Received: 15 October 2020 | Revised: 9 December 2020 | Accepted: 11 December 2020

DOI: 10.1002/pro.4016

FULL-LENGTH PAPERS

THE PROTEIN SOCIETY WILEY

Hexamerization and thermostability emerged very early during geranylgeranylglyceryl phosphate synthase evolution

Cosimo Kropp | **Kristina Straub** | **Mona Linde** | **Patrick Babinger**

Institute of Biophysics and Physical Biochemistry, Regensburg Center for Biochemistry, University of Regensburg, Regensburg, Germany

Correspondence
Patrick Babinger, Institute of Biophysics and Physical Biochemistry, Regensburg Center for Biochemistry, University of Regensburg, 93040 Regensburg, Germany.
Email: patrick.babinger@ur.de

Abstract

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

KEYWORDS

ancestral sequence reconstruction, enzyme, ether lipids, oligomerization, protein evolution, protein stability, thermostability

Abbreviations: AncGGGPS2_N.., sometimes abbreviated as _N.., reconstructed ancestral variant of group II GGGPS at node N..; ASR, ancestral sequence reconstruction; DSC, differential scanning calorimetry; G1P, glycerol 1-phosphate; G3P, glycerol 3-phosphate; GGG, geranylgeranylglycerol; GGGP, geranylgeranylglyceryl phosphate; GGGPS, geranylgeranylglyceryl phosphate synthase; GGPP, geranylgeranyl pyrophosphate; IPTG, isopropyl-β-D-1-thiogalactopyranoside; ITC, isothermal titration calorimetry; LCA, last common ancestor; nanoDSF, differential scanning fluorimetry; SEC, size exclusion chromatography; SLS, static light scattering.

1 | INTRODUCTION

The membrane lipids of Archaea have a quite different composition than those of Bacteria and Eukaryotes. While the latter consist of fatty acids that are linked to a glycerol 3-phosphate (G3P) backbone by ester bonds, archaeal membrane lipids are ether lipids. Their backbone is a glycerol 1-phosphate (G1P), to which polyprenyl moieties are linked.[1] The formation of the first ether bond is catalyzed by the enzyme geranylgeranylglyceryl phosphate synthase (GGGPS), which usually transfers a geranylgeranyl residue (C20; four isoprene units) to the G1P. GGGPS is very interesting in the context of evolution, because it is a key enzyme in the context of this so-called "lipid divide" between the kingdoms of Archaea on the one hand, and Bacteria and Eukarya on the other.[2] Recent studies have unexpectedly revealed that GGGPS-like enzymes also occur in some bacteria. There, ether lipids with either four (Bacteroidetes) or seven (Firmicutes) isoprene units are synthesized,[3–5] but their biological function is still unknown. In Firmicutes, they are further processed by acetylation,[3,6] while a recent study suggests that bacteria of the Fibrobacteres–Chlorobi–Bacteroidetes (FCB) group might synthesize prototypical archaeal ether lipids.[7]

The GGGPS enzyme family can be divided into two groups, both comprising archaeal and bacterial members (Figure 1). While the enzymes of Firmicutes and of a few Euryarchaeota, among them all Halobacteria, belong to group I, most Archaea and the Bacteroidetes possess a group II enzyme. In the recent years, the oligomerization states and structures of many GGGPS representatives have been experimentally determined (as summarized and referenced in Table S1). Up to now, only dimeric variants have been discovered within group I, while most group II enzymes have been described as homohexamers. Recently, we have shown that the $(\beta\alpha)_8$-barrel backbone of several GGGPS enzymes (mtGGGPS, tkGGGPS, taGGGPS, slGGGPS, fjGGGPS) is structurally stable at hyperthermophilic conditions, independent of the oligomerization state. However, some GGGPS like mtGGGPS or fjGGGPS show a higher flexibility of the active site, going along which higher activity. Hexamerization stabilizes this portion of the protein, which otherwise would collapse at elevated temperatures.[8]

The hexamer is formed as a "trimer of dimers", where the conformation of the dimeric building blocks is identical to those of the natively dimeric GGGPS enzymes.[4,8] The conformation of the hexamer is stabilized by three individual contact interfaces between the protomers: The "dimer module interface" is symmetric and is present in dimeric GGGPS as well. The hexamer is stabilized by an additional symmetric interface between the dimer modules which we called "interconnecting interface", plus an asymmetric interface that we called "ring interface",

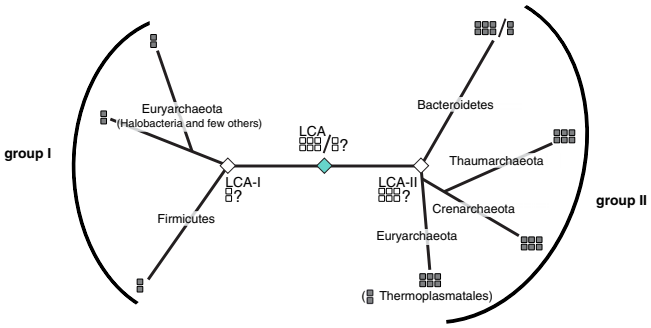

FIGURE 1 Simplified phylogenetic tree of the GGGPS enzyme family. The oligomerization states of a number of representatives have been experimentally identified (Table S1) and are indicated by 2 or 6 grey rectangles (dimeric / hexameric). The oligomerization states of Crenarchaeota and Thaumarchaeota where elucidated within this work. LCA, last common ancestor; LCA-I / LCA-II, LCA of group I / group II. Putative oligomerization states of the LCAs, as discussed within this work, are indicated by white rectangles

because it stabilizes rings of three protomers each. As a hallmark of hexameric GGGPS, the presence of an aromatic amino acid ("aromatic anchor") within the ring interface has been identified. It forms a cation-π interaction with a lysine or arginine from the opposing protomer that is supposed to be essential for hexamerization. However, several group II sequences contain a histidine at the position of the aromatic anchor, particularly a large group of crenarchaeal sequences, which led to the assumption that they might be dimers.[4] The corresponding proteins were not amenable to *in vitro* studies up to now, so that their true oligomerization state remained unclear.

Ancestral sequence reconstruction (ASR) is a powerful method to deduce environmental conditions in the Precambrian era and to study the evolution of proteins in terms of their structural stability as well as their interaction specificities with ligands and other proteins.[9–11] We recently have reconstructed the sequences of ancestral GGGPS enzymes[12] and now provide a comprehensive in vitro characterization of nine of these predecessors to trace back the evolution of hexameric GGGPS. We were able to successfully purify and study crenarchaeal predecessors with a His at the above mentioned "anchor position" and show here that they also form hexamers. Furthermore, we tested the ancestral GGGPS proteins for their thermostability. Our results indicate that hexamerization and thermostability were already present at the very beginning of group II GGGPS evolution, while the above mentioned fine-tuning of active site flexibility developed late and independent of the emergence of hexamerization. Furthermore, we conclude that the few extant dimeric representatives among group II (Thermoplasmatales, Flavobacteriales) emerged by secondary events.

2 | RESULTS

2.1 | Ancestral sequence reconstruction of group II GGGPS

Ancestral sequence reconstruction of GGGPS has been performed to evaluate FitSS4ASR, a tool to prepare representative sequence sets for ASR, and is described in detail elsewhere.[12] We initially strived to perform ASR with a sequence set comprising both GGGPS groups I and II. However, this turned out to be not possible, because the length of the edge that interconnects the nodes representing the ancestors of groups I and II was too long (> 4 substitutions per site).[12] We then concentrated on the more interesting group II with its diverse oligomerization states. For practical reasons, all experiments presented here are based on the ASR from the manually curated sequence set as described earlier,[12] comprising 87 extant GGGPS group II variants. A simplified and annotated phylogenetic tree deduced from these sequences is shown in Figure 2; the complete tree with all nodes and leaves is shown in Figure S1. The names and properties of all GGGPS proteins used in this study are summarized in Table S1 and Table S2. We did not evaluate the properties of the last common ancestor (LCA) of all group II enzymes (LCA-II, N1), because it

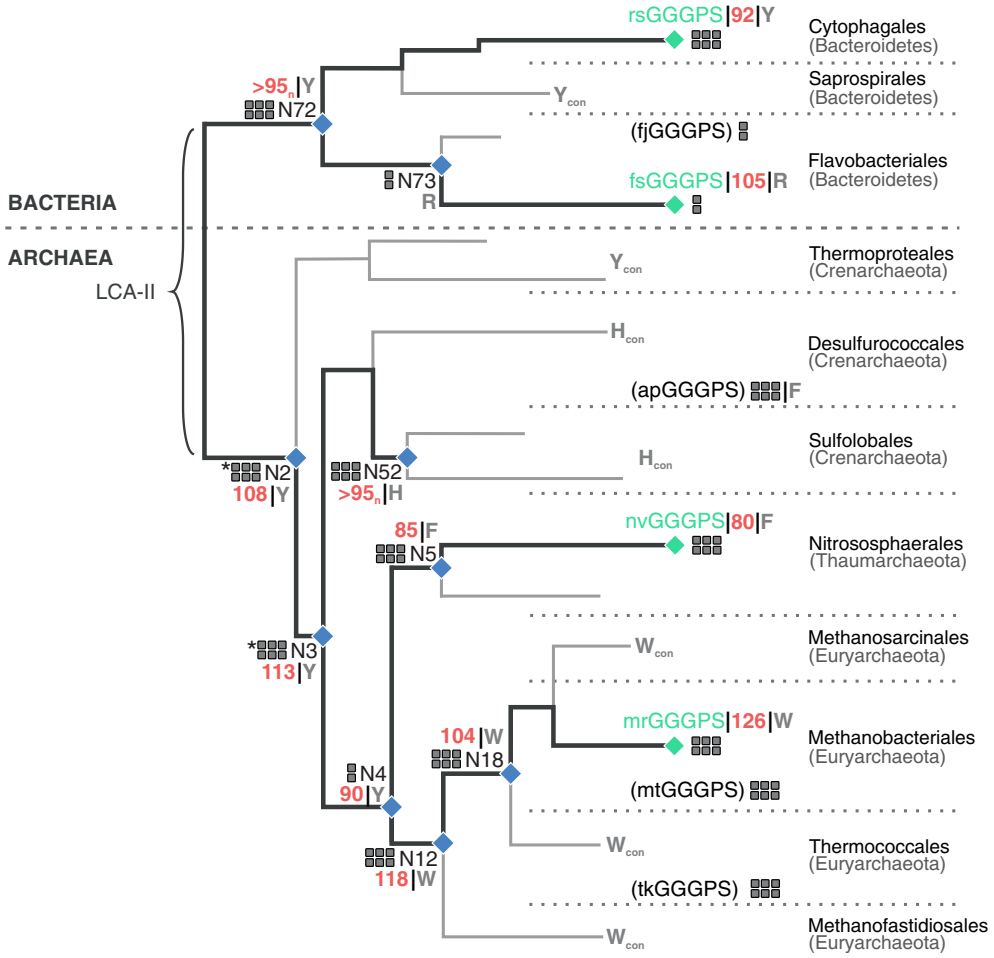

FIGURE 2 Simplified phylogenetic tree used for ASR of GGGPS group II predecessors. Nodes representing ancestral sequences that were characterized experimentally are marked with blue dots and are labeled (N..). Leaves representing extant sequences that were characterized experimentally and that contributed to the calculation of the tree are shown with their names (cf. Table S2) and labeled in green. Protein names in brackets stand for representative proteins of the respective phyla which have been characterized previously (cf. Table S1), but whose sequences have not been used to calculate the tree. Oligomerization states of ancestral and extant proteins are indicated by grey rectangles, dimer-hexamer equilibria are marked by a *. The identity of the amino acid located at the "aromatic anchor" position is given as a grey letter. For branches with no studied extant proteins the most conserved residue at that position is given (index $_{con}$). In red numbers, the transition temperatures T_2 of thermal denaturation [°C] are given. The T_2 values result from DSC experiments (no index), or from nanoDSF experiments (index $_n$; Table 1). Branches that denote evolutionary paths with nodes representing GGGPS sequences that have been characterized within this study are shown in thick lines. Branches with nodes without experimental data are indicated by thin lines or are omitted for clarity. The full tree is shown in Figure S1. Close to the leaves, the names of the orders and phyla of the extant representatives are given

would require to use an outgroup to reconstruct a confident sequence for it. We tested several group I sequences as outgroup, but this resulted in non-robust trees. We use the same nomenclature to address the reconstructed proteins as introduced earlier,[12] "AncGGGPS2_N.." stands for "ancestor of group II GGGPS" at node N.. .

2.2 | Probably all crenarchaeal GGGPS are hexamers

We have postulated in a previous study that an aromatic residue ("aromatic anchor"; Trp, Phe, Tyr) in the so called "ring interface" is a hallmark of hexameric GGGPS.[4] This residue forms a cation-π interaction with a lysine or arginine of the opposite protomer, and when the aromatic residue is mutated to an alanine, the hexamer falls apart into dimers. An analysis of group II GGGPS sequences by calculation of sequence similarity networks (SSN) revealed that in most sequence clusters this aromatic residue is highly conserved, except in two of them. One of the two clusters representing bacterial group II GGGPS (represented by the branch "Flavobacteriales" in Figure 2) has charged residues at this position (Figure 3(a)ii), and accordingly, extant representatives of this cluster (fjGGGPS, zpGGGPS; Table S1) have been demonstrated to be dimeric.[4] The other sequence cluster without the aromatic residue is represented by the branches "Desulfurococcales" and "Sulfolobales" in Figure 2. In these sequences, a histidine is highly conserved at this position (Figure 3(a)iii). Extant proteins of these branches have not been accessible by heterologous expression up to now except hexameric apGGGPS, which is quite diverse in sequence and has a Phe at the anchor position.[4] We now expressed synthetic genes of ancestral GGGPS and were successful in purifying a predecessor protein of the "Sulfolobales" branch, AncGGGPS2_N52 (Figure 2, Table S1, Table S2), which also has a His at the anchor position. AncGGGPS2_N52 displays high activity in an enzymatic assay (Figure S2) and shows a well-defined structure in circular dichroism (CD) spectroscopy (-Figure S3A). Analytical size exclusion chromatography (SEC) revealed that AncGGGPS2_N52 forms a hexamer (Figure 3(b)), indicating that His might replace the conventional aromatic residues Trp, Phe, Tyr as the "aromatic anchor". In fact, His is less frequently observed in cation-π interactions and usually rather plays the cation role in its protonated form, but it can act as the aromatic motif in such interactions.[13–15] In the Sulfolobales sequences, the position of the cationic interaction partner is occupied by a conserved Arg (Figure 3(a)iii).

We have previously shown for mtGGGPS that a mutation of the "aromatic anchor" to alanine (mtGGGPS_W141A;

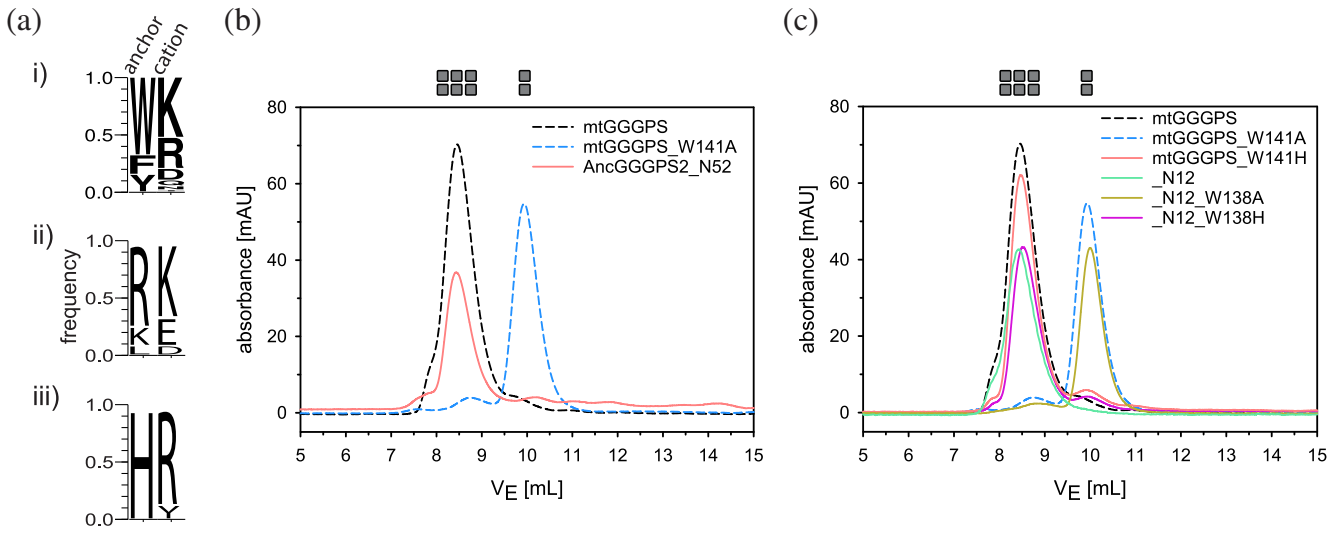

FIGURE 3 Analysis of GGGPS variants with histidine as "aromatic anchor". (a) Relative frequency of amino acid residues at the positions that form a cation-π interaction in hexameric GGGPS variants. Anchor, "aromatic anchor" residue; cation, cationic counterpart. i) Sequences from all extant proteins used in ASR except Flavobacteriales, Sulfolobales and Desulfurococcales (presumed hexamers; Figure 2, Figure S1); ii) sequences from Flavobacteriales, iii) sequences from Sulfolobales and Desulfurococcales. (b) Analytical size exclusion chromatography of AncGGGPS2_N52. (c) Analytical size exclusion chromatography of the variants mtGGGPS_W141H, AncGGGPS2_N12, AncGGGPS2_N12_W138H and AncGGGPS2_N12_W138A. The proteins (40 μM subunit concentration) were applied to a S75 analytical column equilibrated with 50 mM potassium phosphate, pH 7.5, 300 mM KCl. Elution was performed at a flow rate of 0.5 mL/min and followed by measuring the absorbance at 280 nm, which was plotted against the elution volume. mtGGGPS_wt and mtGGGPS_W141A (dashed lines) served as references for the hexameric and dimeric oligomerization states. The oligomerization state is indicated by gray symbols. The oligomerization state of AncGGGPS2_N12 has already been determined in a previous study[12]

serving as reference in Figure 3) breaks the hexamer down into dimers.[4] To demonstrate that histidine does in fact play a similar role for hexamerization like the common aromatic residues do, we first introduced the corresponding His-to-Ala exchange into AncGGGPS2_N52, but the resulting protein strongly tended to aggregation and could not be used for further experiments. As an alternative experiment, we changed the Trp in hexameric mtGGGPS_wt to His. The resulting variant mtGGGPS_W141H still eluted as a hexamer in analytical SEC (Figure 3(c)), which supports that the His can still form a cation-π interaction with the opposing cation (a Lys in mtGGGPS). Additionally, we randomly chose the hexameric ancestral protein AncGGGPS2_N12 (described later in detail) and introduced the W138H exchange, which again did not affect the oligomerization state (Figure 3(c)). In contrast, the hexamer was disrupted in the variant AncGGG PS2_N12_W138A. We conclude that histidine with its heterocyclic aromatic ring also functions as an "aromatic anchor" to stabilize the hexamer, and suspect that all extant sulfolobic GGGPS proteins are hexamers as well. In conclusion, almost all group II GGGPS enzymes must be considered to be hexameric, except the bacterial proteins from Flavobacteriales and a very small number of archaeal proteins, mainly from Thermoplasmatales,[4,16,17] as summarized in Table S1.

2.3 | Hexamerization emerged very early during group II GGGPS evolution

To study the time point of emergence of hexamerization during group II GGGPS evolution, our rationale was to follow the evolutionary paths from the well-characterized extant proteins fjGGGPS (dimeric; Flavobacteriales) and mtGGGPS (hexameric; Methanobacteriales) to LCA-II (Figure 2). We expressed synthetic genes encoding seven ancestral GGGPS proteins in *E. coli*, purified the proteins and analyzed their oligomerization states by SEC (Figure 4). The results are summarized in Table S1 and Figure 2 (grey symbols). Additionally, activity assays (Figure S2) were performed, and the structural integrity of the proteins was analyzed by CD spectroscopy (Figure S3B). AncGGGPS2_N5, _N12, _N18 and _N72 turned out to be active proteins and hexameric (Figure 4(a)). AncGGGPS2_N2 and _N3 are also active proteins, but eluted at a volume between the dimeric and hexameric controls at the initially used protein concentration of 40 μM (Figure 4(b)). SEC experiments with a series of different protein concentrations revealed a dimer-hexamer equilibrium for these variants with tendency to the hexamer under the given experimental conditions (Figure S4). Qualitatively, these variants can still be considered as hexamers, and they possess all characteristics of the hexamer-specific contact interfaces like the "aromatic anchor" and the cation. A recent study on the robustness of reconstructed ancestral protein functions has revealed that quantitative markers of function - like dissociation constants - may vary, while qualitative conclusions on reconstructed protein functions are robust.[18] Only AncGGGPS2_N4 elutes as a dimer in SEC (Figure 4(b)), although it also contains the "aromatic anchor" and the cation. Furthermore, this protein is inactive (Figure S2). As discussed below, this suggests that the dimeric nature of AncGGGPS2_N4 might be an artifact due to sequence reconstruction. From the evolutionary point of

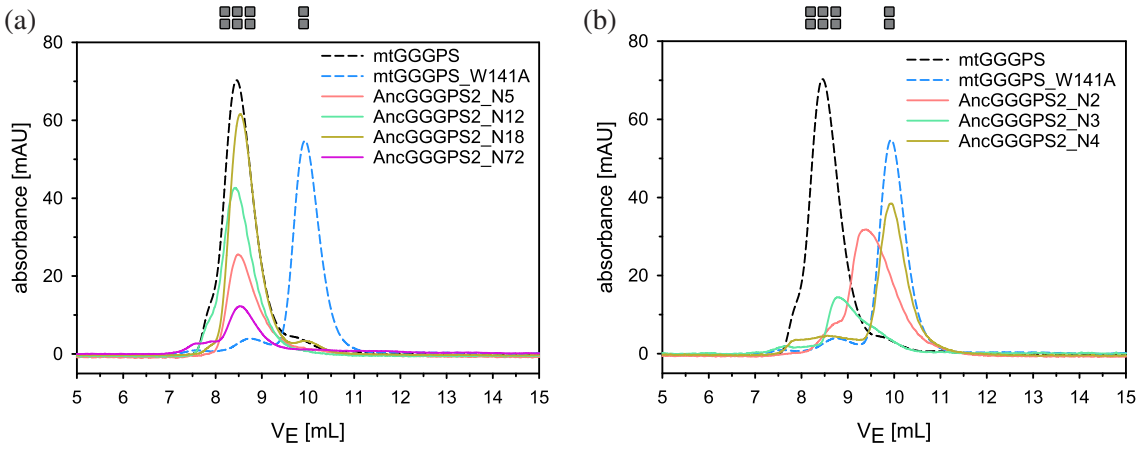

FIGURE 4 Analytical size exclusion chromatography of ancestral GGGPS proteins. (a) Hexameric proteins. (b) Proteins that are dimeric or form dimer-hexamer equilibria at the given concentration. The identities of the proteins are given in the legend. The proteins (40 μM, subunit concentration) were applied to a S75 analytical column equilibrated with 50 mM potassium phosphate, pH 7.5, 300 mM KCl. Elution was performed at a flow rate of 0.5 mL/min and followed by measuring the absorbance at 280 nm, which was plotted against the elution volume. mtGGGPS_wt and mtGGGPS_W141A (dashed lines) served as references for the hexameric and dimeric oligomerization states. The oligomerization state is indicated by gray symbols. The oligomerization states of AncGGGPS2_N4, _N5 and _N12 have already been determined in a previous study[12]

view, there is no rationale why _N4 should be a dimer, while _N3, _N12 and _N5 are hexamers. In summary, these results indicate that hexamerization must have occurred very early in group II GGGPS evolution, but because we could not reconstruct a robust sequence for the group II LCA (LCA-II; as discussed above), we cannot clearly define whether LCA-II already has been hexameric or not (Figure 1).

To shed light on the impairments of AncGGGPS2_N4, we studied this protein in more detail. We have shown previously that a single mutation (W141A) that disrupts the mtGGGPS hexamer, also drastically worsens the catalytic efficiency (k_{cat}/K_M (G1P) 380x lower, K_M (G1P) 50x higher), although the mutated position is not directly associated with the active site. Natively dimeric GGGPS like fjGGGPS are highly active, however.[4,8] A similar correlation between catalytic activity and oligomerization state of modified N4 variants would support the hypothesis that _N4 has a misbalanced hexamerization interface due to artefacts in sequence reconstruction. Along these lines, we transplanted the ring interface of hexameric AncGGGPS_N12 into AncGGGPS2_N4 (Figure S5). The resulting variant AncGGGPS_N4_IF_n12 eluted as a hexamer in analytical SEC (Figure 5) and showed GGGPS activity (Figure S2). As a reverse experiment, we transplanted the ring interface from AncGGGPS2_N4 to AncGGGPS_N12. The resulting variant AncGGGPS_N12_IF_n4 eluted as a dimer in analytical SEC (Figure 5) and showed no GGGPS activity (Figure S2). Furthermore, isothermal titration calorimetry (ITC) experiments indicated that AncGGGPS2_N4 couldn't bind the substrate G1P any more (Figure S6), which is tantamount to the severely increased K_M (G1P) in mtGGGPS_W141A. In summary, these results strongly support that the dimeric nature of AncGGGPS2_N4 is an artefact from ASR due to a misbalanced ring interface. Furthermore, additional experiments indicate that the interplay of ring interface and interconnecting interface might be impaired as well in AncGGGPS2_N4 (data not shown). We have previously shown that both interfaces cooperate to keep the hexamer stable.[8] Ancestral sequence reconstruction, however, calculates each amino acid position independently from other positions. Therefore, mutual interactions of individual residues in an interface, and interactions of the different interfaces in context of oligomerization are difficult to reconstruct, which may occasionally lead to such artifacts. Such effects might also be the reason for the observed dimer-hexamer equilibrium in AncGGGPS2_N2 and _N3. In general, these considerations support the robustness of the reconstruction of the hexameric state of the ancestral GGGPS proteins. It is obvious that it is much more probable to disturb the interface balance and therefore break the hexamer, than to convert a dimer to a hexamer by undesired reconstruction artifacts.

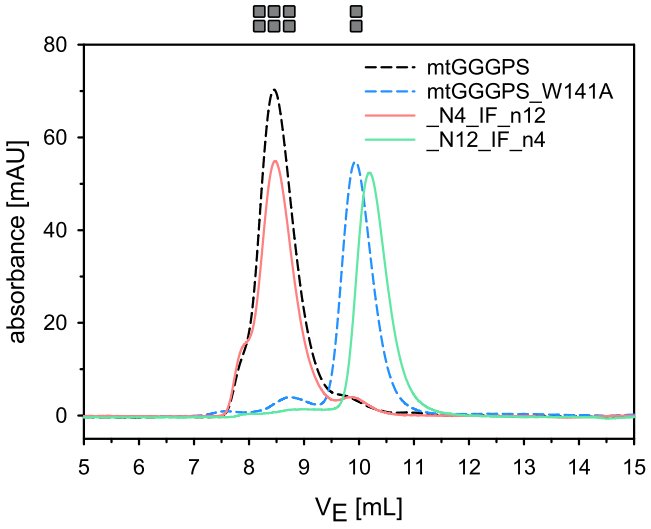

FIGURE 5 Analytical size exclusion chromatography of AncGGGPS2_N4_IF_n12 and AncGGGPS2_N12_IF_n4. The proteins (40 μM, subunit concentration) were applied to a S75 analytical column equilibrated with 50 mM potassium phosphate, pH 7.5, 300 mM KCl. Elution was performed at a flow rate of 0.5 mL/min and followed by measuring the absorbance at 280 nm, which was plotted against the elution volume. mtGGGPS_wt and mtGGGPS_W141A (dashed lines) served as references for the hexameric and dimeric oligomerization states. The oligomerization state is indicated by gray symbols

Additionally, we were able to purify and study the first extant protein from Thaumarchaeota, nvGGGPS, which shows a well-defined structure in CD spectroscopy, is enzymatically active, and is also a hexamer (Figure 1, Table S1, Figure S2, Figure S3D, Figure S7).

2.4 | Extant dimeric group II GGGPS are secondary dimers

To our knowledge, group II GGGPS enzymes from only two phyla have been reported to be dimeric based on experimental data (Table S1). These are the bacterial Flavobacteriales (fjGGGPS, zpGGGPS[4]; fsGGGPS, this work) and the archaeal Thermoplasmatales (taGGGPS[4,16,19]; tvGGGPS[17]). To study the emergence of these dimeric group II GGGPS, we expressed synthetic genes for predecessors of the Flavobacteriales and analyzed the purified proteins. AncGGGPS2_N72 is hexameric, as described above. In contrast, AncGGGPS2_N73, the LCA of the Flavobacteriales GGGPS, eluted as a dimer in analytical SEC (Figure 6). The protein showed a well-defined structure in CD spectroscopy (Figure S3C), but only poor activity (Figure S2). This may indicate an impairment of

this protein as discussed above for AncGGGPS2_N4, but AncGGGPS2_N73 has an Arg at the "aromatic anchor" position and a Lys at the position of the cationic partner, like most extant Flavobacteriales enzymes.[4] This supports that the experimentally obtained dimeric oligomerization state of AncGGGPS2_N73 is not an artifact.

Similarly, the sequences of Thermoplasmatales have an Arg at the "aromatic anchor" position. GGGPS sequences from Thermoplasmatales are missing in our ASR, because their somewhat diverse sequences were filtered out within the process of identification of wandering sequences to obtain robust ASR results.[12] An alternative ASR of GGGPS sequences including Thermoplasmatales has been published recently.[17] Based on computational data only, the authors postulate a convergent evolution for the hexamerization of group II GGGPS in the different phyla, which goes along with dimeric predecessors at the deep nodes in the phylogenetic tree. In contrast, our experimental results strongly support that hexamerization occurred very early in evolution, close to the group II LCA. Our data rather implicate that the occurrence of dimeric GGGPS among the group II proteins is a secondary and convergent effect, because these proteins are direct descendants of hexameric predecessors (Figure 2). A horizontal gene transfer of a dimeric GGGPS from group I is very unlikely, because the Flavobacteriales and Thermoplasmatales GGGPS sequences are much more diverse from group I sequences than from any group II sequence. The biological reason for this change to a dimeric oligomerization state remains unclear, however, and needs further elucidation.

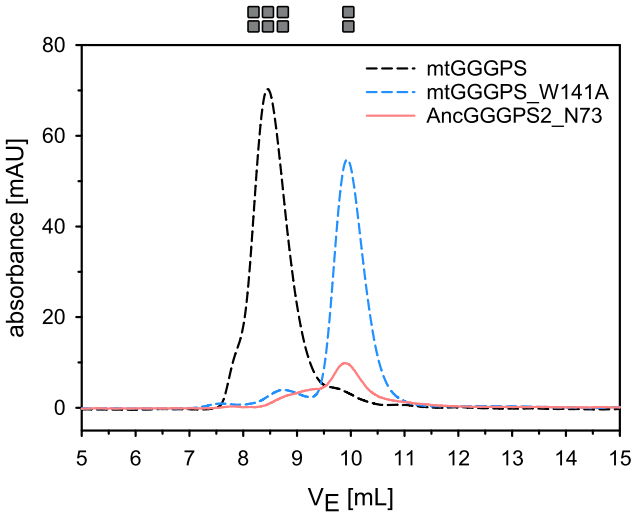

FIGURE 6 Analytical size exclusion chromatography of AncGGGPS2_N73. The proteins (22 μM, subunit concentration) were applied to a S75 analytical column equilibrated with 50 mM potassium phosphate, pH 7.5, 300 mM KCl. Elution was performed at a flow rate of 0.5 mL/min and followed by measuring the absorbance at 280 nm, which was plotted against the elution volume. AncGGGPS2_N73 tends to form aggregates and has a low extinction coefficient, resulting in a quite low peak. mtGGGPS_wt and mtGGGPS_W141A (dashed lines) served as references for the hexameric and dimeric oligomerization states. The oligomerization state is indicated by gray symbols

2.5 | Thermostability emerged very early during group II GGGPS evolution, flexibilization of the active site late

Transition temperatures from thermal denaturation experiments (DSC, nanoDSF) of many extant GGGPS enzymes have been obtained in this and in a previous study (Figure 2, Table 1). Notably, the temperatures of complete denaturation (T_2) are all in the hyperthermophilic (up to 127°C) or at least thermophilic range, although there are proteins from mesophilic species among them (fjGGGPS, fsGGGPS, rsGGGPS, slGGGPS, nvGGGPS). We have previously found that some GGGPS proteins show an additional thermal transition in DSC or nanoDSF experiments (T_1). This transition is hardly visible in CD spectroscopic analysis, indicating that it is not associated with a significant change of secondary structure, but goes along with enzymatic inactivation.[8] For some GGGPS proteins, these T_1 temperatures are in the upper mesophilic range only. Our interpretation of these results was that the protein scaffold of all GGGPS proteins is very thermostable and rigid, whereas the active site region is destabilized in some GGGPS variants to allow a higher flexibility and therefore higher activity at lower temperatures.[8] A localized flexibilization of the active site within a relatively rigid protein scaffold is not unique to GGGPS, but has been frequently observed in enzymes from psychrophilic species to allow high turnover rates at low temperatures.[20,21]

To investigate the emergence of thermostability among group II GGGPS, we have now determined transition temperatures of thermal denaturation of a large number of reconstructed ancestral and extant GGGPS (Figure 2, Table 1). Remarkably, the T_2 temperatures of complete denaturation for all ancestral GGGPS proteins are in the hyperthermophilic range. An exception is AncGGGPS2_N73, which tends to aggregation and where T_2 could not be determined confidently. The representatives at the deepest nodes, AncGGGPS2_N2 and AncGGPPS3_N3 have T_2 temperatures of 108°C and 113°C, respectively. This clearly supports a hyperthermophilic nature of the group II GGGPS LCA. The flexibilization of the active site, going along with the emergence of the T_1 transition, obviously has emerged in

TABLE 1 Temperatures of thermal transitions of different GGGPS proteins as followed by nanoDSF and DSC[a]

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

[a]The first block shows ancestral proteins, the second block proteins studied in this work, the raw data are shown in Figure S8. The third block shows proteins that have been investigated earlier.[8] The given temperatures are the mean of duplicates, standard deviations are given. Note that mtGGGPS shows two T_1 transitions (T_1, T_1*[8]). Only the transition leading to inactivation is given. -, T_1 does not exist; n.d., not determined.

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

some clades only. Besides the hyperthermophilic mtGGGPS and mrGGGPS, only the representatives from mesophilic Bacteroidetes (fsGGGPS, fjGGGPS) showed a T_1 transition. Our results indicate that the flexibilization emerged quite late in evolution, because none of the analyzed predecessors showed a T_1 transition.

To exclude that the reconstructed ancestral GGGPS proteins may not be active at elevated temperatures due to reconstruction artefacts, we analyzed the activity of the most stable ancestral GGGPS in our study, AncGGGPS2_N12 (T_2 = 118°C; no T_1). One of its direct descendants is the well-studied mtGGGPS (T_1 = 102.5°C, T_2 = 127°C), which showed an apparent half life time of 16.5 min at 100°C in heat inactivation studies.[8] We incubated both mtGGGPS and AncGGGPS2_N12 for 60 min at 105°C under pressure in the DSC instrument and tested their activity before and after this incubation step (Figure 7). mtGGGPS completely lost its activity under these conditions, while the apparent activity of AncGGGPS2_N12 was unchanged. This means that AncGGGPS2_N12 is even more thermostable than mtGGGPS and supports that all ancestral group II GGGPS enzymes may have been active at extreme temperatures.

3 | DISCUSSION

Recent studies have shown that many extant representatives of group II GGGPS proteins have a hexameric oligomerization state. A key element to stabilize the hexamer is a cation-π interaction in the so-called ring interface, usually formed by a Trp, Tyr or Phe together with a Lys or Arg at the opposing protein surface.[4,8] A large number of Crenarchaeota, however, have a conserved His at the position of the aromatic moiety, and their GGGPS proteins were not accessible to experiments up to now. Using ancestral sequence reconstruction, we have now characterized a large number of GGGPS predecessors with respect to their oligomerization state, thermostability and activity. Our results demonstrate that the crenarchaeal proteins form hexamers, too. This means that almost all group II GGGPS, and consequently the GGGPS of almost all archaea, have an hexameric oligomerization state.

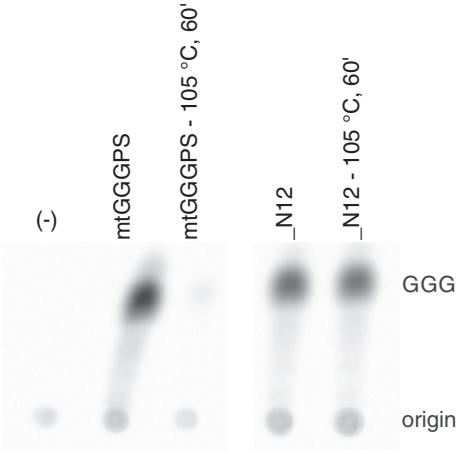

FIGURE 7 Heat inactivation of mtGGGPS and AncGGGPS2_N12. Proteins were incubated for 60 min at 105°C. Heat-treated and untreated samples (1 μM subunit concentration) were incubated with the substrates ^{14}C-G1P and GGPP for 1.5 h at 40°C. The generated products were extracted, dephosphorylated to allow for subsequent separation by thin layer chromatography, and visualized by autoradiography. As negative control (−), no enzyme was added. The origin of the chromatography and the reaction product geranylgeranylglycerol (GGG) are marked

Furthermore, almost all GGGPS predecessors form hexamers, among them those at the deepest nodes of the ASR tree. This makes it very likely that hexameric GGGPS developed close to the emergence of the group II LCA (LCA-II; Figure 1), which itself could not be reconstructed.

In contrast to the group II GGGPS proteins, all studied representatives of group I proteins are reported to be dimers. An exception is hsGGGPS, which eluted as a monomer in SEC, but unpublished results support that this might be an artefact due to the high salt requirements of this halobacterial protein that could not be satisfied during SEC analysis.[4] Because of the divergence of group I and group II GGGPS sequences, we were not able to calculate a combined ASR tree of groups I and II and their common ancestor sequence.[12] However, the exclusive presence of dimers among group I makes it very likely that the LCA of group I (LCA-I; Figure 1) also was a dimer. Under these assumptions (dimeric LCA-I, hexameric LCA-II), the LCA of all GGGPS proteins might have been either dimeric or hexameric. The analysis of large datasets of oligomeric structures has revealed as a common rule that the larger an oligomerization interface is, the earlier it has emerged during the evolution of the complex.[22,23] Furthermore, symmetric interfaces are more likely to emerge during evolution, because they are easier to generate.[24] In GGGPS, the symmetric dimer module interface is by far the largest (1,163 Å^2 in mtGGGPS), while the key interface for hexamerization, the ring interface, is much smaller (256 Å^2 in mtGGGPS) and asymmetric. These facts and the principle of parsimony speak for a dimeric nature of the LCA, and the need for thermostabilization may have driven the evolution towards hexamerization in group II GGGPS by a cyclic trimerization of the dimeric building block. We assume that the few dimeric group II GGGPS like in Flavobacteriales and Thermoplasmatales have emerged convergently as secondary events, because they descend from hexameric predecessors.

Ancestral sequence reconstruction also allowed us to shed light on the evolution of thermostability among group II GGGPS. All reconstructed GGGPS predecessors are highly thermostable. As extreme example, AncGGGPS2_N12 only denatures at temperatures above 118°C and is still fully active after 60 min incubation at 105°C. We have recently presented data which suggest that hexamerization of GGGPS supports thermostability in a rather indirect manner.[8] Even as a monomeric module, the GGGPS protein scaffold is highly thermostable in terms of the temperature of complete denaturation. However, such rigid enzymes tend to be less catalytically active.[25] Similarly to what has been observed for cold-adapted enzymes to allow high turnover rates at low temperatures,[20,21] we have previously found a localized flexibilization of the active site of some GGGPS representatives, which goes along with two transitions in thermal denaturation experiments. The first transition is associated with loss of activity and collapse of the active site, while the second transition marks the complete denaturation of the protein. We could demonstrate that hexamerization elevates the temperature of the first transition, and we assumed that oligomerization serves for a fine tuning of the balance between flexibility and stability in hexameric GGGPS.[8] Ancestral sequence reconstruction now revealed that the flexibilization of the active site, apparent as the occurrence of two transitions in thermal denaturation experiments, obviously emerged quite late in evolution, because only extant GGGPS enzymes showed this phenomenon. This means that the primary stimulus for the emergence of hexamerization was not the need to stabilize the active site, but must have been different, such as a general need for thermostability. The flexibilization of the active site has obviously emerged independently of hexamerization in convergent processes, because it is also present in dimeric fjGGGPS. Its stabilization by hexamerization must rather be regarded as a secondary benefit that made use of the already existing higher oligomeric structure. Consequently, ancestral GGGPS enzymes must have used another still unknown, probably less effective way to balance catalytic activity and stability, like some extant dimeric GGGPS from hyperthermophilic species still do.

4 | MATERIALS AND METHODS

4.1 | Ancestral sequence reconstruction of GGGPS group II sequences

Ancestral sequence reconstruction (ASR) has been described in detail elsewhere.[12] Briefly, the 87 GGGPS group II sequences that have been manually selected by using several filter algorithms in order to obtain high posterior probabilities on the edges of the tree, branch lengths <1.0 mutations per site and a species diversity as high as possible. Filter algorithms as they are the basis of FitS4ASR[12] were used in a manual manner which identify wandering sequences by both RogueNaRok[26] and the described heatmap approach. The resulting sequences are from Crenarchaeota (23 sequences), Euryarchaeota (40 s.), Thaumarchaeota (8 s.), and Bacteroidetes (16 s.). The tree needed for reconstruction was computed by using PhyloBayes,[27] utilizing a site-heterogeneous CAT model. Four independent MCMC samplings of length 50,000 were launched to ensure convergence. After discarding the first 6,000 trees as burn-in, the remaining trees of two MCMC chains - chosen by the quality criteria maximum discrepancy (<0.1; 0.01699) and minimum effective size (> 100; 2,769) (score for a good run; received score) - were concatenated to deduce a consensus tree. By means of FastML,[28] the ancestral sequences were reconstructed with the substitution model JTT and a gamma distribution. The parameter "probability cutoff', which helps to prefer ancestral indel over character, was set to 0.8 in order to adjust the length of the reconstructed sequences to the length of the recent sequences, compensating a bias towards longer than true ancestors.[29] The most probable sequence was determined for each internal node of the tree. A FASTA file with all extant and ancestral sequences as well as a file containing the posterior probabilities for all sites in the ancestral sequences are provided as supplementary material.

4.2 | Cloning and site-directed mutagenesis

As summarized in Table S2, twelve ancestral GGGPS variants and four extant GGGPS proteins were produced and purified. Gene sequences for the reconstructed GGGPS variants (AncGGGPS2_N2, AncGGGPS2_N3, AncGGGPS2_N4, AncGGGPS2_N5, AncGGGPS2_N12, AncGGGPS2_N18, AncGGGPS2_N52, AncGGGPS2_N72, AncGGGPS2_N73, AncGGGPS2_N4_IF_n12, AncGGGPS2_N12_IF_n4) and extant GGGPS proteins of *Roseivirga seohaensis* (rsGGGPS), *Flavobacterium saliperosum* (fsGGGPS), *Methanobrevibacter ruminatium* (mrGGGPS) and *Nitrososphaera viennensis* (nvGGGPS) were optimized in their codon usage for expression in *E. coli* and ordered as GeneArt™ Strings™ DNA fragments from Thermo Fisher Scientific. *Bsa*I restriction sites were incorporated at the 5′ and 3′-ends of the DNA fragments to allow cloning into a modified pET21a expression vector,[30] providing a C-terminal hexahistidine $(His)_6$ tag. AncGGGPS2_N73 was cloned into a modified pMAL-c5T expression vector,[30] providing a C-terminal maltose binding protein (MBP) fusion with a thrombin cleavage site. Other variants of AncGGGPS2_N12 and mtGGGPS were generated by QuickChange mutagenesis[31] with oligonucleotides as given in Table S3. All constructs were verified by sequencing. The wild-type *Methanothermobacter thermautotrophicus* GGGPS (mtGGGPS) and the mtGGGPS_W141A variant have been cloned previously into pET21a.[4] The sequence numbering of mtGGGPS used in this study refers to EMBL ENA entry AAB85058 and pdb-id 4 mm1, which have an N-terminal three amino acid extension compared to UniProt entry O26652.

4.3 | Production and purification of recombinant proteins

Heterologous gene expression was performed in the *E. coli* strain BL21-Gold(DE3) (Agilent Technologies). The transformed cells were grown at 37°C in LB containing ampicillin (150 $\mu g/ml^{-1}$). When OD_{600} reached 0.6–0.8, expression was induced by adding 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG). Growth was continued overnight at 20°C, the cells were harvested by centrifugation, resuspended in 50 mM potassium phosphate, pH 7.5, 300 mM KCl, 10 mM imidazole and disrupted by sonication. The His-tagged proteins were purified from the clarified cell extract by metal chelate affinity chromatography. An ÄKTApurifier system with a HisTrap FF crude column (5 mL, GE Healthcare) was used, and a linear gradient of imidazole (10–500 mM) in 50 mM potassium phosphate, pH 7.5, 300 mM KCl was applied to elute the protein. Interfering imidazole and salt were removed from the purified proteins by dialysis against 50 mM potassium phosphate, pH 7.5 at 4°C. The dialysis buffer for rsGGGPS additionally contained 300 mM KCl. For the MBP fused construct, the dialysis buffer was supplemented with thrombin (2 U/ml^{-1} final concentration) for cleavage. MBP and thrombin were removed by subsequent preparative size exclusion chromatography (SEC) on a Highload™ 26/600 Superdex™ S75 pg column (GE Healthcare). The column was equilibrated with 50 mM potassium phosphate, pH 7.5, 300 mM KCl and was run at a flow rate of 1.5 ml/min. Protein concentrations

were determined by absorbance spectroscopy in the UV range; if the extinction coefficient of the variant was too low, the Bradford assay was used. The molar extinction coefficients ε_{280} and the molecular weight were calculated from the amino acid sequence by means of ProtParam.[32] Purified protein was dropped into liquid nitrogen and stored at −80°C. If high protein concentrations were needed, the samples were concentrated by ultrafiltration (Amicon Ultra-15, mwco 10 kDa, Merck Millipore). The purification yields of all proteins are listed in Table S2, purity is shown in Figure S9.

4.4 | Characterization of oligomerization state of recombinant proteins

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

4.5 | CD spectroscopy

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

4.6 | Differential scanning calorimetry

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

4.7 | Differential scanning fluorimetry (nanoDSF)

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

4.8 | Classification of thermal transitions as T_1 (partial denaturation) or T_2 (complete denaturation)

The existence of two individual thermal transitions, where the first (at temperature T_1) is associated with a loss of enzymatic activity, but with a partial denaturation of the protein only, while the second (at a higher temperature T_2) goes along with a complete denaturation, has been unambiguously demonstrated by a combination of DSC, nanoDSF and CD-spectroscopic experiments as well as enzymatic activity tests for the two proteins mtGGGPS and fjGGGPS.[8] In DSC experiments, the proteins are heated up to 130°C. If only a single transition is visible, it is therefore very likely that this transition goes along with a complete denaturation of the protein and can be addressed as a T_2 transition. This has been demonstrated experimentally by experiments with taGGGPS and slGGGPS.[8] In nanoDSF experiments, however, a complete denaturation cannot be achieved for highly thermostable proteins, because this would exceed the upper temperature limit of the instrument (95°C). This makes it difficult to conclude whether a transition observed in nanoDSF is due to a partial (T_1) or a complete (T_2) denaturation, if only a single transition is visible. Although a decrease of the F_{350}/F_{330} ratio points to a T_1 transition in most cases and an increase of the ratio to a T_2 transition ([8] and this study), we performed DSC experiments wherever possible to clarify the situation. For the three proteins AncGGGPS2_N52, _N72 and _N73, DSC experiments were not possible due to technical reasons. In case of AncGGGPS2_N52 and _N72, no transition could be observed in nanoDSF. We assume that the T_2 transition for these proteins is above

95°C. In case of AncGGGPS2_N73, a transition was observed in nanoDSF at 58.4°C with a decrease in F_{350}/F_{330} ratio, which would speak for a T_1 transition, but the protein strongly tended to aggregation and was difficult to handle. Therefore, it is impossible to decide whether this transition is due to a partial or complete denaturation.

4.9 | Radiometric GGGPS activity assays

^{14}C-G1P was synthesized as described by Guldan et al.[3] To test the activity of purified GGGPS enzymes, 1 μM of protein (subunit concentration) was incubated with 20 μM GGPP and 20 μM ^{14}C-G1P (302 nCi) in 10 mM $MgCl_2$, 0.2% Tween80, 50 mM Tris–HCl, pH 8.0, in a total volume of 100 μL for 1.5 h at 40°C. The products were dephosphorylated by adding 1 U calf intestinal alkaline phosphatase (New England Biolabs) for 30 min at 40°C and extracted according to the method of Bligh and Dyer[33] as modified by Kates.[34] The solvent was evaporated to dryness in a rotary evaporator and the products were analyzed by TLC on Silica 60 plates, developed in ethyl acetate/hexane 1:1 (v/v), and visualized with a phosphorimager system (PerkinElmer Life Sciences).

4.10 | Heat inactivation

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

4.11 | Isothermal titration calorimetry

To follow G1P binding to GGGPS variants, a MicroCal PEAQ-ITC microcalorimeter (Malvern Instruments) with a cell volume of 280 μL was used. Degassed protein (100 μM subunit concentration) and G1P (1 mM) solutions were prepared from the identical buffer batch (50 mM Tris, pH 8.0, 10 mM $MgCl_2$). The G1P solution was titrated to the protein in 2 μL aliquots for a total of 18 injections at 2.5 min intervals at 25°C. During the course of the titration, the reaction mixture was continuously stirred. As controls, titrations of buffer with buffer, protein solution with buffer and buffer with ligand solution were performed. Each titration experiment was baseline corrected, and the change of heat per injection was calculated by integrating the area under each peak using the Origin software provided by MicroCal. The experimentally observed signals for the ligand binding experiment were corrected for the signals of the control experiments, a ΔH versus molar ratio plot was generated and a fit curve was plotted with the "single set of identical sites" option provided by the software.

4.12 | Supplementary material

As supplementary material, we provide a file (Supinfo.pdf) containing three tables that summarize the features of the studied proteins and the used oligonucleotides, as well as nine figures showing a detailed phylogenetic tree as used for ASR and detailed characterizations of the studied proteins like activity assays, CD spectra, additional SEC analyses, ITC, nanoDSF and DSC data as well as SDS-PAGE analysis displaying their purity. A FASTA file with all extant and ancestral sequences (AppendixS2.doc) as well as a file containing the posterior probabilities for all sites in the ancestral sequences (AppendixS1.xlsx) are provided separately.

ACKNOWLEDGEMENTS

We thank 2bind GmbH for access to the Prometheus NT.48 instrument (NanoTemper Technologies) and support in nanoDSF experiments. We are grateful to Christiane Endres, Sonja Fuchs, Sabine Laberer and Jeannette Ueckert for technical assistance. We thank Rainer Merkl and Reinhard Sterner for helpful discussions and critical reading of the manuscript. Open Access funding enabled and organized by ProjektDEAL.

AUTHOR CONTRIBUTIONS

Cosimo Kropp: Conceptualization; investigation; methodology; writing-original draft. **Kristina Straub:** Methodology; software. **Mona Linde:** Investigation; methodology. **Patrick Babinger:** Conceptualization; supervision; writing-review and editing.

ORCID

Patrick Babinger https://orcid.org/0000-0002-7490-6732

REFERENCES

1. Jain S, Caforio A, Driessen AJ. Biosynthesis of archaeal membrane ether lipids. Front Microbiol. 2014;5:641.

2. Villanueva L, Schouten S, Damste JS. Phylogenomic analysis of lipid biosynthetic genes of Archaea shed light on the 'lipid divide'. Environ Microbiol. 2017;19:54–69.
3. Guldan H, Matysik FM, Bocola M, Sterner R, Babinger P. Functional assignment of an enzyme that catalyzes the synthesis of an Archaea-type ether lipid in bacteria. Angew Chem Int Ed. 2011;50:8188–8191.
4. Peterhoff D, Beer B, Rajendran C, et al. 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. 2014;92:885–899.
5. Guldan H, Sterner R, Babinger P. Identification and characterization of a bacterial glycerol-1-phosphate dehydrogenase: Ni^{2+}-dependent AraM from *Bacillus subtilis*. Biochemistry. 2008;47:7376–7384.
6. Linde M, Peterhoff D, Sterner R, Babinger P. Identification and characterization of heptaprenylglyceryl phosphate processing enzymes in *Bacillus subtilis*. J Biol Chem. 2016;291:14861–14870.
7. Villanueva L, von Meijenfeldt FAB, Westbye AB, et al. Bridging the membrane lipid divide: Bacteria of the FCB group superphylum have the potential to synthesize archaeal ether lipids. ISME J. 2020. https://doi.org/10.1038/s41396-41020-00772-41392.
8. Linde M, Heyn K, Merkl R, Sterner R, Babinger P. Hexamerization of geranylgeranylglyceryl phosphate synthase ensures structural integrity and catalytic activity at high temperatures. Biochemistry. 2018;57:2335–2348.
9. Hochberg GKA, Thornton JW. Reconstructing ancient proteins to understand the causes of structure and function. Annu Rev Biophys. 2017;46:247–269.
10. Pillai AS, Chandler SA, Liu Y, et al. Origin of complexity in haemoglobin evolution. Nature. 2020;581:480–485.
11. Merkl R, Sterner R. Ancestral protein reconstruction: Techniques and applications. Biol Chem. 2016;397:1–21.
12. Straub K, Linde M, Kropp C, Blanquart S, Babinger P, Merkl R. Sequence selection by FitSS4ASR alleviates ancestral sequence reconstruction as exemplified for geranylgeranylglyceryl phosphate synthase. Biol Chem. 2019;400:367–381.
13. Liao SM, Du QS, Meng JZ, Pang ZW, Huang RB. The multiple roles of histidine in protein interactions. Chem Cent J. 2013;7:44.
14. Kumar K, Woo SM, Siu T, Cortopassi WA, Duarte F, Paton RS. Cation-pi interactions in protein-ligand binding: Theory and data-mining reveal different roles for lysine and arginine. Chem Sci. 2018;9:2655–2665.
15. Cauet E, Rooman M, Wintjens R, Lievin J, Biot C. Histidine-aromatic interactions in proteins and protein-ligand complexes: Quantum chemical study of X-ray and model structures. J Chem Theory Comput. 2005;1:472–483.
16. Nemoto N, Oshima T, Yamagishi A. Purification and characterization of geranylgeranylglyceryl phosphate synthase from a thermoacidophilic archaeon. Thermoplasma acidophilum J Biochem. 2003;133:651–657.
17. Blank PN, Barnett AA, Ronnebaum TA, et al. Structural studies of geranylgeranylglyceryl phosphate synthase, a prenyltransferase found in thermophilic Euryarchaeota. Acta Cryst D. 2020;76:542–557.
18. Eick GN, Bridgham JT, Anderson DP, Harms MJ, Thornton JW. Robustness of reconstructed ancestral protein functions to statistical uncertainty. Mol Biol Evol. 2016;34:247–261.
19. Nemoto N, Miyazono KI, Tanokura M, Yamagishi A. Crystal structure of (S)-3-O-geranylgeranylglyceryl phosphate synthase from *Thermoplasma acidophilum* in complex with the substrate sn-glycerol 1-phosphate. Acta Cryst F. 2019;75:470–479.
20. Socan J, Purg M, Aqvist J. Computer simulations explain the anomalous temperature optimum in a cold-adapted enzyme. Nat Commun. 2020;11:2644.
21. D'Amico S, Marx JC, Gerday C, Feller G. Activity-stability relationships in extremophilic enzymes. J Biol Chem. 2003;278:7891–7896.
22. Levy ED, Boeri Erba E, Robinson CV, Teichmann SA. Assembly reflects evolution of protein complexes. Nature. 2008;453:1262–1265.
23. Venkatakrishnan AJ, Levy ED, Teichmann SA. Homomeric protein complexes: Evolution and assembly. Biochem Soc Trans. 2010;38:879–882.
24. Villar G, Wilber AW, Williamson AJ, et al. Self-assembly and evolution of homomeric protein complexes. Phys Rev Lett. 2009;102:118106.
25. Karshikoff A, Nilsson L, Ladenstein R. Rigidity versus flexibility: The dilemma of understanding protein thermal stability. FEBS J. 2015;282:3899–3917.
26. Aberer AJ, Krompass D, Stamatakis A. Pruning rogue taxa improves phylogenetic accuracy: An efficient algorithm and webservice. Syst Biol. 2013;62:162–166.
27. Lartillot N, Lepage T, Blanquart S. PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics. 2009;25:2286–2288.
28. Ashkenazy H, Penn O, Doron-Faigenboim A, et al. FastML: A web server for probabilistic reconstruction of ancestral sequences. Nucleic Acids Res. 2012;40:W580–W584.
29. Vialle RA, Tamuri AU, Goldman N. Alignment modulates ancestral sequence reconstruction accuracy. Mol Biol Evol. 2018;35:1783–1797.
30. Rohweder B, Semmelmann F, Endres C, Sterner R. Standardized cloning vectors for protein production and generation of large gene libraries in *Escherichia coli*. Biotechniques. 2018;64:24–26.
31. Zheng L, Baumann U, Reymond JL. An efficient one-step site-directed and site-saturation mutagenesis protocol. Nucleic Acids Res. 2004;32:e115.
32. Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the ExPASy server. In: Walker JM, editor. The proteomics protocols handbook. Totowa, N.J: Humana Press, 2005; p. 988.
33. Bligh EG, Dyer WJ. A rapid method of total lipid extraction and purification. Can J Biochem Physiol. 1959;37:911–917.
34. Kates M. Techniques of lipidology: Isolation, analysis, and identification of lipids. In: Burdon RH, van Knippenberg PH, editors. Laboratory techniques in biochemistry and molecular biology vol 3, part 2. Amsterdam: Elsevier, 1986; p. 1–464.
35. Payandeh J, Fujihashi M, Gillon W, Pai EF. The crystal structure of (S)-3-*O*-geranylgeranylglyceryl phosphate synthase reveals an ancient fold for an ancient enzyme. J Biol Chem. 2006;281:6070–6078.
36. Badger J, Sauder JM, Adams JM, et al. Structural analysis of a set of proteins resulting from a bacterial genomics project. Proteins. 2005;60:787–796.

37. Ren F, Feng X, Ko TP, et al. Insights into TIM-barrel prenyl transferase mechanisms: Crystal structures of PcrB from *Bacillus subtilis* and *Staphylococcus aureus*. Chembiochem. 2013;14:195–199.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Kropp C, Straub K, Linde M, Babinger P. Hexamerization and thermostability emerged very early during geranylgeranylglyceryl phosphate synthase evolution. *Protein Science*. 2021;30:583–596. https://doi.org/10.1002/pro.4016