Published in final edited form as:

J Mol Biol. 2010 December 3; 404(3): 403–417. doi:10.1016/j.jmb.2010.09.032.

Insights into substrate specificity of geranylgeranyl reductases revealed by the structure of digeranylgeranylglycerophospholipid reductase from *Thermoplasma acidophilum*, an essential enzyme in the biosynthesis of archaeal membrane lipids

Qingping Xu^{1,2}, Tadashi Eguchi³, Irimpan I. Mathews¹, Christopher L. Rife^{1,2}, Hsiu-Ju Chiu^{1,2}, Carol L. Farr^{2,4}, Julie Feuerhelm^{2,5}, Lukasz Jaroszewski^{2,6,7}, Heath E. Klock^{2,5}, Mark W. Knuth^{2,5}, Mitchell D. Miller^{1,2}, Dana Weekes^{2,7}, Marc-André Elslinger^{2,4}, Ashley M. Deacon^{1,2}, Adam Godzik^{2,6,7}, Scott A. Lesley^{2,4,5}, and Ian A. Wilson^{2,4,*}

¹Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

²Joint Center for Structural Genomics, <http://www.jcsg.org>

³Department of Chemistry & Materials Science, Tokyo Institute of Technology, O-okayama, Meguro-ku, Tokyo 152-8551, Japan

⁴Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

⁵Protein Sciences Department, Genomics Institute of the Novartis Research Foundation, San Diego, CA 92121, USA

⁶Center for Research in Biological Systems, University of California, San Diego, La Jolla, CA 92093, USA

⁷Program on Bioinformatics and Systems Biology, Sanford-Burnham Medical Research Institute, La Jolla, CA 92037, USA

Abstract

Archaeal membrane lipids consist of branched, saturated hydrocarbons distinct from those found in bacteria and eukaryotes. Digeranylgeranylglycerophospholipid reductase (DGGR) catalyzes the hydrogenation process that converts unsaturated 2,3-di-O-geranylgeranylglyceryl phosphate to saturated 2,3-di-O-phytylglyceryl phosphate as a critical step in the biosynthesis of archaeal membrane lipids. The saturation of hydrocarbon chains confers the ability to resist hydrolysis and oxidation and helps archaea withstand extreme conditions. DGGR is a member of the geranylgeranyl reductase (GGR) family that is also widely distributed in bacteria and plants, where the family members are involved in the biosynthesis of photosynthetic pigments. We have determined the crystal structure of DGGR from the thermophilic heterotrophic archaea *Thermoplasma acidophilum* at 1.6 Å resolution, in complex with FAD and a bacterial lipid. The DGGR structure can be assigned to the well-studied, *para*-hydroxybenzoate hydroxylase (PHBH) SCOP superfamily of flavoproteins that include many aromatic hydroxylases and other enzymes with diverse functions. In the DGGR complex, FAD adopts the IN conformation (closed) previously observed in other PHBH flavoproteins. DGGR contains a large substrate-binding site that extends across the entire ligand-binding domain. Electron density corresponding to a bacterial

*To whom correspondence should be addressed. The Scripps Research Institute, 10550 North Torrey Pines Road, BCC206, La Jolla, CA 92037. wilson@scripps.edu.

lipid was found within this cavity. The cavity consists of a large opening that tapers down to two narrow curved tunnels that closely mimic the shape of the preferred substrate. We identified a sequence motif, PxxYxWxFP, that defines a specificity pocket in the structure and precisely aligns the double bond of the geranyl group with respect to the FAD cofactor, thus providing a structural basis for the substrate specificity of GGRs. DGGR is likely to share a common mechanism with other PHBH enzymes in which FAD switches between two conformations that correspond to the reductive and oxidative half cycles. The structure provides evidence that substrate binding likely involves conformational changes, which are coupled to the two conformational states of the FAD.

Introduction

The structure of their membrane lipids is one of the distinguishing features that separate archaea from bacteria and eukaryotes^{1,2}. Bacterial and eukaryotic membrane lipid structures typically consist of two linear, hydrocarbon chains attached via an ester linkage to an *sn*-glycerol-3-phosphate backbone that is linked to polar head groups to form glycolipids. The lipid structures in archaea are distinct in that two linear, branched phytanyl chains are linked via an ether linkage to a *sn*-glycerol-1-phosphate (G-1-P) backbone (Fig. 1a).

The basic core structure of the archaeal membrane is 2,3-di-O-phytanyl-*sn*-glycerol, whose biosynthesis pathway involves phosphorylation and then alkylation of a glycerol backbone that contains isoprenoid lipid chains¹. An archaea-specific *sn*-glycerol-1-phosphate dehydrogenase (G-1-PD) catalyzes formation of G-1-P from dihydroxyacetone phosphate. Isoprenoid chains are synthesized via the mevalonate pathway, as in bacteria and eukaryotes^{3,4}. Two prenyltransferases, (S)-3-O-geranylgeranylgeranyl phosphate synthase and (S)-2,3-digeranylgeranylgeranyl phosphate synthase, work in tandem to add the isoprenoid lipid chains to C-3 and then C-2 of the glycerophosphate backbone. These steps result in the core diether intermediate (S)-2,3-diether geranylgeranyl phosphate. Further modifications, such as attachment of polar head groups, hydrogenation and cyclization, result in a variety of mature lipids.

Saturation of the geranylgeranyl chains of biosynthetic intermediates is an important reaction that contributes to the chemical stability of archaeal lipids, which helps organisms survive in extreme environments, such as extreme temperatures, pH, or osmotic conditions⁵. Recently, the enzyme that catalyzes the hydrogenation of geranylgeranyl chains in archaeal lipid biosynthesis was identified in *T. acidophilum*^{6,7}, *Archaeoglobus fulgidus*⁸ and *Sulfolobus acidocaldarius*⁹. Digeranylgeranylgeranylphospholipid reductase (DGGR) converts 2,3-di-O-geranylgeranylgeranyl phosphate (DGGGP) to 2,3-di-O-phytanylgeranyl phosphate (archaetidic acid) (Fig. 1b). Additionally, it also catalyzes less favorably the hydrogenation of geranylgeranyl chain(s) in other compounds, such as 3-O-geranylgeranylgeranyl phosphate and geranylgeranyl diphosphate. Interestingly, DGGR is a flavoenzyme that is capable of reducing all, non-saturated, double bonds of DGGGP in a stereospecific manner⁶. DGGR is remotely homologous to geranylgeranyl reductases (GGRs) in cyanobacteria and plants, which are involved in the hydrogenation of geranylgeranyl chains in biosynthesis of photosynthetic pigments^{10,11}. The GGR family consists of more than 600 members (INTERPRO¹² accession number IPR011777), and are part of the aromatic-ring hydroxylase-like protein family (IPR003042).

To gain further insights into DGGR function and to provide a structural basis for the reduction of geranylgeranyl groups in general, we determined the crystal structure of *T. acidophilum* DGGR at 1.6 Å resolution, in complex with flavin adenine dinucleotide (FAD) and a bacterial lipid (PGX). The crystal structure confirms that DGGR adopts a fold that is common in flavin monooxygenases, such as the *p*-hydroxybenzoate hydroxylase (PHBH),

but with a modified substrate-binding site that can accommodate larger and more complex substrates. The endogenous bacterial lipid enabled us to identify a unique substrate-binding cavity that extends across the entire substrate-binding domain. We present structural analyses, which uncovered the basis for substrate specificity of DGGR and shed light on its catalytic mechanism.

Results and Discussion

Structure determination and quality of the model

The full-length DGGR from *T. acidophilum* contains 396 residues with a molecular weight of 43.4 kD and calculated pI of 7.0. The crystal structure was determined using the high-throughput structural genomics pipeline implemented at the Joint Center for Structural Genomics (JCSG, <http://www.jcsg.org>)¹³. The selenomethionine derivative of DGGR was expressed in *E. coli* with an N-terminal TEV cleavable His-tag and was purified by metal affinity chromatography. Three-wavelength MAD data were collected to 1.6 Å resolution at SSRL BL11-1. The data were indexed and processed in orthorhombic space group P2₁2₁2₁. Data processing statistics are summarized in Table 1. The structure was determined with one molecule per asymmetric unit using the MAD phasing method. The electron density for the main chain was well defined throughout the entire molecule, except for two regions (residues 85-91, 385-396), which had weak, but interpretable density. The final model was refined to an R_{cryst} of 15.3% and an R_{free} of 17.1%. The mean residual error of the coordinates was estimated to be 0.07 Å by a diffraction-component precision index method (DPI)¹⁴. The model of DGGR displays good geometry with an all-atom clash score of 5.9, and the Ramachandran plot produced by MolProbity¹⁵ shows that all residues are in allowed regions, with 98.2 % in favored regions. Most of the side chains are well resolved (5 rotamer outliers) and 39 residues were modeled with multiple conformations. Additionally, 13 residues contain incomplete side chains due to lack of interpretable density. All of the disordered side chains are located on the protein surface. The final model of DGGR contains residues 0 to 396 (residue Gly0 is the residue that remains after cleavage from the N-terminal purification tag), one FAD, one bacterial lipid phosphatidylglycerol (PGX), 11 ethylene glycols, 2 glycerols and 332 waters. A summary of the refinement and model statistics is shown in Table 2.

DGGR is a member of PHBH family of flavoenzymes

Analytical gel-filtration data suggested that DGGR exists as a monomer in solution (data not shown). Interfaces in the crystal lattice between symmetry-related molecules are in the range of ~100-500 Å² (< 3% of total solvent accessible surface), which are typical for crystal contacts rather than specific protein-protein interfaces further suggesting that DGGR functions as a monomer. The DGGR structure (Fig. 2) is formed by two functional domains: an FAD-binding domain (residues 1-70, 93-173, and 268-335) and a ligand-binding domain (residues 71-92, 174-267 and 336-396). The FAD-binding domain adopts a Rossmann-type fold. The ligand-binding domain is inserted in the FAD-binding domain and folds into an extensive, seven-stranded, mixed β-sheet. The C-terminus of DGGR is composed of a helical subdomain that includes the C-terminal half of kinked helix αK and helices αL-αN and forms the lid of the substrate binding site. The active site is located at the interface between these two domains. A search of the structural database with the DGGR coordinates using the DALI server¹⁶ returned a significant number of similar structures, most of which were hydroxylases or halogenases that belong to the PHBH flavoenzyme family and contain the same two-domain architecture as DGGR. A comparison of the closest structural homologs is presented in Table S1. The best match is to CndH¹⁷, which can be superimposed to DGGR with an rmsd of 3.0 Å and 19% sequence identity for 349 C_α atoms. Extensive structural and functional studies of PHBH family members have shown that the

PHBH flavoproteins are highly versatile and adopt a diverse set of functions. Thus, DGGR is another example of evolutionary adaption to a new function by a common fold.

The overall structure of DGGR is very similar to the prototypical PHBH with highly conserved secondary structure elements (Fig. S1). There is a disruption of the standard hydrogen bonding in helix α K of DGGR at the designated interface between the functional domains (residues 333-334), which results in a kink ($\Omega=50^\circ$). As a result, the second half of helix α K, as well as helices α L and α M display larger structural differences. Overall, the different arrangement of these helices results in enlargement of the internal cavity to accommodate significantly larger substrates. Additionally, the C-terminal residues in DGGR point away from the cavity opening where the head group of the bound lipid is located, while the equivalent location is blocked by the equivalent residues in PHBH and, thus, inaccessible. Two other regions that differ in DGGR compared to PHBH include the β 13- β 14 loop (residues 200-209) and the α I- α J loop (residues 286-295), which are also in close proximity to the bound lipid.

FAD binding site

Previous studies have indicated that DGGR is an FAD-dependent flavoprotein, which likely uses NADH/NADPH as a reducing agent *in vivo*⁷. The crystal structure of DGGR contains a tightly bound FAD molecule. The cofactor FAD is bound to the enzyme in an extended conformation and interacts extensively with the protein (Fig. S2a), as in other PHBH enzymes. The riboflavin of FAD primarily interacts with two protein regions, one near the α I- α J loop (residues 283-296), and an extended region near β 3 (residues 42-49). Arg100 and Asp283 make hydrogen bond contacts with the ribityl chain. The isoalloxazine ring is involved in extensive hydrogen bonds with the protein backbone via its N1, O2, O4, N3, and N5 atoms.

The FAD bound to DGGR appears to be in the oxidized state initially, as judged from the yellow-colored crystal. However, the refined FAD exhibits planar distortion with a butterfly bend along the N5-N10 axis of the isoalloxazine ring, more consistent with a partially reduced state¹⁸ (Fig. S2b). The angle (ϕ) between the dimethylbenzene and pyrimidine rings is 167.6° . The O4 atom deviates most from the dimethylbenzene plane by around 0.73 Å. The distortion of FAD observed in DGGR could be induced by photo-reduction during data collection¹⁹.

PHBH flavoenzymes typically share a common feature in which FAD adopts two different conformations, associated with the two half-reactions: reduction of the FAD cofactor by NADH (or NADPH) and oxidation of the substrate. An IN (or closed) conformation buries the isoalloxazine ring within the active site and the substrate binding pocket is inaccessible to bulk solvent. An OUT (or open) conformation exposes the isoalloxazine ring, through a torsional rotation of bonds in the ribityl chain, allowing it to be reduced by NADH (or NADPH). When compared to structural homologs where both of these conformations have been observed, such as PHHY²⁰ and RebC^{21,22}, the FAD in DGGR is similar to the IN conformation (Fig. 3). In this conformation, the isoalloxazine ring is inaccessible to solvent. Thus, we expect that DGGR adopts this conserved mechanism of switching FAD conformations during the two half reactions.

Interaction with bacterial lipid

Although DGGR was prepared without addition of any substrate, electron density in the active site clearly indicated the presence of a lipid-like ligand, which we assigned as a phosphatidylglycerol (PGX, Fig. 4a). The exact identity of this bacterial lipid is unknown as the density for the head group and the tails of the hydrocarbon chains are poorly defined

(Fig. 4a) and may represent an average of bacterial lipids of variable lengths (and head groups). Two double bonds were tentatively modeled at C9 and C6 of each lipid chain based on the presence of kinks in the hydrocarbon chains. These bacterial lipids are bound to DGGR, likely as a surrogate ligand, due to the resemblance of their chemical properties and shape to archaeal lipids. The presence of foreign lipids acquired from bacterial expression hosts was previously reported for other proteins, for example, CD1²³ and MD2²⁴. It is not clear whether DGGR can reduce unsaturated bacterial lipids. In bacteria, double bonds at specific locations of hydrocarbon chains are introduced into saturated lipids by desaturases²⁵, which are unrelated to DGGR. The average B-value of PGX (56 Å²) is higher than the protein (28 Å²) and FAD (14 Å²). This increased thermal motion reflects the imperfect fit of the bacterial lipid to the substrate-binding pocket, which appears to be optimized for a bulkier archaeal compound.

PGX interacts extensively with the protein and the isoalloxazine ring of FAD through hydrophobic interactions (Fig. 4a). Both hydrocarbon chains (A and B) are buried inside the protein and make contacts with residues from almost all secondary structure elements of the ligand-binding domain (Fig. 4b). Chains A and B are separated by Tyr209 and a loop region from the FAD-binding domain (residues 288-295). The most striking feature of the lipid is the near perfect stacking of C2-C9 hydrocarbons of chain A on the *re* face of the FAD (Fig. 4a). After C9, chain A takes a 90° turn perpendicular to the FAD ring. N5 of FAD is 3.4 Å from C8, and 3.5 Å from C7 of PGX. Similar parallel, substrate-isoalloxazine ring stacking has been observed in other flavin reductases, such as Old Yellow Enzyme (OYE)²⁶ and PHBH²⁷. The positioning of the reactive portion of the substrate with respect to the cofactor is generally conserved in flavoenzymes, suggesting the importance of an exact alignment of the substrate and FAD²⁸.

The head group of PGX is located near the loop prior to αM and does not make specific interactions with protein side chains. The phosphoryl group of PGX sits on top of the N-terminus of αM and is likely stabilized by the helix dipole. Lys338, which is conserved in *A. fulgidus* and *S. acidocaldarius*, is also close to the phosphoryl group (3.4 Å). Due to differences in stereochemistry of the glycerol backbone, interaction with glycerol might be different for endogenous archaeal lipids.

Properties of lipid recognition cavity

The lipid-binding cavity of DGGR is sandwiched between the β-sheet (strands 6-5-13-14-15-12-16) of the ligand-binding domain and the helical subdomain consisting of αK, αL and αM (Fig. 4c). This cavity has a distinctive R-shape with two tunnels and openings for the head group and for solvent. Notably, the cavity spans the entire substrate-binding domain (Fig. 4c). Additionally, the FAD and the FAD-binding domain also contribute to formation of the substrate-binding site. The presence of a two-prolonged cavity explains the preference of DGGR for digeranylgeranylglycerophospholipids over substrates with a single geranylgeranyl chain⁷. Plant polyamine oxidase (PAO) has the same PHBH fold as DGGR (PDB code 1b37, Z=14), but it binds a linear substrate²⁹. PAO contains a 30 Å U-shaped tunnel that runs parallel to the C4α-C10 axis of FAD. As a result, the orientation of the PAO and DGGR substrate tunnels are approximately perpendicular to each other.

The cavity for the head group is located next to the region where helix αK is kinked. The structural elements that contribute to the cavity opening include the middle portion of αK (residues 335-342), the N-terminus of αM and the preceding loop (residues 372-375), β3 (residues 47-52) and the αI-αJ loop (residues 289-298). Residues from these regions (such as Ser49, Phe199, Asp289, Thr292, Lys338, Ser373, and Val374) contribute formation of a largely hydrophilic pocket with a volume of ~400 Å³, which appears to be large enough to

accommodate different head groups⁷⁻⁹. Interactions between the protein and the head group may limit the accessibility of the allylic double bond. DGGR from *S. acidocaldarius* only reduces the last three double bonds when geranylgeranyl diphosphate is a substrate⁷. The allylic double bond is also not reduced by GGRs from plants and cyanobacteria. The shape and chemical properties of plant/bacterial GGRs substrate head groups are dramatically different from that of DGGR. Consequently, residues forming the cavity for the head group region are not well conserved between archaeal and plants or bacteria. Weak recognition for the head group might be a feature of DGGR so that it is not limited to specific head groups. Furthermore, a strong interaction with the head group would favor reduction of double bonds close to the head group, which appears to conflict with the observation that DGGR can reduce all of the double bonds of DGGGP.

The two lipid binding tunnels (A and B) are separated by Tyr209 from β 14 on one side and α I- α J loop (residue 287-295) on the other. The Tyr209 OH is within hydrogen bond distance of Thr292 O (2.5 Å). The geometry and chemical properties of the two tunnels are very different (Fig. 4c). Tunnel A is narrow and consists of two sections. The first section is sandwiched between the FAD ring and Tyr209, with additional contributions by Phe199, Trp211, Phe213 and a glycine-rich region (Thr292, Gly293 and Gly294). This section is the most restrictive in shape. The second section is more spacious and runs perpendicular to the FAD ring. The turn between the two sections above is defined by Tyr209 and Trp211 and residues with small side chains (Ala182, Gly225, and Val264). The second section is initially hydrophobic (Ile180, Gly208, and Val266), but becomes more hydrophilic at the end and accessible to solvent, with the presence of several water molecules. Additionally, weak density that appears to be an extension to chain A of PGX is also observed. Preliminary modeling studies suggest that when the substrate is fully inserted (during hydrogenation of the double bond at position 2), this tail portion of the narrow tunnel will be occupied by the hydrocarbon tail of the substrate. Tunnel B is significantly larger than either section of tunnel A and is formed exclusively by hydrophobic residues (Ala72, Ile74, Ile83, Leu85, Phe199, Leu201, Tyr209, Ile291, Leu349, Leu360, Val361, Val364, Ile377, Leu378, and Ile381). No waters were observed in this tunnel. B-values of the PGX hydrocarbon chains correlate well with the size of each tunnel. Chain A (18 carbons) has an average B-value of 39 Å², compared to 56 Å² for chain B (13 carbons). The most ordered portion C3-C9 of chain A (average B-value 31 Å²) resides in the most restrictive part of the tunnel, i.e. first section.

The arrangement and properties of the substrate binding site provide a structural framework for understanding the hydrogenation process carried out by DGGR. First, the larger tunnel (B) is more permissive, and could accommodate different types of hydrocarbon chains (both geranylgeranyl and phytanyl). The larger tunnel may favor, larger volume, hydrogenated chains. On the other hand, the tunnel section that is next to the isoalloxazine ring is very restricted in shape. This tunnel is likely to have a preference for the unsaturated chain. The portion of the tunnel immediately adjacent to the isoalloxazine ring mandates that the substrate be planar and parallel to the isoalloxazine ring, and clearly favors recognition of an unsaturated moiety (also see below). These factors are important for attaining the correct orientation of the substrate in the active site such that the unsaturated double bond is always properly aligned with respect to FAD during the multi-step hydrogenation process.

A conserved sequence motif for substrate specificity in geranylgeranyl reductases

GGRs are also widely present in bacteria and plants^{10,30-33}. Bacterial GGRs are predominately found in actinobacteria, cyanobacteria and proteobacteria. GGRs from cyanobacteria are closely related to plant GGRs, due to their identical function in the biosynthesis of photosynthetic pigments. Interestingly, some bacteria, such as actinobacteria, which have no known photosynthetic members, also contain GGR homologs. DGGR

homologs are present in all archaea except for the parasitic *Nanoarchaeum equitans*, which lacks genes for lipid, cofactor, amino acid, or nucleotide biosyntheses (see Table S2). In fact, most archaea, especially those from *Euryarchaeota*, contain multiple DGGR homologs. The number of homologs in archaea may correlate with variations of core lipids of the host. We analyzed the sequence-structure relationship of the geranylgeranyl reductases in order to obtain further insights into DGGR function. A multiple sequence alignment indicated that these homologous proteins share highly conserved residues for FAD and substrate interactions (Fig. 5a-b and Fig. S3).

All GGRs contain nucleotide binding motifs, GxGxxGx₁₇E/D (residues 10-33) and LxGD (residues 280-283), which are conserved in other members of the PHBH family³⁴. Other GGR-specific sequence motifs are located near the isoalloxazine ring (Fig. 5b). The RxxFD motif (residue 100-104), located at the N-terminus of helix α C, is opposite to the LxGD motif. The aspartate carboxyl hydrogen bonds with the guanidinium group of arginine. Arg100 NH1 interacts with O4' (3.0 Å) and O2' (3.1 Å) of the FAD. These two motifs close to the ribityl group may play a role in FAD's IN/OUT conformational switch. The opposite placement of RxxFD and LxGD motifs does not obstruct the IN/OUT conformational changes in FAD when the isoalloxazine ring swings out sideways. Another GGR-specific FAD binding motif is a short glycine-rich peptide that starts with a strictly conserved cysteine, CGGG (residues 44-47), which directly interacts with the *si* face of the isoalloxazine ring and is close to the RxxFD motif (3.3 Å between Arg100 NH1 and Cys44 SG). The distance between Cys44 SG and C4 α of FAD is 3.5 Å. The CGGG motif also interacts with the pyrimidine ring via multiple hydrogen bonds through main-chain atoms. We surveyed the structural database for a motif similar to (⁴³RCGEG⁴⁸). Archaeal L-proline dehydrogenase³⁵ (PDB code 1y56) catalytic β -subunit has the same fold (Z=22.5) and contains a highly similar region (⁴⁶RCGTG⁵¹) with cysteine in almost the identical conformation with respect to the isoalloxazine ring. The role of this conserved cysteine is unclear. In the L-proline dehydrogenase heterodimer, it participates in the electron transfer pathway between FMN and FAD. Thus, we cannot rule out the possibility that the cysteine may be involved in electron transfer or modulating the reactivity of flavin.

Near the *re* face of the isoalloxazine ring, a sequence motif ²⁰⁶PxxYxWxFP²¹⁴ unique to GGR family was identified. The region forms a β -strand (β 14) with three hydrophobic residues (Tyr209, Trp211 and Phe213) pointing towards the FAD. As described above, these residues defined the most restrictive portion of the substrate binding tunnels, together with the flavin ring and a glycine-rich loop (residue 292-295). These residues are arranged such that two surfaces of the cavity that are parallel to the isoalloxazine ring are planar (between Tyr209 and the isoalloxazine ring) (Fig. 4c). An additional wedge-like space is created by the Trp211 and Phe213 side chains (Fig. 5c). As a result, a cavity that is clearly complementary in shape to that of a geranyl group is formed near the reactive portion of FAD (Fig. 5c). Furthermore, this cavity is too small to accommodate the hydrogenated product. As a result, the PxxYxWxFP motif is critical for correct alignment of the geranyl group with respect to the isoalloxazine ring and, thus, is important for substrate specificity of GGRs. Interestingly, some GGR homologs in archaea and bacteria contain a phenylalanine instead of tyrosine at position 209 (Fig. 5d). None of the homologs that contain phenylalanine have been characterized, so the biochemical significance of the residue change is unknown.

Substrate entry and release

Substrates for PHBH family enzymes vary greatly in their size, flexibility, and chemical properties. For PHHY and PHBH, whose substrates are relatively small, access to the active site is controlled by conformational changes that allow entry through a tunnel passing the *re* face of the flavin ring^{20,36} ("front side"). For RebC, whose substrate is rigid but

significantly larger, a path for substrate entry/product release has been proposed²¹ involving a disorder/order transition of a peptide stretch at the “backside” of the substrate pocket. Folding of the disordered stretch into an α -helix after binding of substrate closes off the pocket and seals the ligand in its binding site. The DGGR structure suggests that a RebC-like mechanism is likely the case.

DGGR substrates are both bulky and flexible, thus they are more challenging to bind. The only large opening is at the back of the protein and is partially occupied by the head group of PGX (Fig. 6a). Therefore, substrate in DGGR most likely enters the active site through this opening. Backside substrate entry and product release appears to be a common feature in PHBH-like flavoenzymes that process large substrates, such as PAO²⁹, RdmE³⁷, and RebC²¹. In the FAD IN conformation, as observed here, tunnel A is not accessible to the geranylgeranyl hydrocarbon chain due to tight geometric restrictions resulting from the arrangement of FAD, the conserved specificity motif on β 14 and the α I- α J loop (Fig. 6b). Thus, the FAD IN conformation is not conducive to substrate entry. In order to open tunnel A up for binding, one or both of the following need to occur. First, FAD must be switched to the OUT conformation. Second, the hydrogen-bond interaction between Tyr209 and Thr292 must be disrupted through conformational changes in the glycine-rich α I- α J loop. Since the α I- α J loop and FAD interact with each other, their conformational changes are likely coupled. Therefore, it appears that substrate entry in DGGR is likely coupled with structural changes induced by the conformational switch of the FAD. Coupling binding of “fresh” substrate with reduction of the FAD will avoid wasteful reduction cycles (i.e. FAD being reduced with no substrate in the active site)²¹.

Analysis of thermal motion along the protein backbone also suggests that the C-terminal helical subdomain may be involved in substrate binding/release. Two regions, the β 6- β 7 loop (residues 87-94) and the C-terminal helices (residues 370-396), have relatively higher B-values. Both regions are located near the backside entrance (Fig. 6c). As these regions are not involved in crystal packing, higher B-values likely reflect some inherent conformational flexibility. Helices α M and α N are held in place by weak interactions with PGX and other protein residues, and are expected to be more flexible in the absence of substrate. If these helices were to disassociate from the rest of the protein, the entrance would expand to facilitate substrate binding and product release (Fig. 6b). This suggests that the catalytic efficiency of GGRs may be dependent on protein dynamics.

Mechanistic implications

In order to better understand the catalytic mechanism, we modeled a portion of a geranyl group into the active site based on strict geometrical constraints, assuming the protein structure determined here closely mimics an active complex. This docking was achieved by shifting the geranyl group horizontally along the bacterial lipid such that it minimized steric clashes. Indeed, only one possibility for the substrate satisfies this protein environment, which is to fit the methyl group of the substrate into the cleft defined by Trp211 and Phe213 (Fig. 7a). The resultant model sandwiches the substrate between Tyr209 and the isoalloxazine ring. The distance between C3 of the substrate and FAD N5 is 3.4 Å (Fig. 7b). The geometry of the reactive group with regard to FAD is consistent with what is generally expected from flavoenzymes²⁸.

The structural arrangement of the isoalloxazine ring, substrates and the conserved tyrosine (Tyr209) in DGGR bear some resemblance to that of OYE, a well-studied flavin reductase catalyzing the reduction of α,β -unsaturated aldehydes and ketones (Fig. S4a)³⁸. In OYE, the addition of H₂ occurs in a *trans*-fashion where a hydride ion is transferred from the N5 atom of FMN to C β of the substrate while C α accepts a proton from Tyr196³⁹. The reaction catalyzed by DGGR differs from OYE. First, the C α of OYE substrate is bonded to an

electron withdrawing group (EWG) that helps to activate the substrate, while the DGGR substrate is not activated. Second, H₂-addition proceeds in *syn*-manner in DGGR⁶. If Tyr209 were to function as the proton donor, then a *trans*-addition is expected. Additionally, Tyr209 OH is hydrogen bonded to Thr292. Thus, Tyr209 cannot function as a proton donor in DGGR.

Precise alignment and the π stacking between flavin and substrate appear to be the most important factors in the hydrogenation reaction. Tyr209 is one of the critical residues for substrate specificity. However, it is not clear whether Tyr209 has other roles in catalysis. Besides Tyr209, no other polar residues are near the catalytic center. Mechanisms for many flavin enzymes remain controversial. In many cases, a hydride transfer mechanism seems to be most consistent with experimental observations^{28,38,40}. It is possible that DGGR also adopts a hydride transfer mechanism, such that a hydride ion is transferred from the N5 atom of FAD to C3 (or C2) of the substrate while a proton is extracted by C2 (or C3) from a general acid (Fig. S5). The identity of the proton donor is currently unknown. The most likely candidate is N1 of FAD, which is closest to C2 (~4.6 Å). The nearest water to C2 in the crystal structure is ~7.5 Å away. However, it remains possible that a more labile water molecule not seen in the structure could serve as the proton donor.

Another intriguing and unique aspect of the DGGR mechanism is how DGGR catalyze hydrogenation of double bonds in multiple sites of a DGGGP lipid^{7,9} (Fig. 1b). The starting substrate DGGGP must undergo hydrogenation 8 times where, at each ensuing stage, the substrate is different (with addition of two hydrogens) from the previous cycle (substrate recycling). At least two key questions remain to be addressed. First, are the double bonds hydrogenated randomly or in a certain order? Second, does the substrate evacuate the active site chamber completely after each reduction, or remain bound until one lipid chain is fully hydrogenated? The crystal structure reported here offers interesting clues on how these consecutive reduction steps could be accomplished. However, further experiments will be needed to clarify the detailed mechanism of DGGR.

Material and methods

Protein cloning and expression

Clones were generated using the Polymerase Incomplete Primer Extension (PIPE) cloning method (Klock et al, 2008). The gene encoding DGGR (Swiss-Prot: Q9HKS9) was amplified by the polymerase chain reaction (PCR) from *T. acidophilum* DSM1728 genomic DNA using PfuTurbo DNA polymerase (Stratagene) and I-PIPE (Insert) primers (forward primer: 5'-ctgtacttcagggcATGGAAACCTATGATGTTCTTGTGGTTG-3', reverse primer: 5'-ctgtacttcagggcATGGAAACCTATGATGTTCTTGTGGTTG-3', target sequence in upper case) that included sequences for the predicted 5' and 3' ends. The expression vector, pSpeedET, which encodes an amino-terminal tobacco etch virus (TEV) protease-cleavable expression and purification tag (MGSDKIHSHHHENLYFQ/G), was PCR amplified with V-PIPE (Vector) primers (forward primer: 5'-taacgcgacttaattaactgtttaaacggtctccagc-3', reverse primer: 5'-gccctggaagtacaggttttcgtgatgatgatgatg-3'). V-PIPE and I-PIPE PCR products were mixed to anneal the amplified DNA fragments together. *E. coli* GeneHogs (Invitrogen) competent cells were transformed with the V-PIPE / I-PIPE mixture and dispensed on selective LB-agar plates. The cloning junctions were confirmed by DNA sequencing. Expression was performed in a selenomethionine-containing medium with suppression of normal methionine synthesis. At the end of fermentation, lysozyme was added to the culture to a final concentration of 250 µg/ml, and the cells were harvested and frozen. After one freeze/thaw cycle, the cells were homogenized in lysis buffer [50 mM HEPES pH 8.0, 50 mM NaCl, 10mM imidazole, 1 mM Tris(2-carboxyethyl)phosphine-HCl (TCEP)] and the lysate was clarified by centrifugation at 32,500 x g for 30 min. The soluble

fraction was passed over nickel-chelating resin (GE Healthcare) pre-equilibrated with lysis buffer, the resin washed with wash buffer [50 mM HEPES pH 8.0, 300 mM NaCl, 40 mM imidazole, 10% (v/v) glycerol, 1 mM TCEP], and the protein eluted with elution buffer [20 mM HEPES pH 8.0, 300 mM imidazole, 10% (v/v) glycerol, 1 mM TCEP]. The eluate was buffer exchanged with HEPES crystallization buffer [20 mM HEPES pH 8.0, 200 mM NaCl, 40 mM imidazole, 1 mM TCEP] using a PD-10 column (GE Healthcare), and incubated with 1 mg of TEV protease per 15 mg of eluted protein. The protease-treated eluate was passed over nickel-chelating resin (GE Healthcare) pre-equilibrated with HEPES crystallization buffer and the resin was washed with the same buffer. The flow-through and wash fractions were combined and concentrated for crystallization trials to 9.9 mg/ml by centrifugal ultrafiltration (Millipore).

Oligomeric state

The oligomeric state of DGGR in solution was determined using a $1 \times 30\text{cm}^2$ Superdex 200 column (GE Healthcare) coupled with miniDAWN static light scattering (SEC/SLS) and Optilab differential refractive index detectors (Wyatt Technology). The mobile phase consisted of 20mM Tris pH 8.0, 150mM NaCl, and 0.02% (w/v) sodium azide. The molecular weight was calculated using ASTRA 5.1.5 software (Wyatt Technology).

Crystallization

DGGR was crystallized by mixing 200 nl protein with 200 nl crystallization solution over a 50 μ l reservoir volume using the nanodroplet vapor diffusion method⁴¹ with standard JCSG crystallization protocols¹³. The crystallization reagent that produced the DGGR crystal used for structure determination consisted of 15% glycerol, 8.5% isopropanol, 17% PEG-4000, and 0.1M HEPES pH 7.5. A yellow-colored, rectangle crystal of approximate size 90 $\mu\text{m} \times 40 \mu\text{m} \times 30 \mu\text{m}$ was harvested after 20 days at 277K. Ethylene glycol was added to the crystal as a cryoprotectant to a final concentration of 10% (v/v). Initial screening for diffraction was carried out using the Stanford Automated Mounting system (SAM)⁴² at the Stanford Synchrotron Radiation Lightsource (SSRL, Menlo Park, CA).

Data collection, structure solution, and refinement

Multi-wavelength anomalous diffraction (MAD) data were collected at SSRL beamline 11-1. Data were collected at wavelengths corresponding to the inflection, high energy remote, and peak of a selenium MAD experiment at 100 K using Mar CCD 325 detector (Rayonix). Data processing and structure solution were carried out using a structure solution program developed at the JCSG. This script carries out data processing, phasing and initial model building in a fully automatic manner by making uses of third party programs. In summary, the MAD data were integrated and reduced using XDS and then scaled with the program XSCALE⁴³. One sweep data were processed initially in the space group P1, which allowed identification of the Laue group and unit cell parameters that would be subsequently used to process all data. Location of selenium sites, initial phasing, and identification of the space group were carried out using SHELXD and SHELXE⁴⁴. Initial number of heavy atom sites was inferred assuming a solvent content of ~50%. SHELXD was run in parallel for all possible space groups and the best solution was identified in the space group P2₁2₁2₁ at 2.09 Å [CC/CC(weak)=0.56/0.39]. The heavy atom solution was further improved using SHELXE (CC=0.78). Phase refinement and model building were performed using autoSHARP⁴⁵ and wARP⁴⁶. The above process produced excellent density maps and resulted in an initial model that was ~88% complete. Further model completion and refinement were performed manually with COOT⁴⁷ and REFMAC5⁴⁸ of the CCP4 suite⁴⁹. The refinement included experimental phase restraints in the form of Hendrickson-Lattman coefficients and TLS refinement with 1 TLS group that included the protein and the bound ligands (FAD and PGX). Data and refinement statistics are summarized in Tables 1 and 2.

Analysis of the stereochemical quality of the model was accomplished using MolProbity¹⁵. Sequence logos were produced by WEBLOGO⁵⁰. Molecular graphics were prepared with PyMOL (DeLano Scientific).

Accession code

Atomic coordinates and experimental structure factors for DGGR at 1.6 Å resolution have been deposited in the PDB under accession code 3oz2.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Polat Abdubek, Tamara Astakhova, Herbert L. Axelrod, Dennis Carlton, Connie Chen, Thomas Clayton, Debanu Das, Marc C. Deller, Lian Duan, Kyle Ellrott, Joanna C. Grant, Anna Grzechnik, Gye Won Han, Kevin K. Jin, Piotr Kozbial, S. Sri Krishna, Abhinav Kumar, Winnie W. Lam, David Marciano, Andrew T. Morse, Edward Nigoghossian, Amanda Nopakun, Linda Okach, Christina Puckett, Ron Reyes, Henry J. Tien, Christine B. Trame and Henry van den Bedem of the JCSG for their contributions to the protein production and/or structural work and analysis. The JCSG is supported by the NIH, National Institutes of General Medical Sciences, Protein Structure Initiative (U54 GM074898). Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource (SSRL). The SSRL is a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health (National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences). Genomic DNA from *T. acidophilum* DSM1728 (ATCC Number: 25905D) was obtained from the American Type Culture Collection (ATCC). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

References

1. Koga Y, Morii H. Biosynthesis of ether-type polar lipids in archaea and evolutionary considerations. *Microbiol. Mol. Biol. Rev.* 2007; 71:97–120. [PubMed: 17347520]
2. De Rosa M, Gambacorta A. The lipids of archaebacteria. *Prog. Lipid Res.* 1988; 27:153–175. [PubMed: 3151021]
3. Kuzuyama T, Seto H. Diversity of the biosynthesis of the isoprene units. *Nat. Prod. Rep.* 2003; 20:171–183. [PubMed: 12735695]
4. Boucher Y, Kamekura M, Doolittle WF. Origins and evolution of isoprenoid lipid biosynthesis in archaea. *Mol. Microbiol.* 2004; 52:515–527. [PubMed: 15066037]
5. van de Vossenberg JL, Driessen AJ, Konings WN. The essence of being extremophilic: the role of the unique archaeal membrane lipids. *Extremophiles.* 1998; 2:163–170. [PubMed: 9783161]
6. Nishimura Y, Eguchi T. Stereochemistry of reduction in digeranylgeranyl glycerophospholipid reductase involved in the biosynthesis of archaeal membrane lipids from *Thermoplasma acidophilum*. *Bioorg. Chem.* 2007; 35:276–283. [PubMed: 17275067]
7. Nishimura Y, Eguchi T. Biosynthesis of archaeal membrane lipids: digeranylgeranyl glycerophospholipid reductase of the thermoacidophilic archaeon *Thermoplasma acidophilum*. *J. Biochem.* 2006; 139:1073–1081. [PubMed: 16788058]
8. Murakami M, Shibuya K, Nakayama T, Nishino T, Yoshimura T, Hemmi H. Geranylgeranyl reductase involved in the biosynthesis of archaeal membrane lipids in the hyperthermophilic archaeon *Archaeoglobus fulgidus*. *FEBS J.* 2007; 274:805–814. [PubMed: 17288560]
9. Sato S, Murakami M, Yoshimura T, Hemmi H. Specific partial reduction of geranylgeranyl diphosphate by an enzyme from the thermoacidophilic archaeon *Sulfolobus acidocaldarius* yields a reactive prenyl donor, not a dead-end product. *J. Bacteriol.* 2008; 190:3923–3929. [PubMed: 18375567]

10. Giannino D, Condello E, Bruno L, Testone G, Tartarini A, Cozza R, et al. The gene geranylgeranyl reductase of peach (*Prunus persica* [L.] Batsch) is regulated during leaf development and responds differentially to distinct stress factors. *J. Exp. Bot.* 2004; 55:2063–2073. [PubMed: 15286145]
11. Addlesee HA, Gibson LC, Jensen PE, Hunter CN. Cloning, sequencing and functional assignment of the chlorophyll biosynthesis gene, chlP, of *Synechocystis* sp. *PCC 6803*. *FEBS Lett.* 1996; 389:126–130. [PubMed: 8766814]
12. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. InterPro: the integrative protein signature database. *Nucleic Acids Res.* 2009; 37:D211–215. [PubMed: 18940856]
13. Lesley SA, Kuhn P, Godzik A, Deacon AM, Mathews I, Kreusch A, et al. Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. *Proc. Natl. Acad. Sci. USA.* 2002; 99:11664–11669. [PubMed: 12193646]
14. Cruickshank DW. Remarks about protein structure precision. *Acta Cryst. D.* 1999; 55:583–601. [PubMed: 10089455]
15. Davis IW, Murray LW, Richardson JS, Richardson DC. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* 2004; 32:W615–619. [PubMed: 15215462]
16. Holm L, Kaariainen S, Rosenstrom P, Schenkel A. Searching protein structure databases with DaliLite v.3. *Bioinformatics.* 2008; 24:2780–2781. [PubMed: 18818215]
17. Buedenbender S, Rachid S, Muller R, Schulz GE. Structure and action of the myxobacterial chondrochloren halogenase CndH: a new variant of FAD-dependent halogenases. *J. Mol. Biol.* 2009; 385:520–530. [PubMed: 19000696]
18. Lennon BW, Williams CH Jr, Ludwig ML. Crystal structure of reduced thioredoxin reductase from *Escherichia coli*: structural flexibility in the isoalloxazine ring of the flavin adenine dinucleotide cofactor. *Protein Sci.* 1999; 8:2366–2379. [PubMed: 10595539]
19. Rohr AK, Hersleth HP, Andersson KK. Tracking flavin conformations in protein crystal structures with Raman spectroscopy and QM/MM calculations. *Angew. Chem. Int. Ed. Engl.* 2010; 49:2324–2327. [PubMed: 20187055]
20. Enroth C, Neujahr H, Schneider G, Lindqvist Y. The crystal structure of phenol hydroxylase in complex with FAD and phenol provides evidence for a concerted conformational change in the enzyme and its cofactor during catalysis. *Structure.* 1998; 6:605–617. [PubMed: 9634698]
21. Ryan KS, Chakraborty S, Howard-Jones AR, Walsh CT, Ballou DP, Drennan CL. The FAD cofactor of RebC shifts to an IN conformation upon flavin reduction. *Biochemistry.* 2008; 47:13506–13513. [PubMed: 19035832]
22. Ryan KS, Howard-Jones AR, Hamill MJ, Elliott SJ, Walsh CT, Drennan CL. Crystallographic trapping in the rebeccamycin biosynthetic enzyme RebC. *Proc. Natl. Acad. Sci. U. S. A.* 2007; 104:15311–15316. [PubMed: 17873060]
23. Zajonc DM, Striegl H, Dascher CC, Wilson IA. The crystal structure of avian CD1 reveals a smaller, more primordial antigen-binding pocket compared to mammalian CD1. *Proc. Natl. Acad. Sci. U. S. A.* 2008; 105:17925–17930. [PubMed: 19004781]
24. Yoon SI, Hong M, Han GW, Wilson IA. Crystal structure of soluble MD-1 and its interaction with lipid IVa. *Proc. Natl. Acad. Sci. U. S. A.* 2010; 107:10990–10995. [PubMed: 20534476]
25. Buist PH. Fatty acid desaturases: selecting the dehydrogenation channel. *Nat. Prod. Rep.* 2004; 21:249–262. [PubMed: 15042148]
26. Fox KM, Karplus PA. Old yellow enzyme at 2 Å resolution: overall structure, ligand binding, and comparison with related flavoproteins. *Structure.* 1994; 2:1089–1105. [PubMed: 7881908]
27. Schreuder HA, Prick PA, Wierenga RK, Vriend G, Wilson KS, Hol WG, et al. Crystal structure of the *p*-hydroxybenzoate hydroxylase-substrate complex refined at 1.9 Å resolution. Analysis of the enzyme-substrate and enzyme-product complexes. *J. Mol. Biol.* 1989; 208:679–696. [PubMed: 2553983]
28. Fraaije MW, Mattevi A. Flavoenzymes: diverse catalysts with recurrent features. *Trends Biochem. Sci.* 2000; 25:126–132. [PubMed: 10694883]

29. Binda C, Coda A, Angelini R, Federico R, Ascenzi P, Mattevi A. A 30 Å-long U-shaped catalytic tunnel in the crystal structure of polyamine oxidase. *Structure*. 1999; 7:265–276. [PubMed: 10368296]
30. Shpilyov AV, Zinchenko VV, Shestakov SV, Grimm B, Lokstein H. Inactivation of the geranylgeranyl reductase (ChlP) gene in the cyanobacterium *Synechocystis* sp. *PCC 6803*. *Biochim. Biophys. Acta*. 2005; 1706:195–203. [PubMed: 15694347]
31. Gomez Maqueo Chew A, Frigaard NU, Bryant DA. Identification of the bchP gene, encoding geranylgeranyl reductase in *Chlorobaculum tepidum*. *J. Bacteriol.* 2008; 190:747–749. [PubMed: 17993528]
32. Keller Y, Bouvier F, d'Harlingue A, Camara B. Metabolic compartmentation of plastid prenyllipid biosynthesis—evidence for the involvement of a multifunctional geranylgeranyl reductase. *Eur. J. Biochem.* 1998; 251:413–417. [PubMed: 9492312]
33. Addlesee HA, Hunter CN. Physical mapping and functional assignment of the geranylgeranyl-bacteriochlorophyll reductase gene, bchP, of *Rhodobacter sphaeroides*. *J. Bacteriol.* 1999; 181:7248–7255. [PubMed: 10572128]
34. Dym O, Eisenberg D. Sequence-structure analysis of FAD-containing proteins. *Protein Sci.* 2001; 10:1712–1728. [PubMed: 11514662]
35. Tsuge H, Kawakami R, Sakuraba H, Ago H, Miyano M, Aki K, et al. Crystal structure of a novel FAD-, FMN-, and ATP-containing L-proline dehydrogenase complex from *Pyrococcus horikoshii*. *J. Biol. Chem.* 2005; 280:31045–31049. [PubMed: 16027125]
36. Wang J, Ortiz-Maldonado M, Entsch B, Massey V, Ballou D, Gatti DL. Protein and ligand dynamics in 4-hydroxybenzoate hydroxylase. *Proc. Natl. Acad. Sci. U. S. A.* 2002; 99:608–613. [PubMed: 11805318]
37. Lindqvist Y, Koskiniemi H, Jansson A, Sandalova T, Schnell R, Liu Z, et al. Structural basis for substrate recognition and specificity in aklavinone-11-hydroxylase from rhodomycin biosynthesis. *J. Mol. Biol.* 2009; 393:966–977. [PubMed: 19744497]
38. Massey V. The chemical and biological versatility of riboflavin. *Biochem. Soc. Trans.* 2000; 28:283–296. [PubMed: 10961912]
39. Basran J, Harris RJ, Sutcliffe MJ, Scrutton NS. H-tunneling in the multiple H-transfers of the catalytic cycle of morphinone reductase and in the reductive half-reaction of the homologous pentaerythritol tetranitrate reductase. *J. Biol. Chem.* 2003; 278:43973–43982. [PubMed: 12941965]
40. Fitzpatrick PF. Oxidation of amines by flavoproteins. *Arch. Biochem. Biophys.* 2009; 493:13–25. [PubMed: 19651103]
41. Santarsiero BD, Yegian DT, Lee CC, Spraggon G, Gu J, Scheibe D, et al. An approach to rapid protein crystallization using nanodroplets. *J. Appl. Cryst.* 2002; 35:278–281.
42. Cohen AE, Ellis PJ, Miller MD, Deacon AM, Phizackerley RP. An automated system to mount cryo-cooled protein crystals on a synchrotron beamline, using compact samples cassettes and a small-scale robot. *J. Appl. Cryst.* 2002; 35:720–726.
43. Kabsch W. XDS. *Acta Cryst. D.* 2010; 66:125–132. [PubMed: 20124692]
44. Sheldrick GM. A short history of SHELX. *Acta Cryst. A.* 2008; 64:112–122. [PubMed: 18156677]
45. Bricogne G, Vonrhein C, Flensburg C, Schiltz M, Paciorek W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Cryst. D.* 2003; 59:2023–2030. [PubMed: 14573958]
46. Cohen SX, Morris RJ, Fernandez FJ, Ben Jelloul M, Kakaris M, Parthasarathy V, et al. Towards complete validated models in the next generation of ARP/wARP. *Acta Cryst. D.* 2004; 60:2222–2229. [PubMed: 15572775]
47. Emsley P, Cowtan K. COOT: model-building tools for molecular graphics. *Acta Cryst. D.* 2004; 60:2126–2132. [PubMed: 15572765]
48. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst. D.* 1997; 53:240–255. [PubMed: 15299926]
49. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Cryst. D.* 1994; 50:760–763. [PubMed: 15299374]

50. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res.* 2004; 14:1188–1190. [PubMed: 15173120]

Fig. 1. Archaeal lipid and chemical reaction catalyzed by DGGR. (a). Chemical structures of archaeal core lipid and a representative bacterial or eukaryotic polar lipids. X represents a head group or hydrogen. (b). Hydrogenation reaction catalyzed by archaeal reductase DGGR.

Fig. 2.

Crystal structure of DGGR from *T. acidophilum*. (a). Stereo ribbon diagram of the structure of DGGR. The structure is colored by functional domains (cyan: FAD binding, and magenta: substrate binding). Helices are labeled through A to N, strands 1 to 18. FAD and a bacterial phosphatidylglycerol lipid (PGX) are shown as ball-and-sticks. (b). Protein sequence of *T. acidophilum* DGGR annotated with the secondary structure elements.

Fig. 3. Comparisons of the FAD conformation in DGGR (yellow) to the IN (cyan)/ OUT (gray) conformations in representative PHBH proteins: PHHY (PDB code 1foh, left) and RebC (PDB codes 2r0c (OUT) and 3ept (IN), right).

Fig. 4. Lipid-protein interactions and lipid binding site. (a). Bacterial phosphatidyl lipid (PGX) found in the active site. Fo-Fc omit map of FAD and PGX is contoured at 3.0σ . A schematic diagram of the chemical structure of PGX and its protein environment is shown on the right. (b). A stereoview of the interaction between PGX (cyan/red), FAD (yellow/orange/red/blue) and DGGR protein. Protein residues within 4.5 Å of PGX are colored magenta. Three protein residues important for substrate-binding are shown in green/red. (c). Substrate binding cavity (shown as surface, green in left panel). Protein residues contributing to the formation of the cavity are colored magenta. The direction that the FAD swings from its IN to OUT conformation is indicated by a yellow arrow. The cavity for the head group is marked by an orange arrow. A close-up surface representation of the lipid binding cavity in left panel (colored green), viewed from the opposite side, is shown on the right. The surface color corresponds to underlying atoms (N: blue, O: red, C: white). The two internal tunnels are marked as A and B.

Fig. 5. Structural basis for recognition of a geranyl group by GGRs. (a). Highly conserved regions of characterized geranylgeranyl reductases from archaea (*T. acidophilum*, *S. acidocaldarius* and *A. fulgidus*), plants (*Prunus persica* and *Arabidopsis thaliana*), and bacteria (*Synechocystis* sp. and *C. tepidum*). Strictly conserved residues are colored yellow on purple, and highly conserved residues white on blue. (b). Structure of the conserved regions in the context of FAD and the active site. The specificity pocket defined by FAD (cyan), α I- α J loop (residues 288-295, green) and YxWxF motif (green) is shown as gray surface. Each sequence logo was generated using 541 unique sequences extracted from INTERPRO entry IPR011777. (c). $^{209}\text{YxWxF}^{213}$ motif and FAD define a pocket (gray surface) that is specific to the geranyl group (magenta) by complementarity in shape and chemical properties. (d). Variations of the YxWxF motif in bacteria (336 sequences), archaea (165 sequences) and eukaryotes (39 sequences). Tyr209 of DGGR is marked with a green arrow.

Fig. 6. Substrate entry and product release. (a). A close up view of the substrate entrance at the back side. FAD and PGX are shown as sticks (cyan). The surface color corresponds to underlying atoms (N: blue, O: red, C: white). (b). Local structural environment near the substrate entrance. α I- α J loop and the C-terminal portion of the protein are shown as ribbons (magenta) and rest of the protein as surface representation. (c). Plot of average main chain per residue B-values (top) and mapping of B-values to protein (bottom). The protein is colored in a gradient from blue (low) to red (high) B-values.

Fig. 7. Model and reaction mechanism. (a). A manually docked geranyl fragment (violet, sticks and dots) fits nicely in the active site. FAD and PGX are shown as thin yellow lines. (b). Geometry of the double bond of the modeled substrate (magenta) with respect to FAD and conserved polar residues (Cys44 and Tyr209, cyan). Distances between atoms are shown as dashed lines in Å.

Table 1

Data collection statistics

Space group	P2 ₁ 2 ₁ 2 ₁		
Unit cell	<i>a</i> =49.6 Å, <i>b</i> =70.7 Å, <i>c</i> =117.9 Å		
Data collection	λ ₁ MADSe - infl	λ ₂ MADSe - remo	λ ₃ MADSe - peak
Wavelength (Å)	0.9792	0.9184	0.9786
Max resolution (Å)	1.6	1.6	1.6
Observations	199,376	199,822	199,226
Unique reflections	55,130	55,071	55,119
Completeness (%)	99.6 (99.3) ^a	99.5 (99.3)	99.6 (99.4)
Mean I/σ(I)	15.9 (2.5) ^a	18.0 (2.9)	15.3 (2.0)
R _{merge} on I (%)	3.9 (44.2) ^a	3.6 (37.5)	4.2 (54.0)

$$R_{\text{merge}} = \frac{\sum_{\text{hkl}} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle|}{\sum_{\text{hkl}} \sum_i I_i(\text{hkl})}$$

^a Highest resolution shell in parentheses. The high resolution cutoff was chosen such that the mean I/σ(I) in the highest resolution shell (1.69-1.6 Å) is ~2.

Table 2

Phasing, refinement statistics

MAD Phasing	
Number of Se sites	16
Figure of merit to 1.6 Å	0.59
Model and refinement statistics	
Data set used in refinement	λ_2 MADSe
Resolution range (Å)	45-1.6
Cutoff criteria	$ F > 0$
R_{cryst}	15.3
R_{free}	17.1
ESU based on R_{free} (Å)	0.074
Non-hydrogen atoms	
Protein	3226
Ligands (FAD,PGX)	101
Solvent	392
Restraints (RMS observed)	
Bond lengths (Å)	0.014
Bond angles (°)	1.61
Average isotropic B-value (Å ²)	29.1
MolProbity statistics	
All-atom clash score	5.9
Ramachandran favored (%)	98.2
Ramachandran outliers (%)	0
Rotamer outliers (%)	1.6

$R_{\text{cryst}} = \frac{\sum |F_{\text{obs}}| - |F_{\text{calc}}|}{\sum |F_{\text{obs}}|}$ where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively.

R_{free} = as for R_{cryst} , but for 5% of the total reflections chosen at random and omitted from refinement.

ESU = Estimated Standard Uncertainty in coordinates.