spectrum of other residues occurs). We could furthermore show, that in case of an aromatic amino acid in the position of the anchor, a positively charged residue in the interacting interface (interface 3b, compare figure 15) is correlated (publication B, figure 5). The distribution of amino acids at the anchor position and the correlated position

for all clusters in the sequence similarity network is given in figure 13 (and partially in publication B, figure 5).

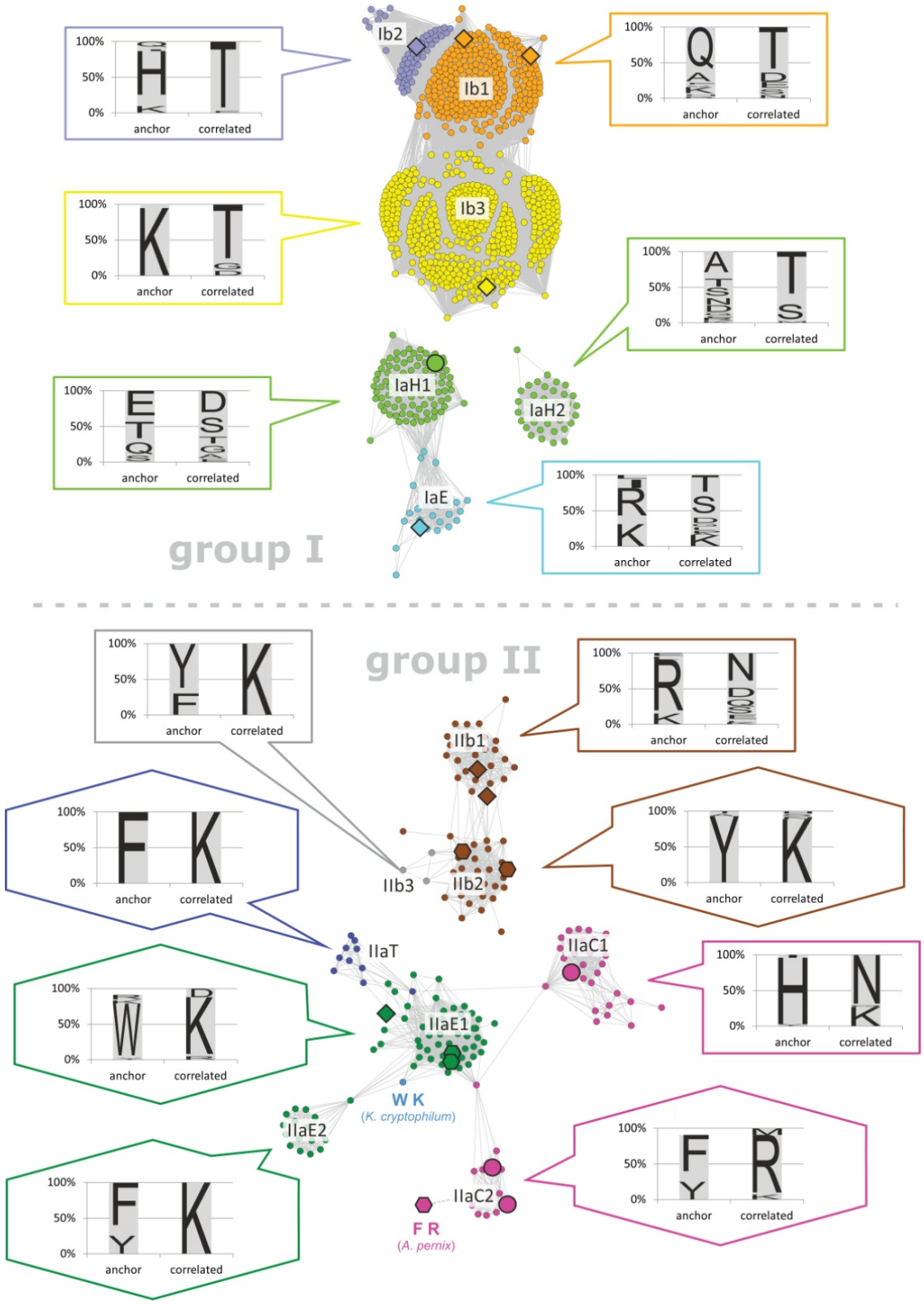


Figure 13: Correlation of similarity network clustering and oligomerization states of group II GGGPS-like proteins.
The network shown here is identical to figure 6. Variants that have been characterized in this study are shown in large symbols, and their quaternary structure is illustrated by shaping: hexagons symbolize

hexameric proteins, diamonds symbolize dimeric proteins, and quaternary structures of variants shown in circles were not analyzed. The speech bubbles show the sequence variation within each cluster at the "aromatic anchor" position ("anchor", green bar) and the position of the correlated lysine ("correlated", red bar), and the shape of the bubbles (square, hexagon) symbolized the assumed oligomerization state (dimeric, hexameric) of the proteins in the cluster. The hexagon close to cluster IIaC2 represents the GGGPS of *A. pernix*, and the dashed line indicates that it clusters with IIaC2 if the network threshold E-value is reduced to 10^{-58}. Note that the dimeric protein in group IIaE1, taGGGPS, has an R at the "anchor" position.

Given that an aromatic amino acid (tryptophan, tyrosine or phenylalanine) at the anchor position mediates hexamerization, most group IIa variants are hexamers. Only the Crenarchaeota subgroup IIaC1 bears a positively charged histidine in the anchor position, which might be an aspect of the group splitting. The group IIb variants bear either an aromatic or a positively charged residue at the anchor position. We characterized two hexameric proteins from the subgroup with an aromatic residue and two dimeric proteins from the subgroup with a positively charged residue at the anchor position, which is consistent with our hypothesis. This raises the question, whether the Bacteroidetes obtained their GGGPS in two separate LGT events from two ancestral subpopulations with different oligomerization states.

To investigate the reason for the different oligomeric states, we examined the influence of hexamerization on thermal stability and activity. To this end, we checked the stability (publication B, figure S6), activity (publication B, figure S3 and S7), and substrate specificity (publication B, figure S7, compare figure 9) of our disrupted hexamers and found only a slightly reduced activity, while stability and substrate specificity were unaltered. When comparing the B-factors of several structures of members from the GGGPS-like family as a measure for structural flexibility (Acharya et al. 2005), we found the regions nearby the interfaces to be the most rigid areas in the proteins (figure 14).

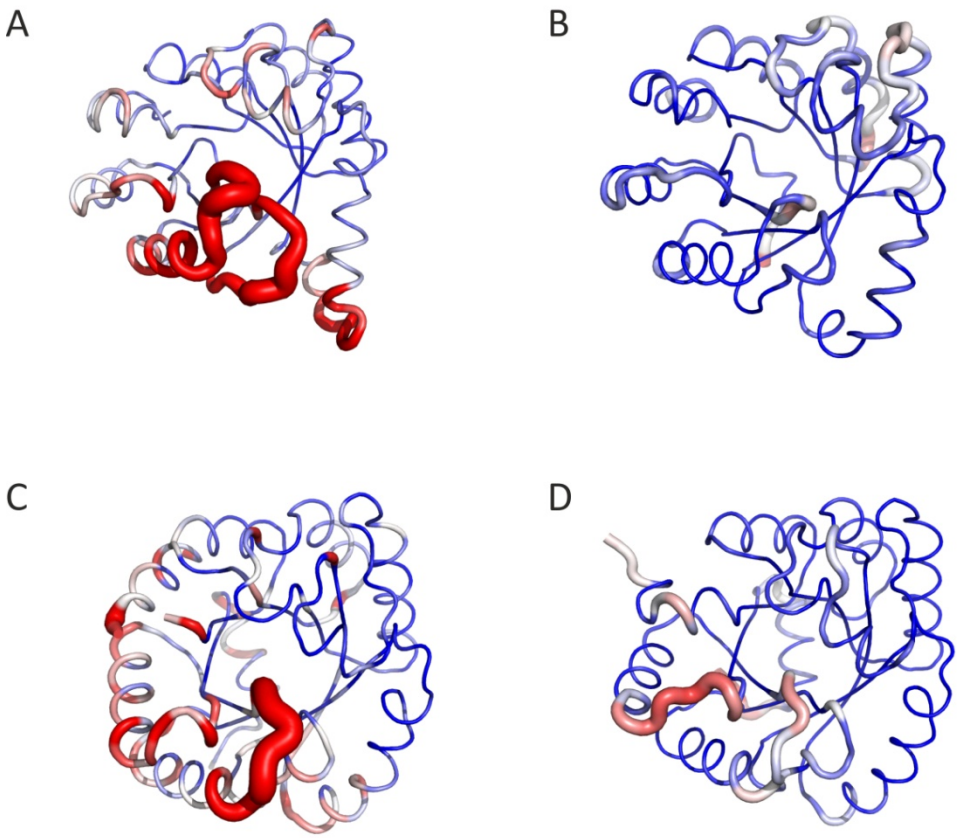


Figure 14: B-factors of representative structures of the GGGPS family
A: afGGGPS chain B (pdb-code: 2F6X); B: bsPcrB chain B (pdb-code: 3VZY); C: mtGGGPS chain A (pdb-code: 4MM1); D: fjGGGPS (pdb-code: 4JEJ). Increased flexibility, as illustrated by red color and thickened ribbon representation, is realized in the region of the "swinging door" and around the Mg^{2+} binding site. Increasingly rigid regions are given in white to blue coloring.

We assume the hexamerization to ensure structural integrity as a prerequisite for activity. This could be of benefit particularly under high temperature conditions.

Moreover it appeared, that the additional three oligomerization interfaces, which enable hexamerization, are identical or at least largely overlapping with the predicted alternative interfaces from PcrB (*B. subtilis*) in publication A (figure 15).

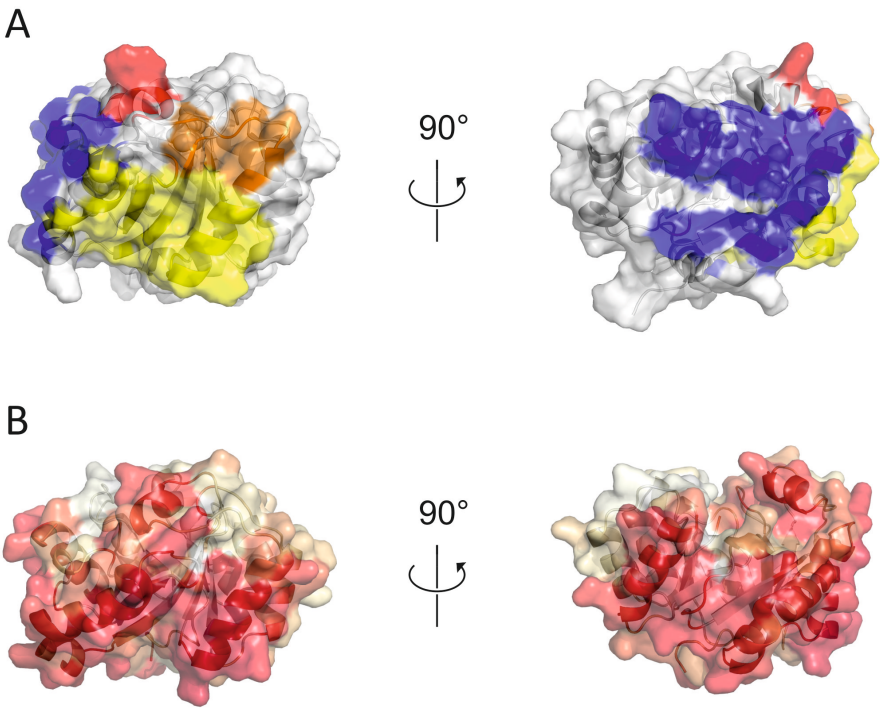


Figure 15: Additional oligomerization interfaces of the hexameric mtGGGPS and their localization compared to the localization of the predicted interfaces of bsPcrB in publication A. A) The symmetric interface 1 (yellow) represents the prototypical interface of the dimeric variants. Interface 2 (blue) is a second symmetric interface, while interface 3a (red) is located on top of the barrel and includes helix α5'. Interface 3b (orange) is located at the loop region α5'α5. Within the hexamer, two interfaces 1 are in contact with each other, as are two interfaces 2. Moreover, interface 3a is in contact with interface 3b (compare publication B, figure S5). B) PresCont predictions of the probability for being an interface, projected on the surface of bsPcrB. High probability is depicted by dark red coloring while white coloring symbolizes low possibility to form an interface. The position of the subunits shown in A) and B) is parallel according to a structural alignment.

Whether the signal for alternative interfaces is a remnant from hexameric ancestor variants or a prerequisite for the evolution of hexameric variants, which took place only in some members of the family, is unclear.

1.2.7 Functional and physiological context of GGGPS-like enzymes

On the substrates and products of the reaction of the GGGPS-like enzymes

It has been mentioned before, how GGGPS-like enzymes act as key enzymes in the synthesis of membrane lipids in Archaea (see chapter 1.2.1). Figure 16 gives an overview on the biochemical pathways were GGGPS-like enzymes are involved, outlining similarities and differences among the domains of Archaea and Bacteria.

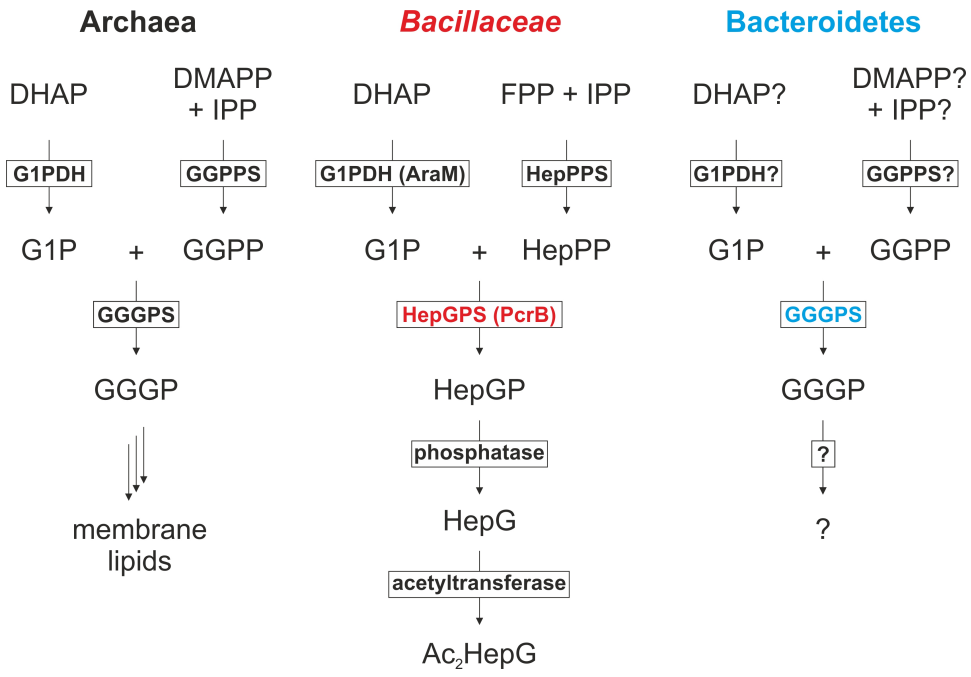


Figure 16: Synthesis pathways in Archaea and Bacteria, where GGGPS-like enzymes are involved
G1P is provided by reduction of DHAP, catalyzed by a G1P dehydrogenase (G1PDH, AraM in *B. subtilis*). The polyprenyl diphosphate substrates GGPP and HepPP are provided by a geranylgeranyl diphosphate synthase (GGPPS) and a heptaprenyl diphosphate synthase (HepPPS) from isoprene precursors (dimethylallyl diphosphate DMAPP, isopentenyl diphosphate IPP, farnesyl diphosphate FPP). While GGGP is a precursor of archaeal membrane lipids, heptaprenylglyceryl phosphate in *B. subtilis* is subsequently dephosphorylated by a phosphatase and acetylated by an acetyltransferase (Linde 2013). Details of the synthesis-pathways of the substrates and the further processing of the product of the GGGPS-reaction in Bacteroidetes are unknown (symbolized by "?").

All GGGPS-like enzymes transfer a polyprenyl diphosphate to G1P which is provided by a Zn^{2+}-dependent G1PDH in Archaea (Han et al. 2005, EC 1.1.1.261) and by a Ni^{2+}-dependent G1PDH in *Bacillacaeae*, as shown for *B. subtilis* by (Guldan et al. 2008). We could not identify an AraM or G1PDH homolog in all other Bacteria where GGGPS-like enzymes exist (*Listeria*, *Clostridia*, *Staphylococcus*, Bacteroidetes, Chlorobi, Caldithrix, Ignavibacteria, and Fibrobacteres). Hence, G1P in these taxa must be provided by an unknown enzyme or the environment.

The hydrophobic polyprenyl diphosphate substrates of different chain-length are provided by ubiquitous short- and medium-chain isoprenyl diphosphate synthases (Ogura et al. 1998, Takahashi et al. 1980; Takahashi et al. 1981; Takahashi et al. 1982; Wang et al. 2000; Oldfield et al. 2012). A special medium chain prenyltransferase, which provides the farnesylgeranyl diphosphate in *A. pernix* has been described (Tachibana et al. 2000). Besides this, recent investigations have demonstrated abiotic formation of polyprenyl phosphates on mineral surfaces from small precursor molecules (Nakatani et al. 2012). This underlines the

particular importance of the GGGPS and G1PDH as key enzymes in the early evolution of Archaea. A GGPPS in Bacteroidetes has not been described yet (marked with a "?" in figure 16).

The product of the reaction in Archaea (GGGP or FGGP) serves as precursor for a large variety of membrane lipids (De Rosa et al. 1986; De Rosa et al. 1988; Matsumi et al. 2011). In *B. subtilis*, the product HepGP is subsequently unspecifically dephosphorylated and specifically acetylated (Guldan et al. 2011; Linde 2013), but a physiological function of the ether lipids in Bacteria is still unknown. Guldan et al. (2011) described a cloggy growth of the *B. subtilis* Δ*pcrB* strain, which may be evidence for the contribution of ether lipids in *B. subtilis* to the cell envelope architecture and function.

2 Abbreviations

1-acyl-GAT	1-acyl-*sn*-glycerol-3-phosphate acyltransferase
Å	Ångström (10^{-10} m)
Ac_2HepG	di-acetyl-(*S*)-3-*O*-heptaprenylglycerol
A. fulgidus	*Archaeoglobus fulgidus*
A. pernix	*Aeropyrum pernix*
AraM	G1PDH from *B. subtilis*
BLAST	Basic Local Alignment Search Tool
bp	base pair(s)
B. subtilis	*Bacillus subtilis*
B. subtilis Δ*pcrB*	*B. subtilis* strain lacking the *pcrB* gene
°C	degree celsius
C1, C2, C3	carbon atom 1, 2, 3
cat	gene for chloramphenicol acetyltransferase
CD	circular dichroism
cDNA	complementary DNA
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-Propanesulfonate, a zwitterionic detergent
cm	centimeter (10^{-2} m)
CoA	Coenzyme A
Cre	recombinase derived from Bacteriophage P1
C-terminal	carboxyterminal end of a polypeptide chain
DGGGPS	(*S*)-2,3-di-*O*-geranylgeranylglyceryl phosphate synthase
DHAP	dihydroxyacetone phosphate
DMAPP	dimethylallyl diphosphate
DNA	deoxyribonucleic acid
DXXDD	amino acid motif consisting of three aspartetes (D) and two variable amino acids (X)
EC	Enzyme Commission
E. coli	*Escherichia coli*
ecOPPS	octaprenyl diphosphate synthase from *E. coli*
EDTA	ethylenediaminetetraacetic acid
E-value	Expect value

FGGP	(*S*)-3-*O*-farnesylgeranylglyceryl phosphate
FGPP	farnesylgeranyl diphosphate
F. johnsoniae	*Flavobacterium johnsoniae*
FPP	farnesyl diphosphate
fst	frame selectable transposon
FsPP	*S*-thiolo-farnesyl diphosphate
g	gram (10^{-3} kg)
GAT	*sn*-glycerol-3-phosphate acyltransferase
G1P	*sn*-glycerol-1-phosphate
G1PDH	*sn*-glycerol-1-phosphate dehydrogenase
G3P	*sn*-glycerol-3-phosphate
G. kaustophilus	*Geobacillus kaustophilus*
GGGPS	*sn*-glycerol-1-phosphate geranylgeranyltransferase
GGGP	(*S*)-3-*O*-geranylgeranylglyceryl phosphate
GGPP	geranylgeranyl diphosphate
GGPPS	geranylgeranyl diphosphate synthase
GPP	geranyl diphosphate
h	hour
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid, a zwitterionic organic chemical buffering agent
HepG	(*S*)-3-*O*-heptaprenylglycerol
HepGP	heptaprenylglyceryl phosphate
HepGPS	heptaprenylglyceryl phosphate synthase
HepPP	heptaprenyl diphosphate
HepPPS	heptaprenyl diphosphate synthase
HPTLC	high performance thin layer chromatography
IPP	isopentenyl diphosphate
Kan^R (*kanR*)	kanamycin resistence gene
kbp	kilo base pairs (1000 bp)
lacP	promoter of the *lac* operon
LB	lysogeny (luria) broth
LUCA	last universal common ancestor
LGT	lateral gene transfer
loxP	34 bp recognition site of Cre

µl	microliter (10^{-6} l)
µg	microgram (10^{-6} g)
m	meter
min	minute
mm	millimeter (10^{-3} m)
M. thermautotrophicus	*Methanothermobacter thermautotrophicus*
MoeO5	enzyme, responsible for the first step in moenomycin biosynthesis
MuA	transposase from bacteriophage Mu
MPD	(4S)-2-methyl-2,4-pentanediol
n	natural number
ng	nanogram (10^{-9} g)
NADH	nicotinamide adenine dinucleotide
N-terminal	aminoterminal end of a polypeptide chain
ORF	open reading frame
ori	origin of replication
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PcrA	a DNA helicase form *B. subtilis*, encoded in the same operon like PcrB
PcrB	heptaprenylglyceryl phosphate synthase from *B. subtilis*
pMB1	ori, derived from pMB1 vector
pdb	protein data bank
PEG	polyethylene glycol
Pfam	database of protein families
pM	picomolar (10^{-12} molar)
pmol	picomol (10^{-12} mol)
R1/R2	transposase recognition sequence
R6K	antibiotic resistance plasmid from *E. coli*
RT	room temperature
SOC	Super Optimal Broth medium
RT-PCR	reverse transcription polymerase chain reaction
SCOP	structural classification of proteins
S. aureus	*Staphylococcus aureus*
T. acidophilum	*Thermoplasma acidophilum*

TAG	amber stop codon
tat-sequence	N-terminal signal peptide of the twin-arginine translocation pathway (tat pathway)
tmhisA	*hisA* gene from *Thermotoga maritima*
TrEMBL	Translated EMBL, a very large protein database
TrpA	α-subunit of tryptophan synthase
UndPP	undecaprenyl diphosphate
UniProtKB	protein knowledgebase
vice versa	with position turned
VMA	vacuolar ATPase subunit of *Saccharomyces cerevisiae*

3 Literature

Acharya, K. R. and M. D. Lloyd (2005). "The advantages and limitations of protein crystal structures." Trends Pharmacol Sci **26**(1): 10-14.

Albers, S. V. and B. H. Meyer (2011). "The archaeal cell envelope." Nat Rev Microbiol **9**(6): 414-426.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). "Basic local alignment search tool." J Mol Biol **215**(3): 403-410.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin and G. Sherlock (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." Nat Genet **25**(1): 25-29.

Atkinson, H. J., J. H. Morris, T. E. Ferrin and P. C. Babbitt (2009). "Using sequence similarity networks for visualization of relationships across diverse protein superfamilies." PLoS One **4**(2): e4345.

Badger, J., J. M. Sauder, J. M. Adams, S. Antonysamy, K. Bain, M. G. Bergseid, S. G. Buchanan, M. D. Buchanan, Y. Batiyenko, J. A. Christopher, S. Emtage, A. Eroshkina, I. Feil, E. B. Furlong, K. S. Gajiwala, X. Gao, D. He, J. Hendle, A. Huber, K. Hoda, P. Kearins, C. Kissinger, B. Laubert, H. A. Lewis, J. Lin, K. Loomis, D. Lorimer, G. Louie, M. Maletic, C. D. Marsh, I. Miller, J. Molinari, H. J. Muller-Dieckmann, J. M. Newman, B. W. Noland, B. Pagarigan, F. Park, T. S. Peat, K. W. Post, S. Radojicic, A. Ramos, R. Romero, M. E. Rutter, W. E. Sanderson, K. D. Schwinn, J. Tresser, J. Winhoven, T. A. Wright, L. Wu, J. Xu and T. J. Harris (2005). "Structural analysis of a set of proteins resulting from a bacterial genomics project." Proteins **60**(4): 787-796.

Baldwin, A. J., K. Busse, A. M. Simm and D. D. Jones (2008). "Expanded molecular diversity generation during directed evolution by trinucleotide exchange (TriNEx)." Nucleic Acids Res **36**(13): e77.

Barabasi, A. L. and Z. N. Oltvai (2004). "Network biology: understanding the cell's functional organization." Nat Rev Genet **5**(2): 101-113.

Berardini, T. Z., D. Li, E. Huala, S. Bridges, S. Burgess, F. McCarthy, S. Carbon, M. C. Lewis SE, Abdulla A, Wood V, Feltrin E, Valle G, Chisholm RL, Fey P, Gaudet P, Kibbe W, Basu S, Bushmanova Y, Eilbeck K, Siegele DA, McIntosh B, Renfro D, Zweifel A, Hu JC, Ashburner M, Tweedie S, Alam-Faruque Y, Apweiler R, Auchinchloss A, Bairoch A, Barrell D, Binns D, Blatter MC, Bougueleret L, Boutet E, Breuza L, Bridge A, Browne P, Chan WM, Coudert E, Daugherty L, Dimmer E, Eberhardt R, Estreicher A, Famiglietti L, Ferro-Rojas S, Feuermann M, Foulger R, Gruaz-Gumowski N, Hinz U, Huntley R, Jimenez S, Jungo F, Keller G, Laiho K, Legge D, Lemercier P, Lieberherr D, Magrane M, O'Donovan C, Pedruzzi I, Poux S, Rivoire C, Roechert B, Sawford T, Schneider M, Stanley E, Stutz A, Sundaram S, Tognolli M, Xenarios I, Harris MA, Deegan JI, Ireland A, Lomax J, Jaiswal P, Chibucos M, Giglio MG, Wortman J, Hannick L, Madupu R, Botstein D, Dolinski K, Livstone MS, Oughtred R, Blake JA, Bult C, Diehl AD, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Sitnikov D, Collmer C, Torto-Alalibo T, Laulederkind S, Shimoyama M, Twigger S, D'Eustachio P, Matthews L, Balakrishnan R, Binkley G, Cherry JM, Christie KR, Costanzo MC, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Hong EL, Krieger CJ, Miyasato SR, Nash RS, Park J, Skrzypek MS, Weng S, Wong ED, Aslett M, Chan J, Kishore R, Sternberg P, Van Auke K, Khodiyar VK, Lovering

RC, Talmud PJ, Howe D, and M. Westerfield (2010). "The Gene Ontology in 2010: extensions and refinements." Nucleic Acids Res **38**(Database issue): D331-335.

Bligh, E. G. and W. J. Dyer (1959). "A rapid method of total lipid extraction and purification." Can J Biochem Physiol **37**(8): 911-917.

Boucher, Y., C. J. Douady, R. T. Papke, D. A. Walsh, M. E. Boudreau, C. L. Nesbo, R. J. Case and W. F. Doolittle (2003). "Lateral gene transfer and the origins of prokaryotic groups." Annu Rev Genet **37**: 283-328.

Boucher, Y., M. Kamekura and W. F. Doolittle (2004). "Origins and evolution of isoprenoid lipid biosynthesis in archaea." Mol Microbiol **52**(2): 515-527.

Brochier-Armanet, C., P. Forterre and S. Gribaldo (2011). "Phylogeny and evolution of the Archaea: one hundred genomes later." Curr Opin Microbiol **14**(3): 274-281.

Brustad, E. M. and F. H. Arnold (2011). "Optimizing non-natural protein function with directed evolution." Curr Opin Chem Biol **15**(2): 201-210.

Buchholz, F. and M. Bishop (2001). "LoxP-directed cloning: use of Cre recombinase as a universal restriction enzyme." Biotechniques **31**(4): 906-918.

Cadwell, R. C. and G. F. Joyce (1992). "Randomization of genes by PCR mutagenesis." PCR Methods Appl **2**(1): 28-33.

Chen, A., D. Zhang and C. D. Poulter (1993). "(S)-geranylgeranylglyceryl phosphate synthase. Purification and characterization of the first pathway-specific enzyme in archaebacterial membrane lipid biosynthesis." J Biol Chem **268**(29): 21701-21705.

Christianson, D. W. (2006). "Structural biology and chemistry of the terpenoid cyclases." Chem Rev **106**(8): 3412-3442.

Claren, J. (2008). "Design stabiler und katalytisch aktiver $(\beta\alpha)_8$-Barrel Enzyme durch Rekombination von $(\beta\alpha)_4$-Halbbarrel Domänen." (PhD thesis).

Claren, J., C. Malisi, B. Hocker and R. Sterner (2009). "Establishing wild-type levels of catalytic activity on natural and artificial (beta alpha)8-barrel protein scaffolds." Proc Natl Acad Sci U S A **106**(10): 3704-3709.

Cronan, J. E. (2003). "Bacterial membrane lipids: where do we stand?" Annu Rev Microbiol **57**: 203-224.

Daggett, K. A., M. Layer and T. A. Cropp (2009). "A general method for scanning unnatural amino acid mutagenesis." ACS Chem Biol **4**(2): 109-113.

De Rosa, M. and A. Gambacorta (1988). "The lipids of archaebacteria." Prog Lipid Res **27**(3): 153-175.

De Rosa, M., A. Gambacorta and A. Gliozzi (1986). "Structure, biosynthesis, and physicochemical properties of archaebacterial lipids." Microbiol Rev **50**(1): 70-80.

Doud, E. H., D. L. Perlstein, M. Wolpert, D. E. Cane and S. Walker (2011). "Two Distinct Mechanisms for TIM Barrel Prenyltransferases in Bacteria." J Am Chem Soc **133**(5): 1270-1273.

Friedberg, I., M. Jambon and A. Godzik (2006). "New avenues in protein function prediction." Protein Sci **15**(6): 1527-1529.

Gaudet, P., M. S. Livstone, S. E. Lewis and P. D. Thomas (2011). "Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium." Brief Bioinform **12**(5): 449-462.

Gerlt, J. A. and P. C. Babbitt (2000). "Can sequence determine function?" Genome Biol. **1**(5): reviews0005.

Gerth, M. L., W. M. Patrick and S. Lutz (2004). "A second-generation system for unbiased reading frame selection." Protein Eng Des Sel **17**(7): 595-602.

Ghosh, K. and G. D. Van Duyne (2002). "Cre-loxP biochemistry." Methods **28**(3): 374-383.

Glansdorff, N. (2000). "About the last common ancestor, the universal life-tree and lateral gene transfer: a reappraisal." Mol Microbiol **38**(2): 177-185.

Glansdorff, N., Y. Xu and B. Labedan (2008). "The Last Universal Common Ancestor: emergence, constitution and genetic legacy of an elusive forerunner." Biol direct **3**: 29.

Goldsmith, M. and D. S. Tawfik (2012). "Directed enzyme evolution: beyond the low-hanging fruit." Curr Opin Struct Biol **22**(4): 406-412.

Guldan, H., F. M. Matysik, M. Bocola, R. Sterner and P. Babinger (2011). "Functional Assignment of an Enzyme that Catalyzes the Synthesis of an Archaea-Type Ether Lipid in Bacteria." Angew Chem Int Ed Engl **50**: 8188-8191.

Guldan, H., R. Sterner and P. Babinger (2008). "Identification and characterization of a bacterial glycerol-1-phosphate dehydrogenase: Ni(2+)-dependent AraM from Bacillus subtilis." Biochemistry **47**(28): 7376-7384.

Haapa, S., S. Taira, E. Heikkinen and H. Savilahti (1999). "An efficient and accurate integration of mini-Mu transposons in vitro: a general methodology for functional genetic analysis and molecular biology applications." Nucleic Acids Res **27**(13): 2777-2784.

Haldimann, A. and B. L. Wanner (2001). "Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria." J Bacteriol **183**(21): 6384-6393.

Han, J. S. and K. Ishikawa (2005). "Active site of Zn^{2+}-dependent sn-glycerol-1-phosphate dehydrogenase from *Aeropyrum pernix* K1." Archaea **1**(5): 311-317.

Hanahan, D. (1983). "Studies on transformation of *Escherichia coli* with plasmids." J Mol Biol **166**(4): 557-580.

Hellinga, H. W. (1997). "Rational protein design: combining theory and experiment." Proc Natl Acad Sci U S A **94**(19): 10015-10017.

Hemmi, H., K. Shibuya, Y. Takahashi, T. Nakayama and T. Nishino (2004). "(S)-2,3-Di-O-geranylgeranylglyceryl phosphate synthase from the thermoacidophilic archaeon *Sulfolobus solfataricus*. Molecular cloning and characterization of a membrane-intrinsic prenyltransferase involved in the biosynthesis of archaeal ether-linked membrane lipids." J Biol Chem **279**(48): 50197-50203.

Henn-Sax, M., R. Thoma, S. Schmidt, M. Hennig, K. Kirschner and R. Sterner (2002). "Two (betaalpha)(8)-barrel enzymes of histidine and tryptophan biosynthesis have similar reaction mechanisms and common strategies for protecting their labile substrates." Biochemistry **41**(40): 12032-12042.

Hunter, S., P. Jones, A. Mitchell, R. Apweiler, T. K. Attwood, A. Bateman, T. Bernard, D. Binns, P. Bork, S. Burge, E. de Castro, P. Coggill, M. Corbett, U. Das, L. Daugherty, L. Duquenne, R. D. Finn, M. Fraser, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, C. McMenamin, H. Mi, P. Mutowo-Muellenet, N. Mulder, D. Natale, C. Orengo, S. Pesseat, M. Punta, A. F. Quinn, C. Rivoire, A. Sangrador-Vegas, J. D. Selengut, C. J. Sigrist, M. Scheremetjew, J. Tate, M. Thimmajanarthanan, P. D. Thomas, C. H. Wu, C. Yeats and S. Y. Yong (2012). "InterPro in 2011: new developments in the family and domain prediction database." Nucleic Acids Res **40**(Database issue): D306-312.

Huson, D. H. and D. Bryant (2006). "Application of phylogenetic networks in evolutionary studies." Mol Biol Evol **23**(2): 254-267.

Hyde, C. C., S. A. Ahmed, E. A. Padlan, E. W. Miles and D. R. Davies (1988). "Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from Salmonella typhimurium." J Biol Chem **263**(33): 17857-17871.

Jackel, C., P. Kast and D. Hilvert (2008). "Protein design by directed evolution." Annu Rev Biophys **37**: 153-173.

Jahn, U., R. Summons, H. Sturt, E. Grosjean and H. Huber (2004). "Composition of the lipids of Nanoarchaeum equitans and their origin from its host Ignicoccus sp. strain KIN4/I." Archives of microbiology **182**(5): 404-413.

Jia, F., S. S. Gampala, A. Mittal, Q. Luo and C. D. Rock (2009). "Cre-lox univector acceptor vectors for functional screening in protoplasts: analysis of Arabidopsis donor cDNAs encoding ABSCISIC ACID INSENSITIVE1-like protein phosphatases." Plant Mol Biol **70**(6): 693-708.

Jones, D. D. (2005). "Triplet nucleotide removal at random positions in a target gene: the tolerance of TEM-1 beta-lactamase to an amino acid deletion." Nucleic Acids Res **33**(9): e80.

Kalyanaraman, C., H. J. Imker, A. A. Fedorov, E. V. Fedorov, M. E. Glasner, P. C. Babbitt, S. C. Almo, J. A. Gerlt and M. P. Jacobson (2008). "Discovery of a dipeptide epimerase enzymatic function guided by homology modeling and virtual screening." Structure **16**(11): 1668-1677.

Kates, M. (1986). Techniques of lipidology : isolation, analysis, and identification of lipids. Laboratory techniques in biochemistry and molecular biology v. 3, pt. 2. Amsterdam ; New York, Elsevier**:** 464.

Kates, M. (1993). Membrane lipids in Archaea. The biochemistry of archaea. M. Kates, D. J. Kushner and A. T. Matheson. Amsterdam, Elsevier**:** 261-295.

Kayushin, A. L., M. D. Korosteleva, A. I. Miroshnikov, W. Kosch, D. Zubov and N. Piel (1996). "A convenient approach to the synthesis of trinucleotide phosphoramidites--synthons for the generation of oligonucleotide/peptide libraries." Nucleic Acids Res **24**(19): 3748-3755.

Kobayashi, K., S. D. Ehrlich, A. Albertini, G. Amati, K. K. Andersen, M. Arnaud, K. Asai, S. Ashikaga, S. Aymerich, P. Bessieres, F. Boland, S. C. Brignell, S. Bron, K. Bunai, J. Chapuis, L. C. Christiansen, A. Danchin, M. Debarbouille, E. Dervyn, E. Deuerling, K. Devine, S. K. Devine, O. Dreesen, J. Errington, S. Fillinger, S. J. Foster, Y. Fujita, A. Galizzi, R. Gardan, C. Eschevins, T. Fukushima, K. Haga, C. R. Harwood, M. Hecker, D. Hosoya, M. F. Hullo, H. Kakeshita, D. Karamata, Y. Kasahara, F. Kawamura, K. Koga, P. Koski, R. Kuwana, D. Imamura, M. Ishimaru, S. Ishikawa, I. Ishio, D. Le Coq, A. Masson, C. Mauel, R. Meima, R. P. Mellado, A. Moir, S. Moriya, E. Nagakawa, H. Nanamiya, S. Nakai, P. Nygaard, M. Ogura, T. Ohanan, M. O'Reilly, M. O'Rourke, Z. Pragai, H. M. Pooley, G. Rapoport, J. P. Rawlins, L. A. Rivas, C. Rivolta, A. Sadaie, Y. Sadaie, M. Sarvas, T. Sato, H. H. Saxild, E. Scanlan, W. Schumann, J. F. Seegers, J. Sekiguchi, A. Sekowska, S. J. Seror, M. Simon, P. Stragier, R. Studer, H. Takamatsu, T. Tanaka, M. Takeuchi, H. B. Thomaides, V. Vagner, J. M. van Dijl, K. Watabe, A. Wipat, H. Yamamoto, M. Yamamoto, Y. Yamamoto, K. Yamane, K. Yata, K. Yoshida, H. Yoshikawa, U. Zuber and N. Ogasawara (2003). "Essential *Bacillus subtilis* genes." Proc Natl Acad Sci U S A **100**(8): 4678-4683.

Koga, Y., T. Kyuragi, M. Nishihara and N. Sone (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**(1): 54-63.

Koga, Y. and H. Morii (2005). "Recent advances in structural research on ether lipids from archaea including comparative and physiological aspects." Biosci Biotechnol Biochem **69**(11): 2019-2034.

Koga, Y. and H. Morii (2007). "Biosynthesis of ether-type polar lipids in archaea and evolutionary considerations." Microbiol Mol Biol Rev **71**(1): 97-120.

Koga, Y. and M. Nakano (2008). "A dendrogram of archaea based on lipid component parts composition and its relationship to rRNA phylogeny." Syst Appl Microbiol **31**(3): 169-182.

Koga, Y., M. Nishihara, H. Morii and M. Akagawa-Matsushita (1993). "Ether polar lipids of methanogenic bacteria: structures, comparative aspects, and biosyntheses." Microbiol Rev **57**(1): 164-182.

Krieger, E., T. Darden, S. B. Nabuurs, A. Finkelstein and G. Vriend (2004). "Making optimal use of empirical energy functions: force-field parameterization in crystal space." Proteins **57**(4): 678-683.

Leaver-Fay, A., M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y. E. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popovic, J. J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker and P. Bradley (2011). "ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules." Methods Enzymol **487**: 545-574.

Leopoldseder, S., J. Claren, C. Jurgens and R. Sterner (2004). "Interconverting the catalytic activities of (betaalpha)(8)-barrel enzymes from different metabolic pathways: sequence requirements and molecular analysis." J. Mol. Biol. **337**(4): 871-879.

Liang, P. H., T. P. Ko and A. H. Wang (2002). "Structure, mechanism and function of prenyltransferases." Eur J Biochem **269**(14): 3339-3354.

Linde, M. (2013). "Identifizierung und Charakterisierung der HepGP-prozessierenden Enzyme aus *Bacillus subtilis*." (master thesis).

Liu, C. C. and P. G. Schultz (2010). "Adding new chemistries to the genetic code." Annu Rev Biochem **79**: 413-444.

Liu, J. and T. A. Cropp (2012). "Experimental methods for scanning unnatural amino acid mutagenesis." Methods Mol Biol **794**: 187-197.

Loewenstein, Y., D. Raimondo, O. C. Redfern, J. Watson, D. Frishman, M. Linial, C. Orengo, J. Thornton and A. Tramontano (2009). "Protein function annotation by homology-based inference." Genome Biol **10**(2): 207.

Lombard, J., P. Lopez-Garcia and D. Moreira (2012). "The early evolution of lipid membranes and the three domains of life." Nat Rev Microbiol **10**(7): 507-515.

Lombard, J., P. Lopez-Garcia and D. Moreira (2012). "Phylogenomic investigation of phospholipid synthesis in archaea." Archaea **2012**: 630910.

Long, S. B., P. J. Casey and L. S. Beese (2002). "Reaction path of protein farnesyltransferase at atomic resolution." Nature **419**(6907): 645-650.

Luckow, B., R. Renkawitz and G. Schutz (1987). "A new method for constructing linker scanning mutants." Nucleic Acids Res **15**(2): 417-429.

Lukk, T., A. Sakai, C. Kalyanaraman, S. D. Brown, H. J. Imker, L. Song, A. A. Fedorov, E. V. Fedorov, R. Toro, B. Hillerich, R. Seidel, Y. Patskovsky, M. W. Vetting, S. K. Nair, P. C. Babbitt, S. C. Almo, J. A. Gerlt and M. P. Jacobson (2012). "Homology models guide discovery of diverse enzyme specificities among dipeptide epimerases in the enolase superfamily." Proc Natl Acad Sci U S A **109**(11): 4122-4127.

Matsumi, R., H. Atomi, A. J. Driessen and J. van der Oost (2011). "Isoprenoid biosynthesis in Archaea--biochemical and evolutionary implications." Res Microbiol **162**(1): 39-52.

Mehta, M. M., S. Liu and J. J. Silberg (2012). "A transposase strategy for creating libraries of circularly permuted proteins." Nucleic Acids Res **40**(9): e71.

Metcalf, W. W., W. Jiang and B. L. Wanner (1994). "Use of the rep technique for allele replacement to construct new Escherichia coli hosts for maintenance of R6K gamma origin plasmids at different copy numbers." Gene **138**(1-2): 1-7.

Morii, H., H. Yagi, H. Akutsu, N. Nomura, Y. Sako and Y. Koga (1999). "A novel phosphoglycolipid archaetidyl(glucosyl)inositol with two sesterterpanyl chains from the aerobic hyperthermophilic archaeon Aeropyrum pernix K1." Biochim Biophys Acta **1436**(3): 426-436.

Murakami, H., T. Hohsaka and M. Sisido (2002). "Random insertion and deletion of arbitrary number of bases for codon-based random mutation of DNAs." Nat Biotechnol **20**(1): 76-81.

Nagano, N., C. A. Orengo and J. M. Thornton (2002). "One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions." J Mol Biol **321**(5): 741-765.

Nakatani, Y., N. Ribeiro, S. Streiff, L. Desaubry and G. Ourisson (2012). "Search for the most primitive membranes: some remaining problems." Orig Life Evol Biosph **42**(5): 497-501.

Nemoto, N., T. Oshima and A. Yamagishi (2003). "Purification and characterization of geranylgeranylglyceryl phosphate synthase from a thermoacidophilic archaeon, *Thermoplasma acidophilum*." J Biochem (Tokyo) **133**(5): 651-657.

Nie, L., J. J. Lavinder, M. Sarkar, K. Stephany and T. J. Magliery (2011). "Synthetic approach to stop-codon scanning mutagenesis." J Am Chem Soc **133**(16): 6177-6186.

Ogura, K. and T. Koyama (1998). "Enzymatic Aspects of Isoprenoid Chain Elongation." Chem Rev **98**(4): 1263-1276.

Oldfield, E. and F. Y. Lin (2012). "Terpene biosynthesis: modularity rules." Angew Chem Int Ed Engl **51**(5): 1124-1137.

Olsen, G. J. and C. R. Woese (1993). "Ribosomal RNA: a key to phylogeny." FASEB J **7**(1): 113-123.

Ono, A., A. Matsuda, J. Zhao and D. V. Santi (1995). "The synthesis of blocked triplet-phosphoramidites and their use in mutagenesis." Nucleic Acids Res **23**(22): 4677-4682.

Pan, J. J., G. Ramamoorthy and C. D. Poulter (2013). "Dependence of the Product Chain-Length on Detergents for Long-Chain E-Polyprenyl Diphosphate Synthases." Biochemistry.

Payandeh, J., M. Fujihashi, W. Gillon and E. F. Pai (2006). "The crystal structure of (S)-3-O-geranylgeranylglyceryl phosphate synthase reveals an ancient fold for an ancient enzyme." J Biol Chem **281**(9): 6070-6078.

Payandeh, J. and E. F. Pai (2007). "Enzyme-driven speciation: crystallizing Archaea via lipid capture." J Mol Evol **64**(3): 364-374.

Peisajovich, S. G. and D. S. Tawfik (2007). "Protein engineers turned evolutionists." Nat Methods **4**(12): 991-994.

Pereto, J., P. Lopez-Garcia and D. Moreira (2004). "Ancestral lipid biosynthesis and early membrane evolution." Trends Biochem Sci **29**(9): 469-477.

Petit, M. A., E. Dervyn, M. Rose, K. D. Entian, S. McGovern, S. D. Ehrlich and C. Bruand (1998). "PcrA is an essential DNA helicase of *Bacillus subtilis* fulfilling functions both in repair and rolling-circle replication." Mol Microbiol **29**(1): 261-273.

Poulter, C. D., J. C. Argyle and E. A. Mash (1978). "Farnesyl pyrophosphate synthetase. Mechanistic studies of the 1'-4 coupling reaction with 2-fluorogeranyl pyrophosphate." J Biol Chem **253**(20): 7227-7233.

Punta, M., P. C. Coggill, R. Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E. L. Sonnhammer, S. R. Eddy, A. Bateman and R. D. Finn (2012). "The Pfam protein families database." Nucleic Acids Res **40**(Database issue): D290-301.

Radivojac, P., W. T. Clark, T. R. Oron, A. M. Schnoes, T. Wittkop, A. Sokolov, K. Graim, C. Funk, K. Verspoor, A. Ben-Hur, G. Pandey, J. M. Yunes, A. S. Talwalkar, S. Repo, M. L. Souza, D. Piovesan, R. Casadio, Z. Wang, J. Cheng, H. Fang, J. Gough, P. Koskinen, P. Toronen, J. Nokso-Koivisto, L. Holm, D. Cozzetto, D. W. Buchan, K. Bryson, D. T. Jones, B. Limaye, H. Inamdar, A. Datta, S. K. Manjari, R. Joshi, M. Chitale, D. Kihara, A. M. Lisewski, S. Erdin, E. Venner, O. Lichtarge, R. Rentzsch, H.

Yang, A. E. Romero, P. Bhat, A. Paccanaro, T. Hamp, R. Kassner, S. Seemayer, E. Vicedo, C. Schaefer, D. Achten, F. Auer, A. Boehm, T. Braun, M. Hecht, M. Heron, P. Honigschmid, T. A. Hopf, S. Kaufmann, M. Kiening, D. Krompass, C. Landerer, Y. Mahlich, M. Roos, J. Bjorne, T. Salakoski, A. Wong, H. Shatkay, F. Gatzmann, I. Sommer, M. N. Wass, M. J. Sternberg, N. Skunca, F. Supek, M. Bosnjak, P. Panov, S. Dzeroski, T. Smuc, Y. A. Kourmpetis, A. D. van Dijk, C. J. ter Braak, Y. Zhou, Q. Gong, X. Dong, W. Tian, M. Falda, P. Fontana, E. Lavezzo, B. Di Camillo, S. Toppo, L. Lan, N. Djuric, Y. Guo, S. Vucetic, A. Bairoch, M. Linial, P. C. Babbitt, S. E. Brenner, C. Orengo, B. Rost, S. D. Mooney and I. Friedberg (2013). "A large-scale evaluation of computational protein function prediction." Nat Methods **10**(3): 221-227.

Ren, F., X. Feng, T. P. Ko, C. H. Huang, Y. Hu, H. C. Chan, Y. L. Liu, K. Wang, C. C. Chen, X. Pang, M. He, Y. Li, E. Oldfield and R. T. Guo (2013). "Insights into TIM-barrel prenyl transferase mechanisms: crystal structures of PcrB from Bacillus subtilis and Staphylococcus aureus." Chembiochem **14**(2): 195-199.

Ren, F., T. P. Ko, X. Feng, C. H. Huang, H. C. Chan, Y. Hu, K. Wang, Y. Ma, P. H. Liang, A. H. Wang, E. Oldfield and R. T. Guo (2012). "Insights into the mechanism of the antibiotic-synthesizing enzyme MoeO5 from crystal structures of different complexes." Angew Chem Int Ed Engl **51**(17): 4157-4160.

Samish, I., C. M. MacDermaid, J. M. Perez-Aguilar and J. G. Saven (2011). "Theoretical and computational protein design." Annu Rev Phys Chem **62**: 129-149.

Schnoes, A. M., S. D. Brown, I. Dodevski and P. C. Babbitt (2009). "Annotation error in public databases: misannotation of molecular function in enzyme superfamilies." PLoS Comput Biol **5**(12): e1000605.

Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson and D. G. Higgins (2011). "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega." Mol Syst Biol **7**: 539.

Smoot, M. E., K. Ono, J. Ruscheinski, P. L. Wang and T. Ideker (2011). "Cytoscape 2.8: new features for data integration and network visualization." Bioinformatics **27**(3): 431-432.

Soderberg, T., A. Chen and C. D. Poulter (2001). "Geranylgeranylglyceryl phosphate synthase. Characterization of the recombinant enzyme from *Methanobacterium thermoautotrophicum*." Biochemistry **40**(49): 14847-14854.

Stemmer, W. P. (1994). "Rapid evolution of a protein in vitro by DNA shuffling." Nature **370**(6488): 389-391.

Sterner, R., A. Dahm, B. Darimont, A. Ivens, W. Liebl and K. Kirschner (1995). "(Beta alpha)8-barrel proteins of tryptophan biosynthesis in the hyperthermophile Thermotoga maritima." EMBO J **14**(18): 4395-4402.

Sterner, R. and B. Hocker (2005). "Catalytic versatility, stability, and evolution of the (betaalpha)8-barrel enzyme fold." Chemical reviews **105**(11): 4038-4055.

Tachibana, A., Y. Yano, S. Otani, N. Nomura, Y. Sako and M. Taniguchi (2000). "Novel prenyltransferase gene encoding farnesylgeranyl diphosphate synthase from a hyperthermophilic archaeon, Aeropyrum pernix. Molecularevolution with alteration in product specificity." Eur J Biochem **267**(2): 321-328.

Takahashi, I. and K. Ogura (1981). "Farnesyl pyrophosphate synthetase from Bacillus subtilis." J Biochem **89**(5): 1581-1587.

Takahashi, I. and K. Ogura (1982). "Prenyltransferases of Bacillus subtilis: undecaprenyl pyrophosphate synthetase and geranylgeranyl pyrophosphate synthetase." J Biochem **92**(5): 1527-1537.

Takahashi, I., K. Ogura and S. Seto (1980). "Heptaprenyl pyrophosphate synthetase from Bacillus subtilis." J Biol Chem **255**(10): 4539-4543.

Tarshis, L. C., P. J. Proteau, B. A. Kellogg, J. C. Sacchettini and C. D. Poulter (1996). "Regulation of product chain length by isoprenyl diphosphate synthases." Proc Natl Acad Sci U S A **93**(26): 15018-15023.

Tee, K. L. and T. S. Wong (2013). "Polishing the craft of genetic diversity creation in directed evolution." Biotechnol Adv.

Ulrih, N. P., D. Gmajner and P. Raspor (2009). "Structural and physicochemical properties of polar lipids from thermophilic archaea." Appl Microbiol Biotechnol **84**(2): 249-260.

van de Vossenberg, J. L., A. J. Driessen and W. N. Konings (1998). "The essence of being extremophilic: the role of the unique archaeal membrane lipids." Extremophiles **2**(3): 163-170.

Van den Brulle, J., M. Fischer, T. Langmann, G. Horn, T. Waldmann, S. Arnold, M. Fuhrmann, O. Schatz, T. O'Connell, D. O'Connell, A. Auckenthaler and H. Schwer (2008). "A novel solid phase technology for high-throughput gene synthesis." Biotechniques **45**(3): 340-343.

Van Duyne, G. D. (2001). "A structural view of cre-loxp site-specific recombination." Annu Rev Biophys Biomol Struct **30**: 87-104.

van Meer, G., D. R. Voelker and G. W. Feigenson (2008). "Membrane lipids: where they are and how they behave." Nat Rev Mol Cell Biol **9**(2): 112-124.

Vega, M. C., E. Lorentzen, A. Linden and M. Wilmanns (2003). "Evolutionary markers in the (beta/alpha)8-barrel fold." Curr Opin Chem Biol **7**(6): 694-701.

Virnekas, B., L. Ge, A. Pluckthun, K. C. Schneider, G. Wellnhofer and S. E. Moroney (1994). "Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis." Nucleic Acids Res **22**(25): 5600-5607.

Wächtershäuser, G. (2003). "From pre-cells to Eukarya--a tale of two lipids." Mol Microbiol **47**(1): 13-22.

Wang, K. C. and S. Ohnuma (2000). "Isoprenyl diphosphate synthases." Biochim Biophys Acta **1529**(1-3): 33-48.

Waterhouse, A. M., J. B. Procter, D. M. Martin, M. Clamp and G. J. Barton (2009). "Jalview Version 2--a multiple sequence alignment editor and analysis workbench." Bioinformatics **25**(9): 1189-1191.

Woese, C. (1998). "The universal ancestor." Proc Natl Acad Sci U S A **95**(12): 6854-6859.

Zellner, H., M. Staudigel, T. Trenner, M. Bittkowski, V. Wolowski, C. Icking and R. Merkl (2012). "PresCont: predicting protein-protein interfaces utilizing four residue properties." Proteins **80**(1): 154-168.

Zhang, D. and C. D. Poulter (1993). "Biosynthesis of Archaeal Ether Lipids. Formation of Ether Linkages by Prenyltransferases." J Am Chem Soc(115): 1270 - 1277.

4 List of publications

A. **Peterhoff, D., Zellner, H., Guldan, H., Merkl, R., Sterner, R., Babinger, P.** (2012). Dimerization determines substrate specificity of a bacterial prenyltransferase. Chembiochem. 13(9):1297-303. doi: 10.1002/cbic.201200127.

B. **Peterhoff, D., Beer, B., Rajendran, C., Kumpula, E.-P., Kapetaniou, E., Guldan, H., Wierenga, R. K., Sterner, R., Babinger, P.** (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(4):885-99. doi: 10.1111/mmi.12596.

5 Personal contribution

A. The experimental work was performed by myself. Computational analyses were performed by Hermann Zellner. Harald Guldan provided some strains. The work was supervised by Patrick Babinger and Reinhard Sterner, and the publication was written by myself, Patrick Babinger, Rainer Merkl, Hermann Zellner and Reinhard Sterner.

B. The experimental work was performed by myself. Barbara Beer and Harald Guldan provided some strains and performed preliminary experiments. Esa-Pekka Kumpula and Evangelia Kapetaniou crystallized PcrB from *G. kaustophilus* and collected the X-ray datasets. Chitra Rajendran collected the X-ray datasets from the GGGPS from *F. johnsoniae* and *M. thermautotrophicus*, solved all X-ray structures and collected static light scattering data. The work was supervised by Patrick Babinger, Reinhard Sterner and Rikkert Wierenga, and the publication was written by myself, Patrick Babinger and Reinhard Sterner.

6 Publications

Publication A

Dimerization determines substrate specificity of a bacterial prenyltransferase.

David Peterhoff, Hermann Zellner, Harald Guldan, Rainer Merkl, Reinhard Sterner, Patrick Babinger (2012).

Chembiochem, Volume 13, Issue 9, pages 1297–1303.

CHEMBIOCHEM FULL PAPERS

DOI: 10.1002/cbic.201200127

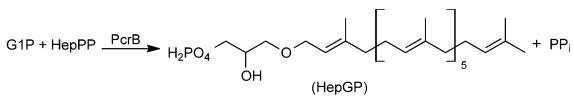
Dimerization Determines Substrate Specificity of a Bacterial Prenyltransferase

David Peterhoff, Hermann Zellner, Harald Guldan, Rainer Merkl, Reinhard Sterner, and Patrick Babinger*[a]

We have identified the native dimer interface of heptaprenylglyceryl phosphate synthase PcrB from the bacterium *Bacillus subtilis* and analyzed the significance of oligomer formation for stability and catalytic activity. Computational methods predicted two different surface regions of the PcrB protomer that could be responsible for dimer formation. These bona fide interfaces were assessed both in silico and experimentally by the introduction of amino acid substitutions that led to monomerization, and by incorporation of an unnatural amino acid to allow cross-linking of the two protomers. The results showed that, in contrast to previous assumptions, PcrB uses the same interface for dimerization as the homologous geranylgeranylglyceryl phosphate synthase from Archaea. Thermal unfolding demonstrated that the monomeric proteins are only slightly less stable than wild-type PcrB. However, activity assays showed that monomerization limits the length of accepted polyprenyl pyrophosphates to three isoprene units, whereas the native PcrB substrate contains seven isoprene entities. We provide a plausible hypothesis as to how dimerization determines substrate specificity of PcrB.

Introduction

The enzyme PcrB from Gram-positive bacteria[1] catalyzes the formation of an ether bond between *sn*-glycerol-1-phosphate (G1P) and heptaprenyl pyrophosphate (HepPP), a polyprenylic compound with seven isoprene units[2] (Scheme 1).



Scheme 1. PcrB catalyzes the formation of an ether bond between *sn*-glycerol-1-phosphate (G1P) and heptaprenyl pyrophosphate (HepPP) to yield heptaprenylglyceryl phosphate (HepGP) and pyrophosphate (PP_i).

As demonstrated for *Bacillus subtilis*,[2b] the product, heptaprenylglyceryl phosphate (HepGP), is subsequently dephosphorylated and acetylated in vivo. The physiological role of these ether lipids in the phylogenetic domain of Bacteria is unknown. Interestingly, however, PcrB is a homologue of the archaeal geranylgeranylglyceryl phosphate synthase (GGGPS), which uses geranylgeranyl pyrophosphate (GGPP, four isoprene units) as substrate. The product geranylgeranylglyceryl phosphate (GGGP) is subsequently converted to the ether lipids which constitute the archaeal cell membrane.[1b,3] It has been postulated that the indole ring of a conserved tryptophan residue acts as a "hydrocarbon ruler" in the binding groove and is responsible for substrate length limitation in GGGPS compared to PcrB.[1b] In support of this hypothesis, the substitution of Trp99 with Ala (small aliphatic side chain) allowed binding and turnover of HepPP by *Archaeoglobus fulgidus* GGGPS (afGGGPS).[2b] Along the same lines, PcrB can process GGPP and other polyprenyl pyrophosphates with fewer isoprene units, albeit with lower efficiency.[2b,4]

Both PcrB and GGGPS form dimers of identical subunits that adopt the frequently encountered $(\beta\alpha)_8$-barrel protein fold.[1,2b,5] We were interested to elucidate the significance of oligomer formation in this enzyme family, which can be manifold.[6] Dimerization has been shown to 1) improve the stability of proteins,[7] 2) influence their catalytic efficiency,[8] 3) facilitate the regulation of enzymatic activity,[9] and 4) allow substrate channeling.[10] Moreover, it has been demonstrated in a few cases that oligomer formation can be an important determinant of substrate specificity.[11]

In order to study the impact of dimerization on stability and enzymatic specificity of the PcrB/GGGPS family, we decided to monomerize the PcrB complex by protein engineering. However, inspection of the crystal structure suggested that the published dimer interface is non-native, and an in silico analysis identified two alternative interfaces. We introduced amino acid exchanges into each candidate interface to inhibit dimer formation. The characterization of the different variants and cross-linking experiments with the unnatural amino acid *p*-azido-L-phenylalanine (AzF) unambiguously identified the native interface and demonstrated that the

[a] D. Peterhoff, Dr. H. Zellner, Dr. H. Guldan, Dr. R. Merkl, Prof. Dr. R. Sterner, Dr. P. Babinger
Institute of Biophysics and Physical Biochemistry
University of Regensburg
Universitätsstrasse 31, 93053 Regensburg (Germany)
E-mail: patrick.babinger@biologie.uni-r.de

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cbic.201200127.

quaternary structure is conserved between GGGPS and PcrB. Whereas the thermal stability of the monomeric variants was only slightly reduced compared to dimeric wild-type PcrB, the length of the accepted polyprenyl pyrophosphates was limited to a maximum of three isoprene units. These findings show that dimerization of PcrB, and most likely also GGGPS, is a prerequisite for the binding of the natural substrates.

Results

In silico assessment of potential PcrB dimerization interfaces

The starting point for our in silico analysis was the crystal structure of dimeric PcrB from *B. subtilis*, which has been determined at 1.85 Å resolution (PDB ID: 1VIZ).[1a] Due to symmetry considerations, three putative interfaces (I1, I2, and I3) can be deduced from the protein crystal (Figure 1), which correspond to the dimers C1, C2, and C3, respectively. According to the PDB dataset and the related PDBsum page,[12] I1 is the native interface. However, relative to the protein size, this interface is extremely small and differs from the dimerization interface of the homologous enzyme afGGGPS (PDB IDs: 2F6U and 2F6X),[1b] which corresponds to I3. For a quantitative comparison of I1, I2, and I3, we used our software PresCont,[13] which determines for each surface residue of a given protein the probability of belonging to an interface. According to PresCont, I2 and I3 are more likely the native interface than I1 (Figure S1). In order to support this finding by a second independent in silico method, we utilized the Rosetta force field[14] and computed the stability of the three putative complexes C1, C2, and C3. C3 had the lowest (best) total energy *E*, whereas C2 and C1 scored 14 and 22 Rosetta energy units (REU) worse than C3. In summary, these in silico results suggested C3 as the most probable complex and C2 as a less probable alternative. Thus we concentrated on I2 and I3 when testing our computational predictions by site-directed mutagenesis experiments. Here, our rationale was that disturbance of the native interface would result in monomerization of PcrB.

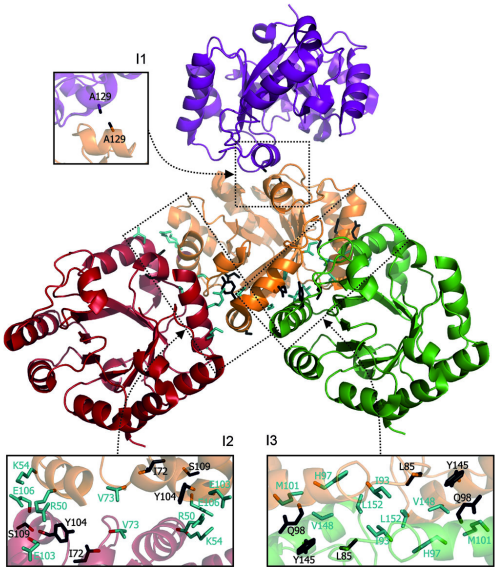


Figure 1. The crystal structure of PcrB suggests three putative dimerization interfaces (I1, I2, and I3). The boxes depict the contact areas in detail. In order to identify the native interface, we introduced electrostatic repulsion by substituting the residues shown in cyan. Residues which were chosen for cross-linking experiments and substituted by AzF are shown in black.

Selection of amino acid exchanges to monomerize PcrB

Visual inspection of I2 suggested that salt bridges formed by the residue pairs R50/E103 and K54/E106, which are located at each side of the two-fold symmetry axis of C2, might be crucial for the stability of this putative dimer (Figure 1). In order to loosen the interface by introducing electrostatic repulsion,[7] we considered individual substitutions of R50 and K54 with E or D, and E103 and E106 with K or R. In order to avoid the introduction of steric clashes within the individual protomers, we assessed all possible substitutions with the mutation wizard of PyMOL,[15] and consequently decided to experimentally test the effect of the exchanges R50E, K54E, and E103K. We further noted that V73 is located close to the dimer symmetry axis of I2 and thus faces V73 of the second protomer. We therefore expected a strong electrostatic repulsion by replacing V73 with a negatively charged residue. As PyMOL signaled a steric clash for the V73E exchange, we opted for testing V73D. Using Rosetta,[14] we determined the impact of the considered substitutions on complex stability in terms of $\Delta\Delta E$ values (see the Experimental Section). Rosetta calculated a minor destabilization of the C2 complex upon introduction of the single exchanges R50E or E103K, while the substitutions K54E and V73D were predicted to be slightly stabilizing (Figure S2).

Visual inspection of I3 led us to expect that replacement of the hydrophobic residues I93, H97, V148, and L152 with negatively charged glutamate would weaken the dimer interface by introducing electrostatic repulsion (Figure 1). As M101 faces the positively charged R147 of the complementary protomer, we opted for the M101K exchange to maximize the repulsion effect. For all but L152E, the $\Delta\Delta E$ values calculated by Rosetta predicted destabilization of C3 (Figure S2). Interestingly, the mean $\Delta\Delta E$ value was significantly larger for I3 (3.75) than for I2 (0.38), thus indicating that the C3 interface would be disturbed more severely by the considered single exchanges than the interface of C2.

Monomerizing amino acid substitutions and cross-linking experiments identify I3 as the native interface

The nine amino acid exchanges R50E, K54E, V73D, E103K (for I2) and I93E, H97E, M101K, V148E, L152E (for I3) were intro-

duced individually into PcrB by site-directed mutagenesis. Following expression of the mutants in *Escherichia coli*, the recombinant proteins were purified to homogeneity (Figure S3). The purified M101K variant was prone to aggregation and therefore could not be characterized. For all other variants, the oligomerization state was analyzed by size exclusion chromatography. The elution profiles showed that variants with exchanges at I2 were still dimeric, whereas proteins with substitutions at I3 formed monomers (Figure 2). These findings suggested that I3 is the native interface of PcrB.

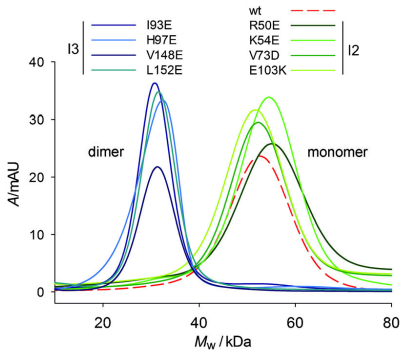


Figure 2. Analytical gel filtration chromatography of wild-type PcrB and variants carrying single amino acid exchanges. Protein was applied at a subunit concentration of 7–10 μM. Elution from BioSep-SEC-S-3000 column resulted in an approximately eightfold dilution of the protein. The runs were performed on different days, and the elution volume profiles were converted to molecular weight profiles for normalization based on updated calibration runs, resulting in an inversion of the *x*-axis. The calculated molecular weight of the PcrB protomer is 26.3 kDa.

It has been shown that protein–protein interactions can be analyzed by site-specifically incorporated unnatural amino acids with UV-induced cross-linking properties.[16] Based on these findings, we used AzF[16b] (Figure 3A) to confirm the results of our mutational analysis. The rationale was that the two protomers of PcrB can only be cross-linked if AzF is incorporated in the native interface. By using the molecular modeling software Yasara,[17] AzF was introduced in silico at several candidate sites. Each potential substitution was then evaluated to minimize the resulting structural perturbation. For this purpose, AzF was compared with the native amino acid for 1) bulkiness, 2) physicochemical properties of the side chain, and 3) calculated effects on complex stability. Based on the Yasara analysis, we selected residues I72, Y104, and S109 (I2), L85, Q98, and Y145 (I3), and A129 (I1) for substitution with AzF (Figure 1). Residue A193, for which PresCont determined a low interface probability, was used as a negative control. An *E. coli* strain expressing a designed orthogonal $tRNA^{AzF}_{CUA}$/aminoacyl-tRNA synthetase pair[18] was used to incorporate AzF into PcrB at substituted amber stop codons at the selected positions. The resulting PcrB variants were subjected to cross-linking experiments in vitro and in *E. coli* cells. Western blot analysis showed that the protomers were cross-linked when Q98 (interface I3) was replaced by AzF (Figure 3B). All other variants as well as the wild-type protein remained monomeric after UV irradiation (Figure S4). These results confirm that the native interface of the PcrB dimer is I3, which corresponds to the dimer interface of the homologous archaeal afGGGPS.[1b]

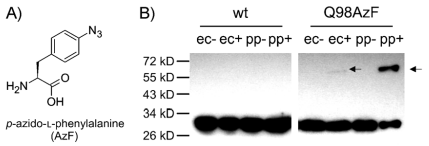


Figure 3. UV-induced cross-linking of dimeric PcrB incorporation of AzF. A) Structure of AzF. B) *E. coli* cells expressing the *pcrB* gene (ec) or purified protein (pp) samples were irradiated (+) with UV light ($\lambda = 366$ nm) or left untreated (−). PcrB monomers and cross-linked dimers (arrows) were detected by Western blot analysis. PcrB variant Q98AzF protomers could be cross-linked, whereas no cross-link was observed for wild-type (wt) PcrB.

Monomerization of PcrB has a minor effect on stability, but limits the size of accepted polyprenyl substrates

To test whether the introduced amino acid exchanges at I3 result in perturbation of the secondary or tertiary structure of PcrB, we analyzed the monomeric variants I93E, H97E, V148E, and L152E by far-UV circular dichroism (CD) and fluorescence spectroscopy. The results showed that monomerization led to only minor modifications of the CD spectrum and caused no significant shift of the fluorescence emission maximum of intrinsic Trp and Tyr residues (Figure S5A, B). The consequences of monomerization for protein stability were analyzed by thermal denaturation experiments, where loss of far-UV CD signal was monitored as a function of temperature. Compared to the wild-type protein, the unfolding transitions of the monomeric variants were only slightly less cooperative, and the melting temperature (51 °C) was decreased by no more than 2–3 °C (Figure S5C). It can be concluded from these findings that the dimeric association state of PcrB has only a minor impact on the conformational stability of the protein.

To study whether monomerization alters the substrate specificity of PcrB, activity tests were performed at an enzyme concentration of 1 μM, a concentration at which wild-type and I2 variants were shown to be dimeric, whereas I3 variants were monomeric. The native PcrB polyprenyl substrate HepPP contains seven isoprene units and is difficult to handle in assays performed in aqueous solution because of its distinct hydrophobicity. We therefore used the soluble substrates for which PcrB shows side activity, GGPP (four isoprene units), FPP (three isoprene units), and GPP (two isoprene units),[2b,4] and incubated them with the PcrB variants and the radiolabeled second substrate [^{14}C]G1P. Analysis of the ^{14}C-labeled products by thin layer chromatography showed that three of the four dimeric variants that contained substitutions in I2 were active with all tested substrates. The exception was variant V73D, which accepted only GPP. This exchange is located in the so-called

"swinging door" region, movements of which have been proposed to control substrate binding and product release for GGGPS.[1b] The inactivation caused by the V73D substitution supports this model. Remarkably, none of the monomeric variants with substitutions in I3 accepted GGPP as a substrate, and only variant I93A accepted FPP. However, all variants were able to catalyze the reaction of [^{14}C]G1P with GPP, albeit with lower efficiency than the wild-type enzyme (Figure 4). We conclude that dimerization of PcrB is necessary for binding of polyprenyl pyrophosphate substrates that contain more than three isoprene units.

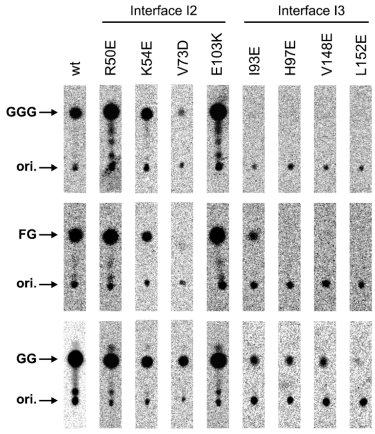


Figure 4. Activity assays of wild-type PcrB and variants with amino acid exchanges at interfaces I2 (dimers) and I3 (monomers). The proteins were incubated with the substrates [^{14}C]G1P and either GGPP (four isoprene units), FPP (three isoprene units), or GPP (two isoprene units). The generated products were dephosphorylated to geranylgeranylglycerol (GGG), farnesylglycerol (FG), or geranylglycerol (GG), respectively, then separated by thin layer chromatography and visualized by autoradiography. The origin of the TLC spots (ori.) and the product spots are marked by arrows.

Discussion

Substrate specificity appears to be modulated by structural movements induced by dimerization

The object of this study was to unravel the significance of dimer formation for the function and stability of PcrB. We found that subunit interaction does not provide much extra stability, as the apparent melting temperatures of the monomeric variants and the wild-type enzyme did not differ significantly. However, drastic differences were observed with respect to the substrate spectra: in contrast to wild-type PcrB and most dimeric variants, all monomerized variants accept only short polyprenyl pyrophosphates (no more than three isoprene units). None of the residues exchanged in I2 or I3 contributes to the substrate binding pocket. Thus, it is unlikely that the I3 exchanges are, in themselves, the reason for the altered substrate specificity of PcrB; rather, it is their monomerizing effect. Although the native PcrB substrate HepPP could not be tested (because of its pronounced hydrophobicity), our results imply that monomerization also precludes binding of its seven isoprene units to the active site.

Inspection of the crystal structure of PcrB with and without modeled HepPP provides a plausible explanation for our findings. The upper part of the extended binding groove, which harbors the pyrophosphate moiety and isoprene units 1–3, is exposed to the solvent (Figure 5). However, between the binding sites for isoprene units 3 and 4 the groove is "kinked" by approximately 45°.[1b] Further down the groove, isoprene units 4–6 are flanked by helix α4a (nomenclature according to Payandeh et al.).[1b] Superposition with the apo crystal structure shows that the distal part of the binding pocket is partially collapsed when no substrate is present. We deduced trajectories for the movements of secondary structure elements upon HepPP binding: these are most pronounced for swinging door elements and for helix α4a (Figure 5). This helix, which is also present in afGGGPS,[1b] forms a central element of the dimer interface and harbors two (I93, H97) of the four amino acids, substitutions of which induced monomerization of PcrB. We therefore speculate that hydrophobic interactions with the adjoining protomer could pull helix α4a away from the protein core, which would result in widening of the bottom part of the groove and allow the binding of long-chain polyprenyl pyrophosphates. In this model, the absence of the second protomer restricts the mobility of helix α4a and thus prevents the accommodation of substrates containing more than three isoprene units. The conservation of quaternary structure between PcrB and GGGPS suggests that the archaeal enzyme must also dimerize to enable binding of its native polyprenyl pyrophosphate substrate (four isoprene units).

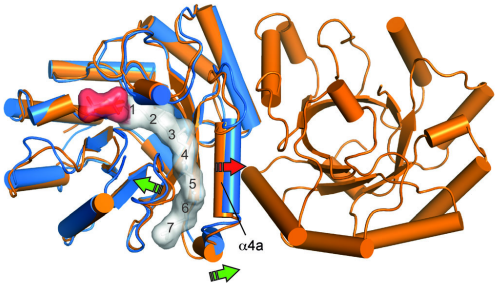


Figure 5. Superposition of the dimeric PcrB apo crystal structure 1VIZ (orange) and a 3D model[2b] (blue). The model includes HepPP (light gray, space-filling model), with pyrophosphate (red) at the catalytic center of the enzyme (isoprene units are numbered). By comparing the position of secondary structure elements, trajectories were determined. Putative movements (green arrows) are related to elements of the "swinging door".[1b] The postulated movement of helix α4a towards the adjoining protomer is also indicated (red arrow).

Oligomerization can influence catalysis and substrate binding of enzymes in different manners. 1) Enzymes such as aspartate transcarbamoylase[19] and alanine racemase[20] contain "shared active sites" that are constituted of residues belonging

to different protomers. 2) For some enzymes of isoprenoid metabolism, dimerization determines substrate specificity, as the boundary of the binding pocket is formed by residues from both protomers.[11] 3) In PcrB, dimerization modulates substrate specificity in an indirect manner by influencing the structural flexibility of the long polyprenylic binding groove. This principle somewhat resembles that for triosephosphate isomerase (TIM),[21] in which each monomer harbors a complete active site that is stabilized by dimer formation. However, whereas in TIM the catalytic residues are affected by dimerization, in PcrB substrate specificity is modulated by the neighboring protomer.

In silico analysis suggests oligomerization states beyond the structurally conserved dimer within the PcrB/GGGPS family

We applied two independent algorithms for the in silico prediction of the native interface of PcrB: Rosetta[14] and PresCont.[13] Rosetta was used for the calculation of the relative stabilities of the putative C1, C2, and C3 complexes, and the effect of amino acid substitutions introduced at interfaces I2 and I3. Notably, C3 was correctly predicted as the native complex, and the calculated mean $\Delta\Delta E$ value, which quantifies the relative (de)stabilization of a complex caused by an amino acid exchange, was considerably higher for the I3 than for the I2 variants. Thus, the predictions made by Rosetta were highly reliable in our case, which is remarkable given the problems that are generally associated with predicting the effects of amino acid substitutions on protein stability.[22] In comparison, tools predicting protein–protein interfaces have generally reached a good performance level.[13] One of the state-of-the-art algorithms is PresCont, which requires as input the three-dimensional structure of one protomer and a multiple sequence alignment (MSA) of homologous sequences. The analysis of the MSA improves prediction quality, because the localization of interfaces is conserved even in remote structural homologues.[23] Remarkably, besides correctly predicting I3 as the native interface, PresCont identified I2 as another surface area with a high potential for forming a protein–protein interaction (Figure S1). As I2 and I3 are located perpendicular to each other on the protein surface, members of the PcrB/GGGPS family could form higher oligomeric complexes by using both interfaces. Consistent with this notion, tetrameric and pentameric complexes have been reported for some archaeal GGGPS enzymes.[2b,3b] We will elucidate the physiological significance for the formation of higher-order oligomers by disrupting I2 in these GGGPS enzymes and testing the consequences for protein stability and substrate specificity.

Experimental Section

Prediction of protein–protein interfaces using PresCont: To predict potential interface regions of a protomer, the algorithm PresCont is supplied with the 3D structure of a protein and an MSA of homologues. We utilized chain A of PDB ID: 1VIZ, and this was preprocessed by Rosetta's *fast relax* protocol[14] in order to resolve unnatural conformations of the crystal structure. By means of BLAST[24] and the non-redundant protein sequence database of the NCBI[25] (version as of April 2011) homologous sequences were collected. BLAST hits were eliminated until all pairwise residues sequence comparisons ranged between 20 and 90% identity. By using *muscle*,[26] the resulting 109 sequences were aligned in an MSA.

Assessing the consequence of individual amino acid substitutions for complex stability by using Rosetta: Single amino acid exchanges were rated in silico by using Rosetta's *fixbb* protocol.[14] The energy status was computed for monomers and homodimers, with and without the considered amino acid exchanges. 3D models for the monomeric wild-type (WT_monomer), the monomers containing a single exchange (Mut_monomer), and the respective dimers (WT_dimer, Mut_dimer) were created. All 3D structures were minimized by utilizing *fast relax*. For every variant, 100 3D models were generated, and the resulting energies were averaged. Based on Rosetta's *score12* energy function, the difference in complex stability $\Delta\Delta E$ was determined according to:

$$\Delta\Delta E = (E_{\mathrm{Mut\_dimer}} - E_{\mathrm{WT\_dimer}}) - 2(E_{\mathrm{Mut\_monomer}} - E_{\mathrm{WT\_monomer}})$$

Thus, by comparing the respective energy values for wild-type and variant proteins, the consequence of each single exchange for complex stability was predicted. Analogously, the stabilities of the putative wild-type complexes C1, C2, and C3 were determined as $\Delta\Delta E = E_{\mathrm{WT\_dimer}} - 2E_{\mathrm{WT\_monomer}}$.

Site-directed mutagenesis and cloning: The *pcrB* gene was amplified by PCR from *B. subtilis* 168 genomic DNA (all primer sequences are listed in Table S1), and the reaction product was cloned into the expression vector pET21a (Novagen) at NdeI/XhoI restriction sites such that there was addition of a C-terminal hexahistidine (His_6) tag to the recombinant protein. For cross-linking with AzF, *pcr*B mutant genes carrying the amber stop codon (TAG) at selected positions were generated by overlap extension PCR.[27]

To introduce nucleotide exchanges for monomerization, the *pcrB* gene was first cloned at BamHI/SalI sites in pUC19 to obtain a small plasmid template for subsequent QuikChange mutagenesis.[28] The mutant genes were then introduced into pET21a_B/S, which is a pET21a derivate containing a linker inserted between the NdeI and XhoI sites in the multiple cloning site (B/S-linker, Table S1).

Production of recombinant proteins, cotranslational incorporation of the unnatural amino acid *p*-azido-L-phenylalanine, and crosslinking: Heterologous expression of genes cloned into pET21a was performed in *E. coli* strain T7 Express I^q (#C2833, Cam^S; New England Biolabs). For standard expressions, the transformed cells were grown at 37 °C in lysogeny broth (LB) containing ampicillin (150 μg mL^{-1}). When the OD_{600} reached 0.6–0.8, expression was induced with isopropyl-β-D-1-thiogalactopyranoside (IPTG; 500 μM), and growth was continued for 4 h.

For site-specific incorporation of the unnatural amino acid AzF,[16b] the host strain was first transformed with the plasmid pEVOL-AzF,[18] to provide the orthogonal $tRNA^{AzF}_{CUA}$/aminoacyl-tRNA synthetase pair for incorporation of AzF. Then, one of the pET21a plasmids containing the *pcr*B amber mutants was introduced into the host by a second transformation. Transformants were grown at 37 °C in LB (50 mL) containing ampicillin (150 μg mL^{-1}), chloramphenicol (30 μg mL^{-1}), and AzF (1 mM, Bachem). The unnatural amino acid was dissolved directly in the medium, accelerated by ultrasonication. Gene expression was induced at OD_{600} = 0.6 by adding L-arabinose (0.02% (*w*/*v*)) and IPTG (1 mM). After a further

1 h of incubation, two aliquots (5 mL) of the culture were transferred into 6 cm petri dishes for in vivo cross-linking experiments. One dish was irradiated for 30 min with a UV handheld lamp (λ = 366 nm, CAMAG) under constant stirring, while the other dish served as negative control. Cells were then harvested, ruptured by boiling, and whole cell extracts were subjected to Western blot analysis. For in vitro cross-linking experiments, the remaining 40 mL of the culture was incubated for a further 2 h at 37 °C, followed by purification of PcrB as described below. Purified protein samples (20 μL) were placed into a 96-well microtiter plate and irradiated for 15 min using a UV handheld lamp (λ = 366 nm, CAMAG). Proteins were separated by SDS-PAGE in 12.5 % polyacrylamide gels, then PcrB was detected by Western blot analysis with a monoclonal Anti-His_6 peroxidase-conjugated antibody (Roche). In large scale experiments, purification yields with AzF-incorporated protein (1 mg protein per liter cell culture, corresponding to 40 nmol protein per 3–4 mL wet cell volume), the PcrB concentration in cross-linking experiments was estimated to be approximately 10 μM in vivo, and 4 μM for in vitro experiments.

Protein purification: Cells were harvested by centrifugation and disrupted by ultrasonication. The His_6-tagged proteins were purified from the crude extract by metal chelate affinity chromatography. In large scale experiments (≥ 0.5 L growth medium), an ÄKTA-purifier system with a HisTrap FF crude column (5 mL, GE Healthcare) was used, and a linear gradient of imidazole (10–500 mM) in potassium phosphate (50 mM, pH 7.5) and KCl (300 mM) was applied to elute the protein. For small scale purifications, SpinTrap columns (GE Healthcare) were used according to the manufacturer's instructions. If necessary, interfering imidazole and salt was removed from the purified proteins by dialysis against potassium phosphate (50 mM, pH 7.5) at 4 °C. To obtain highly purified samples for subsequent analytical gel filtration, preparative size-exclusion chromatography was performed on a Superdex S75 column (GE Healthcare), which was equilibrated with potassium phosphate (50 mM, pH 7.5) and NaCl (500 mM), at a flow rate of 0.5 mL min^{-1}. Protein concentrations were determined either by absorbance spectroscopy (ε_{280} = 25 900 $M^{-1}cm^{-1}$, calculated from the amino acid sequence),[29] or by a Bradford assay (Bio-Rad Protein Assay Kit I; Bio-Rad). The purification yields of the variants are listed in the Supporting Information (Figure S3). Purified protein was characterized immediately or placed in liquid nitrogen and stored at −80 °C.

Activity assay: To test the activity of the PcrB variants with different substrates, [^{14}C]G1P was synthesized, and radio assays were performed as described previously.[2b] Briefly, GPP, FPP, or GGPP (20 μM; Sigma–Aldrich) was incubated with [^{14}C]G1P and PcrB variant (1 μM) to produce radiolabeled polyprenylglyceryl phosphates. Following dephosphorylation, the products were analyzed by thin layer chromatography, and visualized by autoradiography.

Analytical methods: Association states of the recombinant proteins were determined by analytical size exclusion chromatography. A BioSep-SEC-S-3000 column (Phenomenex) was run with potassium phosphate (50 mM, pH 7.5) and NaCl (500 mM) in a LaChrom L-7100 HPLC System (Merck–Hitachi) at 25 °C and a flow-rate of 0.8 mL min^{-1}. Protein elution was followed by measuring the absorbance at 280 nm.

Far-UV circular dichroism and fluorescence spectra of the purified variants (4 μM subunit concentration) were recorded in potassium phosphate (50 mM, pH 7.5) at 25 °C, with a J-815 spectropolarimeter (d = 0.1 cm; JASCO) or an FP-6500 spectrofluorometer (d = 1 cm; JASCO). Emission spectra were recorded by exciting tryptophan and tyrosine fluorescence (λ_{ex} = 280 nm, λ_{em} = 295–400 nm) or by exciting exclusively tryptophan fluorescence (λ_{ex} = 295 nm, λ_{em} = 300–400 nm).

Thermal denaturation of purified proteins (10 μM subunit concentration) was monitored in potassium phosphate (50 mM, pH 7.5) with the JASCO J-815 circular dichroism spectrometer (d = 0.1 cm) by following the loss of ellipticity at 220 nm. Unfolding was induced by raising the temperature (1 °C min^{-1}). The measured ellipticity was normalized, and the apparent midpoint temperature (T_M^{app}) of the irreversible unfolding transition was determined as an operational measure of protein stability.

Acknowledgements

We thank Prof. Peter G. Schultz (Scripps Research Institute) for the plasmid pEVOL-AzF and Dietmar Birzer for helpful discussions on the manuscript.

Keywords: enzymes · ether lipids · oligomerization · protein engineering · unnatural amino acids

[1] a) J. Badger, J. M. Sauder, J. M. Adams, S. Antonysamy, K. Bain, M. G. Bergseid, S. G. Buchanan, M. D. Buchanan, Y. Batiyenko, J. A. Christopher, S. Emtage, A. Eroshkina, I. Feil, E. B. Furlong, K. S. Gajiwala, X. Gao, D. He, J. Hendle, A. Huber, K. Hoda et al., *Proteins Struct. Funct. Bioinf.* **2005**, *60*, 787–796; b) J. Payandeh, M. Fujihashi, W. Gillon, E. F. Pai, *J. Biol. Chem.* **2006**, *281*, 6070–6078.

[2] a) H. Guldan, R. Sterner, P. Babinger, *Biochemistry* **2008**, *47*, 7376–7384; b) H. Guldan, F.-M. Matysik, M. Bocola, R. Sterner, P. Babinger, *Angew. Chem.* **2011**, *123*, 8338–8341; *Angew. Chem. Int. Ed.* **2011**, *50*, 8188–8191.

[3] a) A. Chen, D. Zhang, C. D. Poulter, *J. Biol. Chem.* **1993**, *268*, 21701–21705; b) T. Soderberg, A. Chen, C. D. Poulter, *Biochemistry* **2001**, *40*, 14847–14854; c) N. Nemoto, T. Oshima, A. Yamagishi, *J. Biochem.* **2003**, *133*, 651–657; d) J. Payandeh, E. F. Pai, *J. Mol. Evol.* **2007**, *64*, 364–374; e) Y. Koga, *J. Mol. Evol.* **2011**, *72*, 274–282.

[4] E. H. Doud, D. L. Perlstein, M. Wolpert, D. E. Cane, S. Walker, *J. Am. Chem. Soc.* **2011**, *133*, 1270–1273.

[5] R. Sterner, B. Höcker, *Chem. Rev.* **2005**, *105*, 4038–4055.

[6] D. S. Goodsell, A. J. Olson, *Annu. Rev. Biophys.* **2000**, *29*, 105–153.

[7] T. Schwab, D. Skegro, O. Mayans, R. Sterner, *J. Mol. Biol.* **2008**, *376*, 506–516.

[8] a) L. C. Thompson, J. Walters, J. Burke, J. F. Parsons, R. N. Armstrong, H. W. Dirr, *Biochemistry* **2006**, *45*, 2267–2273; b) R. Sterner, W. Liebl, *Crit. Rev. Biochem. Mol. Biol.* **2001**, *36*, 39–106.

[9] M. B. Ottosen, O. Bjornberg, S. Nørager, S. Larsen, B. A. Palfey, K. F. Jensen, *Protein Sci.* **2002**, *11*, 2575–2583.

[10] a) S. Beismann-Driemeyer, R. Sterner, *J. Biol. Chem.* **2001**, *276*, 20387–20396; b) X. Huang, H. M. Holden, F. M. Raushel, *Annu. Rev. Biochem.* **2001**, *70*, 149–180.

[11] a) M. Noike, T. Katagiri, T. Nakayama, T. Koyama, T. Nishino, H. Hemmi, *FEBS J.* **2008**, *275*, 3921–3933; b) T.-H. Chang, F.-L. Hsieh, T.-P. Ko, K.-H. Teng, P.-H. Liang, A. H.-J. Wang, *Plant Cell* **2010**, *22*, 454–467; c) Y.-W. Zhang, X.-Y. Li, T. Koyama, *Biochemistry* **2000**, *39*, 12717–12722.

[12] R. A. Laskowski, *Nucleic Acids Res.* **2009**, *37*, D355–359.

[13] H. Zellner, M. Staudigel, T. Trenner, M. Bittkowski, V. Wolowski, C. Icking, R. Merkl, *Proteins Struct. Funct. Bioinf.* **2012**, *80*, 154–168.

[14] A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. W. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y.-E. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov et al., *Methods Enzymol.* **2011**, *487*, 545–574.

[15] *PyMOL*, Schrödinger, LLC.

[16] a) J. W. Chin, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11020–11024; b) J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *J. Am. Chem. Soc.* **2002**, *124*,

9026 – 9027; c) N. Hino, M. Oyama, A. Sato, T. Mukai, F. Iraha, A. Hayashi, H. Kozuka-Hata, T. Yamamoto, S. Yokoyama, K. Sakamoto, *J. Mol. Biol.* **2011**, *406*, 343 – 353; d) M. Zhang, S. Lin, X. Song, J. Liu, Y. Fu, X. Ge, X. Fu, Z. Chang, P. R. Chen, *Nat. Chem. Biol.* **2011**, *7*, 671 – 677; e) E. M. Tippmann, W. Liu, D. Summerer, A. V. Mack, P. G. Schultz, *ChemBioChem* **2007**, *8*, 2210 – 2214; f) H.-w. Ai, W. Shen, A. Sagi, P. R. Chen, P. G. Schultz, *ChemBioChem* **2011**, *12*, 1854 – 1857.

[17] E. Krieger, G. Koraimann, G. Vriend, *Proteins Struct. Funct. Bioinf.* **2002**, *47*, 393 – 402.

[18] T. S. Young, I. Ahmad, J. A. Yin, P. G. Schultz, *J. Mol. Biol.* **2010**, *395*, 361 – 374.

[19] W. N. Lipscomb, E. R. Kantrowitz, *Acc. Chem. Res.* **2012**, *45*, 444 – 453.

[20] P. LeMagueres, H. Im, A. Dvorak, U. Strych, M. Benedik, K. L. Krause, *Biochemistry* **2003**, *42*, 14752 – 14761.

[21] R. K. Wierenga, E. G. Kapetaniou, R. Venkatesan, *Cell. Mol. Life Sci.* **2010**, *67*, 3961 – 3982.

[22] V. Potapov, M. Cohen, G. Schreiber, *Protein Eng. Des. Sel.* **2009**, *22*, 553 – 560.

[23] Q. C. Zhang, D. Petrey, R. Norel, B. H. Honig, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 10896 – 10901.

[24] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **1990**, *215*, 403 – 410.

[25] S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, *Nucleic Acids Res.* **1997**, *25*, 3389 – 3402.

[26] R. C. Edgar, *Nucleic Acids Res.* **2004**, *32*, 1792 – 1797.

[27] S. N. Ho, H. D. Hunt, R. M. Horton, J. K. Pullen, L. R. Pease, *Gene* **1989**, *77*, 51 – 59.

[28] W. Wang, B. A. Malcolm, *Biotechniques* **1999**, *26*, 680 – 682.

[29] C. N. Pace, F. Vajdos, L. Fee, G. Grimsley, T. Gray, *Protein Sci.* **1995**, *4*, 2411 – 2423.

Received: February 21, 2012
Published online on May 21, 2012

Supporting Information for Publication A:

CHEMBIOCHEM

Supporting Information

Dimerization Determines Substrate Specificity of a Bacterial Prenyltransferase

David Peterhoff, Hermann Zellner, Harald Guldan, Rainer Merkl, Reinhard Sterner, and Patrick Babinger*[a]

cbic_201200127_sm_miscellaneous_information.pdf

Table S1: Primer sequences

Table S1. Primers

Denotation	Sequence
5'*NdeI-pcrB*	5'-gatccatatgtacgatgtaacggagtggaa-3'
3'*XhoI-pcrB*	5'-gtacctcgagctcgcctttcacagccgcta-3'
5'*BamHI-pcrB*	5'-cgcggatcctacgatgtaacggagtggaa-3'
3'*SalI-pcrB*	5'-acgcgtcgacctcgcctttcacagccgcta-3'
B/S-linker_fw	5'-tatgggatccgaattcgtcgacctcgaga-3'
B/S-linker_rev	5'-ctagtctcgaggtcgacgaattcggatccca-3'
PcrB_R50E_fw	5'-gaggataatgtcctggaaatgatgtctaaggtaag-3'
PcrB_R50E_rev	5'-cttaccttagacatcatttccaggacattatcctc-3'
PcrB_K54E_fw	5'-cggatgatgtctgaagtaagacgg-3'
PcrB_K54E_rev	5'-ccgtcttcattcagacatcatccg-3'
PcrB_V73D_fw	5'-gcgattgaagcgattgatcccggctttgactta-3'
PcrB_V73D_rev	5'-taagtcaaagccgggatcaatcgcttcaatcgc-3'
PcrB_I93E_fw	5'-caaaaacgcggattgggaagtcggaatgcaccag-3'
PcrB_I93E_rev	5'-ctggtgcattccgacttcccaatccgcgtttttg-3'
PcrB_H97E_fw	5'-gattgtcggaatggaacagaaagccatg-3'
PcrB_H97E_rev	5'-catggctttctgttccattccgacaatc-3'
PcrB_M101K_fw	5'-gaatgcaccagaaagccaaaaaggaatacggagaac-3'
PcrB_M101K_rev	5'-gttctccgtattcctttttggctttctggtgcattc-3'
PcrB_E103K_fw	5'-cagaaagccatgaagaaatacggagaactgatg-3'
PcrB_E103K_rev	5'-catcagttctccgtatttcttcatggctttctg-3'

Denotation	Sequence
PcrB_V148E_fw	5'-gtcgcctatgcgcgtgaatcggagctgctgcaac-3'
PcrB_V148E_rev	5'-gttgcagcagctccgattcacgcgcataggcgac-3'
PcrB_L152E_fw	5'-cgtgtgtcggagctggaacaactgccgattttt-3'
PcrB_L152E_rev	5'-aaaaatcggcagttgttccagctccgacacacg-3'
PcrB_I72*_fw	5'-tcagcgattgaagcgtaggttcccggctttg-3'
PcrB_I72*_rev	5'-caaagccgggaacctacgcttcaatcgctga-3'
PcrB_L85*_fw	5'-attccgagtgtatagaacagcaaaaacgcg-3'
PcrB_L85*_rev	5'-cgcgtttttgctgttctatacactcggaat-3'
PcrB_Q98*_fw	5'-gtcggaatgcactagaaagccatgaag-3'
PcrB_Q98*_rev	5'-cttcatggctttctagtgcattccgac-3'
PcrB_Y104*_fw	5'-gccatgaaggaatagggagaactgatgtc-3'
PcrB_Y104*_rev	5'-gacatcagttctccctattccttcatggc-3'
PcrB_S109*_fw	5'-cggagaactgatgtagatggaagaaatcgtg-3'
PcrB_S109*_rev	5'-cacgatttcttccatctacatcagttctccg-3'
PcrB_A129*_fw	5'-gattgcaaagccgcgtagctgactgaagcg-3'
PcrB_A129*_rev	5'-cgcttcagtcagctacgcggctttgcaatc-3'
PcrB_Y145*_fw	5'-gatattgtcgcctaggcgcgtgtgtcgg-3'
PcrB_Y145*_rev	5'-ccgacacacgcgcctaggcgacaatatc-3'
PcrB_A193*_fw	5'-atcaaggatgcgtagacggccaagcagtac-3'
PcrB_A193*_rev	5'-gtactgcttggccgtctacgcatccttgat-3'

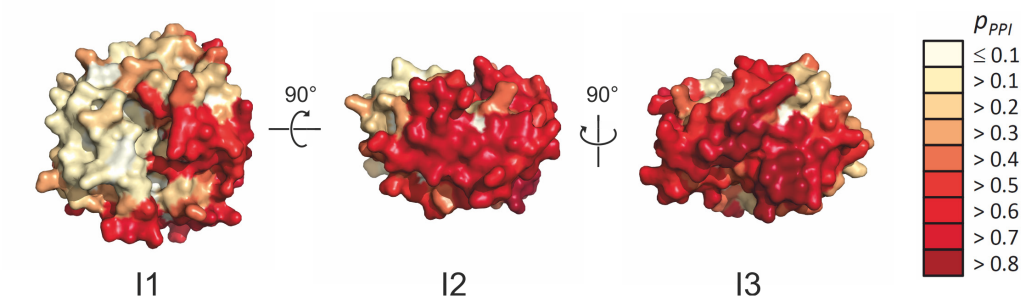


Figure S1: Assessment of putative PcrB dimerization interfaces by means of PresCont. PresCont computes for each surface residue the probability p_{PPI} of being an element of a protein-protein interface. A color code was used to map p_{PPI}-values onto the surface: The darker the red color is, the more likely the residue belongs to an interface. I2 and I3 contain more extended deep-red areas than I1.

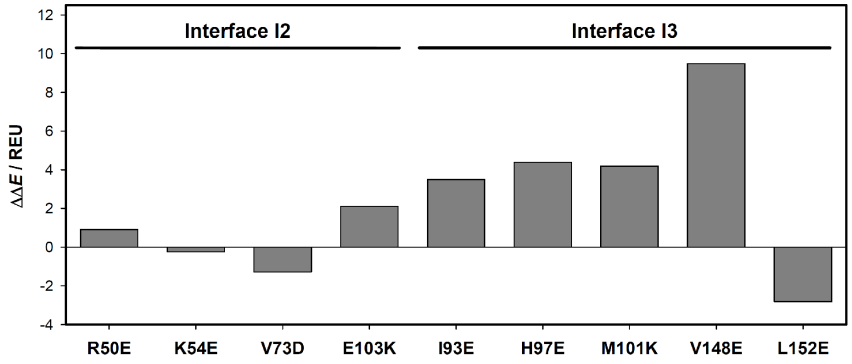


Figure S2: Assessment by Rosetta of the effect of individual amino acid exchanges in I2 and I3 for the stability of the putative C2 and C3 dimers. For each substitution, the energy difference ($\Delta\Delta E$) resulting from a comparison of complexes consisting of wild-type and variant protomers was calculated. A positive $\Delta\Delta E$ value indicates that the substitution destabilizes the complex, whereas a negative value indicates that the exchange is stabilizing. Energy values are plotted in Rosetta Energy Units (REU).

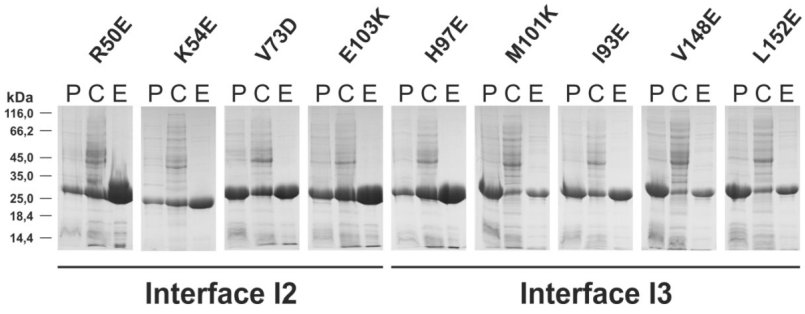


Figure S3: Solubility and homogeneity of purified PcrB variants with amino acid exchanges in I2 and I3 as judged by SDS-PAGE (12.5 % acrylamide). Following the disruption of *E. coli* host cells and centrifugation, insoluble protein appears in the pellet (P), and soluble protein in the crude extract (C). The recombinant PcrB variants carry a C-terminal His_6-tag and were purified from the crude extract by metal chelate affinity chromatography (E) with the following yields (mg per liter of cell suspension): R50E (50), K54E (34), V73D (30), I93E (25), H97E (47), M101K (14), E103K (57), V148E (13), L152E (11). Wild-type PcrB was purified with a yield of 55 mg per liter of cell suspension.

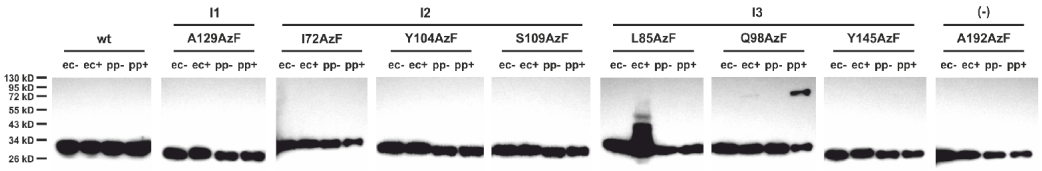
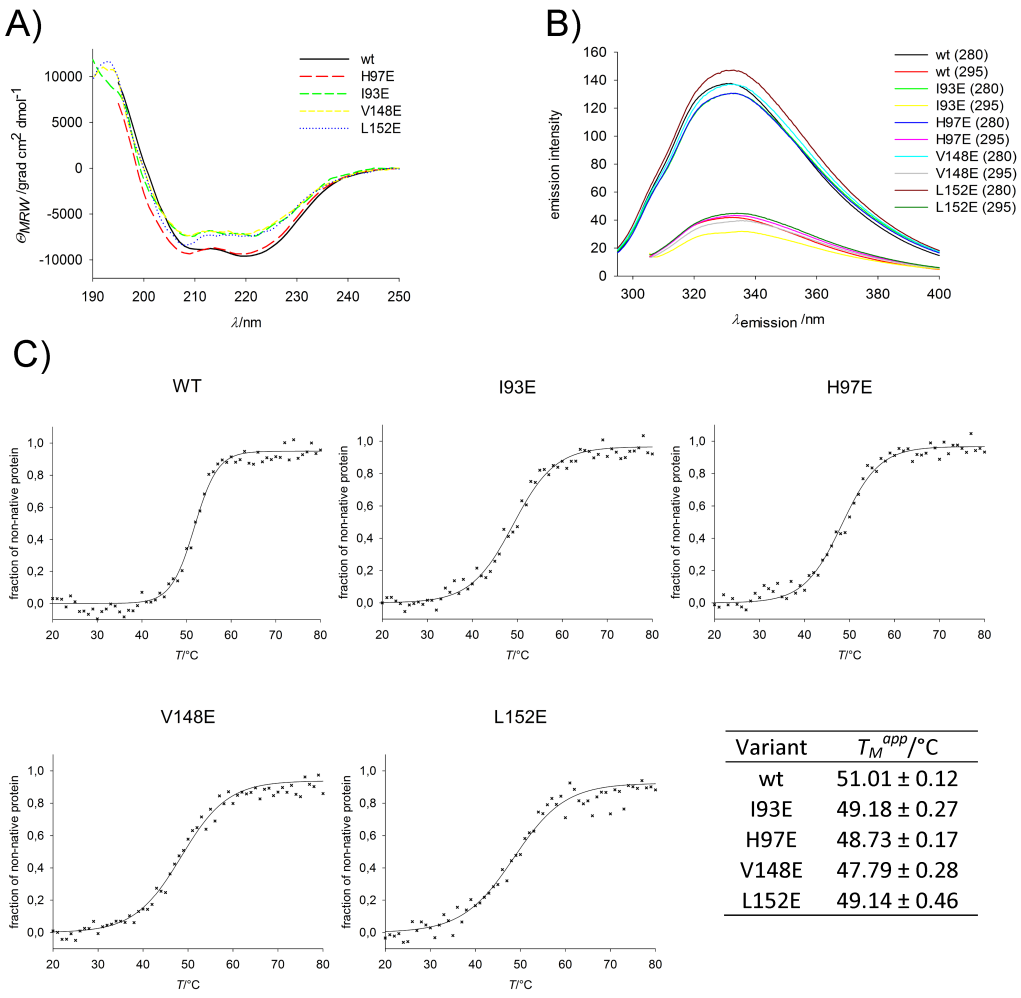


Figure S4: UV-induced cross-linking of dimeric PcrB after incorporation of AzF at various positions within the putative interfaces I1, I2, and I3 (see Figure 1). *E. coli* cells expressing the *pcrB* gene (ec) or purified PcrB protein (pp) were irradiated (+) with UV light (λ = 366nm) or left untreated (-). PcrB monomers and cross-linked dimers were subjected to SDS-PAGE (12.5 % acrylamide) and visualized by Western blot analysis. The protomers of PcrB variant Q98AzF could be cross-linked, whereas no cross-link was observed for the other variants and the wild-type (wt) protein.



Variant	T_M^{app}/°C
wt	51.01 ± 0.12
I93E	49.18 ± 0.27
H97E	48.73 ± 0.17
V148E	47.79 ± 0.28
L152E	49.14 ± 0.46

Figure S5: Structural integrity and stability of wild-type PcrB and its monomeric variants. A) Far-UV circular dichroism spectra. B) Fluorescence emission spectra after excitation at 280 (Tyr + Trp residues) or 295 nm (only Trp residues). C) Thermal denaturation experiments monitored by the loss of the CD signal at 220 nm. The listed apparent melting temperatures (T_M^{app}) correspond to the temperature at which 50 % of the signal is lost and were used as an operational measure of protein stability. The experiments were performed in 50 mM potassium phosphate, pH 7.5, with PcrB subunit concentrations of 4 µM (CD spectra and fluorescence) and 10 µM (thermal unfolding), respectively.

4

CHEMBIOCHEM

ChemPubSoc Europe

CORRIGENDUM

The labels on the two sets of peaks in Figure 2 on p. 1299 were inadvertently swapped over. The correct figure appears here.

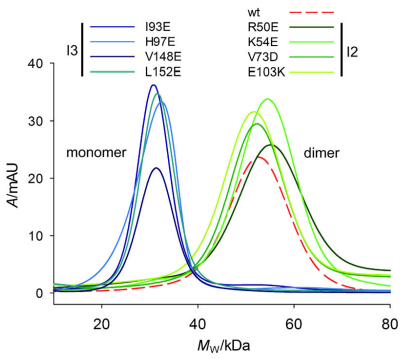


Figure 2. Analytical gel filtration chromatography of wild-type PcrB and variants carrying single amino acid exchanges. Protein was applied at a subunit concentration of 7–10 μM. Elution from BioSep-SEC-S-3000 column resulted in an approximately eightfold dilution of the protein. The runs were performed on different days, and the elution volume profiles were converted to molecular weight profiles for normalization based on updated calibration runs, resulting in an inversion of the *x*-axis. The calculated molecular weight of the PcrB protomer is 26.3 kDa.

The authors apologise for this mistake.

*D. Peterhoff, H. Zellner, H. Guldan, R. Merkl, R. Sterner, P. Babinger**

Dimerization Determines Substrate Specificity of a Bacterial Prenyltransferase

ChemBioChem **2011**, *12*, 1297–1303

DOI 10.1002/cbic.201200127

Publication B

A comprehensive analysis of the geranylgeranylglyceryl phosphate synthase enzyme family identifies novel members and reveals mechanisms of substrate specificity and quaternary structure organization.

David Peterhoff, Barbara Beer, Chitra Rajendran, Esa-Pekka Kumpula, Evangelia Kapetaniou, Harald Guldan, Rik K. Wierenga, Reinhard Sterner, Patrick Babinger (2014).

Molecular Microbiology, Volume 92, Issue 4, pages 885–899.

Molecular Microbiology (2014) 92(4), 885–899 ■
doi:10.1111/mmi.12596
First published online 16 April 2014

A comprehensive analysis of the geranylgeranylglyceryl phosphate synthase enzyme family identifies novel members and reveals mechanisms of substrate specificity and quaternary structure organization

David Peterhoff,[1†] Barbara Beer,[1‡] Chitra Rajendran,[1] Esa-Pekka Kumpula,[2] Evangelia Kapetaniou,[2] Harald Guldan,[1§] Rik K. Wierenga,[2] Reinhard Sterner[1] and Patrick Babinger[1]*

[1]*Institute of Biophysics and Physical Biochemistry, University of Regensburg, Regensburg 93040, Germany.*

[2]*Department of Biochemistry, Biocenter Oulu, University of Oulu, Oulu 90014, Finland.*

Summary

Geranylgeranylglyceryl phosphate synthase (GGGPS) family enzymes catalyse the formation of an ether bond between glycerol-1-phosphate and polyprenyl diphosphates. They are essential for the biosynthesis of archaeal membrane lipids, but also occur in bacterial species, albeit with unknown physiological function. It has been known that there exist two phylogenetic groups (I and II) of GGGPS family enzymes, but a comprehensive study has been missing. We therefore visualized the variability within the family by applying a sequence similarity network, and biochemically characterized 17 representative GGGPS family enzymes regarding their catalytic activities and substrate specificities. Moreover, we present the first crystal structures of group II archaeal and bacterial enzymes. Our analysis revealed that the previously uncharacterized bacterial enzymes from group II have GGGPS activity like the archaeal enzymes and differ from the bacterial group I enzymes that are heptaprenylglyceryl phosphate synthases. The length of the isoprenoid substrate is determined in group II GGGPS enzymes by 'limiter residues' that are different from those in group I enzymes, as shown by site-directed mutagenesis. Most of the group II enzymes form hexamers. We could disrupt these hexamers to stable and catalytically active dimers by mutating a single amino acid that acts as an 'aromatic anchor'.

Accepted 26 March, 2014. *For correspondence. E-mail patrick.babinger@ur.de; Tel. (+49) 941 943 1634; Fax (+49) 941 943 2813. †Present address: Institute of Medical Microbiology and Hygiene, University of Regensburg, Regensburg 93042, Germany. ‡Present address: Chemistry of Biogenic Resources, Straubing Centre of Science, Technische Universität München, Schulgasse 16, Straubing 94315, Germany. §Present address: Lophius Biosciences GmbH, Josef-Engert-Straße 13, Regensburg 93053, Germany.

Introduction

A main difference between the phylogenetic superkingdoms of the Bacteria and the Archaea is the composition of their membrane lipids. While bacterial (and eukaryotic) phospholipids are based on *sn*-glycerol-3-phosphate (G3P) being esterified with two fatty acids, in archaeal phospholipids the enantiomer *sn*-glycerol-1-phosphate (G1P) forms two ether-bonds with isoprenoids, typically with a C20 chain length (Kates, 1993; Koga *et al.*, 1993; Wächtershäuser, 2003; Matsumi *et al.*, 2011). Consequently, the enzyme that catalyses the stereo-specific transfer of the polyprenyl moiety from geranylgeranyl diphosphate (GGPP; 20 C-atoms) to G1P, geranylgeranylglyceryl phosphate synthase (GGGPS), has been regarded to be archaea-specific and presumably played a key role in the evolutionary separation between the superkingdoms (Pereto *et al.*, 2004; Payandeh and Pai, 2007; Glansdorff *et al.*, 2008; Koga, 2011; Lombard *et al.*, 2012a,b). The enzymatic characteristics of archaeal GGGPS have been studied (Chen *et al.*, 1993; Zhang and Poulter, 1993; Soderberg *et al.*, 2001; Nemoto *et al.*, 2003) and the crystal structure of the *Archaeoglobus fulgidus* enzyme (afGGGPS) has been solved (Payandeh *et al.*, 2006).

Interestingly, the progress in genome sequencing identified a number of GGGPS family enzymes also within the Bacteria. Recently, the orthologue from *Bacillus subtilis*, bsPcrB, has been found to be a heptaprenylglyceryl phosphate (HepGP; 35 C-atoms) synthase (HepGPS; Guldan *et al.*, 2011), and the molecular and structural features of bacterial PcrBs have been studied (Badger *et al.*, 2005; Doud *et al.*, 2011; Peterhoff *et al.*, 2012; Ren *et al.*, 2013).

AfGGGPS and PcrB share the frequently encountered $(\beta\alpha)_8$-barrel fold (Sterner and Höcker, 2005) and show a high structural similarity. Their active-site residues are mostly conserved, the main difference being the existence of 'limiter residues' that restrict the length of the isoprenoid substrate in GGGPS to 20 C-atoms (Guldan *et al.*, 2011; Ren *et al.*, 2013). Furthermore, the crystal structure of the related prenyltransferase MoeO5 from *Streptomyces ghanaensis* has been solved (Ren *et al.*, 2012). Like afGGGPS and PcrB, MoeO5 is a $(\beta\alpha)_8$-barrel protein and shares several structural features like a modified version of helix α3, and a similar substrate binding site. However, the reaction catalysed by MoeO5 shows some significant differences to that of GGGPS and PcrB. A *trans*-to-*cis* isomerization occurs within the farnesyl moiety when it is transferred from farnesyl diphosphate to 3-phosphoglycerate, while in case of GGGPS and PcrB the isoprenoid chain remains in an all-*trans* configuration when transferred to G1P (Doud *et al.*, 2011).

Sequence comparisons have shown that the archaeal GGGPS from some Euryarchaeota such as *A. fulgidus* and the Halobacteria share identities of up to 45% with the bacterial PcrB sequences. We term these enzymes GGGPS family group I. In addition, there exists a group II of GGGPS family enzymes, which share a sequence identity of less than 20% with the group I enzymes and contain representatives from all archaeal phyla (Nemoto *et al.*, 2003; Boucher *et al.*, 2004; Guldan *et al.*, 2011). While the enzymes from group II were the first GGGPS to be described (Chen *et al.*, 1993; Zhang and Poulter, 1993; Soderberg *et al.*, 2001; Nemoto *et al.*, 2003), little is known about their structural features, and a rare occurrence of bacterial representatives has been reported also within these sequences (Boucher *et al.*, 2004; Doud *et al.*, 2011; Guldan *et al.*, 2011). For this reason, we decided to study the diversity of the whole family of GGGPS-like enzymes and to systematically characterize representatives with a focus on group II and bacterial enzymes.

Results

Occurrence of GGGPS family enzymes in Archaea and Bacteria

In order to analyse the phylogenetic distribution of GGGPS family enzymes, we followed two complementary approaches by calculating both sequence similarity networks and phylogenetic trees (Fig. 1). Basis for phylogenetic analysis were BLAST searches with different query sequences within the family. The sequence similarity network (Fig. 1A) was calculated with all 1205 sequences from the InterPro database entry IPR008205 ('Geranylgeranylglyceryl phosphate synthase/Heptaprenylglyceryl phosphate synthase'; release 43.1; 7/2013) by using the algorithms developed by Atkinson *et al.* (2009). In such a network, nodes represent individual sequences, and edges represent the E-values from pairwise alignments, calculated by the BLAST algorithm. Edges are drawn when a specified BLAST E-value cut-off (here 10^{-60}) is outvalued. Sequence similarity networks have been shown to correlate well with phylogenetic trees (Kalyanaraman *et al.*, 2008; Lukk *et al.*, 2012), but the two-dimensional representation frequently reveals additional sequence characteristics. The trees in Fig. 1B and Fig. S1 are built from a representative selection of GGGPS family sequences that cover every relevant order plus additional sequences that are discussed in this work.

Our analysis shows that all Archaea except the Nanoarchaeota have a GGGPS gene, which is coherent with the hypothesis that GGGPS is a key enzyme in the evolution of the Archaea (Payandeh and Pai, 2007; Glansdorff *et al.*, 2008) and the finding that *Nanoarchaeum equitans* obtains its membrane lipids from its host *Ignicoccus hospitalis* (Jahn *et al.*, 2004). It has been discussed that the GGGPS family enzymes are split into two distinct phylogenetic groups (group I and group II), both comprising bacterial and archaeal sequences (Nemoto *et al.*, 2003; Boucher *et al.*, 2004; Payandeh and Pai, 2007; Guldan *et al.*, 2011). Group I assembles sequences from some Euryarchaeota and the Firmicutes (*Bacillaceae, Staphylococcaceae, Listeriaceae* and some *Clostridiales*). Most Archaea have a group II GGGPS, but there are also bacterial species, mainly Bacteroidetes, with an enzyme from this group. We term these subgroups in the following group Ia (group I archaeal) and group Ib (group I bacterial; equivalent to PcrB) as well as group IIa (group II archaeal) and group IIb (group II bacterial). The sequence similarity network revealed further splittings within the family: the halobacterial group Ia enzymes are not only separated from the euryarchaeotic ones (IaE), but are divided into two clusters (IaH1, IaH2). Similarly, the group IIa euryarchaeotic sequences are split in two clusters (IIaE1, IIaE2), as do the crenarchaeotic (IIaC1, IIaC2) and the group IIb bacterial ones (IIb1, IIb2).

Systematic expression and characterization of GGGPS family enzymes

While the GGGPS activity has initially been described for group II enzymes (Chen *et al.*, 1993; Zhang and Poulter, 1993; Soderberg *et al.*, 2001; Nemoto *et al.*, 2003), both group Ia and group Ib variants were recently studied in detail with respect to their structure and function (Badger *et al.*, 2005; Payandeh *et al.*, 2006; Payandeh and Pai, 2007; Guldan *et al.*, 2011; Peterhoff *et al.*, 2012; Ren *et al.*, 2013). However, up to now no structure of a group II enzyme has been available, and the rare occurrence of

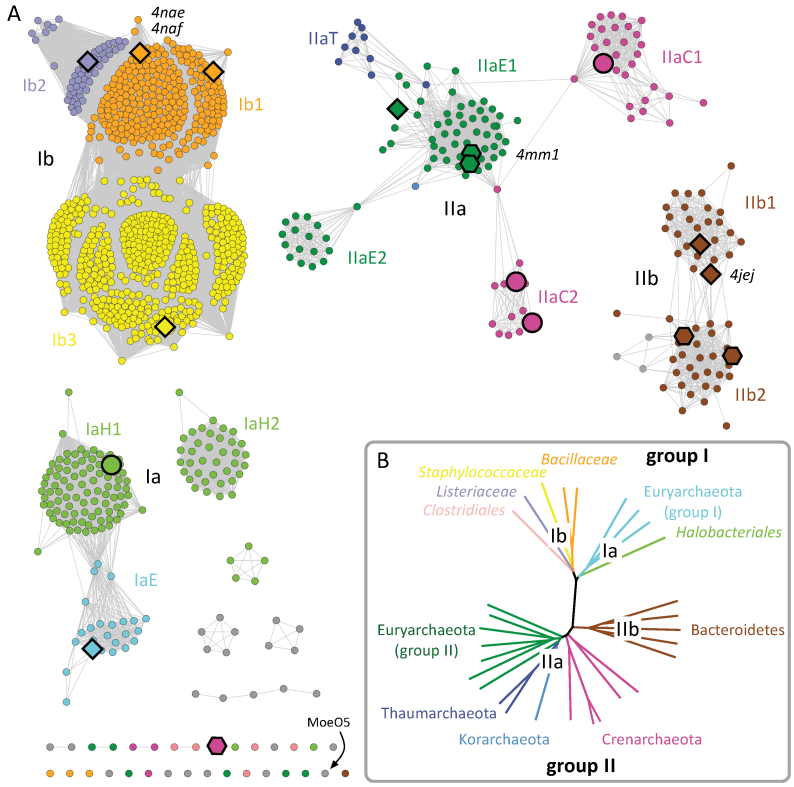


Fig. 1. Representation of the GGGPS enzyme family as a sequence similarity network (A) and a phylogenetic tree (B).
A. The network, which is based on the 1205 sequence members of Interpro entry IPR008205, is thresholded at a BLAST E-value of 10^{-60}. Note that sequences of *Staphylococcaceae*, *Bacillaceae* and *Listeriaceae* are largely overrepresented because they are frequent objects of sequencing projects.
B. Representative sequences were selected from every order, a multiple sequence alignment was calculated, and a phylogenetic tree was generated. The same tree with full species names is shown in Fig. S1.
Subclusters are labelled in corresponding colour and named as follows: Latin number: group I and II; lowercase letter: archaeal, bacterial; capital letter: Halobacteria, Euryarchaeota, Crenarchaeota, Thaumarchaeota; the Arabic number counts the subclusters. Variants characterized in this study (as listed in Table S1) are shown in large symbols, and their quaternary structure is illustrated by shaping: hexagons symbolize hexameric proteins, diamonds symbolize dimeric proteins, quaternary structures of variants shown in circles were not analysed. The pdb-codes of the four structures solved in this study are given next to their symbols in italics. Sequences that are represented only in (A) but not in (B) are shown in grey. The related prenyltransferase MoeO5 (arrow) is also a member of Interpro entry IPR008205, but it does not cluster with GGGPS or PcrB sequences.

bacterial group IIb enzymes has been considered to be non-significant.

In order to shed light onto the enzymatic function and the structural features in particular of the group II GGGPS family enzymes, we opted for a systematic and comparative characterization of enzymes from all subgroups. We decided to examine 17 enzymes from all branches of the phylogenetic tree (Fig. 1, Fig. S1). Following heterologous expression in *Escherichia coli*, we were able to purify 14 of the recombinant proteins (Fig. S2, Table S1). Some of them (afGGGPS, bsPcrB, gkPcrB, saPcrB, lmPcrB, mtGGGPS, tkGGGPS, taGGGPS) have already been partially characterized previously (Chen *et al.*, 1993; Soderberg *et al.*, 2001; Nemoto *et al.*, 2003; Payandeh *et al.*, 2006; Guldan *et al.*, 2011; Ren *et al.*, 2013).

Oligomerization states of the 14 purified variants were determined by static light scattering (SLS) and size exclusion chromatography (SEC) experiments. As reported

earlier (Nemoto *et al.*, 2003; Payandeh *et al.*, 2006; Guldan *et al.*, 2011; Ren *et al.*, 2013), all group I enzymes are dimeric proteins, except the enzyme from *Halobacterium salinarum*, which elutes corresponding to the monomer (Table S2). This deviating behaviour could be due to the specific properties of halobacterial proteins (Lanyi, 1974; Madern *et al.*, 2000; Mevarech *et al.*, 2000) and might be an artefact under the chosen experimental conditions. Group II enzymes have either dimeric or hexameric oligomerization states. This finding is in contrast to previous studies of mtGGGPS, where pentamers have been determined (Chen *et al.*, 1993; Soderberg *et al.*, 2001). The structural basis of hexamerization will be discussed later, based on crystallographic data.

Bacterial group II enzymes exhibit GGGPS, but not HepGPS activity

For a rapid test of prenyltransferase activity, we incubated all purified variants with commercially available geranyl diphosphate (GPP; 10 C-atoms), farnesyl diphosphate (FPP; 15 C-atoms) and geranylgeranyl diphosphate (GGPP; 20 C-atoms) as well as radiolabelled G1P as substrates. All enzymes showed prenyltransferase activity with these substrates (Fig. S3). This finding is the first experimental proof that group IIb enzymes from Bacteroidetes have polyprenylglyceryl phosphate synthase activity.

Longer substrates were tested in an *in vivo* system. *B. subtilis* cells synthesize the polyprenyl diphosphates FPP (15 C-atoms) and GGPP (20 C-atoms) as well as heptaprenyl diphosphate (HepPP; 35 C-atoms) and undecaprenyl diphosphate (UndPP; 55 C-atoms) by specific short-, medium- and long-chain prenyltransferases (Takahashi and Ogura, 1981; 1982; Kabisch *et al.*, 2013). The genes of the 17 selected GGGPS family enzymes were expressed in a *B. subtilis* Δ*pcrB* strain (Kobayashi *et al.*, 2003; Guldan *et al.*, 2011) in presence of radiolabelled G1P. The recombinant proteins used their preferred polyprenyl diphosphates provided by the cells as substrates. When analysing the results, one has to keep in mind that in *B. subtilis* the produced polyprenylglyceryl phosphates are dephosphorylated and then acetylated at one or both hydroxyl groups (Guldan *et al.*, 2011). After cultivation of the *B. subtilis* cells, the lipids were extracted, completely dephosphorylated to allow their separation by thin-layer chromatography, and detected by autoradiography (Fig. 2). The dephosphorylated products (GGG, HepG) as well as the single and double acetylated derivates (Ac_1HepG, Ac_2HepG) result in spots on the autoradiogram. While faint bands of HepG, Ac_1HepG, and Ac_2HepG can be seen in lane 1 comprising an extract of *B. subtilis* wild-type cells, these bands are missing in the Δ*pcrB* knockout strain (lane 2) as well as in the knockout with an empty expression vector control (lane 3). All lanes with extracts from group Ib variants show strong bands for the C35 product and its derivates (lanes 4–7). HsGGGPS (lane 8) shows no activity, presumably because no soluble enzyme is produced in *B. subtilis*. AfGGGPS (lane 9) produces both products C20 and C35, which is consistent with previous findings (Guldan *et al.*, 2011). Group II enzymes from Euryarchaeota and Crenarchaeota mainly produce short chained (C20) products, but also a faint C35 band can be seen in some cases (lanes 10–16). Likewise, mostly short chained products appear in the extracts containing group IIb enzymes (lanes 17–20). This finding clearly demonstrates that the bacterial group IIb enzymes exhibit a substrate specificity different from bacterial group I enzymes, but similar to the archaeal enzymes.

Crystal structures of group II GGGPS enzymes

Variants with acceptable yields in protein expression were subjected to crystallization screens, and four new structures could be solved (Table S3). We obtained two data sets from gkPcrB (a dimeric group Ib variant from *Geobacillus kaustophilus*) with and without bound G1P at a resolution of 2.0 Å, one data set from fjGGGPS (a dimeric group IIb variant from *Flavobacterium johnsoniae*) with bound G1P at a resolution of 1.5 Å, and one from mtGGGPS (a hexameric group IIa variant from *Methanothermobacter thermautotrophicus*) with bound G1P at a resolution of 2.8 Å. The gkPcrB structure could be resolved using the closely related bsPcrB structure (pdb-code 1viz) as a search model. In case of fjGGGPS the molecular replacement worked with afGGGPS and bsPcrB (pdb-codes 2f6x and 1viz) as search models when excluding helix α3*, being consistent with the fact that this is the most flexible and variable region within all members of the GGGPS family enzymes. The mtGGGPS structure was solved using fjGGGPS (pdb-code 4jej) as search model. With the new fjGGGPS (Fig. 3A and B) and mtGGGPS (Fig. 3C) data, structures from the four main subgroups in the GGGPS family (Ia, Ib, IIa, IIb) are now available, including dimeric and hexameric proteins, and allow for a comparative discussion of their features.

The four novel structures show the $(\beta\alpha)_8$-barrel fold with the characteristic variations of GGGPS family enzymes. The active site is located at the C-terminal face of the β-barrel, but the long hydrophobic binding groove for the polyprenylic substrate extends to the rim of the protein, and the adjacent helix α3* is reduced or even completely replaced by a strand (Fig. 3A). All GGGPS family enzymes have an additional helix α0 that forms the 'bottom' or 'plug' of the barrel (Payandeh *et al.*, 2006). In group II enzymes, this helix is considerably longer than in group I enzymes. Interestingly, the α-subunit of tryptophan synthase (TrpA) has such a long helix α0 as well, and based on sequence similarity (BLAST; Altschul *et al.*,

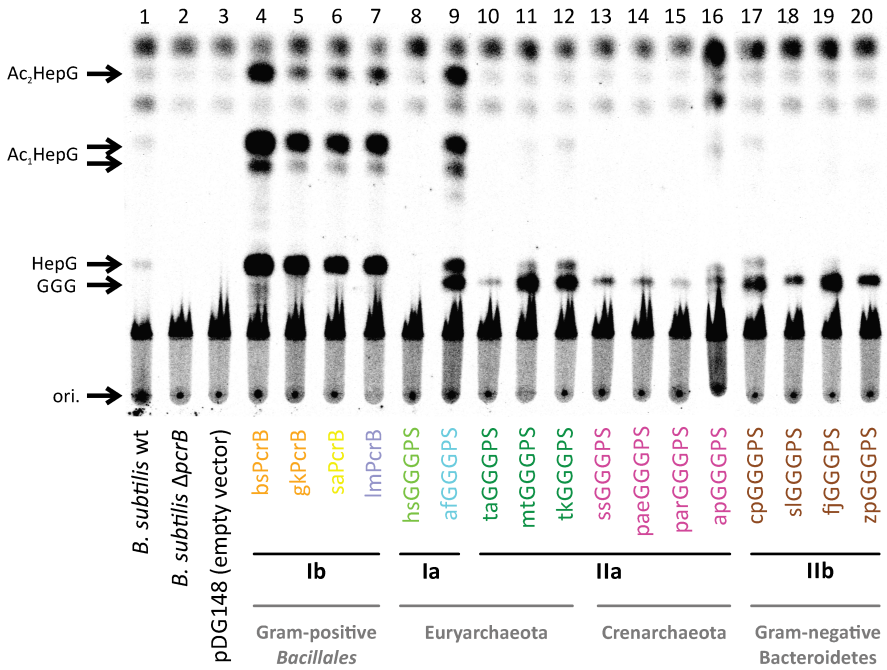


Fig. 2. Analysis of the substrate specificity of selected GGGPS family enzymes. GGGPS variants were expressed in a *B. subtilis* Δ*pcrB* strain in presence of radiolabelled G1P. Lipids were extracted from the cell extract, dephosphorylated, separated by thin-layer chromatography (TLC) using silica 60 plates, and detected by autoradiography. The positions of single or double acetylated heptaprenylglycerol (Ac_1HepG, Ac_2HepG), heptaprenylglycerol (HepG), geranylgeranylglycerol (GGG), and the TLC origin (ori.) are marked with arrows. Because the substrate HepPP is available in excess in *B. subtilis* cells, some GGPP-specific enzymes accept it in a side reaction. The *in vivo* activities of bsPrcB, gkPcrB, saPcrB, lmPrcB and afGGGPS have already been analysed previously (Guldan *et al.*, 2011).

1997), the TrpA protein is the closest relative to the GGGPS family enzymes. Based on structural similarity (DALI; Holm and Sander, 1996), MoeO5 and other phosphate-binding $(\beta\alpha)_8$-barrel enzymes such as the phosphoribosyl anthranilate isomerase (TrpF), the phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (HisA), and the cyclase subunit of the imidazole glycerol phosphate synthase (HisF) are the closest relatives (Payandeh *et al.*, 2006).

The G1P binding site of group II enzymes provides new insights in the mechanism of stereo-selectivity

When discussing the GGGPS structures, we will first focus on three aspects: the anchoring of the G1P via its phosphate group, the catalytic activation of G1P, and the mechanistic basis of stereo-selectivity. In all available GGGPS structures, the anchoring of the G1P phosphate group is accomplished by the $(\beta\alpha)_8$-barrel 'standard phosphate binding motif' (Nagano *et al.*, 2002; Vega *et al.*, 2003), which is constituted by side-chains and backbone amino groups of the loops βα6 (S169, G170), βα7 (G194, G195) and βα8 (V214, G215, N216) (Fig. 4, numbering according to afGGGPS). Interestingly, N216 of the group I enzymes (Fig. 4A and B) is replaced by a conserved threonine in the group II structures (Fig. 4C and D). Furthermore, the loop βα6 is elongated in group II and comes closer to loop βα8, allowing for a direct contact between the side-chain of S177 (mtGGGPS)/S183 (fjGGGPS) and G1P. As for afGGGPS (Payandeh *et al.*, 2006), a network of water molecules involved in the co-ordination of the phosphate moiety could be found in the fjGGGPS structure. However, compared to afGGGPS, the side-chain hydroxy group of S183 from fjGGGPS replaces one of the water molecules (Fig. 4A and D).

It has been proposed for group I enzymes that three strictly conserved tyrosine and glutamate residues (Y124/Y165/E167 in afGGGPS; Y118/Y158/E160 in bsPcrB) are

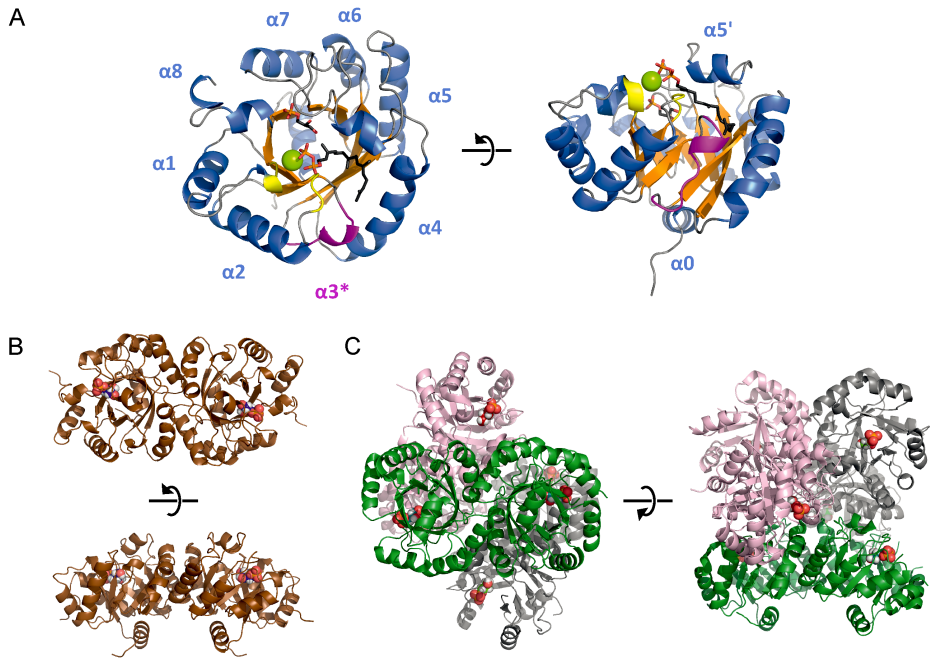


Fig. 3. Structural features of GGGPS family enzymes.
A. Top view (left panel) and side view (right panel) onto the central eight-stranded β-barrel of the protomer of fjGGGPS with bound G1P (pdb-code: 4jej). GGPP and Mg^{2+} were modelled using YASARA Structure V. 13.4.21, employing the YAMBER3 force field (Krieger *et al.*, 2004), based on a previous model of HepPP in bsPcrB (Guldan *et al.*, 2011) and the position of the resolved Mg^{2+} in the close relative MoeO5 (pdb-code: 3vkb). G1P and GGPP are shown in stick representation with carbon atoms in black, oxygen atoms in red, phosphorus atoms in orange. Mg^{2+} is shown as a green sphere. The surface-exposed α-helices are coloured in blue and labelled, the central β-sheets are coloured in orange, the loops involved in Mg^{2+} binding in yellow, and the flexible 'swinging door' helix α3* in magenta.
B. Quaternary structure of dimeric fjGGGPS with G1P shown as red spheres.
C. The quaternary structure of hexameric mtGGGPS combines three dimers in the configuration of native dimeric GGGPS variants (compare B). The three dimers are depicted in green, pink and grey. G1P is shown as red spheres. The catalytic C-terminal face of the β-barrel is oriented towards the centre of the hexamer.

responsible for both stereo-selectivity and catalytic activation (Payandeh *et al.*, 2006; Ren *et al.*, 2013). These residues are strictly conserved in group II enzymes, too (Y130/Y172/E174 in mtGGGPS; Y134/Y178/E180 in fjGGGPS). The glutamate is positioned by the amide of a serine (S169/S162) in group I enzymes, or, due to the elongated βα6 loop in the group II enzymes, by the amide of the adjacent glycine (G176/G182). The glutamate is assumed to polarize the C3 hydroxyl group of G1P, making it an excellent prenyl acceptor. It could even accept the proton to allow for the nucleophilic attack on the preformed carbocation at C1 of the polyprenyl diphosphate, which is the proposed reaction mechanism for GGGPS family enzymes (Payandeh *et al.*, 2006; Ren *et al.*, 2012; 2013). The same function can be assumed for the hydroxyl groups of the two strictly conserved tyrosine residues. Because they are located slightly more remote, they may rather accept the proton than bind the C3-OH for co-ordination.

Stereo-selectivity for G1P versus G3P has been discussed to be realized by the co-ordination of the C2 hydroxy substituent of G1P through the conserved glutamate and tyrosines. But interestingly, a significant difference between the G1P binding pocket of group I and group II enzymes reveals an additional strategy to provide stereo-specificity. In group I enzymes, a conserved lysine (K11 in afGGGPS; K12 in bsPcrB) is located next to C2 and C3 of G1P and has been presumed to be in contact with the C3 hydroxyl group of G1P (Payandeh *et al.*, 2006; Ren *et al.*, 2013). This lysine is substituted by a conserved leucine in group II enzymes (L23 in mtGGGPS; L27 in fjGGGPS). We presume that the hydrophobic part of the

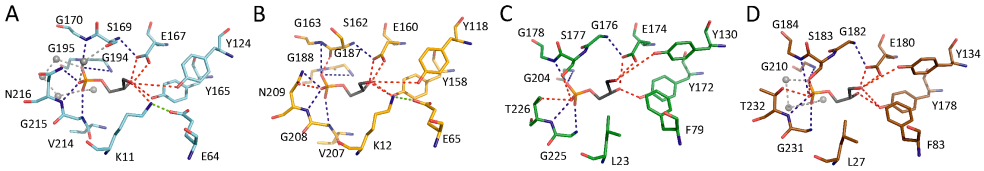


Fig. 4. Organization of G1P binding sites in GGGPS family enzymes.
A. afGGGPS (pdb-code 2f6x), representing group Ia.
B. bsPcrB (pdb-code 3vzy), representing group Ib.
C. mtGGGPS (pdb-code 4 mm1), representing group IIa.
D. fjGGGPS (pdb-code 4jej), representing group IIb.
G1P is coloured with carbon atoms in black, oxygen atoms in red, and the phosphorus atom in orange. Amino acids are coloured according to Fig. 1 and Fig. S1 with amides in blue and oxygens in red. Possible H-bonds that involve amino acid side-chains are shown as red dashed lines, H-bonds to backbone amides are shown as blue dashed lines. The contact of the lysine-glutamate pair is symbolized by green dashed lines. Water molecules are given as grey spheres and their network and connection to the phosphate moiety of G1P is symbolized by grey dashed lines.

lysine side-chain in group I enzymes, like the completely hydrophobic side-chain of the leucine in group II enzymes, establishes the hydrophobic environment for the C2-hydrogen and thereby contributes to the stereoselectivity of the enzyme. The ε-amino group of the lysine may not interact with G1P, but with a closely located group I-conserved glutamate in the βα3 loop region (E64 in afGGGPS; E65 in bsGGGPS). In group II, this lysine–glutamate couple is replaced by a leucine–phenylalanine couple (F79 in mtGGGPS and F83 in fjGGGPS).

Different limiter residues determine substrate specificity in group I and group II enzymes

To achieve specificity for a polyprenyl diphosphate of specific length, the GGGPS fold (i) needs to position the diphosphate head group at the active site in a polar environment and (ii) limit the length of the long hydrophobic binding pocket. The diphosphate moiety has been discussed to bind in complex with magnesium to GGGPS family enzymes (Payandeh *et al.*, 2006; Ren *et al.*, 2013), like to other closely (MoeO5; Ren *et al.*, 2012) and more distantly related prenyltransferases (Christianson, 2006; Oldfield and Lin, 2012). An aspartate residue from the conserved DP(D) motif in loop βα1 and a Ser or Thr from the conserved GGS/T motif in loop βα2 is most probably involved in Mg^{2+} binding (Fig. S4). Residues D13 and T37 (afGGGPS) as well as D14 and S40 (bsPcrB) superimpose with an Asp and Ser in the closely related structure of MoeO5 that co-ordinate the resolved Mg^{2+} (Ren *et al.*, 2012). Furthermore, a D13A mutant in afGGGPS (Payandeh and Pai, 2007) and D14A mutants in bsPcrB and saPcrB (Ren *et al.*, 2013) have severely reduced or no detectable activity, which indicates the importance of this residue for catalysis. Although we found no density indicating that magnesium is bound in our group II structures, D29 (whose side-chain is not resolved), D31 and S58 in fjGGGPS form a polar cluster comparable to the group I structures. In mtGGGPS, the negatively charged cluster is even more pronounced by the presence of an extra glutamate (DPEE25–28 together with S54).

Along with the publication of the first group I structure, afGGGPS, a long hydrophobic cavity, formed by helix α3*, β-strand β4 and partially β5 as well as helix α4 and α5′, has been discussed to bind the isoprenoid chain of the hydrophobic substrate (Payandeh *et al.*, 2006). In a recent study, Ren *et al.* (2013) confirmed this hypothesis by co-crystallizing the substrate analogue *S*-thiolofarnesyl diphosphate (FsPP). When modelling the native substrate HepPP into the predicted binding pocket of bsPcrB (Guldan *et al.*, 2011), however, we ended up with a localization of the substrate that was shifted by one isoprene unit towards the Mg^{2+} ion compared to the FsPP from the co-crystallized structure (Fig. 5A). We presume that this result is due to the chain length of the substrate analogue, which is shorter by five C-atoms compared to the native substrate. In the novel group II structures of fjGGGPS and mtGGGPS, a PEG molecule is bound to the hydrophobic binding pocket, as it has been observed in group Ib gkPcrB and saPcrB (Ren *et al.*, 2013). However, the PEG molecule in the group II structures is located more proximal than the PEG in the group Ib structures. This finding is consistent with the fact that group Ib enzymes have a deeper hydrophobic groove than the other GGGPS family enzymes (Fig. 2), and indicates that hydrophobic forces pull the substrate to the distal end of the binding pocket. In group II, the PEG ends exactly with the fourth isoprene unit of the modelled HepPP, which corresponds to the chain length of GGPP (Fig. 5A). We therefore presume that the distal end of the PEG molecule marks the bottom of the binding pocket of fjGGGPS and mtGGGPS.

Our tests of substrate specificity showed that most probably GGPP is the substrate of group Ia, IIa and IIb

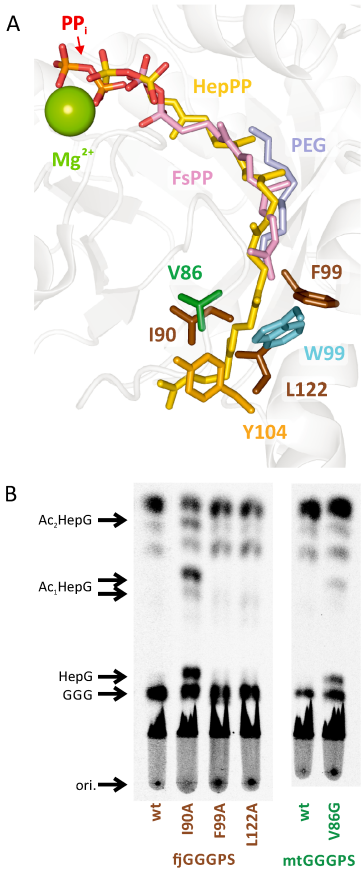


Fig. 5. Analysis of 'limiter' residues from different GGGPS family proteins.
A. Location of potential 'limiter' residues in relation to ligand molecules. The structures of afGGGPS (pdb-code 2f6u), bsPcrB with *S*-thiolo-farnesyl diphosphate (FsPP; pdb-code 3vzz), bsPcrB with modelled heptaprenyl diphosphate (HepPP; Guldan *et al.*, 2011), mtGGGPS (pdb-code: 4 mm1), fjGGGPS with polyethylene glycol (PEG; pdb-code 4jej), and MoeO5 with Mg^{2+} and diphosphate (PP_i; pdb-code 3vkc) were superimposed. Ligands and relevant residues are shown in relation to the backbone of afGGGPS, which is displayed in ribbon representation (grey). The species-specific colour code of amino acid residues corresponds to Fig. 2. HepPP is shown in yellow (oxygens from diphosphate moiety in red), FsPP in pink (oxygens from diphosphate moiety in red), PEG in light blue, PP_i in red (oxygens) and orange (phosphors), and Mg^{2+} is shown as a green sphere.
B. Analysis of the substrate specificity of GGGPS variants with mutated putative limiter residues as shown in (A). For technical details of the assay: see legend to Fig. 2.

enzymes, while group Ib enzymes prefer HepPP (Fig. 2). Previous studies have shown that the bulky side-chain of W99 limits the length of the hydrophobic binding pocket in afGGGPS (Guldan *et al.*, 2011). While it is replaced by an alanine in PcrB, Y104 (bsPcrB numbering) is likely to limit substrate length to C35 in group Ib enzymes (Ren *et al.*, 2013). Somewhat surprisingly, although group II enzymes prefer the short-chained GGPP as a substrate, they have an alanine or a glycine at the afGGGPS-W99 position. We therefore searched the fjGGGPS structure for alternative limiter residues in near proximity to the distal end of the PEG molecule, and identified three promising candidates – I90 in helix α3*, F99 in β-sheet β4 and L122 in helix α4 (Fig. 5A). To analyse the influence of these residues on substrate length limitation, we generated alanine mutants and tested them in the *in vivo* assay. The I90A mutant clearly accepted HepPP as substrate, in contrast to the wild-type enzyme and the F99A and L122A variants (Fig. 5B). Next, we checked the mtGGGPS structure for corresponding residues. At the position corresponding to I90 of fjGGGPS we found a valine (V86) (Fig. 5A). Since the steric hindrance caused by valine compared to alanine is minor, we decided to generate the V86G mutant. It showed an altered product spectrum similar to that of the fjGGGPS-I90A mutant (Fig. 5B). We therefore propose that the I90/V86 position is an essential component of the hydrocarbon ruler in group II GGGPS enzymes. Interestingly, the group II enzyme from the crenarchaeon *Aeropyrum pernix* (apGGGPS), whose membranes exceptionally consist exclusively of C25 polyprenylglyceryl phosphate derivates (De Rosa and Gambacorta, 1988), displays small amino acids at both limiter positions: a glycine at the position of the group Ia tryptophan limiter and an alanine at the group II limiter position. We presume that an alternative limiter residue might exist in this variant, which allows for binding of a C25 polyprenyl substrate.

Hexamerization in group II enzymes is mediated by an aromatic anchor residue

Our SEC and SLS experiments showed that virtually all group I enzymes are dimers, while group II enzymes are either dimers or hexamers (Fig. 1, Table S2). When examining the group II crystal structures, we found that dimeric fjGGGPS from group IIb has the same configuration as the dimeric group I enzymes (Fig. 3B), while the quaternary structure of the hexameric mtGGGPS from group IIa assembles from a combination of three of these dimers (Fig. 3C). In addition to the symmetric standard dimer interface (interface 1), three further interfaces (interface 2, 3a, 3b) are established (Fig. S5A), and each protomer is in contact with three other protomers (Fig. S5B). Interestingly, the location of the additional interfaces has already

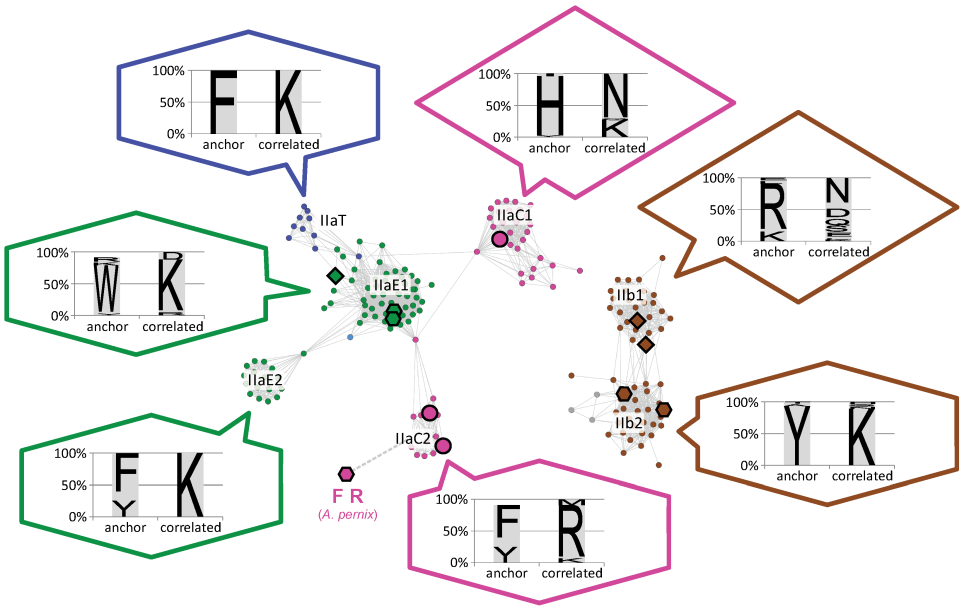


Fig. 6. Correlation of similarity network clustering and oligomerization states of group II GGGPS family proteins. The network shown here is the group II part of Fig. 1, and the same colours and symbols are used. The speech bubbles show the sequence variation within each cluster at the 'aromatic anchor' position ('anchor') and the position of the correlated lysine ('correlated') The shape of the bubbles (diamond, hexagon) symbolizes the assumed oligomerization state (dimeric, hexameric) of the proteins in the cluster. The hexagon close to cluster IIaC2 represents the GGGPS of *A. pernix*, and its F 'anchor' and 'correlated' R residue are given. The dashed line indicates that it clusters with IIaC2 if the network threshold E-value is reduced to 10^{-58}. Note that the dimeric protein in group IIaE1, taGGGPS, has an R at the 'anchor' position like a few other representatives of this cluster.

been predicted in good approximation by the program PresCont (Zellner *et al.*, 2012), which we have used in a recent study to identify the native interface of bsPcrB (Peterhoff *et al.*, 2012).

Interface 2 is symmetric like interface 1 and covers the hydrophobic substrate binding pocket. It includes a hydrogen bond from N82 to Y105 in the adjacent protomer, and a base stacking contact between the two Y105 residues (Fig. S5B). The expanded interface ends at the 'swinging door' of helix α3* (Payandeh *et al.*, 2006). The door is virtually blocked, and a polar contact is formed between R88 in helix α3* and D57 in the βα2 loop of an adjacent protomer. Interfaces 3a and 3b together form a non-symmetric contact. Interface 3a is located at the catalytic face of the barrel. An 'aromatic anchor' (W141) located in helix α5′ plunges into a cavity (interface 3b) in the loop region α5′α5 of an adjacent protomer. This cavity is formed by small hydrophobic amino acids plus the hydrophobic side-chain parts of two lysines, K146 and K152. We assume that this interface is stabilized by hydrophobic interactions and possibly by a cation-pi interaction between W141 and K146. In addition, a polar contact is formed by K146 and D144 of the adjacent protomer, but there is no polar counterpart to K152 (Fig. S5B).

When inspecting all group II members of the sequence similarity network, in those subgroups with a hexameric representative a strong correlation between the occurrence of an aromatic residue (W, Y, F) at the position corresponding to W141 of mtGGGPS and the occurrence of a positively charged residue (K, R) at the position corresponding to K152 was found (Fig. 6). Therefore we assumed that the 'aromatic anchor' is the most essential component for hexamerization. To test this hypothesis, we created an mtGGGPS variant with a W141A mutation. Because the wild-type hexamer is a 'trimer of dimers', we expected to disrupt only the 'trimerization' by the mutation, resulting in a dimeric variant. Additionally, we tested two further hexamers by introducing the analogous mutations Y143A in cpGGGPS and W143A in tkGGGPS. In fact, all three mutants turned out to be dimeric in SEC experiments (Table S4). This result, in combination with the SEC-SLS experiments of all wild-type variants (Table S2),

suggests that the occurrence of the 'aromatic anchor' in correlation with a Lys/Arg is a reliable indicator for hexameric GGGPS family proteins (Fig. 6). Based on this rationale, most of the group II enzymes are hexamers. Interestingly, the sequence similarity network technique is able to discriminate between hexameric and dimeric proteins, leading to a splitting of IIb enzymes into two subclusters (IIb1, IIb2). For the same reason, the crenarchaeotic sequences are split (IIaC1, IIaC2). While most tested GGGPS variants from Crenarchaeota could not be heterologously expressed, apGGGPS from *A. pernix* turned out to be hexameric in SEC-SLS experiments (Table S2). In the similarity network with a clustering cut-off of $E = 10^{-60}$, its sequence has not been integrated into a cluster (Fig. 1), but when the cut-off is changed to $E = 10^{-58}$, apGGGPS clusters with the IIaC2 group (indicated by a dashed line in Fig. 6).

Finally, we tested whether the disruption of the hexamers had consequences for protein stability and activity. Thermal denaturation experiments showed only a slightly decreased apparent melting temperature of dimeric cpGGGPS-Y143A ($T_M^{app} = 58.3°C$) compared to hexameric wild-type cpGGGPS ($T_M^{app} = 60.6°C$) (Fig. S6). The melting temperatures of dimeric mtGGGPS and tkGGGPS were > 100°C, like those of the wild-type proteins (data not shown). Two (mtGGGPS-W141A and tkGGGPS-W143A) of the disrupted hexamers showed almost wild-type catalytic activity in the enzymatic assays whereas the third variant (cpGGGPS-Y143A) was much less active (Figs S3 and S7).

Many Halobacteria have two GGGPS family proteins

As described, the sequence similarity network shows a splitting of halobacterial sequences into two clusters (IaH1 and IaH2 in Fig. 1). We assigned the different variants to species and found that many Halobacteria possess two paralogues of GGGPS enzymes, consistent with former findings of Boucher *et al.* (2004). The IaH1 variant is the prototypical GGGPS which occurs in almost all Halobacteria. Approximately one-third of the Halobacteria (based on the species represented in the network and Interpro entry IPR008205) additionally possess an IaH2 variant. Only two species (*Halovivax asiaticus, Halovivax ruber*) exclusively have a type IaH2 enzyme.

Boucher *et al.* (2004) suggest that the occurrence of two GGGPS paralogues in Halobacteria could be the molecular basis for the existence of C20-C25 diether lipids in some Halobacteria that require the existence of two GGGPS variants with different substrate specificities. Our studies of limiter residues nicely support this hypothesis: A multiple sequence alignment showed that all IaH1 variants have the prototypical group Ia tryptophan limiter (data not shown). Most IaH2 variants have a leucine instead, which could allow for the binding of longer substrates. They occur in species that in fact have C20-C25 diether lipids, like *Haloterrigena*, *Natronobacterium* and *Natrialba* (Kamekura and Kates, 1999). *Haloferax* and *Halobacterium* have two GGGPS paralogues, although these species contain only C20-C20 diether lipids. Strikingly, their IaH2 variant has the bulky aromatic phenylalanine at the W99 position, which could again restrict substrate specificity to C20. Last but not least, there are species with C20–C25 diether lipids that have no IaH2 variant, but our studies show that many archaeal GGGPS enzymes also accept longer substrates if they are present in excess, like HepPP in the *B. subtilis in vivo* system (Fig. 2).

The main difference between IaH1 and IaH2, however, is an insertion of on average 25 amino acids in the variant IaH2, which is located at the $\alpha 4\beta 5$ loop region. Its function is unclear, although it is located close to the binding groove for the polyprenylic substrate. We speculate that it might be involved in protein–protein interactions with substrate providing or product receiving enzymes, or might even be the structural basis for an entirely altered function of these enzymes.

Discussion

About 20 years ago, detailed studies on GGGPS family enzymes started with the purification and characterization of the GGGPS from *M. thermautotrophicus* (Chen *et al.*, 1993). Since then, a significant amount of information has been collected about these enzymes, and it has become obvious that there exist at least two phylogenetic subgroups within the family. However, a comprehensive study of the distinctive structural and functional features has been missing. We have now combined phylogenetic studies with a sequence similarity network analysis to visualize the variability within the family of GGGPS-like enzymes and to select candidates for experimental analysis. The similarity network detected previously unknown splittings within the family (Fig. 1), and our experimental results revealed that this reflects different oligomerization states in case of the groups IIb1 and IIb2, as well as IIaC1 and IIaC2. The two halobacterial groups IaH1 and IaH2 presumably differ in substrate specificity. Only the separation of the IIaE1 and IIaE2 clusters remains without a functional or structural explanation, up to now.

The growing number of sequenced genomes has made clear that there exist bacterial GGGPS family enzymes besides those in *Bacillales*, and we studied their substrate specificity. While bsPcrB from the Gram-positive *B. subtilis* has been shown to be a heptaprenylglyceryl phosphate synthase, and the produced HepGP is further processed by dephosphorylation and acetylation (Guldan *et al.*,

2011), the enzymes from the Gram-negative Bacteroidetes have GGGPS activity like the archaeal enzymes. Unfortunately, Bacteroidetes are not genetically accessible, and *in vivo* experiments by feeding them with radiolabelled G1P were not successful, presumably because G1P is not taken up by the cells (data not shown). Therefore, the activity of these enzymes and the further processing of ether lipids in the native species could not be analysed, and the physiological function remains an enigma. Another mystery with Bacteroidetes is that in contrast to *Bacillales*, most of these species do not possess a glycerol-1-phosphate dehydrogenase gene. So if the GGGPS family enzymes from Bacteroidetes really produce G1P-based ether lipids *in vivo*, they need an alternative source for G1P.

Although most Archaea possess a group II GGGPS, little has been known about their structural features up to now, because most molecular studies have been done with group I enzymes (Badger *et al.*, 2005; Payandeh *et al.*, 2006; Payandeh and Pai, 2007; Guldan *et al.*, 2011; Peterhoff *et al.*, 2012; Ren *et al.*, 2013). We were now able to solve four new crystal structures from GGGPS family enzymes, among them two from group II enzymes. A comparison of group I and group II G1P binding sites shed new light on the mechanism to ensure G1P stereoselectivity, and the analysis of the binding groove for the polyprenylic substrate revealed that group I and group II GGGPS enzymes use alternative amino acid residues to limit the chain length of the substrate. But our activity assays also reveal that although limiter residues direct substrate specificity towards a chain length of C20, most GGGPS enzymes with predominant C20 specificity also use longer substrates if they are available in excess, as it is the case in the *B. subtilis in vivo* assay system. In consequence, the product specificity of these enzymes *in vivo* depends on the concentration of the different polyprenyl diphosphates in the cell. It will be interesting to study how in organisms like *A. pernix*, which have ether lipids with polyprenyl chains longer than C20, the *in vivo* concentrations of longer-chained polyprenyl diphosphates and limiter residues in GGGPS interplay to ensure the accurate synthesis of their lipids.

Interestingly, the crystal structure of mtGGGPS and SEC-SLS analysis of several group II enzymes revealed that most group II enzymes might be hexamers, while all group I enzymes are dimers. This is in contrast to a previous finding that describes mtGGGPS to be a pentamer (Chen *et al.*, 1993). From sequence comparisons we could deduce that an aromatic anchor (Tyr, Trp or Phe) located in helix $\alpha 5'$ is essential for holding the hexamer together. By mutating this residue to an alanine, we disrupted three hexameric representatives to dimers in the prototypical group I architecture. The mutants still showed activity and high thermal stability, and therefore the biological function of hexamer formation remains currently unknown.

Experimental procedures

Sequence similarity networks and phylogenetic analysis

A sequence similarity network of the GGGPS family proteins [1205 sequences, IPR008205 from InterPro release 43.1 (Hunter *et al.*, 2012)] was created using the methods developed by Atkinson *et al.* (2009). The number of BLAST hits was limited to 250 to improve resolution within group Ib, without affecting other clusters. A network with an E-value cut-off of 10^{-60} was visualized using the organic layout in Cytoscape 2.8.3 (Cline *et al.*, 2007; Smoot *et al.*, 2011).

To generate a representative phylogenetic tree, phyla possessing GGGPS family enzymes were identified by BLAST (Altschul *et al.*, 1997). Representative sequences were manually selected for relevant taxa, a multiple sequence alignment was calculated using Clustal Omega (Sievers *et al.*, 2011), and the tree was generated by means of SplitsTree (Huson and Bryant, 2006), using uncorrected P distances and the BioNJ algorithm.

Cloning and site-directed mutagenesis

The genes of 17 GGGPS variants (Table S1) were amplified from genomic DNA, or codon optimized genes were purchased (life technologies), as listed in Table S5. For heterologous expression in *E. coli*, the variants were cloned into pET21a (Novagen), providing a C-terminal hexahistidine (His6) tag, or into a pET28a derivate (pET28trxN) which provides a C-terminal His6 tag plus an N-terminally fused thioredoxin (TrxA), removable by thrombin digestion leaving an N-terminal Gly–Ser extension to the protein. The construct pET28trxN was generated by amplifying the thioredoxin gene *trxA* from genomic *E. coli* DNA and inserting it between the NcoI and NdeI restriction sites in pET28a (Novagen). For expression experiments in *B. subtilis* Δ*pcrB* (Kobayashi *et al.*, 2003), the variants were re-amplified from the pET constructs, including the sequence coding for the His6 tag, and cloned into the pDG148*Stu*I vector (Joseph *et al.*, 2001) via ligation-independent cloning (Aslanidis and de Jong, 1990). Mutants were generated by QuickChange Mutagenesis (Zheng *et al.*, 2004). All constructs were verified by sequencing.

Production and purification of recombinant proteins

Heterologous expression was carried out in the *E. coli* strains T7 Express I^q (New England Biolabs) or BL21-CodonPlus(DE3)-RIPL (Agilent Technologies). The pET vectors led to the addition of a C-terminal His6 tag to the recombinant proteins. The variants hsGGGPS and apGGGPS additionally had an N-terminally fused thioredoxin to improve their solubility (LaVallie *et al.*, 1993). For standard expression experiments, cells were grown at 37°C in LB medium containing 150 μg ml^{-1} ampicillin. After an OD_{600} of 0.6–0.8 was reached, expression was induced by adding

0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) and growth was continued at 37°C either for 4 h or overnight.

Cells were harvested by centrifugation and disrupted by ultrasonication. The His-tagged proteins were purified from the crude extract using a HisTrap FF crude column (5 ml, GE Healthcare) and were eluted with linear gradient from 10 mM to 500 mM or 1 M imidazole in 50 mM potassium phosphate, pH 7.5, 300 mM KCl. Imidazole and salt was removed from the purified proteins by dialysis against 50 mM potassium phosphate, pH 7.5, at 4°C. To remove the fused thioredoxin, the variants were treated with 20 U thrombin (Sigma) per mg protein overnight at room temperature. Uncleaved protein, thioredoxin and thrombin were removed by size exclusion chromatography (Superdex S200 column, GE Healthcare) at 4°C in 50 mM potassium phosphate, pH 7.5, 300 mM KCl. Samples were concentrated by ultrafiltration (Amicon Ultra-15, mwco 10 kDa, Merck Millipore), and protein concentrations were determined either by absorbance spectroscopy or by Bradford Assay (Bio-Rad Protein Assay, Bio-Rad). The molar extinction coefficients at 280 nm were calculated from the amino acid sequence by means of ProtParam (Gasteiger *et al.*, 2005). Purified protein was characterized immediately or dropped into liquid nitrogen and stored at −80°C. The purification yields of the variants are listed in Table S1.

Protein analytics

Average molecular weights of the proteins were determined by static light scattering (SLS) and size exclusion chromatography (SEC) experiments. For SLS experiments the tested proteins had concentrations of 0.5–3.5 mg ml^{-1} (applied volume 100 µl, compare Table S2) and were filtered through 0.22-µm-pore-size PVDF membranes (Ultrafree Centrifugal Filters, Millipore) before being applied to the instrument. The instrument consisted of a Superdex 200 PC 3.2/30 column (GE Healthcare) operated on a ÄKTAmicro system (GE Healthcare) equipped with a Viscotek TDA 305 triple detector array (Malvern) including right-angle light scattering (RALS), viscosimetry, and refractive index (RI) detectors. The system was operated at 25°C in degassed buffer (50 mM potassium phosphate pH 7.5, 500 mM NaCl) at a flow rate of 0.5 ml min^{-1}. The data were analysed using the OmniSec software (Viscotek, version 4.7.0; Malvern).

Standard SEC experiments were performed on a Superdex S75 column (GE Healthcare), which was operated with 50 mM potassium phosphate, pH 7.5, 500 mM NaCl at a flow rate of 0.5 ml min^{-1}. The applied volumes and protein concentrations of the analysed variants are given in Tables S2 and S4.

The heat stability of proteins was investigated by thermal denaturation experiments. To this end, 8 µM protein (subunit concentration) was heated in degassed 50 mM potassium phosphate pH 7.5 at a ramp rate of 1°C min^{-1} in a JASCO J-815 circular dichroism spectrometer ($d = 0.1$ cm). The loss of ellipticity with raising temperature was followed at 220 nm. Ellipticity was normalized, and the apparent midpoint temperature (T_M^{app}) of the irreversible unfolding transition was determined as an operational measure of protein stability.

In vitro *GGGPS activity assays*

^{14}C-G1P was synthesized as described by Guldan *et al.* (2011). To test the activity of purified GGGPS enzymes, 1 µM of protein (subunit concentration) was incubated with 50 µM of different commercially available substrates (GPP, FPP, and GGPP; Sigma-Aldrich) and 100 µM ^{14}C-G1P (150 nCi) in 5 mM $MgCl_2$, 1 mM 2-sulphanyl ethanol, 100 mM HEPES, pH 7.5, for 2 h at 37°C as described previously (Guldan *et al.*, 2011). The products were extracted according to the method of Bligh and Dyer (1959) as modified by Kates (1986). Following dephosphorylation with calf intestinal alkaline phosphatase (New England Biolabs), the products were analysed by TLC on Silica 60 plates, developed in ethyl acetate/hexane 1:1 (v/v), and visualized by autoradiography (overnight exposition). Under these conditions, turnover can be detected only with significantly active variants, while some PcrB mutants showed no turnover under similar conditions (Peterhoff *et al.*, 2012).

In vivo *^{14}C-labelling experiments*

B. subtilis Δ*pcrB* (Kobayashi *et al.*, 2003) was transformed with the pDG148*Stu*I vector constructs containing the different GGGPS variants via electroporation (Xue *et al.*, 1999). Successful transformation was verified by PCR. The transformants were inoculated at an OD_{600} of 0.01 and incubated in 5 ml LB medium containing 1 µg ml^{-1} erythromycin, 5 µg ml^{-1} kanamycin, 1 mM IPTG and 1–2 µCi (25–50 µM) ^{14}C-G1P overnight at 37°C. Lipids were extracted and analysed by thin-layer chromatography (TLC) on silica 60 plates with concentrating zone (Merck Millipore), as described before (Bligh and Dyer, 1959; Kates, 1986; Guldan *et al.*, 2011).

Crystallization

Crystallization experiments with fjGGGPS and mtGGGPS were performed using the commercial sparse matrix screens PGA and Structure screen I + II (Molecular Dimensions) by hanging drop vapour diffusion methods. The proteins (approximately 15–20 mg ml^{-1} in 50 mM potassium phosphate buffer pH 7.5) were supplemented with 1 mM $MgCl_2$ and 1 mM glycerol phosphate (rac.) and incubated at 4°C for 15 min before adding the crystallization buffer, and incubation was continued at 18°C. Crystals of fjGGGPS were obtained with 0.2 M ammonium sulphate, 0.1 M sodium acetate, pH 4.6 and 30% w/v PEG 2000. Crystals of mtGGGPS were obtained using 1.0 M ammonium formate, 0.1 M sodium acetate, pH 5.0, 8% w/v PGA (low molecular weight). After transferring the crystals to ethylene glycol as cryoprotectant, they were flash frozen in liquid nitrogen. Data of single crystals was collected at synchrotron beamline X06DA (Swiss Light Source).

Crystals of gkPcrB were obtained with 19% PEG4000, 0.166 M ammonium acetate, 0.1 M Citrate buffer pH 6.05. Data for apo gkPcrB were collected in house from a single crystal at 100 K with a microfocus Bruker X8 Proteum X-ray generator, equipped with a CCD detector and an Oxford Cryostream 700 series cryosystem. From a single crystal obtained by co-crystallization with G1P, a data set at 2.0 Å

resolution was collected at beamline I911-2 at MAX-lab Lund, Sweden, using a MARCCD detector at wavelength 1.04 Å at 100 K.

The data processing was done using XDS (Kabisch *et al.*, 2013) for all the data sets except for gkPcrB which was processed using the PROTEUM2 software suite (Bruker AXS). The data quality assessment was done using phenix.xtriage (Adams *et al.*, 2002). Molecular replacement was performed with MOLREP within the CCP4i suite (Potterton *et al.*, 2004) using 2f6x and 1viz as search model for fjGGGPS as well as 4jej search model for mtGGGPS. Initial refinement was performed using REFMAC (Murshudov *et al.*, 1997). The model was further improved in several refinement rounds using automated restrained refinement with the program PHENIX (Adams *et al.*, 2002) and interactive modelling with Coot (Emsley and Cowtan, 2004). G1P was identified in the first electron density map based on molecular replacement phases and was included in the model using program Ligandfit embedded in PHENIX (Adams *et al.*, 2002). The refinement statistics are listed in Table S3. The final model was analysed using the program MolProbity (Davis *et al.*, 2007).

Diffraction data and model co-ordinates have been deposited at the PDB with entry codes 4jej (fjGGGPS), 4mm1 (mtGGGPS), 4nae (gkPcrB with bound G1P), and 4naf (gkPcrB apo structure).

Acknowledgements

We thank Kathirvel Alagesan, Richard Greifzu, Patricia Seidel and Mona Linde for experimental assistance, Patrick Löffler for modelling structures, and Harald Huber for helpful discussions. We are grateful to John Gerlt for introducing us to sequence similarity networks and access to a computing cluster to calculate them. This work was supported by a research grant from the Deutsche Forschungsgemeinschaft (BA 3943/2-1).

References

Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., *et al.* (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58:** 1948–1954.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389–3402.

Aslanidis, C., and de Jong, P.J. (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res* **18:** 6069–6074.

Atkinson, H.J., Morris, J.H., Ferrin, T.E., and Babbitt, P.C. (2009) Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS ONE* **4:** e4345.

Badger, J., Sauder, J.M., Adams, J.M., Antonysamy, S., Bain, K., Bergseid, M.G., *et al.* (2005) Structural analysis of a set of proteins resulting from a bacterial genomics project. *Proteins* **60:** 787–796.

Bligh, E.G., and Dyer, W.J. (1959) A rapid method of total lipid extraction and purification. *Can J Biochem Physiol* **37:** 911–917.

Boucher, Y., Kamekura, M., and Doolittle, W.F. (2004) Origins and evolution of isoprenoid lipid biosynthesis in archaea. *Mol Microbiol* **52:** 515–527.

Chen, A., Zhang, D., and Poulter, C.D. (1993) (S)-geranylgeranylglyceryl phosphate synthase. Purification and characterization of the first pathway-specific enzyme in archaebacterial membrane lipid biosynthesis. *J Biol Chem* **268:** 21701–21705.

Christianson, D.W. (2006) Structural biology and chemistry of the terpenoid cyclases. *Chem Rev* **106:** 3412–3442.

Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., *et al.* (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* **2:** 2366–2382.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35:** W375–W383.

De Rosa, M., and Gambacorta, A. (1988) The lipids of archaebacteria. *Prog Lipid Res* **27:** 153–175.

Doud, E.H., Perlstein, D.L., Wolpert, M., Cane, D.E., and Walker, S. (2011) Two distinct mechanisms for TIM barrel prenyltransferases in Bacteria. *J Am Chem Soc* **133:** 1270–1273.

Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60:** 2126–2132.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005) Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook*. Walker, J.M., and Totowa, N.J. (eds). Totowa, NJ: Humana Press, pp. 571–607.

Glansdorff, N., Xu, Y., and Labedan, B. (2008) The Last Universal Common Ancestor: emergence, constitution and genetic legacy of an elusive forerunner. *Biol Direct* **3:** 29.

Guldan, H., Matysik, F.M., Bocola, M., Sterner, R., and Babinger, P. (2011) Functional assignment of an enzyme that catalyzes the synthesis of an archaea-type ether lipid in Bacteria. *Angew Chem Int Ed* **50:** 8188–8191.

Holm, L., and Sander, C. (1996) Mapping the protein universe. *Science* **273:** 595–603.

Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T.K., Bateman, A., *et al.* (2012) InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res* **40:** D306–D312.

Huson, D.H., and Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23:** 254–267.

Jahn, U., Summons, R., Sturt, H., Grosjean, E., and Huber, H. (2004) Composition of the lipids of *Nanoarchaeum equitans* and their origin from its host *Ignicoccus* sp. strain KIN4/I. *Arch Microbiol* **182:** 404–413.

Joseph, P., Fantino, J.R., Herbaud, M.L., and Denizot, F. (2001) Rapid orientated cloning in a shuttle vector allowing modulated gene expression in *Bacillus subtilis*. *FEMS Microbiol Lett* **205:** 91–97.

Kabisch, J., Thurmer, A., Hubel, T., Popper, L., Daniel, R., and Schweder, T. (2013) Characterization and optimization

of *Bacillus subtilis* ATCC 6051 as an expression host. *J Biotechnol* **163:** 97–104.
Kalyanaraman, C., Imker, H.J., Fedorov, A.A., Fedorov, E.V., Glasner, M.E., Babbitt, P.C., *et al.* (2008) Discovery of a dipeptide epimerase enzymatic function guided by homology modeling and virtual screening. *Structure* **16:** 1668–1677.
Kamekura, M., and Kates, M. (1999) Structural diversity of membrane lipids in members of *Halobacteriaceae*. *Biosci Biotechnol Biochem* **63:** 969–972.
Kates, M. (1986) Techniques of lipidology: isolation, analysis, and identification of lipids. In *Laboratory Techniques in Biochemistry and Molecular Biology v. 3, Pt. 2.* Amsterdam; New York: Elsevier, p. 464.
Kates, M. (1993) Membrane lipids in Archaea. In *The Biochemistry of Archaea.* Kates, M., Kushner, D.J., and Matheson, A.T. (eds). Amsterdam: Elsevier, pp. 261–295.
Kobayashi, K., Ehrlich, S.D., Albertini, A., Amati, G., Andersen, K.K., Arnaud, M., *et al.* (2003) Essential *Bacillus subtilis* genes. *Proc Natl Acad Sci USA* **100:** 4678–4683.
Koga, Y. (2011) Early evolution of membrane lipids: how did the lipid divide occur? *J Mol Evol* **72:** 274–282.
Koga, Y., Nishihara, M., Morii, H., and Akagawa-Matsushita, M. (1993) Ether polar lipids of methanogenic bacteria: structures, comparative aspects, and biosyntheses. *Microbiol Rev* **57:** 164–182.
Krieger, E., Darden, T., Nabuurs, S.B., Finkelstein, A., and Vriend, G. (2004) Making optimal use of empirical energy functions: force-field parameterization in crystal space. *Proteins* **57:** 678–683.
Lanyi, J.K. (1974) Salt-dependent properties of proteins from extremely halophilic bacteria. *Bacteriol Rev* **38:** 272–290.
LaVallie, E.R., DiBlasio, E.A., Kovacic, S., Grant, K.L., Schendel, P.F., and McCoy, J.M. (1993) A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Biotechnology* **11:** 187–193.
Lombard, J., Lopez-Garcia, P., and Moreira, D. (2012a) The early evolution of lipid membranes and the three domains of life. *Nat Rev Microbiol* **10:** 507–515.
Lombard, J., Lopez-Garcia, P., and Moreira, D. (2012b) Phylogenomic investigation of phospholipid synthesis in archaea. *Archaea* **2012:** 630910.
Lukk, T., Sakai, A., Kalyanaraman, C., Brown, S.D., Imker, H.J., Song, L., *et al.* (2012) Homology models guide discovery of diverse enzyme specificities among dipeptide epimerases in the enolase superfamily. *Proc Natl Acad Sci USA* **109:** 4122–4127.
Madern, D., Ebel, C., and Zaccai, G. (2000) Halophilic adaptation of enzymes. *Extremophiles* **4:** 91–98.
Matsumi, R., Atomi, H., Driessen, A.J., and van der Oost, J. (2011) Isoprenoid biosynthesis in Archaea – biochemical and evolutionary implications. *Res Microbiol* **162:** 39–52.
Mevarech, M., Frolow, F., and Gloss, L.M. (2000) Halophilic enzymes: proteins with a grain of salt. *Biophys Chem* **86:** 155–164.
Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53:** 240–255.
Nagano, N., Orengo, C.A., and Thornton, J.M. (2002) One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J Mol Biol* **321:** 741–765.
Nemoto, N., Oshima, T., and Yamagishi, A. (2003) Purification and characterization of geranylgeranylglyceryl phosphate synthase from a thermoacidophilic archaeon, *Thermoplasma acidophilum*. *J Biochem* **133:** 651–657.
Oldfield, E., and Lin, F.Y. (2012) Terpene biosynthesis: modularity rules. *Angew Chem Int Ed* **51:** 1124–1137.
Payandeh, J., and Pai, E.F. (2007) Enzyme-driven speciation: crystallizing Archaea via lipid capture. *J Mol Evol* **64:** 364–374.
Payandeh, J., Fujihashi, M., Gillon, W., and Pai, E.F. (2006) The crystal structure of (S)-3-O-geranylgeranylglyceryl phosphate synthase reveals an ancient fold for an ancient enzyme. *J Biol Chem* **281:** 6070–6078.
Pereto, J., Lopez-Garcia, P., and Moreira, D. (2004) Ancestral lipid biosynthesis and early membrane evolution. *Trends Biochem Sci* **29:** 469–477.
Peterhoff, D., Zellner, H., Guldan, H., Merkl, R., Sterner, R., and Babinger, P. (2012) Dimerization determines substrate specificity of a bacterial prenyltransferase. *Chembiochem* **13:** 1297–1303.
Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P., *et al.* (2004) Developments in the CCP4 molecular-graphics project. *Acta Crystallogr D Biol Crystallogr* **60:** 2288–2294.
Ren, F., Ko, T.P., Feng, X., Huang, C.H., Chan, H.C., Hu, Y., *et al.* (2012) Insights into the mechanism of the antibiotic-synthesizing enzyme MoeO5 from crystal structures of different complexes. *Angew Chem Int Ed* **51:** 4157–4160.
Ren, F., Feng, X., Ko, T.P., Huang, C.H., Hu, Y., Chan, H.C., *et al.* (2013) Insights into TIM-barrel prenyl transferase mechanisms: crystal structures of PcrB from *Bacillus subtilis* and *Staphylococcus aureus*. *Chembiochem* **14:** 195–199.
Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7:** 539.
Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L., and Ideker, T. (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27:** 431–432.
Soderberg, T., Chen, A., and Poulter, C.D. (2001) Geranylgeranylglyceryl phosphate synthase. Characterization of the recombinant enzyme from *Methanobacterium thermoautotrophicum*. *Biochemistry* **40:** 14847–14854.
Sterner, R., and Höcker, B. (2005) Catalytic versatility, stability, and evolution of the $(\beta\alpha)_8$-barrel enzyme fold. *Chem Rev* **105:** 4038–4055.
Takahashi, I., and Ogura, K. (1981) Farnesyl pyrophosphate synthetase from *Bacillus subtilis*. *J Biochem* **89:** 1581–1587.
Takahashi, I., and Ogura, K. (1982) Prenyltransferases of *Bacillus subtilis*: undecaprenyl pyrophosphate synthetase and geranylgeranyl pyrophosphate synthetase. *J Biochem* **92:** 1527–1537.
Vega, M.C., Lorentzen, E., Linden, A., and Wilmanns, M.

(2003) Evolutionary markers in the $(\beta/\alpha)_8$-barrel fold. *Curr Opin Chem Biol* **7:** 694–701.
Wächtershäuser, G. (2003) From pre-cells to Eukarya – a tale of two lipids. *Mol Microbiol* **47:** 13–22.
Xue, G.-P., Johnson, J.S., and Dalrymple, B.P. (1999) High osmolarity improves the electro-transformation efficiency of the gram-positive bacteria *Bacillus subtilis* and *Bacillus licheniformis*. *J Microbiol Methods* **34:** 183–191.
Zellner, H., Staudigel, M., Trenner, T., Bittkowski, M., Wolowski, V., Icking, C., and Merkl, R. (2012) PresCont: predicting protein–protein interfaces utilizing four residue properties. *Proteins* **80:** 154–168.
Zhang, D., and Poulter, C.D. (1993) Biosynthesis of archaeal ether lipids. Formation of ether linkages by prenyltransferases. *J Am Chem Soc* **115:** 1270–1277.
Zheng, L., Baumann, U., and Reymond, J.L. (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res* **32:** e115.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Supporting Information for Publication B:

Supplementary information

A comprehensive analysis of the geranylgeranylglyceryl phosphate synthase enzyme family identifies novel members and reveals mechanisms of substrate specificity and quaternary structure organization

David Peterhoff, Barbara Beer, Chitra Rajendran, Esa-Pekka Kumpula, Evangelia Kapetaniou, Harald Guldan, Rik K. Wierenga, Reinhard Sterner, Patrick Babinger

1

Table S1. Enzymes characterized in this study with their purification yields following expression in *E. coli*.

species	protein name	yield (mg/l culture volume)
Archaeoglobus fulgidus	afGGGPS[1]	21.0
Halobacterium salinarum	hsGGGPS	0.9
Bacillus subtilis	bsPcrB[1]	66.3
Geobacillus kaustophilus	gkPcrB[1]	31.8
Staphylococcus aureus	saPcrB[1]	0.3
Listeria monocytogenes	lmPcrB[1]	0.6
Chitinophaga pinensis	cpGGGPS	10.0
Chitinophaga pinensis	cpGGGPS Y143A	9.8
Spirosoma linguale	slGGGPS	3.7
Flavobacterium johnsoniae	fjGGGPS	19.2
Flavobacterium johnsoniae	fjGGGPS I90A	9.2
Zunongwangia profunda	zpGGGPS	2.6
Sulfolobus solfataricus	ssGGGPS	could not be purified
Pyrobaculum aerophilum	paeGGGPS	could not be purified
Pyrobaculum arsenaticum	parGGGPS	could not be purified
Aeropyrum pernix	apGGGPS	1.8
Methanothermobacter thermautotrophicus	mtGGGPS[1]	13.9
Methanothermobacter thermautotrophicus	mtGGGPS V86G	49.0
Methanothermobacter thermautotrophicus	mtGGGPS W141A	41.0
Thermococcus kodakaraensis	tkGGGPS[1]	44.0
Thermococcus kodakaraensis	tkGGGPS W143A	18.6
Thermoplasma acidophilum	taGGGPS	28.1

[1] The genes of these variants have already been cloned and expressed previously (Guldan *et al.*, 2011).

Table S2. SEC-SLS and SEC experiments with GGGPS family proteins.

	MW_{calc} /Da	c_{used} /(mg/ml)	$MW_{apparent}$ /kDa	deduced oligomerization state
bsPcrB[1]	26152.8	2.60	51.2	dimer
gkPcrB[1]	27676.2	1.50	55.3	dimer
saPcrB[1]	26937.6	3.46	45.2	dimer
lmPcrB[1,2]	26417.9	0.50	51.7	dimer
hsGGGPS	26352.1	1.41	29.5	monomer
afGGGPS[1]	27208.9	1.25	53.3	dimer
fjGGGPS	26384.5	1.50	53.0	dimer
zpGGGPS	25220.5	1.30	48.7	dimer
cpGGGPS	26551.8	2.00	156.3	hexamer
slGGGPS	26391.5	0.70	154.9	hexamer
tkGGGPS[3]	27893.2	1.25	163.6	hexamer
mtGGGPS[3]	27590.4	1.07	162.0	hexamer
apGGGPS	27223.3	0.93	159.6	hexamer
taGGGPS[1]	28316.8	1.50	56.9	dimer

[1] The dimeric oligomerization states of these variants have been determined previously by SEC (Guldan *et al.*, 2011, Nemoto *et al.*, 2003) and crystallization (Payandeh *et al.*, 2006, Ren *et al.*, 2013)

[2] lmPcrB was tested in a standard SEC experiment without SLS.

[3] The oligomerization states of these proteins have previously been determined to be tetrameric by SEC only (Guldan *et al.*, 2011) or pentameric by SEC and analytical ultracentrifugation (Chen *et al.*, 1993, Soderberg *et al.*, 2001).

Table S3. Crystal structure determination: data collection and refinement statistics.

	fjGGGPS G1P	mtGGGPS G1P	gkPcrB G1P	gkPcrB apo
Data collection				
Wavelength (Å)	1.0	1.0	1.04	1.54
Unit cell (Å) (Space group)	a=82.27 b=43.01 c=75.927 α=90 β=117.41 γ=90 (C 1 2 1)	a=143.6 b=148.7 c=91.5 α=90 β=109.5 γ=90 (C 1 2 1)	a=85.35 b=157.97 c=38.09 α=90 β=90 γ=90 (P 21 21 1)	a=85.114 b=157.84 c=38.17 α=90 β=90 γ=90 (P 21 21 1)
Unique reflections	34719 (2517)	42401	28722	41523
Redundancy	2.0	2.2	5.3	5.3
Mosaicity	0.563	0.200		
Resolution (Å)	37.06 - 1.524 (1.579 - 1.524)	2.8	2.0	2.0
Completeness (%)	95.89 (70.15)	95.10	80.27	99.95
$I/\sigma(I)$	11.68 (2.80)	2.52	26.9	
R_{merge}	0.058	0.11	0.038	0.038
Refinement				
Resolution (Å)	37.06 - 1.524 (1.579 - 1.524)	48.86 - 2.8	29.02 - 2.0	29.02 - 2.0
R_{work} / R_{free} (%)	0.1771 (0.3171) / 0.1976 (0.3508)	0.2004 / 0.2708	0.2023 / 0.2428	0.2125 / 0.2408
Wilson B-factor	17.60	42.10	25.84	25.76
Average B-factor (Å^2)	26.00	41.20	29.70	28.40
Number of atoms	3878	20136	6521	3634
Macromolecules	1803	9986	3252	3344
Ligands	27	120	46	
Water	181	71	212	290
R.m.s.deviations				
Bond (Å)	0.010	0.006	0.007	0.005
Angles (°)	1.23	1.04	1.10	0.95
Ramachandran plot				
Residues in most favorable and allowed regions (%)	99	96	96	95
Outliers (%)	-	0.98	0.48	0.69
PDB code	4jej	4mm1	4nae	4naf

Table S4. SEC experiments with disrupted hexameric GGGPS variants and the corresponding wild-type proteins.

	MW_{calc} /Da	c_{used} /(µM)	applied volume /µl	$MW_{apparent}$ /kDa[1]	oligomerization state
cpGGGPS wt	27617.0	17.0	100	>100	hexamer
cpGGGPS Y143A	27524.9	16.0	100	53.1	dimer
tkGGGPS wt	27893.2	16.0	100	>100	hexamer
tkGGGPS W143A	27778.1	17.5	100	59.4	dimer
mtGGGPS wt	27590.4	19.1	100	>100	hexamer
mtGGGPS W141A	27475.2	19.6	100	53.8	dimer

[1] A Superdex S75 column was used to ensure proper discrimination between monomers, dimers, and higher oligomers. Since the resolution of the S75 column is low at MW > 100 kDa, refer to Table S2 for exact values.

Table S5. Primer sequences.

A) Primers used for cloning

	Primer for cloning in pET21a	
variant	**5'primer**	**3'primer**
afGGGPS	GATCCATATGCGATGGAGAAAGTGGAG	CTAGCTCGAGAGGTAGCGTTTCAAGGAACG
hsGGGPS	gene synthesis (codon usage optimized to *E. coli*)	
bsPcrB	GATCCATATGTACGATGTAACGGAGTGGAA	GTACCTCGAGCTCGCCTTTCACAGCCGCTA
gkPcrB	gene synthesis (codon usage optimized to *Pyrococcus furiosus*)	
saPcrB	GCTAGCTAGCATGTATGACATTAAAAAATG	GCTACTCGAGTTTACTAGACTCCTTTATTT
lmPcrB	CTAGCATATGAAGCATTTATTCAAGTTAGATC	CTAGCTCGAGAACCGTTTTCTTCCTAAAAA
cpGGGPS	ATCGCATATGCACAATAAAATATACAATTC	ATCGCTCGAGTTTTAACACAGGTGCAGCT
slGGGPS	ATCGCATATGACTATACTCCGGGATTAT	ATCGCTCGAGCGCTTGAACAACGGATTG
fjGGGPS	ATCGCATATGGAGCAAAAAATACTTACC	ATCGCTCGAGTGAATCGAAAAAATGGCTGT
zpGGGPS	ATCGCATATGCCTAAAATTTTAGATGCC	ATCGCTCGAGTTTAAATTCTCCATTTTCAAAAGC
ssGGGPS	gene synthesis (codon usage optimized to *E. coli*)	
apGGGPS	CAGGGCATATGATGACCGTCCTCGAAGACATG	CCGGTACTCGAGTTGTATTTTTGATATTCCTAC
paeGGGPS	CAGGGCATATGAACCTATATGAATATTTACTTG	CCGGTACTCGAGAGCTTGTTTTATCGCCCGCAC
parGGGPS	CAGGGCATATGAATCTCTATGAGTATCTCTC	CCGGTACTCGAGGCGTTGCTTAATTGCGCGC
mtGGGPS	ATCGCATATGTTCAAAATGAAGGTTGAAGATT	ATCGCTCGAGGACGGACCCCATACCCTCAACGATCTCC
tkGGGPS	CTAGCATATGCTCAAGCTTGGAAAAG	CTAGCTCGAGTCCCTTAACGCCCCTGT
taGGGPS	CAGGGCATATGATGACCGTCCTCGAAGACATG	CCGGTACTCGAGTTGTATTTTTGATATTCCTAC

	Primer for cloning in pDG148	
variant	**5'primer**	**3'primer**
afGGGPS	AAGGAGGAAGCAGGTATGCGATGGAGAAAGTGGAG	GACACGCACGAGGTTCAGTGGTGGTGGTGGTG
hsGGGPS	AAGGAGGAAGCAGGTCATATGACTGCGCCGTGGGCGGACTG	GACACGCACGAGGTTCAGTGGTGGTGGTGGTG
bsPcrB	AAGGAGGAAGCAGGTATGTACGATGTAACGGAGT	GACACGCACGAGGTTCAGTGGTGGTGGTGGTG
gkPcrB	AAGGAGGAAGCAGGTATGGAAGAGATTAGGGCGTG	GACACGCACGAGGTTCAGTGGTGGTGGTGGTG
saPcrB	AAGGAGGAAGCAGGTATGTATGACATTAAAAAATGGC	GACACGCACGAGGTTCAGTGGTGGTGGTGGTG
lmPcrB	AAGGAGGAAGCAGGTATGAAGCATTTATTCAAGTT	GACACGCACGAGGTTCAGTGGTGGTGGTGGTG
cpGGGPS	AAGGAGGAAGCAGGTATGCACAATAAAATATACAATTCT	GACACGCACGAGGTTCAGTGGTGGTGGTGGTG
slGGGPS	AAGGAGGAAGCAGGTATGACTATACTCCGGGATTAT	GACACGCACGAGGTTCAGTGGTGGTGGTGGTG
fjGGGPS	AAGGAGGAAGCAGGTATGGAGCAAAAAATACTTACC	GACACGCACGAGGTTCAGTGGTGGTGGTGGTG
zpGGGPS	AAGGAGGAAGCAGGTATGCCTAAAATTTTAGATGCC	GACACGCACGAGGTTCAGTGGTGGTGGTGGTG
ssGGGPS	AAGGAGGAAGCAGGTATGCGCATCCTGAAAAAAAAAATG	GACACGCACGAGGTTCAGTGGTGGTGGTGGTG
apGGGPS	AAGGAGGAAGCAGGTATGGCAGTTAAACGTCGTCG	GACACGCACGAGGTTCAGTGGTGGTGGTGGTG
paeGGGPS	AAGGAGGAAGCAGGTATGAACCTATATGAATATTTACTTG	GACACGCACGAGGTTCAGTGGTGGTGGTGGTG
parGGGPS	AAGGAGGAAGCAGGTATGAATCTCTATGAGTATCTCTC	GACACGCACGAGGTTCAGTGGTGGTGGTGGTG
mtGGGPS	AAGGAGGAAGCAGGTATGTTCAAAATGAAGGTTG	GACACGCACGAGGTTCAGTGGTGGTGGTGGTG
tkGGGPS	AAGGAGGAAGCAGGTATGCTCAAGCTTGGAAAAGTCG	GACACGCACGAGGTTCAGTGGTGGTGGTGGTG
taGGGPS	AAGGAGGAAGCAGGTCATATGATGACCGTCCTCGAAGACATG	GACACGCACGAGGTTCAGTGGTGGTGGTGGTG

B) Primers used for site-directed mutagenesis

variant	**5'primer**	**3'primer**
fjGGGPS_I90A	GGAGATCCATCTCAGGCGTCACCAAAGGCCGATG	CATCGGCCTTTGGTGACGCCTGAGATGGATCTCC
fjGGGPS_F99A	GCCGATGCGATTTTGGCGTTATCATTATTGTCTG	CAGACAATAATGATAACGCCAAAATCGCATCGGC
fjGGGPS_L122A	CAGGCAGCGCCAATAGCGAAAAAAACAAATCTTG	CAAGATTTGTTTTTTTCGCTATTGGCGCTGCCTG
mtGGGPS_V86G	GGTAACACCACGGGTGGCAGCCGCTACGCCGATG	CATCGGCGTAGCGGCTGCCACCCGTGGTGTTACC
cpGGGPS_Y143A	GCGCCGACTACTGTATCTGCGATCAGTAATACC	GGTATTACTGATCGCAGATACAGTAGTCGGCGC
mtGGGPS_W141A	GGGGGTACGGTTGGAGCGGTCGGTGACACCAAG	CTTGGTGTCACCGACCGCTCCAACCGTACCCCC
tkGGGPS_W143A	GGAGAAACCGTGGGCGCGGTCGGGGACGCAAAG	CTTTGCGTCCCCGACCGCGCCCACGGTTTCTCC

C) Primers used for construction of pET28trxN

fw_NcoI_trxA	CAGGGCCATGGGCAGCGATAAAATTATTCA
rev_trxA_Th_NdeI	CAGGGCATATGGCTGCCGCGCGGCACCAGCGCCAGGTTAGCGTCGAGGA

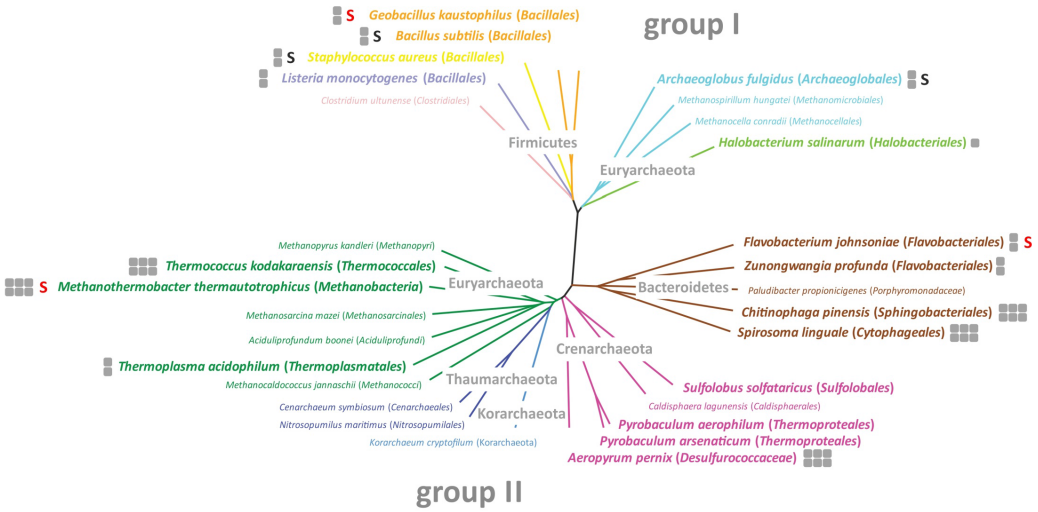


Figure S1. All taxa possessing GGGPS family enzymes were identified by BLAST searches, representative sequences were selected for every order, a multiple sequence alignment was calculated using Clustal Omega (Sievers *et al.*, 2011), and a phylogenetic tree was generated by means of SplitsTree (Huson & Bryant, 2006). Species whose GGGPS variants were characterized during this work are given in large font size. The oligomerization states of some variants were determined by static light scattering experiments (grey symbols). Solved crystal structures are indicated by an "S" (red: this work; black: previous publications).

8

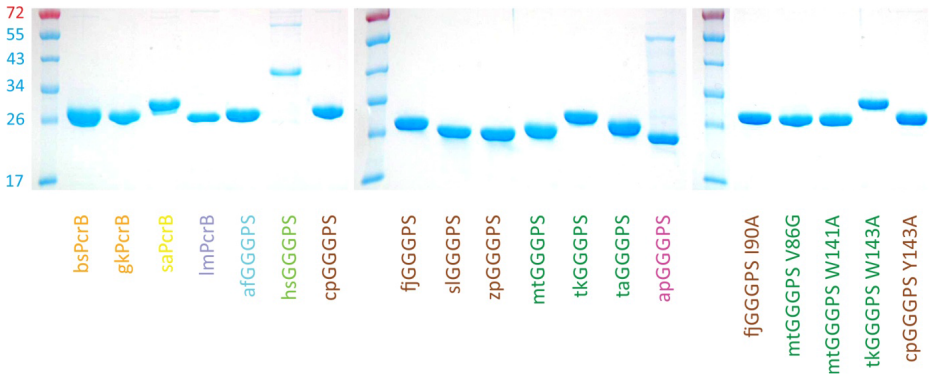


Figure S2. Homogeneity of all purified proteins analyzed in this study as judged by SDS-PAGE (12.5 % acrylamide). Left lanes: marker proteins with molecular weights in kDa. The recombinant proteins carry a C-terminal His6 tag and were purified from the crude extract by metal chelate affinity chromatography with the yields given in Table S1. hsGGGPS and apGGGPS were purified with an N-terminal thioredoxin fusion which was removed by thrombin digestion and subsequent gel filtration in 50 mM potassium phosphate pH 7.5 and 300 mM NaCl. All proteins were dialyzed against 50 mM potassium phosphate, dipped into liquid nitrogen, and stored at -80 °C.

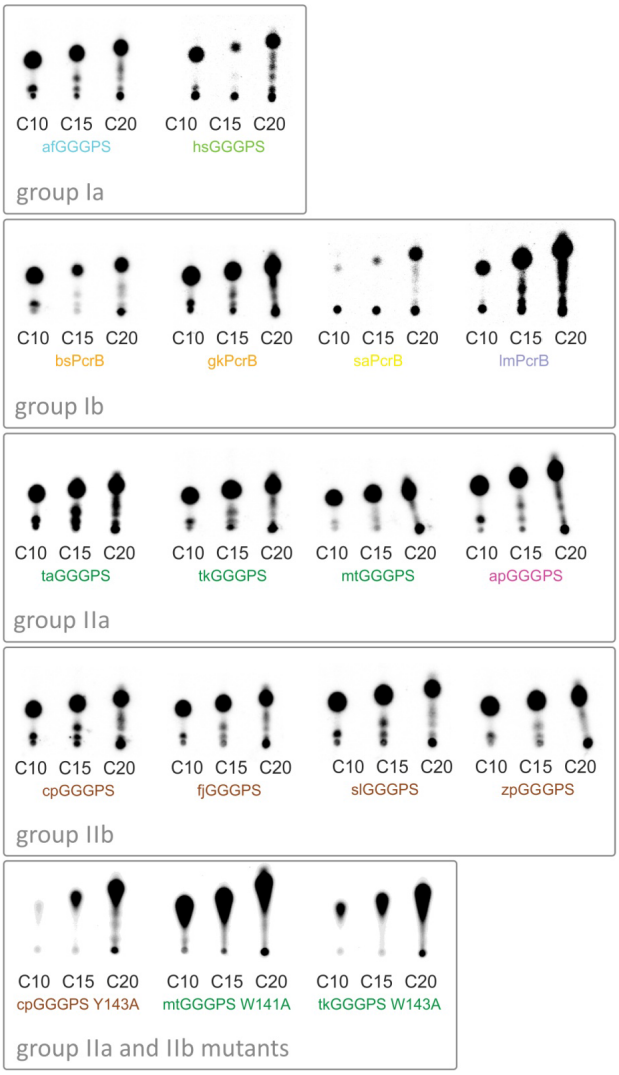


Figure S3. *In vitro* activity assay of purified GGGPS family enzymes with different short chained polyprenyl diphosphate substrates. The proteins were incubated with 50 µM GPP (C10), FPP (C15), or GGPP (C20) and 100 µM ^{14}C-G1P (150 nCi). The products were dephosphorylated, extracted, separated by thin layer chromatography using silica 60 plates, and visualized by autoradiography. The *in vitro* activities of afGGGPS, bsPcrB, gkPcrB, saPcrB and lmPcrB with the substrate GGPP have already been analyzed previously (Guldan *et al.*, 2011).

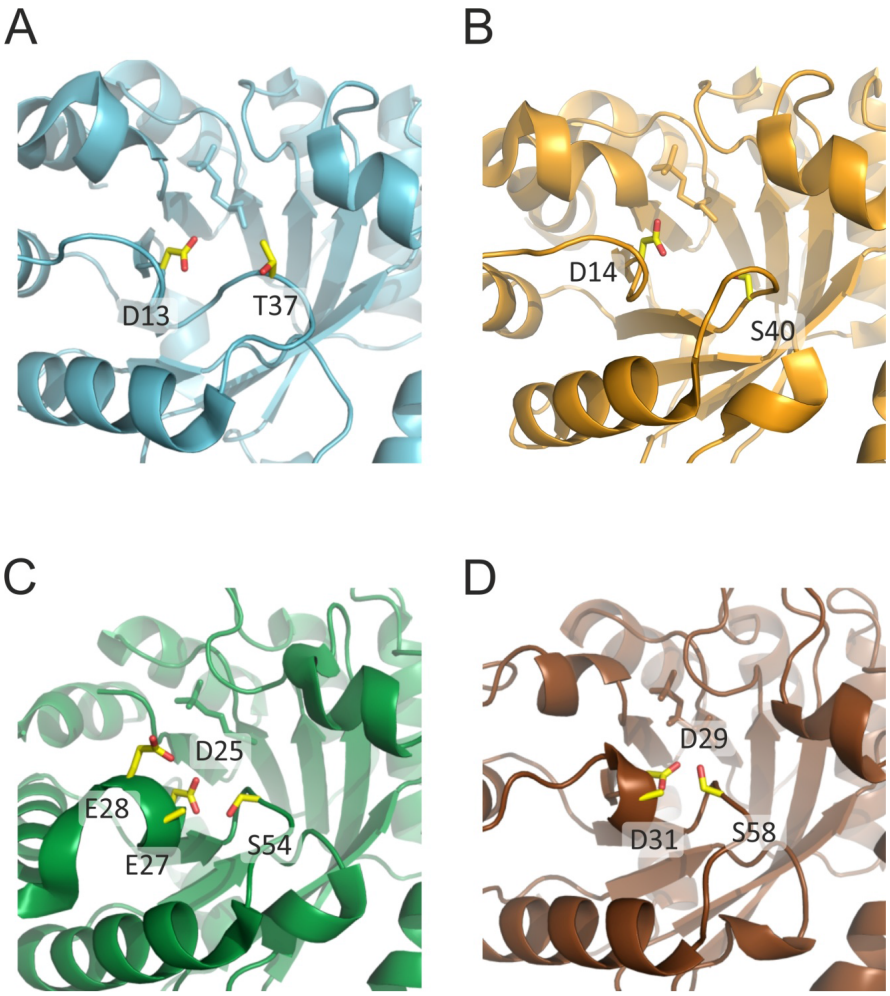


Figure S4. Putative Mg^{2+} binding sites in GGGPS family enzymes. Side chains that are presumably involved in Mg^{2+}-binding are shown as yellow sticks with oxygen atoms marked in red. A) afGGGPS; B) bsPcrB; C) mtGGGPS; D) fjGGGPS. The side chains of E27 in mtGGGPS and D31 in fjGGGPS are not resolved.

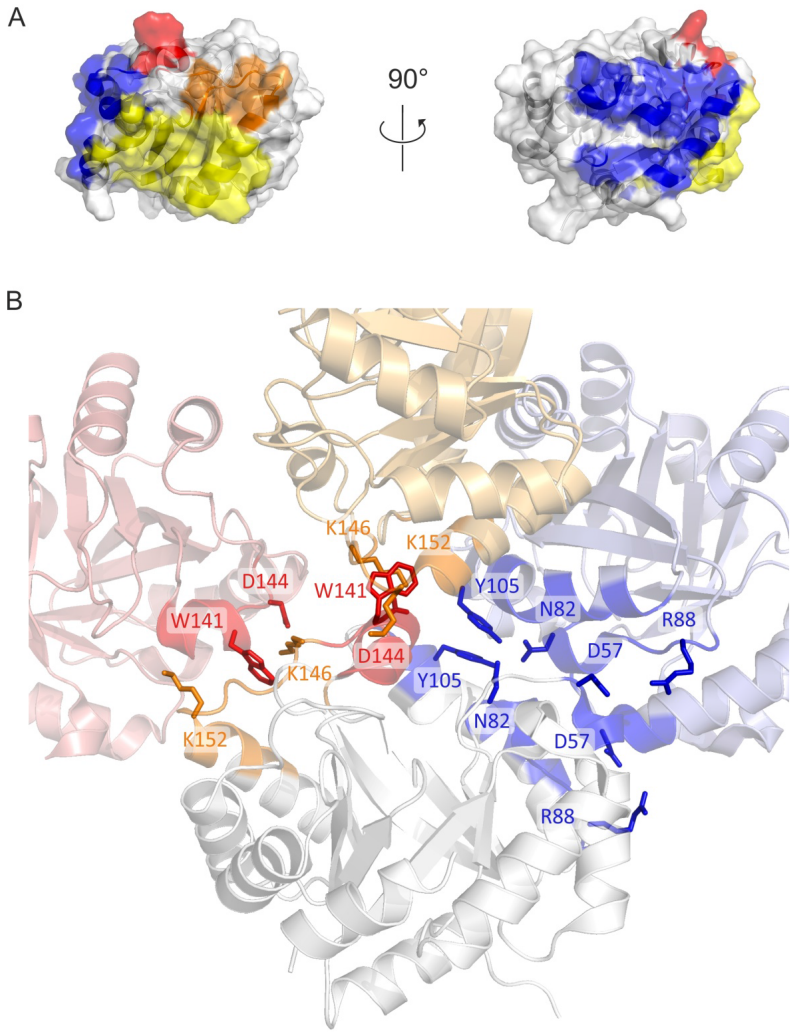


Figure S5. Oligomerization interfaces of hexameric mtGGGPS. A) Localization of the oligomerization interfaces, mapped on the surface of one protomer of mtGGGPS. The symmetric interface 1 (yellow) represents the prototypical interface of the dimeric variants. Interface 2 (blue) is a second symmetric interface. Interface 3a (red) is located at the catalytic face of the barrel and includes helix α5'. Interface 3b (orange) is located at the loop region α5'α5. Within the hexamer, two interfaces 1 are in contact with each other, as are two interfaces 2. Moreover, interface 3a is in contact with interface 3b. B) Important residues in the hexamerization interfaces 2, 3a, and 3b of mtGGGPS. Only four of six subunits shown, and interface 1 is not visible. Coloring of the interfaces corresponds to A). The symmetric contact of interface 2 is stabilized by D57, R88, N82, and Y105. W141 from interface 3a is buried in interface 3b, a hydrophobic cleft which is formed by small hydrophobic amino acids plus a conserved lysine (K152) and a further lysine (K146) which additionally forms a polar contact with D144 from interface 3a.

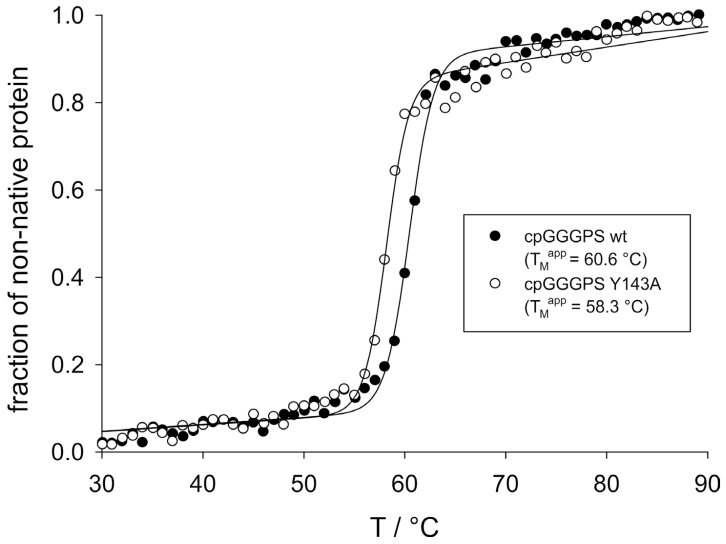


Figure S6. Thermal denaturation of cpGGGPS wt and cpGGGPS-Y143A. Protein at a subunit concentration of 8 µM protein was heated at a ramp rate of 1 °C min^{-1} in degassed 50 mM potassium phosphate, pH 7.5 in a JASCO J-815 circular dichroism spectrometer. The loss of ellipticity with raising temperature was followed at 220 nm. Ellipticity was normalized, and the apparent midpoint temperature (T_M^{app}) of the irreversible unfolding transition was determined as an operational measure of protein stability.

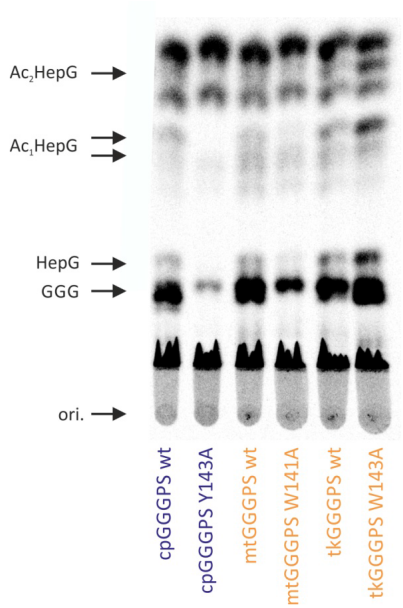


Figure S7. Analysis of the substrate specificity of disrupted hexamer variants, compared to the wild-type proteins. The color code corresponds to Figure 1. For technical details of the assay: see legend to Figure 2. The positions of single or double acetylated heptaprenylglycerol (Ac_1HepG, Ac_2HepG), heptaprenylglycerol (HepG), geranylgeranylglycerol (GGG) and the TLC origin (ori.) are marked with arrows.

References

Chen, A., Zhang, D., and Poulter, C.D. (1993) (S)-geranylgeranylglyceryl phosphate synthase. Purification and characterization of the first pathway-specific enzyme in archaebacterial membrane lipid biosynthesis. *J. Biol. Chem.* **268**: 21701-21705.

Guldan, H., Matysik, F.M., Bocola, M., Sterner, R., and Babinger, P. (2011) Functional assignment of an enzyme that catalyzes the synthesis of an archaea-type ether lipid in Bacteria. *Angewandte Chemie Int. Ed.* **50**: 8188-8191.

Huson, D.H., and Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**: 254-267.

Nemoto, N., Oshima, T., and Yamagishi, A. (2003) Purification and characterization of geranylgeranylglyceryl phosphate synthase from a thermoacidophilic archaeon, *Thermoplasma acidophilum*. *J. Biochem.* **133**: 651-657.

Payandeh, J., Fujihashi, M., Gillon, W., and Pai, E.F. (2006) The crystal structure of (S)-3-O-geranylgeranylglyceryl phosphate synthase reveals an ancient fold for an ancient enzyme. *J. Biol. Chem.* **281**: 6070-6078.

Ren, F., Feng, X., Ko, T.P., Huang, C.H., Hu, Y., Chan, H.C., *et al.* (2013) Insights into TIM-barrel prenyl transferase mechanisms: crystal structures of PcrB from *Bacillus subtilis* and *Staphylococcus aureus*. *Chembiochem* **14**: 195-199.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst.Biol.* **7**: 539.

Soderberg, T., Chen, A., and Poulter, C.D. (2001) Geranylgeranylglyceryl phosphate synthase. Characterization of the recombinant enzyme from *Methanobacterium thermoautotrophicum*. *Biochemistry* **40**: 14847-14854.

7 Appendix: Generation of codon scanned gene libraries

This PhD thesis included a side project with the aim to generate efficient and small gene libraries for either scanning mutagenesis or the random co-translational incorporation of unnatural amino acids into proteins, using the amber suppression technique developed by Schultz and co-workers (Liu et al. 2010). As a result of this, an optimized system for the generation of codon scanned gene libraries was developed, which will be described in the following section.

7.1 Introduction: Methods for protein engineering

In protein engineering man is mimicking *in vitro* what nature has done for billions of years *in vivo*. In an evolutionary process variability is introduced, followed by selection for the fittest. Thus, new or improved functions are established based on well-tried. Virtually two equivalent strategies exist to accomplish this: rational protein design (Hellinga 1997) and directed evolution (Jackel et al. 2008; Brustad et al. 2011). In contrast to rational attempts, where small numbers of variants are tested based on theoretical considerations (often supported by bioinformatical methods (Samish et al. 2011)), in directed evolution large libraries of variants are explored though screening and selection systems. In principle, such libraries can be generated directly via combinatorial chemistry (e.g. solid-phase peptide synthesis) or indirectly by introduction of genetic diversity in the *in vivo* process of protein production. Numerous methods for construction of gene libraries encoding protein variants have been developed during the last decades (Tee et al. 2013). Frequently used approaches, like error-prone PCR (Cadwell et al. 1992) or DNA shuffling (Stemmer 1994), produce libraries with high complexity, often exceeding screening capabilities. Furthermore, they suffer from limitations based on amplification biases and masking of beneficial mutations by deleterious ones. Hence a need for relatively small and well defined gene libraries exists. Along this line, novel concepts in directed evolution favor such well defined libraries based on well chosen target sequences, carrying compensatory stabilizing and neutral mutations (Peisajovich et al. 2007; Goldsmith et al. 2012).

Scanning mutagenesis leads to libraries containing exclusively single spot amino acid (codon triplet) exchanges, insertions and deletions that cover the whole target sequence. Such codon scanned libraries comply with the constraints discussed before. Besides, they enable for unnatural amino acid scanning mutagenesis using the amber stop codon suppression system developed by Schultz and co-workers (Liu et al. 2010). Two different strategies exist in principle to generate such libraries: synthesis via solid phase chemistry and use of molecular

biology methods. While in the current solid phase synthesis procedures, trinucleotide building blocks are implemented in the synthesis process (Virnekas et al. 1994; Ono et al. 1995; Kayushin et al. 1996; Nie et al. 2011), molecular biology approaches use several DNA-modifying enzymes (Luckow et al. 1987; Murakami et al. 2002; Baldwin et al. 2008; Daggett et al. 2009; Liu et al. 2012; Mehta et al. 2012). Successful combinations of both principles have been published (Van den Brulle et al. 2008). While solid phase synthesis requires extensive infrastructure, molecular biology approaches are relatively cost-effective but complicated.

7.2 Optimizing a protocol for scanning mutagenesis

The most recent methods for the generation of codon scanned libraries with scanning mutagenesis are based upon the initial random insertion (transposition) of a mobile DNA fragment (transposon) into the target sequence by means of a transposase (Baldwin et al. 2008; Daggett et al. 2009; Mehta et al. 2012). For this purpose, the transposase MuA (originating from the bacteriophage Mu) is used, which lacks sequence specificity *in vitro*, meaning it catalyzes the transposition of its transposon at random (Rice et al. 1995; Savilahti et al. 1995; Aldaz et al. 1996; Yuan et al. 2005; Saariaho et al. 2006; Kim et al. 2009; Haapa et al. 1999; Jones 2005). The sequence of the MuA transposon is variable but has to feature the two inverted transposase recognition sequences (R1/R2) at its ends. Due to these two characteristics and its high transposition frequency *in vitro*, MuA and its transposon has been used in the first step of scanning mutagenesis approaches to enter the target sequence (Jones 2005; Baldwin et al. 2008; Daggett et al. 2009; Liu et al. 2012). A current protocol (Daggett et al. 2009) for scanning mutagenesis with random in-frame introduction of a codon triplet (shown for the amber stop codon TAG), is outlined in in figure A1.

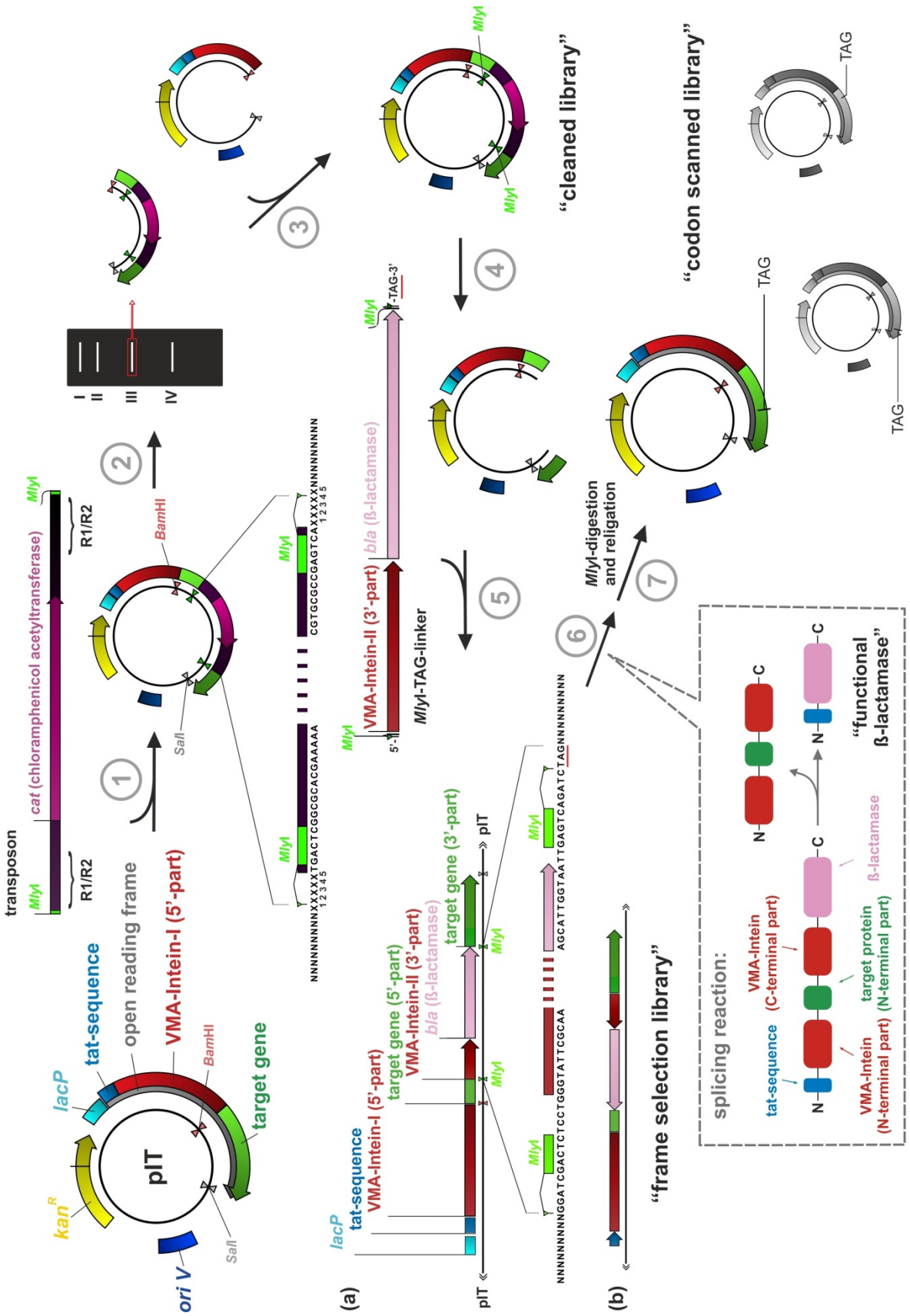


Figure A1: Elements and workflow in scanning mutagenesis of a target gene with the TAG stop codon (Daggett et al. 2009).
In six steps a triplet codon (TAG) scanned library is produced. For a detailed description see main text.

The vector of the system (pIT) replicates under the *oriV* origin and is selectable by kanamycine (*kan*^{R}). It contains an open reading frame, controlled by the lac-promoter (*lacP*), which assembles the tat-sequence for protein export to the periplasm, the 5'part of an intein (VMA-Intein-I, the VMA intein gene from *Saccharomyces cerevisiae*), and the target gene. In the initial reaction the MuA transposase randomly inserts the transposon, consisting of the *cat* gene as selectable marker and the R1/R2 MuA recognition sequences which contain flanking *Mly*I restriction sites, into the vector (step 1). The insertion results in a duplication of five nucleotides of the target sequence (depicted with X_1-X_5), which flank the integration site of the transposon. Digestion with *Bam*HI and *Sal*I (step 2) produces four bands in a subsequent agarose gel electrophoresis: vector backbone containing the transposon (I), vector backbone (II), target gene containing the transposon (III) and target gene (IV). Fragment III (target gene containing the transposon) is religated in the pIT-vector via *Bam*HI and *Sal*I restriction sites leading to a "cleaned library" comprising the transposon randomly inserted at each position of the target gene and a vector backbone free of transposon insertions (step 3). Digestion with the type II restriction endonuclease *Mly*I removes the transposon, results in a deletion of three nucleotides from the target gene sequence (corresponding to X_2-X_4), and produces blunt ends (step 4). For selection of in frame deletions and their replacement with the TAG stop codon the *Mly*I-TAG-linker is ligated into the disrupted target gene (step 5). The *Mly*I-TAG-linker is a fusion of the 3'-part of the VMA-Intein gene to the ß-lactamase gene. In case of in-frame ligation of the *Mly*I-TAG-linker in correct orientation an open reading frame containing tat-sequence, 5'-part of the VMA-Intein, 5'-part of the target gene, 3'-part of the VMA-Intein and the ß-lactamase gene results (a). The non-functional product of reverse ligation is depicted in (b). *E. coli* is transformed with the ligation mixture ("frame selection library") and plated on LB-Agar containing ampicillin and kanamycin. Posttranslational splicing of N- and C-terminal intein (including N-terminal part of target gene-product) results in a functional β-lactamase providing the N-terminal tat-sequence for protein export (see box "splicing reaction"). This splicing occurs only in variants containing the *Mly*I-TAG-linker in frame and in correct orientation. Consequently only such variants survive the ampicillin selection (step 6). Plasmids from the surviving colonies are digested with *Mly*I to remove the *Mly*I-TAG-linker from the vector DNA leaving back the TAG codon to replace an in-frame codon triplet (step 7). The religation of the vector results in a "codon-scanned library".

Using the above described system we found several bottlenecks in the procedure to be worth optimizing. First we found the efficiency of the blunt end ligation step of the *Mly*I-TAG-linker to be low (step 5). We therefore opted for a transposon which combines the features of

the transposon and the *Mly*I-TAG-linker, thereby avoiding the blunt end ligation step. This "frame selectable transposon" required a completely new restriction endonuclease strategy. Second, we decided to redesign the vector to reduce its size (down to 80 %), thus being more efficient in the transposition step (more transposition events in the target gene) and in transformation (Hanahan 1983). Figure A2 shows the redesigned plasmid (pIG, a **pI**T derivate based upon the ori **g**amma) and the novel "frame selectable transposon".

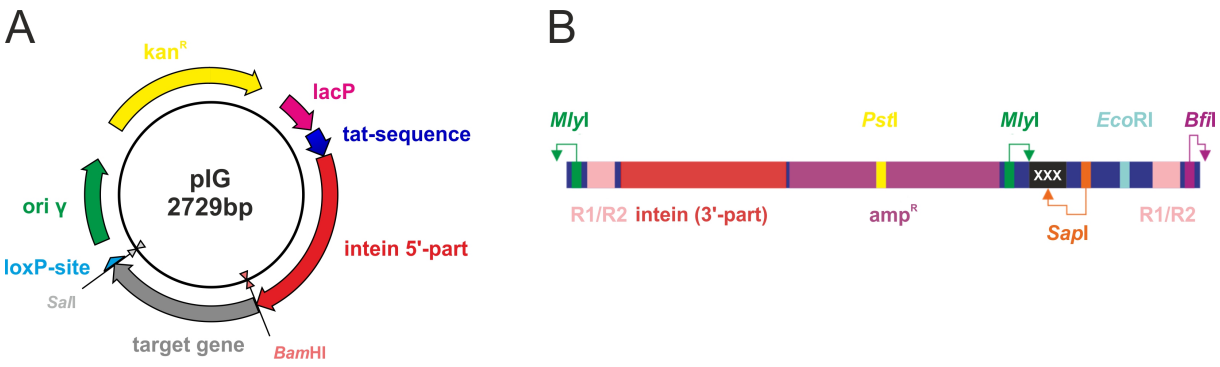


Figure A2: The pIG plasmid and the "frame selectable transposon".
A: Vector chart of the pIG plasmid. The size of the vector is given without a target gene. B: The novel "frame selectable transposon" combines the components of the transposon (R1/R2 sites for random transposition by the MuA transposase) and the MlyI-TAG-linker (elements for frame selection). It contains the scanning triplet (symbolized by XXX) and requires a different restriction endonuclease strategy (necessary restriction sites are shown).

The pIG plasmid (figure A2 A) is based upon the extremely small conditional replicating γ replication origin of R6K (Metcalf et al. 1994). Ori γ requires the *trans*-acting Π protein for replication, which is encoded by the *pir*-gene in special *E. coli* host strains (*E. coli* BW23473, *E. coli* BW23474, and others (Haldimann et al. 2001)). By choosing different strains, providing different mutants of the *pir*-gene, the copy number of the ori γ based plasmids can be influenced. Ori γ plasmids replicate at a medium copy number in *pir*$^+$ strains (15 per cell, *E. coli* BW23473) or high copy number in *pir-116* strains (250 per cell, *E. coli* BW23474). The ori γ has been used for construction of the pIG-plasmid due to its relative small size (390bp). The pIG plasmid furthermore provides the before described components of the frame-selection system (lacP-controlled open reading frame, containing tat-sequence and 5'-part of intein) as well as a selection marker (kanamycin resistance).

The frame selectable transposon (figure A2 B) combines the components of the above described *Mly*I-TAG-linker with the flanking R1/R2 transposase recognition sites (compare figure A1). This combines step 1-6 of the system published by Daggett et al. to a single step, and avoids a blunt end ligation (compare figure A1). As a consequence, the restriction

endonuclease strategy had to be changed. The modified restriction endonuclease strategy is demonstrated in figure A3.

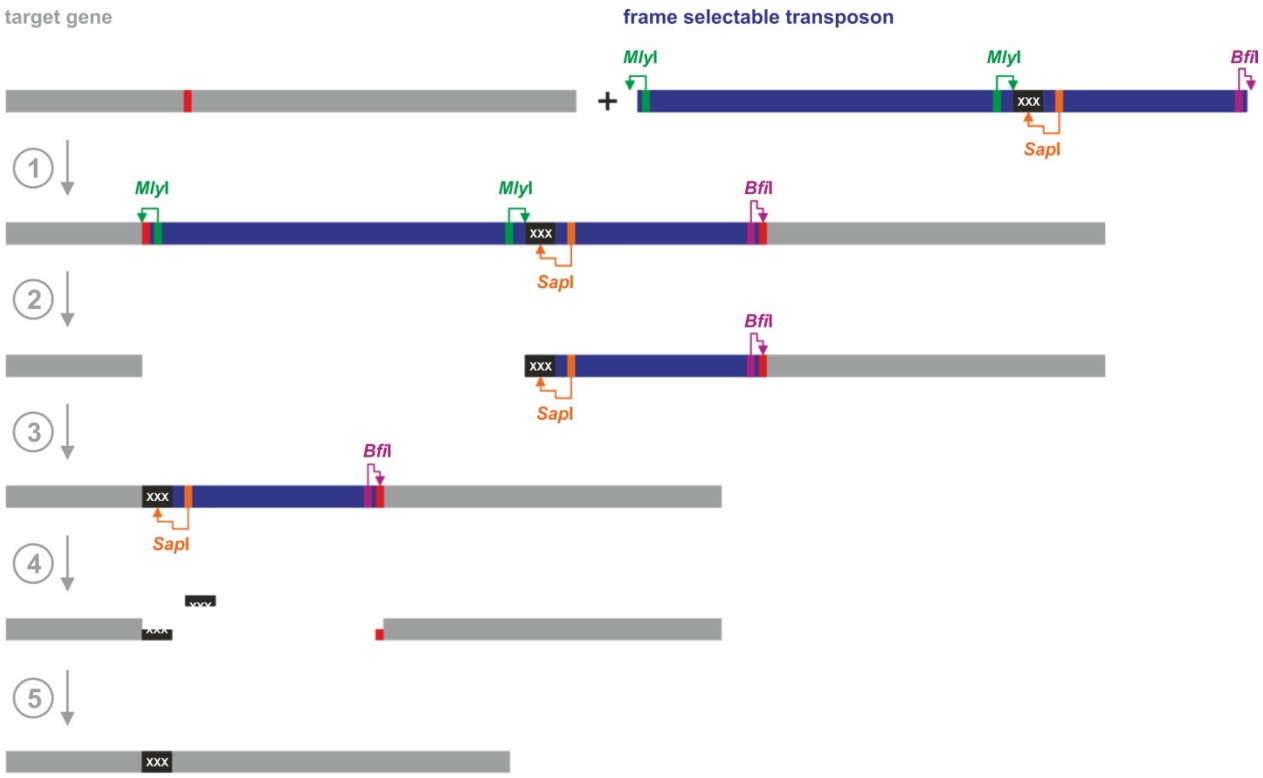


Figure A3: Restriction endonuclease strategy of the optimized system for the generation of a codon scanned library using the frame selectable transposon.
The gene (grey) is cloned into the pIG-vector (not shown). The position in the gene, where the frame selectable transposon (dark blue) is inserted by the transposase is marked in red. The scanning codon triplet is symbolized by XXX. For detailed descriptions on the workflow see main text.

Initially the random transposition of the frame selectable transposon by the MuA transposase into the target gene (subcloned in the pIG-plasmid) is performed (step 1). As described before, this leads to a duplication of five nucleotides of the target sequence, which flank the transposon. Cells transformed with a pIG plasmid, where the transposon was inserted in frame and in correct orientation into the target gene sequence, survive the following selection step due to their resistance to ampicillin. Digestion of the library with *Mly*I (step 2) removes the 5'-part of the transposon and leads to a linear product which is religated in step 3. A second restriction digestion is performed (combining *Sap*I and *Bfi*I) to remove the second part of the transposon and the original codon triplet, while leaving behind the new scanning codon triplet (step 4). Due to the fact that no second type IIS restriction enzyme which cuts blunt ended is commercially available, the sticky ended linear product has to be blunted (remove 3'-overhangs (here from the *Bfi*I site) and fill-in 5'-overhangs (here from the *Sap*I site)) using

a suitable polymerase (e.g. Pfu polymerase, which has proofreading activity), before relegation yields the final codon scanned library (step 5).

The scanning triplet in the frame selectable transposon can be exchanged after transposition by a small synthetic linker, thus providing an easy and low cost feasibility of scanning with every decided codon, without the need of different transposons. Therefore the system can be used for total amino acid exchange randomization of the target. The implementation of the synthetic linker in the scanning process is illustrated in figure A4.

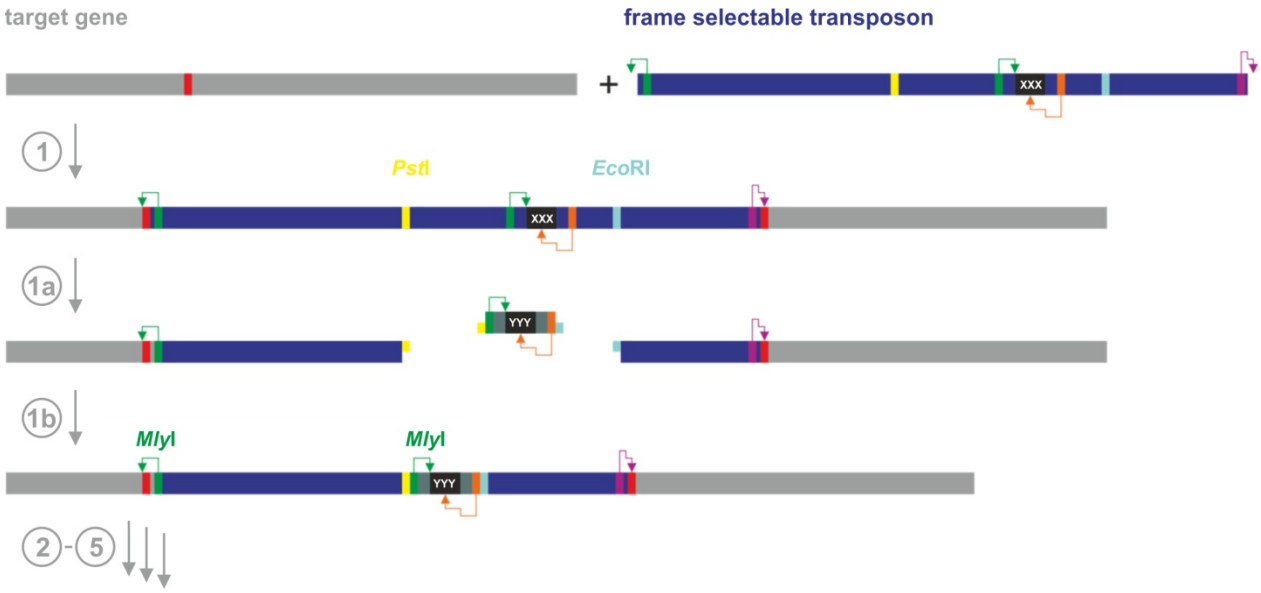


Figure A4: Additional steps for the exchange of the codon triplet in the initial library After transposition of the frame selectable transposon in step 1, its scanning codon triplet can be exchanged. Therefore the original triplet (XXX) has to be removed by a *Pst*I-*Eco*RI-digestion (step 1a) and a small synthetic linker containing the substituting triplet (YYY) has to be ligated (step 1b).

The pIG vector furthermore provides an additional module for the removal of frame-shift mutants from the final scanning library, which may occur during the optimized scanning process. Therefore, at the 3' end of the cloning site for the target gene, a loxP site (34 nucleotides) is inserted 5' to the stop codon which closes the open reading frame. At this loxP site, a second vector (likewise providing a loxP site) can be fused to the pIG-vector by *in vitro* recombination catalyzed by the Cre-recombinase (Buchholz et al. 2001; Van Duyne 2001; Ghosh et al. 2002; Jia et al. 2009). The composition of this supplementary vector (pDF), which allows for elimination of frame shift mutants, and the product of its recombination with the pIG-plasmid is depicted in figure A5.

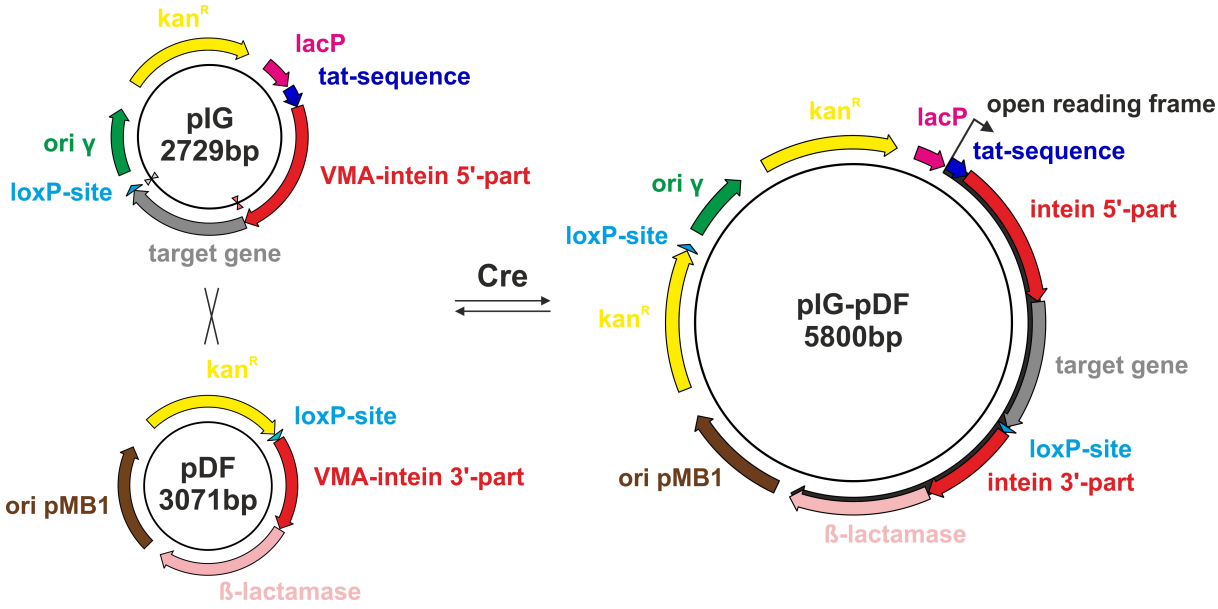


Figure A5: Elimination of frame shift variants from the final scanned library.
By means of Cre-mediated recombination of the pDF vector with the scanned library (in pIG-vector) and subsequent selection on ampicillin resistance, frame shift variants can be removed.

The emerging (re)combination plasmid "pIG-pDF" assembles an open reading frame which again mediates ampicillin resistance, when no frame shift occurred during the library generation. The mechanisms are similar to the before described, including splicing of the two intein parts and subsequent export of the β-lactamase to the periplasma. This is a simple alternative to subcloning the library into a vector allowing for the removal of frame shift variants, as published by Lutz and co-workers (pInSALect vector, Gerth et al. 2004).

As a proof of concept, the novel system for codon scanning mutagenesis was used to reproduce a recent *in vitro* molecular evolution experiment, which was performed in our group. It has been shown, that the enzyme N'-[(5'-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide isomerase (HisA) from the histidine synthesis pathway can be activated for the chemically analogous reaction from the tryptophan pathway, which is catalyzed by the enzyme phosphoribosyl anthranilate isomerase (TrpF, Leopoldseder et al. 2004). Both reactions are Amadori-rearrangements of an aminoaldose into an aminoketose (figure A6).

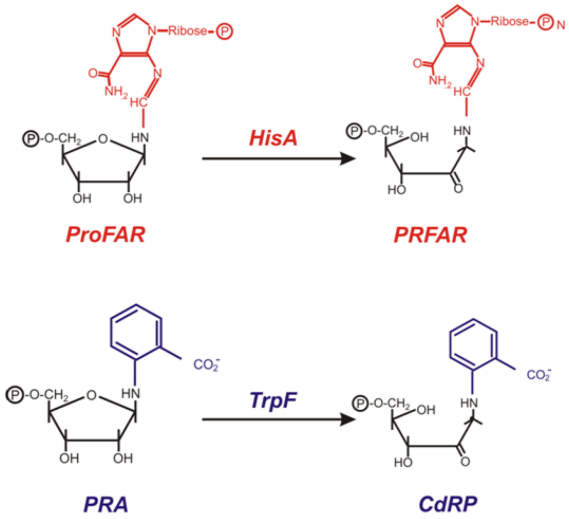


Figure A6: The two $(\beta\alpha)_8$-barrel enzymes HisA and TrpF catalyze analogous reactions.
HisA und TrpF catalyze the Amadori-rearrangement of an aminoaldose into an aminoketose. CdRP: 1-(o-carboxyphenylamino)-1-deoxyribulose-5-phosphate; HisA: ProFAR isomerase; PRA: phosphoribosyl anthranilate; PRFAR: N'-[(5'-phosphoribulosyl)-formimino]-5-aminoimidazol-4-carboxamid ribonucleotide; ProFAR: N'-[(5'-phosphoribosyl)-formimino]-5-aminoimidazol-4-carboxamid ribonucleotide; TrpF: PRA isomerase. (modified from: Claren 2008).

A single amino acid exchange in HisA is sufficient for this interconversion (either D127V or D169V in HisA from *Thermotoga maritima*). We now tested our system by scanning the *hisA* gene (from *T. maritima*) with the valine codon (GTG). In a selection step in a tryptophan auxotroph *E. coli* Δ*trpF* strain (Sterner et al. 1995) on tryptophan lacking minimal media only variants comprising an TrpF-activated HisA variant survive. When selecting the valine codon scanned *hisA*-library on minimal media plates, we found a variant carrying the D127V substitution with two additional deletions of codon triplets directly downstream to the substitution (HisA_D127VΔΔ). A kinetic characterization of the HisA_D127VΔΔ variant gave a k_{cat} of 0.12 min^{-1} and a K_M of 16 µM leading to a k_{cat}/K_M of 123 M^{-1} s^{-1} (figure A7). In comparison, the D127V variant without deletions had a k_{cat} of 0.52 min^{-1}, a K_M of 74 µM and a k_{cat}/K_M of 117 M^{-1} s^{-1} (Leopoldseder et al. 2004).

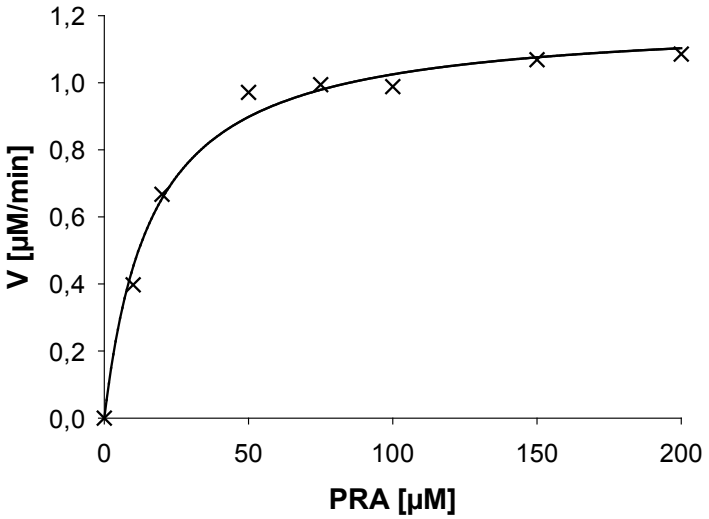


Figure A7: Steady state kinetics of the HisA variant D127VΔΔ from *T. maritima*.
Kinetic measurements were performed with 10 µM enzyme as described before (Claren et al. 2009). The data were fit using the hyperbolic function and the program Sigma Plot.

Although with this experiment we could show that our novel system works in principle, we found several steps in our procedure which still need to be optimized. In particular, the blunting of the sticky ends varies in efficiency, the restriction enzymes show a low but considerable star activity and the initial transposition step still has to be proven to be totally random.

7.3 Protocol

7.3.1 Requirements and materials:

- 5 µg of the target gene cloned into pIG via *Bam*HI/*Sal*I restriction sites (must not contain the following restriction sites: *Mly*I/*Sap*I/*Bfi*I/*Bam*HI/*Sal*I)
- 5 µg pUC19_fst_VAL (pUC19 vector containing the frame selectable transposon [fst] providing the valine codon GTG for scanning)
- 5 µg pDF or pInSal-B/S (pInSALect vector derivate, redesigned for cloning with *Bam*HI/*Sal*I)
- 5 µg of pTNA-B/S (optional, pTNA vector derivate, for cloning with *Bam*HI/*Sal*I)
- MuA-Transposase (Thermo Scientific), Cre-Rekombinase (NEB), *Mly*I-FD (Fermentas), *Bfi*I (Fermentas), *Sap*I-FD (Fermentas), *Bam*HI-HF (NEB), *Sal*I-HF (NEB), *Eco*RI-HF (NEB), *Pst*I-HF (NEB), CIP (NEB), Pfu (Promega), T4-Ligase (Fermentas), Quick Ligation Kit (NEB)
- Electro-competent *E. coli* BW23474, electro-competent *E. coli* DH5α, electro-competent *E. coli* NEB Turbo.
- LB agar plates: 21 LB agar plates (150/25 mm, GREINER bio-one) containing 50 µg/ml Kanamycin and 80 µg/ml Carbenicillin (a Carboxypenicillin which is more stable than Ampicillin), 2 LB agar plates (94/16 mm, GREINER bio-one) containing 50 µg/ml Kanamycin and 80 µg/ml Carbenicillin, 17 LB agar plates (94/16 mm) containing 50 µg/ml Kanamycin, 8 LB agar plates (94/16 mm) containing 150 µg/ml Ampicillin, 4 LB agar plates (94/16 mm) containing 30 µg/ml Chloramphenicol and 80 µg/ml Carbenicillin.
- SOC medium
- Sequencing primer:
  pIG 5’ sequencing primer 5’-GTATAAGGACAGAAAAGAAC-3’
  pIG 3’ sequencing primer 5’-CGCGAAGAAAAGCAGATAAACTC-3’
- Oligonucleotides for the small linker for exchanging the scanning triplet:
  connector_fw 5’PO_4-GAGTCTTCAGYYYTGAAGAGC-3’
  connector_rev 5’PO_4-AATTGCTCTTCAYYYCTGAAGACTCTGCA-3’
  (YYY represents the interchanging codon triplet [fw-primer] and its anticodon triplet [rev-primer])

7.3.2 Procedure:

Prearrangements:
Digest 5 µg of pUC19_fst_VAL with *Bgl*II. Purify the frame selectable transposon (fst, 1.5 kbp) by agarose gel electrophoresis.
Anneal the connector (optional) by mixing 50 µl of both 100 pM oligonucleotides (connector_fw and connector_rev, dissolved in water). Incubate the mixture in a water bath at 95 °C for 10 min and let it slowly cool down.

1st working day: Transposition reaction.

To calculate the amount (X) of the transposon (fst) for the transposition reaction, use the following equation (n = size of target gene).

$$\text{X ng (fst)} = \frac{800 \text{ ng} \cdot 1500 \text{ bp}}{1{,}3 \cdot (2800 + \text{n}) \text{ bp}}$$

Perform transposition reaction using the MuA manufacturer's buffer system:

800 ng pIG_X (pIG with subcloned target gene X)
X ng fst
MuA reaction buffer
ad 47.5 µl with H_2O
2.5 µl MuA (0.22 µg/µl)

Incubate for 4 h at 30 °C followed by a 10 min inactivating heat step at 75 °C. Dialyze the mixture > 3 h at 4 °C against double distilled water (Ø = 2.5 cm dialyze filter [Millipore, pore size 0,025 µm]).
Transformation: Electroporate ten 100 µl aliquots of electro-competent *E. coli* BW23474 with 5 µl of the mixture each. Therefore incubate the cells for 5 min on ice, add the transformation mixture, transfer the suspension into a pre-cooled electroporation cuvette and apply the electrical pulse (2500 V, 25 µF, 200 Ω) in the electroporator (time should be < 6.0 ms). Transfer the cells into 1.5 ml SOC medium, and incubate for 1 h at 37 °C while shaking.

Plate on 4-8 LB agar plates (150/25 mm) containing 50 µg/ml Kanamycin and 80 µg/ml Carbenicillin and incubate at 30 °C for selection (at least 40 h).
The transformation efficiency should be determined by plating several dilutions of the mixture on selective agar containing 50 µg/ml Kanamycin.

2nd working day: Library cleaning.

Pool the library by washing the plates with PBS and extracting the plasmids (use ten miniprep columns for plasmid preparation [GeneJET Plasmid Miniprep Kit, Thermo Scientific]). Digest 5 µg of the library with *Bam*HI and *Sal*I for 1 h at 37 °C in 20 µl and run an agarose gel. Extract the fragments at 2700 bp und 1500 bp+n (n = size of target gene) together (!) from the agarose gel (elution with 30 µl water). Religate the preparation in 40 µl with T4-DNA-Ligase (1 h at RT).
Transformation: Electroporate four 100 µl aliquots of electro-competent *E. coli* BW23474 with 2.5 µl of the mixture each, as described before.
Plate on two LB agar plates (150/25 mm) containing 50 µg/ml Kanamycin and 80 µg/ml Carbenicillin and incubate at 30 °C for selection (at least 40 h).

3rd working day: Transposon removal – part 1.

Pool the library by washing the plates with PBS and extracting the plasmids.

Optional: At this point the scanning triplet may be exchanged by ligation of the small synthetic linker (connector).

Digest 5 µg of the library with *Pst*I and *Eco*RI for 1h at 37 °C in 20 µl. Add 2 µl CIP and incubate again for 1 h at 37 °C. Run an agarose gel. Extract the fragment at 3870+n bp (n = size of target gene) from the agarose gel.

Ligation:	Ligate 5 pmol connector together with 50 ng digested library preparation using the NEB Quick ligation Kit. Perform a religation control without connector!
Transformation:	Electroporate two 100 µl aliquots of electro-competent *E. coli* BW23474 with 10 µl of the mixture each as described before.
Selection:	Plate on (at least) two LB agar plates (94/16 mm) containing 50 µg/ml Kanamycin and a 1/100 dilution of the suspension on a further plate and incubate at 30 °C for selection (at least 40 h).

Digest 5 µg of the library with *Mly*I for 20 min at 37 °C in buffer GoGreen (Fermentas) and 20 µl. Run an agarose gel. Extract the fragment at 2800+n bp (n = size of target gene) from the agarose gel (elution with 30 µl water). Religate 50-100 ng of the preparation with the NEB Quick Ligation Kit (5 min bei RT) in 20 µl.
For transformation electroporate two 100 µl aliquots of electro-competent *E. coli* BW23474 with 10 µl of the religation mixture each, as described above.
Plate 95 % of the suspension on two LB agar plates (94/16 mm) containing 50 µg/ml Kanamycin and 5 % of the suspension as a control one LB agar plate (94/16 mm) containing 50 µg/ml Kanamycin and 80 µg/ml Carbenicillin and incubate 24 h at 37 °C for selection.

4th working day: Transposon removal – part 2.

Pool the library by washing the plates with PBS and extracting the plasmids. Digest each 1 µg of the library with …

sample 1: *Sap*I (control)
sample 2: *Bfi*I (control)
sample 3: *Sap*I and *Bfi*I

… for 1 h at 37 °C in 1 x Tango buffer (Fermentas) in 20 µl, followed by a heat inactivation step for 20 min at 65 °C. Run an agarose gel (sample 1 and 2 are controls to test the fidelity of the restriction enzymes). Extract the fragment at 2720+n bp (n = size of target gene) from

sample 3 from the agarose gel (elution with 30 µl water). Use 100 ng of the preparation in the subsequent blunting reaction:

5 µl 10X Pfu buffer
1 µl dNTPs (10mM each)
ad 49 µl with water
1 µl Pfu polymerase (Promega)

Incubate the mixture for 5 min at 72 °C and chill on ice. Dialyze the mixture > 3 h at 4 °C against double distilled water (Ø = 2.5 cm dialyze filter [Millipore, pore size 0,025 µm]). Religate the library (after dialysis approximately 100 µl) with 4 µl T4-DNA-ligase for 1 h at 25 °C [a]. As a control sample, religate 20 ng of library preparation without blunting using the same procedure and 1 µl ligase to quantify the efficiency of the blunting reaction [b].

Transformation:
[a] Electroporate ten 100 µl aliquots of electro-competent *E. coli* BW23474 with 10 µl of the relegation mixture each, as described before.
[b] Electroporate one 100 µl aliquot of electro-competent *E. coli* BW23474 with 10 µl of the control sample (corresponding to 10 % of [a])

Plate [a] on ten LB agar plates (94/16 mm) containing 50 µg/ml Kanamycin. Spread a 1/10 and 1/100 dilution of the sample on two additional plates for determination of the transformation efficiency, which should be at least one order of magnitude higher than the number of codons in the gene. Plate [b] on one LB agar plate (94/16 mm) containing 50 µg/ml Kanamycin (to test the efficiency of the blunting reaction). Incubate 24 h at 37 °C for selection.

5th working day: Reading frame selection.

Pool the library by washing the plates with PBS and extracting the plasmids.

Two strategies can be pursued to remove frame shift variants from the library. The library can be subcloned into a pInSal-derivate (a) or a recombination with the frame selection vector pDF can be performed (b).

a) Subcloning: Digest 5 µg of the library with *Bam*HI-HF, *Sal*I-HF and *Eco*RI-HF (to remove variants still containing a transposon fragment) for 20 min at 37 °C in 50 µl with 1 µl enzyme each for 3h at 37 °C. In parallel, digest 5 µg pInSal-B/S with 1 µl *Bam*HI-HF and 1 µl *Sal*I-HF for 2 h at 37 °C. Add 1 µl CIP to the mixture containing pInSal-B/S and continue the incubation for 1h. Run an agarose gel with both samples and extract the fragment corresponding to the size of the target gene (elution with 30 µl water) as well as the linearized vector (4.6 kbp). Mix 50 ng of the vector preparation with X ng of the insert 20 µl.

$$\text{X ng (Insert)} = \frac{50\ \text{ng} \cdot 3 \cdot \text{Y bp (Insert)}}{4600}$$

Use T4-DNA-Ligase (Fermentas, 1 h at RT) or the NEB Quick Ligation Kit for ligation (5 min at RT).

b) Recombination: Mix 280 ng pDF and 250 ng of the library plasmid in 50 µl "buffer Cre" (NEB) and add 1 µl Cre Recombinase (NEB). Incubate for 1 h at 37 °C followed by a 10 min inactivating heat step at 70 °C.

Transformation:
a) Electroporate two 100 µl aliquots of electro-competent *E. coli* DH5α with 10 µl of the mixture each, as described above.
b) Electroporate five 100 µl aliquots of electro-competent *E. coli* DH5α with 10 µl of the mixture each, as described above.

Plate a) on four LB agar plates (94/16 mm) containing 30 µg/ml Chloramphenicol and 80 µg/ml Carbenicillin. Plate a 1/10 and a 1/100 dilution of the sample on two further plates (30 µg/ml Chloramphenicol, 80 µg/ml Carbenicillin) for determination of the transformation efficiency, which should be at least one order of magnitude higher than the number of codons in the gene. Plate b) on seven LB agar plates (94/16 mm) containing 50 µg/ml Kanamycin and 80 µg/ml Carbenicillin. One should be a 1/10 dilution and one should be a 1/100 dilution for

determination of the transformation efficiency. Incubate all plates > 40 h at 30 °C for selection.

6th working day: Subcloning (optional).

Pool the library by washing the plates with PBS and extracting the plasmids. Digest 5 µg of the library with *Bam*HI-HF and *Sal*I-HF for 2 h at 37 °C 20 µl with 1 µl enzyme each for 3 h at 37 °C. In parallel digest 5 µg pTNA-B/S with 1 µl *Bam*HI-HF and 1 µl *Sal*I-HF for 2 h at 37 °C. Add 1 µl CIP to the mixture containing 5 µg pTNA-B/S and continue the incubation for 1 h. Run an agarose gel with both samples and extract the fragment corresponding to the size of the target gene as well as the linearized vector (3.4 kbp). Mix 50 ng of the vector preparation with X ng of the insert 20 µl.

$$\text{X ng (insert)} = \frac{50\ \text{ng} \cdot 3 \cdot \text{Y bp (insert)}}{3420}$$

Use T4-DNA-Ligase (Fermentas, 1h at RT) or the NEB Quick Ligation Kit for ligation (5 min at RT).

Electroporate two 100 µl aliquots of electro-competent *E. coli* DH5α (or *E. coli* NEB Turbo) with 10 µl of the mixture each, as described above. Plate on four LB agar plates (94/16 mm) containing 150 µg/ml Ampicillin. Plate a 1/10 and a 1/100 dilution of the sample on two further plates (150 µg/ml Ampicillin) for determination of the transformation efficiency. Incubate 24 h at 37 °C for selection.

7.3.3 Sequences

pIG-plasmid

Annotation table:

kan^R	join (2904..3452, 1..267)
lacP	368..547
ORF	561..2285
tat-sequence	561..681
Intein 5'-part	684..1511
tmhisA (target gene)	1515..2249
loxP-site	2250..2283
Ori γ	2367..2756
pIG 5' sequencing primer	1439..1458
pIG 3' sequencing primer	complement (2322..2344)

Sequence:

AAGCTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATT
GATGTTGGACGAGTAGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCT
TTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGAATTGGTTAATTGGTT
GTAACACTGGCAGAGCATTACGCTGACTTGACGGGACGGCGGCTTTGTTGAATAAATCGAACTTTTGCTGAGTTGAAGGATCTCCCCGCGCGTTG
GCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAG
GCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTGGTGTGGAATTGTGAGCGGATAACAATTTCAAGGAGATATACATATGAACAATAA
CGATCTCTTTCAGGCATCACGTCGGCGTTTTCTGGCACAACTCGGCGGCTTAACCGTCGCCGGGATGCTGGGGCCGTCATTGCTAACGCCGCGAC
GTGCGACTGCGCATAATGGGTGCTTTGCCAAGGGTACCAATGTTTTAATGGCGGATGGGTCTATTGAATGTATTGAAAACATTGAGGTTGGTAAT
AAGGTCATGGGTAAAGATGGCAGACCTCGTGAGGTAATTAAATTGCCCAGAGGAAGAGAAACTATGTACAGCGTCGTGCAGAAAAGTCAGCAC
AGAGCCCACAAAAGTGATTCAAGTCGTGAAGTGCCAGAATTACTCAAGTTTACGTGTAATGCGACCCATGAGTTGGTTGTTAGAACACCTCGTA
GTGTCCGCCGTTTGTCTCGTACCATTAAGGGTGTCGAATATTTTGAAGTTATTACTTTTGAGATGGGCCAAAAGAAAGCCCCCGACGGTAGAATT
GTTGAGCTTGTCAAGGAAGTTTCAAAGAGCTACCCAATATCTGAGGGGCCTGAGAGAGCCAACGAATTAGTAGAATCCTATAGAAAGGCTTCAA
ATAAAGCTTATTTTGAGTGGACTATTGAGGCCAGAGATCTTTCTCTGTTGGGTTCCCATGTTCGTAAAGCTACCTACCAGACTTACGCGCCAATTC
TTTATGAGAATGACCACTTTTTCGACTACATGCAAAAAAGTAAGTTTCATCTCACCATTGAAGGTCCAAAAGTACTTGCTTATTTACTTGGTTTAT
GGATTGGTGATGGATTGTCTGACAGGGCAACTTTTTCGGTTGATTCCAGAGATACTTCTTTGATGGAACGTGTTACTGAATATGCTGAAAAGTTG
AATTTGTGCGCCGAGTATAAGGACAGAAAAGAACCACAAGTTGCCAAAACTGTTAATTTGTACTCTAAAGTTGTCAGAGGTACCATGGCCGGAT
CCATGCTCGTTGTCCCGGCGATAGATCTCTTCAGAGGAAAGGTAGCGAGGATGATAAAAGGAAGAAAAGAGAACACCATATTTTACGAAAAAG
ATCCCGTAGAACTGGTGGAAAAACTCATCGAAGAGGGCTTCACACTGATCCACGTGGTGGATCTCTCGAATGCGATAGAAAACAGCGGCGAAA
ATCTTCCAGTTCTCGAGAAACTCTCTGAATTTGCCGAGCACATACAGATCGGAGGCGGGATCAGATCGCTCGATTACGCGGAAAAACTCCGAAA
GCTGGGATACAGAAGACAGATCGTGAGCTCAAAGGTTCTGGAAGATCCTTCTTTCCTGAAATCCCTGAGAGAAATCGATGTGGAGCCCGTGTTC
AGTCTGGACACTCGAGGTGGAAGAGTAGCGTTCAAAGGGTGGCTGGCGGAAGAGGAAATCGACCCTGTTTCTCTTCTGAAGAGACTGAAAGAAT
ACGGCCTTGAAGAGATCGTACACACGGAGATCGAAAAAGATGGCACTCTTCAGGAGCACGATTTTTCTCTCACCAAAAAGATAGCGATCGAAGC
TGAAGTGAAAGTACTCGCAGCGGGTGGTATCTCTTCGGAGAACTCTTTGAAAACAGCGCAGAAGGTTCACACAGAAACGAACGGGCTTCTCAAA
GGTGTGATCGTGGGAAGGGCGTTTCTGGAGGGCATTCTCACAGTTGAGGTGATGAAGAGATATGCTCGCGTCGACATAACTTCGTATAATGTAT
GCTATACGAAGTTATAATGATAGCATGCATCGACCATCATCATCATCATCATTGAGTTTATCTGCTTTTCTTCGCGCTGCAGCCATGGCTAATTCC
CATGTCAGCCGTTAAGTGTTCCTGTGTCACTCAAAATTGCTTTGAGAGGCTCTAAGGGCTTCTCAGTGCGTTACATCCCTGGCTTGTTGTCCACAA
CCGTTAAACCTTAAAAGCTTTAAAAGCCTTATATATTCTTTTTTTTCTTATAAAACTTAAAACCTTAGAGGCTATTTAAGTTGCTGATTTATATTAA
TTTTATTGTTCAAACATGAGAGCTTAGTACGTGAAACATGAGAGCTTAGTACGTTAGCCATGAGAGCTTAGTACGTTAGCCATGAGGGTTTAGTT
CGTTAAACATGAGAGCTTAGTACGTTAAACATGAGAGCTTAGTACGTGAAACATGAGAGCTTAGTACGTACTATCAACAGGTTGAACTGCTGAT
CTTCAGATCCTCTACGCCGGACGCATCGTGGCCGGATCTTGCGGCCGCATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTA
AGGGATTTTGGTCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCTTGCTC
GAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGA
TTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACT
GGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGGAAAACA
GCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGT
AATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTAGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGA
GCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCAT

pDF-plasmid

Annotation table:

kan^R	873..1688
loxP	1695..1728
Intein 3'-part	1731..2273
ß-lactamase gene	2280..3071
ori-pMB1	91..705

Sequence:

GGTACCGATATCTGCTTTTCTTCGCGAATTAATTCCGCTTCGCAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCG
TTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA
GATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAA
GCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAG
CCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAATCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGAT
TAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCT
CTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCA
GCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGG
ATTTTGGTCATGAGTTGTGTCTCAAAATCTCTGATGTTACATTGCACAAGATAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAA
ACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCTTGCTCGAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGG
GTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACAT
GGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTAT
CCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGGAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATA
TTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTAGCTCAGGC
GCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCAT
AAGCTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATT
GATGTTGGACGAGTAGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCT
TTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAACTGCAGATAACTTCGTATAA
TGTATGCTATACGAAGTTATCTGGTATTCGCAATAATCTTAATACTGAGAATCCATTATGGGACGCTATTGTTGGCTTAGGATTCTTGAAGGACG
GTGTCAAAAATATTCCTTCTTTCTTGTCTACGGACAATATCGGTACTCGTGAAACATTTCTTGCTGGTCTAATTGATTCTGATGGCTATGTTACTG
ATGAGCATGGTATTAAAGCAACAATAAAGACAATTCATACTTCTGTCAGAGATGGTTTGGTTTCCCTTGCTCGTTCTTTAGGCTTAGTAGTCTCGG
TTAACGCAGAACCTGCTAAGGTTGACATGAATGGCACCAAACATAAAATTAGTTATGCTATTTATATGTCTGGTGGAGATGTTTTGCTTAACGTT
CTTTCGAAGTGTGCCGGCTCTAAAAAATTCAGGCCTGCTCCCGCCGCTGCTTTTGCACGTGAGTGCCGCGGATTTTATTTCGAGTTACAAGAATTG
AAGGAAGACGATTATTATGGGATTACTTTATCTGATGATTCTGATCATCAGTTTTTGCTTGCCAACCAGGTTGTCGTCCATAATTGCGCTAGTCAC
CCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTG
AGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAG
CAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAG
AATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTG
CACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTG
CAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAA
AGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTG
CAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGCCAGGCAACTATGGATGAACGAAATAGACAGATCGC
TGAGATAGGTGCCTCACTGATTAAGCATTGGTAA

Frame selectable transposon

Annotation table:

R1/R2	complement (13..54)
Intein 3'-part	65..607
ß-lactamase gene	614..1405
valin codon (GTG)	1422..1424
R1/R2	1448..1489

Sequence:

AGATCTGACTCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAACGGATCGATCGGTATTCGCAATAATCTTAATACTGAGAAT
CCATTATGGGACGCTATTGTTGGCTTAGGATTCTTGAAGGACGGTGTCAAAAATATTCCTTCTTTCTTGTCTACGGACAATATCGGTACTCGTGAA
ACATTTCTTGCTGGTCTAATTGATTCTGATGGCTATGTTACTGATGAGCATGGTATTAAAGCAACAATAAAGACAATTCATACTTCTGTCAGAGA
TGGTTTGGTTTCCCTTGCTCGTTCTTTAGGCTTAGTAGTCTCGGTTAACGCAGAACCTGCTAAGGTTGACATGAATGGCACCAAACATAAAATTA
GTTATGCTATTTATATGTCTGGTGGAGATGTTTTGCTTAACGTTCTTTCGAAGTGTGCCGGCTCTAAAAAATTCAGGCCTGCTCCCGCCGCTGCTT
TTGCACGTGAGTGCCGCGGATTTTATTTCGAGTTACAAGAATTGAAGGAAGACGATTATTATGGGATTACTTTATCTGATGATTCTGATCATCAG
TTTTTGCTTGCCAACCAGGTTGTCGTCCATAATTGCGCTAGTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACG
AGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTC
TGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCA
CCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTAC
TTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCT
GAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACG
CTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGC
TGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGA
CGGGGAGCCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAAACTAGTGAGTCTTCAGG
TGTGAAGAGCGAATTCATCGATCCGTTTTCGCATTTATCGTGAAACGCTTTCGCGTTTTTCGTGCGCACTGGGAGATCT

The frame selectable transposon is subcloned in the *Bgl*II site of the pUC19-vector.

pTNA-B/S

For construction of the pTNA-B/S plasmid a synthetic linker was subcloned between the *Bam*HI and *Hind*III sites in the pTNA plasmid (Henn-Sax et al. 2002). The synthetic linker was generated by annealing the oligonucleotides with the following sequences:

Linker_fw 5'-GATCCGTCGACCATCACCATCACCATCACTAAACCGGTGTCATTCCGTA-3'

Linker_rev 5'-AGCTTACGGAATGACACCGGTTTAGTGATGGTGATGGTGATGGTCGACG-3'

pInSal-B/S

This vector is a derivate of pInSALect (Gerth et al. 2004).

Annotation table:

cam^R	complement(300..959)
tat-sequence	2268..2387
VMA Intein 5'-part	2394..3005
Endonuclease fragment I	3006..3215
Endonuclease fragment II	3249..3584
VMA Intein 3'-part	3585..3788
ß-lactamase gene	3798..4589
pIG 5' sequencing primer	3146..3165
pInSal 3'-Seq-primer	complement(3288..3306)

Sequence:

TTTTCCAATGATGAGCACTTTTGATCAGAAAAAAAGGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCAGGCAT
TTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACC
GACGACCGGGTCGAATTTGCTTTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGGGCACCAATA
ACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGG
CATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCAT
ATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCC
AGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTC
AGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGATGAGCAT
TCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACG
GTCTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGAT
TTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGACAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGA
ACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGA
AGTGATCTTCCGTCACAGGTATTTATTCGGTCGAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTT
TCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAA
ACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATA
CTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTG
CTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTG
CACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAA
GGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGT
CGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTAC
GGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGA
TACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCG
TTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCAT
TAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGATCTCGACATGAACA
ATAACGATCTCTTTCAGGCATCACGTCGGCGTTTTCTGGCACAACTCGGCGGCTTAACCGTCGCCGGGATGCTGGGGCCGTCATTGTTAACGCCG
CGACGTGCGACTGCGCATAATGGGTGCTTTGCCAAGGGTACCAATGTTTTAATGGCGGATGGGTCTATTGAATGTATTGAAAACATTGAGGTTGG
TAATAAGGTCATGGGTAAAGATGGCAGACCTCGTGAGGTAATTAAATTGCCCAGAGGAAGAGAAACTATGTACAGCGTCGTGCAGAAAAGTCA
GCACAGAGCCCACAAAAGTGACTCAAGTCGTGAAGTGCCAGAATTACTCAAGTTTACGTGTAATGCGACCCATGAGTTGGTTGTTAGAACACCT
CGTAGTGTCCGCCGTTTGTCTCGTACCATTAAGGGTGTCGAATATTTTGAAGTTATTACTTTTGAGATGGGCCAAAAGAAAGCCCCCGACGGTAG
AATTGTTGAGCTTGTCAAGGAAGTTTCAAAGAGCTACCCAATATCTGAGGGGCCTGAGAGAGCCAACGAATTAGTAGAATCCTATAGAAAGGCT
TCAAATAAAGCTTATTTTGAGTGGACTATTGAGGCCAGAGATCTTTCTCTGTTGGGTTCCCATGTTCGTAAAGCTACCTACCAGACTTACGCTCCA
ATTCTTTATGAGAATGACCACTTTTTCGACTACATGCAAAAAAGTAAGTTTCATCTCACCATTGAAGGTCCAAAAGTACTTGCTTATTTACTTGGT
TTATGGATTGGTGATGGATTGTCTGACAGGGCAACTTTTTCGGTTGATTCCAGAGATACTTCTTTGATGGAACGTGTTACTGAATATGCTGAAAA
GTTGAATTTGTGCGCCGAGTATAAGGACAGAAAAGAACCACAAGTTGCCAAAACTGTTAATTTGTACTCTAAAGTTGTCAGAGGTCATATGGGA
TCCGAATTCGTCGACCTCGAGACTAGTGGTATTCGCAATAATCTTAATACTGAGAATCCATTATGGGACGCTATTGTTGGCTTAGGATTCTTGAA
GGACGGTGTCAAAAATATTCCTTCTTTCTTGTCTACGGACAATATCGGTACTCGTGAAACATTTCTTGCTGGTCTAATTGATTCTGATGGCTATGT
TACTGATGAGCATGGTATTAAAGCAACAATAAAGACAATTCATACTTCTGTCAGAGATGGTTTGGTTTCCCTTGCTCGTTCTTTAGGCTTAGTAGT
CTCGGTTAACGCAGAACCTGCTAAGGTTGACATGAATGGCACCAAACATAAAATTAGTTATGCTATTTATATGTCTGGTGGAGATGTTTTGCTTA
ACGTTCTTTCGAAGTGTGCCGGCTCTAAAAAATTCAGGCCTGCTCCCGCCGCTGCTTTTGCACGTGAGTGCCGCGGATTTTATTTCGAGTTACAAG
AATTGAAGGAAGACGATTATTATGGGATTACTTTATCTGATGATTCTGATCATCAGTTTTTGCTTGCCAACCAGGTTGTCGTCCATAATTGCGCTA
GTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGA
TCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGC
AAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGT
AAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCT
TTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGA
TGCCTGCAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGC
GGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTA
TCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACA
GATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAAGAATTCGAT

Acknowledgement

First, I would like to gratefully thank my supervisor Prof. Dr. Reinhard Sterner for his constant, reliable and encouraging guidance throughout this thesis. I am very grateful for his generous and pleasant support in all scientific as well as personal projects. Further on, I would like to thank Dr. Patrick Babinger for his encouragement, expertise, practical advice and support. In his friendly, open minded and patient way he was an excellent supervisor and most valuable discussion partner.

I would like to thank Prof. Dr. Christine Ziegler for enabling protein crystallization in Regensburg, and for reviewing this dissertation. In addition, I am grateful to Prof. Dr. Harald Kolmar, who mentored this thesis in the context of the International Graduate School of Life Science.

I am very thankful to Prof. Dr. Rainer Merkl for his ongoing interest in all my projects and many helpful and most cheerful as much as perceptive discussions.

Dr. Chitra Rajendran introduced me with great expertise and in a very amiable manner to protein crystallography and static light scattering measurements.

I would like to thank my collaboration partners from Finland, Prof. Dr. Rikkert Wierenga, Dr. Evangelia Kapetaniou and Esa-Pekka Kumpula for their valuable contributions to this thesis.

Also, I am very grateful to my research partners Dr. Hermann Zellner, Barbara Beer and Mona Linde, who were excellent co-workers and discussion partners on the PcrB-project and especially to Dr. Harald Guldan, who introduced me to the topic.

My student interns Julia Pickel, Daniel Rackl and Patricia Seidel contributed valuably to my work and more than once made my day with their drive and enthusiasm.

Special thanks for the perfect teamwork and all the personal support to my dear friend and brilliant research collaboration partner Florian Busch.

I also would like to thank for pleasant and reliable assistance by Christiane Endres, Hermine Reisner, Sonja Fuchs, Barbara Kellerer and Klaus-Jürgen Tiefenbach as well as for administrative support by Claudia Pauer.

Thanks to my office colleagues for a homey atmosphere with plenty of coffee and fun.

Cordial thanks to all current and former members of the Sterner and Merkl research group for an excellent working atmosphere, mutual support and lots of valuable discussions. Thanks furthermore for marvelous excursions, wonderful parties, and countless hilarious opportunities to improve my table soccer skills!

Finally, I am deeply grateful to my parents and Yasmin for their unlimited encouragement and support. Without their help all this would not have been possible.